Inner and outer approximation of belief structures using a hierarchical clustering approach[1]

Thierry Denœux

UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France
email: Thierry.Denoeux@hds.utc.fr

March 30, 2015

[1] A short version of this paper was presented at IPMU'2000, Madrid, July 2000 [6].

Abstract

A hierarchical clustering approach is proposed for reducing the number of focal elements in a crisp or fuzzy belief function, yielding strong inner and outer approximations. At each step of the proposed algorithm, two focal elements are merged, and the mass is transfered to their intersection or their union. The resulting approximations allow the calculation of lower and upper bounds on the belief and plausibility degrees induced by the conjunctive or disjunctive sum of any number of belief structures. Numerical experiments demonstrate the effectiveness of this approach.

Keywords: Belief functions, Dempster-Shafer theory, Fuzzy belief structures, Approximation.

1 Introduction

In the last twenty years, the theory of belief functions (BF's), sometimes referred to as "Dempster-Shafer Theory", has gained increasing recognition as one of the major frameworks for representing and manipulating uncertain and partial knowledge. The seminal work of Dempster [4] and Shafer [23] has been followed by a large of number of important theoretical and practical contributions. In particular, the Transferable Belief Model (TBM) introduced by Smets [28] has emerged has a coherent, well-founded interpretation of the mathematical apparatus of BF's, with clear axiomatic justification. Belief functions have been applied to a wide range of problems including medical diagnosis, data fusion, pattern recognition and function approximation [29, 5, 7, 21].

Despite its success as a model of human reasoning under uncertainty, one of the arguments often raised against the theory of BF's is its relatively high computational complexity, specially as compared to other approaches such as Bayesian probability theory or Possibility theory. Indeed, the complexity of combining BF's while aggregating pieces of evidence increases exponentially with the number of sources [33], which may lead to serious problems when both the number of steps in the reasoning process and the size of the frame of discernment become large.

The problem of limiting the complexity of BF manipulation has, of course, been addressed by many authors. Existing approaches rely either on efficient procedures for performing exact computations, or on approximation techniques. In the first category, an optimal algorithm for computing Dempster's rule of combination was described by Kennes [16], and efficient algorithms for manipulating restricted classes of BF's were proposed by several authors [1, 24]. The second category of approaches is composed of Monte-Carlo techniques [32, 20, 19], and BF approximation procedures. The latter approach is adopted in this paper.

Several techniques have been proposed for approximating a BF bel by a "simpler" BF bel'. A first strategy is to constrain bel' to belong to a predefined class $\mathcal{B}$ of BF's having a relatively simple form. Such a strategy was advocated by Voorbraak [31], Dubois and Prade [11], and, more recently, by Grabisch [12], the set $\mathcal{B}$ being composed, respectively, of probability functions, possibility measures, and k-additive BF's (i.e., BF's whose focal elements are at most of size k). This approach is particularly interesting when it is regarded as a means to build a bridge between BF theory and another theory, as is the case in Dubois and Prade's work. A nice property of Voorbraak's method is that it commutes with Dempster's rule of combination; however, a lot of information is usually lost when approximating a general BF by a probability function. The set of k additive BF's is larger, but its relevance in the context of uncertainty modeling remains to be clarified [12].

Another strategy is to simplify a BF by removing and/or aggregating focal elements, without imposing *a priori* a particular structure for the approximating BF. Reducing the number of focal elements is generally a good strategy because the combination of two BF's bel_1 and bel_2 can be performed in time proportional to $|\mathcal{F}(\text{bel}_1)||\mathcal{F}(\text{bel}_2)||\Omega|$, where $\mathcal{F}(\text{bel}_i)$ is the set of focal elements of bel_i ($i = 1, 2$), and Ω is the frame of discernment. One of the simplest way to reduce the complexity of a BF is to replace the k focal elements with the smallest mass by their union, leaving the other focal elements unchanged. This is the principle of the summarization method introduced by Lowrance et al [14]. More sophisticated methods were subsequently proposed by Tessem [30], Bauer [2], Harmanec [13] and Petit-Renaud and Denœux [21]. In particular, Harmanec and Petit-Renaud independently introduced the idea of using a systematic *clustering* strategy for grouping similar or unimportant focal elements. This idea will be further explored in this paper.

An important feature of an approximation method is the way in which a BF bel is related to its approximation bel′. When the associated plausibility functions pl and pl′ are such that pl $\leq$ pl′, then bel′ is said to be less committed than (or to weakly include) bel: in some sense, it is less informative than bel, and it thus constitutes a "cautious" or conservative approximation. In particular, the summarization algorithm [14], the possibilistic outer approximation method [11] as well as the procedures described in [13] generate conservative approximations.

In this paper, we propose a technique for approximating any crisp or fuzzy BF bel by two BF's bel^- and bel^+, called strong inner and outer approximations [11], such that $\widehat{\mathrm{pl}}^- \leq \mathrm{pl} \leq \widehat{\mathrm{pl}}^+$. The interesting feature of this approach is that it allows the calculation of lower and upper bounds on belief and plausibility degrees when combining BF's in a conjunctive or disjunctive fashion, allowing the user to control the quality of the approximation. Our approach thus partially shares the same goals as the possibilistic techniques proposed by Dubois and Prade. However, it does not impose any particular constraint on the structure of the BF's, allowing more precise approximations, and our method may be applied to belief structures with fuzzy focal elements.

The rest of this paper is structured as follows. Section 2 summarizes the main concepts related to crisp and fuzzy belief functions. The notions of weak and strong inclusions between belief structures are then recalled in Section 3, where some new results are also established. Finally, our approach to BF approximation is described in Section 4, and experimental results are presented in Section 5. Section 6 concludes the paper.

2 Background

Let Ω denote a finite set called the frame of discernment. A *belief structure* (BS) is a function m from 2^Ω to $[0,1]$, verifying:

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

The subsets A of Ω such that $m(A) > 0$ are the *focal elements* of m. A BS m such that $m(\emptyset) = 0$ is said to be normal. This condition was originally imposed by Shafer [23], but it is relaxed in the TBM, the allocation of a positive belief number to the empty set being interpreted as a consequence of the open-world assumption [25].

The BF induced by m is the function bel : $2^\Omega \mapsto [0,1]$ verifying:

$$\mathrm{bel}(A) \triangleq \sum_{\emptyset \neq B \subseteq A} m(B). \tag{2}$$

for all $A \subseteq \Omega$. Such a function may be shown to have the property of complete monotonicity [23]. Its use for representing degrees of belief was justified by Smets on an axiomatic basis [27]. Whereas bel(A) represents the amount of support given to A, the *potential* amount of support that *could be* given to A is represented by the plausibility of A defined as:

$$\mathrm{pl}(A) \triangleq \mathrm{bel}(\Omega) - \mathrm{bel}(\bar{A}) \tag{3}$$

where $\bar{A}$ denotes the complement of A. There is therefore a kind of duality between belief and plausibility functions.

Two BS's representing distinct items of evidence may be combined using the conjunctive sum $\cap$ or the disjunctive sum $\cup$ operations defined, respectively, as:

$$(m_1 \cap m_2)(A) \triangleq \sum_{B \cap C = A} m_1(B) m_2(C), \tag{4}$$

$$(m_1 \cup m_2)(A) \triangleq \sum_{B \cup C = A} m_1(B) m_2(C) \tag{5}$$

for all $A \subseteq \Omega$. These operations are commutative and associative. The choice of one of them for aggregating evidence may be guided by "metaknowledge" concerning the reliability of the two sources [26]. Note that the conjunctive sum as described by Eq. (4) may produce a subnormal BS (i.e., it is possible to have $(m_1 \cap m_2)(\emptyset) > 0$). Under the closed-world assumption, some kind of normalization thus has to be performed. The *Dempster normalization procedure* converts a subnormal BS m into a normal BS m^* by dividing each belief number by $1 - m(\emptyset)$ [23].

The above concepts may be generalized to allow the assignment of degrees of beliefs to ambiguous propositions such as typically expressed in verbal statements, and represented by fuzzy subsets of the frame of discernment. As defined by Zadeh [38] and Yager [34], a fuzzy belief structure (FBS) may be defined as a function m assigning belief numbers to a finite set of fuzzy focal elements F_i, $i = 1, \ldots, n$. As in the standard case, the condition

$$\sum_{i=1}^{n} m(F_i) = 1$$

is imposed. A FBS is normal iff its focal elements are normal fuzzy sets (i.e., they have unit height). Following Zadeh [38], the concept of plausibility of a fuzzy subset A may then be generalized[1] as the expectation of the conditional possibility measure of A given F_i, defined as:

$$\mathrm{pl}(A) \triangleq \sum_{i=1}^{n} m(F_i) \Pi(A|F_i) \tag{6}$$

with

$$\Pi(A|F_i) \triangleq \max_{\omega \in \Omega} \min[\mu_A(\omega), \mu_{F_i}(\omega)].$$

Similarly, the credibility of a fuzzy subset A induced by a FBS m may be defined as the expectation of the conditional necessity of A:

$$\mathrm{bel}(A) \triangleq \sum_{i=1}^{n} m(F_i) N(A|F_i), \tag{7}$$

[1] Other generalizations are possible, such as the one proposed by Yen [37]. The comparison of these generalizations is outside the scope of this paper.

where the conditional necessity of A given F_i is defined as

$$N(A|F_i) \triangleq \Pi(\Omega|F_i) - \Pi(\bar{A}|F_i), \tag{8}$$

to account for the possible subnormality of F_i, as suggested by Dubois and Prade [9]. Note that (7) is then a valid generalization of (2).

As proposed by Yager [35], the conjunctive and disjunctive sums may be readily extended to fuzzy belief structures by replacing the crisp intersection and union in (4) and (5) by fuzzy counterparts, defined, for example, using the min and max operations. Note that the conjunctive combination of two normal FBS's may produce a subnormal FBS. If necessary, the normalization of a FBS m may be performed using Yager's soft normalization procedure [36] generalizing Dempster's normalization and defined as:

$$m^*(A) \triangleq \frac{\sum_{B^*=A} h(B)m(B)}{\sum_{B \in \mathcal{F}(m)} h(B)m(B)} \tag{9}$$

where $h(B) = \max_\omega \mu_B(\omega)$ denotes the height of B, B^* is the normal fuzzy set defined by $\mu_{B^*}(\omega) = \mu_B(\omega)/h(B)$, and $\mathcal{F}(m)$ is the set of focal elements of m.

3 Inclusion of belief structures

The set $\mathcal{S}(\Omega)$ of crisp or fuzzy BS's on Ω can be equipped with partial ordering relations, which may be used to assess the coherence of two BS's and compare their "information content" [8, 11, 26]. Although such ordering relations were initially proposed for crisp BS's, they can be extended without any difficulty to FBS's. In the sequel, no distinction will be made between crisp and fuzzy BS's, unless explicitly stated.

Let m and m' be two elements of $\mathcal{S}(\Omega)$. Then m is said to be more committed than m', or to be weakly included in m', iff the associated plausibility functions verify $\mathrm{pl} \leq \mathrm{pl}'$. When m and m' are normal, this condition is equivalent to $\mathrm{bel} \geq \mathrm{bel}'$. This weak inclusion relation (hereafter noted $\preccurlyeq$) plays a central role in the TBM, in which the Principle of Least Commitment [26] (commanding to choose the least committed belief function compatible with a set of constraints) plays a role similar to the Maximum Entropy principle in Bayesian theory.

Although weak inclusion has a simple definition, it is difficult to work with, because the condition $m \preccurlyeq m'$ is difficult to express as a function of the belief numbers. This is one of the reasons why a more "operational" definition of inclusion, called strong inclusion, was proposed by Yager and discussed by Dubois and Prade [10], among others. Let m and m' be two BS's with focal elements $\mathcal{F}(m) = \{F_1, \ldots, F_n\}$ and $\mathcal{F}(m') = \{F'_1, \ldots, F'_q\}$. Then m is said to be strongly included in m', or to be a *specialization* of m' (noted $m \subseteq m'$), iff there exists a non-negative matrix W with entries w_{ij} $(i = 1, \ldots, n; j = 1, \ldots, q)$ such that

$$\sum_{j=1}^{q} w_{ij} = m(F_i), \quad i = 1, \ldots, n, \tag{10}$$

$$\sum_{i=1}^{n} w_{ij} = m'(F'_j), \quad j = 1, \ldots, q \tag{11}$$

and

$$w_{ij} > 0 \Rightarrow F_i \subseteq F'_j.$$

The relationship between m and m' may be seen as a transfer of mass from each focal element F_i of m to supersets $F'_j \supseteq F_i$, the quantity w_{ij} denoting the part of $m(F_i)$ transferred to F'_j. Alternatively, one may define a matrix G whose general term $g_{ij} = w_{ij}/m(F_i)$ denotes the relative proportion of $m(F_i)$ assigned to F'_j. Matrix G is called a *generalization matrix* by Klawonn and Smets [17]. It verifies

$$\sum_{j=1}^{q} g_{ij} = 1 \quad i = 1, \ldots, n.$$

and it allows to give a simple expression of m' as a function of m:

$$m'(F'_j) = \sum_{i=1}^{n} m(F_i) g_{ij} \quad j = 1, \ldots, q.$$

The terms "strong" and "weak" inclusion are justified by the following proposition.

PROPOSITION 1
For any two BS's on Ω, $m \subseteq m' \Rightarrow m \preccurlyeq m'$.

Proof: The proof was given by Dubois and Prade [10] for the crisp case. It may easily be extended to the more general case of fuzzy BS's as follows. Using the definition of the plausibility of a fuzzy event given by (6) and the definition of matrix G, we have, for all crisp or fuzzy subset A of Ω

$$\begin{aligned} \mathrm{pl}'(A) &= \sum_j m'(F'_j)\Pi(A|F'_j) \\ &= \sum_j \sum_i m(F_i) g_{ij} \Pi(A|F'_j) \\ &= \sum_i m(F_i) \sum_j g_{ij} \Pi(A|F'_j). \end{aligned}$$

For all $g_{ij} > 0$, we have $F_i \subseteq F'_j$, and consequently

$$\Pi(A|F'_j) \geq \Pi(A|F_i).$$

Hence,

$$\sum_j g_{ij} \Pi(A|F'_j) \geq \Pi(A|F_i),$$

and

$$\mathrm{pl}'(A) = \sum_i m(F_i) \sum_j g_{ij} \Pi(A|F'_j) \geq \sum_i m(F_i) \Pi(A|F_i) = \mathrm{pl}(A). \tag{12}$$

□

Strong inclusion also has an interesting property with regard to conjunctive and disjunctive combination, as expressed by the following proposition.

Proposition 2

Let m and m' be two BS's such that $m \subseteq m'$. Then, for all $m'' \in \mathcal{S}(\Omega)$,

$$(m \cap m'') \subseteq (m' \cap m''), \text{ and}$$

$$(m \cup m'') \subseteq (m' \cup m'').$$

Proof: We give only the proof for the conjunctive case. Since $m \subseteq m'$, there exists a matrix W with general term (w_{ij}) verifying Eqs. (10) and (11). Let F_i, F'_j, and F''_k denote the focal elements of m, m' and m'', respectively. Similarly, let $D_h, h = 1, \ldots, r$ be the focal elements of $m \cap m''$,and $C_\ell, \ell = 1, \ldots, s$ the focal elements of $m' \cap m''$. Let U denote the matrix of size $r \times s$, with general term

$$u_{h\ell} = \sum_{i,j} \sum_{\{k | F_i \cap F''_k = D_h, F'_j \cap F''_k = C_\ell\}} w_{ij} m''(F''_k).$$

We have

$$\begin{aligned}
\sum_h u_{h\ell} &= \sum_{i,j} \sum_h \sum_{\{k | F_i \cap F''_k = D_h, F'_j \cap F''_k = C_\ell\}} w_{ij} m''(F''_k) \\
&= \sum_{i,j} \sum_{\{k | F'_j \cap F''_k = C_\ell\}} w_{ij} m''(F''_k) \\
&= \sum_j \sum_{\{k | F'_j \cap F''_k = C_\ell\}} \left(\sum_i w_{ij} \right) m''(F''_k) \\
&= \sum_j \sum_{\{k | F'_j \cap F''_k = C_\ell\}} m'(F'_j) m''(F''_k) \\
&= (m' \cap m'')(C_\ell),
\end{aligned}$$

and

$$\begin{aligned}
\sum_\ell u_{h\ell} &= \sum_{i,j} \sum_\ell \sum_{\{k | F_i \cap F''_k = D_h, F'_j \cap F''_k = C_\ell\}} w_{ij} m''(F''_k) \\
&= \sum_{i,j} \sum_{\{k | F_i \cap F''_k = D_h\}} w_{ij} m''(F''_k) \\
&= \sum_i \sum_{\{k | F_i \cap F''_k = D_h\}} \left(\sum_j w_{ij} \right) m''(F''_k) \\
&= \sum_i \sum_{\{k | F_i \cap F''_k = D_h\}} m(F_i) m''(F''_k) \\
&= (m \cap m'')(D_h).
\end{aligned}$$

Moreover, $u_{h\ell} > 0$ implies the existence of i, j and k such that $D_h = F_i \cap F''_k$, $C_\ell = F'_j \cap F''_k$, and $F_i \subseteq F'_j$, and hence that $D_h \subseteq C_\ell$.

This completes the proof that $(m \cap m'')$ is strongly included in $(m' \cap m'')$. $\square$

Remark 1 As mentioned above, the strong inclusion relation is a partial ordering which allows to compare the "information content" of BS's. However, this interpretation poses some difficulty in the case of subnormal BS's, as shown by the following example. Let us assume that we have a BS $m_1(\{\omega_0\}) = 1$ for some $\omega_0 \in \Omega$, and a BS m_2 such that $m_2(\{\omega_0\}) = 0.5$ and $m_2(\emptyset) = 0.5$. Then, m_2 is strictly included in m_1 (in the strong sense), i.e., $m_2 \subseteq m_1$. However, under the usual interpretation of the mass given to the empty set [28], m_2 cannot be considered as more informative than m_1, whose information content is maximal (it corresponds to complete certainty). Hence, some caution should be exercised when interpreting strong inclusion between m_1 and m_2 in terms of information content, in the general case where $m_1(\emptyset) \neq m_2(\emptyset)$. The same remark applies in the case of BS's with subnormal fuzzy focal elements.

4 Approximation of belief structures

4.1 Clustering approximations of a BS

The idea of grouping, or *clustering* similar elements of a large set is central to many abstraction and information compression mechanisms. As already mentioned in the introduction, the conjunctive or disjunctive combination of two belief structures m_1 and m_2 can be performed in time proportional to the sizes of $\mathcal{F}(m_1)$ and $\mathcal{F}(m_2)$. A natural way to decrease the complexity of the combination is therefore to decrease the number of focal elements by grouping similar, or unimportant ones. This was already the strategy employed in the summarization method [14], in which the focal elements with the smallest mass were aggregated, the sum of their belief numbers being transfered to their union. Recently, more sophisticated schemes for approximating a belief structure by clustering its focal elements were proposed independently by Harmanec [13] and Petit-Renaud and Denœux [21, 22]. The basic mechanism for obtaining strong inner and outer approximations from a partition of the focal elements is explained in the sequel.

Let $\mathcal{F}(m) = \{F_1, \ldots, F_n\}$ be the set of (crisp or fuzzy) focal elements of $m \in \mathcal{S}(\Omega)$. Let $\mathcal{P} = \{I_1, \ldots, I_K\}$ be a partition of $\mathbb{N}_n = \{1, \ldots, n\}$, i.e., a family of subsets of $\mathbb{N}_n$ such that

$$I_k \cap I_\ell = \emptyset \quad \forall k \neq \ell$$

and

$$\bigcup_{k=1}^{K} I_k = \mathbb{N}_n.$$

An approximation of m may be constructed by transferring each belief number $m(F_i)$ such that $i \in I_k$ to

$$G_k = \bigcup_{j \in I_k} F_j. \tag{13}$$

We then obtain a new BS $\widehat{m}_{\mathcal{P}}^{+}$ with (at most) K focal elements $\mathcal{F}(\widehat{m}_{\mathcal{P}}^{+}) = \{G_1, \ldots, G_K\}$ and

$$\widehat{m}_{\mathcal{P}}^{+}(G_k) = \sum_{i \in I_k} m(F_i) \quad k = 1, \ldots, K.$$

As remarked by Harmanec [13] and Petit-Renaud [21], $\widehat{m}^+_{\mathcal{P}}$ strongly includes m: it is a strong outer approximation of m according to the terminology of Dubois and Prade [11].

In the same way, given a partition $\mathcal{P}' = \{I'_1, \ldots, I'_K\}$ of $\mathbb{N}_n$, it is possible to define a strong inner approximation $\widehat{m}^-_{\mathcal{P}'}$ with focal elements

$$H_k = \bigcap_{i \in I_k} F_i,$$

and such that

$$\widehat{m}^-_{\mathcal{P}'}(H_k) = \sum_{i \in I_k} m(F_i) \quad k = 1, \ldots, K.$$

It is clear that $\widehat{m}^-_{\mathcal{P}'} \subseteq m$.

Example 1 Let us consider the following BS on $\Omega = \{a, b, c, d, e\}$, taken from [2]:

$$\begin{array}{ll} F_1 = \{a, c, d\} & m(F_1) = 0.3 \\ F_2 = \{c, d\} & m(F_2) = 0.05 \\ F_3 = \{c\} & m(F_3) = 0.1 \\ F_4 = \{d, e\} & m(F_4) = 0.05 \\ F_5 = \{a, b\} & m(F_5) = 0.5 \end{array}$$

Let $\mathcal{P} = \{I_1, I_2, I_3\}$ be the following partition of $\mathbb{N}_5$:

$$\begin{array}{lll} I_1 & = & \{1, 4\} \\ I_2 & = & \{2, 3\} \\ I_3 & = & \{5\}. \end{array}$$

The strong outer approximation $\widehat{m}^+_{\mathcal{P}}$ induced by $\mathcal{P}$ is defined by

$$G_1 = F_1 \cup F_4 = \{a, c, d, e\}, \qquad \widehat{m}^+_{\mathcal{P}}(G_1) = m(F_1) + m(F_4) = 0.35,$$

$$G_2 = F_2 \cup F_3 = \{c, d\}, \qquad \widehat{m}^+_{\mathcal{P}}(G_2) = m(F_2) + m(F_3) = 0.15,$$

$$G_3 = F_5, \qquad \widehat{m}^+_{\mathcal{P}}(G_3) = m(F_5) = 0.5.$$

Similarly, the strong inner approximation $\widehat{m}^-_{\mathcal{P}}$ induced by $\mathcal{P}$ is defined by

$$H_1 = F_1 \cap F_4 = \{d\}, \qquad \widehat{m}^-_{\mathcal{P}}(H_1) = m(F_1) + m(F_4) = 0.35,$$

$$H_2 = F_2 \cap F_3 = \{c\}, \qquad \widehat{m}^-_{\mathcal{P}}(H_2) = m(F_2) + m(F_3) = 0.15,$$

$$H_3 = F_5, \qquad \widehat{m}^-_{\mathcal{P}}(H_3) = m(F_5) = 0.5.$$

4.2 Determination of inner and outer clustering approximations

As shown above, given two partitions $\mathcal{P}$ and $\mathcal{P}'$ of $\mathbb{N}_n$ in K classes, it is possible to define two BS's $\widehat{m}^-_{\mathcal{P}'}$ and $\widehat{m}^+_{\mathcal{P}}$ with at most K focal elements, and such that

$$\widehat{m}^-_{\mathcal{P}'} \subseteq m \subseteq \widehat{m}^+_{\mathcal{P}}.$$

It may then be wondered how to choose $\mathcal{P}$ and $\mathcal{P}'$ so as to obtain "good" approximations of m. Clearly, $\widehat{m}^+_{\mathcal{P}}$ is less *specific*, or less *precise* than m. Klir [18] and Harmanec [13] proposed to measure this loss of precision between m and $\widehat{m}^+_{\mathcal{P}}$ by:

$$D_{\mathrm{bel}}(m, \widehat{m}^+_{\mathcal{P}}) \triangleq \sum_{A \subseteq \Omega} \mathrm{bel}(A) - \widehat{\mathrm{bel}}^+_{\mathcal{P}}(A),$$

where $\widehat{\mathrm{bel}}^+_{\mathcal{P}}$ denotes the belief function induced by $\widehat{m}^+_{\mathcal{P}}$. The use or this error measure requires the normality assumption, because weak inclusion of bel in $\widehat{\mathrm{bel}}^+_{\mathcal{P}}$ then implies $\mathrm{bel} \geq \widehat{\mathrm{bel}}^+_{\mathcal{P}}$. An alternative measure could be

$$\begin{aligned} D_{\mathrm{pl}}(m, \widehat{m}^+_{\mathcal{P}}) &\triangleq \sum_{A \subseteq \Omega} \widehat{\mathrm{pl}}^+_{\mathcal{P}}(A) - \mathrm{pl}(A) \\ &= 2^{|\Omega|}[\widehat{\mathrm{bel}}^+_{\mathcal{P}}(\Omega) - \mathrm{bel}(\Omega)] + D_{\mathrm{bel}}(m, \widehat{m}^+_{\mathcal{P}}), \end{aligned}$$

which remains valid even in case of subnormality of m or its approximation. Both distance measures are obviously equivalent under the normality assumption.

A different approach to measure the quality of an approximation is to use some measure of imprecision or "information content" of belief functions. In [11], Dubois and Prade propose to measure the imprecision of a BS m by its generalized cardinality, defined as

$$|m| \triangleq \sum_{i=1}^{n} m(F_i)\ |F_i|, \tag{14}$$

where $|F_i|$ is the number of elements in F_i, or its sigma-count cardinality if F_i is fuzzy. Since $m \subseteq \widehat{m}^+_{\mathcal{P}}$ obviously implies $|m| \leq |\widehat{m}^+_{\mathcal{P}}|$, the following criterion can be used to measure the quality of the approximation of m by $\widehat{m}^+_{\mathcal{P}}$:

$$\Delta(\widehat{m}^+_{\mathcal{P}}, m) \triangleq |\widehat{m}^+_{\mathcal{P}}| - |m| = \sum_{k=1}^{K} \sum_{i \in I_k} m(F_i)(|G_k| - |F_i|),$$

where, as before, $I_1, \ldots, I_K$ are the elements of partition $\mathcal{P}$, and G_k is defined by (13). Note that, in the above expression, cardinality could be replaced by other measures of uncertainty, such as nonspecificity [18]. In the sequel, cardinality will be preferred for its simplicity.

A similar line of reasoning may be used to measure the quality of the inner approximation. Clearly, $\widehat{m}^-_{\mathcal{P}'}$ is more specific than m. However, this gain of precision is spurious because it is not supported by any evidence. It should therefore be minimized, which may be achieved by minimizing $\Delta(m, \widehat{m}^-_{\mathcal{P}'}) = |m| - |\widehat{m}^-_{\mathcal{P}'}|$.

Let $\mathcal{P}_K(\mathbb{N}_n)$ denote the set of all partitions of $\mathbb{N}_n$ in K classes. Adopting the cardinality difference Δ as a quality criterion, the best inner and outer clustering K-approximations of m can then be defined as $\widehat{m}^-_{\mathcal{P}_*}$ and $\widehat{m}^+_{\mathcal{P}^*}$ such that:

$$\Delta(m, \widehat{m}^-_{\mathcal{P}_*}) = \min_{\mathcal{P} \in \mathcal{P}_K(\mathbb{N}_n)} \Delta(m, \widehat{m}^-_{\mathcal{P}})$$

and

$$\Delta(\widehat{m}^+_{\mathcal{P}^*}, m) = \min_{\mathcal{P} \in \mathcal{P}_K(\mathbb{N}_n)} \Delta(\widehat{m}^+_{\mathcal{P}}, m).$$

It is well known that the number $S(n, K)$ of partitions of a set of n elements in K classes is

$$S(n, K) = \frac{1}{K!} \sum_{i=1}^{K} (-1)^{K-i} \binom{K}{i} i^n$$

which rapidly explodes even for moderate values of n [15]. Consequently, an exhaustive search in the space of all partitions becomes quickly prohibitive, even for small values of n, and one has to resort to heuristic search techniques. Among these, hill-climbing strategies, converging to local minima of the objective function, are possible candidates. Such an approach was tested by Harmanec [13], but its efficiency seemed to be limited for this problem. An alternative technique is to use a hierarchical clustering algorithm [15], an approach successfully applied by Petit-Renaud [21] and Harmanec [13], whose "pair approximation" algorithm is implicitly based on the hierarchical clustering principle. In this approach, pairs of focal elements are grouped sequentially to decrease the complexity of the belief structure, until the desired number of focal elements has been reached. A similar idea was proposed by Moral and Salmeron [19] as an approximate pre-computation step in a Monte Carlo algorithm.

More precisely, this approach may be described, in the case of outer approximation, as follows. In order to decide which focal elements will be aggregated first, a distance, or dissimilarity has first to be defined. For each pair (F_i, F_j) of focal elements, transferring the mass $m(F_i) + m(F_j)$ to $F_i \cup F_j$ increases the cardinality by

$$\delta_\cup(F_i, F_j) = [m(F_i) + m(F_j)]|F_i \cup F_j| - m(F_i)|F_i| - m(F_j)|F_j|.$$

This quantity measures the impact of replacing F_i and F_j by their union, and can therefore be interpreted as a "distance" between these two focal elements. It may also be written

$$\begin{aligned} \delta_\cup(F_i, F_j) &= m(F_i)(|F_i \cup F_j| - |F_i|) + m(F_j)(|F_i \cup F_j| - |F_j|) \\ &= m(F_i) d_H(F_i \cup F_j, F_i) + m(F_j) d_H(F_i \cup F_j, F_j), \end{aligned}$$

where $d_H(\cdot, \cdot)$ denotes the Hamming distance. The quantity $\delta_\cup(F_i, F_j)$ can thus be viewed as a weighted sum of the Hamming distances from F_i and F_j to their union.

Once a distance between focal elements has been defined, a simple clustering procedure is then to compute the $n(n-1)/2$ distances, aggregate the two closest focal elements, and iterate until the predefined number of focal elements has been reached (or until the acceptable quality of approximation has been reached). Needless to say,

Table 1: Hierarchical clustering algorithm for computing a strong outer approximation of a BS with a fixed number K of focal elements.

Store	belief structure as list $\mathtt{m} = \{[F_i, m(F_i)], i = 1, \ldots, n\}$
Pick	K = maximum number of focal elements.
Initialize	Dissimilarity matrix $\delta_\cup(F_i, F_j)$ for $i, j \in \{1, \ldots, n\}$ `mhat`= `m`
Iterate	`While` $n > K$, Find i^* and j^* such that $\delta_\cup(F_{i^*}, F_{j^*}) = \min_{i \neq j} \delta_\cup(F_i, F_j)$ Delete $[F_{i^*}, m(F_{i^*})]$ and $[F_{j^*}, m(F_{j^*})]$ from `mhat` Add $[F_{i^*} \cup F_{j^*}, m(F_{i^*}) + m(F_{j^*})]$ to `mhat` $n \leftarrow n - 1$ Update dissimilarity matrix $\delta_\cup$ `End While`

exactly the same approach may be used to construct an inner approximation, the distance between F_i and F_j being then measured by the decrease of cardinality resulting from the transfer of $m(F_i) + m(F_j)$ to the intersection of F_i and F_j:

$$\delta_\cap(F_i, F_j) = m(F_i)|F_i| + m(F_j)|F_j| - [m(F_i) + m(F_j)]|F_i \cap F_j|,$$

which is also a weighted sum of the Hamming distances from F_i and F_j to their intersection. A detailed description of the algorithm in the case of the outer approximation is given in Table 1. Note that the stopping criterion used in this algorithm (based on the number of focal elements) may be replaced by a condition on the approximation quality, measured, e.g., by the relative increase of cardinality. This variant of the algorithm is described in Table 2.

Remark 2 Note that this hierarchical clustering procedure does not yield just a single partition, but a hierarchy, i.e., a family of partitions linked by inclusion relations (each aggregation of two focal elements yields a new, finer partition). Denoting $\varphi_k^-(m)$ and $\varphi_k^+(m)$ the inner and outer approximations of m with k focal elements obtained by the above method, we have

$$\varphi_1^-(m) \subseteq \ldots \subseteq \varphi_{n-1}^-(m) \subseteq m \subseteq \varphi_{n-1}^+(m) \subseteq \ldots \subseteq \varphi_1^+(m).$$

If $F_1, \ldots, F_n$ denote the focal elements of m, then the BS's $\widehat{m}_1^-$ and $\widehat{m}_1^+$ with only one focal element are defined, respectively, as

$$\varphi_1^-(m)\left(\bigcap_{i=1}^n F_i\right) = 1$$

$$\varphi_1^+(m)\left(\bigcup_{i=1}^n F_i\right) = 1.$$

Table 2: Hierarchical clustering algorithm for computing a strong outer approximation of a BS with a maximum approximation error.

Store	belief structure as list `m` $= \{[F_i, m(F_i)], i = 1, \dots, n\}$
Pick	$\epsilon =$ maximum approximation error.
Initialize	Dissimilarity matrix $\delta_\cup(F_i, F_j)$ for $i, j \in \{1, \dots, n\}$ `mhat`= `m`, `mhat1`= `m`

Iterate	`Begin Loop`
	Find i^* and j^* such that $\delta_\cup(F_{i^*}, F_{j^*}) = \min_{i \neq j} \delta_\cup(F_i, F_j)$
	Delete $[F_{i^*}, m(F_{i^*})]$ and $[F_{j^*}, m(F_{j^*})]$ from `mhat1`
	Add $[F_{i^*} \cup F_{j^*}, m(F_{i^*}) + m(F_{j^*})]$ to `mhat1`
	Compute approximation error `err` between `m` and `mhat1`
	`if err` $> \epsilon$,
	`stop`
	`else`
	`mhat` $\leftarrow$ `mhat1`
	Update dissimilarity matrix $\delta_\cup$
	`Endif`
	`End Loop`

Remark 3 As shown in Section 3, we have, for any inner and outer approximations $\widehat{m}^-$ and $\widehat{m}^+$: $\widehat{\text{pl}}^- \leq \text{pl} \leq \widehat{\text{pl}}^+$, but we do not have in general $\widehat{\text{bel}}^+ \leq \text{bel} \leq \widehat{\text{bel}}^-$ in the case of subnormal belief structures. However, a bracketing of bel may be obtained even in this case by noticing that

$$\text{bel}(A) = \text{pl}(\Omega) - \text{pl}(\bar{A}) \quad \forall A \in [0, 1]^\Omega.$$

from which we can derive the following inequalities for all $A \in [0, 1]^\Omega$:

$$\underline{\text{bel}}(A) \leq \text{bel}(A) \leq \overline{\text{bel}}(A) \tag{15}$$

with

$$\underline{\text{bel}}(A) = \max[0, \widehat{\text{pl}}^-(\Omega) - \widehat{\text{pl}}^+(\bar{A})] \tag{16}$$

and

$$\overline{\text{bel}}(A) = \widehat{\text{pl}}^+(\Omega) - \widehat{\text{pl}}^-(\bar{A}). \tag{17}$$

Remark 4 As a consequence of Remark 1, an outer approximation $\varphi_k^+(m)$ of m is not necessarily less informative in a strict sense if $m(\emptyset) > 0$, because the mass assigned to the empty set may have disappeared in the course of the hierarchical clustering process. A simple remedy for this problem might be to modify the inner and outer approximation algorithms, so as to prevent any change to the mass given to the empty set. However, a similar problem occurs in the case of fuzzy BS's, with no such obvious solution. Anyway, our main objective when approximating BS's is to speed up the combination process (as will be shown in Section 4.3), and these subnormality problems need not be considered in this context.

Table 3: Construction of strong inner approximations of a BS by hierarchical clustering of focal elements. The matrices of dissimilarities $\delta_\cap(F_i, F_j)$ between focal elements at each step are displayed on the right-hand side of the table, the lowest dissimilarity being shown in bold characters. The two focal elements merged at each step are marked by an asterisk.

		F_i	$m(F_i)$	$\delta_\cap(F_i, F_j)$				
Step 1		$\{a, c, d\}$	0.3	-	-	-	-	-
	*	$\{c, d\}$	0.05	0.3	-	-	-	-
	*	$\{c\}$	0.1	0.6	**0.05**	-	-	-
		$\{d, e\}$	0.05	0.65	0.1	0.2	-	-
		$\{a, b\}$	0.5	1.1	1.1	1.1	1.1	-
Step 2		$\{a, c, d\}$	0.3	-	-	-	-	
	*	$\{d, e\}$	0.05	0.65	-	-	-	
		$\{a, b\}$	0.5	1.1	1.1	-	-	
	*	$\{c\}$	0.15	0.6	**0.25**	1.25	-	
Step 3	*	$\{a, c, d\}$	0.3	-	-	-		
		$\{a, b\}$	0.5	1.1	-	-		
	*	$\emptyset$	0.2	**0.9**	1	-		
Step 4	*	$\{a, b\}$	0.5					
	*	$\emptyset$	0.5					
Step 5		$\emptyset$	1					

EXAMPLE 2 Let us consider the construction of an inner hierarchy $\varphi_1^-(m) \subseteq \ldots \subseteq \varphi_{n-1}^-(m) \subseteq m$ for the BS of Example 1. The four steps of the algorithm are described in Table 3. At the first step, F_2 and F_3 have the smallest $\delta_\cap$ distance and are replaced by their intersection $F_2 \cap F_3 = \{c\}$, yielding an inner approximation $\varphi_4^-(m)$ with 4 focal elements. In a second step, F_4 is grouped with $\{c\}$ to form the empty set, which subsequently absorbs F_1 and F_5. The highest level of the hierarchy thus corresponds to a BS focused on the intersection of the focal elements of m, the empty set in this case. The final result is summarized in Table 4. Similar results for the outer approximation procedure are shown in Table 5.

4.3 Approximate combination of BS's

The above scheme may be used to compute strong inner and outer approximations for the conjunctive or disjunctive sum of any number of belief structures. Let us assume that we have N belief structures $m^1, \ldots, m^N$, and we want to approximate

$$m = m^1 \nabla \ldots \nabla m^N,$$

where $\nabla \in \{\cap, \cup\}$ is a combination operator. A simple procedure to compute strong inner and outer approximations of m is combine the BS's one at a time, and perform the approximations at each step, for a predefined maximal number K of focal elements.

Table 4: Successive strong inner approximations of a belief structure using the hierarchical clustering approach.

<table>
<tr><td>m</td><td>$\{c\}$
0.1</td><td>$\{c, d\}$
0.05</td><td>$\{d, e\}$
0.05</td><td>$\{a, c, d\}$
0.3</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_4^-(m)$</td><td colspan="2">$\{c\}$
0.15</td><td>$\{d, e\}$
0.05</td><td>$\{a, c, d\}$
0.3</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_3^-(m)$</td><td colspan="3">$\emptyset$
0.2</td><td>$\{a, c, d\}$
0.3</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_2^-(m)$</td><td colspan="4">$\emptyset$
0.5</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_1^-(m)$</td><td colspan="5">$\emptyset$
1</td></tr>
</table>

Table 5: Successive strong outer approximations of a belief structure using the hierarchical clustering approach.

<table>
<tr><td>m</td><td>$\{a, c, d\}$
0.3</td><td>$\{c, d\}$
0.05</td><td>$\{c\}$
0.10</td><td>$\{d, e\}$
0.05</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_4^+(m)$</td><td colspan="2">$\{a, c, d\}$
0.35</td><td>$\{c\}$
0.10</td><td>$\{d, e\}$
0.05</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_3^+(m)$</td><td colspan="3">$\{a, c, d\}$
0.45</td><td>$\{d, e\}$
0.05</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_2^+(m)$</td><td colspan="4">$\{a, c, d, e\}$
0.5</td><td>$\{a, b\}$
0.5</td></tr>
<tr><td>$\varphi_1^+(m)$</td><td colspan="5">$\{a, b, c, d, e\}$
1</td></tr>
</table>

Table 6: Algorithm for the combination of N belief structures, with a predetermined maximal number K of focal elements.

Store	$m^1, \ldots, m^N$ N belief structures to be combined.
Pick	∇ = conjunctive or disjunctive sum operator
	K = maximum number of focal elements.
Initialize	$\widehat{m}^- \leftarrow m^1$
	$\widehat{m}^+ \leftarrow m^1$
Iterate	`For` $i = 2 : N$,
	$\quad \widehat{m}^- \leftarrow \varphi_K^-(\widehat{m}^- \nabla m^i)$
	$\quad \widehat{m}^+ \leftarrow \varphi_K^+(\widehat{m}^+ \nabla m^i)$
	`Next` i

The corresponding algorithm is described in Table 6. It follows immediately from Proposition 2 that $\widehat{m}^- \subseteq m \subseteq \widehat{m}^+$.

Note that, if m is subnormal, normalizing $\widehat{m}^-$ and $\widehat{m}^+$ does not, in general, yield inner or upper approximations of the normalized form m^* of m. However, if m is crisp, it may be noticed that

$$\text{bel}^*(A) = \frac{\text{bel}(A)}{\text{pl}(\Omega)}$$

and

$$\text{pl}^*(A) = \frac{\text{pl}(A)}{\text{pl}(\Omega)},$$

for all $A \subseteq \Omega$. Consequently, we have, using (15):

$$\frac{\max[0, \widehat{\text{pl}}^-(\Omega) - \widehat{\text{pl}}^+(A)]}{\widehat{\text{pl}}^+(\Omega)} \leq \text{bel}^*(A) \leq \frac{\widehat{\text{pl}}^+(\Omega) - \widehat{\text{pl}}^-(A)}{\widehat{\text{pl}}^-(\Omega)}$$

and

$$\frac{\widehat{\text{pl}}^-(A)}{\widehat{\text{pl}}^+(\Omega)} \leq \text{pl}^*(A) \leq \frac{\widehat{\text{pl}}^+(A)}{\widehat{\text{pl}}^-(\Omega)}.$$

Example 3 Let us consider a data fusion experiment in which six sensors deliver six trapezoidal possibility distributions $\pi^i = (a^i, b^i, c^i, d^i), i = 1, 6$ about some parameter of interest in $\Omega = \{1, 2, \ldots, 32\}$. For all $\omega \in \Omega$, we thus have

$$\pi^i(\omega) = \begin{cases} 0 & \omega \notin [a^i, d^i] \\ \dfrac{\omega - a^i}{b^i - a^i} & \omega \in [a^i, b^i) \\ 1 & \omega \in [b^i, c^i) \\ \dfrac{d^i - \omega}{d^i - c^i} & \omega \in [c^i, d^i] \end{cases}$$

Each of these possibility distributions can be transformed into a consonant belief structure m^i. To take into account one's knowledge regarding the reliability of the sensors, each BS m^i is discounted by some factor α^i. The problem is to approximate

Table 7: Parameters of the possibility distributions and discounting factors used in Example 3.

a^i	b^i	c^i	d^i	α^i
1.55	10.39	16.04	25.58	0.51
0.53	9.21	12.00	23.93	0.87
-0.77	10.03	14.39	24.62	0.72
-3.26	7.03	13.37	24.85	0.97
4.71	13.41	17.73	27.66	0.73
-1.49	9.01	13.16	24.65	0.71

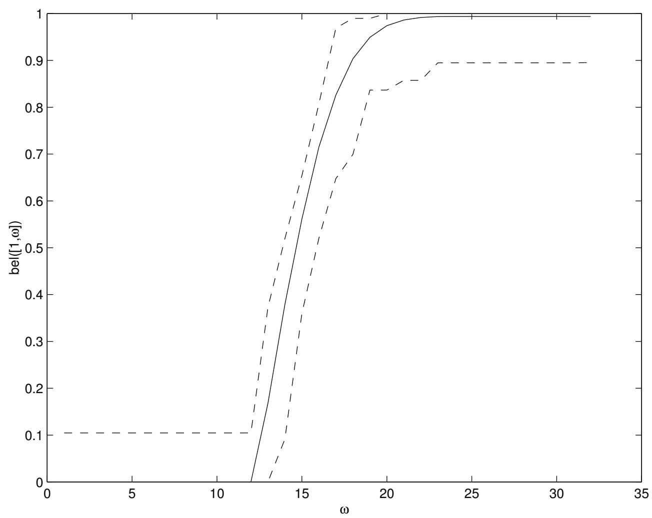

Figure 1: Degree of belief bel($\{1,\ldots,\omega\}$) as a function of $\omega \in \Omega$ (solid line), as well as its lower and upper approximations (dashed line) computed using (15), for the data of example 3.

the conjunctive sum of the six discounted belief structures, using only a small number of focal elements.

The parameters of the possibility distributions as well as the discounting factors are shown in Table 7. The final BS $m = m^1 \cap \ldots \cap m^6$ contains 139 focal elements. Figures 1 and 2 show, respectively, bel($\{1,\ldots,\omega\}$) and pl($\{1,\ldots,\omega\}$) as a function of $\omega \in \Omega$, as well as their lower and upper bounds computed using the above algorithm, with only $K = 15$ focal elements.

5 Numerical experiments

Numerical experiments were conducted to verify the effectiveness of the proposed scheme for BF approximation, as compared to simpler approaches. To provide a simple reference method against which to compare our approach, we considered the following variants of the summarization method. Let m be a crisp or fuzzy BS with

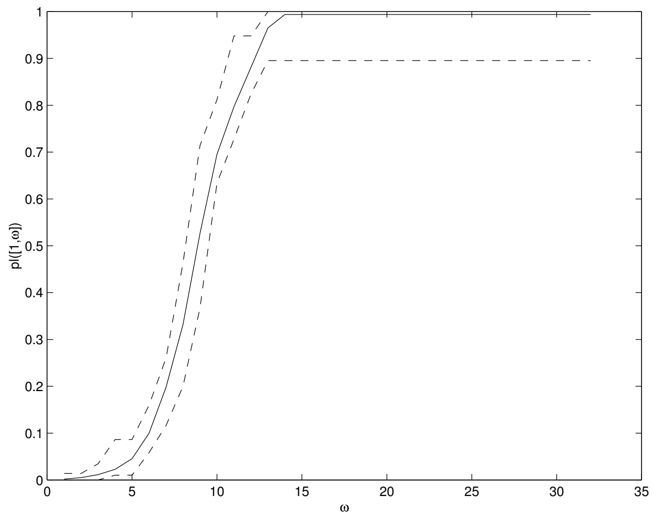

Figure 2: Degree of plausibility $\mathrm{pl}(\{1,\ldots,\omega\})$ as a function of $\omega \in \Omega$ (solid line), as well as its lower and upper approximations (dashed line), for the data of example 3.

focal elements $F_1, \ldots, F_n$ such that $m(F_1) \geq m(F_2) \geq \ldots \geq m(F_n)$. For any integer $K < n$, we define the inner summarization of m with K focal elements as the BS $\widehat{m}_S^-$ such that

$$\widehat{m}_S^-(F_i) = m(F_i) \quad i = 1, \ldots, K-1$$

and

$$\widehat{m}_S^-(\bigcap_{i=K+1}^{n} F_i) = \sum_{i=K+1}^{n} m(F_i).$$

The outer summarization $\widehat{m}_S^-$ is obtained using the usual summarization method as

$$\widehat{m}_S^+(F_i) = m(F_i) \quad i = 1, \ldots, K-1$$

and

$$\widehat{m}_S^+(\bigcup_{i=K+1}^{n} F_i) = \sum_{i=K+1}^{n} m(F_i).$$

It is clear that $\widehat{m}_S^-$ and $\widehat{m}_S^+$ are, respectively, inner and outer clustering approximations of m.

Another reference approximation method that does not generate inner or outer approximations is Bauer's D1 algorithm, which was shown in [2] to outperform Tessem's k-l-x method [30] as well as the summarization procedure according to several error criteria based on pignistic probabilities.

Following Tessem [30], Bauer [2] and Harmanec [13], we used a probabilistic model to randomly generate BS's on a frame of size $|\Omega| = 32$. Our model is close to Bauer's, except for the generation of focal elements. All three aforementioned authors generate the focal elements in a purely random fashion (i.e., each non empty subset of Ω has the same probability of becoming a focal element of m), which induces much more conflict among focal elements than usually encountered in real applications. To alleviate this problem, we propose a procedure in which:

Table 8: Algorithm for randomly generating crisp belief structures.

Pick	$n =$ number of focal elements $q =$ size of frame. $\mu, \sigma =$ parameters for the generation of focal elements.
Initialize	`For` $i = 1 : q$, $p_i \leftarrow 0.8 * \exp[-(i-\mu)^2/(2*\sigma^2)]$, `Next` i `For` $k = 1 : n$, $F_k \leftarrow \emptyset$ `Next` k $r \leftarrow 1$

Iterate	`For` $k = 1 : n$,
	  generate q realizations $x_1, \ldots, x_q$ of $X \sim \mathcal{U}_{[0,1]}$
	  `For` $i = 1 : q$,
	    `if` $x_i \leq p_i$, $F_k \leftarrow F_k \cup \{\omega_i\}$ `Endif`
	  `Next` i
	  generate a realization u from $U \sim \mathcal{U}_{[0,1]}$
	  `If` $k < n$,
	    $m(F_k) \leftarrow u \cdot r$
	    $r \leftarrow r - m(F_k)$
	  `Else`
	    $m(F_k) \leftarrow r$
	  `Endif`
	`Next` k

1. a probability $p_i \in [0, 1]$ is assigned to each element ω_i of Ω;

2. a focal element F is constructed by drawing each element ω_i from Ω with probability p_i;

3. the process is iterated until the desired number of focal elements is reached.

Using this procedure, elements ω_i with larger probability p_i have more chance to be included in F, and tend to appear more often in the focal elements of the generated structure m. In our simulations, the p_i were defined as

$$p_i = 0.8 \exp\left[-\frac{(i-16)}{50}\right] \quad i = 1, \ldots, 32.$$

For generating BS's with fuzzy focal elements, we used a variant of this method in which the membership degree $\mu_F(\omega_i)$ of each $\omega_i \in \Omega$ to F is randomly generated using a uniform distribution on the interval $[0, p_i]$, after which F is normalized. The complete algorithm used for generating crisp BS's is summarized in Table 8.

To simulate the common situation in which BF's independently provided by several experts are combined, we generated 5 BS's with $n = 10$ focal elements on a frame of size $|\Omega| = 32$ using the above procedure, and combined them using the conjunctive sum operation. The exact result was compared to the inner and outer approximations with $K = 15$ focal elements computed using the procedure described in Table 6, as well as to the D1 approximation with the same number of focal elements, computed in a similar way.

Table 9: Results with crisp BS's. The three methods are: hierarchical clustering (HC), summarization (S), and Bauer's method (D1). The mean and standard deviation (std) of 50 trials are given for each of the three error criteria (see details in text).

	D^-		D^+		D	
	mean	std	mean	std	mean	std
HC	0.095	(0.052)	0.091	(0.061)	0.026	(0.016)
S	0.156	(0.057)	0.183	(0.107)	0.076	(0.044)
D1					0.027	(0.021)

Since degrees of belief are usually of main interest when working with BF's, we measured the quality of a pair $(\widehat{m}^-, \widehat{m}^+)$ of inner and outer approximations of a BS m as the mean differences between $\mathrm{bel}(A)$ and its lower and upper bounds computed from (15), for all $A \subseteq \Omega$:

$$D^-(m, \widehat{m}^-, \widehat{m}^+) = 2^{-|\Omega|} \sum_{A \subseteq \Omega} \mathrm{bel}(A) - \underline{\mathrm{bel}}(A) \tag{18}$$

$$D^+(m, \widehat{m}^-, \widehat{m}^+) = 2^{-|\Omega|} \sum_{A \subseteq \Omega} \overline{\mathrm{bel}}(A) - \mathrm{bel}(A) \tag{19}$$

with $\underline{\mathrm{bel}}(A)$ and $\overline{\mathrm{bel}}(A)$ defined according to Eqs (16) and (17), respectively.

An additional error measure is obtained by comparing each degree of belief $\mathrm{bel}(A)$ to the center of its approximating interval, which is a meaningful point approximation:

$$D(m, \widehat{m}^-, \widehat{m}^+) = 2^{-|\Omega|} \sum_{A \subseteq \Omega} \left| \frac{\underline{\mathrm{bel}}(A) + \overline{\mathrm{bel}}(A)}{2} - \mathrm{bel}(A) \right|. \tag{20}$$

For the $D1$ method which only generates a point approximation $\widehat{m}$, we used

$$D(m, \widehat{m}) = 2^{-|\Omega|} \sum_{A \subseteq \Omega} \left| \widehat{\mathrm{bel}}(A) - \mathrm{bel}(A) \right|,$$

where $\widehat{\mathrm{bel}}$ is the belief function induced by $\widehat{m}$. The exact computation of the above error measures requires the calculation of $2^{|\Omega|}$ degrees of belief, which for $|\Omega| = 32$ is not practically feasible. So, each error measure was estimated by an average computed over 100 randomly selected subsets of Ω.

The results are summarized in Tables 9 and 10 for the crisp and fuzzy case, respectively (results for the D1 are not available for the fuzzy case, because the method has been developed for the approximation of crisp BS's only). Our hierarchical clustering method clearly outperforms the summarization method according to all three criteria, and appears to be roughly equivalent to the D1 method as far as point approximation of crisp BS's is concerned. Results for a particular trial in the crisp case are shown graphically in Figures 3 to 5.

Table 10: Results with fuzzy BS's. The three methods are: hierarchical clustering (HC) and summarization (S). The mean and standard deviation (std) of 50 trials are given for each of the three error criteria (see details in text).

	D^-		D^+		D	
	mean	std	mean	std	mean	std
HC	0.076	(0.021)	0.185	(0.080)	0.057	(0.033)
S	0.089	(0.023)	0.391	(0.126)	0.154	(0.061)

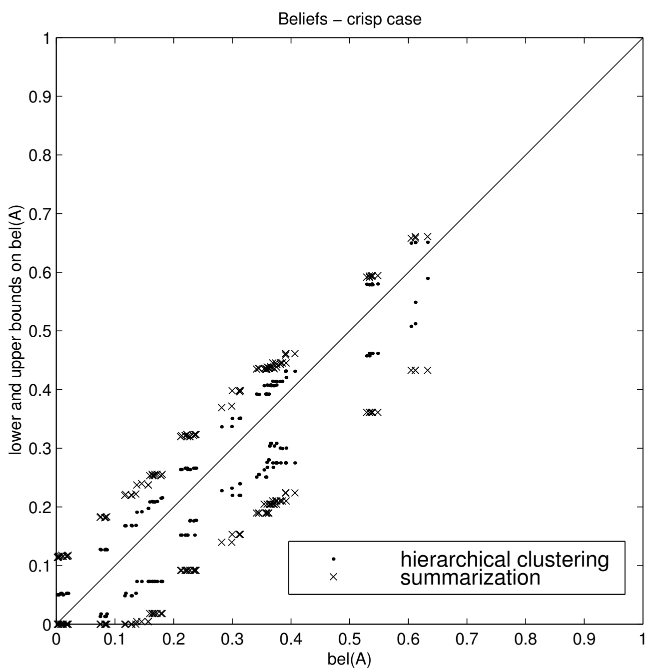

Figure 3: Lower and upper bounds on bel(A) as a function of bel(A) for 100 randomly selected subsets A of Ω.

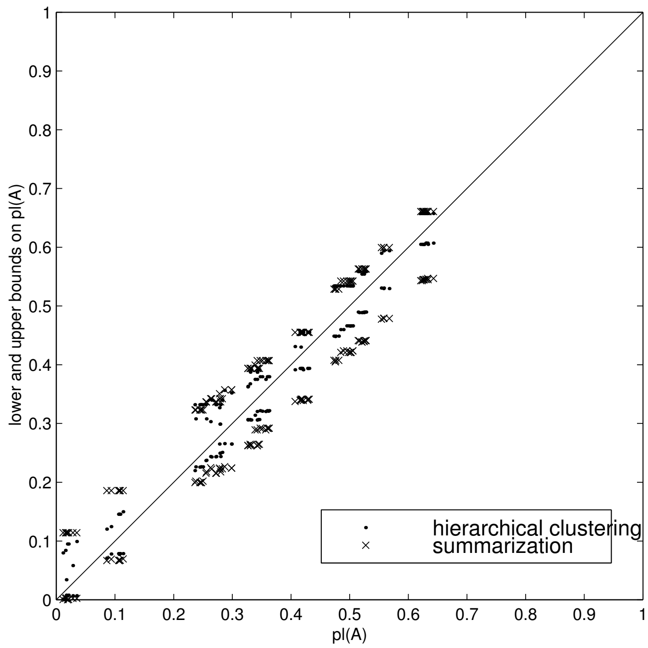

Figure 4: Lower and upper bounds on pl(A) as a function of pl(A) for 100 randomly selected subsets A of Ω.

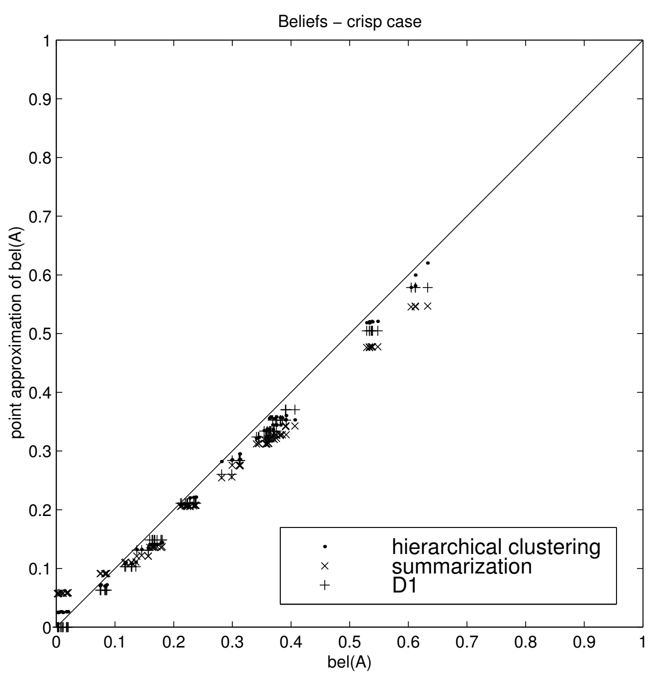

Figure 5: Point approximations of bel(A) as a function of bel(A) for 100 randomly selected subsets A of Ω, using the hierarchical clustering, summarization and D1 methods.

6 Conclusions

Whereas the belief functions elicited from experts or inferred from observation data are usually simple, their combination often leads to a proliferation of focal elements which causes an exponential increase of computation time. A natural way to reduce the complexity of a BS is to merge "similar" or "unimportant" focal elements, and replace them by their union or their intersection, leading, respectively, to a strong outer or inner approximation of the original belief function. This process may be iterated until a given complexity level (or a given approximation quality) has been reached. When combining several BS's, this approach can be applied at each intermediate step of the computation, which allows to obtain lower and upper bounds on the belief and plausibility degrees at a reasonable computational cost.

The idea of joining pairs of focal elements for approximation purposes was already present in the work of Petit-Renaud and Denœux [21], Harmanec [13], and Moral and Salmeron [19], and the concepts of inner and outer approximations were first proposed by Dubois and Prade [11]. These two ideas have been combined in this paper, leading to the notions of inner and outer clustering approximations, which were applied to both crisp and fuzzy BS's. A different, but related approach to the simplification of BS's consists in reducing the size of the frame Ω by clustering its elements. This approach, the investigation of which has just started [3], may be a valuable alternative to the technique presented in this paper, especially when the number of focal elements is very large.

References

[1] J. A. Barnett. Computational methods for a mathematical theory of evidence. In *Proceedings IJCAI-81*, pages 868–875, Vancouver, 1981.

[2] M. Bauer. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence – an empirical study. *International Journal of Approximate Reasoning*, 17:217–237, 1997.

[3] A. Ben Yaghlane, T. Denœux, and K. Mellouli. Coarsening approximations of belief functions. In *Proc. of ECSQARU'2001 (to appear)*. Springer-Verlag, Toulouse, September 2001.

[4] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS-38:325–339, 1967.

[5] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

[6] T. Denœux. Inner and outer clustering approximations of belief structures. In *Proceedings of IPMU'2000*, volume 1, pages 125–132, Madrid, July 2000.

[7] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics A*, 30(2):131–150, 2000.

[8] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12:193–226, 1986.

[9] D. Dubois and H. Prade. An alternative approach to the handling of subnormal possibility distributions. *Fuzzy Sets and Systems*, 24:123–126, 1987.

[10] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.*, 4:244–264, 1988.

[11] D. Dubois and H. Prade. Consonant approximations of belief measures. *International Journal of Approximate Reasoning*, 4:419–449, 1990.

[12] M. Grabisch. Upper approximation of non-additive measures by k-additive measures – the case of belief functions. In *Proc. of the 1st Int. Symposium on imprecise probabilities and their applications*, Ghent, Belgium, June 1999.

[13] D. Harmanec. Faithful approximations of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 15 (UAI99)*, Stockholm, Sweden, 1999.

[14] Lowrance J. D, T. D. Garvey, and T. M. Strat. A framework for evidential-reasoning systems. In T. Kehler et al., editor, *Proceedings of AAAI'86*, volume 2, pages 896–903, Philadelphia, August 1986. AAAI.

[15] A. K. Jain and R. C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Englewood Cliffs, NJ., 1988.

[16] R. Kennes. Computational aspects of the Möbius transform of graphs. *IEEE Trans. SMC*, 22:201–223, 1992.

[17] F. Klawonn and P. Smets. The dynamic of belief in the transferable belief model and specialization-generalization matrices. In D. Dubois et al., editor, *Proc. of the 8th conference on Uncertainty in Artificial Intelligence*, pages 130–137. Morgan Kaufmann, San Mateo, CA, 1992.

[18] G. J. Klir and M. J. Wierman. *Uncertainty-Based Information. Elements of Generalized Information Theory*. Springer-Verlag, New-York, 1998.

[19] S. Moral and A. Salmeron. A Monte Carlo algorithm for combining Dempster-Shafer belief based on approximate pre-computation. In A. Hunter and S. Pearsons, editors, *Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99)*, pages 305–315, London, June 1999. Springer Verlag.

[20] S. Moral and N. Wilson. Markov-chain Monte-Carlo algorithms for the calculation of Dempster-Shafer belief. In *Proc. of the Twelfth National Conference on Artificial intelligence (AAAI-94)*, volume 1, pages 269–274, 1994.

[21] S. Petit-Renaud and T. Denœux. Handling different forms of uncertainty in regression analysis: a fuzzy belief structure approach. In A. Hunter and S. Pearsons, editors, *Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99)*, pages 340–351, London, June 1999. Springer Verlag.

[22] S. Petit-Renaud and T. Denœux. Regression analysis using fuzzy evidence theory. In *Proceedings of FUZZ-IEEE'99*, volume 3, pages 1229–1234, Seoul, August 1999.

[23] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[24] G. Shafer and R. Logan. Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987.

[25] P. Smets. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

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

[27] P. Smets. The normative representation of quantified beliefs by belief functions. *Artificial Intelligence*, 92(1–2):229–242, 1997.

[28] P. Smets. The Transferable Belief Model for quantified belief representation. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, 1998.

[29] P. Smets. Practical uses of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 15 (UAI99)*, pages 612–621, Stockholm, Sweden, 1999.

[30] B. Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61:315–329, 1993.

[31] F. Voorbraak. A computationally efficient approximation of Dempster-Shafer theory. *Int. J. Man-Machine Studies*, 30:525–536, 1989.

[32] N. Wilson. A Monte Carlo algorithm for Dempster-Shafer belief. In B. D'Ambrosio, P. Smets, and P. Bonissone, editors, *Proc. of the 7th Conference on Uncertainty in AI.* Morgan Kaufmann, 1991.

[33] N. Wilson. Algorithms for Dempster-Shafer theory. In D. M. Gabbay and P. Smets, editors, *Hanbook of defeasible reasoning and uncertainty management. Volume 5: Algorithms for Uncertainty and Defeasible Reasoning*, pages 421–475. Kluwer Academic Publishers, Boston, 2000.

[34] R. R. Yager. Generalized probabilities of fuzzy events from fuzzy belief structures. *Information Sciences*, 28:45–62, 1982.

[35] R. R. Yager. Arithmetic and other operations on Dempster-Shafer structures. *Int. J. Man-Machines Studies*, 25:357–366, 1986.

[36] R. R. Yager. On the normalization of fuzzy belief structure. *International Journal of Approximate Reasoning*, 14:127–153, 1996.

[37] J. Yen. Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):559–569, 1990.

[38] L. A. Zadeh. Fuzzy sets and information granularity. In R. K. Ragade M. M. Gupta and R. R. Yager, editors, *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North-Holland, Amsterdam, 1979.