

Brice Mora<sup>1</sup>, Richard A. Fournier<sup>1</sup>, Samuel Foucher<sup>2</sup>

1: Centre d'Application et de Recherche en Télédétection (CARTEL),  
Université de Sherbrooke, Québec, Canada.

2: Centre de Recherche en Informatique de Montréal, Québec, Canada.  
{brice.mora,richard.fournier}@usherbrooke.ca,samuel.foucher@crim.ca

# Map regenerating forest stands based on DST and DSMT combination rules

Published in:

Florentin Smarandache & Jean Dezert (Editors)

**Advances and Applications of DSMT for Information Fusion**

(Collected works), Vol. III

American Research Press (ARP), Rehoboth, 2009

ISBN-10: 1-59973-073-1

ISBN-13: 978-1-59973-073-8

Chapter XXI, pp. 529 - 547

**Abstract:** *Our results demonstrated the ability of the Free Dezert-Smarandache (DSm) model to improve thematic classification of forest regeneration over the use of Dempster-Shafer Theory (DST) and a classical Maximum Likelihood Algorithm (MLA). Overall, a classification accuracy of 82.75% was obtained with the reference method, MLA but it was improved by 7.4% by applying the fusion method DST (90.14%). Further improvement of 1% (to 91.13%), compared to those from the DST, was modest but noticeable when using the free DSm model. The study also showed the critical aspect of the design of the mass functions of each ancillary source and the difficulty to model the associated vagueness and uncertainty. Finally, the ability of the algorithms to take advantage of data fusion provided an excellent tool to test various combinations. After testing series of potential inputs, we found that drainage and surface deposit were the two best ancillary inputs in addition to spectral information to improve classification on the growth potential of regenerating forest stands in Southern Québec.*

## 21.1 Introduction

This impetus of our work spurred from the necessity to improve map accuracy in the regenerating forest stands of the Province of Québec in Canada to facilitate forest management in general and more specifically field operations. Current forest inventory maps divide the landscape into polygons of uniform characteristics based on stand type, tree density and average height as delineated by an experienced photo-interpret. The polygon dimensions are larger than 2ha and we wish to develop a method that provides information at a finer resolution. Another limitation is the update frequency of the maps: inventory cycles imply production of a new map every 8 years. We wish to develop a method that can provide information between inventory update.

The Maximum Likelihood Algorithm (MLA) is commonly used in forest mapping context [5, 14] and thus can be used as a reference result for the fusion algorithms. The limitation of the MLA for thematic classification lies with its incapacity to deal with heterogeneous data (nominal and ordinal data). Thus, only satellite imagery is usually used to produce such maps. We selected the Dempster-Shafer Theory (DST) and the Dezert-Smarandache Theory (DSmT) with its free DSm model to improve mapping accuracy of regeneration for their ability to fuse satellite imagery with heterogeneous and complementary data but also for their ability to deal with data uncertainty and vagueness. We therefore compared the results from the DST and those when free DSm model was used. Results obtained in [18, 19] suggested that free DSm model was more adapted to deal with conflicting fusion cases compared with DST and we felt it needed to be tested further for our purpose.

The main objective of our study is to test if DST and DSmT based on the free DSm model allow improving map accuracy for area under regeneration. In such case, specific objectives are included to compare results with MLA and also to assess the best supplementary input for data fusion to improve the results. This work was extracted from a study with extended objectives which will be submitted by Mora *et al.* [11]. This chapter below focused only on the fusion case that provided the best results.

## 21.2 Reasoning theories

### 21.2.1 Dempster-Shafer theory (DST)

Unlike the theory proposed by Bayes [1], the works from Dempster [4] and Shafer [17] allows fusing sources of information. Data fusion using DST takes into account the uncertainty and the vagueness linked to the data and the knowledge that we have about their influence on a given purpose. The following description is a reminder of the theoretical bases of the fusion method.

The first step of data fusion with DST involves defining the frame of discernment  $\Theta$  that includes all the classes of the stratification:

$$\Theta = \{\theta_1, \theta_i, \dots, \theta_n\}. \tag{21.1}$$

Then a power set  $2^\Theta$  is deduced from  $\Theta$  including all the subsets of  $\Theta$  and the empty set  $\emptyset$ . For instance, for 3 singleton hypotheses we have :

$$2^\Theta = \{\theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3, \emptyset\}. \tag{21.2}$$

In DST, Dempster’s combination rule allows fusing information sources with mass functions describing all states of each source. These mass functions can be equated to a confidence level given to each focal element, i.e. each element of  $2^\Theta$  with a non null mass. Thus, the mass functions  $m(\cdot)$  of each hypothesis of  $2^\Theta$  will comply with the following requirements, for a given source:

$$\begin{aligned} m: 2^\Theta &\rightarrow [0,1], \\ \sum_{A \in 2^\Theta} m(A) &= 1, \\ m(\emptyset) &= 0. \end{aligned} \tag{21.3}$$

The combination rule (21.4) combines the sources two by two according to the mass functions defined at the previous step. If we fuse three sources, a second iteration will fuse the third source with the results of the first combination. The same process can be expanded to larger number of sources. The combination rule is associative and commutative. This means that the order for which the sources are combined is not important. Thus for two distinct sources characterized by their belief masses  $m_1(\cdot)$  and  $m_2(\cdot)$ , the combination rule is written as  $m(\emptyset) = 0$  and  $\forall C \in 2^\Theta \setminus \{\emptyset\}$ :

$$m(C) = [m_1 \oplus m_2](C) = \frac{\sum_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A)m_2(B)}. \tag{21.4}$$

The denominator of (21.4), also represented by the letter  $K$ , equals zero if the sources are completely contradictory. In such case it means that conflict between sources, symbolized by  $k$ , equals 1 knowing that:

$$K = 1 - k. \tag{21.5}$$

Zadeh [20] showed that in the case of highly conflicting combinations, the DST can provide counterintuitive results. Some authors proposed different solutions to solve this problem. We decided to test the DS<sub>m</sub>T which has been designed specifically to assess conflicting cases.

### 21.2.2 Dezert-Smarandache theory (DSmT)

DSmT is a generalization of the DST for dealing with conflicts and/or paradoxical hypotheses [18, 19]. This generalization brings a more adapted framework to take into account conflicts existing between sources. There is a suite of DSm models available according to the application [18, 19]. Among all the different adaptations of DSm models we selected the free DSm model for our study because of its ability to deal with conflicts and for its simplicity of implementation.

The classic DSm combination rule (DSmC) works with the free DSm model and keeps the properties of commutativity and associativity of the DST. A hyper-power set is now derived from the frame of discernment  $\Theta$ . This set is built with disjunctive and conjunctive operators  $\cup$  and  $\cap$ . Consequently for the frame of discernment presented in (21.1), the derived hyper-power set  $D^\Theta$  will be as follows:

$$D^\Theta = \{\theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_1 \cap \theta_2, \emptyset\}. \tag{21.6}$$

The requirements to build mass functions for each focal element of  $D^\Theta$  are identical to what was presented for the DST. The DSmC rule of combination for two distinct sources is defined as  $m(\emptyset) = 0$  and  $\forall C \in D^\Theta \setminus \{\emptyset\}$ :

$$m(C) = [m_1 \oplus m_2](C) = \sum_{A, B \in D^\Theta, A \cap B = C} m_1(A)m_2(B). \tag{21.7}$$

As we can see in (21.7), the parameter  $k$  representing the conflict in the DST combination rule (21.4) disappeared. Now the conflict (or the paradox) is represented by every composed class resulting in the intersection of two singleton hypothesis.

### 21.2.3 Decision rule

Various decision rules are proposed in the literature. The most common are the maximum credibility and maximum plausibility and the pignistic probability. For our study we choose to deal only with two singletons hypotheses. Consequently, the maximum credibility decision rule was chosen for its simplicity of implementation. Indeed in this case, all the other common decision rules cited above will provide the same decision. For a hypothesis  $A$ , it is computed as:

$$Cr(A) = \sum_{B \subseteq A} m(B). \tag{21.8}$$

## 21.3 Information used in this work

### 21.3.1 Study area

The study area is located in the Watopeka forest located in Southern Québec, Canada with a center latitude and longitude at 45°35'00" N 71°46'00" W. The study area can

be delimited by a square of 50km<sup>2</sup> in which 2.5km<sup>2</sup> is occupied by regenerating forest stands. This forest is mainly composed of maple species, yellow birch and coniferous species like balsam fir, jack pine and black spruce. Due to this species composition, this forest is dedicated to wood production for a paper factory. The local climate can be defined as "continental, sub-humid". The mean altitude of the area varies from 250 to 400m above sea level. The growth season varies from 170 to 190 days per year and the cumulative number of day degrees varies from 2400 to 3400°C.

### 21.3.2 Satellite imagery

We used a multispectral SPOT-5 HRG image taken on September 9 2002 and covering the study area. The image was orthorectified using a DEM (Digital Elevation Map which was interpolated with the Spline method applied to contour lines (1:20 000) extracted from the Québec topographic database. The SPOT-5 image was composed of pixels with 10m spatial resolution. Such spatial resolution is a good compromise between the lower resolution provided by Landsat images at 30m and the very high spatial resolution images (e.g. QuickBird, IKONOS) ranging from 0.6 to 4m. Landsat images do not provide sufficient spatial resolution to identify efficiently spatial patterns of regenerating forest often in stripes. In contrast, very high spatial resolution satellite images at the level of 1 to 4m provide a sufficient level of details but are far more complex to process. In addition to supplying with a suitable spatial resolution to identify regeneration areas, multispectral SPOT-5 image also offers a good compromise between cost and total surface covered.

### 21.3.3 Sample plots

We collected field sample plots for the three classes of stand regeneration: Deciduous commercial species, Non commercial, Conifers. The main commercial deciduous found in the study area were maple sugar and yellow birch. Non commercial species included shrubs, ferns and typical species from humid sites like lycopods, horsetails. The conifer class included balsam fir, jack pine and black spruce. For each of the three classes, plot localization was chosen at random inside known areas having regenerating stands in the study area. During the field visit, a GPS reference value was taken at the center of each sample plot for location and the following attributes were recorded as an average considering all trees in the plot: species composition, density, age, and height, approximate radius of stand homogeneity from the center. Sample plots diameter could vary depending on the homogeneity of the species distribution. This did not add any difficulty in the analysis as we did not need to compare these plot diameter and all plots had a minimum radius of 12m which allowed all plots to be used in the analysis. We superimposed the satellite image over the sample plots to assigned pixels that corresponded to each plots. Table 21.1 provides the number of sample plots and their associated number of pixels obtained for each class.

|              | Commercial deciduous | Non Commercial | Conifers |
|--------------|----------------------|----------------|----------|
| Sample plots | 61                   | 53             | 36       |
| Pixels       | 334                  | 168            | 164      |

Table 21.1: Numerical description of each class of the stratification.

The spectral separability of the three classes was evaluated by using the pixel radiometric values at the location of each plot. We first applied the Jarque-Bera test [9] which examined the normality of radiometric values of each class. Each class was evaluated for each of the four bands of the satellite image. Half of the combinations were proved to be normal. The others were rejected with an alpha coefficient of 5 or 1%. Then we computed the Bhattacharyya distance to examine the separability of the class distributions. Results provided a good separability between classes from 1.31 to 1.46 (Table 21.2) knowing that a perfect separability is equal to 2. According to these tests we decided to use the maximum likelihood algorithm (MLA) as the first reference test to compare its results from those of the fusion algorithms. We divided randomly the datasets into two parts in order to obtain first a dataset for the training of the MLA (66% of the sample plots) and second another dataset to evaluate the results of the classification (34% remaining).

|                | Commercial deciduous | Non Commercial |
|----------------|----------------------|----------------|
| Non Commercial | 1.43                 | /              |
| Conifers       | 1.31                 | 1.46           |

Table 21.2: Bhattacharyya distance on the sample plots distributions.

### 21.3.4 Drainage and surface deposit

In our first series of tests to map regeneration of forest stands, we only considered two pedological attributes, surface deposit and drainage, among all the potential biophysical parameters involved in the growth potential of forest stands. The value of these two attributes will be added to the spectral values as input of the classification methods using evidential reasoning. We focused on these two pedological attributes because they have been identified in the scientific and professional literature as the major explanatory variables for the stand growth for ecosystems. Roy *et al.* [15, 16], Gagnon and Roy [6], Robitaille [13] and MRNQ [12] provide more details on the importance of surface deposit and drainage on the spatial distribution of the species in Eastern Canada. Both attributes were available from the maps published by the provincial government of Québec which served as a base for forest inventory. These maps are produced from the interpretation of aerial photography taken at a scale of 1 : 15000.

## 21.4 Methods

A flowchart of the steps required to apply our method is given in Figure 21.1. Before applying the classification methods we identified the areas with regenerating stands from the interpreted provincial forest inventory and created a spatial mask to apply the analysis over that area only. Then we processed to image classification using the MLA. The result of this reference method allowed mapping the coniferous areas in the regenerating stands. We therefore identified pixels of the image and in the regeneration stands that were dominated by conifer trees. These pixels can therefore be treated separately in the analysis. Once the delimitation of regeneration area is completed, we start the analysis for belief assignment, i.e., we defined the mass functions for the two remaining classes: Commercial deciduous and Non commercial. Then we processed the data fusion according to the DST and DS<sub>m</sub>T with the decision rule of the maximum credibility. The results are validated by comparing with a reference method (MLA) or with the validation plots.

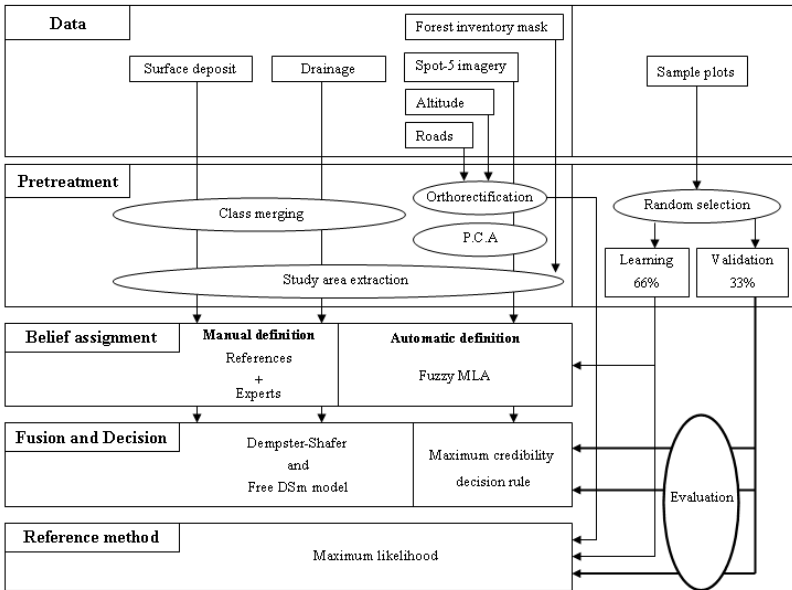


Figure 21.1: Flowchart of the methodology.



### 21.4.1 Reference methods

The first step of the process consisted in pre-identifying the regenerating stands with the provincial forest inventory. This allowed reducing the number of classes including only the regeneration stands to be considered in the study area. Then we were able to mask the satellite image only on these areas which were relevant for the analysis. We applied the MLA on the masked image by using the training sample plots. The results were not only used as a reference to evaluate the performance of the fusion methods but were also used to identify the coniferous pixels in the regenerating stands.

### 21.4.2 Data fusion methods

We were not able to define mass functions for the conifer class of the stratification because of a lack of references leading to support mass function values. Therefore the fusion process was applied only to pixels of the two other remaining parts of the regenerating stands identified as "Commercial deciduous" or "Non commercial".

The belief assignment was processed according two specific ways, one for the satellite image and one for the ancillary sources. We used the Fuzzy Statistical Expectation Maximization algorithm (FSEM) [7] to define the mass functions for the satellite image. This supervised multi-iterative method is based on Gaussian distribution classes and compute posterior probabilities. The use of the FSEM requires having strictly independent sources. Consequently we used as data input the first two principal components (90% of the variance) of a principal component analysis (PCA) applied to the four spectral bands of the satellite imagery. The FSEM has the ability to produce fuzzy classes. This automatic way to design the mass functions was only applied to the satellite image because it was not possible to obtain normal distributions with the other two ancillary sources of input: surface deposit and drainage.

We designed the mass functions of the two ancillary sources manually according to the references and some expert interviews. Corgne [3] and Cayuela *et al.* [2] also adopted this way to define mass functions of their models. Firstly, the references indicated in what way each source had a positive influence on the growth development of the deciduous species of interest. This focuses mostly on the sugar maple, the most common species of interest in the area. Secondly, we designed the mass functions so that the sum of all masses of pure classes was equal to 1. A normalization occurred later in order to integrate the mass of fuzzy classes to follow the rule defined by (21.3). Thus, at this step we have the following relationship for two pure classes:

$$m(\theta_2) = 1 - m(\theta_1). \quad (21.9)$$

In the case of our study, we can replace the class name "Deciduous species" by  $\theta_1$  and "Non commercial species" by  $\theta_2$ . At this point we had values for mass functions only for pure classes. However the next step impose that we define new masses for union classes or fuzzy classes and that we renormalize with these new mass values.

We defined the mass functions for fuzzy classes in the discounting framework [10, 17]. This helps also to weaken the bba's associated with sources believed to be less reliable or of lesser importance for the fusion procedure. The discounting method is defined in our 2D case as follows:

$$m'(\theta_1) = \alpha \cdot m(\theta_1), \quad (21.10)$$

$$m'(\theta_2) = \alpha \cdot m(\theta_2),$$

$$m'(\theta_1 \cup \theta_2) = (1 - \alpha) + \alpha \cdot m(\theta_1 \cup \theta_2). \quad (21.11)$$

Given the vagueness and uncertainty related to the sources (scale digitization, quality of the manufacturing process), we fixed the coefficient  $\alpha$  to a value of 0.5 empirically. According to (21.11), this is equivalent to considering the mass of the fuzzy class as the mean of the masses of the pure classes before the normalization. On the one hand this choice appeared as the best compromise to model the vagueness and the uncertainty of the sources. On the other hand, we were not able to define the masses of the fuzzy class manually on a scientific basis. Then we applied a linear normalization to follow the requirement of (21.3).

References from the scientific literature provided the necessary information to define the influence of each state of the sources on the growth development of the deciduous species. In other words, we were able to design the general shape of the mass functions. Then, we had to interpret numerically the specific influence of each source when the reference did not provide such information. This was done empirically so we applied a sensitivity analysis to assess the influence of the variation of the mass values on the quality of the fusion.

- *Drainage*

Roy *et al.* [16] established a curve linking the drainage with dieback rate of forty deciduous forest stands in Southern Québec. We used this curve to quantify the influence of soil drainage on growth development of the sugar maple (see Table 21.3). We noticed that the two levels 'Excessive' and 'Fast' were not found in our study area. The codes of Table 21.3 correspond to the provincial forest inventory standards.

- *Surface deposit*

According to Roy *et al.* [15, 16], Gagnon and Roy [6], Robitaille [13] and MRNQ [12] and also to expert interviews, we translated the influence of the amount of clay and the thickness of the soils on the growth of sugar maple as indicated in Table 21.4. These codes also corresponded to the Québec provincial forest inventory standards.

- *Data fusion*

As motivated above, the fusion model did not consider conifer pixels. Identification of conifer pixels was accomplished with the MLA. When applying

| Level     | Code | Mass of $\theta_1$ | Mass of $\theta_2$ |
|-----------|------|--------------------|--------------------|
| Excessive | 0    | 0                  | 1                  |
| Fast      | 1    | 0.69               | 0.31               |
| Good      | 2    | 0.77               | 0.23               |
| Moderate  | 3    | 0.77               | 0.23               |
| Imperfect | 4    | 0.60               | 0.40               |
| Bad       | 5    | 0                  | 1                  |
| Very poor | 6    | 0                  | 1                  |

Table 21.3: Mass values for the drainage.

| Type   | Code              | Mass of $\theta_1$ | Mass of $\theta_2$ |
|--|-------------------|--------------------|--------------------|
| Thin organic deposits  | 7T                | 0.15               | 0.85               |
| Thin glacial deposits  | 1aM               | 0.3                | 0.7                |
| Medium thickness glacial deposits  | 1aY               | 0.4                | 0.6                |
| Thick glacial deposits   | 1a                | 0.5                | 0.5                |
| Juxtaglacial deposits,<br>Proglacial deposits,<br>Ancient fluvial deposits | 2A,<br>2B,<br>3AN | 0.7                | 0.3                |
| Glaciolacustral deposits   | 4GS               | 0.9                | 0.1                |

Table 21.4: Mass values for the surface deposit.

the DST algorithm we tested the possible combinations of the results of the PCA from the SPOT-5 image with one and two ancillary sources. The results from data fusion using the DST were compared with those from the reference method (MLA).

We applied a Hill-Smith test [8] to study the link between the masses, the quality of the result obtained by the DST and the conflict level (Figures 21.2 and 21.3). The result shows a positive correlation between the conflict and the misclassified pixels which justified the use of DSMT with the free DSMT model to fuse the sources. In the application of the free DSMT model we followed the same procedure as the DST fusion process. Here, we fused the sources with a total transfer of fuzzy masses to the paradoxical class as in Corgne [3].

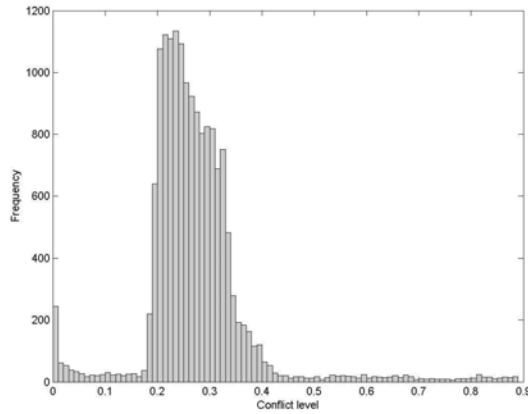


Figure 21.2: Histogram of the conflict for the best source combination.

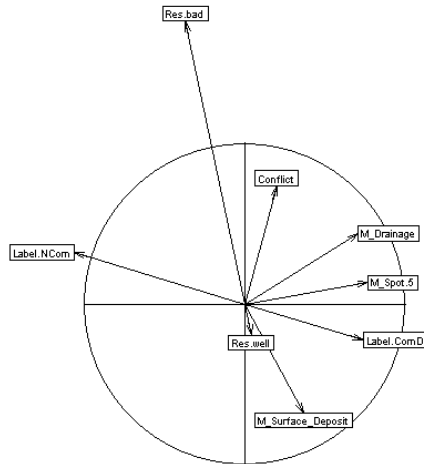


Figure 21.3: Correlation circle of the Hill-Smith test on the best DST fusion parameters. The prefix "Res" means "Result", "NCom" refers to the Non commercial class and "ComD" refers to the Commercial deciduous class.

## 21.5 Results and their interpretation

### 21.5.1 Results based on the maximum likelihood algorithm

Table 21.5 presents the correct classification results obtained using the MLA with the three classes of the original stratification and those obtained after adding the conifers to the mask. These results served as a base of comparison for those that were obtained with the fusion algorithms. The "two classes" case provided better results (82.75%) than the "three classes" case (70.03%). This can be explained by the reduction of the confusion produced by the removal of the Conifers class. For the "two class" case both classes were well classified; above 90% for the Commercial deciduous class and above 70% for the Non commercial class. Such good results were expected in view of the normality of spectral values within the sample plots and the Bhattacharrya distance obtained for the classes (Table 21.2).

|                      | ComD   | NCom   | Conifers | Mean   |
|----------------------|--------|--------|----------|--------|
| "three classes" case | 85%    | 40.96% | 81.48%   | 70.03% |
| "two classes" case   | 90.83% | 71.08% | /        | 82.75% |

Table 21.5: Results of the MLA with according to the number of classes. "NCom" refers to the Non commercial class and "ComD" refers to the Commercial deciduous class.

### 21.5.2 Results based on the fusion in DST framework

Comparison of results obtained with the MLA with those obtained with the FSEM (Table 21.6), shows that the FSEM was less efficient than the MLA to classify the satellite image. Consequently we decided to stop the FSEM after one iteration in order to obtain a fuzzy MLA classification. In fact the first iteration of the FSEM fixed the prior probabilities for each of the  $n$  classes at  $1/n$  which is equivalent to applying the MLA. When compared to the FSEM (Table 21.6), the fuzzy MLA provided an improvement of 6.4% on the overall accuracy. The result provided by the MLA used as a reference result was lower than the one obtained with the fuzzy MLA by about 1%. This can be explained by the fact that the last method computed the masses for a third class (the fuzzy class). This induced a new distribution of mass values which led to a new hierarchy between the singleton classes. Moreover when using the fuzzy MLA, the input data were not the spectral bands but the bands provided by a PCA. Therefore this can lead to a slight difference in the results from the MLA applied to the spectral bands of the satellite image.

|           | Commercial deciduous | Non commercial | Mean   |
|-----------|----------------------|----------------|--------|
| MLA       | 90.83%               | 71.08%         | 82.75% |
| FSEM      | 96.67%               | 49.40%         | 77.34% |
| Fuzzy MLA | 85%                  | 81.93%         | 83.74% |

Table 21.6: Results obtained by the MLA and the FSEM.

Table 21.7 presents the results obtained when the satellite image was fused with one and both ancillary sources. The fusion of the satellite image with the Surface deposit or the Drainage provided better results compared with those from the MLA and the fuzzy MLA, respectively by +3.94% and +3.45%. The best results were obtained with the fusion of Surface deposit with the PCA values of the SPOT-5 image. Adding the second ancillary source (Drainage) to the combination provided a small but noticeable improvement of +2.46% on the overall accuracy.

|                               | Commercial deciduous | Non commercial | Mean   |
|-------------------------------|----------------------|----------------|--------|
| Drainage                      | 95%                  | 75.90%         | 87.19% |
| Surface deposit               | 90%                  | 84.34%         | 87.68% |
| Drainage /<br>Surface deposit | 95.83%               | 81.93%         | 90.14% |

Table 21.7: Results obtained using DST framework.

Table 21.8 provides information about the conflict level for the whole area and within the validation sample plots for the fusion of the satellite image and both ancillary sources. As shown in Figure 21.2 the mean conflict level in the image is not high (0.27) but some pixels have high values (until 0.89). The Hill-Smith test showed a positive correlation between the conflict level and the misclassified pixels. From the table we noticed that the validation sample plots were not within the highest conflicting areas (maximum conflict value of 0.42). Nonetheless, the fusion with DSmT and the free DSm model remains relevant. Next section presents the results obtained with this fusion method.

|                            | Minimum | Mean  | Standard deviation | Maximum |
|----------------------------|---------|-------|--------------------|---------|
| Whole area                 | 0       | 0.27% | 0.1                | 0.89    |
| Validation<br>sample plots | 0       | 0.24% | 0.07               | 0.42    |

Table 21.8: Conflict levels for the fusion with both ancillary sources.

### 21.5.3 Results based on the fusion in DSMT framework

Table 21.9 presents the results obtained for the fusion of the satellite image with the ancillary sources. The use of the free DSMT model induced a small improvement of 0.49% of the overall accuracy for the DST fusion using the combination of the satellite image with the Drainage. Applying the free DSMT model using only Surface deposit with the PCA bands provided worse results for the Commercial deciduous class and also induced a slight decrease of the overall accuracy by 0.49%. The best results were obtained with the combination of the PCA values of satellite image with the Drainage and the Surface deposit. It induced an improvement of 1% on the overall accuracy compared to the DST.

|                               | Commercial deciduous | Non commercial | Mean   |
|-------------------------------|----------------------|----------------|--------|
| Drainage                      | 93.33%               | 79.51%         | 87.68% |
| Surface deposit               | 89.16%               | 84.33%         | 87.19% |
| Drainage /<br>Surface deposit | 95%                  | 85.54%         | 91.13% |

Table 21.9: Best results obtained using DST framework and the free DSMT model.

## 21.6 Sensitivity analysis

### 21.6.1 Mass functions of the ancillary sources

Because some mass function values were determined empirically, we decided to apply sensitivity tests. This analysis implied varying the mass values through their potential range to study the impact of the initial choice. We aimed at preserving the shape of the curves which represents the hierarchy between the state values of the source. Figure 21.4 represents the evolution of the overall accuracy according to the mass variations. Note that the mass variation displayed in the x-axis is the variation of the Commercial deciduous class. When the mass values of one hypothesis are increased, this automatically leads to an improvement of the overall accuracy of the class, and a decrease in the other one. Figure 21.4 also shows that the initial mass values provided the best overall accuracy which is the best compromise for the quality of the detection of both classes.

### 21.6.2 Discounting coefficient

Previously we justified why we chose to fix the  $\alpha$  coefficient at a value of 0.5. Nonetheless we studied the variation of this parameter that could influence the overall accu-

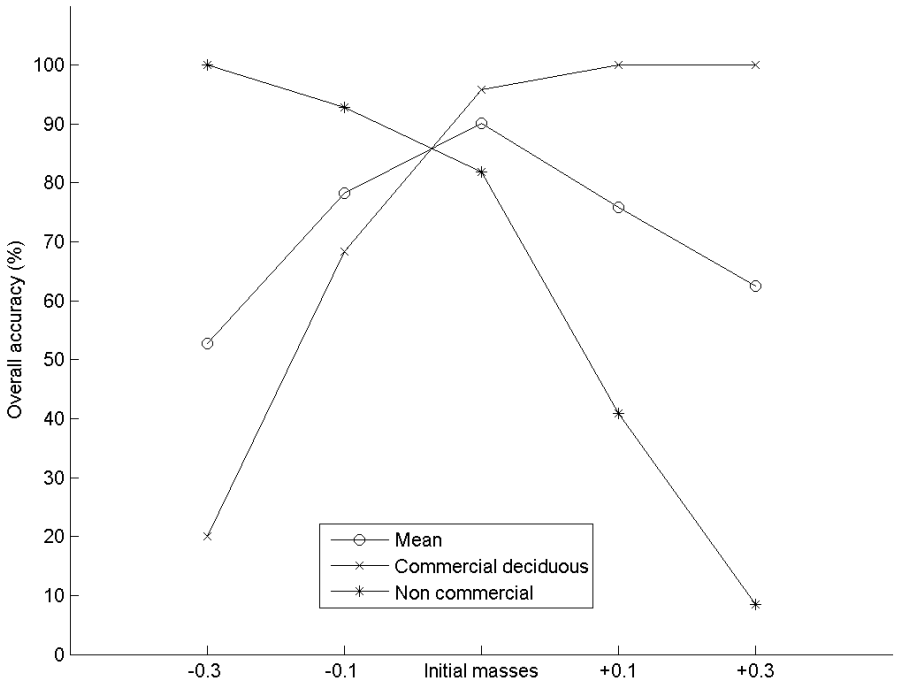


Figure 21.4: Evolution of the overall accuracy according to the variation of the masses for the combination with two ancillary sources.

racy. Figure 21.5 represents the evolution of the results according to the variation of this coefficient. A zero value for  $\alpha$  means that the discounted mass functions will be equivalent to a Bayesian belief structure and will be very specific whereas a value equals to 1 will transform the mass functions to a non-informative belief structure. It shows the small influence of the discounting coefficient on the overall accuracy. Only a slight improvement is obtained with the highest values ( $\alpha = 0.8$  and  $0.9$ ) of the coefficient. With a value of  $\alpha = 0.8$ , the identification of the Non commercial class is improved by 1.35%. With a value of  $\alpha = 0.9$ , the identification of the Non commercial class is improved by 4% and the ability to identify Commercial deciduous decreased by a value of -0.82%. Thus we realized the small impact of the discounting coefficient on the overall results for our study.



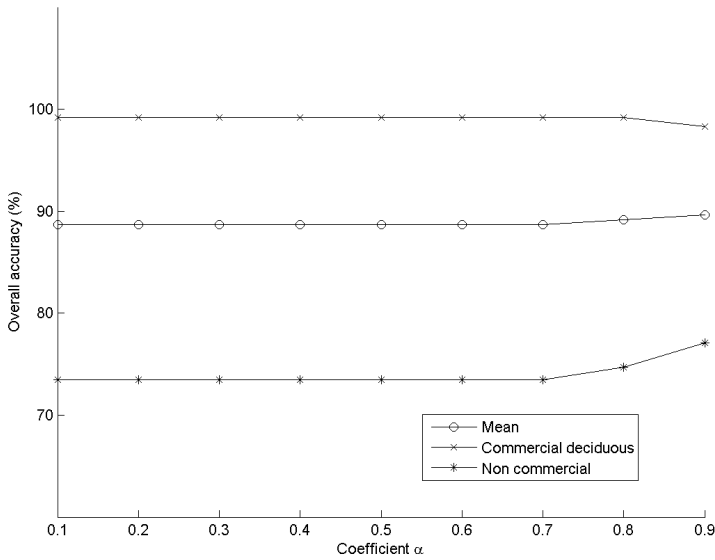


Figure 21.5: Evolution of the overall accuracy according to the value of  $\alpha$  for the combination with two ancillary sources.

## 21.7 Discussion

From our preliminary runs of the algorithms we quickly realized that results from multiple iteration of the FSEM were not as useful as using only the first iteration (which is equivalent to the fuzzy MLA). It seems to be an interesting way to compute automatically the masses of the spectral bands of a satellite image. We think that this automatic way is preferable than an empirical one. However the normality of the spectral information within the training samples has to be considered. Also, the classification resulting from the fuzzy MLA gave an interesting overall accuracy (83.74%) and balanced results for each class (both above 80%). This also confirmed the interest of the fuzzy MLA.

Our study also showed the difficulty to establish the mass functions of the ancillary sources. This is due to the lack of scientific knowledge about the influence of the attributes on the growth potential of regenerating forest stands. Moreover the sensitivity tests showed the high sensitivity of the results to the design of the mass functions. We also showed the low influence of the discounting coefficient on the global quality of the fusion but we still think the way we used the discounting frame

by fixing  $\alpha$  at a value of 0.5 is a best way to deal with the lack of knowledge we have about the influence of each ancillary source and their quality. This value is a compromise because it equals the mean of the masses of the pure hypotheses before the linear normalization that considers the mass of the fuzzy class. Nonetheless, we advise to test different  $\alpha$  values for each source to be fused according to its quality and the uncertainty about the design of the mass functions. For example the scale of each source could be considered for this. We could not test this in our study because all the sources had the same scale.

Note that for this first study about the regenerating forest stands, we only considered pedological attributes. Some references and experts, also cite topographic and hydrographical information as attributes of interest for our purpose. Thus, in order to improve the results, we should review all potential sources of information. Also, on a technical point of view, a way to improve the results may be to condition the transfer of the fuzzy mass to the "intersection class" according to conflict level encountered during the DST fusion. Lastly we benefited in our study from only modest improvements while applying the free DS<sub>m</sub> model. This might be partially due to the fact that validation plots were not taken in high conflict areas (see Table 21.8). This situation shows the importance of plot distribution to make sure they are also present in areas presenting high conflict. Therefore, the level of conflict should always be tested as a prior indication to choose the most relevant fusion method.

## 21.8 Conclusion

Our study showed the ability of DS<sub>m</sub>T and the free DS<sub>m</sub> model to improve the classification results (91.14%) compared to those from DST (90.14%) and also those from reference method like MLA that is typically used in forestry (82.75%). DS<sub>m</sub>T using the free DS<sub>m</sub> model gave better results than the DST but only with a small improvement of 1% which indicates that the DST provided most of the improvements in accuracy that was expected for the purpose of mapping stand regeneration. Traditional methods like MLA use satellite image as their only source of information. Data fusion methods proposed in DST and DS<sub>m</sub>T allow the inclusion of other parameters that are known to explain forest regeneration. In our case it allowed to model the influence of surface deposit and drainage which are both known by forester to influence the growth potential of regenerating forest stands in Southern Québec. As a continuation of this contribution, further studies should focus on the mass function structuring prior to the fusion. This is a recurring issue for any project wishing to adopt DST model framework. An adapted uncertainty to each source should provide better results. This would be done according to the quality of the data (statistics about its accuracy, the scale . . .).

## 21.9 References

- [1] T. Bayes, *An Essay towards solving a Problem in the Doctrine of Chances*, By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S, Philosophical Transactions of the Royal Society of London, 53, 370-418, 1763.
- [2] L. Cayuela, J.D. Golicher, J. Salas Rey, J.M. Rey Benayas, *Classification of a complex landscape using Dempster-Shafer theory of evidence*, International Journal of Remote Sensing. Vol. 27, No. 10, pp. 1951-1971, 2006.
- [3] S. Corgne, *Modélisation prédictive de l'occupation des sols en contexte agricole intensif : application à la couverture hivernale des sols en Bretagne*, Thèse de doctorat, Université de Rennes 2 & Haute Bretagne, 2004.
- [4] A.P. Dempster, *A generalization of Bayesian inference*, Journal of the Royal Statistical Society, Series B 30, pp. 205-247, 1968.
- [5] D. Ducrot, *Méthodes d'analyse et d'interprétation d'images de télédétection multi-sources - Extraction de caractéristiques du paysage*, Habilitation à diriger des recherches, INP Toulouse, France, 2005. [http://www.cesbio.ups-tlse.fr/data/\\_all/theses/HDR/\\_DD/\\_2005.pdf](http://www.cesbio.ups-tlse.fr/data/_all/theses/HDR/_DD/_2005.pdf).
- [6] G. Gagnon, G. Roy, *Le dépérissement de l'éradle à sucre (Acer saccharum Marsh.) au Québec*, Revue Forestière Française (Nancy) 46(5), pp. 512-521, 1994.
- [7] M. Germain, M. Voorons, J.M. Boucher, G.B. Benie, *Fuzzy statistical classification method for multiband image fusion*, Proceedings of the Fifth International Conference on Information Fusion, Vol. 1, pp. 178-184, 2002.
- [8] M.O. Hill, A.J.E. Smith, *Principal component analysis of taxonomic data with multistate discrete characters*, Taxon, 25, pp. 249-255, 1976.
- [9] C.M. Jarque, A.K. Bera, *A test for normality of observations and regression residuals*, International Statistical Review 55, pp. 163-172, 1987.
- [10] D. Mercier, T. Denoeux, M.H. Masson, *General Correction Mechanisms for Weakening or Reinforcing Belief Functions*, 9th International conference on Information fusion, pp. 1-7, Florence, Italy, 2006.
- [11] B. Mora, R.A. Fournier, S. Foucher, *Application of evidential reasoning to improve regenerating forest stands mapping*, to be submitted.
- [12] MRNQ, *Rapport de classification écologique - Érablière à tilleul de l'Est*, Programme de connaissance des écosystèmes forestiers du Québec méridional, Service de l'évaluation de l'offre/Service de la recherche appliquée/Direction des inventaires forestiers/Direction de la recherche forestière/Forêt Québec. Ministère des Ressources Naturelles du Québec, Février 1999.

- [13] A. Robitaille, J.P. Saucier, *Paysages régionaux du Québec méridional*, Direction de la gestion des stocks forestiers et la direction des relations publiques du ministère des Ressources naturelles du Québec, 220 pages, 1998.
- [14] J. Rogan, S.R. Yool, *Mapping fire-induced vegetation depletion in the Peloncillo Mountains, Arizona and New Mexico*, International Journal of Remote Sensing, Vol. 22, No. 16, pp. 3101–3121, 2001.
- [15] G. Roy, A. Sauvesty, F. Pagé, R. Van Hulst, C. Ansseau, *A comparison of soil fertility and leaf nutrient status of sugar maples (*Acer saccharum*) in relation to microrelief in two forests in Québec*, Canadian Journal of Soil Science, 82 (1), pp. 23–31, 2002.
- [16] G. Roy, L. Robitaille, G. Gagnon, *Étude des principaux facteurs du dépérissement des érablières au Québec*, Phytoprotection, 66, pp. 91–99, 1985.
- [17] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, 1976.
- [18] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion, (Collected Works)*, American Research Press, Rehoboth, June 2004.  
<http://fs.gallup.unm.edu/DSMT-book1.pdf>.
- [19] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion, (Collected Works)*, American Research Press, Rehoboth, 2006.  
<http://fs.gallup.unm.edu/DSMT-book2.pdf>.
- [20] L.A. Zadeh, *On the Validity of Dempster's rule of Combination of Evidence*, Memo M79/24, Univ. of California, Berkeley, 1979.