Improvement of Land Cover Map from Satellite Imagery Using DST and DSmT

Radja Khedam*, Abdenour Bouakache*, Grégoire Mercier**, and Aichouche Belhadj-Aissa*
** Image Processing and Radiation Laboratory, Faculty of Electronic and Computer Science, University of Science and Technology Houari Boumediene (USTHB).*
BP. 32, El Alia, Bab Ezzouar, 16111, Algiers, Algeria.
*** ITI Dpt, GET/ENST Bretagne CS 83818, 29238 Brest Cedex3, France*
radjakhedam@lycos.com, h.belhadj@lycos.com, Gregoire.Mercier@enst-bretagne.fr

Abstract

The aim of this paper is to show that Dempster-Shafer Theory (DST) and a recent theory of plausible and paradoxical reasoning introduced by Dezert and Smaradache and thus called Dezert-Smarandache Theory (DSmT), can be successfully applied to improve a supervised classification of remotely sensed data. Notice that application fields of these two theories are related on multisensor/multitemporal/multiscale data fusion. In this study, our contribution lies in developing a new multispectral data classification process which can be seen as a multisensor fusion process where each thematic class is considered as one source of information.

1. Introduction

Given the current available techniques, remote sensing is recognized as a timely and cost-effective tool for earth observation and land monitoring. It constitutes the most feasible approach to both land surface change detection, and land-cover information required for the management of natural resources. The extraction of land-cover information is usually achieved through supervised or unsupervised classification methods. In unsupervised classification, an algorithm such as K-means or Isodata, is chosen that will take a remotely sensed data set and find a pre-specified number of statistical clusters in multispectral space. Although these clusters are not always equivalent to actual classes of land cover, this method can be used without having prior knowledge of the ground cover in the study site. Supervised classification, however, does require prior knowledge of the ground cover in the study site. The process of gaining this prior knowledge is known as ground-truthing. With supervised classification algorithms such as Maximum Likelihood or minimum of distance, the researcher locates areas on the unmodified image for which he knows the type of land cover, defines a polygon or a polyline around the known area, and assigns that land cover class to the pixels within the polygon or the polyline. This process known as training step is continued until a statistically significant number of pixels exist for each class in the classification scheme. Then, the multispectral data from the pixels in the sample polygons are used to train a classification algorithm. Once trained, the algorithm can then be applied to the entire image and a final classified image is obtained. In this work we propose a novel supervised classification approach.

Conventional supervised classifiers are statistical and very often based on the Bayesian theory which has been proved as a theoretically robust foundation for satellite image classification [3] [13]. However, the main limitation of a Bayesian formalism is that it cannot represent imprecision about uncertainty measurement and is able to consider only single (or individual) classes, which may lead to misclassification especially face to mixed pixels. To overcome these problems, Dempster-Shafer Theory (DST) [6] [14] and a new theory of plausible and paradoxical reasoning introduced by Dezert and Smaradache [7] [15] and thus called Dezert-Smarandache Theory (DSmT) were used as they offer an appropriate mathematical framework for the modeling of both imprecision and uncertainty and have the ability to consider not only singletons but also compound classes such as union of classes in DST's model and intersection of classes in DSmT's model.

The remainder of the paper is organised as follows. In the next section, we recall the mathematical basis of DST and DSmT and their application to fusion process. Section 3 deals with the way DST and DSmT can be used to multispectral classification of remotely sensed data. Section 4 shows the obtained results when applying the proposed classification methodology on a real satellite data acquired by ETM+ sensor of LandSat 7 satellite. These results are discussed and compared to a Bayesian result. Finally, Section 5 gathers our conclusions.

2. DST and DSmT basis

The evidence theory developed by Dempster [6] and better formalized by Shafer [14] enables to represent both uncertainty and imprecision, and was initially introduced to fuse a conflicting information sources. The plausible and paradoxical theory [7] [15] is a generalization of the classical DST which allows to formally combining any types of information sources: rational, uncertain or paradoxical. Notice that the two theories are based on the definition of an elementary

0-7803-9521-2/06/$20.00

mass function, from which are derived plausibility and belief (or credibility) functions.

2.1. DST Basis

2.1.1. Elementary notions

The theory of evidence needs the definition of a frame of discernment Θ including k exclusive hypothesis θ_i, the k classes in our case where $1 \le i \le k$. A referential 2^Θ represents the set of all subsets of Θ. Plausibility and credibility functions can be expressed with a unique function, the mass function *m( )*. Mass, plausibility and credibility, which are all defined from 2^Θ on the interval [0, 1], characterize the likelihood of any subset A_i where $1 \le i \le 2^\Theta$ of 2^Θ. The mass function is defined as:

$$\begin{cases} m : 2^\Theta \rightarrow [0,1] \\ A_i \rightarrow m(A_i) \end{cases} \quad (1)$$

and

$$\begin{cases} m(\phi) = 0 \\ \sum_{A_i \in 2^\Theta} m(A_i) = 1 \\ m(A_i) \ge 0, \quad A_i \subset 2^\Theta \end{cases} \quad (2)$$

Where ϕ represents the empty set.

Focal elements are the elements A_i of 2^Θ in which the mass function $m(A_i)$ is not null.

The plausibility and belief functions are given by:

$$Pl(A_i) = \sum_{A_i \in 2^\Theta / A_i \cap A_j \neq \phi} m(A_j), \quad \forall A_i \in 2^\Theta \quad (3)$$

$$Bel(A_i) = \sum_{A_j \in 2^\Theta / A_j \subseteq A_i} m(A_j) \quad \forall A_i \in 2^\Theta \quad (4)$$

The notion of plausibility can be introduced in relation to the notion of credibility. It is defined as:

$$Pl(A_i) = 1 - Bel(\overline{A}_i) \quad (5)$$

The uncertainty about a focal element A is represented by the values of the interval *[Bel(A), Pl(A)]*, which is called the "belief interval" and the length of this interval gives a measurement of the imprecision about the uncertainty value.

2.1.2. Evidential combination rule of Dempster

Once the evidence functions (masse, Plausibility and Belief) associated to each n independent information sources S_i $(1 \le i \le n)$ defined on the same frame of discernment are defined, it is then possible to combine them according to the DST's orthogonal combination rule symbolized by the operator $\oplus$. This rule results in:

$$m_1 \oplus m_2 \oplus \quad \oplus m_p(A_i) =$$
$$m_\oplus(A_i) = \bigoplus_{i=1}^{n}(A_i) = \frac{(m_c)(A_i)}{1 - m_c(\phi)} \quad \forall A_i \neq \phi \;\; et \;\; A_i \subseteq 2^\Theta \quad (6)$$

Where $m_c(\phi)$ représents the mass assigned to empty set Φ and is often interpreted as a measure of conflict between the different n sources. It is given as follows:

$$m_c(\phi) = \sum_{A_1 \cap \ldots \cap A_n = \phi} \prod_{j=1}^{n} m_j(A_j) \quad \forall A_i \in 2^\Theta \quad (7)$$

More details about the mathematical properties of DST can be found in reference [14]. In particular, it is shown that the DST's rule of combination is commutative and associative, which allows one to combine the available sources in any order.

2.1.3. Evidential decision rule

After combination of the different sources, a decision is made according to a certain criteria. Several decision rules have been proposed: 1) maximum of plausibility which is judged as the best by some authors [3] [4] [11] [12], 2) maximum of belief over the simple hypothesis which is the most used [11], and 3) maximum of belief without overlapping of belief intervals which is very strict and called absolute decision rule [4] [11] [12].

2.2. DSmT Basis

The DSmT of plausible, uncertain and paradoxical reasoning [7] [8] [9] [15] is a generalization of the classical DST [6] [14] which allows to formally combine any types of sources of information (rational, uncertain or paradoxical). The DSmT is able to solve complex data/information fusion problems where the DST usually fails, especially when conflicts (paradoxes) between sources become large and when the refinement of the frame of discernment Θ is inaccessible because of the vague, relative and imprecise nature of Θ elements. The foundation of DSmT is based on the definition of the hyperpowerset D^Θ (Dedekind's lattice) [8] [9] of a general frame of discernment Θ.

2.2.1. Notion of hyper-powerset D^Θ

The foundation of DSmT is based on the definition of the hyper-powerset D^Θ [8] [9]. Let Θ be a set of k elements θ_i, 2^Θ commonly named a power-set is a set of subsets of Θ when all θ_i are disjoints. The hyper-powerset D^{Θ} is defined as the set of all composite propositions built from elements of Θ with $\cup$ and $\cap$ operators such that:

$$\forall A_i \in D^{\Theta},\ \forall A_j \in D^{\Theta},\ (A_i \cup A_j) \in D^{\Theta} \text{ and } (A_i \cap A_j) \in D^{\Theta}. \tag{8}$$

From a general frame of discernment Θ, is defined a quantity *m(A)* called the generalized basic belief mass for A such that:

$$m(\phi) = 0 \quad et \quad \sum_{A_i \in D^{\Theta}} m(A_i) = 1 \tag{9}$$

The plausibility and belief functions are defined in almost the same manner as within the DST, it means:

$$Pl(A_i) = \sum_{A_j \in D^{\Theta} / A_i \cap A_j \neq \phi} m(A_j), \quad \forall A_i \in D^{\Theta} \tag{10}$$

$$Bel(A_i) = \sum_{A_j \in D^{\Theta}, A_j \subseteq A_i} m(A_j) \quad \forall A_i \in D^{\Theta} \tag{11}$$

These definitions are compatible with the DST definitions when the sources of information become uncertain but rational (they do not support paradoxical information). We still have:

$$\forall A_i \in D^{\Theta},\ Bel(A_i) \leq Pl(A_i) \tag{12}$$

2.2.2. Paradoxical combination rule of Dezert

Let $Bel_1(.)$ and $Bel_2(.)$ be two belief functions over the same frame of discernment Θ and their corresponding generalized basic belief mass $m_1(.)$ and $m_2(.)$ provided by two distinct but potentially paradoxical sources of evidences S_1 et S_2. Then the combined global belief function $Bel(.) = Bel_1(.) \oplus Bel_2(.)$ associated to the fusion process of the two sources, is obtained by combining the information granules $m_1(.)$ and $m_2(.)$ through the Dezert's rule of combination given by:

$$\forall A_i \in D^{\Theta}, \quad m(A_i) \cong [m_1 \oplus m_2](A_i) = \sum_{A_j, A_k \in D^{\Theta},\ A_j \cap A_k = A_i} m_1(A_j) m_2(A_k) \tag{13}$$

For n sources of evidence S_j, the generalized form is given by :

$$\forall A_i \in D^{\Theta}, \quad m(A_i) \cong \bigoplus_{j=1}^{n}(A_i) = \sum_{\substack{A_j \in D^{\Theta} \\ A_1 \cap \ldots \cap A_n = A_i}} \prod_{i=1}^{n} m_j(A_j) \tag{14}$$

This rule of combination is commutative and associative and can always be used for the fusion of paradoxical or rational sources of information (bodies of evidence).

2.2.3. Paradoxical decision rule

The decision rule in the framework of DSmT fusion process is defined in almost the same manner as within the DST, it means by choosing one of the three mentioned criteria.

2.3. Definition of DST and DSmT mass functions

The determination of mass functions in DST and DSmT represents a crucial step in a fusion process and remains a largely unsolved problem, which did not yet find a general answer. In image processing, Bloch [2] [3] dresses three different levels from where a mass function may be derived: at the highest level where information representation is used in a way similar to that in artificial intelligence and masses are assigned to propositions, at an intermediate level, masses are computed from attributes, and may involve simple geometrical models, at the pixel level, mass assignment is inspired from statistical pattern recognition. Recall that the difficulty increases when we are interested on the compound hypothesis and their mass functions. The most widely used approach is to assign to simple hypotheses masses that are computed from conditional probabilities. Then a transfer model is introduced to distribute the initial masses over all compound hypothesis (union of classes in DST and intersection of classes in DSmT). This transfer operation is done through a coarsening (discounting) factor and/or a conditioning factor applying to the conditional probabilities (initial masses).

The literature reported several transfer models: the transferable belief model [16], the upper and lower probability model [6], the parametric model [16], the consonant model [11] [12], the dissonant model [1], etc. In this paper, the mass functions are estimated using a dissonant model of Appriou [1] that was initially developed for two classes only as follows:

$$m_i^j(\{\theta_i\})(x) = \frac{\alpha_i . R_j . P(x/\theta_i)}{1 + R_j . P(x/\theta_i)}, \tag{15}$$

$$m_i^j(\{\overline{\theta_i}\})(x) = \frac{\alpha_i}{1 + R_j . P(x/\theta_i)}, \tag{16}$$

$$m^j(\Theta)(x) = 1 - \alpha_i, \tag{17}$$

Where $P(x/\theta_i)$ is the conditional probability, α_i is a coarsening factor, and R_j represents a normalization factor that is introduced in the axiomatic approach in order to respect the mass and plausibility definitions, and is given by:

$$R_j \in \left[0, \left\{\max_{i \in [1,K]} [\sup(P(x/\theta_i)]\right\}\right] \tag{18}$$

3. DST and DSmT classification approach

In remote sensing, first applications of DST and DSmT were developed in the framework of multisensor/multitemporal/multiscale data fusion [10] [11] [12] [15]. However, in recent studies, the thematic application of DST and DSmT concerns land use and land cover mapping, sometimes, by considering temporal changes [5].

In the present section, we describe the proposed supervised classification approach based on DST and DSmT with the main objective to improve a Bayesian classification. The adopted methodology is as follows [4]:

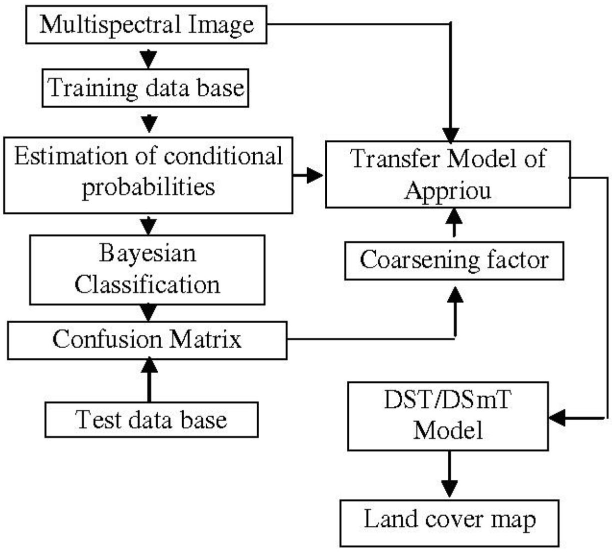

Figure 1. The proposed DST and DSmT classification methodology

1. According to an *"a prior"* knowledge, two data bases are constructed: a training base to be used in a supervised classification process, and a test base to be used during the assessment of the classification accuracy.
2. A Bayesian classification is performed using a maximum likelihood algorithm.
3. A confusion matrix is established between a Bayesian classification result and a test data base.
4. For each class, a coarsening factor is obtained from the confusion matrix and it can be seen as the accuracy of that class which is computed by dividing the total number of correct pixels in that class by either the total number of pixels in that category as derived from the test data base.
5. Mass functions of the individuals and the compound classes are estimated through a transfer model of Appriou that we have generalized and extrapolated for more than two classes as follows:

$$m_j(\theta_i) = \frac{\alpha_i . R_j . P(x/\theta_i)}{1 + R_j . P(x/\theta_i)} - \frac{(2^\Theta - k - 1) * \varepsilon}{k} \quad (19)$$

$$m_j(\theta_1) = m_j(\theta_2) = m_j(\theta_{i-1}) = m_j(\theta_{i+1}) =$$

$$m_j(\theta_k) = \frac{\alpha_i / (k-1)}{1 + R_j . P(x/\theta_i)} - \frac{(2^\Theta - k - 1) * \varepsilon}{k} \quad (20)$$

$$m_j(\theta_1 \cup \theta_2) = m_j(\theta_1 \cup \theta_3) = \ldots = m_j(\theta_1 \cup \theta_2 \cup \ldots \cup \theta_{k-1}) = \varepsilon, \ \forall \varepsilon > 0 \quad (21)$$

$$m_j(\Theta) = 1 - \alpha_i \quad (22)$$

Where k is the number of the considered classes and ε is a sensitivity factor that weighted the mass functions in order to have their sum over all the hypothesis equal to 1.

6. A combination rule of DST or DSmT is applied. For each pixel to classify, its mass functions are combined as follows :

$$m_\oplus(A_i) = (m_1 \oplus m_2 \oplus m_3 \oplus m_4)(A_i) \cong \frac{(m_1 \circ m_2 \circ m_3 \circ m_4)(A_i)}{1 - m_c(\phi)} \quad \forall A_i \neq \phi \quad et \quad A_i \subseteq 2^\Theta \quad (23)$$

Where $m_c(\phi)$ is the global conflict degree between all sources (classes in our case) and is computed for each pixel to be classified.

7. Finally, a multispectral classification is released according to a decision rule. We have chosen a maximum of belief criterion.

4. Results and discussion

The DST/DSmT classification algorithm we described was applied to improve the Bayesian classification result, using a multispectral ETM+ image acquired by Landsat 7 satellite on June 2001. The RGB composition of the data set which covers the north-eastern part of Algiers (Algeria), and the selected data bases of training and testing our algorithms are respectively given by Figure 2, Figure 3, and Figure 4. Four thematic classes dominate the study site: Dense Urban (DU), Less Dense Urban (LDU), Vegetation (V), and Bare Soil (BS).

The Bayesian classification result based on a maximum likelihood algorithm is shown on Figure 5. The assessment of this result relatively to the considered test data gives a confusion matrix of Table 1 on which it is clearly shown a large confusion between DU and LDU, and between BS and LDU. A conflict degree between the considered classes belongs to [0.44, 0.49] and is given by the image of Figure 6. The DST and DSmT classification results are given respectively on Figure 7 and Figure 8. Table 2 shows the different land cover types present on the study site, obtained respectively from a Bayesian, a DST, and a DSmT approaches.

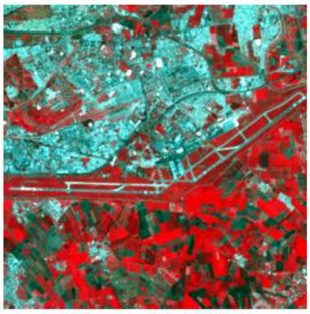

Figure 2. RGB composition

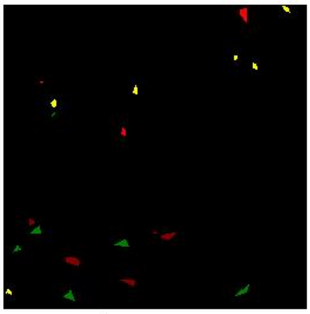

Figure 3. Training data base

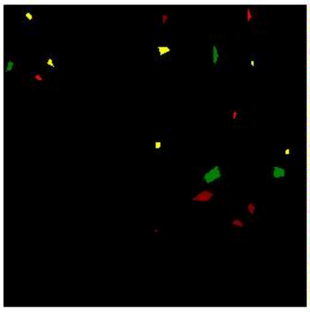

Figure 4. Test data base

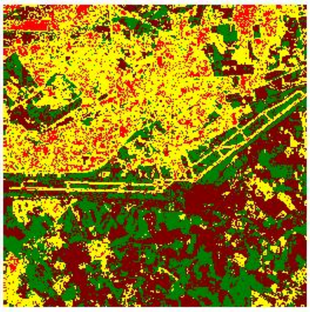

Figure 5. Bayesian result

Table 1. Confusion matrix

	$S_1(DU)$	$S_2(LDU)$	$S_3(V)$	$S_4(BS)$
DU	**28**	14	00	00
LDU	23	**71**	00	17
V	00	00	**212**	13
BS	01	37	13	**125**

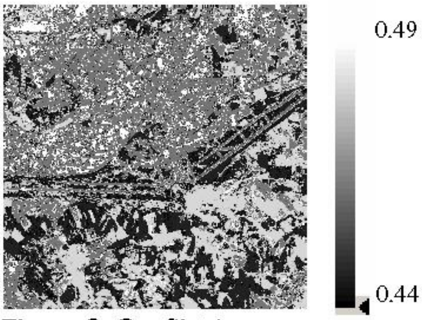

Figure 6. Conflict image

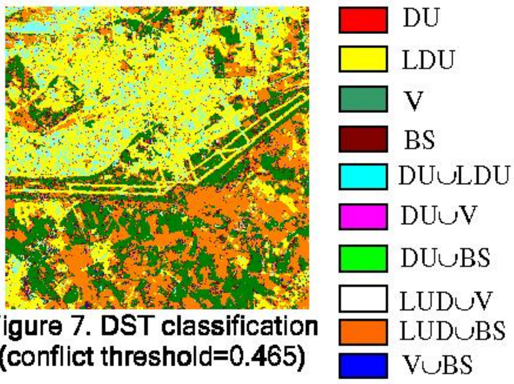

Figure 7. DST classification (conflict threshold=0.465)

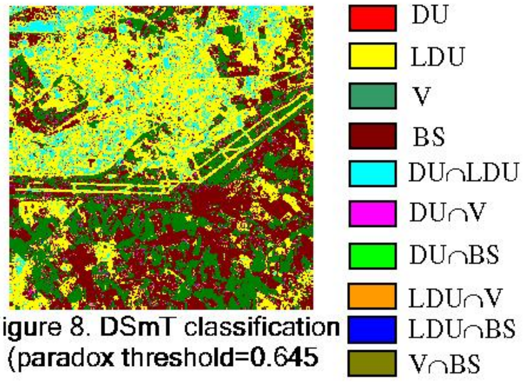

Figure 8. DSmT classification (paradox threshold=0.645

Table 2. Percentage of the different land cover types in a Bayesian, DST, and DSmT approaches

Class	Bayesian class cover (%)	DST class cover (%)	DSmT class cover (%)
DU	8.59	0.77	2.03
LDU	35.18	32.22	23.15
V	24.24	24.00	31.57
BS	31.99	5.41	4.50
DU∪LDU	0	8.93	0
DU∪V	0	0.04	0
DU∪BS	0	0.08	0
LDU∪V	0	0.72	0
LDU∪BS	0	**26.51**	0
V∪BS	0	1.31	0
DU∩LDU	0	0	**17.36**
DU∩V	0	0	0.00
DU∩BS	0	0	0.00
LDU∩V	0	0	4.29
LDU∩BS	0	0	**14.46**
V∩BS	0	0	2.64

It is known that a Bayesian classification result has often a "salt-and-pepper" noise appearance due to many miss-classified pixels especially those located at the segment borders or extremities of the classes. The suggested DST and DSmT classifiers aim to improve the Bayesian land cover map by tacking into account the imprecision and the uncertainty of the acquired data.

DST classifier leads to a land-cover map constituted of “pure zones” being to individual classes (DU, LDU, V, and BS) and “mixed zones” (or ambiguous zones) being to the union of classes (DU∪LDU, DU∪V, DU∪BS, LDU∪V, LDU∪BS, and V∪BS). A decision rule is based on a maximum of belief according to a threshold chosen by the user to decide of the desired conflict degree. As it is seen on Table 2, the land cover types which are the most conflicting on the site are LDU and BS, and they represent **26.51%** of the site.

DSmT classifier leads to a land-cover map constituted of “pure zones” being to individual classes (DU, LDU, V, and BS) and “paradoxical zones” (or very conflicting zones) being to the intersection of classes (DU∩LDU, DU∩V, DU∩BS, LDU∩V, LDU∩BS, and V∩BS). A decision rule is based on a maximum of belief according to a specified-user threshold to decide of the desired degree about the uncertainty on the paradox between the classes. Notice that unlike DST classification result where the union of classes represents the ignorance existing between these classes, in DSmT classification result, the intersection of classes represents new spectral classes having a common spectral response between those of classes of the intersection. For example, DU∩LDU is a thematic class between a dense urban and a less dense urban, which is surely an urban zone but with an average density. Thanks to the DSmT classifier, finer heterogeneous classes may be detected. As it is shown on Table 2, the land cover types which are the most paradoxical on the site are DU and LDU, and they represent **17.36%** of the site, and LDU and BS, and they represent **14.46%** of the site.

5. Conclusion

Two supervised classifiers of multispectral remotely sensed data have been presented in this paper. The first one is based on Dempster-Shafer Theory (DST) and the second based on Dezert-Smarandach Theory (DSmT). The main purpose of these classifiers is to improve the result of the Bayesian classification by modeling the conflicting/paradoxical nature of the considered classes.

The particularity of the proposed methodology is that we are dealing with a thematic class as one source of information or a sensor. So, DST and DSmT have been adapted to consider this modification in order to design multispectral classifiers through a multisource fusion process. The most important step in the framework of DST/DSmT fusion process is the definition of the mass function. In this work we have adopted a transfer model of Appriou that we have generalized for more than two sources of information (thematic classes). The combination rule of Dempsetr/Dezert allows to combine the mass functions of individual and compound classes, and using a criterion of maximum of belief a most realistic class of each pixel is selected. In this manner, DST classifier attributes union of classes to the conflicting pixels, and DSmT classifier attributes intersection of classes to the paradoxical (very conflicting) pixels. These proposed classifiers have effectively improved a Bayesian classification result on which only about 60% of land cover types has been confirmed as “a pure zone”, the remainder 40% has been detected as “a mixed zone”. For further work, it would have been interesting to see a spectral behavior of the compound classes through an analysis of their spectral signature.

Finally, this study shows that DST and DSmT represent a powerful mathematical tool to design successfully multispectral classifiers of satellite data.

6. References

[1] A. Appriou, “Multisensor signal processing in the framework of the theory of evidence”, *NATO/RTO, Application of Mathematical Signal Processing Techniques to Mission Systems*, 1999.

[2] I. Bloch, “Information Combination Operators for Data Fusion: A comparative review with classification”, *IEEE, Trans. Sys. Man Cybern,* Vol. 26, 1996, pp. 52-67.

[3] Bloch I., *Fusion d’informations en traitement du signal et des images*, Edition Hermès Science, Paris, France, 2003, 319 p.

[4] A. Bouakache, “Fusion des images satellitaires par la thèorie d’évidence et la théorie du raisonnement plausible et paradoxal“, Thèse de Magister en traitement du signal et de l’image, Université des Sciences et de la Technologie Houari Boumediene (USTHB), Alger, Algérie, 2005, 95p.

[5] S. Corgne, L. Hubert-Moy, J. Dezert, and G. Mercier, “Land cover change prediction with a new theory of plausible and paradoxical reasoning”, CNRS, Université Rennes 2, ISIF, France, 2003, pp. 1141-1148.

[6] A. P. Dempster, “Uper and lower probabilities induced by a multivalued maping”, *Annals of Mathematical Statistics,* Vol. 38, n° 2, 1976, pp. 325-339.

[7] J. Dezert, “An introduction to the theory of plausible and paradoxical reasoning”, Proc. of NM&A 02. Conf., Borovetz, Bulgaria, Aug, 2002, pp. 20-24.

[8] J. Dezert, and F. Smarandache, “On the generation of hyperpowersets for DSmT”, Proc. Of Information fusion 2003. Conf., Cairns, Australia, 2003.

[9] J. Dezert, and F. Smarandache, “Partial ordering of hyper-powersets and matrix representation of belief functions within DSmT”, ISIF, 2003, pp. 1230-1238.

[10] L. Hubert-Moy, S. Corgne, G. Mercier, and B. Solaiman, “Land use and land cover change prediction with the theory of evidence. A study case in Intensive Agricultural Region in France”, proceedings of Fusion 2002 conference, Annapolis, MD, USA, 2002, pp. 114-122.

[11] S. Le Hégarat-Mascle, I. BLOCH, and D. Vidal-Madjard, “Application of Dempster Shafer evidence theory to unsuprvised classification in multisource remote sensing”, *IEEE Transactions on Geosciences and Remote Sensing,* Vol.35, N°. 4, 1997, pp. 1018-1031.

[12] S. Le Hégarat-Mascle, D. Richards, and C. Ottlé, “Multi-scale data fusion using Dempster-Shafer evidence theory”, *Integrated Computer-Aided Engineering*, IOS press, 2003.

[13] Richards, J.A., *Remote Sensing Digital Image Analysis, An Introduction,* 2nd Edition, Springer-Verlag, Berlin, 1993.

[14] Shafer, G., *A Mathematical Theory of Evidence,* Princeton University Press, Princeton (NJ), 1976, 312 p.

[15] Smarandache, F., and J. Dezert, *Advances and Application of DSmT for Information Fusion*, American Research Press, 2004, 418 p.

[16] PH. Smets, and R. Kennes, “The transferable belief model”, *Artificial Intelligence 66*, 1994, pp.191-234.