Evaluation, analysis and synthesis of multiple source information: an application to nuclear computer codes.

Sebastien Destercke & Eric Chojnacki

Institut de Radioprotection et de Sûreté Nucléaire, BP3, 13115, St Paul-lez-Durance, Cadarache, France

ABSTRACT: in this paper, we're interested in the problem of evaluating, analyzing and synthesizing information delivered by multiple sources about the same badly known variable. We focus on two approaches that can be used to solve the problem, a probabilistic and a possibilistic one. They are first described and then applied to the results of uncertainty studies performed in the framework of the OECD BEMUSE project. Usefulness and advantages of the proposed methods are discussed and emphasized in the lights of obtained results.

1 INTRODUCTION

In the field of nuclear safety, the value of many variables are tainted with uncertainty. This uncertainty can be due to a lack of knowledge, of experimental value or simply because a variable cannot be directly observed and must be evaluated by some mathematical models. Two common problems encountered in such situations are the following:

- 1. The difficulty to build synthetic representations of our knowledge of a variable;
- 2. The need to compare, analyse and synthesize results coming from different mathematical models modeling a common physical phenomenon.

Both issues can be viewed as problems of information fusion in presence of multiple sources. In the first case, the information can come from multiple experts, sensors, or from different experimental results. Taking into account these multiple sources to model input uncertainty is therefore desirable. In the second case, the output of each single mathematical model or computer code can be considered as a single source of information. The synthesis and analysis of the different outputs can then be treated as an information fusion problem.

Both probability and possibility theories offer formal frameworks to evaluate, analyze and synthesize multiple sources of information. In this paper, we remind the basics of each methodology derived from these theories and then apply them to the results of the BEMUSE (Best Estimate Methods - Uncertainty and Sensitivity Evaluation) OECD/CSNI program (OCDE 2007), in which the IRSN paticipated.

The rest of the paper is divided in two main sections. Section 2 details the ideas on which are based the methods and then gives some basics about the formal settings of the two approaches. Section 3 describes the BEMUSE program and the application of each methodology to its results. The benefits of using these approaches are then discussed in the lights of the obtained results.

2 METHODS

Most of the formal approaches proposing to handle information provided by multiple sources consist in three main steps: modeling the information, evaluating the sources by criteria measuring the quality of the provided information, and synthesizing the information.

In this paper, we will only remind the basic ideas of each methodology, and will focus on the results obtained with the BEMUSE program. More details are given in an extended paper (Destercke and Chojnacki 2007), while the probabilistic and possibilistic approaches are fully motivated respectively in (Cooke 1991) and in (Sandri, Dubois, and Kalfsbeek 1995), respectively.

2.1 Modeling information

In this paper, we consider that the information is provided in term of percentiles, which are surely the commonest type of probabilistic information encountered in safety studies. Other kinds of information cover characteristics of distribution (mean, median, ...), comparative assessments, (see (Cooke 1991) and (Walley 1991, ch 4.) for extensive reviews).

Classical approaches consist of singling out a probability distribution that correspond to the information, usually by maximizing an information measure (e.g., the entropy). Nevertheless, many arguments converge to the fact that single probabilities cannot adequately account for incompleteness, imprecision or unreliability in the information (See (Ferson and Ginzburg 1996) for a short discussion). Other uncertainty theories, such as possibility theory, allows to explicitly account for such features of the information. Such theories are less precise than probability theory, but ensure that no extra assumptions are added to the available information.

The probability distribution that fits a set of percentiles $q_{k\%}^{1}$ and maximize entropy simply corresponds to a linear interpolation between percentiles. Figure 1 represents a cumulative distribution function (CDF) corresponding to the peak clad temperature (Maximal temperature value reached during an accidental transient phase) temperature of a fuel rode in a nuclear reactor core for which available information is $q_{0\%} = 500 K$, $q_{5\%} = 600 K, q_{50\%} = 800 K, q_{95\%} = 900 K$, $q_{100\%} = 1000 K$. The corresponding probability density is pictured in dashed lines.



Figure 1: Examples of probabilistic modeling



Figure 2: Examples of possibilistic modeling

A possibility distribution (Dubois and Prade 1988) over the reals is formally defined as a mapping $\pi : \mathbb{R} \to [0, 1]$. For a given value $\alpha \in [0, 1]$, the (strict) α -cut of π is defined as the set $\pi_{\alpha} = \{x \in \mathbb{R} | \pi(x) > \alpha\}$. Given a possibility distribution π , possibility II and necessity N measures of an event A are respectively defined as:

$$\Pi(A) = \max_{x \in A} \pi(x) \text{ and } N(A) = 1 - \pi(A^c)$$

with A^c the complements of A. We have, for any event A, $N(A) \leq \Pi(A)$, and possibility and necessity measure are respectively interpreted as upper and lower confidence level given to an event. They can be compared to classical probabilities, where the confidence in an event is given by a single (precise) measure. Actually, the possibility and necessity measures can be interpreted as lower and upper probability measures (Dubois and Prade 1992), thus defining a set \mathcal{P}_{π} of probability distributions such that

$$\mathcal{P}_{\pi} = \{ P | \forall A \subseteq \mathbb{R}N(A) \le P(A) \le \Pi(A) \}$$

where P are probability measures over R. This set of probabilities is also related to α -cuts in the following sense

$$\mathcal{P}_{\pi} = \{ P | \forall \alpha \in [0, 1], \ P(\pi_{\alpha}) \ge 1 - \alpha \}.$$

This relation indicates that possibility distributions allows to model information given in terms of nested intervals associated to confidence levels (the narrower the interval, the less the confidence in it).

¹the percentile $q_{k\%}$ of the probability distribution P of a variable X is the deterministic value x s.t. $P(X \le x) = k\%$

It can thus model information given by a finite number of percentiles, as well as cases where we have partial information about characteristics of an unknown distribution (e.g. mean, percentiles, mode, ..., see (Baudrit and Dubois 2006)). Figure 2 represents a possibility distribution corresponding to the peak clad temperature of a fuel rode in a nuclear reactor core where information consists of four intervals [750 K, 850 K], [650 K, 900 K], [600 K, 950 K], [500 K, 1000 K] which have respective confidence levels of 10%, 50%, 90% and 100%.

2.2 Evaluating sources

Once information has been given by a source, it is desirable to evaluate the quality of this information and of the sources. In each approach, this quality is given by two numerical values computed on the basis of rationality requirements:

Informativeness (Inf): evaluate the precision of the information, by comparing it to the model representing ignorance. The more informative is a source, the more useful the information is, and the higher is the associated score.

Calibration (Cal): evaluate the coherence between the provided information and some observed experimental values. The higher this coherence with observed values, the higher the calibration score. Variables on which are computed calibration are called seed variables (that is, variables for which sources have given information and for which experimental data are or will be available)

In the probabilistic approach, informativeness and calibration are computed by the means of the Kullbach-Leibler (KL) divergence, which can be interpreted as a distance between two probabilities. The informativeness is obtained by comparing the probability distribution p_X derived from the source information to the uniform probability distribution u_X defined on the whole variation domain of the variable. Calibration is obtained by comparing probability p_X to an empirical distribution r_X built from the observations. If distributions are discretized in *B* elements, then the KL divergence used to compute informativeness and calibration of a source respectively read:

$$I(p, u) = \sum_{i=1}^{B} p_i \log\left(\frac{p_i}{u_i}\right)$$

and

$$I(r,p) = \sum_{i=1}^{B} r_i \log\left(\frac{r_i}{p_i}\right)$$

And are then transformed to obtain, for all sources, non-negative scores summing up to one. In the probabilistic approach, calibration is based on a convergence argument and requires about 10 experiment to ensure a good stability. It is argued by Sandri *et al.* (Sandri, Dubois, and Kalfsbeek 1995) that the probabilistic approach tends to confuse variability and imprecision.

In the possibilistic approach, informativeness is evaluated by comparing the distribution built from the source information to the interval covering the whole variation domain of a variable. Calibration is simply the extent to which experimental value are judged plausible by the built distribution. In this case, no argument of convergence is used. Let X_r denote the variation domain of a variable X, I_{X_r} the indicator function of X_r (i.e., has value one in X_r , zero elsewhere), and π_X the possibility distribution built from the source information. Informativeness is given by:

$$I(\pi_X) = \frac{\int_{X_r} (I_{X_r} - \pi_X) dx}{\int_{X_r} I_{X_r} dx}$$

and if x^* denote the observed value for X, calibration score $C(\pi_X)$ is simply given by the value $\pi_X(x^*)$ (the upper confidence degree given to x^* .

Once calibration and informativeness scores for every source and for all variables are computed, these scores are then normalized so that they are non-negative and sums up to one.

2.3 Synthesizing the information

Synthesizing the information consists of aggregating multiple models built from the information given by different sources to get a single models. This model can be used in subsequent treatments or analyzed to get information about the sources. Three main kinds of operators are usually used:

Conjunction : equivalent to set intersection. Supposes that all sources are reliable. Conjunction gives poorly reliable results in case of disagreement between sources, but allows to detect such disagreement.

Disjunction : equivalent to set union. Supposes that at least one source is reliable. Disjunction gives reliable results that are often very

imprecise (hence of limited usefulness).

Arithmetic mean : equivalent to a statistical counting of the sources. Supposes that sources are independent, and gives a result that is between conjunction and disjunction. With this operator, sources can also be weighted by scores obtained during the evaluation phase.



Figure 3: Probabilistic synthesis illustration



Figure 4: Possibilistic synthesis illustration

Disjunctive and conjunctive operators are not applicable to the probabilistic approach, and it is commonly recognized that the weighted arithmetic mean is the best approach to aggregate probability distributions. We don't consider Bayesian methods here, because we do not assume we have prior information (see (Clemen and Winkler 1999) for a recent review of such methods). Let p_1, \ldots, p_N be the probability distributions corresponding to the information delivered by N different sources, and $\lambda_1, \ldots, \lambda_N$ be the non-negative weights summing to one attached to these sources (possibly provided by the evaluation procedure briefly described in Section 2.2). The probability distribution p_{Σ} obtained by arithmetic weighted mean is:

$$p_{\sum} = \sum_{i=1}^{N} \lambda_i p_i$$

This is not the case for the possibilistic approach, for which conjunctive (π_{\cap}) , disjunctive operators (π_{\cup}) and the arithmetic mean (π_{mean}) are

well defined, allowing for a greater flexibility in the synthesis and analysis. Let π_1, \ldots, π_N be the probability distributions corresponding to the information delivered by N different sources, and $\lambda_1, \ldots, \lambda_N$ be the non-negative weights summing to one attached to these sources (possibly provided by the evaluation procedure briefly described in Section 2.2). Then, classical conjunctions, disjunctions and arithmetic mean are given, for all $x \in \mathbb{R}$, by:

$$\pi_{\cap}(x) = \min_{i=1,\dots,N} \pi_i(x) \tag{1}$$

$$\pi_{\cup}(x) = \max_{i=1,\dots,N} \pi_i(x) \tag{2}$$

$$\pi_{\sum}(x) = \sum_{i=1}^{N} \lambda_i \pi_i(x) \tag{3}$$

Note that the above conjunctive and disjunctive operators belongs to a broad family of mathematical operators respectively called t-norms and tconorms (Klement, Mesiar, and Pap 2000).

3 APPLICATION TO BEMUSE BENCH-MARK

To show the usefulness and potential applications of the methodology, we apply them to the results of the BEMUSE (Best Estimate Methods - Uncertainty and Sensitivity Evaluation) programme (OCDE 2007) performed by the NEA (Nuclear Energy Agency). Our study will focus on the results of the first step of the programme, in which nine organisations were brought together in order to compare their respective uncertainty analysis with experimental data coming from the experiment L2-5 performed on the loss-of-fluid test (LOFT) facility, for which an accidental transient was simulated.

We will focus on four scalar variables for which each participant had to provide a lower bound (Low), a reference value (Ref) and an upper bound (Upp). These variables are the first (PCT1) and second (PCT2) peak clad temperature (respectively corresponding to the peak of the blowdown and of the reflood phase), the time of accumulator injection (T_{inj}) and the time of complete quenching (T_q) . These four variables are amongst the more critical values that have to be surveyed in case of nuclear accident (this is particularly true for the peak clad temperatures). Values resulting from the

	1PCT(K)			2PCT(K)			T_{inj} (s)			T_q (s)		
	Low	Ref	Up	Low	Ref	Up	Low	Ref	Up	Low	Ref	Up
CEA	919	1107	1255	674	993	1176	14.8	16.2	16.8	30	69.7	98
GRS	969	1058	1107	955	1143	1171	14	15.6	17.6	62.9	80.5	103.3
IRSN	872	1069	1233	805	1014	1152	15.8	16.8	17.3	41.9	50	120
KAERI	759	1040	1217	598	1024	1197	12.7	13.5	16.6	60.9	73.2	100
KINS	626	1063	1097	608	1068	1108	13.1	13.8	13.8	47.7	66.9	100
NRI1	913	1058	1208	845	1012	1167	13.7	14.7	17.7	51.5	66.9	87.5
NRI2	903	1041	1165	628	970	1177	12.8	15.3	17.8	47.4	62.7	82.6
\mathbf{PSI}	961	1026	1100	887	972	1014	15.2	15.6	16.2	55.1	78.5	88.4
UNIPI	992	1099	1197	708	944	1118	8.0	16.0	23.5	41.4	62.0	81.5
UPC	1103	1177	1249	989	1157	1222	12	13.5	16.5	56.5	63.5	66.5
Exp. Val.		1062			1077			16.8			64.9	

Table 1: Results of the BEMUSE program.

uncertainty studies achieved by each participant are summarized in Table 1.

For each participant and each variable, the chosen probabilistic model was to take the lower bound as $q_{1\%}$, the reference value as $q_{50\%}$ (median) and the upper bound as $q_{99\%}$. The possibilistic model was taken as $\pi(Low) = \pi(Upp) = 0.02$ (98% confidence interval), $\pi(Ref) = 1$ (most plausible value). Figure 5 illustrates both models built from the information of NRI2 concerning the second PCT.



Figure 5: Probability and possibility dist. of NRI1 for the 2PCT

3.1 Evaluation

Table 2 shows the results of the evaluation steps performed on the results of the BEMUSE programme, with the models described above. From a methodological point of view, we can notice that the scores and the ranking between sources are globally in agreement, even if there are some differences coming from the differences between formalisms. From a practical standpoint, interesting things can be said from the analysis of results. First, our results are in accordance with informal observations made in previous reports (OCDE 2007): PSI and UNIPI have high informative scores, which reflects their narrow uncertainty bands, and have very low calibration scores, due to the fact that, for each of them, two experimental values are outside interval [Low, Upp]. This consistency between conclusions drawn from our methods and informal observations confirms that using formal methods to analyze information is meaningful.

Another noticeable result is that participants using the same code can have very different scores (both high and low, e.g. global scores of RELAP5 users can range from 0.025 to 0.59), which illustrates and confirms the well-known user influence on the result of a given computer code. Also note that, since scores are built to be directly comparable between them, they can also be used as code validation tools (the better the global result, the better the information delivered by the code). We will see in the next section that using the results of the evaluation can improve the results of the synthesis.

3.2 Synthesis

Figure 6 shows some results of the synthesis for the PCT2. Since this variable is of critical importance in accidental transient and is difficult to estimate, it is of particular interest in the current problem.

Figure 6.A shows the synthetic probabilities when we consider subgroup of participant using the same code. This figures indicate that, while

	0		(/	
	Pr	ob. app	roach	Poss. approach			
Participant	Used code	Inf.	Cal.	Global	Inf.	Cal.	Global
CEA	CATHARE	0.77	0.16	0.12	0.71	0.55	0.40
GRS	ATHLET	1.23	0.98	1.21	0.84	0.52	0.44
IRSN	CATHARE	0.98	0.75	0.73	0.73	0.83	0.60
KAERI	MARS	0.68	0.16	0.11	0.70	0.48	0.34
KINS	RELAP5	1.29	0.16	0.21	0.72	0.67	0.49
NRI1	RELAP5	0.79	0.75	0.59	0.75	0.63	0.47
NRI2	ATHLET	0.79	0.13	0.10	0.78	0.72	0.56
\mathbf{PSI}	TRACE	1.6	0.004	0.008	0.88	0.25	0.22
UNIPI	RELAP5	0.53	0.75	0.4	0.69	0.67	0.46
UPC	RELAP5	1.44	0.02	0.025	0.87	0.28	0.24

 Table 2:
 Scores resulting from evaluation (Inf.: informativeness ; Cal.: Calibration)

CATHARE and RELAP5 users seem to underestimate the experimental value, ATHLET users tend to overestimate it. Figure 6.B shows the benefits of weighting sources or of selecting a subgroup of sources judged better by the evaluation step. Such a selection and weighting shift the curves towards the experimental value (resulting in a better global calibration) and tighten their uncertainty bounds (resulting in a better global informativeness). We also see that the arithmetic mean tends to average the result, and that using probabilistic modeling do not allow us to see possible disagreements between sources. This can be problematic, since it is often desirable to detect and investigate the sources of such disagreements, particularly when synthesis tools are used to analyze the information.

Figure 6.C and 6.D show synthetic possibility distributions resulting from the application of a conjunctive operator (Equation (1)). In this case, the disagreement between sources of a particular subgroup is directly visible, both graphically and quantitatively (disagreement is measured by the maximal height of a distribution: the lower the distribution, the higher the disagreement). We can thus see that information given by ATHLET users are more conflicting than those given by CATHARE users (this could be explained by the higher number of input data parameters in ATH-LET code). Similarly, Figure 6.D shows that all sources strongly disagreeing when considered as a whole, but that the best sources globally agree together, and that taking only their information into account gives a more reliable synthesis.

Figure 6.E and 6.F respectively illustrate the synthetic possibility distributions resulting from the application of the disjunction (Equation (2)) and of the arithmetic weighted mean (Equation (3)) over all sources. Anew, we can see on Fig-

ure 6.F that the arithmetic mean averages the result, thus smoothing the resulting curves. Figure 6.E well illustrates the potential high imprecision resulting from the disjunction. Although the resulting uncertainty model is reliable, its informative content appears of poor interest (e.g., the 50% confidence interval for the 2PCT temperature is [800, 1200], which is very broad).

4 CONCLUSIONS

We have applied methods to evaluate, synthesize and analyze information coming from multiple sources to results of uncertainty studies on various computer codes. By using formal methods based on rational requirements, evaluations are made as objective as possible.

Proposed methods allow to take uncertainty (either aleatory or coming from imprecision in the data) explicitly into account in the evaluation process. they provide interesting tools to evaluate sources. In the particular case of computer codes, they give new instrumental tools for code validation procedures (Trucano, Swiler, Igusa, Oberkampf, and Pilch 2006), a problem particularly important for nuclear safety institute as the IRSN. The consistency between conclusions drawn from our results and informal observations confirms that using formal methods to analyze information is meaningful and can be useful. Compared to such informal observations, presented methods allow for a more subtle analysis, allowing to quantify disagreement among sources, to detect biases, underestimated uncertainty, ...

We have also illustrated the potential advantages offered by the use of possibility theory. In terms of information evaluation, probabilistic and

Figure 6: Results of synthesis for PCT2: probabilistic and possibilistic approaches (---: experimental value)







6.F : Possibilities: arithmetic weighted mean

possibilistic approaches have comparable results (which is not surprising, since they are based on similar rational requirements). However, the possibilistic approach has more flexibility to synthesis and analyze the information, offering a wider range of tools. The fact that both probabilities and possibilities can be seen as special cases of imprecise probabilities could be used to build a generalized approach, possibly by using some recent research results about measure of divergence for sets of probabilities (Abellan and Gomez 2006). Such a generalization remains the subject of further research.

Also, since results given by basic synthesizing operators can sometimes be found too rough, sometimes more complex tools that allow for a finer analysis are needed. This is why the IRSN is working on methods that are more complex but remain tractable and interpretable (Destercke, Dubois, and Chojnacki 2007).

References

- Abellan, J. and M. Gomez (2006). Measures of divergence on credal sets. *Fuzzy Sets and* System 157(11).
- Baudrit, C. and D. Dubois (2006). Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis* 51(1), 86–108.
- Clemen, R. and R. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* 19(2), 187–203.
- Cooke, R. (1991). *Experts in uncertainty*. Oxford, UK: Oxford University Press.
- Destercke, S. and E. Chojnacki (2007). Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes. *Accepted for publication in*

Nuclear Eng. and Design.

- Destercke, S., D. Dubois, and E. Chojnacki (2007). Possibilistic information fusion using maximal coherent subsets. In *Proc. IEEE Int. Conf. On Fuzzy Systems (FUZZ'IEEE).*
- Dubois, D. and H. Prade (1988). Possibility Theory: An Approach to Computerized Processing of Uncertainty. New York: Plenum Press.
- Dubois, D. and H. Prade (1992). On the relevance of non-standard theories of uncertainty in modeling amd pooling expert opinions. *Reliability Engineering and System* Safety 36, 95–107.
- Ferson, S. and L. R. Ginzburg (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety* 54, 133–144.
- Klement, E., R. Mesiar, and E. Pap (2000). *Triangular Norms*. Dordrecht: Kluwer Academic Publisher.
- OCDE (2007, May). Bemuse phase iii report: Uncertainty and sensitivity analysis of the loft l2-5 test. Technical Report NEA/NCIS/R(2007)4, NEA.
- Sandri, S., D. Dubois, and H. Kalfsbeek (1995, August). Elicitation, assessment and pooling of expert judgments using possibility theory. *IEEE Trans. on Fuzzy Systems* 3(3), 313– 335.
- Trucano, T., L. Swiler, T. Igusa, W. Oberkampf, and M. Pilch (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering and System Safety 91*, 1331–1357.
- Walley, P. (1991). Statistical reasoning with imprecise Probabilities. New York: Chapman and Hall.