

Models of belief functions - Impacts for patterns recognitions

P.-E. Doré, A. Fiche, A. Martin

E3I2 EA3876 - ENSIETA

2 rue Francois VERNY,

29806 BREST Cedex 09, FR

Pierre-Emmanuel.Dore, Anthony.Fiche,

Arnaud.Martin@ensieta.fr

Abstract – *In a lot of operational situations, we have to deal with uncertain and inaccurate information. The theory of belief functions is a mathematical framework useful to handle this kind of imperfection. However, in most of the cases, uncertain data are modeled with a distribution of probability. We present in this paper different principles to induce belief functions from probabilities. Hence, we decide to use these functions in a pattern recognition problem. We discuss about the results we obtain according the way we generate the belief function. To illustrate our work, it will be applied to seabed characterization.*

Keywords: Continuous belief functions, least commitment, maximum of necessity, decision making, seabed characterization.

1 Introduction

The theory of belief functions [3, 4, 19] is a powerful formalism to describe the imperfections of data given by an information's source. It is widely used in classification to merge the decisions coming from several classifiers [15]. Recently, some works deal with the question of representation of belief functions on real numbers [25, 14, 18, 24]. They give us the opportunity to use directly the framework of belief functions with raw data coming from sensors.

In this paper, we will focus on the way that belief functions have been modeled in [6]. It allows us to induce belief functions from probabilities. According to the nature of the source of information, the appropriate way to construct belief is not the same. To illustrate this phenomenon, we will use the work presented in [18, 10] and compare the classification result of seabed's sediments according the modeling of information.

This paper is organized as follow. In the section 2, we present the framework proposed in [7] to model belief function on real number and we recall different ways to induce belief functions from probabilities [6]. In the following section, we applied these functions to pattern recognition using a reasoning similar to the one found in [18].

2 Beliefs induced by probabilities

The theory of belief functions is a framework richer than the theory of probability. It allows us to represent information in a finnier way. In this section, we present an approach of belief functions on real numbers and some principles to induce belief functions from probabilities.

2.1 Belief on real numbers

In [24], Ph. Smets works on continuous belief functions on real numbers. He proposes to put basic belief assignement only on the intervals of $\mathbb{R} = \mathbb{R} \cup \{-\infty, \infty\}$. With this assumption, he had an efficient way to describe all the focal elements and made a link between belief functions on \mathbb{R} and probability density functions on \mathbb{R}^2 using the concept of basic belief densities. However, this framework is quite restrictive and for some applications, we need to express beliefs on more complex sets. In [7], the authors explicit an index function to scan \mathcal{F} , the set of all focal elements of a belief function on Ω , using a set I , the index space:

$$\begin{aligned} f^I : I &\longrightarrow \mathcal{F} \\ y &\longmapsto f^I(y) \end{aligned} \quad (1)$$

Hence, a credal measure μ^Ω is a positive measure such as $\int_I d\mu^\Omega(y) \leq 1$ and a credal space is defined by the brace (f^I, μ^Ω) . In order to compute belief functions, we need to define for all A in $\mathcal{P}(\Omega)$ (a family of subset of Ω):

$$F_{\subseteq A} = \{y \in I \mid f^I(y) \subseteq A\} \quad (2)$$

$$F_{\cap A} = \{y \in I \mid (f^I(y) \cap A) \neq \emptyset\} \quad (3)$$

$$F_{\supseteq A} = \{y \in I \mid A \subseteq f^I(y)\} \quad (4)$$

It can be used to model a belief on Borel algebra $\mathcal{B}(\overline{\mathbb{R}^n})$.

If $F_{\supseteq A}, F_{\subseteq A}, F_{\cap A}$ belong to $\mathcal{B}(I)$ for all A in $\mathcal{B}(\overline{\mathbb{R}^n})$, we define:

- The belief function:

$$bel^\Omega(A) = \int_{F_{\subseteq A}} d\mu^\Omega(y) \quad (5)$$

- The plausibility function:

$$p^{I\Omega}(A) = \int_{F_{\cap A}} d\mu^\Omega(y) \quad (6)$$

- The communality function:

$$q^\Omega(A) = \int_{F_{\supseteq A}} d\mu^\Omega(y) \quad (7)$$

In this framework, we define some basic tools. Let (f_1^I, μ_1^Ω) and (f_2^I, μ_2^Ω) be two credal spaces. Let combine them thanks to the conjunctive rule of combination [21]. We obtain the credal space $(f_{1\odot 2}^I, \mu_{1\odot 2}^\Omega)$ [7] such as:

$$q_{1\odot 2}^\Omega(A) = q_1^\Omega(A) \cdot q_2^\Omega(A) \quad (8)$$

There are other ways to combine informations within the theory of belief functions. One of them is the cautious rule [13] used to combine correlated sources of information. Let f^{I_1} and f^{I_2} be two index functions linked to the credal measures μ_1^Ω et μ_2^Ω . Let φ be a change of variables such as $\varphi(y_1) = y_2$ implies $f^{I_1}(y_1) = f^{I_2}(y_2)$. The braces (f^{I_1}, μ_1^Ω) and (f^{I_2}, μ_2^Ω) represent the same belief if [7]:

$$d\mu_1^\Omega(y_1) = |\det(\varphi'(y_1))| d\mu_2^\Omega(\varphi(y_1)) \quad (9)$$

Within this framework, we will study a particular type of belief functions, the consonant ones.

Example: Consonant belief functions. A belief function whose the focal elements are nested is a consonant belief function. This allows us to create a total ordering on \mathcal{F} linked to the \subseteq relation. Hence, we can define an index function f from I , a subset of \mathbb{R}^+ , to \mathcal{F} such as $(y \geq x) \implies (f(y) \subseteq f(x))$ [24]. The α -cuts of g , a continuous function from $\overline{\mathbb{R}^n}$ to $g(\overline{\mathbb{R}^n}) = I \subset \mathbb{R}^+$ are the set such as:

$$f_{cs}^I(\alpha) = \{x \in \mathbb{R}^n | g(x) \geq \alpha\} \quad (10)$$

We have the property that $F_{\subseteq A}^{cs}$ is an element of Borel algebra. Indeed:

$$\begin{aligned} F_{\subseteq A}^{cs} \neq \emptyset &\implies \exists \alpha_{\inf} = \inf \{ \alpha \in I | f_{cs}^I(\alpha) \cap A \neq \emptyset \} \\ &\implies F_{\subseteq A}^{cs} = [\alpha_{\inf}, \alpha_{\max}] \end{aligned} \quad (11)$$

Using a similar argument, we can prove that $F_{\supseteq A}^{cs}$ and $F_{\cap A}^{cs}$ are elements of Borel algebra. Hence, we can define the index function:

$$\begin{aligned} f_{cs}^I : I = [0, \alpha_{\max}] &\longrightarrow \{ f_{cs}^I(\alpha) | \alpha \in I \} \\ \alpha &\longmapsto f_{cs}^I(\alpha) \end{aligned} \quad (12)$$

If we consider a probability measure $\mu^{\overline{\mathbb{R}^n}}$ on \mathcal{I} , we obtain a credal space $(f_{cs}^I, \mu^{\overline{\mathbb{R}^n}})$.

Within this framework, we can build belief functions on real number with complex focal sets. In most of the cases, information of a source on real number is represented with a probability distribution. Hence, the problem is to find some criteria to build belief functions induced by a probability.

2.2 Isopignistic set and minimal commitment

2.2.1 Pignistic transformation

In most of the cases, imperfect knowledge is modeled with probability. Unfortunately, this framework is not suitable to represent phenomenons such as ignorance or uncertainty [23]. Therefore, we want to associate a belief function to a probability. There are operations to derivate a probability from a belief function. One of them is the pignistic transformation. Ph. Smets [20] has given a justification of this transformation in the transferable belief model. He proposes to use it in order to take decision on singletons. We propose to describe this transformation with the equation:

$$BetP(A) = \int_{F_{\cap A}} \frac{\nu(A, y)}{\nu(f^I(y), y)} d\mu^\Omega(y) \quad (13)$$

If we work on a discrete frame of discernement, $\nu(A, y)$ gives the cardinality of $A \cap f^I(y)$. We define on real numbers:

$$\nu(A, y) = \lambda(A \cap f^I(y)) + \omega(A, y) \cdot \delta(\lambda(f^I(y))) \quad (14)$$

where λ is the Lebesgue measure, δ is the Dirac measure and $\omega(A, y)$ is a number in $[0, 1]$ used to split the basic belief assignment of the focal elements on singletons of Ω . It is equal to 1 if $f^I(y) \subseteq A$ and 0 if $A \cap f^I(y) = \emptyset$. Generally, we have $\omega(A, y) + \omega(\complement A, y) = 1$ ¹. The opposite operation is quite more complexe. We will present several approaches of this problem according the context. The aim is to associate a belief function to a probability according the type of source which delivers the probability.

2.2.2 Least commitment principle

In the theory of belief functions, the minimal commitment principle is frequently used [12, 9]. The idea is quite simple. To choose a belief function among a set of belief functions and when there is no reason to prefer one to another, we have to choose the least informative one. That assumes there is an ordering to decide which one is the least informative. One ordering commonly used is funded on the communality function. Indeed, we can consider that communality function is a way to measure the non specificity of a belief function and that this function is linked to the conditioning process [24]. It is an interesting *criterion* if we know that a *posteriori* information will be use in the fusion process. Let assume that a source of information delivers a continuous probability density function $Betf$, but that this probability is induced by a bet. That implies this kind of data comes from a subjective point of view. We note $\mathcal{B}Iso(BetP)$, the set of belief functions whose the pignistic transformation gives $BetP$. There is a belief function [7] belonging to $\mathcal{B}Iso(BetP)$ whose the focal elements are the α -cuts of $Betf$ and the credal measure $\mu^{\mathcal{B}(\overline{\mathbb{R}^n})}$ is such as:

$$d\mu^{\mathcal{B}(\overline{\mathbb{R}^n})}(\alpha) = \lambda(f_{cs}^I(\alpha)) d\lambda(\alpha) \quad (15)$$

¹ $\complement A$ is the set such as $\complement A \cap A = \emptyset$ and $A \cup \complement A$ is the frame of discernement.

Theorem 2.1 Among the set of belief functions $\mathcal{B}Iso(BetP)$, the belief function defined by equation (15) is the least committed one for the commuality ordering.

We can build the least committed belief function linked to $\mathcal{B}Iso(BetP)$ when the associated probability density function is continuous. For discret frames of discernement or in particular cases of continuous belief functions, this kind of result has already been obtained [24]. When an expert models a phenomenon with a probability density function, we can use this transformation to combine a belief function with a given distribution of probability. Indeed, we assume that the opinion of an expert is quite subjective.

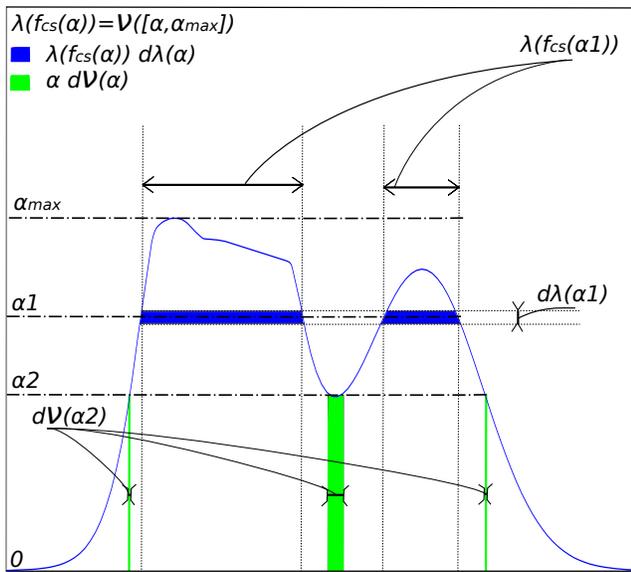


Figure 1: Two different ways to build belief functions from a probability density function $Betf$. We remark that $\lambda(f_{cs}^I(\alpha))$ refers to the volume of an α -cut and that $dV(\alpha)$ corresponds to the variation of the volume for a given α . The green and blue areas symbolize the infinitesimal quantity we have to integrate in order to compute the belief functions.

2.3 Maximum of necessity

When we work with an “objective” source of information, we can apply the principle of maximum of necessity. This principle comes from the theory of possibility [16]. The idea is to work with the most informative distribution of possibility (for the necessity ordering) which fulfils the following assumptions. The first one is that the possibility dominate the probability, *i.e.* for all A measurable $\Pi(A) \geq P(A)$. The second one is that the ordering must be kept, *i.e.* $P(A) \geq P(A') \Leftrightarrow \Pi(A) \geq \Pi(A')$. These conditions can be transposed in the framework of belief functions by setting that the plausibility function is equal to the possibility and the necessity is equal to the belief function if

we work with a consonant belief function [8]. Finding a belief function which verifies these properties is equivalent to find a nested focal sets family such as for all A belonging to this family, A is the smallest set (for the inclusion ordering) such as $P(A) = \beta$. This sets family corresponds to the confidence sets in theory of probability. If we have as input a continuous probability density function $Betf$, the focal sets can be described with the α -cuts of this function. We obtain a belief function defined by $(f_{cs}^I, \mu^{\mathcal{B}(\overline{\mathbb{R}}^n)})$ such as if we adapt the result obtained in [16]:

$$pl^{\mathcal{B}(\overline{\mathbb{R}}^n)}(x) = 1 - BetP(f_{cs}^I(Betf(x))) \quad (16)$$

i.e.:

$$d\mu^{\mathcal{B}(\overline{\mathbb{R}}^n)}(\alpha) = \alpha dV(\alpha) \quad (17)$$

with $V([\alpha, \alpha_{max}]) = \lambda(f_{cs}^I(\alpha))$ (*cf.* figure 1).

3 Application to recognition

To illustrate the benefit of using belief functions to make classification with data sensor's, we compare the result obtain in a probabilistic case and in belief functions cases.

3.1 Decision making

In [2], the authors decide to use credal inference in order to deal with the lack of information in sensor context. In our study, we prefer to keep the framework of the general Bayesian theorem (GTB) [22, 17] because it fulfils the least commitment principle within the transferable belief model. For the sake of simplicity, we set for all y in I that $\mu^T[X](y) = \mu^T[X](f(y))$. If we have the *a posteriori* information that the value of the parameter is in X , we have according the GTB for all A included in \mathcal{T} :

$$\mu^T[X](A) = \prod_{T_i \in A} pl^{\mathcal{B}(\overline{\mathbb{R}}^n)}[T_i](X) \prod_{T_i \in \overline{A}} (1 - pl^{\mathcal{B}(\overline{\mathbb{R}}^n)}[T_i](X)) \quad (18)$$

In order to take a decision on singletons, we can use the pignistic transformation. Hence, we have:

$$BetP(T_i) = \int_{F \cap T_i} \frac{\nu(T_i \cap f^I(y))}{\nu(f^I(y))} d\mu^T[X](y) \quad (19)$$

To illustrate, we use two different mixtures of 2 Gaussians. Firstly, according the method, we generate the belief functions induced by these mixtures (*cf.* figure 2). Hence, we suppose that there is two kinds of object whose a parameter could be describe by the pdf print figure 2. We use the method describes in [18] to decide, knowing the parameter, the type of object observed. As we can see (*cf.* figure 3), the consonant belief function obtained thanks to the maximum of necessity principle amplifies the decision of probability when the least commitment principle moderate this result. However, in this case, we observe that the belief function induce by Caron's method [1] inverses for some values of the parameter the decision which is taken in the probabilistic case.

3.2 Sediments characterization

The aim of this part is to compare automatic recognition results obtained by a Bayesian probabilist approach [26] and by several different belief functions approaches.

3.2.1 Presentation of data

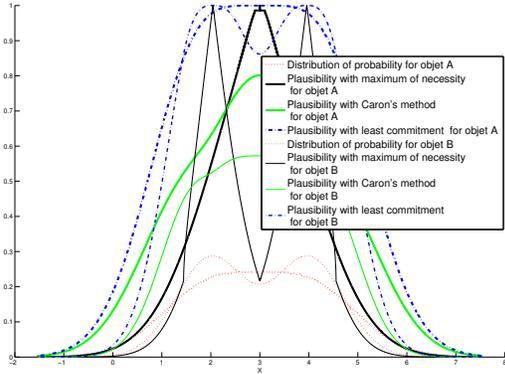


Figure 2: Different belief functions induced by two probability density functions.

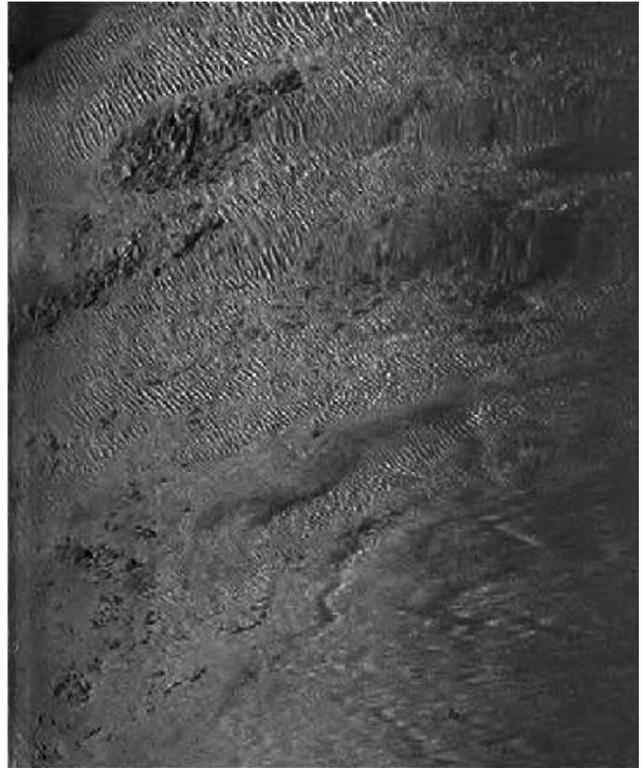


Figure 4: A sample of sonar images.

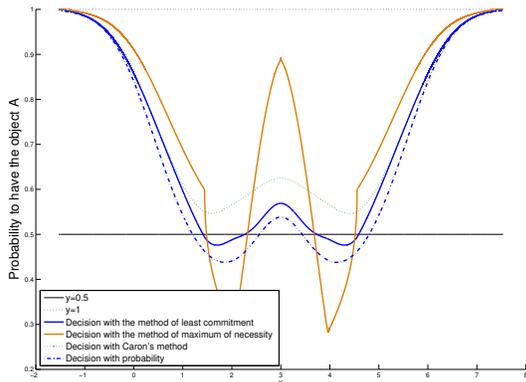


Figure 3: Decisions according the framework.

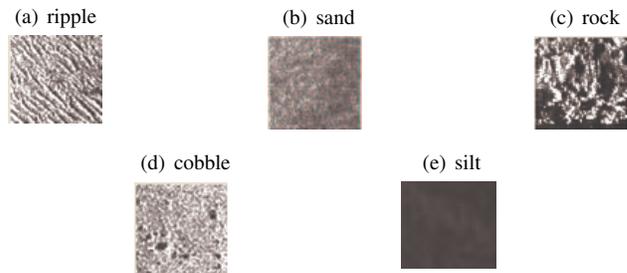


Table 1: Examples of tiles.

We use a data base of 42 images from a Klein 5400 sonar given by the GESMA (Groupe d'Études Sous-Marines de l'Atlantique) [15] (*cf.* image 4). A pre-processing have been done in order to reduce the speckle and the variation of gain in the images. An expert has labeled the images to indicate the type of sediments seen. He distinguishes five kinds of sediments: rock, cobble, ripple, sand, and silt (*cf.* table 1). In order to use an automatic classifier, we divide the images

in tiles of 32×32 pixels. To characterize a sediment tank to the tiles, we compute six parameters [11]: the homogeneity, the entropy, the contrast, the correlation, the directionality and the uniformity. A learning set help us to construct a mixture of five Gaussian curves linked to each class and parameters thanks to the EM algorithm [5].

3.2.2 Result

In [10], belief functions approach using Caron's method to generate belief function has already been developed to recognize seabed sediments. However, the belief functions describe in section 2 are quite different from this one. Hence, we decide to compare the results obtained by using belief functions induced by Caron's method with the ones obtained when we use the belief functions build by applying the least commitment principle and the maximum of necessity principle. This test is done with Gaussian mixtures computed using EM algorithm applied on data set of different sizes. Hence, the reliability of the Gaussian mixtures increases with the size of the learning data set. We observe (*cf.* figure 5) that the method which derives from the maximum of necessity emphasizes the behavior we have with the Bayesian probability approach when the methods funded on the least commitment principle and the Caron's approach reduce this phenomenon. Hence, when we lack of information, these methods seem to be best-suited.

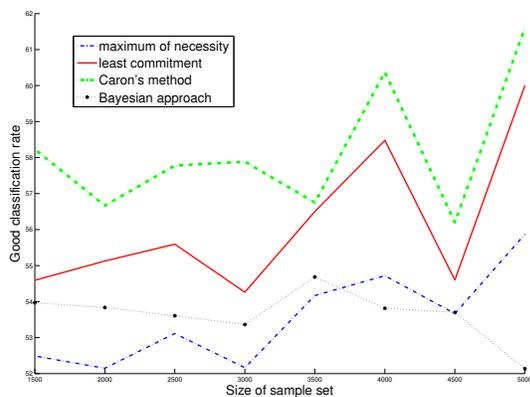


Figure 5: Rates of good recognition according the sample set size.

4 Conclusion

The recent advances in theory of belief functions, especially about beliefs on real numbers, provide us tools to estimate and merge continuous parameters. It allows us to use belief functions directly with raw data issued from sensors. Thanks to these breakthroughs, the framework of belief functions is a match for the theory of probabilities. It is even possible to take into account phenomenons we cannot handle with the probabilities as the ignorance. It has a great consequence when we have to take decision using sensor's

data, especially in an uncertain context and when we lack of *prior* information about the environment. However, we have to be careful when we generate a belief functions to model a source of information. Indeed, according the way we consider the source, the method used to build the belief function will not be the same. The method funded on the maximum of necessity confirms the information contained by a probability when the principle of least commitment or the Caron's method discounts it. Hence, when we use these functions to make seabed characterization, we observe that the good classification rates depends on the size of the learning set. When the size of the learning set is small, the least committed belief function gives the best results. However, the bigger the learning set is, the better the results for Bayesian and maximum of necessity principle methods are. Even if the result we have obtained are encouraging. There is still a lot of works left. A first step is to find an efficient way to compute the least committed belief function induced by a probability density function. Another step would be to propose combination rule in order to keep a consonant belief function when we combine two consonant belief functions. Ending, we could use this framework and these tools to make multi-sensor's data fusion.

References

- [1] F. Caron, B. Ristic, E. Duflos, and P. Vanheeghe. Least committed basic belief density induced by a multivariate Gaussian: formulation with applications. *International Journal of Approximate Reasoning*, 48(2):419–436, 2008.
- [2] F. Caron, P. Smets, E. Duflos, and P. Vanheeghe. Multisensor data fusion in the frame of the TBM on reals. Application to land vehicle positioning. 2, 2005.
- [3] A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *The Annals of Mathematical Statistics*, pages 325–339, 1966.
- [4] A.P. Dempster. Upper and lower probabilities generated by a random closed interval. *The Annals of Mathematical Statistics*, pages 957–966, 1968.
- [5] A.P. Dempster, NM Laird, and D.B. Rubin. Maximum likelihood from incomplete data using the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.
- [6] P.-E. Doré and A. Martin. About using beliefs induced by probabilities. *Workshop on the theory of belief functions (Brest France)*, 2010.
- [7] P.-E. Doré, A. Martin, and A. Khenchaf. Constructing consonant belief function induced by a multimodal probability. *COGNITIVE systems with Interactive Sensors (COGIS 2009)*, 2009.
- [8] D. Dubois and H. Prade. A set-theoretic view of belief functions. *International Journal of General Systems*, 12(3):193–226, 1986.

- [9] D. Dubois and H. Prade. The principle of minimum specificity as a basis for evidential reasoning. In *Processing and Management of Uncertainty in Knowledge-Based Systems on Uncertainty in knowledge-based systems. International Conference on Information*. Springer-Verlag, 1987.
- [10] A. Fiche and A. Martin. Bayesian approach and continuous belief functions for classification. *Rencontre francophone sur la Logique Floue et ses Applications (LFA09)*, 2009.
- [11] R.M. Haralick, I. Dinstein, and K. Shanmugam. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6):610–621, 1973.
- [12] Y.T. Hsia. Characterizing belief with minimum commitment. In *Intern. joint conf. on artificial intelligence*, pages 1184–1189, 1991.
- [13] A. Kallel and S.L. Hégarat-Masclé. Combination of partially non-distinct beliefs: The cautious-adaptive rule. *International Journal of Approximate Reasoning*, 2009.
- [14] L. Liu. A theory of Gaussian belief functions. *International Journal of Approximate Reasoning*, 14(2-3):95–126, 1996.
- [15] A. Martin. Comparative study of information fusion methods for sonar images classification. In *The Eighth International Conference on Information Fusion, Philadelphia, USA*, pages 25–29, 2005.
- [16] G. Mauris. Transformation of bimodal probability distributions into possibility distributions. *Instrumentation and measurement, IEEE Transactions*, Accepted for inclusion in a future issue of this journal.
- [17] D. Mercier, B. Quost, and T. Denœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258, 2006.
- [18] B. Ristic and P. Smets. Belief function theory on the continuous space with an application to model based classification. pages 4–9, 2004.
- [19] G. Shafer. *A mathematical theory of evidence*. Princeton University Press Princeton, NJ, 1976.
- [20] Ph. Smets. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*, 5:29–39, 1990.
- [21] Ph. Smets. The combination of evidence in the transferable belief model. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 12(5):447–458, 1990.
- [22] Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–1, 1993.
- [23] Ph. Smets. Probability, possibility, belief: Which and where. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 1–24, 1998.
- [24] Ph. Smets. Belief functions on real numbers. *International journal of approximate reasoning*, 40(3):181–223, 2005.
- [25] T.M. Strat. Continuous belief functions for evidential reasoning. In *Proceedings of the National Conference on Artificial Intelligence, University of Texas at Austin*, 1984.
- [26] D.P. Williams. Bayesian Data Fusion of Multi-View Synthetic Aperture Sonar Imagery for Seabed Classification. *IEEE Transactions on Image Processing*, 18:1239–1254, 2009.