FOI-D--0216--SE
Maj 2005

Johan Schubert, Per Svensson

Methodology for guaranteed and robust high level fusion performance: a literature study

Command and Control Systems
SE-581 11 Linköping

SWEDISH DEFENCE RESEARCH AGENCY
Ledningssystem
Box 1165
581 11 Linköping

FOI-D--0216--SE
Maj 2005

Johan Schubert, Per Svensson

Methodology for guaranteed and robust high level fusion performance: a literature study

Issuing organization	Report number, ISRN	
FOI – Swedish Defence Research Agency Command and Control Systems P.O. Box 1165 SE-581 11 Linköping	FOI-D--0216--SE	
	Programme Area 4 C4ISTAR	
	Month year Maj 2005	**Project no.** I7164
	Subcategories 41 C4I	
	Subcategories 2	
Author/s (editor/s) Johan Schubert Per Svensson	**Project manager** Per Svensson	
	Approved by	
	Sponsoring agency	
	Scientifically and technically responsible	

Report title (In translation)

Methodology for guaranteed and robust high level fusion performance: a literature study

Abstract (not more than 200 words)

A currently hot topic in Information Fusion research is fusion performance, including relevance, reliability, robustness, and trustworthiness, in particular for higher level fusion methods such as situation and impact refinement. Unless methodologies are found and generally applied which enable us to achieve and demonstrate trustworthiness, chances are slim that future commanders would be willing to trust high-level fusion systems.

On the other hand, quite extensive work has been performed in other fields during the last two decades, in particular in risk analysis and robust Bayesian analysis, which indicates that robust and trustworthy behaviour may in principle be achieved from probabilistic simulations and therefore from information fusion methods in decision support systems. It is likely, however, that complex systems which were not designed from the outset to satisfy specified trustworthiness and robustness criteria will be hard or impossible to validate later using this methodology.

Thus we believe that this partly formal research needs to be evaluated and, if found appropriate, adapted for use in information fusion systems design. Since it is not yet known for certain if or how these new approaches can be applied to information fusion algorithms and systems, we perceive an urgent need to start serious research and education in this area. We know that the answer to questions of reliability and trustworthiness will not be simple; instead, they will to a large degree have to evolve over time through scientific debate. We also know many situations where computer-supported military decision making works well, but we do not know how to deal with problems near the fluid boundary between those situations we know how to automate and those we believe must be reserved for human decision-making.

Keywords

Information fusion, performance evaluation, robustness

Further bibliographic information	Language English
	Pages 37 p.

1. Background

A currently hot topic in Information Fusion research is *fusion performance*, including relevance, reliability, robustness, and trustworthiness, in particular for higher level fusion methods such as situation and impact refinement. Unless methodologies are found and generally applied which enable us to achieve and demonstrate trustworthiness, chances are slim that future commanders would be willing to trust high-level fusion systems.

On the other hand, quite extensive work has been performed in other fields during the last two decades, in particular in risk analysis and robust Bayesian analysis, which indicates that robust and trustworthy behaviour may in principle be achieved from probabilistic simulations and therefore from information fusion methods in decision support systems. It is likely, however, that complex systems which were not designed from the outset to satisfy specified trustworthiness and robustness criteria will be hard or impossible to validate later using this methodology.

Thus we believe that this partly formal research needs to be evaluated and, if found appropriate, adapted for use in information fusion systems design. Since it is not yet known for certain if or how these new approaches can be applied to information fusion algorithms and systems, we perceive an urgent need to start serious research and education in this area. We know that the answer to questions of reliability and trustworthiness will not be simple; instead, they will to a large degree have to evolve over time through scientific debate. We also know many situations where computer-supported military decision making works well, but we do not know how to deal with problems near the fluid boundary between those situations we know how to automate and those we believe must be reserved for human decision-making.

Feynman's minority report

"Uncertainties appear everywhere! When using a mathematical model, careful attention must be given to uncertainties in the model."

--- Richard Feynman, 1989, Space Shuttle Challenger Inquiry "*Macroscopic approach to risk estimation*".

2. Introduction

The C4I system development methodology proposed by the Swedish Armed Forces has been evolutionary development in integrated product teams. Today, however, we see little application of that philosophy, and instead the responsibility for technology development has mainly been transferred to the industrial consortium in charge of developing the first prototypes, and the user/purchaser is represented only indirectly by the so-called LedsystT project, run by the Defence Materiel Administration. By pursuing in parallel a so-called methodology project (LedsystM), the Armed Forces hope to catch up with the LedsystT technology development and be able to specify in time what the user requirements are. The new methods which are to be studied in LedsystM are of course tactical and operative network-based processes, not technical ones.

Still one clings to the notion that technology development in LedsystT is to take place in flexibly specified phases followed by technology demonstrations after each phase. The posture originally advocated by the Swedish Armed Forces that technology demonstrators need to be forward-looking and therefore will have to involve considerable risk-taking is falling into the background, however. Of course, we who work with IF research in Sweden see a potential problem here. We claim that without adequate information fusion, investments in a dynamic and mobile digital communication network with high security, reliability, and capacity will be largely wasted, since users will then be able to absorb and act on only a small fraction of the information flowing in the network.

The only things we can analyze in depth are (more or less formal) models. Many people probably believe that they are able to analyze and discuss reality itself, but what we are able to analyze and discuss is in fact more or less well-specified, more or less common abstract models of situations which are "similar" to real events and phenomena. It is fundamentally important to understand that a model has to be created in order to make it possible to describe and analyze any phenomenon (abstracted from reality, let us assume) with sufficient "fidelity" and precision to reach a certain goal which may or may not be well specified in advance. A "model of the model" can often be implemented in software and/or hardware to get products for various purposes. Again, to evaluate the finished product there exists a number of different, more or less standardized problem formulations, technologies, and modelling methods. These evaluation technologies usually need to possess a higher degree of sophistication (higher fidelity, precision, etc.) than the final product in order to make the evaluation process credible. A number of other criteria also need to be satisfied, in particular, the availability of expertise in various related areas.

2.1. Reliability dimensions

Reliability and trustworthiness of a complex information fusion product presupposes reliability of (almost) all its component subsystems and is an issue with many dimensions:

- degree of robustness is needed as a measure of sensitivity to changes in input data and, when possible, validity conditions (B Meyer: "robustness is the ability of a [software] system to react appropriately to abnormal conditions")
- information fusion should be based on proven methods for management of uncertainty, to a large degree Bayesian statistical inference, but sometimes other techniques such as possibility theory or evidence theory; these theories then need to be interpreted in terms of a common reliability language and methodology
- Bayesian models are based on a prior distribution, a likelihood function, and when used in decision making, also a utility function; these will need to have known degree of robustness; most discussions, however, have focused on robustness to perturbance of priors. One speaks about *aleatory* vs. *epistemic* uncertainty, and about local and global robustness approaches

- information fusion reliability is a special case of modelling software reliability methodology where trust must be based on guaranteed performance; this guarantee should include precise and preferably complete (all-encompassing) conditions for validity, and be based on verification of these conditions
- performance should include character, value, precision and degree of robustness of the estimated set of parameters
- correctness is the ability of [software] products to perform their exact tasks, as defined by their specification; theoretically, guaranteed correctness is a necessary condition for both performance and robustness; however, in practice, 100% correctness can rarely be guaranteed. Can one speak of robustness anyway? Only if there is built-in surveillance to detect and redundancy to exit from emergency events
- to discuss the reliability of a particular IF system we need to assume that all those involved subsystems that do not form a part of the IF software proper are perfect, or preferably, have a known and sufficient reliability, enabling us to focus on the IF subsystem itself and its contribution to the credibility and robustness of the total system. In addition, if, for example, in a certain situation sensors or databases are unable to deliver the quality of information needed by the IF system to draw proper conclusions, maintaining trustworthiness would require that the IF system itself discovers the insufficient quality of its input
- depending on application, high robustness may seem more or less important than known robustness; but, of course, you can not know that robustness is high if it is not known. Cf. the 2003 space shuttle disaster, where it was claimed (at least in the press) that the shuttle's mechanical robustness against plastic foam spallation was believed by management to be adequate, both before and a long time after the disaster.

Many of these aspects seem obvious to engineers and scientists but are nonetheless not always respected in practice. It is simply very hard and therefore costly to solve all reliability and result quality issues before a new kind of system is fielded. But of course, it should not be up to the IF designers themselves to decide which degree of reliability is built into the systems. Instead, these requirements will have to be elicited from the customer, who in turn must be expected to represent the end user in an effective manner. Since the supply chain from scientists and system designers to end users is a long and indirect one, there is considerable risk that user needs are distorted in the process. When scientists and engineers are going to build the first information fusion systems for a network-enabled defence, based on new sensor, communication, and modelling technology, this dilemma will become acute, since the systems will eventually play survival-critical roles but might become fielded before extensive user experience has been accumulated. Thus, trustworthiness needs to be recognized as a system feature of primary importance, and ways to plausibly demonstrate system robustness will have to be developed.

We note that there is an entire scientific field of reliability engineering, with its own methods and literature, some of which will be referred to in the sequel.

The approaches we have surveyed are:

- sensitivity analysis (Frey, Kleijnen, Helton, Saltelli,...)
- robust bayesianism (Jaynes, Berger, Rios Insua, Moreno, ...)
- imprecise probabilities (Walley, de Cooman,...)
- interval probabilities (Weichselberger, Cozman, Ferson, Kreinovich, Berleant, Williamson, ...)

3. Sensitivity analysis

Sensitivity analysis of simulation models [47][48][50][51][52] can be used to identify the most significant exposure or risk factors in a model, as an aid in identifying the important uncertainties for the purpose of prioritizing additional data collection or research, and it can play an important role in model verification and validation throughout the course of model development and refinement. Sensitivity analysis also can be used to provide insight into the robustness of model results when making decisions. In [52], a number of sensitivity analysis methods are surveyed.

Kleijnen [47] gives a survey on the use of statistical designs for what-if analysis in simulation, including sensitivity analysis, optimization, and validation/verification. According to this paper, sensitivity analysis is divided into two phases. The first phase is a pilot stage, which consists of screening or searching for important factors among possibly hundreds of potentially important factors. The second phase uses regression analysis to approximate the input/output transformation that is implied by the simulation model; the resulting regression model is also known as a *metamodel* or *response surface*. Regression analysis gives better results when the simulation experiment is well-designed, using either classical statistical designs (such as fractional factorials) or optimal designs. To optimize the simulated system, the analysts may apply *Response Surface Methodology* (RSM); RSM combines regression analysis, statistical designs, and steepest-ascent hill climbing. To validate a simulation model, again regression analysis and statistical designs may be applied.

In military applications of modelling and simulation, there is a specified process of verification, validation, and accreditation (VV&A) [44][45][46]; and, in a sense, information fusion may be seen as a special case of M&S, since information fusion is partly based on situation modelling using theories of uncertainty, and frequently employs Monte Carlo simulation techniques in model evaluation and decision-making.

4. Robust Bayesian analysis

There is a common perception that foundational arguments lead to subjective Bayesian analysis as the only coherent method of behavior. According to Berger [6], non-Bayesians often recognize this, but feel that the subjective Bayesian approach is too difficult to implement, and hence they ignore the foundational arguments. Subjective Bayesian analysis is, indeed, the only coherent mode of behavior, but only if it is assumed that one can make arbitrarily fine discriminations in judgment about unknowns and utilities. In reality, it is very difficult to discriminate between, say, 0.10 and 0.15 as the subjective probability, *P(E)*, to assign to an event *E*, much less to discriminate between 0.10 and 0.100001. Yet standard Bayesian axiomatics assumes that the latter can (and will) be done. Non-Bayesians intuitively reject this, and hence reject subjective Bayesian theory.

It is perhaps less well known that realistic foundational systems exist (see, e.g., [11], and further references in [6]), based on axiomatics of behavior which acknowledge that arbitrarily fine discrimination is impossible. For instance, such systems allow the possibility that *P(E)* can only be assigned the range of values from 0.08 to 0.13; reasons for such restrictions range from possible psychological limitations to constraints on time for elicitation. The conclusion of these foundational systems is that a type of *robust Bayesian analysis* is the coherent mode of behavior. Roughly, coherent behavior corresponds to having *classes* of models, priors, and utilities, which yield a range of possible Bayesian answers. If this range of answers is too large, the question may not be settled: if the inputs are too uncertain, one cannot expect certain outputs! Indeed, if one were to perform ordinary Bayesian analysis without checking for robustness, one could be seriously misled about the accuracy of the conclusion.

Thus, robust Bayesian analysis is concerned with the sensitivity of the results of a Bayesian analysis to the inputs for the analysis. Excellent texts describing the basics of Bayesian statistics and robust Bayesian analysis are [34][35]. More recent texts are [1][2][4]. According to Berger *et al.* [5], the early 90's was the golden age of robust Bayesian analysis. In the last half of the 90s, robust Bayesian analysis shifted from being a hot topic to being a mature field within Bayesian analysis, with continued gradual development, but with less of a sense of urgency. During this period, the need to consider Bayesian robustness has increased dramatically, in that the modeling that is now routinely utilized in Bayesian analysis is of such complexity that inputs (such as priors) can be elicited only in a very casual fashion. Thus, in the opinion of the authors of [5], it is now time to focus again on Bayesian robustness and to attempt to bring its ideas into the Bayesian mainstream. New opportunities are offered by the developments in algorithms [7][8][9], the possibility of using MCMC methods and the need for sensitivity analysis in other fields. Bayesian robustness is playing a role in SAMO (*Sensitivity Analysis of Model Output*), a group interested in investigating the relative importance of model input parameters on model predictions [51] in many applied areas.

David Rios Insua recently wrote the note [3], where he explains the Bayesian approach to inference and decision analysis as follows:

- Modelling beliefs about a parameter of interest through a prior which, in presence of further information, is updated to the posterior
- Modelling preferences and risk attitudes about (multicriteria) consequences, with a multiattribute utility function
- Associate with each alternative its (multiattribute) posterior expected utility
- Propose the alternative which maximises the posterior expected utility

As in any quantitative approach, there are many reasons to check the sensitivity of the output (the optimal alternative) with respect to the inputs (model, beliefs and preferences). In addition, since in this framework inputs to the analysis encode the decision makers' judgements, she should wish to explore their implications and possible inconsistencies. The need for sensitivity analysis is further emphasized by the fact that the assessment of beliefs and preferences is a difficult task. This is an important point, since her judgements will evolve through the analysis until they are adequate. Robust Bayesian analysis guides this process.

The usual practical motivation underlying robust Bayesian analysis is the difficulty in assessing the prior distribution. Consider the simplest case in which it is desired to elicit a prior over a finite set of states s_i, $i=1,...,I$. A common technique to assess a precise prior $\Pr(s_i) = p_i$ with the aid of a reference experiment proceeds as follows: one progressively bounds $\Pr(s_i)$ above and below until no further discrimination is possible and then takes the midpoint of the resulting interval as the value of p_i. Instead, however, one could directly operate with the obtained constraints $a_i \leq \Pr(s_i) \leq b_i$, acknowledging the cognitive limitations.

The same situation holds when modelling preferences. One might assess the utility of some consequences through, say, the certainty equivalent method [77], and then fit a utility function. However, in reality we only end up with upper and lower constraints on such utilities, possibly with qualitative features such as monotonicity and concavity, if preferences are increasing and risk averse. These constraints can often be approximated by an upper and a lower utility function, leading to the consideration of all utility functions that lie between these bounds. If a parameterized utility function is assessed, the constraints are typically placed on the parameters of the utility, say the risk aversion coefficient. Of course, in developing the model for the data itself there is a typically great imprecision, and a need for careful study of the model robustness.

When there are several decision makers and/or experts involved in the elicitation, it is not necessarily theoretically possible to obtain a single model, prior, or utility; one might be left with only classes of each, corresponding to differing expert opinions.

Robust Bayesian analysis provides tools to check the impact of the utility function, the prior and the model on the optimal alternative, and its posterior expected utility. We distinguish three main approaches to Bayesian robustness. We illustrate it considering robustness with respect to changes in the prior, but similar issues are raised when considering likelihoods and utilities. A "guided tour" through these approaches is presented in Berger et al. [5] and the references therein.

4.1. Informal approach

The first approach is the informal one, which considers several priors and compares the quantity of interest (e.g., the posterior mean) under each of them. The approach is very popular because of its simplicity. Its rationale is that since we shall be dealing with messy computational problems, why not analyze sensitivity by trying only some alternative pairs of utilities and priors? While this is a healthy practice and a good way to start a sensitivity analysis, in general this will not be enough and we should undertake more formal analyses: the limited number of priors chosen might not include some which are compatible with the prior knowledge and could lead to very different values of the quantity.

It is worth mentioning that the consideration of a finite number of utilities links clearly with multi-objective decision making problems.

4.2. Global robustness

The most popular approach in Bayesian robustness is called *global sensitivity.* All probability measures compatible with the prior knowledge available are considered and robustness measures are computed as the prior varies in a class. Computations are not always easy since they require the evaluation of good upper and lower bounds of quantities of interest. The robustness measures provide, in general, a number that in principle should be interpreted in the following way:

1. if the measure is "small", then robustness is achieved and any prior in the class (possibly one computationally simple) can be chosen without relevant effects on the quantity of interest,
2. if the measure is "large", then new data should be acquired and/or further elicitation narrows the class, recomputing the robustness measure and stopping as in the previous item,
3. if the measure is "large" and the class can not be modified, then a prior can be chosen in the class but we should be wary of the relevant influence of our choice on the quantity of interest. Alternatively, we may use an ad hoc method such as *G-minimax* [3] to select an alternative.

4.3. Local robustness

The last approach looks for *local sensitivity* and studies the rate of change in inferences and decisions, using functional analysis differential techniques such as Fréchet derivatives of posterior expected utilities and their norms, total derivatives, or Gâteaux differentials.

4.4. Decision and utility robustness

An important and occasionally controversial issue is the distinction between decision robustness and expected utility robustness. A variety of situations may occur. For instance, when performing sensitivity analysis, it may happen that expected utility changes considerably with virtually no change in the optimal Bayes action, even if the utility is fixed.

4.5. Discussion

A number of results show that we may model imprecision in beliefs and preferences through a class of probability distributions and a class of utility functions. These results have two basic implications. First, they provide a qualitative framework for sensitivity analysis, describing under what conditions we may undertake the standard and natural sensitivity analysis approach of perturbing the initial probability-utility assessments, within some reasonable constraints. Second, they point to the basic solution concept of robust approaches, thus indicating the basic computational objective in sensitivity analysis, as long as we are interested in decision analytic problems: that of *non-dominated alternatives*. This corresponds to a *Pareto ordering* of decision rules, based on inequalities on the posterior expected utility.

As a consequence of this model, the solution concept is the set of non-dominated alternatives. In some cases, non-dominance is a very powerful concept leading to a unique non-dominated alternative. However, in most cases the non-dominated set will be too large to imply a final decision. It may happen that there are several non-dominated alternatives and differences in expected utilities are non-negligible. If such is the case, we should look for additional information that would help us to reduce the classes, and, perhaps, reduce the non-dominated set.

Stability theory provides another unifying general sensitivity framework, formalizing the idea that imprecisions in elicitation of beliefs and preferences should not affect the optimal decision greatly. When *strong stability* holds, careful enough elicitation leads to decisions with expected utility close to the smallest achievable; when *weak stability* holds, at least one stabilized decision will have this property. However, when neither concept of stability applies, even small elicitation errors may lead to disastrous results in terms of large losses in expected utility.

4.6. Conclusion

The approach proposed by Berger *et al.* in [5] may be summarized as follows: at a given stage of analysis, one elicits information on the decision maker's beliefs and preferences, and consider the class of all priors and utilities compatible with such information. One approximates the set of non-dominated solutions; if these alternatives do not differ too much in expected utility, one may stop the analysis; otherwise, one needs to gather additional information, possibly with the tools outlined above. This would further constrain the class: in this case the set of non-dominated alternatives will be smaller and one could hope that this iterative process would converge until the non-dominated set is small enough to reach a final decision.

It is conceivable in this context that at some stage one might not be able to gather additional information, yet there remain several non-dominated alternatives with very different expected utilities. In these situations, *L x G - maximin* solutions may be useful as a way of selecting a single robust solution. One associates with each alternative its worst expected utility; then suggests the alternative with the maximum worst expected utility.

From an engineering perspective of trustworthy information fusion, robust Bayesian concepts are certainly likely to be of fundamental importance. It is less clear at this point how these concepts can be applied in practical robust fusion algorithms. Few papers address these issues (but see [16] and Section 9 below), and much more research is needed. However, application of analytical robust Bayesian analysis to complex engineering systems, such as information fusion systems, is likely to be very difficult. Perhaps numerical approaches to robustness such as *probability bounds analysis*, briefly described in Section 10 below, will be more tractable. On the other hand, experience with closely related methods in numerical analysis [30][31] indicates that they are likely to provide very pessimistic estimates unless great care is taken in algorithm design.

5. Imprecise probabilities

Recently, there has been a flurry of interest in theories of imprecise probabilities. In particular, the biannual conference series ISIPTA (*International Symposia on Interval Probabilities and Their Applications*) started in 1999. A selection of 10 papers from the first ISIPTA conference was published in 2000 as a special issue [18] of the *International Journal of Approximate Reasoning*. The fourth ISIPTA conference will take place in Pittsburgh in July 2005. Another recent effort, of particular relevance for engineering applications, is the *Sandia Workshop on Alternative Representations of Epistemic Uncertainty*, held in August 2002. This workshop has been documented in a special issue of the journal *Reliability Engineering & System Safety* [19]. Both these journal issues are highly relevant to this study, as are the proceedings of the ISIPTA conferences.

Imprecise probability is a generic term used to describe mathematical models that measure uncertainty without precise probabilities. This is certainly the case with robust Bayesian analysis, but there are many other imprecise probability theories, including *upper and lower probabilities, belief functions, Choquet capacities, fuzzy logic*, and *upper and lower previsions*, see [11][12][13]. A pioneering researcher in theories of imprecise probabilities is Peter Walley, who in 1991 published a book (now out of print) with nearly 800 pages entitled *Statistical reasoning with Imprecise Probabilities*, [11]. Since then, Walley has made additional contributions, and in particular, explanations of the theory [12][13], and some of them will be briefly surveyed below.

Some of these theories, such as fuzzy logic and belief functions, are only tangentially related to robust Bayesian analysis (but see [16] and section 9 below, where results of some of these methods are compared within a robust Bayesian framework, using a difficult but perhaps somewhat artificial example), while others are closely related; for example, some classes of probability distributions that are considered in robust Bayesian analysis, such as distribution band classes, can also be interpreted in terms of upper and lower probabilities. Classes of probability distributions used in robust Bayesian analysis will typically generate upper and lower previsions as their upper and lower envelopes. Walley [11] describes the connection between robust Bayesian analysis (in terms of sensitivity to the prior) and the theory of coherent lower previsions.

Seen from the robust Bayesian perspective of [5], in a crude sense, the major difference between robust Bayesian analysis and these alternative theories is that robust Bayesian analysis stays with ordinary Bayesian intuition, considering classes (of priors, say) that consist only of those priors that are individually compatible with prior beliefs. In contrast, the alternative theories view the classes themselves (not the individual priors) as the basic elements of the theory.

Walley's view, on the other hand [13], is that a general theory of imprecise probability can accommodate all the kinds of uncertainty and partial ignorance that are currently being studied, including vague or qualitative judgements of uncertainty, models for complete ignorance or near ignorance, random sets and multivalued mappings, and partial information about an unknown probability measure. But it is not yet clear what level of mathematical generality will be needed in a unified theory of imprecise probability. Walley argues in [13] that none of the four mathematical models that are most popular at present (see below) is sufficiently general, and proposes other models that do seem to be sufficiently general but have received less attention.

5.1. Importance of a common mathematical model

Each problem in a set of basic "challenge problems" (see [27] and Sec. 10.2. below) analyzed at the Sandia workshop provides expert assessments for the value of two parameters, *a* and *b*. It is important to note that each of the expert assessments deals with a single parameter. Moreover, it is either of the

- *vacuous type*, such as the interval information "*a* belongs to A = [0.1, 0.9]"
- *Bayesian type*, such as "*b* has the log-normal distribution with parameters *m* = 0.5 and *s* = 0.5"
- *Bayesian type with vacuous parameters*, such as "*b* has a log-normal distribution with parameters m belonging to the interval *M* = [0.0, 1.0], and *s* belonging to the interval *S* = [0.1, 0.5]"

Assessments of the last type are *hierarchical*: an assessment of the variable *b* is made through an assessment about variables *m* and *s* and an assessment about the variable *b* conditional on the variables *m* and *s*. A first step toward combining these assessments is to express them using mathematical models of the same type. This allows one to deal with all sources of information in a uniform way. In fact, all the given expert assessments can be modelled by specific imprecise probability models, called *coherent lower previsions*.

Coherent upper and lower probabilities, Choquet capacities of order 2, belief functions and possibility measures are amongst the most popular mathematical models for uncertainty and partial ignorance. However, according to Walley [13], these models are not sufficiently general to represent some common types of uncertainty. In particular, they are not sufficiently informative about expectations and conditional probabilities. Coherent lower previsions and sets of probability measures are considerably more general, but they may not be sufficiently informative for some purposes. Two other models for uncertainty, which involve *partial preference orderings* and sets of desirable gambles, are discussed in [13]. These are more informative and more general than the previous models, and they may provide a suitable mathematical foundation for a unified theory of imprecise probability.

The mathematical models of uncertainty considered by Walley in the article [13] are, in order of increasing generality:

1. possibility measures and necessity measures

2. belief functions and plausibility functions

3. Choquet capacities of order 2

4. coherent upper and lower probabilities, also known as interval-valued, or interval, or non-additive probabilities

5. coherent upper and lower previsions

6. sets of probability measures

7. sets of desirable gambles

8. partial preference orderings, with an interesting special case, called partial comparative probability orderings

Again according to Walley [13], the models listed are all appropriate and useful in particular types of application. Some well-known examples are: (1) vague judgements of uncertainty in natural language; (2) multivalued mappings and non-specific information; (3) some types of statistical neighborhood in robustness studies, and various economic applications; (4) personal betting rates, and upper and lower bounds for expectations, and envelopes of expert opinions, (6) partial information about an unknown probability measure, and robust statistical models; (7,8) preference judgements in decision making; and (9) qualitative judgements of uncertainty.

Walley argues that the most promising candidates for a single sufficiently general model to accommodate all the listed models would be (7) or (8). Of course, this does not mean that it is always desirable in practice to switch to these methods, unless they are also at least as easy to use as their less general alternatives. But it is not at all uncommon that, e.g., belief functions and possibility measures are unsuitable models, since they are useful (more specifically, provide coherent expectations) only in cases where lower probabilities are 2-monotone, a rather strong condition. However, there are important applications in which belief functions or possibility measures are good models. Some of these applications (see [14][15]) are especially important because they are clear instances where probability measures are inadequate models and imprecise probabilities are needed.

5.2. Choquet capacities of order 2

Let O denote the set of possible outcomes under consideration. Suppose that lower probabilities $P(A)$ are defined for all elements A in K, where K is a collection of subsets of O. Here, K is assumed to be an algebra. For models 1-4 in the list above, lower probabilities determine conjugate upper probabilities through $P(A) = 1 - \underline{P}(A^c)$, so it suffices to consider lower probabilities (upper and lower probabilities are also known as interval-valued or interval or non-additive probabilities). Let $\varnothing$ denote the empty set. Assume that $\varnothing \leq P(A) \leq 1$ for all A in K, $P(\varnothing) = 0$ and $P(O) = 1$. The lower probability P is said to be *2-monotone*, or a *Choquet capacity of order 2* or a *convex capacity*, when it also satisfies, whenever A and B are in K,

$$P(A \vee B) + P(A \wedge B) \geq P(A) + P(B)$$

It is well known that probability measures, belief functions and necessity measures (the conjugates of possibility measures) are 2-monotone lower probabilities.

A simple method of constructing 2-monotone lower probabilities is to apply a convex transformation to the probability interval: if P_0 is a probability measure on *K*, *f* is a convex function from [0, 1] into [0, 1] with *f*(0) = 0, and $\underline{P}$ is defined by $\underline{P}(O) = 1$ and $\underline{P}(A) = f(P_0(A))$ when *A* is a member of *K* and $A \neq O$, then $\underline{P}$ is a 2-monotone lower probability.

Walley shows by a simple example that there are coherent lower probabilities which are not 2-monotone. Therefore, he concludes that order-2 capacities are not sufficiently general to provide a basis for a general theory of imprecise probability.

5.3. Coherent lower and upper previsions

Let us now consider an agent who is uncertain about something, say, the value of the variable *a* that takes values in the set *A* [21]. A *gamble* is a bounded mapping from *A* to the set of real numbers *R*, and it is interpreted as an uncertain reward: if some $\hat{a}$ in *A* would turn out to be the true value of the variable *a* then the agent would receive the amount $X(\hat{a})$, expressed in units of some (predetermined) linear utility. Gambles play a similar part in the theory of imprecise probabilities as events do in the classical, or Bayesian, theory of probability. In fact, any event can be interpreted as a very simple game that only allows the modeler to distinguish between two situations: the event either occurs, or it does not, and the reward depends only on whether or not it does. So, an *event*, modelled as a subset $\underline{A}$ of the space *A* of possible parameter values corresponds to a gamble $I_{\underline{A}}$ (its indicator) that yields one unit of utility if it occurs, i.e., if *a* belongs to $\underline{A}$, and zero units if it does not, i.e., if *a* belongs to the complement of $\underline{A}$. In other words, there is a natural correspondence between events and zero-one-valued gambles. The concept of a gamble can therefore be seen as a generalization of the concept of an event. The set of all gambles associated with the variable *a* is denoted by *L*(*A*). It is a real linear space under the point-wise addition of gambles and the scalar point-wise multiplication of gambles with real numbers.

The information the agent has about the value of the parameter a will lead him to accept or reject transactions whose reward depends on this value, and we can formulate a model for his uncertainty by looking at a specific type of transaction: buying gambles. The agent's *lower prevision* (or supremum acceptable buying price, or lower expectation) $\underline{P}(X)$ for a gamble *X* is the greatest real number *s* such that he is disposed to buy the gamble *X* for any price strictly lower than *s*. If the agent assesses a supremum acceptable buying price for every gamble *X* in some subset *K* of *L*(*O*), the resulting mapping $\underline{P}$: $K \rightarrow R$ is called the agent's lower prevision. $\overline{P}$ will denote the conjugate *upper prevision* for *X*. $\overline{P}(X)$ represents the agent's infimum acceptable price for selling the gamble X.

Lower and upper previsions for gambles are a natural generalization of probabilities for events. Indeed, any assessment of a probability of an event can be translated into an assessment of a supremum buying price and an infimum selling price for a zero-one-valued gamble. Suppose that the probability of the event *A* is known to be *p*. The reward we expect from I_A is then equal to $0*(1-p) + 1*p = p$. Therefore, we are willing to buy I_A for any price less than *p*, and we are willing to sell I_A for any price greater than *p*. We infer that $\underline{P}(I_A) = \overline{P}(I_A) = p$. The power of lower and upper previsions, compared to classical probability theory, is that lower and upper previsions allow for far more generality. In particular, the theory does not require that an agent's supremum buying price should be equal to his infimum selling price.

A particular benefit from this way of modelling available information (or uncertainty) is that it leads naturally to a theory of decision making under uncertainty. For example, making a particular decision d from a set D of alternatives is behaviorally equivalent to accepting a gamble X_d, which represents the (possibly negative) utility received as a function of the value of the (unknown) parameter a of the decision problem. The agent should strictly prefer one action d_1 over an alternative d_2 if $P(X_{d1} - X_{d2}) > 0$: this means that he is willing to pay some strictly positive amount of utility for exchanging the rewards of making decision d_2 with those of making d_1.

Since a lower prevision $\underline{P}$ represents an agent's commitments to act in certain ways -- to buy gambles X in its domain K up to certain prices $P(X)$ -- it should satisfy a number of requirements that ensure that his behaviour is rational. The strongest such rationality criterion is that of coherence. It is easiest to understand and define if the domain K is the set of all gambles $L(A)$. A lower prevision P on $L(A)$ is called coherent if it satisfies the following three requirements:

1. *Accepting sure gains*: The agent should always be willing to buy a gamble X for a price equal to the lowest possible reward he may expect from X, that is inf[X]. Hence, it should hold that $Q(X) \geq \inf[X]$ for all gambles X.

2. *Positive homogeneity*: Next, since we are working with a linear utility, buying prices should be independent of the choice of scale of the utility. Mathematically, this means that $Q(cX) = c\,Q(X)$ for each gamble X and $c > 0$.

3. *Superadditivity*: Finally, since we are working with a linear utility, if the agent is willing to buy X for price $Q(X)$ and Y for price $Q(Y)$, he should be willing to buy $X + Y$ for at least $Q(X + Y) \geq Q(X) + Q(Y)$ for all gambles X and Y.

A lower prevision P on an arbitrary domain K in $L(A)$ is called coherent if it is the *restriction* of -- can be extended to -- some coherent lower prevision on $L(A)$.

5.4. Walley's updating principle

Walley's updating principle states that "Any gamble Z is B-desirable if and only if BZ is desirable" [11]. A gamble being desirable here means that You are disposed to accept it. B-desirable mean that You intend to accept Z provided You observe only the event B (that happens or fails), and the payoff BZ is Z if B happens, otherwise 0.

A justification of Walley's updating principle is given by Shafer, Gillet and Scherl in [20]. They differentiate between a House and a Gambler in a betting situation. In traditional terms a person's beliefs are revealed by the bets he is willing to accept. The odds of the bet that he is willing to accept then defines his probabilities. Shafer *et al*. associate You with the House but take the view of the Gambler, asking the Gambler if he believes the process can make him very rich in the long run. By taking the position of a Gambler and asserting that certain numbers do not offer the Gambler any opportunity to get very rich, then these numbers are objective probabilities. When we advance these same numbers as subjective probabilities we assert that they do not offer us (the Gambler), with the knowledge we have, any opportunity to get very rich. The issue is what we (the Gambler) then can do with the number, not how we got them. This way we bring the concept of subjective probability closer to the concept of objective probability.

5.5. Sharing epistemic probabilities

When sharing intelligence and epistemic probabilities between systems in a system-of-systems it will be crucial to investigate the underlying model of the different systems and the possibilities to adjust probabilities when passing information from one system to another. As Shafer points out "a proposition attains its full meaning only as part of its frame, and its degree of support or epistemic probability is always assessed relative to that frame. When two incompatible frames are compared it may be possible (...) to find a close resemblance between a proposition in one of the frames and a proposition in the other, but no matter how close this resemblance is, the two propositions will be formally different - and their degrees of support may be very different indeed" [14], p. 284.

5.6. Conclusion

While advocating more general models, Walley [13] recognizes the need and usefulness of research in upper and lower probabilities, Choquet capacities, belief functions, possibility measures and other special kinds of model. Each of these models is useful in some kinds of application, but there is a need to spend more effort to study more general models which are needed in many applications.

Coherent lower and upper previsions are needed in a general theory because they are direct generalizations of the most commonly used models (coherent lower and upper probabilities, order-2 capacities, belief functions, possibility measures, linear previsions and probability measures), so that a theory of imprecise probability can be applied directly to these special models. Lower previsions are much more general and informative than lower probabilities, and they seem to be adequate models in the great majority of applications that are concerned with uncertainty but not with utility, and those applications in which utilities are precisely known. They also have the advantage of being relatively close to well-established concepts of probability and expectation, and especially to de Finetti's concept of expectation.

Sets of probability measures are also needed in a general theory because, at present, most examples of coherent models are presented in this form. Since upper and lower envelopes of a set of probability measures are always coherent upper and lower previsions, specifying a set of probability measures is a simple way of constructing a coherent model. For example, after receiving new information, a set can be updated by using Bayes' rule to update each probability measure in the set. This is the approach in the robust Bayesian theory, which uses a set of probability measures as the canonical model for uncertainty.

According to Walley [13], the robust Bayesian approach has some serious defects, and sets of probability measures are not an adequate foundation for a general theory of imprecise probability. There are many applications in which it is mistaken and misleading to regard a set of probability measures as a set of hypotheses about the "correct" probabilities, because it is meaningless to talk of "correct" probabilities. That is the case in many applications of belief functions and possibility measures. Another disadvantage of sets of probability measures is that they are not as closely related as the other models to decision making and observable behaviour. To understand the practical implications of a set of probability measures, it is often necessary to calculate upper and lower probabilities, previsions or preferences.

Sets of desirable gambles and partial preference orderings are the most informative of the mathematical models discussed by Walley in [13], and he claims that they seem to be able to model all the common types of uncertainty. They uniquely determine upper and lower previsions, and they contain all the information about preferences that is relevant in making decisions. In many ways they are the simplest and most natural mathematical models. The coherence axioms and rules of inference (natural extension) for sets of desirable gambles are especially simple. Partial preference orderings are direct generalizations of partial comparative probability orderings, and they are essential in a general theory of decision. The main difficulty with these models is that, because they can be more informative than coherent lower previsions and sets of probability measures, they can also be more complex and difficult to specify. Further research is needed to develop special types of sets of desirable gambles that can be easily specified, such as the finitely generated models.

Coherent lower previsions, sets of probability measures and sets of desirable gambles are each useful for different purposes. Walley concludes that a unified theory of imprecise probability will need to make use of all three models and to exploit the duality relationships between them. In a general theory, it may be appropriate to adopt the most general of these models, sets of desirable gambles or partial preference orderings, as the fundamental model.

6. Weichselberger's axiomatic approach to interval probabilities

As mentioned above, the concept of interval probabilities is usually used synonymously with that of upper and lower probabilities. In [17], Weichselberger discusses an axiomatic approach to *interval-probability*, as he calls it, and it seems that in the way he develops this concept, it may be more generally applicable than upper and lower probabilities. The goals of this theory are specified as follows:

1. Different kinds of uncertainty should be treated by the same concept. This applies to:
- imprecise probability and uncertain knowledge
- imprecise data
- the use of Choquet capacities
- the concept of ambiguity and its employment in decision theory
- belief functions and related concepts
- interpretation of interval-estimates in classical theory
- the study of experiments with possibly diverging relative frequencies
- non-additive measures (fuzzy measures)

2. As a special case, classical probability must fit into the theory

3. A simple system of axioms must describe the fundamentals of the theory

4. All statements of the theory must be derivable from the given axioms and appropriate definitions

5. The domain of application must neither be limited to purely formal aspects nor be bound by a certain interpretation of probability.

The axiom system developed by Weichselberger is directly related to Kolmogorov's classical probability axioms.

There is an obvious limitation for any theory of interval-probability: only those assessments which assign intervals to random events qualify as genuine subjects of the theory. The benefits and the power of the theory are due to the duality between a set of interval-limits and the corresponding set of classical probabilities. These qualities distinguish the approach from those admitting more general types of probability assignments, such as Walley's.

According to [17], a general approach to decision problems with respect to behavioral viewpoints is made possible by the theory. Behaviour under ambiguity can be analyzed and classified. An interesting issue in this context is discussed by Augustin in [72].

7. Information-gap Decision Theory

Information-gap theory developed by Ben-Haim [58] is radically different from the probabilistic approaches discussed above. He models uncertainty as an information gap rather than a probability. The info-gap is the disparity between what *is known* and what *needs to be known* in order to make a well-founded decision. The need for information gap modelling and management of uncertainty arises in dealing with severe lack of information and highly unstructured uncertainty. This theory can be used to measure the robustness of a classifier. Let a classification algorithm be denoted as a function $C(u)$ where u is a measurement vector and $C(u) = n$ represents the decision that u arose from class n. However, if u actually arose from class m then $\|C(u) - m\|$ represent a quantitative assessment of the error.

The robustness function for C is the greatest value of the uncertainty parameter for which the normed error in the classification is no greater than a classification error r_c:

$$\hat{\alpha}(C, r_c) = \max\{\alpha : \|C(u) - n\|, \forall(n, u \in \mathcal{U}_n(\alpha, \tilde{u}_n))\}$$

When the robustness $\hat{\alpha}(C, r_c)$ is large, then the classification error is $\leq r_c$ for any class, regardless of uncertainty. As usual, by reducing the classification error the robustness is also reduced.

8. Robustness and contradictory propositions

An important issue in robustness concerns partially contradictory propositions. Zadeh has pointed out that in general, a probability-bound interpretation of belief functions is inconsistent with normalization in Dempster's rule [53]. In a comment Shafer points out that belief functions can be made to define consistent probability bounds with an appropriate discounting before combination [54]. Zadeh also criticizes Dempster's rule for leading to paradoxes, such as in the case with two belief functions with a low supported intersection, e.g., m_1(A)=0.95, m_1(B)=0.05, m_1(C)=0, m_2(A)=0, m_2(B)=0.05, m_2(C)=0.95 with an {A,B,C} frame, to yield m_{12}(B)=1 and m_{12}(A)=m_{12}(C)=0 after combination, regardless of the assigned basic belief to A and B by m_1 and to B and C by m_2.

We believe, however, that the critique is somewhat misguided. If you accept Dempster's rule as it is, and a static frame of discernment, you have also accepted the idea of a closed world and normalization. The result yielding m_{12}(B)=1 is after all completely consistent with logic as the first belief function states in logical terms $\neg C$ and the second $\neg A$, with $(A \wedge B \wedge C) \wedge \neg A \wedge \neg C \Rightarrow B$, or to quote Sir Arthur Conan Doyle "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

However, in many real world applications a closed world assumption is not a realistic modelling assumption. We can often not claim with certainty that our representation (frame of discernment) is exhaustive and mutually exclusive. Thus, the problem seems to lie not with the rule but with our model, i.e., representation.

To accept the modelling problem we might adopt an open world assumption and the TBM idea of not making any normalization in Dempster's rule as a way to indirectly deal with the modelling problem. As long as the conflicts are small it does not matter whether we normalize or not. However, more important is to use the conflict as a warning signal of a robustness problem.

As an alternative to the TBM approach we may instead consider the model perfect and view the robustness problem of a high conflict as a problem regarding the sources. Instead of going from normalization to no normalization we might choose to make systematic discounts of highly conflicting sources in such a way as to bring the conflict down to whatever seem reasonable.

A third more general option would be to trust neither the sources nor the frame and discount sources whenever the conflict is too large within TBM.

Whichever approach is taken, problem solving should be viewed as two simultaneous problems, that of fusing information and a parallel conflict management problem.

When constructing a frame "we can pick out two considerations that influence it:

(1) we want our evidence to interact in an interesting way, and

(2) we do not want it to exhibit too much internal conflict. Two items of evidence can always be said to interact, but they interact in an interesting way only if they jointly support a proposition more interesting than the propositions supported by either alone."

"[S]ince interesting interactions can always be destroyed by loosening relevant assumptions and thus enlarging our frame, it is clear that our desire for interesting interaction will incline us towards abridging or tightening our frame. Our desire to avoid excessive internal conflict in our evidence will have precisely the opposite effect: it will incline us towards enlarging or loosening our frame. For internal conflict is itself a form of interaction – the most extreme form of it. And it too tends to increase as the frame is tightened, decrease as it is loosened."

"Notice, too, that we will tend to enlarge our frame as more evidence becomes available. For as we accumulate evidence we are likely to find more and more interactions that persist even in a looser frame and more and more conflicts that force us to looser frames" [14], p. 280–281.

In a presentation at *Fusion 2004,* Haenni [55] discussed Zadeh's critique and proposed two ways to achieve robustness by viewing the high conflict as a modelling problem. In the example the frame consists of three diseases: meningitis, concussion, or tumor: $\Theta_1 = \{M,C,T\}$. In the first instantiation of the example an underlying modelling assumption is that a patient can only have one illness at a time, since a frame of discernment consists of mutually exclusive alternatives, i.e., exactly one alternative is the true disease. In a second more realistic model no such assumption is made. We assume that the patient can have any number of illnesses or none, $\Theta_2 = \{\emptyset, M, C, T, MC, MT, CT, MCT\}$, *i.e.*, exactly one alternative is the true combination of diseases.

If you have two highly conflicting pieces of evidence (as in Zadeh's example) you will get a high conflict. Let the first doctor state that the patient has meningitis (M) with 99% but a small 1% chance of concussion (C). The second doctor states a 99% chance of tumor (T) but with a small 1% chance of concussion (C).

$$m_1(A) = \begin{cases} 0.99, & A = \{M\} \\ 0.01, & A = \{C\} \\ 0 & \text{for all other } A \subseteq \Theta \end{cases} \qquad m_2(A) = \begin{cases} 0.99, & A = \{T\} \\ 0.01, & A = \{C\} \\ 0 & \text{for all other } A \subseteq \Theta \end{cases}$$

After combination and normalization our conclusion is 100% for concussion (C) and 0% for the other alternatives:

$$m_{1 \oplus 2}(A) = \begin{cases} 1, A = \{C\} \\ 0, A \subseteq \Theta \end{cases}$$

This is a perfectly valid conclusion given that you accept that a patient can only have one illness (not zero, two or three) and that both doctors are completely reliable. If you don't accept this, then you have a modelling problem that should be addressed up front and possibly online and not be swept under the carpet. We notice that here we also have an obvious robustness problem. Moving just 1% of the support for M in the first evidence to T will change the conclusion to almost surely tumor (T):

$$m_{1 \oplus 2}(A) = \begin{cases} 0.99, & A = \{T\} \\ 0.01, & A = \{C\} \\ 0, & \text{for all other } A \subseteq \Theta \end{cases} .$$

In the second instantiation of the problem with $\Theta_2 = \{\emptyset, M, C, T, MC, MT, CT, MCT\}$ the two pieces of evidence from the two doctors are simply represented as

$$m_1(A) = \begin{cases} 0.99, & A = \{M, MC, MT, MCT\} \\ 0.01, & A = \{C, MC, CT, MCT\} \\ 0, & \text{for all other } A \subseteq \Theta \end{cases} \quad m_2(A) = \begin{cases} 0.99, & A = \{T, MT, CT, MCT\} \\ 0.01, & A = \{C, MC, CT, MCT\} \\ 0, & \text{for all other } A \subseteq \Theta \end{cases} .$$

After combination we have (with no conflict)

$$m_{1 \oplus 2}(A) = \begin{cases} 0.9801, & A = \{MT, MCT\} \\ 0.0099, & A = \{MC, MCT\} \\ 0.0099, & A = \{CT, MCT\} \\ 0.0001, & A = \{C, MC, CT, MCT\} \\ 0, & \text{for all other } A \subseteq \Theta \end{cases}$$

and may conclude

$$Bel_{1\oplus 2}(''M'') = Bel_{1\oplus 2}(\{M, MC, MT, MCT\}) = 0.99$$
$$Bel_{1\oplus 2}(''T'') = Bel_{1\oplus 2}(\{T, MT, CT, MCT\}) = 0.99$$
$$Bel_{1\oplus 2}(''C'') = Bel_{1\oplus 2}(\{C, MC, CT, MCT\}) = 0.0199$$

i.e., that the patient has both meningitis and tumor. Thus, a model refinement completely changes the conclusion and eliminates the robustness problem (a change of 1% from M to T in evidence m_1 will hardly change the belief at all):

$$Bel_{1\oplus 2}(''M'') = Bel_{1\oplus 2}(\{M, MC, MT, MCT\}) = 0.98$$
$$Bel_{1\oplus 2}(''T'') = Bel_{1\oplus 2}(\{T, MT, CT, MCT\}) = 0.9901 \ .$$
$$Bel_{1\oplus 2}(''C'') = Bel_{1\oplus 2}(\{C, MC, CT, MCT\}) = 0.0199$$

9. On the relationship between Bayesian inference and evidence theory

Although many papers discuss the abstract relationship between different approaches to reliable uncertainty management, papers which compare different approaches in concrete examples are less abundant. Notable exceptions are some of the papers from the Sandia workshop, presented in [19], as well as the paper [16] where Arnborg discusses the relationship between robust Bayesian inference and evidence theory, using Zadeh's example in its original form to graphically illustrate the effects of different rules of combination.

Arnborg notes that to obtain bodies of evidence, likelihoods and priors are needed, and therefore an analysis of a hypothetical Bayesian obtainment of bodies of evidence might shed light on problems in evidence and aggregation theory. The natural approach to impreciseness used in this paper is that impreciseness in conclusions is caused by impreciseness in sampling functions and priors. An assessment that the sampling function is imprecise gives the same effect on the body of evidence, regardless of what the reason is. Particularly, a body of evidence represented by a DS-structure (bpa) has an interpretation as a set of possible probability distributions, and combining or aggregating two such structures can be done in robust Bayesian analysis.

The Dempster-Shafer (DS) combination rule is computationally equivalent to allowing the operands as well as the result to be non-empty, not necessarily singleton, random sets. The combination of evidence - likelihood functions normalized so they can be seen as probability distributions - and a prior over a finite space is done in this paper by component-wise multiplication followed by normalization. The resulting combination operation agrees with the DS and the MDS rules, the latter proposed by Fixsen and Mahler [10] and involving a re-weighting of the operands, for precise beliefs. The robust Bayesian version of this would replace the probability distributions by sets of probability distributions, for example represented as DS beliefs.

A set of distributions which is not a Choquet capacity can be approximated by *rounding* it to a minimal Choquet capacity that contains it, and this rounded set can be represented by a DS-structure. Thus, imprecise distributions can, if constrained by rounding to Choquet capacities, be viewed as random sets. The random sets can be combined by taking the intersection of the participating random sets on condition that the result is non-empty (i.e., component-wise multiplication followed by normalization) and the resulting random set can be regarded as a Choquet capacity.

Arnborg introduces the concepts of *robust Bayesian combination operator* and *rounded robust Bayesian combination operator* and notes that they are both monotone with respect to imprecision:

Let F_1 and F_2 be two probability distributions over a common space Λ. The *robust Bayesian combination operator* $\times$ is defined by:

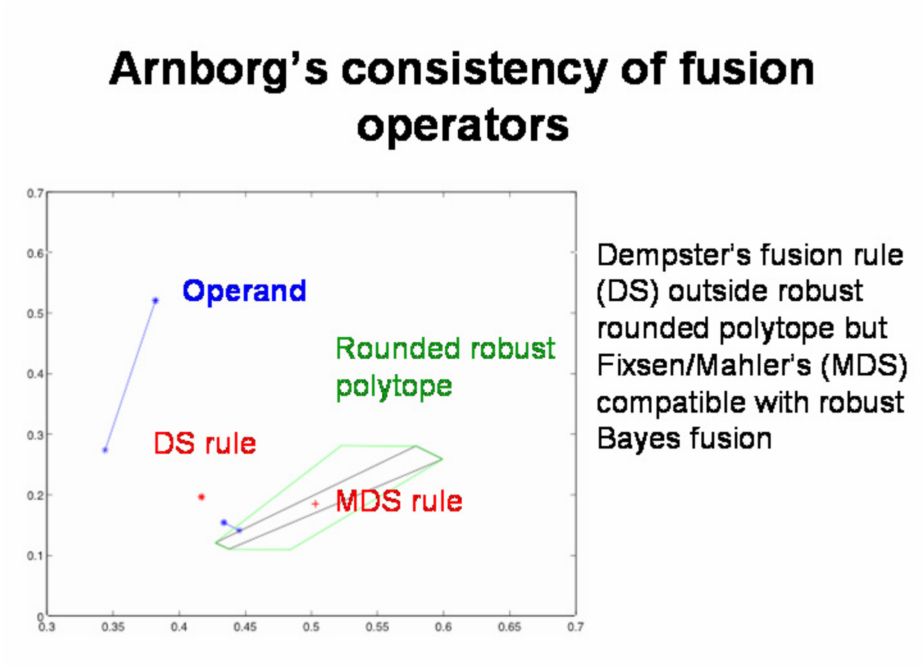
$$F_1 \times F_2 = \left\{ cf_1f_2 | f_1 \in F_1, f_2 \in F_2, c = 1/\sum_{\lambda \in \Lambda} f_1(\lambda)f_2(\lambda) \right\}$$

The rounded robust operation is defined as applying the robust operator to the rounded operands, then rounding the result.

If F_i and F'_i are (rounded) robust operators, and if $F'_i \subseteq F_i$, then $F'_1 \times F'_2 \subseteq F_1 \times F_2$. This property is called *monotonicity with respect to imprecision*.

Arnborg shows that for any combination operator $\times'$ that is monotone wrt imprecision and is equal to the Bayesian (Dempster's) rule for precise arguments, $F_1 \times F_2 \subseteq F_1 \times' F_2$, where $\times$ is the robust rule.

When interpreting DS-structures as Choquet capacities, it is highly desirable that the combination gives a capacity that is contained in the robust rule result. It can be shown that the MDS rule, viewed as a capacity, is contained in the robust Bayesian fusion result. This is not true in general for Dempster's rule, however.



Arnborg finally notes that, unlike the robust and rounded robust Bayesian combination operators he proposes, the DS and MDS operators are not monotone with respect to imprecision. Therefore, they either underestimate imprecision or eliminate imprecision in a way that can not easily be defended, whereas the maximum entropy principle can be given a rational game interpretation, and gives a quite different result in many cases. Thus, evidence theory and robust Bayesianism are very different in their conclusions. Further work is needed for understanding the basis for assessing uncertainty objectively, so that a given problem will not have incompatible solutions in the two frameworks.

10. Application of imprecise probabilities in systems modelling

In [21], de Cooman and Troffaes discuss why coherent lower previsions provide a good uncertainty model for solving generic uncertainty problems involving possibly conflicting expert information. They review the definition and meaning of important concepts in imprecise probability models, adding up to a concise and readable introduction to the subject. Finally, they apply their proposed approach to the set of "challenge problems" around which the Sandia workshop was organized [27], arguing that the theory of coherent lower previsions is eminently suited for solving the first set of problems posed in [27].

10.1. Computations and algorithms

Computing posterior upper expectations. In an article by Cozman [25], this problem is discussed and a number of algorithms are compared. The paper focuses on specifying an underlying theory more suitable for computations than the original theory of lower and upper previsions, emphasizing an interpretation of imprecise probabilities that relies on convex sets of probability measures, similar to the quasi-Bayesian theory of Giron and Rios [26].

Coherent assessment by iterated natural extension of imprecise probabilities or previsions. In [22], Dickey presents an interactive open-source computer program which implements *coherent assessment* by iterated *natural extension* (see below), of imprecise probabilities or previsions, conditional and unconditional. The method is based on a generalization of de Finetti's *Fundamental Theorem of Probability*, developed by Lad, Dickey, and Rahman [22][23][24]. He considers previsions of random quantities, loosely, expectations of random variables, a probability being the prevision of an event, or 0-1 random quantity.

Prevision assessments can either be intended as estimates of frequencies, more generally averages, or they can be intended as mere quantitative expressions of human uncertainty. In either case, they should be coherent, that is, extendible to at least one full probability distribution. Thus, estimates of frequencies or averages must not be impossible when interpreted together as limiting frequencies or limiting averages in an experiment. For previsions intended as expressions of uncertainty, coherence is a kind of rationality, a direct generalization of non-contradiction for statements of fact, a self-consistency in the sense that, if taken as a person's betting prices, the person could not be made a sure-loser by combining a finite number of bets at such prices. For a sequence of mathematically related random quantities, if coherent prevision values are given for an initial segment of the sequence, the available cohering values for the prevision of the next quantity comprise an interval whose endpoints can be computed by linear programming. Walley [11] calls this interval the *natural extension* of the given coherent previsions.

The linear-programming variables are interpretable as the probabilities of the "constituent" events, the events of the joint-range points of the random quantities. Coherence restricts the prevision vector of the quantities to the convex hull of the joint-range set, that is, the prevision point must be some weighted average of the joint-range points. The assessed previsions impose additional linear constraints. Coherent previsions are always capable of being extended coherently with the value for any further random quantity assignable in an extend-assess cycle. If supplementary calculations are made of the extension interval for a random quantity of special interest, the interval will be seen to shrink to a subinterval whenever a further coherent prevision is assessed.

The method generalizes to include conditional previsions, as inputs and/or outputs. In addition, since prevision is a linear operator, a linear combination of previsions can be assessed directly as the prevision of a linear combination of random quantities.

The interactive program proceeds in a number of steps, each in the form of an extend-assess cycle:

1. Based on all the prevision bounds assessed so far, the program computes natural extensions, the implied extension interval(s), for the previsions of one or more user-selected quantities.

2. The user assesses a lower and/or upper bound (or a point value) for a prevision, cohering with its computed extension interval. To calculate the extension interval for the unspecified prevision of a quantity, the program must determine the convex hull of the joint range set of the considered quantities, and then impose the linear constraints on the assessed prevision values and bounds.

Arithmetic with uncertain numbers. Ferson and Hajagos [28] solve the Sandia challenge problems [27] using *probability bounds analysis*, a combination of the methods of standard interval analysis [30][31] and classical probability theory [32][33]. Probability bounds analysis is closely allied in spirit with robust Bayes techniques, in particular Bayesian sensitivity analysis. In this approach, an analyst's uncertainty about which prior distribution should be used is expressed by replacing a single prior distribution by an entire class of prior distributions. The analysis proceeds by studying the variety of outcomes as each possible distribution is considered. The inputs are first expressed as interval bounds on cumulative distribution functions. Each uncertain input variable is then decomposed into a list of pairs (interval, probability). A Cartesian product of these lists, reflecting both the independence among inputs and the mathematical expression that binds them together, creates another list, which is recomposed to form the resulting uncertain number as upper and lower bounds on a cumulative distribution function. Ancillary techniques are also employed, such as *condensation*, which is necessary to keep the length of the list from growing inordinately in sequential operations, and *subinterval reconstitution*, which is needed to solve interval arithmetic problems involving repeated parameters. *Moment propagation* formulas are simultaneously used to bound mean and variance estimates accompanying the bounds on the cumulative distribution function. Generalizations of this approach are also described that allow for dependencies other than independence, completely unknown dependence, and model uncertainty more generally.

Based on work by Williamson and Downs [36], who developed an approach that computes rigorous bounds on the cumulative distribution functions of convolutions without having to assume independence between the operands, the methods of probability bounds analysis used by Ferson *et al.* have been implemented in software packages [37][38]. Williamson and Downs used their method to estimate rigorous bounds on the distributions of sums, products, differences and quotients of random variables specified only by their marginal distributions, without any information about the dependence among the variables. They also described a method to compute bounds on distributions under independence assumptions. Because their approach uses interval bounds to represent discretization and dependency errors, it can also account for uncertainty about the shape of input distributions themselves. In the software packages [37][38], these methods have been extended to transformations such as logarithms and square roots, other convolutions such as minimum, maximum and powers, and other dependence assumptions.

Using techniques of mathematical programming, Berleant *et al.* [39][40][41][42] independently derived and implemented algorithms to compute convolutions of bounded probability distributions, both with and without independence assumptions. The application of these methods to the Sandia challenge problems is reported in [29]. Results are of comparable quality to those reported in [28].

A third, interesting and recent, but less mature arithmetic approach is that of Lodwick and Jamison [43], who present a method for estimating and validating the cumulative distribution of a function of random variables (independent or dependent). The method creates a sequence of bounds that will converge to the distribution function in the limit for functions of independent random variables or of random variables of known dependencies. The bounds are possibility and necessity distributions, consistent with the underlying probability distribution. An approximation is constructed from, and contained in, these bounds. Preliminary numerical experiments indicate that this approximation converges quickly to the actual distribution when the number of variables is moderate.

The Sandia challenge problems embody several issues that beset all technologies for uncertainty propagation, whether probabilistic or not, distributional or not, approximative or rigorous. These issues are:

- aggregation information from different sources (such as expert judgements)
- combination of probabilistic and non-probabilistic uncertainty
- repetition of uncertain parameters

According to [28], although probability bounds analysis does not prescribe a general slution for the question of how to aggregate information from disparate sources, it does offer what may be the definitive solution for a problem requiring the combination of probabilistic and non-probabilistic uncertainty. It also provides a workable, albeit computationally intensive, strategy for handling repretitions of uncertain parameters in expressions. When subinterval reconstitution is used to remedy the computational problem introduced by repeated uncertain variables, the results are only asymptotically best possible when the number of subintervals become very large. A comprehensive and flexible methodology to obtain best possible bounds on moments awaits development.

10.2. Summary of approaches to the Sandia "challenge problems"

The Sandia challenge problems are presented in [27] as directly addressing issues in the representation and aggregation of information concerning model parameters. The information can be of different types and from a number of sources, including measurements and expert opinion. Given a representation and aggregation of the information into the vector **x**, the vector is propagated through the simulator *f*. *f* is assumed to be deterministic; that is, for one realization of **x**, there is only one realization of **y** = *f*(**x**), which could be a function of space and/or time. The authors of [27] state: "Given a representation and aggregation of uncertainty through *f*, we also wish to address the issue of how to interpret the resultant uncertainty representation in **y**. ... We believe that a synthesis is needed of the strategies that have been developed separately within various communities... . It is our opinion that if some coherence concerning these simple problem sets can be achieved, then there is some hope that these methods could be extended to realistic target systems."

The two problem sets are given by:

1. A simple algebraic system of the form $y = (a + b)^a$
2. A simple dynamic system in the form of an initial value problem given by an ordinary differential equation.

Most of the results and discussion turned out to focus around the results for the several variants of the algebraic problem set, which is the only one we will present here. Within this problem set, six main variants and some subvariants were proposed, as follows:

Problem 1. a and b are contained in the closed real interval A and B, respectively.

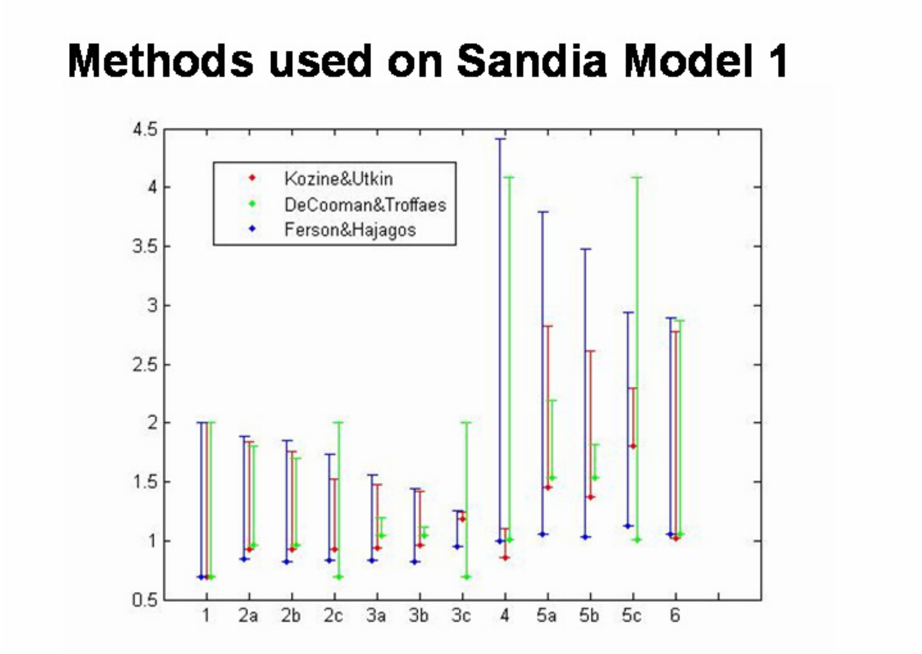
Problem 2. a is contained in the closed interval A, and the information concerning b is given by n independent and equally credible sources. Each source specifies a closed interval B_i of possible values for b. Three subvariants are defined by allowing this set of intervals to be either *consonant* (nested), *consistent* (the intersection of all the intervals is non-empty), or *arbitrary.*

Problem 3. The information concerning both a and b is given by independent and equally credible sources of information, m sources for a and n sources for b. Here, too, consonant, consistent, and arbitrary collections of intervals are considered.

Problem 4. a is contained in the closed interval A, and b is given by a log-normal distribution, $\ln b \sim N(\mu, \sigma)$. The value of the mean, μ, and the standard deviation, σ, are given by the closed intervals M and S, respectively.

Problem 5. The information concerning a is given by m independent and equally credible sources of information. Each source specifies a closed interval A_i that contains the value of a. The information concerning b is given by n independent and equally credible sources of information Each source agrees that b is given by a log-normal probability distribution, however, each source specifies closed intervals, M_j and S_j, of possible values of the mean and the standard deviation, respectively. Consonant, consistent, and arbitrary collections of intervals are considered.

Problem 6. a is contained in the closed interval A and b is given by a log-normal distribution. The values of both the mean and the standard deviation are precisely known.



For each of these cases numerical values of the stated free parameters were also specified, to allow easy comparison between results arising from different approaches.

Table 1: Comparison of bounds on expected values

Problem	Kozine and Utkin [79]	De Cooman and Troffaes [21]	Ferson and Hajagos [28]	Helton et al. [80]
1	[0.69, 2.0]	[0.692201, 2.0]	[0.692, 2]	---
2a	[0.93, 1.84]	[0.956196, 1.8]	[0.84, 1.89]	---
2b	[0.93, 1.76]	[0.956196, 1.7]	[0.82, 1.85]	---
2c	[0.93, 1.52]	[0.692201, 2.0]	[0.83, 1.73]	---
3a	[0.944, 1.473]	[1.04881, 1.2016]	[0.83, 1.56]	---
3b	[0.964, 1.418]	[1.04881, 1.1156]	[0.82, 1.44]	---
3c	[1.187, 1.242]	[0.692201, 2.0]	[0.946, 1.25]	---
4	[0.859, 1.108]	[1.00966, 4.08022]	[0.9944, 4.416]	[1, 3.7]
5a	[1.45, 2.824]	[1.54027, 2.19107]	[1.05, 3.79]	(graph)
5b	[1.373, 2.607]	[1.54027, 1.81496]	[1.03, 3.48]	(graph)
5c	[1.802, 2.298]	[1.00966, 4.08022]	[1.12, 2.94]	(graph)
6	[1.019, 2.776]	[1.05939, 2.86825]	[1.052, 2.89]	[1.05, 3]

Although these problems were extremely simple (and were for that reason heavily criticized by some of the workshop participants), the table shows rather limited quantitative agreement between the different approaches. However, the different result intervals for the same problems presented in the table are indeed consistent, using the terminology of the problem statements.

10.3. Advantages and disadvantages of different approaches

In [38], Ferson gives the following summary of the pros and cons of the various methods being used for environmental risk assessments. Most of these aspects are independent of application, or easily translatable to other application areas.

Deterministic (best estimate) calculation

Why not?
- doesn't express reliability of results

What-if studies and sensitivity analysis

Why?
- very general and flexible approach
- can work for all kinds of uncertainty

Why not?
- cumbersome to design and implement
- computationally expensive (sometimes impossible)
- hard to explain when elaborated

Worst-case analysis

Why?
- accounts for uncertainty by being conservative
- protective of human health and ecosystem integrity
- under ignorance, shifts burden of proof to industry
- especially useful in a screening assessment

Why not?
- level of conservatism not consistent from analysis to analysis
- impossible to compare risks from different analyses
- possibly hyperconservative
- unfair and wasteful regulation that is burdensome to industry
- estimates biased

Interval analysis

Why?
- natural for scientists
- very simple and easy to explain
- generalizes and refines worst-case analysis
- works no matter where uncertainty comes from
- especially useful in a screening assessment

Why not?
- ranges can grow very quickly
- often too conservative
- paradoxical (no exact value, but exact bounds)

Monte-Carlo analysis (classical probability theory)

Why?
- simple to implement
- fairly simple to explain
- characterizes impacts of all possible magnitudes
- can use information about correlations among variables

Why not?
- requires a lot of empirical information (or assumptions)
- analysts usually need to guess some things
- routine assumptions lead to non-protective conclusions
- confounds ignorance with variability
- may be inappropriate for non-statistical uncertainty
- requires careful attention about which population uncertainty refers to
- may not be okay to merge subjective estimates from different sources

Fuzzy arithmetic (possibility theory)

Why?
- computations are simple and easy to explain
- acceptable to assign distributions subjectively
- doesn't require detailed empirical information
- doesn't require knowledge of dependencies or correlations among variables
- maintains conservatism under uncertainty about dependencies among variables
- intermediate in conservatism between analogous Monte Carlo and worst case/interval approaches
- fuzzy numbers are robust representations when empirical information is very sparse
- characterizes impacts of all possible magnitudes
- generalizes and refines interval analysis
- works with non-statistical uncertainty

Why not?
- not yet widely known
- may be overly conservative
- repeated parameters can be a computational problem
- not clear it's okay to merge numbers whose conservatisms are different
- not clear how to merge with Monte Carlo analysis

Probability bounds analysis

Why?
- handles uncertainty about parameter values, distribution shapes, dependencies, and model form
- faithful to frequentist interpretation of probability
- bounds get narrower with better empirical information
- provides quality assurance for Monte Carlo results
- bounds are rigorous

Why not?
- displays must be cumulative
- must truncate infinite tails
- optimal bounds expensive to compute when parameters are repeated
- cannot handle two-dimensional probabilities

Two-dimensional Monte Carlo simulation

Why?
- allows for a comprehensive expression of parametric uncertainty
- can handle model uncertainty in a limited way
- only requires (nested) Monte Carlo methods

Why not?
- cannot handle uncertainty about distribution shape
- can be daunting to parameterize
- calculations can be cumbersome
- confounds frequentist and subjectivist interpretations of probability

Hybrid arithmetic

Why?
- does not confound different kinds of uncertainty
- faithful to both probability and possibility interpretations

Why not?
- optimal results expensive when parameters are repeated
- displays can be complex and difficult to explain to the public
- information about correlations (other than independence) cannot be used to tighten results

11. Some approaches to fusion performance evaluation

Theil, Kester, and Bossé [63] describe several objective measures to characterize the effectiveness of detection, tracking and classification. They propose three categories of measure of performance (MOP) to measure the performance of detection, tracking and classification, respectively. Each category may consist of several different measures of performance. Which measure that is most applicable in each category depends on the task of the fusion algorithms.

A performance evaluation method based on a leave-one-out method is developed by Cremer, de Jong and Schutte [61]. In this method a classifier is trained on all but one sample and tested on the remaining sample. They repeat this process until every sample has been part of the evaluation set. However, the leave-one-out method does not have a way to generate ROC curves by itself. An extension of the method that uses a range of cost functions solves this problem [62].

Goebel [70] used a two phase development method. In a first conceptual phase of tool development, he used the sum of squared error as a simple metric for a Design-of-Experiment phase. In order to confirm that the fusion tool led to actual performance improvement in a second phase, he devised a benchmark algorithm that performed a maximum-win strategy. He developed an overall performance index by weighing false positives (FP), false negatives (FN), and false classified (FC) where the weights were set by the application as $0.6(1-FP)+0.3(1-FN)+0.1(1-FC)$. An increase in performance was measured as the fraction of improvement from that baseline to perfect performance, expressed in percent.

Chang, Song and Liggins [59] present a fusion performance model for a general multisensor fusion system. Their model includes kinematic and classification measures with a focus on positional and classification error. The performance model is based on Bayesian theory and uses a combination of simulation and analytical approaches. In a second paper [60] the same authors focus on distributed tracking and classification. The model provides high level performance bounds given the sensor distribution as well as information on sensor quality for system control. These issues have broad implications for fusion system architecture design, and evaluation of different system alternatives.

The scenario dependence of the performance of a hybrid estimation algorithm is elaborated by Li and Bar-Shalom [57]. They develop a performance predictor approach. This is a computer-aided approach, in which the performance prediction of a stochastic algorithm is made by using a specifically developed deterministic algorithm. Thus, no simulation is used. The accuracy of this method can be much better than a performance analysis approach where relations between a small number of key parameters and the performance of the algorithm is established. Since these relations can not be complicated many assumptions and approximations need to be done.

The cumulative effect of all these assumptions and approximations make it extremely difficult to obtain an analytic model of good accuracy. Performance prediction seems to be the only reasonable alternative to Monte Carlo simulation. In the performance predictor approach very complex relationships may be modeled. A technique for performance prediction of hybrid algorithms is called the *HYbrid Conditional Averaging* (HYCA). This hybrid technique is obtained by conditional expectation operation through which the randomness of the algorithm performance due to uncertainties in the continuos-valued state space is averaged out, whereas the dependence of the performance on the scenario is retained. HYCA can be used to predict the performance of multiple models such as IMMs, but also PDA and Nearest Neighbor filters, etc.

While simulation is one of the most powerful tools for evaluating the performance of complex systems, it is computationally slow. Panayiotou et al. [65] develop a metamodel to overcome this limitation. They generate a model of the system that accurately captures the relationships between input and output. Using this model is much more efficient than simulation. Neural networks (NN) are good function approximators and thus make good metamodels. In a training phase, the NN is presented with several input/output examples, and learn the relationship between inputs and outputs of the simulation model. Thus, the NN has the ability to generalize when presented with novel input and predict the output. As the ability to generalize requires many training examples that are obtained through simulation developing a metamodel is slow. Often it is possible to use perturbation analysis to obtain sensitivity information for the input parameters. Using sensitivity information Panayiotou et al. [65] investigate how to reduce the simulation effort required for training a feed-forward neural network. Using sensitivity information can significantly reduce the number of input/output training examples needed.

Several other authors have also addressed simulation metamodeling. A metamodel for a *Tactical Electronic Reconnaissance Simulation Model* was developed by Zeimer and Tew [66]. It estimates the number of ground-based radar sites detected by a reconnaissance aircraft. Other authors have used statistical analysis. Santos and Nova [67], for instance, used least squares estimation for estimation of non-linear metamodels and Cheng [68] applied regression using Bayesian methods. As NN are generally good function approximators they make good candidates for surrogate functions. Jablunovsky et al. [69] used a back-propagation neural net to capture the behaviour of a Command and Control (C2) network.

Hoffman *et al.* [71] develop an approach based on *Finite Set STatistics* (FISST) for performance evaluation of multisensor-multitarget threat assessment algorithms. Performance is here evaluated by measuring the Kullback-Leibler distance between ground truth and a contour map from the threat prediction output.

McGirr [64] discusses resources available for the design of interoperable systems. He differentiates between lower level *Measures of Performance* (MOPs) and higher level *Measures of Evaluation* (MOEs). An MOE can be the timeliness of information to support a decision. The performance evaluation process typically incorporates a combination of MOPs and MOEs such as detailed analysis, Monte Carlo simulation, laboratory based testing, and operational evaluation.

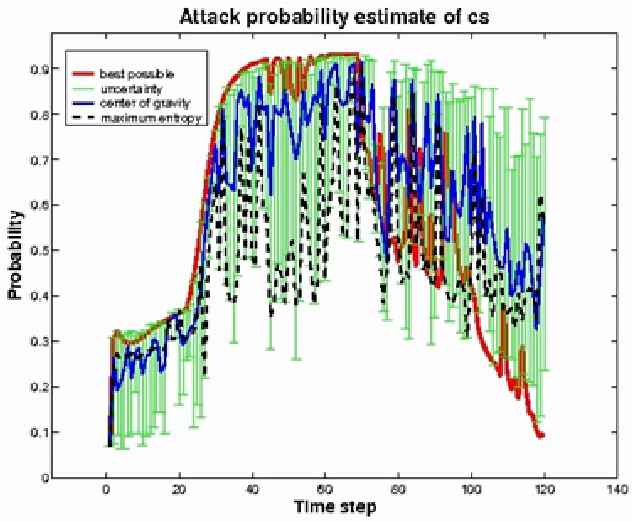
12. Examples

Set-valued estimation offers a way to account for imprecise knowledge of the prior distribution of a Bayesian statistical inference problem. The set-valued Kalman filter [73], which propagates a set of conditional means corresponding to a convex set of conditional probability distributions of the state of a linear dynamic system, is a general solution for linear Gaussian dynamic systems. In [74], the set-valued Kalman filter is extended to the non-linear case by approximating the non-linear model with a linear model that is chosen to minimize the error between the non-linear dynamics and observation models and the linear approximation. The conventional extended Kalman filter is a well-accepted and practical solution for point-valued estimates, but it does not apply to the set-valued case. The extended set-valued Kalman filter provides an approximate solution to the non-linear set-valued dynamic state estimation problem that is computationally feasible. An application is presented to illustrate and interpret the estimator results. However, the paper does not provide arguments in support of this approach as a means to achieve fusion robustness, nor are bounds on the approximate solution presented.

Another approach to generalizing the Kalman filter to fit an imprecise probability framework, in this case TBM, is presented in [75].

None of these papers discuss the use of set-valued filters to achieve robustness or probabilistic error bounds. Perhaps the TBM-based algorithm could be combined with probability bounds analysis to achieve this [28, 76].

In a manuscript submitted to the Fusion 2005 conference [78], our colleagues Ronnie Johansson and Robert Suzic study information acquisition for robust plan recognition. Tactical commanders want to obtain predictive situation awareness. To do this effectively, they need real-time decision support tools which can both recognize basic tactical plans of the opponent, such as an imminent attack, and proactively control limited sensor resources by prioritizing dangerous plan alternatives in a sensible way. Johansson and Suzic introduce a particle filter that maintains a state estimate even when observations are lacking. The particle filter produces a multi-model state representation with each particle as a mode.



Initially, the suspected attacker (cs) is observed by both UAVs and observers on the ground. Estimated attack probability is close to best possible and uncertainty is small. As estimated attack probability increases, this attracts the interest of the sensor control, which manages to keep the uncertainty relatively small by prioritizing this objective. Near the end of the scenario, attack probability has decreased (because cs is moving away from the own targets) and the interest of the sensor control has been lowered.

They compare two approaches experimentally, one (`cg`) which is based on finding the mean value (center of gravity) of the probability for each plan alternative given all plan distributions, the other (`me`) by choosing the estimate that has the maximum entropy compared to other distributions. As a baseline, they also compute the best possible (`bp`) estimates of the attack probability by assuming continuous and accurate observations by all agents. Initially, the suspected attacker (`cs`) is observed by both UAVs and observers on the ground. Estimated attack probability is close to `bp` and the uncertainty is small. As estimated attack probability increases, this attracts the interest of the sensor control, which manages to keep the uncertainty relatively small by prioritizing this objective. Near the end of the scenario, attack probability has decreased (because `cs` is moving away from the own targets) and the interest of the sensor control has been lowered. Together with a built-in "gravity" bias of the model which pulls particles towards own targets, this explains why the uncertainty interval fails to cover the `bp` estimate during the last time steps. In this experiment, the `cg` approach appears to approximate `bp` better than does `me`.

13. Conclusion

Although imprecise probabilities is still an area of active research, it seems that its methods have reached a level of maturity that is sufficient for development of robust test applications in our own field of application, information fusion.

14. References

[1] Rios Insua, D., F. Ruggeri (eds.) (2000). *Robust Bayesian Analysis*. Lecture notes in statistics 152, Springer-Verlag, New York, NY.

[2] French, S. and D. Rios Insua (2000). *Statistical Decision Theory*. Arnold.

[3] Rios Insua, D., Robustness Analysis (2004). A Bayesian point of view. www.inescc.pt/~ewgmcda/ForDRios.html

[4] Ruggeri, F., D. Rios Insua, J. Martin (2004). Robust Bayesian Analysis, in Dey (ed.) *Handbook of Statistics: Bayesian Analysis*, North-Holland, Amsterdam.

[5] Berger, J. O., D. Rios Insua, F. Ruggeri (2000). Bayesian Robustness, in [1], 1-32.

[6] Berger, J. O. (1993) An Overview of Robust Bayesian Analysis, Technical Report #93-53C, Dept. of Statistics, Purdue University.

[7] Betro, B., A. Guglielmo (2000). Methods for global prior robustness under generalized moment conditions, in [1], 273-293.

[8] Lavine, M., M. Perone Pacifico, G. Salinetti, G. Tardella (2000). Linearization techniques in Bayesian robustness, in [1], 261-272.

[9] MacEachern, M., P. Mueller (2000). Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models, in [1], 295-315.

[10] Fixsen, D., R. P. S. Mahler (1997) The modified Dempster-Shafer approach to classification. *IEEE Trans. SMC-A*, 27(1):96-104.

[11] Walley, P. (1991) *Statistical reasoning with Imprecise Probabilities*, Chapman & Hall, London.

[12] Walley, P. (1996) Measures of uncertainty in expert systems, *Artificial Intelligence* 83, 1-58.

[13] Walley, P. (2000) Towards a unified theory of imprecise probability, in [18], 125-148.

[14] Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.

[15] Zadeh, L. A. (1986) Is probability theory sufficient for dealing with uncertainty in AI: a negative view, in: L. N. Kanal, J. F. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, 103-116.

[16] Arnborg, S. (2004) Robust Bayesianism: Imprecise and Paradoxical Reasoning, in P. Svensson and J. Schubert (eds.) *Proceedings of the 7th Int. Conf. on Information Fusion*, 407-414, International Society of Information Fusion, Mountain View, CA.

[17] Weichselberger, K. (2000) The theory of interval-probability as a unifying concept for uncertainty, in [18], 149-170.

[18] Cozman, F. C., Moral, S. (eds.) (2000) Reasoning with imprecise probabilities, special issue of *Int. J Approx Reasoning* 24, 121-299.

[19] Helton, J. C., Oberkampf, W. L. (eds.) (2004) Alternative Representations of Epistemic Uncertainty, special issue of *Reliability Engineering & System Safety* 85, 1-369.

[20] Shafer, G., Gillett, P. R., Scherl, R. B. (2003) A new understanding of subjective probability and its generalization to lower and upper prevision, *Int J Approx Reasoning* 33, 1-49.

[21] de Cooman, G., Troffaes, M. C. M. (2004) Coherent lower previsions in systems modelling: products and aggregation rules, in [19], 113-134.

[22] Dickey, J. (2003), Convenient interactive computing for coherent imprecise prevision assessments, in *Proceedings of Third International Symposium on Interval Probabilities and Their Applications (ISIPTA '03)*.

[23] Lad, F., Dickey, J., Rahman (1990), The fundamental theorem of prevision. *Statistica*, 50, 15-38.

[24] Lad, F., Dickey, J., Rahman (1992), Numerical application of the fundamental theorem of prevision. *J. of Statistical Computation and Simulation* 40, 135-151.

[25] Cozman, F. G. (2000) Computing posterior upper expectations, in [18] 191-205.

[26] Giron, F. J., S. Rios (1980) Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. deGroot, D. V. Lindley, A. F. M. Smith (eds.) *Bayesian Statistics*, University Press, Valencia, 17-38.

[27] Oberkampf, W. L., J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, S. Ferson (2004) Challenge problems: uncertainty in system response given uncertain parameters, in [19] 11-19.

[28] Ferson, S., J. G. Hajagos (2004) Arithmetic with uncertain numbers: rigorous and (often) best possible answers, in [19] 135-152.

[29] Berleant, D., J. Zhang (2004) Representation and problem solving with Distribution Envelope Determination (DEnv), in [19] 153-168.

[30] Moore, R. E. (1966) *Interval analysis*. Prentice Hall, Englewood Cliffs, NJ.

[31] Moore, R. E. (1979) *Methods and applications of interval analysis*, SIAM, Philadelphia, PA.

[32] Feller, W. (1968) *An introduction to probability theory and its applications*, vol. 1, Wiley, New York, NY.

[33] Feller, W. (1971) *An introduction to probability theory and its applications*, vol. 2, Wiley, New York, NY.

[34] Jaynes, E.T. (2003) *Probability theory. The logic of science*, University Press Cambridge, UK.

[35] Berger, J. O. (1980) *Statistical decision theory and Bayesian analysis*, Springer-Verlag New York, Inc.

[36] Williamson, R. C., Downs, T. (1990) Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds, *Int J Approx Reasoning* 4, 89-158.

[37] Ferson, S, Long, T. F. (1995) Conservative uncertainty propagation in environmental risk assessments, in Hughes, J. S., Biddinger, G. R., Mones, E. (eds.) *Environmental toxicology and risk assessment*, ASTM STP 1218, vol. 3, 97-110, ASTM, Philadelphia, PA.

[38] Ferson, S. (2002) *RAMAS Risk Calc 4.0 software: risk assessment with uncertain numbers*, Lewis Publishers, Boca Raton, FL.

[39] Berleant, D. (1993) Automatically verified reasoning with both intervals and probability density functions, *Interval Computing* 2, 48-70.

[40] Berleant, D. (1996) Automatically verified arithmetic on probability distributions and intervals, in Kearfott, B., Kreinovich, V. (eds.) *Applications of interval computations*, 227-244, Kluwer Publications, Dordrecht.

[41] Berleant, D., Cheng, H. A. (1998) A software tool for automatically verified operations on intervals and probability distributions, *Reliable Computing* 4, 71-82.

[42] Berleant, D., Goodman-Strauss, C. (1998) Bounding the results of arithmetic operations on random variables of unknown dependency using intervals, *Reliable Computing* 4, 147-165.

[43] Lodwick, W. A., Jamison, K. D. (2003) Estimating and validating the cumulative distribution of a function of random variables: toward the development of distribution arithmetic, *Reliable Computing* 9, 127-141.

[44] Balci, O. (2001) A Methodology for Certification of Modeling and Simulation Applications. *ACM Trans. on Modeling and Computer-based Simulation* 11(4), 352-377.

[45] Brade, D. (2003) *A generalized process for the verification and validation of models and simulation results*. Ph D thesis, Universität der Bundeswehr, Munich.

[46] Kleijnen, J. P. C. (1995) Verification and validation of simulation models, *European Journal of Operational Research* 82, 145-162.

[47] Kleijnen, J. P. C. (1998) Experiment design for sensitivity analysis, optimization, and validation of simulation models. In J. Banks (ed), *The Handbook of Simulation*, Wiley, New York, 1998.

[48] Kleijnen, J. P. C. (1999) Validation of models: statistical techniques and data availability. In *Proc. of the 1999 Winter Simulation Conference*, 647-654, Society for Computer Simulation International.

[49] Sargent, R. G.. (1999) Validation and verification of simulation models. In *Proc. of the 1999 Winter Simulation Conference*, 39-48, Society for Computer Simulation International.

[50] Kleijnen, J. P. C. and R. G. Sargent (2000) A methodology for fitting and validating metamodels in simulation, *European Journal of Operational Research* 120, 14-29.

[51] Saltelli, A., K. Chan, and E. M. Scott, Eds (2000) *Sensitivity Analysis*, Wiley, West Sussex, UK.

[52] Frey, H. C. and S. R. Patil (2002) Identification and review of sensitivity analysis methods, *Risk Analysis* 22(3), 553-578.

[53] Zadeh, L. (1986) A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination, *AI Magazine* 7(2), 85–90.

[54] Shafer, G. (1990) Perspectives on the theory and practice of belief functions, *International Journal of Approximate Reasoning* 4(5/6), 323–362.

[55] Haenni, R. (2004) Is DSmT a reasonable alternative to DST?, Panel presentation at the *Seventh International Conference on Information Fusion*, 1 July 2004, Stockholm.

[56] Walley, P., R. Pelessoni, and P. Vicig (1999) Direct algorithms for checking coherence and making inferences from conditional probability assessments, *Quaderni del Dipartimento di Matematica Applicata alle Scienze Economiche. Statistiche e Attuariali "B. de Finetti"* 6.

[57] Li, X.R., and Y. Bar-Shalom (1995) Performance Prediction of Hybrid Algorithms, in *Control and Dynamic Systems: Advances in Theory and Applications,* (C.T. Leondes, ed.), Vol. 72, pp. 99–151, New York: Academic Press, 1995.

[58] Ben-Haim, Y. (2001) *Information-gap Decision Theory - Decisions Under Severe Uncertainty,* Academic Press, London.

[59] Chang, K.C., Y. Song, and M. Liggins (2003) Performance Modeling for Multisensor Data Fusion, in Proceedings of SPIE Vol. 5096 *Signal Processing, Sensor Fusion, and Target Recognition XII,* (I. Kadar, ed.), SPIE, Bellingham, Washington, USA, pp. 354–363.

[60] Chang, K.C., Y. Song, and M. Liggins (2003) Fusion Performance Model for Distributed Tracking and Classification, in *Proceedings of the Sixth International Conference on Information Fusion,* Cairns, Australia, pp. 567–574.

[61] Cremer, F., W. de Jong, and K. Schutte (2003) Fusion of polarimetric infrared features and GPR features for landmine detection, in *Second International Workshop on Advanced Ground Penetrating Radar*, Delft, The Netherlands, pp. 222–227.

[62] Cremer, F., W. de Jong, and K. Schutte (2003) Processing of polarimetric infrared images for landmine detection, in *Second International Workshop on Advanced Ground Penetrating Radar*, Delft, The Netherlands, pp. 216–221.

[63] Theil, A., L. Kester, and É. Bossé (2000) On measures of performance to assess sensor fusion effectiveness, in *Proceedings of the Fifth International Conference on Information Fusion,* Paris, France, pp. ThB5/3–7.

[64] McGirr, S.C. (2001) Resources for the design of data fusion systems, in *Proceedings of the Fourth Annual Conference on Information Fusion,* Montréal, Québec, Canada, pp. FrB2/9–15.

[65] Panayiotou, C.G.., C. G.. Cassandras, and W.-B. Gong (2000) Model abstraction for discrete event systems using neural networks and sensitivity information, in *Proceedings of the 32nd Conference on Winter Simulation,* Orlando, Florida, USA, pp. 335–341.

[66] Zeimer, M.A., and J. D. Tew (1995) Metamodel applications using TERSM. In *Proceedings of 27th Winter Simulation Conference*, Arlington, Virginia, USA, pp. 1421–1428.

[67] Santos, M.I.R. d., and A. M. O. P. Nova (1999) The main issues in nonlinear simulation metamodel estimation. In *Proceedings of the 31st Conference on Winter Simulation*, Phoenix, Arizona, USA, pp. 502–509.

[68] Cheng, R.C.H. (1999) Regression metamodeling in simulation using Bayesian methods. In *Proceedings of The 31st Conference on Winter Conference*, Phoenix, Arizona, USA, pp. 330–335.

[69] Jablunovsky, G.., C. Dorman, and P. Yaworsky (2000) A neural network sub-model as an abstraction tool: Relating network performance to combat outcome, in Proceedings of SPIE Vol. 4026 *Enabling Technology for Simulation Science IV,* (A.F. Sisti, ed.), SPIE, Bellingham, Washington, USA, pp. 192–200.

[70] Goebel, K. (2003) Sensitivity of fusion performance to classifier model variations (2003) *Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* 15, 335–348.

[71] Hoffman, J., E. Sorensen, C. Stelzig, R. Mahler, A. El-Fallah, and Alford, M. (2002) Scientific performance estimation of robustness and threat, in Proceedings of SPIE Vol. 4729 *Signal Processing, Sensor Fusion, and Target Recognition XI,* (I. Kadar, ed.), SPIE, Bellingham, WA, 2002, pp. 248-258.

[72] Augustin, T. (2003) On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: A cautionary note on updating imprecise priors, in *Proceedings of Third International Symposium on Interval Probabilities and Their Applications (ISIPTA '03)*.

[73] Morrell, D. R. and W. C. Stirling (1991) Set-valued filtering and smoothing. *IEEE Trans. SMC* 21(1), pp. 184-193.

[74] Morrell, D. R. and W. C. Stirling (2003) An extended set-valued Kalman filter, in *Proceedings of Third International Symposium on Interval Probabilities and Their Applications (ISIPTA '03)*.

[75] Smets, P. and B. Ristic (2004) Kalman filter and joint tracking and clasification in the TBM framework, in P. Svensson and J. Schubert (eds.) *Proceedings of the 7th Int. Conf. on Information Fusion*, 46-53, International Society of Information Fusion, Mountain View, CA.

[76] Ferson, S., V. Kreinovich, L. Ginzburg, D. Myers, and K. Sentz (2003) Constructing probability boxes and Dempster-Shafer structures. SAND2002-4015, Sandia National Laboratories, Albuquerque, NM, USA.

[77] Farquhar, P. H. (1984) Utility assessment models. *Management Science*, 30, 1283-1300.

[78] Johansson, L. R. M. and R. Suzic (2005) Particle filter-based information acquisition for robust plan recognition, s*ubmitted*.

[79] Kozine, I. O. and L. V. Utkin (2004) An approach to combining unreliable pieces of evidence and their propagation in a system response analysis, in [19] 103-112.

[80] Helton, J. C., J. D. Johnson, and W. L. Oberkampf (2004) An exploration of alternative approaches to the representation of uncertainty in model predictions, in [19] 39-71.