Combination of supervised and unsupervised classifiers based on belief functions

Na Li

▶ **To cite this version:**

Na Li. Combination of supervised and unsupervised classifiers based on belief functions. Image Processing [eess.IV]. University of Rennes I, 2020. English. NNT : . tel-03029781

HAL Id: tel-03029781
https://hal.science/tel-03029781

Submitted on 29 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Informatique*

Par

Na LI

Combination of supervised and unsupervised classifiers based on belief functions

Thèse présentée et soutenue à Lannion, le 28/09/2020
Unité de recherche : IRISA
Thèse N° :

Rapporteurs avant soutenance :

Sylvie LE HEGARAT-MASCLE	Professeure des universités, Paris-Sud 11
Cédric WEMMERT	Professeur des universités, Université de Strasbourg

Composition du Jury :

Président :	Sébastien LEFEVRE	Professeur des universités, IRISA, Université de Bretagne Sud
Examinateurs :	Sylvie LE HEGARAT-MASCLE	Professeure des universités, Paris-Sud 11
	Cédric WEMMERT	Professeur des universités, Université de Strasbourg
	Zhunga LIU	Professeur, Northwestern Polytechnical University, China
	Jean DEZERT	Senior Research Scientist, DTIS, ONERA
	Rémi ESTIVAL	Ingénieur de Recherche, Total, Pau
Dir. de thèse :	Arnaud MARTIN	Professeur des universités, IRISA, Université de Rennes 1

List of Tables

List of Figures

Chapter 1

Introduction

Resume

Land cover classification based on multi-sensor data fusion has great potential for practical applications, especially for missions on complex nature areas. A large number of labeled samples is important in land cover classification. However, collecting sufficient ground information, especially in hard-to-access areas, is usually difficult and expensive. Moreover, labels and remote sensing data are often collected in different periods, so that objects on the ground can be mislabeled. Therefore, our research interest focuses on how to improve the accuracy of land cover classification by fusion of multiple sources when limited labeled samples are available.

In this chapter, we present the challenges and limitations in industrial problems that arose in the METIS (Multiphysics Exploration Technology Integrated System) project initiated by TOTAL. Problems in a natural or industrial context are often difficult to define precisely and clearly. On the contrary, problems in an academic context need to be well described and located. Therefore, the main objective of this chapter is to explain how we transfer multiple industrial requirements into academic problems.

1.1 Introduction

Land cover relates to the biophysical cover of the Earth's terrestrial surface, identifying vegetation, water, bare soil, or impervious surfaces, *etc.* Identifying land cover is essential for planning and managing natural resources (*e.g.* development, protection), understanding the distribution of habitats, and for modeling environmental variables. Identification of land cover types provides basic information for the generation of other thematic maps and establishes a baseline for monitoring activities. Therefore, land cover classification using satellite data is one of the most important applications of remote sensing.

A great deal of ground information (*e.g.* labeled samples) is usually required to generate high-quality land cover classification. However, in complex natural areas, collecting information on the ground can be time-consuming and extremely expensive. Nowadays, multiple sensor technologies have gained great attention in land cover classification. They bring different and complementary information—spectral characteristics that may help to overcome the limitations caused by inadequate ground information. In the METIS project initiated by TOTAL, multiple sensors are available, including sensors for multi-spectral and hyper-spectral imaging, synthetic aperture radar (SAR), and light detection and ranging (LiDAR). Nevertheless, labeled samples in the study area are severely inadequate. Multiple sensors can generate different land cover maps to provide heterogeneous yet supplementary information of the objects on the ground. Accordingly, in our thesis, we manage to improve the accuracy of land cover classification with limited labeled data by the combination of heterogeneous information from multiple sensors.

Another problem caused by the lack of ground information is the ambiguous relations between land cover maps and land use maps. Land cover maps provide information on natural features that can be directly observed on the Earth's surface. Land use maps refer to how people are using information on landscapes for different purposes. Without adequate ground information, generating land use maps by land cover maps is difficult for complex areas. For example, in the METIS project, we are interested in a land use map (*i.e.* accessibility map) to describe the difficulty to move to some location or to traverse some area on the ground. However, the relations between accessibility and land covers are still difficult to determine because the term accessibility involves many factors, such as time, risk, cost *etc.* Therefore, many ground surveys are required to exploit the real-time information on the ground so that the concept of accessibility can be well-defined in terms of users' objectives, and that the relations between accessibility and land covers can be determined. Therefore, when combining multiple heterogeneous land cover maps, we have to consider how to enable users to synthesize the scheme for land use maps, such as accessibility with different criteria.

In our research, we focus on the fusion of heterogeneous information from different sources. The combination system aims to solve the problems caused by limited labeled samples and can thus be used in land cover classification for hard-to-access areas. The semantic labels for the land cover classification from each sensor can be different, and may not corresponds to the final scheme of labels that users await. For instance, land cover classification methods of different sensors provide semantic labels for the ground. However, based on these land cover maps, an accessibility map is supposed to be generated to meet users' needs. Therefore, another objective of the combination is to provide an

interface with a final scheme probably different from the input land cover maps present.

In this chapter, we present the challenges and limits of the real problems in section 1.2. In section 1.3, we present our motivations and contributions, focusing on the modeling of these challenges and limitations. We give an overview of this thesis in section 1.4. More details on the METIS project of TOTAL are presented in the appendix A.

1.2 The real-world challenges and limitations

This section explains some major challenges and limitations related to our research in real-world applications. Ground surveys often play an essential role, however, it is usually time-consuming and expensive, sometimes even risky. Limited ground surveys bring many inconveniences, such as difficulties in the labeling process, detailed in section 1.2.1, and problems in land use maps, detailed in section 6.2.4. Besides, fusion of multi-sensor data is also a challenging task in reality because of many factors, such as heterogeneity of data. Related details are presented in section 1.2.3.

1.2.1 Difficulties in labeling process

Sufficient labeled data with high quality is one of the prerequisites in land cover classification. Despite various proposed land cover scheme, none of the current schemes has been internationally accepted because the definition of land covers greatly changes due to different purposes. The major process to define land cover scheme includes two phases: an initial dichotomous phase, and a subsequent modular-hierarchical phase, shown in Figure 1.1. In the dichotomous phase, major land cover types (*e.g.*, ice/snow, water, bare) are defined. In modular-hierarchical phase, environmental and technical attributes are incorporated in the definition to refine the major land cover types defined in the dichotomous phase. Environmental attributes (*e.g.*, climate, altitude, time, and erosion) easily influence land covers possibly leading to mislabeled. Technical attributes (*e.g.*, floristic aspect, crop types, and soils types) are also sometimes considered in defining land cover types for specific purposes. However, specific technical disciplines are often required during a ground survey, making the labeling process expensive and time-consuming. We take a recent land cover scheme proposed by NASA [142] as an example in Figure 1.2. The major land cover classes contain seven types: ice/snow, water, bare, developed, forest, herbaceous, and shrub. Below each of them, more refined class types are defined based on environmental attributes, marked in blue text, and/or specific technical attributes, marked in green text. Some classes related to environmental attributes, such as open shrub and sparse shrub, possibly change in a short time, and their boundaries become less clear to observe on the ground. Furthermore, specific knowledge is required to distinguish phenological sub-classes of forest or shrub, types of bare soil, and rocks.

In multi-sensor systems, data are often collected with different resolutions. Therefore, it is highly possible that the resolution of data required to label does not correspond to the scale of ground survey areas. For example, if the ground survey is conducted by the areas with 30m×30m, collected labels possibly contain information from more than one land covers in each square. Remote sensed data with higher resolution than 30m

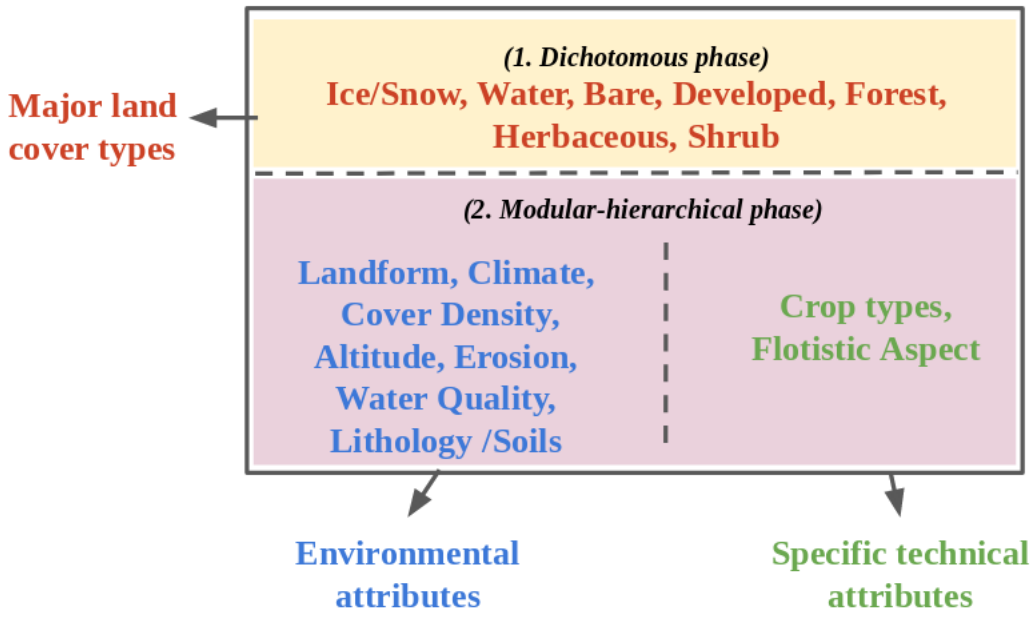


Figure 1.1. Overview of the land cover classification scheme.

are easily incorrectly labeled because more detailed information lacks within squares less than 30m×30m. Furthermore, remotely sensed data is also possible to be localized in an incorrect geographic location, especially when the scale of ground survey areas and data resolution is highly different.

Limited accessibility usually occurs in hard-to-access areas such as mountain areas, which makes ground survey extremely difficult. The collection of labeled data in this situation is usually so expensive, risky, and time-consuming that gathering sufficient labeled data is impossible. Besides manually labeling process on the ground, airborne photos and images captured by drones can also help for generating semantic labels. This process, however, is usually expensive in real-life applications, so that is more suitable for labeling simple landscapes such as vegetation area, water, bare soil, *etc.* More complicated details, related to vegetation type, density, soil humidity, and ground situations, are usually hard to manually describe merely based o airborne photos. It is also highly possible to generate wrong labels in this process.

For some complicated landscapes, the labeling process is usually difficult due to many reasons including but not limited the following:

1) No clear boundary exists between two land cover classes so that they are difficult to distinguish on the ground.

2) Rapid changes of land covers lead to the difference between the ground labels and corresponding remote sensed data.

3) Specific knowledge/techniques are required to distinguish some land cover types on the ground.

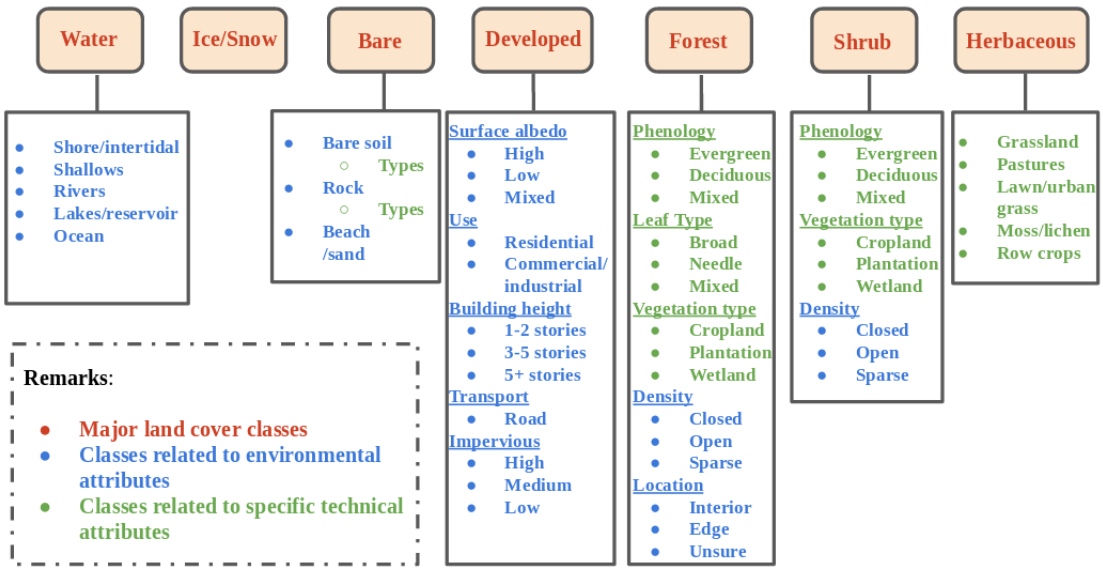


Figure 1.2. Land cover scheme with multiple refined class types based on environmental and specific technical attributes.

4) The scale of ground survey areas does not correspond to the resolution of remotely sensed data, especially in multi-sensor applications.

5) Geographic location on the ground does not correspond to the remotely sensed data.

6) Limited accessibility on hard-to-access areas makes the ground survey difficult or even impossible.

Therefore, collecting sufficient and high quality labeled data is often difficult due to technical or economical limitations. Classification results generated by a limited amount of training data are usually uncertain and incorrect. The limitations caused by insufficient labeled data is one of the major focus of the thesis.

1.2.2 Generation of land use maps by land cover maps

Land use and land cover are two commonly used classification systems describing the circumstance on a given location. Land use and land cover are often confused as interchangeable terms in remote sensing applications. However, they describe different characteristics of the landscape and have different usages in reality. Land cover maps provide information on nature features (*e.g.* vegetation) or man-made constructions (*e.g.* buildings) that can be directly observed on the Earth's surface. For example, a land cover map can document how a region is covered by physical land types, including water, forests, wetlands, impervious surfaces, and other land covers. Land use maps refer to how people are using the information on landscapes for different purposes, such as wildlife habitat, conservation, *etc.* Land use maps do not describe directly the surface cover on the ground whereas they are related to land covers to some extent.

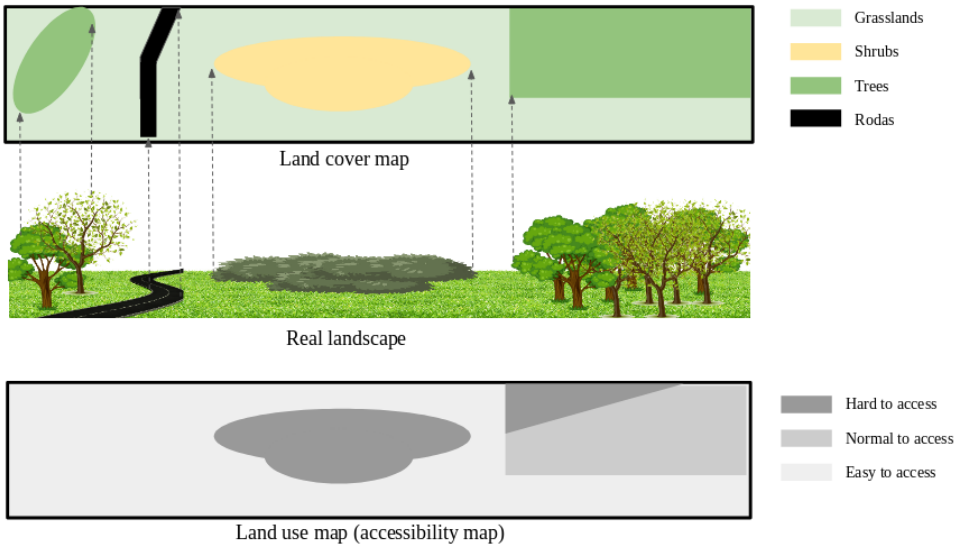


Figure 1.3. Land cover map VS. land use map.

Land cover data usually are interpreted from remote sensing satellite imagery and/or aerial photos because they are merely related to physical characteristics on the surface. Land use data, by contrast, sometimes can be determined from satellite imagery or aerial photos, but not usually. Land use maps often incorporate ancillary data, such as elevation, topographic information, and ground surveys since useful information can be difficult to interpret from imagery, as shown in Figure 1.3.

Classes contained in a land use/land cover map is called as scheme in remote sensing classification, and it also refers to the frame of discernment in belief functions, which are detailed in section 2.4. Each class in the land use/land cover scheme should neither be conflicting nor overlapping in their definitions.

In our research, land cover data are provided by different sensors, such as multi-spectral satellites, SAR or LiDAR, based on which land cover maps can be extracted. However, land use data related to accessibility cannot be directly interpreted by sensors, and even is hard to define by ancillary data.

Determining the accessibility of different land covers is difficult for the following reasons:

1) The definition of accessibility is based on multiple criteria and thus could be ambiguous or even contradictory sometimes. For example, accessibility can be related to the cost of energy, the consumption of time, the risk of traversing, and the expense of budget.

2) A land cover types may have different accessibility in the same landscape or different ones. For example, grassland on low elevation could be easier to traverse than that on a high elevation.

3) Different land covers could also have the same accessibility in the same landscape or different ones. For example, different vegetation could have the same accessibility

based on energy cost.

Up till now, the definition of accessibility has still been ambiguous and vague, because its determination depends on further observations and surveys on the ground. Therefore, the scheme of land use maps related to the displacement model has not been available. What we can directly obtain from satellites and other sensors is merely the land cover maps. Therefore, to transfer the different land cover maps to the land use map is of great importance.

1.2.3 Multi-sensor and heterogeneous information fusion

The information from different sensors such as multi-spectral satellites, SAR or LiDAR is highly different. It is thus essential to combine heterogeneous information from multiple sensors. Unlike traditional fusion system that is designed to improve the accuracy of land cover classification, besides improvement of accuracy, we also focus on generating land use maps that can be later defined by users. As we presented in section 1.2.1 and 6.2.4, the accessibility is neither well-defined nor used to label data on for a learning process. All information provided by sensors and ground survey is used to generate land cover maps. A clear gap exists between what the users are looking for and what sensors can provide. Therefore, in this thesis, we focus on developing a fusion system which cannot only combine heterogeneous information but also handle the gap, so that land use classes can be defined by users after fusion.

Figure 1.4 describe the overview of the fusion system required in METIS. This fusion system can achieve the combination of data from different sensors and more important, the fusion is supposed to be sensor-independent so that any input sensor can be replaced by others. The processing of data from multiple sensors is extremely different and they are usually managed by different experts separately. Focusing on land cover maps after processing makes the fusion system become sensor-independent and the related details are displayed in section 2.3.

The major difficulties in the fusion system are the fusion of classification maps on different land cover schemes into the same scheme. Not all sensors use the same land cover scheme to make their land cover maps. It is thus still possible to have some conflicts in different land cover schemes. Land cover maps, in this case, are also heterogeneous information. Land use classes, moreover, may be different from any land cover schemes from sensors due to users' definition. The fusion system has to consider these and meanwhile to handle insufficient and low-quality training samples, as referred in section 1.2.1.

Another difficulty in our research, as we mentioned in section 6.2.4, is the ambiguous relations between different land covers and their accessibility. Tracking the problem of accessibility would require lots of ground measurements. Defining, measuring, and processing the accessibility with multiple criteria of land covers is a time and cost consuming research work. Therefore, as an input of the fusion system, the land use scheme (*i.e.* accessibility classes) has not been available in our research. This difficulty has been considered in the modeling of real-world problems. Besides, despite the availability of multiple sensors, no ground truth has been available in the study area provided by TOTAL. Meanwhile, processing techniques of SAR and LiDAR are extremely complicated

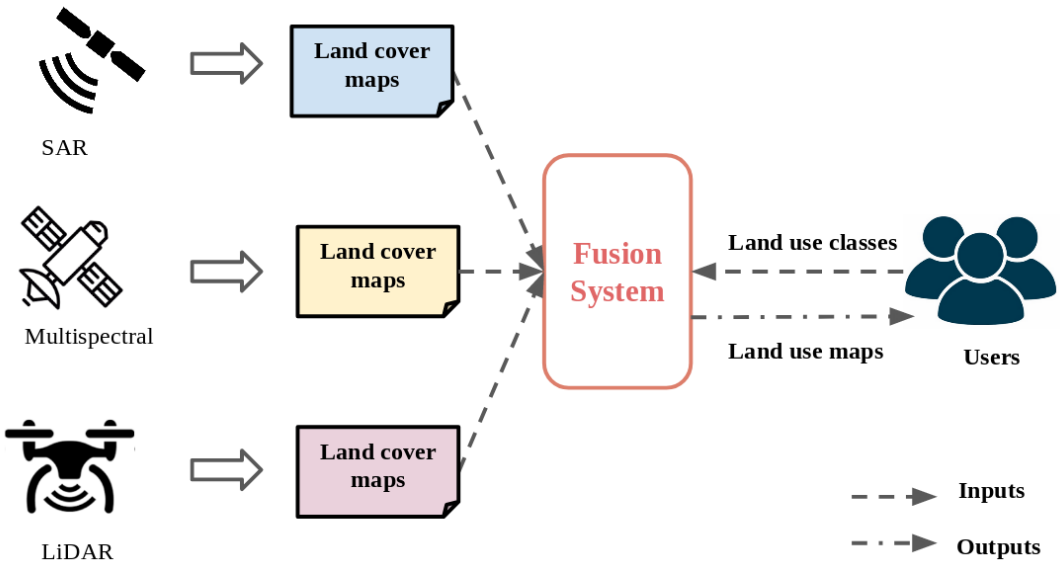


Figure 1.4. Overview of the fusion system.

and time-consuming. Due to limited time and energy in our research, we focus more on the architecture of the fusion system rather than processing different data. In our model, data related to multi-sensors is also simplified. Details on how to model the real-world challenges are presented in the following section 1.3.

1.3 Motivations and contributions

This section focuses on modeling the real-world challenges and limitations as research problems. As mentioned in the previous section 1.2, the major problems arrive in our research including:

(1) Only limited and low-quality labeled data are available in the real situation;

(2) Complicated and time-consuming processing of data from SAR and LIDAR;

(3) Heterogeneity (*i.e.* using different land cover schemes) of land cover maps make the fusion difficult;

(4) Concepts and important factors related to accessibility are not available.

In section 1.3.1, we give our proposals to handle the problem (1), we simplify the problem (2) and (3) and formulate a specific research problem in section 1.3.2. The problem (4) is considered in section 1.3.3. In addition, specific academic contributions are detailed in section 1.3.4.

1.3.1 Handling limited labeled data

Limited and low-quality labeled data cause many problems in real applications, which also has a major influence on land cover classification accuracy. Due to many constraints such as economic factors, we can encounter the limited labeled data problem in METIS. Considering the specific context of our research, we propose some directions to compensate for the inconveniences caused by labeled data.

The fusion of multiple land cover maps is an efficient way to improve the overall land cover classification accuracy when each land cover map shows poor performance. Related details about fusion process are presented in the following section 1.3.2 and 1.3.3. It is important to notice that overcoming some limitations from insufficient labeled data and improving accuracy to some extent is one of the benefits provided by fusion processes.

Furthermore, we also consider handling uncertainty and imprecision to generate a soft fusion framework. Uncertainty indicates that some object belongs to certain land cover class with some belief degree. Imprecision describes some objects could belong to the union of two or more land cover class with some belief degree. When lacking sufficient labeled data, we manage to take advantage of information from uncertainty and imprecision. Considering the low accuracy of each land cover map, it is more reasonable not to decide on the label of an object in a crisp way. Therefore, we focus on a soft fusion framework to handle uncertainty and imprecision.

1.3.2 Fusion of multiple supervised and unsupervised methods

As we mentioned previously, complicated data processing techniques of SAR and LiDAR and heterogeneity of land cover maps are two major difficulties in reality. In this section, we explain how to simplify and model these difficulties into a specific research problem which becomes the core of our research. The fusion system is supposed to be sensor-independent, which indicates that we focus more on land cover maps from different sensors instead of data processing and classification methods themselves. Therefore, we use land cover maps generated by different classifiers based on multi-spectral sensors to replace those from multi-sensors, as shown in Figure 1.5. Different sensors capture distinct information on the ground, and have their own data processing techniques, and apply specific classification methods to generate land cover maps. The inputs of the fusion system are different land cover maps instead of initial raw data, intermediate features, or classification methods themselves. That means the fusion system handles directly the high-level information — land cover maps and ignores the low-level information — sensor types, data, features, and classifiers. Accordingly, processes to generate land cover maps by multiple sensors are so complicated yet subordinate that we choose to simplify them as land cover classification by different classifiers based on multi-spectral data.

No matter how we generate different land cover maps, one thing we have to keep in mind is that land cover maps in reality could be incompatible on account of the use of different land cover schemes. In other words, a class defined in one scheme may be conflicting or overlapping with another one in a different scheme. For example, one land cover scheme has *Forest*, *Non-forest* as classes; and another scheme includes *Vegetation* and *Non-vegetation*. Obviously, a class in one scheme is not the refined class in the other

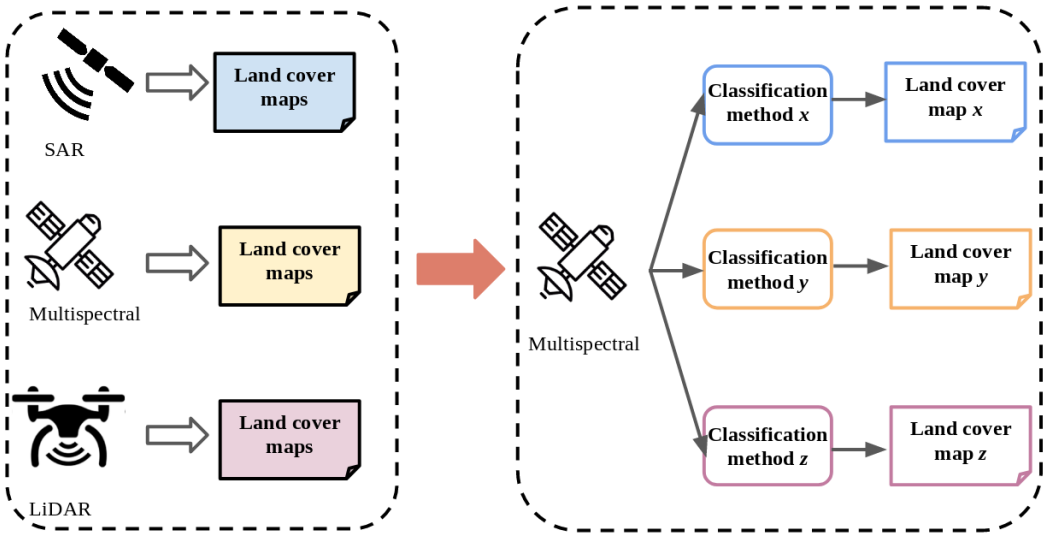


Figure 1.5. Simplify fusion of land cover maps from multiple sensors as the fusion of land cover maps from multiple classifiers based on multi-spectral data.

one, which brings conflicts. An object belonging to *Vegetation* could be *Forest* but also *Non-forest*. On the contrary, however, an object labeled as *Non-forest* may belong to *Vegetation* but could also be *Non-Vegetation*. Intersections between the two schemes (*i.e. Forest*, *Non-vegetation*) are not the refinement of the original classes and they are apparently not exclusive. Accordingly, finding an agreement between these two schemes becomes impossible, which brings a great deal of difficulties to handle heterogeneity in fusion process.

Let us continue with the previous example. Due to conflicts between the two schemes, the semantic meaning of land cover classes in one scheme becomes meaningless for those in the other. We can thus only keep the scheme that we are interested in and consider the other one as a result of clustering, which separates data by their spectral properties yet regardless of semantic labels. In this way, we keep the information that helps to distinguish these land cover classes while ignoring the semantic meaning of labels which brings conflicts with other land cover maps. Therefore, the fusion system focuses on the fusion of multiple supervised and unsupervised classification methods as shown in Figure 1.6, which is also the essence of our research.

1.3.3 A fusion architecture possible to define land use maps by users

In the previous section 6.2.4, we have presented the gap existing between land cover map and land use map (accessibility map). The inputs of the fusion system are various land cover maps that describe situations on the ground from multiple perspectives. The output for users, however, is supposed to be a land use map that provides information on the accessibility of land covers. The difficulty is that, as we previously mentioned, concepts related to accessibility have still been vague and ambiguous, so that we cannot

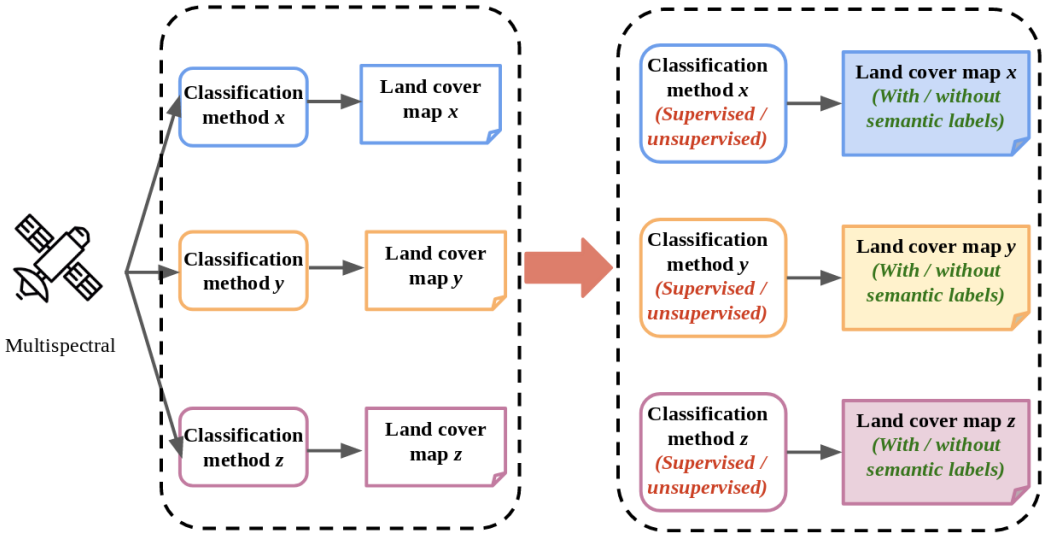


Figure 1.6. Model heterogeneity caused by incompatible land cover schemes by fusion of supervised and unsupervised classification methods.

directly use them in our research. Since our major efforts concentrate on the fusion of multiple supervised and unsupervised methods, we have not done deep and far researches on accessibility. Due to the required amount of ground surveys, we leave the related definitions and concepts on accessibility to users. The fusion system, as shown in Figure 1.7, is supposed to take land use maps defined by users as an optional input when it is available.

Land cover maps generated by multiple sensors may have incompatible schemes so that they cannot reach an agreement. What we can do is to keep the schemes we are interested in, and then remove the semantic meanings of land cover classes in those incompatible ones, that is to say, regarding them as clustering results. When the optional input — accessibility information is available, we can incorporate it into the kept schemes to generate the final accessibility map for users. More details are discussed in chapter 6 on future work.

1.3.4 Contributions

In this section, we briefly resume the major contributions of this dissertation. We classify them according to the area they naturally belong to.

In the area of belief functions, the main contribution is:

• A new methodology to transfer mass functions into another discernment. The main properties of our method that we want to emphasize are the following ones: (i) the transformation can preserve both uncertainty and imprecision from the original frame of discernment and transfer them to the target frame, (ii) the transformation is based on similarity measures, which have multiple choices, (iii) the transformation can also be extended to a learning version, using training samples to learn the similarities.

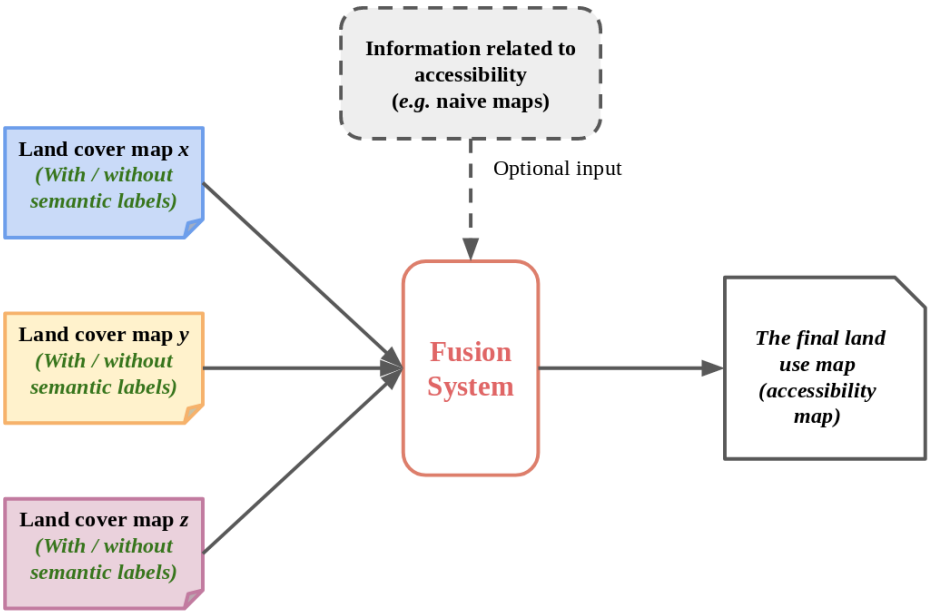


Figure 1.7. The fusion system takes accessibility information as an optional input.

In the area of fusion of multiple supervised and unsupervised methods, our contribution is:

• A new soft fusion framework based on belief functions that can combine heterogeneous information and handle uncertainty. Based on the transformation method we proposed, results from either supervised or unsupervised methods can be projected into the discernment that we are interested in. In this framework, we also develop an iterative fusion strategy that helps gradually to reduce uncertainty and make the overall accuracy converge. We evaluate this fusion framework on remote sensing data in land cover classification, and the experiments show its evident efficiency, especially when training samples for supervised methods are insufficient or mislabeled.

In the area of land cover classification, the contribution is:

• A new water detection method based on belief functions. This method takes into account not only decisions on singletons, but also on total ignorance. In this way, potential water on the ground can also be detected.

• An automatic land cover classification method that considers only spectral indexes so that it can directly generate semantic labels without using training samples.

1.4 Plan of the thesis

In this chapter, we have introduced the industrial context of the problems and the difficulties of working with real data. We have also explained how to model real tasks into specific research problems. In this dissertation, we essentially focus on the fusion of multiple supervised and unsupervised methods and manage to apply the fusion system in land cover classification.

In chapter 2, we present state-of-the-art of the land cover classification using multi-spectral optical data. We review the development of land cover approaches from the 1970s onwards and present the most commonly used methods. More specifically, we discuss their advantages and disadvantages, to localize our proposed frameworks in land cover classification community. We also display a state-of-the-art of fusion techniques at pixel, feature, and decision level, to explain the reason that we focus on the decision level instead of the other two levels. Limited attention has been paid on fusion at the decision level and most common fusion techniques can be separated into two categories: consensus maximization, and uncertainty theories. We briefly present all commonly used approaches based on consensus maximization and the basic concepts of belief functions in this chapter.

In chapter 3, we propose a new water detection method based on belief functions that can make decisions on total ignorance, so that potential water on the ground can also be detected. We also develop an automatic land cover classification method to generate semantic labels without using training samples. Related methods in this chapter show the efficiency and advantages of belief functions in land cover classification.

In chapter 4, we present a new transformation method that can change the discernment of mass functions based on similarity measures. The advantages of this architecture include the preservation of both uncertainty and imprecision and multiple choices of similarity measures. We also propose an iterative fusion strategy for multiple supervised and unsupervised methods, based on the transformation method. We compare the fusion system with this strategy with other state-of-the-art fusion methods at the output level. We evaluate the proposed framework on synthetic data at the output level to study how the quality of supervised and clustering results affect the combination.

In chapter 5, we evaluate the proposed framework on multiple real remote sensing datasets to discuss how limited training samples affect the combination results.

In chapter 6, we give the conclusions of this dissertation and some future research directions related to the problems addressed during this work.

Chapter 2

Foundations of land cover classification and information fusion

Resume

Is this thesis, our research focuses on the combination of multiple supervised and unsupervised methods in land cover classification, involving different research domains.

Therefore, in this chapter, we introduce the foundations of the domains related to our research, including multi-spectral land cover classification, information fusion and ensemble methods, and the fusion theory we used in our research.

The objective of this chapter is: (1) to understand the trend in land cover classification; (2) to present related researches in information fusion and ensemble methods; (3) to introduce the basic concepts of belief functions; (4) to locate our research topic in the community of land cover classification, ensemble methods and belief functions.

2.1 Introduction

In land cover classification, multiple developments haven been marked in the last four decades. Most classification methods were initially proposed in the 1970s and 1980s. However, many advancements in specific classifiers and algorithms have occurred in the last decade. Supervised and unsupervised pixel-based methods are one of the most important land cover classification methods in the 1970s. After 1980, the augmentation of resolution of remote sensing data makes other methods such as subpixel, contextual-based, and object-based, and hybrid methods became common in land cover classification. Most studies have reported the superior performance of object-based methods on different landscapes such as urban areas, agricultural areas, and forests. Nevertheless, object-based methods are sensitive to the optimal segmentation scale, which can result in over or under segmentation. The low spatial resolution of images also severely affects the performance of object-based methods. Nowadays, hybrid classifiers and multiple classifier systems have drawn more and more attention in land cover classification.

As multiple classifier systems show great potential in mapping land covers, many information fusion techniques and ensemble learning methods are applied to land cover classification. Therefore, we also give a brief introduction to information fusion techniques, including different structures and levels of fusion, and some popular ensemble methods. According to the needs in METIS, we locate our research interest in semi-supervised learning at the output level. Not too much effort has been paid on this topic according to our knowledge. We present some important previous works to combine supervised and unsupervised classification results at the output level. Most of the current works achieve the combination by maximizing the consensus of all individual classification, which usually requires the baseline with high accuracy. To solve the problems caused by limited training samples, we give a new perspective to solve the combination of supervised and unsupervised classification results in the framework of belief functions. Therefore, we also introduce the basic concepts in this chapter.

In this chapter, we give a brief introduction of multi-spectral land cover classification in section 2.2, information fusion and ensemble methods in section 2.3, and the basic concepts of belief functions in section 2.4. The conclusions are drawn in section 2.5.

2.2 Multi-spectral land cover classification

With the development of land cover classification methods, the accuracy of land cover maps has been greatly improved since the 1970s. It is obvious that classification methods become more and more complicated, and thus relies more on the quality and quantity of training samples. Popular approaches such as multiple classifier systems and deep learning techniques require sufficient and high-quality labeled data to train classifiers. However, collecting labeled data is always an arduous and hazardous task in land cover applications, especially for complex areas such as mountain areas. Therefore, decreasing the requirement of training samples becomes one of the most difficult challenges in land cover classification. Many efforts have been done to solve this problem, yet fewer achievements have recently been reached. Lots of methods are proposed for specific situations, thus

lacking the generality to some extent. Transfer learning is a novel technique to transfer a well-trained classification system into similar scenes where limited training samples are available. This immature technique is still under development and has many limitations in land cover classification.

Land cover classification based on multi-spectral optical data has gained more and more attention in remote sensing applications since the 1970s, shortly after Landsat 1 was launched. We display the development of different land cover classification methods in Figure 2.1. The early land cover classification methods are based on pixels, analyzing each pixel as an individual unit. At this period, the traditional machine learning approaches, such as supervised (i.e maximum likelihood, support vector machine) and unsupervised (*i.e.* k-means and ISODATA), were first applied and showed great performance in generating land cover maps [125].

These pixel-based approaches have limitations in heterogeneous regions where the size of objects may be much smaller the size of pixels. To address this problem, subpixel based methods, such as fuzzy classification and spectral mixture analysis techniques, have been developed during the 1980s and 1990s. Contextual information indicating the analysis of the relationship between a pixel and its neighborhoods was developed has gained attention in remote sensing society [81]. The early contextual-based methods includes texture extraction and Markov random fields models [63], [144]. It developed as the object-based methods in the late 1990s, especially after the launch of Very- High-Resolution (VHR) sensors, such as QuickBird and IKONOS. Object-based methods gather a group of homogeneous pixels as an object to achieve the classification. Therefore, the development of segmentation methods boosted the performance of object-based classification [159], [6].

The requirement of the accuracy of land cover classification has been greatly increased, so that much more complicated methods are developed. Hybrid methods and multiple classification systems gained the major attentions in real-life applications [22], [61]. Basic supervised classifiers such as maximum likelihood, support vector machine showed satisfying performance in the hybrid system. Ensemble learning methods and some fusion theories, such as belief functions, Bayesian inference, are also applied in combining multiple classifiers and have achieved success in different aspects. After the 2010s, more complicated methods have been proposed on object-level due to the more and more widespread use of VHR images. Many advanced classifiers have also shown their efficiency in real-life applications. Among them, deep learning techniques recently become a hot topic and outperforms other traditional methods in different situations [89].

From the view of land cover classification, our research focus on the multiple classifierssystem while also considering unsupervised methods. Unlike the traditional combination methods which require specific classifiers and rely greatly on training samples, our research focuses on a general fusion system that does not rely on specific classifiers and tries to solve the problem of limited training samples. We present some important techniques in land cover classification in the next sections, to localize our research interest in the community of land cover classification.

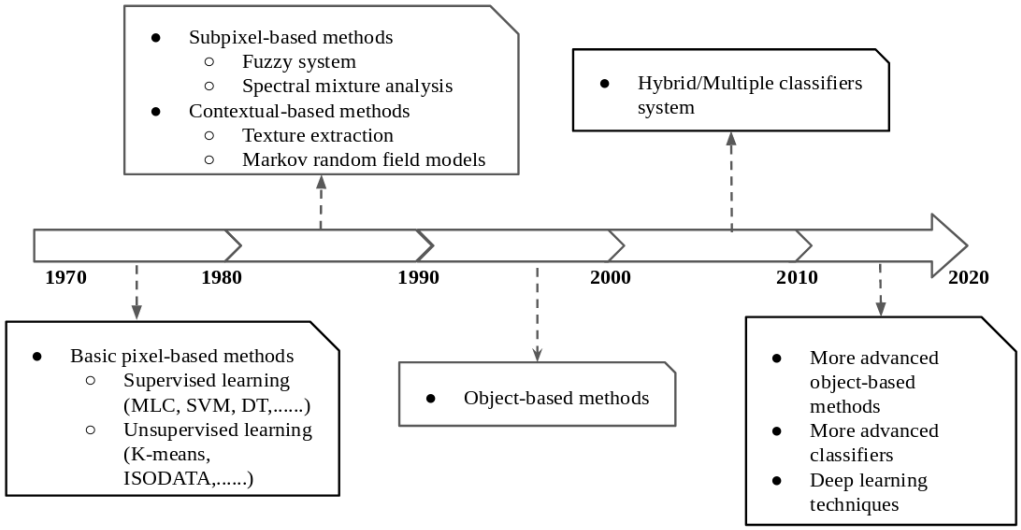


Figure 2.1. Important land cover classification methods in different periods.

2.2.1 Basic pixel-based methods

In the early 1970s, many pixel-based methods on land cover classification were used for medium/low resolution multi-spectral data, such as Landsat. Pixel-based methods assign each pixel to a class by treating each pixel as an individual unit. Pixels classified into the same class have more similar spectral properties than that in different classes. Since major approaches in land cover classification are pixel-based, to make a difference with other recent methods such as deep learning, we refer to this section as basic pixel-based methods. Basic pixel-based methods take advantage of traditional machine learning approaches which are categorized as supervised and unsupervised methods. This section focuses on the most commonly used supervised and unsupervised methods and investigates their advantages and limitations.

The most commonly used supervised classifier in land cover classification involves Maximum Likelihood (ML), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Network (ANN).

Most supervised classifiers are based on assumptions of data distribution, which is also called parametric classifiers, such as Maximum likelihood Classification (MLC). MLC is fast and simple to process and perform well in different scenes of land cover classification [87]. MLC also performs well with a limited number of training samples owing to the assumption of data distribution [54]. Using prior probabilities boosts the land cover classification accuracy. However, the improvement is not significantly large due to the nature of the remotely sensed data set and to the amount of spectral overlapping of class pairs [132]. It is found that MLC can cause over-fitting of the most dominated classes [132]. Even though parametric classifiers have proved to be useful, these classifiers have two major drawbacks in land cover classification: (1) data of high heterogeneous land

covers are usually not normally distributed; (2) a lot of uncertainty is associated with the distribution of land cover surfaces which cannot be described based on data distribution [87].

On the contrary, non-parametric classifier refers to no prior assumption about data distribution is assumed, such as SVM, ANN, and DT. SVM aims to find the optimal boundary in the feature space between the classes. It can provide a good generalization and well control over-fitting problems. Numerous studies [67], [93], [111], [92] have shown that SVM outperforms other classifiers in most cases due to its ability to handle complex features, even with small size training samples. However, SVM can be impacted by the quality of the training samples. For example, a study in [54] shows that the accuracy of SVM decreased by 8% when 20% of training samples are mislabelled. SVM is also dependent on user-defined parameters which is difficult to determine the optimal ones.

ANN is a non-linear classifier considered as a mathematical analog of the brain. ANN has multi-layers composed of numerous elements called neurons. All neurons in one layer are connected to all those in adjacent layers with some weights. ANN has advantages in handling noisy inputs while it depends greatly on user-defined parameters and architectures and suffers from over-fitting problem [54], [93]. DT is a recursive split of the input data depending on whether the value in a certain band is above or below a threshold [110]. Once DT has been trained, it is extremely rapid because no further complex mathematics is required. However, DT may generate a non-optimal solution and get over-fitting. The latter is normally addressed by pruning which reduces the accuracy of classification [110].

For supervised methods, the selection of training samples is critical because on the one hand the samples are supposed to be as representative as possible for all data set, and on the other hand, the sufficiency of samples has to be guaranteed to avoid the over-fitting. It has been reported that sufficient training samples are crucial for supervised classification [125]. For coarse resolution data, the selection of training samples is usually difficult because a pixel may contain more than one land cover type, which is referred to as the mixed pixel problem. The objective of generating training samples is to study statistically their patterns and features to classify each land cover class. According to Kavzoglu and Mather [74], large and accurate training samples are generally preferable. Studies have shown that increasing the training samples can improve classification accuracy [67].

Unfortunately, it is not always easy to collect adequate training samples with high quality due to limited access, time, or cost constraints. For remote sensing data, mislabeled training samples are another concern. Numerous studies [93],[54] find that MLC, SVM, DT, and ANN are sensitive to mislabeled training samples. Directly using traditional supervised methods in land cover classification is thus less effective when training samples are insufficient.

In the context where prior knowledge is unavailable, or in the strict sense that no training samples are available, unsupervised methods are employed. Unsupervised classification focus on the distribution and inter-relationship of data in feature space to classify them into various groups. Therefore, are especially suitable for the situation where labeled data is difficult to obtain, such as mapping land cover in a hard-to-access area. A study in [68] found that the unsupervised classification techniques appear less accurate in comparison to the supervised classification.

Clustering techniques are usually used in land cover classification when semantic la-

bels are unavailable. Among various unsupervised methods, k-means and ISODATA (Iterative self-organizing data analysis technique algorithm) is the most commonly used unsupervised classification method in remote sensing. k-means has to define the number of clusters at the beginning and classifies each pixel into clusters by minimizing the sum of the distance between each pixel to its assigned cluster center. The objective of k-means is to reduce the variability within the cluster. Unlike k-means, ISODATA requires thresholds to split and merge clusters instead of the number of clusters. Several iterations are performed to update clusters until the stop condition is reached. ISODATA eliminates clusters associated with fewer in comparison to the user-specific minimum number of pixels and isolated pixels are either reclassified or ignored as unclassified data. k-means and ISODATA algorithms have the best performance for images with clusters that are spherical and that have the same standard deviation, which is often not true for complex remote sensing images.

2.2.2 Subpixel-based methods

Compared to pixel-basic methods, subpixel-based classification was developed to address the mixed pixel problem that a pixel may contain several land cover classes. These landscapes may not be easily separated when classified by the ordinary pixel-based classification [19]. For low/medium resolution data such as Landsat, several land cover classes may constitute a single pixel. The most common methods of subpixel classification are fuzzy set techniques [53], [19] and Spectral Mixture Analysis (SMA) [87], [136].

Fuzzy set techniques were used in subpixel based methods since the 1980s. Fuzzy representation describes partial membership of each pixel belonging to all classes, thus the extent of each class within each pixel can be estimated [162], [51]. When using this method, each land cover is assigned a fuzzy membership depending on its proportion in each pixel. The proportions are in the form of ratios, percentages, or probabilities which are converted to actual areas on the ground. It is reported that the fuzzy classification method can reach high classification accuracy up to 93% when compared to the maximum likelihood pixel-based method with 61% [162]. subpixel analysis methods were developed to quantify the amount of urban impervious surfaces and urban vegetation and also show great performance [71], [100]. The fuzzy classification has been proved important for solving mixed pixels problems; however, it is not commonly applied in practical terms because it is not easy to use compared to other classification methods.

Spectral mixture analysis (SMA) is another effective method for dealing with subpixel methods, especially for medium resolution imagery such as Landsat [140]. It evaluates each pixel spectrum as a linear combination of a set of endmember spectra [5]. The output of SMA is endmembers represented by a fraction of each land cover type. One image for each endmember spectrum represents the area proportions of the endmembers within the pixel [32]. Endmember selection is one of the most important aspects of SMA, and much previous research has explored the approaches [121]. The common forms of SMA are Linear Spectral Mixture Analysis (LSMA) and Multiple Endmember Spectral Mixture Analysis (MESMA). LSMA is designed to work with a fixed number of endmembers while MESMA can be used on pixels with different numbers of endmembers [114]. Most studies have indicated that SMA is important in improving the area estimation of land cover

types [120]. SMA and subpixel based classification, in general, are important for effective classification of Landsat images as they are of medium resolution and are usually used for large areas that have heterogeneous land cover types and are likely to have mixed pixels. The major challenge for SMA is the errors in the final allocation of fractional endmembers resulting from spectral variability and similarity during the selection of endmembers [140].

2.2.3 Contextual-based and object-based methods

As the spatial resolution increasing, information derived from spatial and spectral relationships among pixels has gained attention to improve the accuracy of land cover classification. Contextual-based and object-methods was developed to deal with the problems of inter-class spectral variation, and make use of spatial information. According to [81], contextual-based methods are usually divided into three categories: (1) Texture extraction, (2) Markov random field models, and (3) object-based images analysis. However, since object-based methods have been greatly developed and many complex approaches have recently been proposed, it has already been regarded as a new independent category.

Texture extraction

Texture describes the placement and spatial arrangement of repetitions of tones and is often employed to quantify the variability of pixels in a neighborhood. In the early 1970s, texture extraction had been already used in remote sensing image classifications [63]. Many studies demonstrated the efficiency of texture-based features in land cover classification [65]. Texture extraction methods can be classified into three major categories: (1) structural (including mathematical morphology), (2) statistical, and (3) transformation. Structural textures can be defined with the primitives and their placement rules [63]. Morphology techniques based on non-linear operators have been applied to generate structural textures. Especially, morphological profiling and morphological attribute filters have been developed to capture geometrical and multi-scale properties. Statistical methods include first-order statistics (*i.e.* mean, standard deviation) and second-order statistics, especially the grey-level co-occurrence matrix (GLCM) proposed by Haralick [63]. Transform methods include Fourier, Wavelet transforms [94], and Gabor features [156], [70]. When compared to Fourier and Gabor, the wavelet transforms perform better as they are based on multiple spatial resolutions, and a wide range of wavelet functions can be chosen to improve the classification accuracy. Texture information can be incorporated in the processes of image pre-classification (*e.g.* as an additional variable) and post-classification (*e.g.* image filtering). Several studies have proven that the integration of textural information into remote sensing image classification can generate better classification accuracy [115].

Markov random fields models

Markov random fields (MRFs) conceptually generalize the notion of Markov chain, a popular model for one-dimensional random sequences, to the two-dimensional framework of image analysis. MRFs are a general family of probabilistic models for two-dimensional

stochastic processes defined over discrete pixel lattices. They represent flexible and powerful models for the contextual information associated with images. Their use in classification allows taking advantage of the dependence among neighboring pixels to maximize the accuracy in land-cover discrimination [99].

For remote sensing image classifications, MRFs incorporate spatial-contextual information into a classifier by modifying the discriminant function with the addition of spatial correlation terms. MRFs can examine the global and local properties of a remote sensing image, and quantify the spatial autocorrelation among pixels through a mathematically rigorous means [99]. Recently, an increasing number of studies have applied MRF-based image classification techniques and reported significantly better results when compared to the conventional non-contextual classification techniques. It is reported that the addition of contextual and edge information through the MRF-based methods improved both the visual interpretation and classification accuracy [144]. Although several MRF-based classification techniques have been successfully applied in land use land cover classifications, the concepts of MRF are considered difficult to many remote sensing scientists, and their implementations involve challenging computational problems [99]. The current development of this research area offers effective automatic parameter-optimization algorithms, which make Markovian classifiers feasible also for end users with no deep image processing knowledge.

Object-based methods

Compared to traditional per-pixel and subpixel classification methods, object-based models provide a new paradigm to classify remote sensing imagery. With object-based models, geographical objects are considered the basic unit for analysis. Instead of considering an image as a collection of individual pixels with spectral properties, object-based methods generate image objects through image segmentation and then conduct image classification on objects rather than pixels. With image segmentation techniques, image objects are formed using spectral, spatial, and textural and contextual information. Then these objects are further classified using spectral and other relevant criteria. Object-based approaches are considered more appropriate for VHR remote sensing images since they assume that multiple image pixels form a geographic object. Many studies have proven that significantly higher accuracy has been achieved with object-based approaches [159], [6]. Since major researches on object-based methods are focused on supervised context [88], we thus briefly introduce the supervised object-based methods.

Object-based methods include two major steps: segmentation and classification. Segmentation is used to generate an object as the basic unit to analyze. Numerous studies demonstrate the importance of the scale parameter because it controls the dimension and size of segmented objects, which may directly affect classification [101]. Generally, the smaller segmentation scales are more suitable for higher spatial resolution. However, it can be difficult to determine the optimization scale given the fact that the variability of the scale is affected by many factors. The scale issue, therefore, has emerged as a major problem in object-based methods. Successful research about scale optimization is to combine Local Variance (LV) and Rates of Change of LV (ROC-LV) to determine appropriate segmentation scales [41]. However, this method is only capable of processing

single-band images. Another optimized segmentation object using the Euclidean Distance measure for the reference object and the actual object is proposed in [151]. This method requires manual interpretation of reference objects for different land-covers. Unsupervised segmentation methods commonly suggest a single optimization scale or require a difficult threshold, which therefore is less used than supervised segmentation methods [47].

Even though the remarkable performance of object-based methods has been reported for high spatial resolution images. There are still some limitations to object-based methods in land cover classification. For instance, extracting objects of land cover requires high levels of spatial detail, which are limited in low spatial resolution images. Moreover, selecting the optimal segmentation scale is a challenge. When the segmentation scale is not appropriate, the image can be under or over segmented. Under segmentation means that the image-objects are larger than the objects on the ground, which thus makes two or more land covers in one large image object. Likewise, over segmenting an image can result in a real-world object being split into two or more objects with different classes. The selection of classifiers is another important issue in object-based methods. It is reported that Random Forest (RF) achieves the best performance in object-based classification, and has attracted significant attention in recent years, followed by SVM, while MLC performing the worst [47]. Although RF and SVM classifiers have also attracted great attention owing to their excellent classification performance, deep learning, which is an excellent classification technique developed in recent years, is still expected to further promote the development of supervised object-based classification techniques [47].

2.2.4 Deep learning techniques

With the development of deep learning (DL) techniques, land cover classification studies are paying increasing attention to feature learning. A comparison of numerous studies has revealed that deep features have a more powerful capability than low-level features in representing and understanding images [164], [82], [10]. Several methods have been proposed for land cover classification with deep features, such as greedy layer-wise unsupervised pre-training [122] and multi-scale CNN algorithm [89], and all of them have shown a proved improvement.

Convolutional Neural Networks (CNN) model has been identified as the most efficient solution in land cover classification [89] although some efforts were initially made on other DL models, such as the Deep Belief Networks (DBNs) [28]. Based on CNN solution, specific improvements have been proposed to obtain a more accurate land cover mapping. A generative adversarial network (GAN) was first introduced as a regularization technique on CNN [166]. Results show better performance in preventing over-fitting [166]. Besides, they proposed 3D-GAN for spectral and spatial characteristics, leading to a significant gain in land cover classification accuracy compared to conventional CNN. With the same intuition, CNN architecture combined with a 3D-DL strategy as a feature processing technique was studied in [15]. In this solution, classical 1D convolution operators are replaced by 3D convolution operations to allow joint processing of spectral and spatial information, and a significant performance gain has also been proven by experiments. Therefore, the 3D-DL strategy seems to be a generally accepted method for processing spectral-spatial features in the CNN model.

Regarding the training of a supervised DL model, the preparation of the required large number of training samples is always an inevitable and time-consuming problem. For this reason, the techniques of the Hybrid/Multiple classifiers system (MCS) have been developed to improve the accuracy of land cover classification to solve the problem of heavy training data. The reduction of training samples tends to be the new topic of land cover classification and it is also one of the objectives of our research.

2.2.5 Hybrid/Multiple classifiers system

Even though deep learning techniques outperform other methods with high accuracy of classification, their performance is severely limited by insufficient labeled samples in real problems. Multiple classifier systems (MCS) are other newly emerged methods that combine multiple classification results from several different classifiers. The purpose of MCS is to achieve a better classification result than that acquired by using only one classifier. The core of MCS is to combine the results provided by different base classifiers, and the earliest method for the combination was through the majority voting. By now, some more approaches have been proposed for classifier combination, such as Bayes approach, Dempster-Shafer theory. Previous studies have shown that MCS are effective for land cover classification [84], [24]. For example, a MCS with six base classifiers was constructed in [84], and the classifier combination was through the voting strategy. Their results show that MCS obtained higher accuracy than those achieved by its base classifiers. For improved performance, base classifiers to be combined in a MCS should be selected from diverse families of pattern recognition [24], and the diversity is measured by the difference among the base classifier's pattern recognition algorithms [77].

The selection of appropriate supervised classifiers, such as SVM, RF, is also important for MCS methods. SVM, RF, ANN and boosted DTs have proven to be very powerful methods for land cover classification and, in general, these methods appear to produce higher accuracy than other classifiers such as DT. However, the best classifier depends on many factors, *e.g.*, classes over the study area, nature of the training samples. There is currently no one classifier that outperforms others in all circumstances. Each classifier has its advantages and limitations. For example, the SVM is powerful for handling complex characteristics while its parameters are generally difficult to determine. Besides, SVM works well with small training samples but is sensitive to mislabelled data. The ANN has high accuracy in general but suffers from the problem of over-fitting. RF is slightly less efficient than SVM in pixel-based methods, but it performs better than SVM in object-based methods. Multiple classifiers are generally more robust than a single classifier.

It has generally been shown that the quality and quantity of training data have a significant impact on the accuracy of classification. Training data may even have a greater impact in comparison to the algorithm used [67]. Therefore, it is preferable to obtain a large number of high-quality training samples that fully characterize the class signatures. Figure 2.2 shows the relationship between training samples and precision for different land cover classification methods. Unsupervised methods do not require training samples and are generally less accurate than supervised methods. For supervised methods, including pixel, subpixel, and object-based approaches, adequate training samples are always preferable. Many supervised classifiers such as SVM, ANN are sensitive to the quality of the

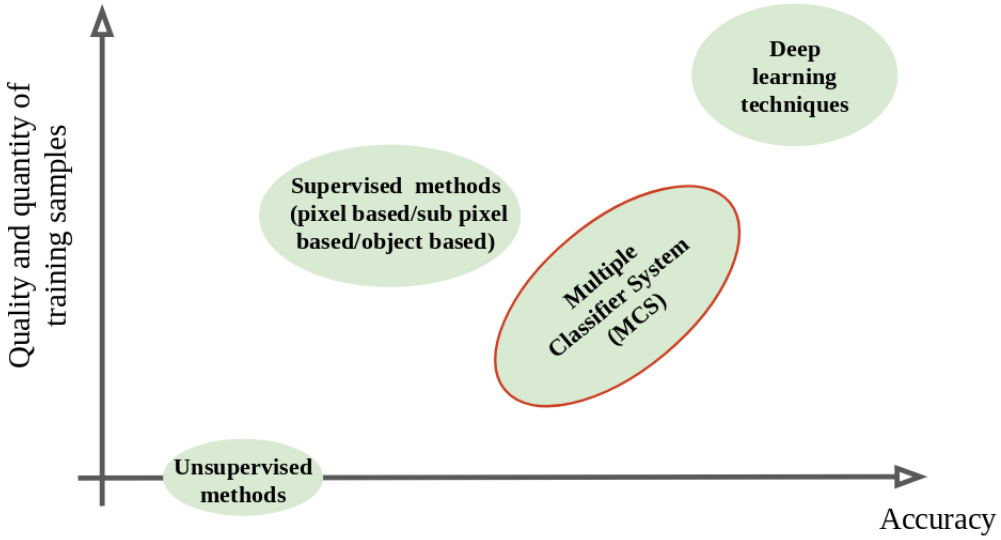


Figure 2.2. Relationship of training samples and accuracy for different categories of land cover classification methods.

training samples. If abundant, high-quality training samples are available, deep learning techniques may be considered the best choice. Their accuracy is higher in comparison to the others and thus becomes a new trend in the land cover classification community.

The combination of multiple classifiers outperforms other methods and should be considered the best choice with limited training samples. Therefore, we focus on fusion and ensemble methods in the next section.

2.3 Information fusion and ensemble methods

Multiple techniques from information fusion and ensemble methods effectively combine different classifiers in MCS methods and achieve high accuracy in land cover classification. In this section, we present related researches in the fields of information fusion and ensemble methods, which are furthered studied in land cover classification and help to design MCS methods.

2.3.1 Structures of fusion

Different classifiers, such as parametric classifiers (*e.g.* maximum likelihood) and non-parametric classifiers (*e.g.* neural network, decision tree), have their strengths and limitations. Multiple Classification System (MCS) has shown great potential for improving the accuracy and reliability of remote sensing image classification. A critical step is to develop appropriate rules for combining classification results from different classifiers. MCS

can be roughly classified in both sequential and parallel structures. The former uses the output of one classifier as the input of another, while the latter applies several classifiers individually and processes their output together. In this section, we provide only a brief introduction to the multiple classifiers used in land use classification. Details on the ensemble and fusion methods are explained in the next section.

Sequential structure

In a sequential structure, the classification result generated by one classifier is used as input to the next classifier, as shown in Figure 2.3. The results obtained by each classifier are also passed on to the next classifier until a result is obtained by the last classifier in the chain. Boosting is the sequential structure most commonly used in land cover classification [22], [61]. Boosting can process data with weights, and the weights of misclassified samples are increased to focus the learning algorithm on specific samples. As a result, boosting can reduce both variance and classification bias, but it is very sensitive to noisy data.

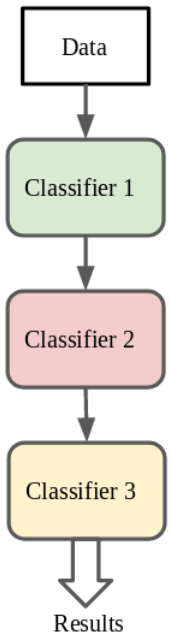


Figure 2.3. Sequential structure.

Boosted Decision Tree (Boosted DTs) is an ensemble method using multiple DTs to overcome the drawbacks of a single DT. Boosted DTs attempt to minimize the errors of the previous trees adaptively. A commonly used boosted DTs is Adaptive boosting (Adaboost). Each tree of Adaboost is a weak learner with poor predictions. An additive model allows combining all weak learners so that the loss function is minimized. Adaboost improves the accuracy of each weak learner by $6\% - 8\%$ and also shows high robustness and low risk of over-fitting [27].

Extreme gradient boosting (Xgboost) is another popular boosting method proposed in [26]. Xgboost was conducted on the data of WorldView-3 using Bayesian parameter

optimization [59]. This study found that Xgboost outperforms RF and SVM by 2–5% in larger sample sizes albeit with increased computational time. Another study [90] compared Xgboost with five non-parametric classifiers on Landsat-8 data, and it showed that Xgboost slightly outperformed SVM by 0.3%.

Parallel structure

In a parallel structure, multiple classifiers are designed independently without any mutual interaction and their results are combined according to certain strategies, as shown in Figure 2.4. One of the most successful parallel ensemble classifiers is Random Forest (RF), which uses a large number of DTs to overcome the weakness of a single DT. The majority vote of all DTs is generally used to assign a final class. This directly overcomes the problem that a single DT may not be optimal. By incorporating several DTs, an overall optimum should be obtained. RF has some advantages such as reduced training time and easy parameterization [113]. Besides, RF parameters seem to have little influence on the accuracy of the classification [113].

Other popular and successful ensemble approaches are majority voting, fuzzy integral, Dempster-Shafer's theory, etc. Some previous studies have shown that simple majority voting of the prediction is an effective strategy [131]. In this scheme, a pixel is classified as the class that has been chosen by the majority of individual classifiers. The simple majority voting rule is used to integrate multiple binary classifiers for mapping a specific class [52]. Some studies have examined the combination of soft classification methods for remotely sensed images and found that the combination of classification methods can improve the accuracy [40].

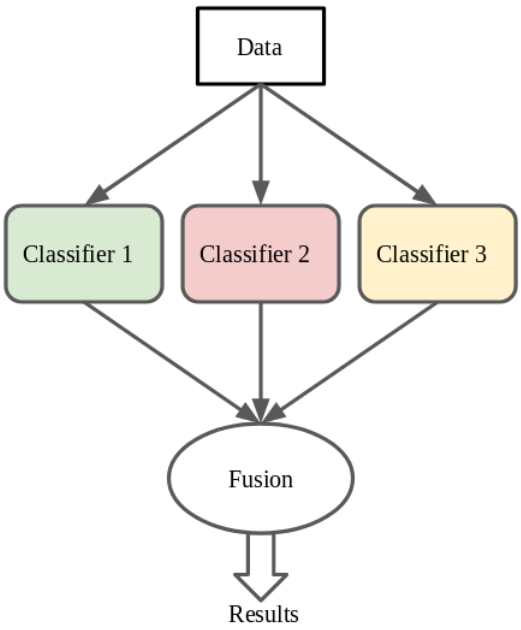


Figure 2.4. Parallel structure.

2.3.2 Levels of fusion

Levels of fusion is another important aspect in MCS methods for land cover classification, because each individual level requires different processing steps and has different advantages and limitations [91]. The levels are distinguished by where the actual fusion step takes place, as shown in Figure 2.5.

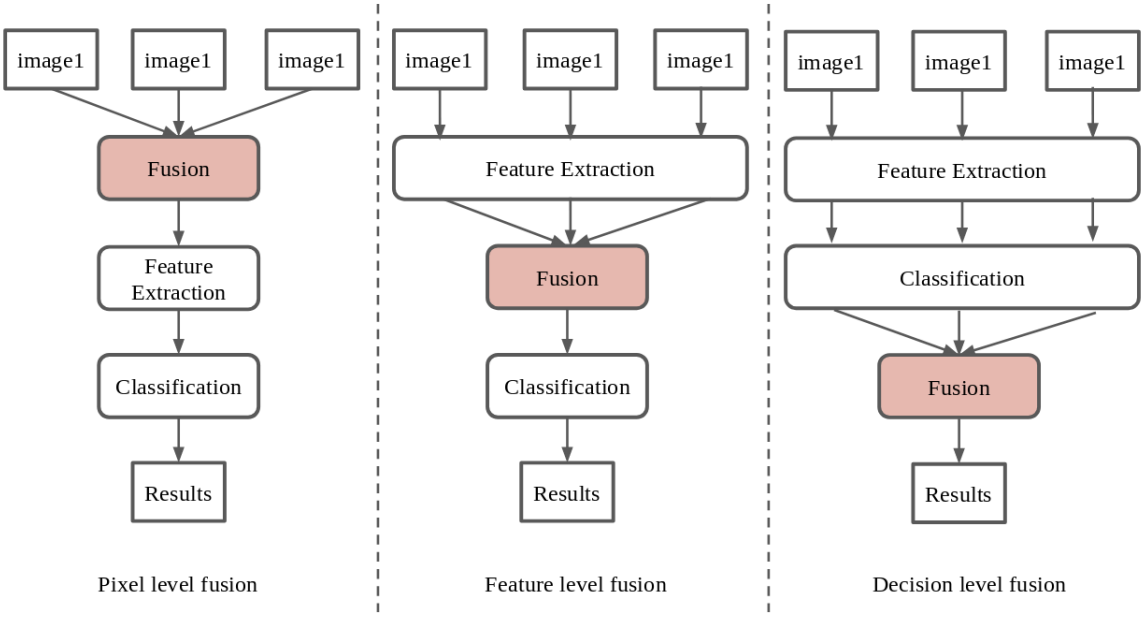


Figure 2.5. General processing steps for different fusion levels.

Pixel level fusion

Pixel level fusion is designed to synthesize multiple input images in pixel levels to generate a new fused image, which is expected to be more informative than any of the inputs. Pixel level fusion is the fusion at the lowest processing level referring to the merging of measured physical parameters. Image fusion taking place at pixel level requires sensor specifically corrected data and alignment. It is critical to remove influences that affect and falsify the values collected by a certain sensor. Besides, the multi-sensor images have to be coregistered and geometrically corrected to ensure that the data coincide on a pixel-by-pixel basis and refers to the same location on the ground. For the integration in a geographic information system (GIS) or the Digital Earth environment, geocoding is indispensable. The actual information extraction takes place after the fusion process. An important fact to be considered in pixel-based fusion is the effect of changes between the different acquisition dates of the input images [139]. A combination of changed objects into one fused image can cause spectral artifacts if these changes are not considered or if image fusion techniques are used in an uncontrolled manner.

Remote sensing image fusion techniques have different sensitivities to misregistration depending on the chosen approach. In remote sensing applications, pixel-based fusion is

essentially used for pansharpening which aims at enhancing spatial resolution through the synthesis of a low resolution multi-spectral (MS) image and a high resolution panchromatic (PAN) image [95]. Different arithmetic combinations have been developed for pixel image fusion. The principal component analysis (PCA) transform [83] converts inter-correlated multi-spectral bands into a new set of uncorrelated components.

Pixel fusion algorithms mentioned above have been widely used for relatively simple and time-efficient fusion schemes [139]. However, several problems must be considered before their application: 1) These fusion algorithms generate a fused image from a set of pixels in the various sources. These pixel-level fusion methods are very sensitive to registration accuracy so that co-registration of input images at subpixel level is required; 2) These image fusion methods are often successful at improving the spatial resolution. However, they tend to distort the original spectral signatures to some extent [14]. More new techniques such as the wavelet transform [153], [60] seem to reduce the color distortion problem and to keep the statistical parameters invariable.

Feature level fusion

Complex object structures require more complex processing. Land cover mapping studies often rely on remote sensing feature-based fusion for classification to obtain a better and more complete description of the objects on the ground. In particular, complex urban areas require additional inputs into the fusion process because the variability of urban objects is based on their inherent spectral heterogeneity [105]. Prior feature extraction of the urban extent can help the mapping, especially if multi-resolution data sets are involved [56]. Feature fusion implies that the images are processed to extract features of interest using feature extraction methods. Feature-level fusion uses a group of image pixels to form contiguous regions and requires the extraction of different features from source data. According to [62], "feature extraction addresses the problem of finding the most compact and informative set of features to improve the efficiency of data storage and processing."

For the objective of land cover classification, features can be spectral signatures, texture features, etc. Compared to pixel-level fusion, feature level fusion is more suitable for a complex environment that requires higher processing levels. Especially in an urban context where high-resolution data are a prerequisite and land cover is complex in nature, feature-level fusion is of interest.

Most existing techniques defined shallow handcrafted features or transform based filters of the input data that are not robust enough to make a deal with classification challenges of remote sensing data [62]. Additionally, traditional fusion methods use either low-level features or score-based fusion to fuse features. Deep learning frameworks have recently enhanced the classification performance by automatic extraction of extremely powerful deep features. In [146], the authors provide a deep supervision strategy to enhance the generalization performance in the intermediate layers of the AlexNet model. In [145], the proposed approach merges the deep learning architecture and classical handcrafted feature extraction approaches. The authors fuse the rich feature extraction capabilities of the Convolutional Neural Networks (CNNs) with the Correlograms features. Despite the significant performance of deep learning feature fusion, generating powerful

fused features usually requires a large number of training samples. Its application thus is limited for inadequate labeled data.

Decision level fusion

Decision fusion is defined as "the highest level of information fusion, which reinforces the common interpretation, resolves differences" while providing "a better understanding of the observed objects" [42]. Decision-level fusion approaches include Bayesian theory, belief functions, abductive reasoning, majority voting *etc.*

Bayesian method combines evidence related to the probability theory rules. It can provide a determination of the probability of a hypothesis being true with given evidence. Furthermore, it not only allows the incorporation of prior knowledge about the likelihood but also utilization of subjective probabilities for a prior probability. Theory of belief functions also known as Dempster-Shafer theory allows representation of incomplete knowledge with uncertainty and imprecision. Based on the theory, authors in [79] improved the land cover accuracy using multi-source remote sensing data. The abductive reasoning approach [11] searches for the best probability of the information to be true following a reasoning pattern and can involve neural networks or fuzzy logic. Semantic methods utilize semantic data of different origin to derive the interpretation. It requires a known set of knowledge, which can be used to match the information.

Compared to pixel-level fusion and feature level fusion, fusion at decision level is more robust yet more complicated [42]. It thus gains less attention in comparison to the other two fusion levels. In our research, we have to consider combining the results generated by different supervised and unsupervised methods. Decision level fusion is more suitable for our situation because we are only interested in the outputs of classifiers rather than the input features or classifiers themselves.

2.3.3 Overview of ensemble methods

Supervised and unsupervised methods have been proved to be outstanding in different contexts: supervised methods show powerful performance when sufficient labeled data are available, while unsupervised methods also called clustering focus on grouping unlabeled data based on specific properties. Both of them have their advantages and drawbacks which provide complementary or conflicting information about the raw data (*e.g.* multi-spectral data). Therefore, a combination of supervised and unsupervised methods have been gradually gained more attention.

According to the research in [57], [58], the entire spectral of learning methods is based on two dimensions: one dimension is the level of supervision (supervised, semi-supervised, unsupervised), and the other is the way the ensemble executes (no ensemble, ensemble at the raw data level and the ensemble at the output level), as shown in Figure 2.6 [57].

The ensemble at raw data level can also be considered as an ensemble by learning. In the supervised context, the most common methods are boosting, bagging, Bayesian model averaging, in which raw data are applied to train a combination process [124], [16]. Weighted majority voting [33] can also be classified into this level because weights are generally trained from labeled data by maximizing some pre-defined objective function.

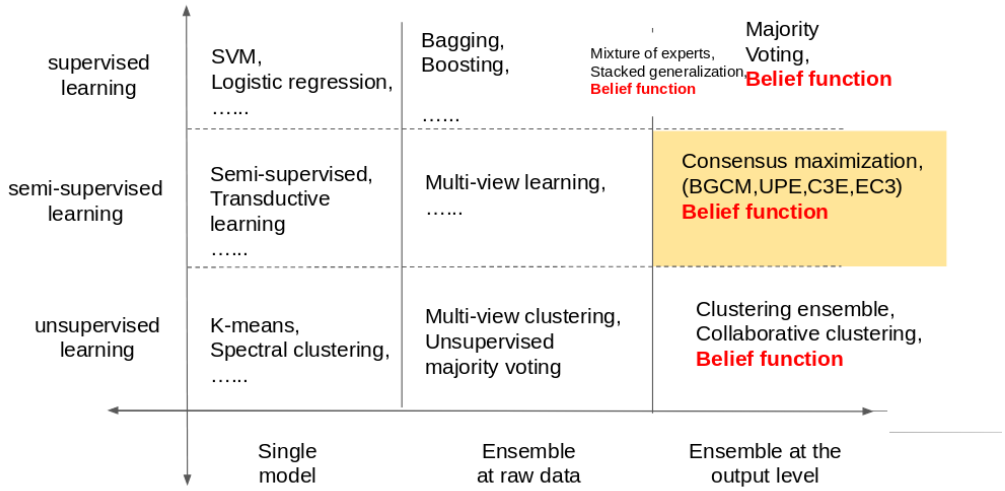


Figure 2.6. Classification of combination strategies proposed in [57].

The theory of belief functions is usually used for multi-source fusion at the output level. However, some recent researches add a training process to obtain optimal weight for each classifier. Authors in [86] proposed a new weighted classifier combination method based on belief functions to enhance classification accuracy. The optimal weighting factors of classifiers are obtained by minimizing the distances between fusion results obtained by Dempster's rule and the target output in training data space to fully take advantage of the complementary information of the classifiers.

For ensemble in the semi-supervised learning context, multi-view learning can learn from both labeled and unlabeled data from multiple sources [21], [163]. It aims to learn one function to model each view and jointly optimizes all the functions to improve the performance. A naive solution for multi-view learning is to concatenate all multiple views into one single view by training process, which yet has the over-fitting problem on small training sets. Recently, a more elaborate solution has been proposed to overcome its drawbacks such as co-EM [106], which are trained alternatively on two distinct views with confident labels for the unlabeled data. Some specific scenarios to achieve the combination are also considered on this level. For instance, in [31], supervised methods are used to extract the dissimilarity from raw data to guide an unsupervised clustering instead of the traditional distance function. A combination of supervised and unsupervised has also been applied to train the activation functions of neural networks in [23]. However, the majority of semi-supervised scenarios usually consider only one supervised method and one unsupervised method [25], which is arduous to combine several supervised and unsupervised sources. Although multi-view learning can combine multiple classifiers, it is easily suffered from the over-fitting problem, which is therefore not appropriate for small training sets [163].

At the output level, major research has been devoted to ensemble learning for either supervise or unsupervised methods. In supervised methods, majority voting [98] can achieve the combination on the output level. A noteworthy point of majority voting based combination is that dependent classifiers can potentially offer a dramatic improvement over independent classifiers [76]. Except for majority, belief functions can also be applied in combining multiple classifiers at this level [17], [85]. Compared to majority voting, methods based on belief functions cope with a more complex problem because several combinations and decision rules are proposed under this framework to manage different situations [128]. The details about belief functions are displayed in section 2.4. Other methods such as stacked generalization [64], [152], and the mixture of experts [3] can be considered between raw data ensemble and output ensemble because they generate a meta-learner on the output level and need the labels of raw data as feedback as well.

In the unsupervised context, the combination is more difficult when the numbers of clusters from the various single model are different. Therefore, the final combined clustering results are generally required to obtain the most agreement with all individual clustering methods, modeled by an objective function using consensus maximization [55]. Multi-view clustering comes from a statistical background of several methods analyzing different views of the same data with different attributes [18]. Collaborative clustering [30], [141], is another close family to multi-view clustering, which allows different types of algorithms to work together. In [150], the authors proposed a multi-view voting strategy to combine multiple clustering. Each result votes for the class it had found for each object and its corresponding class in the other results. The theory of belief functions was also used in combination of unsupervised learning in an early research [78], which takes the intersection from two clustering methods as the final discernment. Recently, a novel method combining evidential clustering methods based on belief functions is proposed [80]. This method combines relational representations of base credal partitions by averaging and minimizes an error function to obtain the consensus solution and it show the good performances on simulated and real datasets.

We make a summary of the advantages and limitations of different methods both from the level of supervision and the level of combination, as shown in Table 2.1. According to our research on state of the art, several latest developments, such as combination methods based on belief functions have not yet drawn much attention in the domain of ensemble methods. Compared to other methods, the theory of belief functions has excellent potential infusing multiple sources, no matter in which level of supervision. Furthermore, majority voting has also been underestimated in the previous state of the art. This technique is not merely suitable to combine supervised learning at the output level but raw data level as well. The theory of belief functions has a more flexible and widely used compared to other methods. It also measures uncertainty and imprecision, which thus becomes the leading theory in our fusion framework.

We consider combining multiple supervised methods which are categorized as semi-supervised methods, marked by yellow rectangular in Figure 2.6. Ensemble methods at raw data, also known as an ensemble by learning, require adequate labeled data to train the learning process, which is not appropriate in our situation. Since we concentrate on the case where training samples are limited or inadequate, it is reasonable to focus on the semi-supervised ensemble method combining several supervised and unsupervised

methods at the output level, which only applies the outputs from every single model without accessing raw data to achieve the training process.

According to our research on the state of the art, there are minimal attempts to combine multiple supervised methods and unsupervised methods at the output level. In section 2.3.4, we present the essential combination methods at the decision level to combine several supervised and unsupervised models.

Table 2.1 – A summary of combination methods.

Supervision level	Combination level	Methods	Advantages	Limitations
Supervised learning	Single model	SVM, KNN, ...	Simple model	1. Can not combine with other methods 2. Over-fitting 3. Require enough labeled data
	Ensemble at raw data	Boosting, Bagging, ...	Combine with multiple methods	1. Require enough labeled data 2. Over-fitting
	Ensemble at the output level	Majority voting, Belief functions	Not require the amount of labeled data	1. Combined results depend a lot on basic classifiers 2. Complicated
Semi-supervised learning	Single model	Semi-supervised learning	Simple model	1. Can not combine with multiple methods
	Ensemble at raw data	Multi-view learning	Combine with classifiers	1. Over-fitting 2. Require amount of labeled data
	Ensemble at the output level	Consensus maximization, Belief functions	Combine classifiers &clusterings	1. Models are complicated 2. Not too much related research
Unsupervised learning	Single model	Kmeans, ...	Used for unlabeled data	1. Can not combine with multiple methods
	Ensemble at raw data	Multi-view clustering, Majority voting	Can combine with multiple clusterings	1. Final results do not have labeled information
	Ensemble at the output level	Collaborative clustering, Belief functions	Can combine with multiple clusterings	1.Final results do not have labeled information

2.3.4 Semi-supervised ensemble methods at the output level

Most of the existing semi-supervised ensemble methods at the output level are based on consensus maximization, transferring the fusion process as an optimization problem. The solution of this optimization problem is used as the final fused result which takes the most agreements among all supervised and unsupervised models.

Suppose we have N objects $X = \{x_1, ..., x_N\}$ to classify into Y classes $\Omega = \{\omega_1, ..., \omega_Y\}$. We have M supervised models $S = \{s_1, s_2, ..., s_M\}$ and a L unsupervised models $C = \{c_1, c_2, ..., c_L\}$ with different numbers of clusters. Both classes and clusters discovered by each supervised and unsupervised model are called groups. The set of groups noted as $G = \{g_1, g_2, ..., g_G\}$, contains all classes and clusters from different methods. For consensus maximization based methods, there are some important concepts to construct the optimization problem, presented as follows:

Definition 1 (*Membership Matrix). Membership matrix is defined as* $\mathscr{A}^m|_{N\times G}$ *where* $\mathscr{A}^m|_{ij} = 1$ *indicates the object* x_i *belongs to group* g_j, 0 *otherwise.*

Definition 2 (*Co-occurrence Matrix). Co-occurrence matrix is defined as* $\mathscr{A}^c|_{N\times N}$ *where* $\mathscr{A}^m|_{ij} \in \mathbb{R}$ *indicates the number of times two objects* x_i *and* x_j *occur together in the same group.*

Definition 3 (*Object-class Matrix). Object-class matrix is defined as* $\mathscr{F}^o|_{N\times Y}$ *where* $\mathscr{F}^o|_{ij}$ *indicates the probability of an object* x_i *belonging to class* ω_j.

Definition 4 (*Group-class Matrix). Group-class matrix is defined as* $\mathscr{F}^g|_{G\times Y}$ *where* $\mathscr{F}^g|_{ij} \in \mathbb{R}$ *indicates the probability of a group* g_i *being labeled as class* ω_j.

Definition 5 (*Average-class-distribution Matrix). Average-class-distribution matrix is defined as* $\mathscr{Y}^o|_{N\times Y}$ *where* $\mathscr{Y}^o|_{ij}$ *indicates the fraction of times object* x_i *is labeled as class* ω_j *by the* M *base classifiers.*

Definition 6 (*Average-group-distribution Matrix). Average-group-distribution matrix is defined as* $\mathscr{Y}^g|_{G\times Y}$ *where* $\mathscr{Y}^g|_{ij}$ *indicates the fraction of times group* g_i *is labeled as class* ω_j *by the* M *base classifiers.*

Based on these matrices, different optimization problems have been constructed to maximize the consensus among all supervised and unsupervised models. We present the existing methods based on consensus maximization in the following.

Graph-based Consensus Maximization (BGCM) Algorithm

One of the early remarkable researches is Bipartite Graph-based Consensus Maximization (BGCM) Algorithm [57]. It considers the results of clustering as constraints and maximizes the consensus among both supervised predictions and unsupervised constraints. The data, models, and outputs are summarized by a bipartite graph, in which the on one side nodes denote the groups' output by the models, and the other side nodes denote objects. A group node and an object node are connected if the object is assigned to the group, no matter if it comes from a supervised or unsupervised model. From the resulting graph, the goal is to predict the class labels so that they agree with the supervised models and also satisfy the constraints enforced by the clustering models, as much as possible. In other words, the authors in [57] aim at consolidating a classification solution by maximizing the consensus among both supervised predictions and unsupervised constraints, casting it as an optimization problem on a bipartite graph.

The consensus agreement optimization problem is constructed as:

$$\min_{\mathscr{F}^g,\mathscr{F}^o}\{\sum_{i=1}^{N}\sum_{j=1}^{G}\mathscr{A}_{ij}^{m}||\mathscr{F}_{i.}^{o}-\mathscr{F}_{j.}^{g}||^2+\varrho\sum_{j=1}^{G}k_j^g||\mathscr{F}_{j.}^{g}-\mathscr{Y}_{j.}^{g}||^2\} \tag{2.1}$$

$$s.t.\ \mathscr{F}_{i.}^{o}\geqslant 0,|\mathscr{F}_{i.}^{o}|=1,i=1,...,N \tag{2.2}$$

$$\mathscr{F}_{j.}^{g}\geqslant 0,|\mathscr{F}_{j.}^{g}|=1,j=1,...,G \tag{2.3}$$

where $||.||$ and $|.|$ denote a vector's $L2$ and $L1$ norm. ϱ is the shadow price payment for violating the constraints. $k_j^g\in[0,1]$ is the weight of the constraints. The first term ensures that if an object x_i is assigned to group g_j by one of the models, their conditional probability estimates for the category label must be close. The second term imposes the constraint that group g_j's consensus class label estimate should not deviate much from its initial class label prediction.

C3E model

C3E [4] is another early ensemble model that combines heterogeneous base methods. It uses multiple classifiers to generate an initial class-level probability distribution for each object. The distribution is then refined using a clustering ensemble. The authors think that the clustering ensemble provides supplementary constraints for classifying the objects of X, with the rationale that similar objects are more likely to share the same class label.

The consensus agreement optimization problem is constructed as:

$$\min_{\mathscr{F}_{i.}^{o}}(\sum_{i=1}^{N}\mathscr{B}(\mathscr{Y}_{i.}^{o},\mathscr{F}_{i.}^{o})+\varrho\sum_{i=1}^{N}\sum_{j=1}^{N}\mathscr{A}_{ij}^{cn}\mathscr{B}(\mathscr{F}_{i.}^{o},\mathscr{F}_{j.}^{o})) \tag{2.4}$$

$$s.t.\ \mathscr{F}_{i.}^{o}\geqslant 0,|\mathscr{F}_{i.}^{o}|=1,i=1,...,N \tag{2.5}$$

$$\mathscr{Y}_{j.}^{o}\geqslant 0,|\mathscr{Y}_{j.}^{o}|=1,j=1,...,G \tag{2.6}$$

where $\mathscr{B}$ can be any Bregman divergence, used as the loss function. ϱ is the shadow price payment for violating the constraints. $\mathscr{A}^{cn}$ with $\mathscr{A}_{ij}^{cn}\in[0,1]$ is the normalized version of $\mathscr{A}^{c}$.

Unconstrained Probabilistic Embedding (UPE) Model

The UPE model [7] casts the combination problem as an unconstrained probabilistic embedding problem. It assumes that both objects and groups have latent coordinates without constraints in a $\mathscr{D}$-dimensional Euclidean space. They also consider the mapping from the embedded space into the space of results from supervised and unsupervised models as a probabilistic generative process. The prediction of an object is then determined by the distance between the object and the classes in the embedded space.

Suppose we have a set of objects $X=\{x_1,...,x_N\}$ and the objects have latent coordinates $X^*=\{x_i^*\}_{i=1}^n$, where x_i^* is a coordinate of the i-th object in the embedded space with dimension $\mathscr{D}$. For each group g_j, it also has latent coordinate g_j^* with dimension $\mathscr{D}$ in the embedded space. The set of all the group coordinates is G^*.

The unknown parameters are a set of object coordinates X^* and a set of group coordinates G^*, which can be estimated by maximizing the posterior $P(X^*, G^*|A^c)$. It is equivalent to the following optimization problem:

$$\min_{X^*,G^*} \sum_{i=1}^{N}\sum_{j=1}^{G} A_{ij}^c logP(g_j|x_i^*, G^*) + \sum_{i=1}^{N} logP(x_i^*) + \sum_{i=1}^{G} logP(g_j^*) \tag{2.7}$$

After obtaining a local optimum solution for $\{X^*, G^*\}$, the label of object x_i can be predicted by:

$$\min_{1\le j\le Y} P(g_j|x_i^*, G^*) \tag{2.8}$$

EC3 model

EC3 [25] focuses on the meta-output level without accessing the raw data. It maps the combination into a convex optimization problem and the objective function is based on a consensus at the object level as well as at the group level, which differs from BGCM and UPE. The proposed objective function ensures that:

1. the group characteristics is similar to the characteristics of its constituent objects, note as $J1$:

$$J1 = \sum_{i=1}^{N}\sum_{j=1}^{G} k_{ij}^m ||\mathscr{F}_{i.}^o - \mathscr{F}_{j.}^g||^2 \tag{2.9}$$

   where k^m is a bi-stochastic matrices corresponding to A^m.

2. the more two objects are part of same base groups, the higher the probability that they are assigned to the same class, which is occurrence principle, noted as $J2$:

$$J2 = \sum_{i=1}^{N}\sum_{j=1}^{N} k_{ij}^c ||\mathscr{F}_{i.}^o - \mathscr{F}_{j.}^o||^2 \tag{2.10}$$

   where k^c is a bi-stochastic matrices corresponding to the co-occurrence matrix A^c.

3. class distribution of an object is similar to its average class distribution obtained from multiple base classifiers,which is called consensus principle, noted as $J3$:

$$J3 = \sum_{i=1}^{N} ||\mathscr{F}_{i.}^o - \mathscr{Y}_{i.}^o||^2 \tag{2.11}$$

4. class distribution of a group is similar to the average class distribution of its constituent objects, noted as $J4$:

$$J4 = \sum_{j=1}^{G} ||\mathscr{F}_{j.}^g - \mathscr{Y}_{j.}^g||^2 \tag{2.12}$$

The optimization problem combines these four hypotheses together as follows:

$$\min_{\mathscr{F}^o,\mathscr{F}^g} \frac{\delta_1}{2}J1 + \frac{\delta_2}{2}J2 + \delta_3 J3 + \delta_4 J4 \tag{2.13}$$

$$s.t. 0 \leq \delta_1, \delta_2, \delta_3, \delta_4 \leq 1 \tag{2.14}$$

$$\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2 + \delta_3 + \delta_4 = 1 \tag{2.15}$$

$$\mathscr{F}_i^o \geqslant 0, |\mathscr{F}_i^o| = 1, i = 1, ..., N \tag{2.16}$$

$$\mathscr{F}_j^g \geqslant 0, |\mathscr{F}_j^g| = 1, j = 1, ..., V \tag{2.17}$$

The majority of the existing methods at the output level, as displayed previously, are based on the consensus maximization, which demands the most possible agreement for every single model. These proposed ensemble methods assume that: 1) labels are generated from hard-classification and hard-clustering; 2) the object in the same class should be in the same cluster; 3) the number of classes should be equal to the number of clusters.

In this section, we introduced existing methods to combine several supervised and unsupervised methods and analyze their pros and cons. Considering our research focusing on the limited training samples, we consider solving this fusion problem by the theory of belief functions, which is detailed in the next section.

2.4 Theory of belief functions

Uncertainty is one of the most important concepts in the real world, which represents a state of limited knowledge or imperfect information. Often, the term uncertainty involves two different sorts: aleatory uncertainty and epistemic uncertainty. The former represents the variability of entities in a certain population, *i.e.* frequency probability. It is an objective description and thus cannot be reduced. The latter, epistemic uncertainty, corresponds to the lack of information and thus can be reduced. It can be described by Bayesian/evidential probability, which represents a subjective degree of belief in events. Many theories, such as Bayesian model, belief functions, fuzzy set theory, possibility theory, and rough set theory are proposed to handle epistemic uncertainty.

Bayesian model employs Bayes' rule to combine a new piece of evidence which is supposed to be reliable and in the form of a certain statement [126]. In the Bayesian fusion framework, information from different sources is integrated as the joint conditions for the interesting object. After the combination based on Bayes'rule, decisions are usually made by the maximum posterior. However, many other criteria, such as maximum likelihood, maximum entropy, have been developed in order to find the best way to adapt the objective and the context of the decision. Bayesian fusion has been widely applied in different applications [148], however it can not deal with imprecise information.

In many real-world problems, however, evidence or data can be uncertain due to ill-defined concepts (*e.g.* fuzzy data), or unreliable sources, *etc.* Belief functions, considered as a generalization of Bayesian theory, can combine existing evidence with new ones

which could be uncertain. Belief functions deal with uncertain evidence by removing the concept of the conditioning in Bayes' rule and represent unreliable measures as discounted probabilities. The theory of belief functions, also called Dempster-Shafer theory, is a powerful method to deal with this sort of uncertainty and also imprecision [130], [34]. Belief functions has great advantages in image and vision fusion because it is able to represent spatial information in images as well as its imprecision on unions and the total ignorance [20], [129].

In contrast to Bayesian fusion and belief functions, fuzzy set theory proposed by [160] more focus on dealing with vague data. Fuzzy reasoning has been widely used in land cover classification based on rules and expert knowledge thanks to its ability representing vague data. A fuzzy-based multi-sensor data fusion classifier is developed in [138] for land cover classification, which integrates multi-sensor, contextual and a prior information in a single and a homogeneous framework. Furthermore, a fuzzy-fusion inference approach for image classification that uses reinforcement aggregation operators to perform the inference reasoning, in which fuzzy rules of land covers have been established based on spectral signature [127]. Compared to probability and belief reasoning, fuzzy set theory is more suitable to describe fuzzy membership of a target in an ill-defined class. The fuzzy membership is also required as prior knowledge. Furthermore, fuzzy set theory has been often integrated with Bayesian fusion [46] and belief functions fusion algorithms [165] in a complementary manner. Based on fuzzy set theory, possibility theory was first founded by Zadeh [161] and later extended by Dubois and Prade [45]. Unlike fuzzy set theory, possibility theory deal with the incomplete data in a similar way as Bayesian model and belief functions only with a different quantification approach. Compared to other theories aforementioned, possibility theory has not been commonly used in the data fusion community, yet it is argued to be most appropriate in poorly informed environments, such as no statistical data available [44].

After Bayesian model, belief functions and fuzzy set theory, rough set theory [112] is a new mathematical tool to represent imprecise data, ignoring uncertainty at different granularity levels. Rough set theory is considered as one of the first non-statistical approaches in data analysis, which is concerned with the classification and analysis of imprecise, uncertain or incomplete information and knowledge. In other words, the rough set approach can be considered as a formal framework for discovering facts from imperfect data. Rough set theory would allow the approximation of possible states of the system based on the granularity of input data. After approximating data pieces as rough sets, information can be fused using classic set theory conjunctive or disjunctive fusion operators as intersection or union. The major advantage of rough set is that it does not require any preliminary or additional information such as data distribution or membership function [158]. Rough set theory allows for fusion of imprecise data approximated based merely on its internal structure [149].

We summarize different uncertainty theories in Table 2.2 and give a little discussion on their advantages and limitations. Bayesian model is the most widely used models yet only focuses on modeling uncertainty. As a generalized version of Bayesian model, belief functions measures uncertainty as well as imprecision, which can better model information from various sensors thus is widely used in data fusion. Compared to belief functions, fuzzy set theory and possibility theory measures vagueness and incompleteness. The fuzzy

membership and possibility distribution can be also represented in the framework of belief functions. Probability and possibility are also proven as special cases in belief functions [135]. Rough set theory focuses on modeling granularity and has not been shown great advantages in data fusion. In our research, we focus on land cover data from different sensors/sources where imperfection of information is usually caused by similar spectral properties of different land covers and different spectral properties of the same land cover. This generates uncertain and imprecise information and is most appropriate to model by belief functions. Therefore, we select belief functions in our research instead of other uncertainty theories.

In this section, we present the basic details of the belief functions to localize our research interest in this framework.

Table 2.2 – Comparison of different uncertainty theories.

	Modeling of imperfection	Advantages	Limitations
Bayesian model	Uncertainty	Has well-established distribution to model uncertainty of data.	Only deal with uncertainty.
Fuzzy set theory	Vagueness	Can be used to express rules or expert knowledge.	Only handle with vague data.
Possibility theory	Incompleteness	Deals with incomplete data in poorly informed environment.	Not commonly used in data fusion.
Rough set theory	Granularity	Deals with data granularity.	Has been rarely applied to data fusion problems.
Belief functions	Uncertainty Imprecision	Enables fusion of uncertainty and imprecise data.	High computational Complexity.

2.4.1 Representation of information

Mass functions

Consider a decision from the source E regarding a variable X. All possible states ω_x of X construct a finite set Ω, called the frame of discernment. The information supporting the decision on X can be quantified by a basic belief assignment (BBA), also called mass function $m^\Omega : 2^\Omega \longrightarrow [0, 1]$, verifying:

$$\sum_{A \subseteq \Omega} m_E^\Omega(A) = 1 \tag{2.18}$$

A mass function m_E^Ω represents the state of knowledge of the source E about the variable X on Ω. When there is no ambiguity about the frame of discernment, Ω can be omitted and a BBA is simplified as m. Given a variable X, for any element $\omega \in \Omega$,

ω is called the *singleton* and the value of $m(\omega)$ represents the belief degree to support the membership of X in ω. For $B \subset \Omega$, the value of $m(B)$ represents the belief degree committed to the hypothesis that X belongs to B.

Definition 1 (*Focal element*). Focal element is a subset A of Ω verifying $m(A) > 0$.

Focal elements include $\emptyset$ in the open-world assumption (the true state is outside of the frame of discernment Ω). For the closed-world assumption (Ω contains all possible state of variables), $\emptyset$ is not a focal element, thus $m(\emptyset) = 0$.

Definition 2 (*Vacuous BBA*). A BBA is vacuous if Ω is the unique focal element, *i.e.* $m(\Omega) = 1$.

Vacuous BBA represents the total ignorance. It implies the true state of variable X always belongs to Ω.

Definition 3 (*Categorical BBA*). A non-vacuous BBA has only one focal element A, *i.e.* $m(A) = 1, A \subset \Omega$. Categorical BBA is usually denoted as $m_{[A]}$.

Categorical BBA represents that the true state of variable X is always A without any uncertainty.

Definition 4 (*Bayesian BBA*). A BBA is Bayesian when all its focal elements are on singletons.

Probability distributions are considered as a kind of Bayesian BBA, which represent precise but uncertain information.

Definition 5 (*Simple BBA (SBBA)*). A BBA is simple if it has no more than two focal elements, Ω being included. In this dissertation, we denote the SBBA as m_A, where A is the focal element besides Ω.

Simple BBA represents that the state of variable X could be A and nothing more. It supports a unique hypothesis.

Definition 6 (*Dogmatic BBA*). A BBA is dogmatic if Ω is not a focal element.

Dogmatic BBA implies a strong assumption that no global imprecise information is involved. It should be cautiously used because dogmatic BBA may sometimes lead to contradictory results in combination.

Other representations

In the theory of belief functions, besides mass functions, many representations of information are available, such as belief, plausibility, commonality, the weight of evidence.

Definition 7 (Belief and plausibility). Belief (Bel) measures all evidences that supports B, $\forall B \subseteq \Omega$. It is defined as:

$$Bel(B) = \sum_{A \subseteq B} m(A) \tag{2.19}$$

Plausibility (Pl) measures the evidences consistent with B, $\forall B \supseteq \Omega$. It is defined as:

$$Pl(B) = \sum_{A \cap B \neq \emptyset} m(A) \tag{2.20}$$

The uncertainty on a proposition B can be represented by both Bel(B) and Pl(B), with $Bel(B) \leq Pl(B)$. The interval of $[Bel(B), Pl(B)]$ represents the upper and lower

bound of probability. For Bayesian BBA, the interval is reduced to a point and Bel becomes the probability measure.

Definition 8 (*Commonality*). The commonality function q is defined as:

$$q(A) = \sum_{B \supseteq A} m(B) \tag{2.21}$$

The commonality $q(A)$ represents the belief degree that could potentially support A with further information.

Mass functions (BBA), belief, plausibility and commonality are four equivalent representations of a piece of evidence. Mass functions can be calculated by Bel, Pl or q as:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} Bel(B) \tag{2.22a}$$

$$= \sum_{B \subseteq A} (-1)^{|A|-|B|+1} (Pl(\bar{B})) \tag{2.22b}$$

$$= \sum_{B \supseteq A} (-1)^{|A|-|B|} q(B) \tag{2.22c}$$

Definition 9 (*Weight of evidence*) For a simple BBA as:

$$m_A(A) = s \tag{2.23a}$$

$$m_A(\Omega) = 1 - s \tag{2.23b}$$

with $s \in [0, 1]$, the weight of evidence $w \in [0, \infty]$ is defined as:

$$w(A) = -ln(m_A(\Omega)) \tag{2.24}$$

The weight of evidence w can be calculated by commonality q as:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}, \forall A \subset \Omega \tag{2.25}$$

Discounting

In the theory of belief functions, we can use discounting operation to model the reliability of the source regarding a piece of information. Its basic process is to weaken a mass functions by transferring uncertainty on singletons and local ignorance on unions to the total ignorance, as following:

$$^{\alpha}m(A) = \alpha m(A) + (1 - \alpha)m(\Omega), \forall A \subseteq \Omega \tag{2.26}$$

with $\alpha \in [0, 1]$. A discounting coefficient of α reflects the reliability of the source. $\alpha = 1$ means the source is reliable and the information it provides can be entirely taken into account. On the contrary, a null α indicates that the source is not reliable at all and thus we obtain total ignorance.

***Least Commitment Principle* (LCP)**

The *Least Commitment Principle* plays a significant role in belief functions, as does the principle of maximum entropy in Bayesian Probability Theory. Given several BBAs compatible with a set of constraints, the least informative also called the least committed, should be selected [134]. This principle requires partial orderings to compare belief functions by their information content. Many partial orderings are proposed by Yager [154], Dubois and Prade [43], and Denœux [36], defined as:

1. Pl-ordering: For all $A \subseteq \Omega$, if two BBAs m_1 and m_2 verify $Pl_1(A) \leq Pl_2(A)$, it implies m_1 is Pl-more committed than m_2, noted as $m_1 \sqsubseteq_{Pl} m_2$;

2. q-ordering:For all $A \subseteq \Omega$, if two BBAs m_1 and m_2 verify $q_1(A) \leq q_2(A)$, it implies m_1 is q-more committed than m_2 by, noted as $m_1 \sqsubseteq_q m_2$;

3. w-ordering: For all $A \subseteq \Omega$, if two BBAs m_1 and m_2 verify $w_1(A) \leq w_2(A)$, it implies m_1 is w-more committed than m_2, noted as $m_1 \sqsubseteq_w m_2$;

4. s-ordering: For all $A \subseteq \Omega$, if there exists a square matrix S verifying:

$$\sum_{B \subseteq \Omega} S(A, B) = 1 \tag{2.27}$$

$$S(A, B) > 0 \Rightarrow A \subseteq B \tag{2.28}$$

such that

$$m_1(A) = \sum_{B \subseteq \Omega} S(A, B) m_2(B) \tag{2.29}$$

then m_1 is said to be a specification of m_2, noted as $m_1 \sqsubseteq_s m_2$, indicating m_1 is s-more committed than m_2.

For two non-dogmatic BBAs m_1 and m_2, we have:

$$m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{Pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{cases} \tag{2.30}$$

2.4.2 Combination and decision

Combination process focuses on synthesizing multiple pieces of imperfect information modeled by mass functions. Many combination rules are available in belief functions to meet different requirements in the real world problems. In this section, we focus on the most commonly used combination rules in belief functions.

Given two BBAs m_1 and m_2 induced by two independent and reliable sources, we can combined them by the conjunctive rule to compute a new BBA as:

$$m_1 \textcircled{\cap} m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \forall A \subseteq \Omega \tag{2.31}$$

The conjunctive rule may yield an unnormalized BBA with $m(\emptyset) > 0$ in the open-world assumption. In this dissertation, we only consider close-world assumption where BBAs expressing pieces of evidence are always normalized, *i.e.* $m(\emptyset) = 0$.

The normalized conjunctive rule, also referred as Dempster's rule is used to combine evidence in the closed-world assumption, defined for all $A \subseteq \Omega$ by:

$$m_1 \oplus m_2(A) = \begin{cases} \frac{m_1 \textcircled{\cap} m_2(A)}{1 - m_1 \textcircled{\cap} m_2(\emptyset)} & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{2.32}$$

The term $\sum_{B \cap C = A} m_1(B) m_2(C), \forall A \neq \emptyset$ measures the inconsistency between m_1 and m_2. Dempster's rule is valid if and only if the conflict is less than 1. Otherwise, the two BBAs are incompatible and thus cannot be combined by Dempster's rule. Dempster's rule is commutative, associative, and has the vacuous BBA as the unique neutral element. When combining different pieces of evidence, it requires that the combined pieces of evidence induced by cognitively independent and reliable sources.

If we only know at least one source is reliable, we can use the disjunctive rule. For all $A \subseteq \Omega$, we have:

$$m_1 \textcircled{\cup} m_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C) \tag{2.33}$$

If the two sources are dependent, we can use the average rule of combination, calculated as:

$$m_{1,2}(A) = \frac{1}{2}(m_1(A) + m_2(A)) \tag{2.34}$$

Decision is the last step in the framework of belief functions, based on combined mass functions from multiple sources. For an object x, a subset of the frame of discernment has to be chosen to maximize/minimize certain criterion. In belief functions, several decision rules are possible and most of the time applied to the choice of a singletons. *The maximum plausibility*:

$$x \in A, \quad \text{if } Pl(A) = \max_{A \in \Omega}\{Pl(A)\}, \forall A \in \Omega \tag{2.35}$$

The maximum credibility:

$$x \in A, \quad \text{if } Bel(A) = \max_{A \in \Omega}\{Bel(A)\}, \forall A \in \Omega \tag{2.36}$$

The maximum pignistic probability:

$$x \in A, \quad \text{if } Betp(A) = \max_{A \in \Omega}\{Betp(A)\}, \forall A \in \Omega \tag{2.37}$$

where $Betp$ represents the pignistic probability that dissipates the mass values associated with focal elements to a specified focal element, given by:

$$BetP(A) = \sum_{A \in B \subseteq \Omega} \frac{m(B)}{|B|(1 - m(\emptyset))} \tag{2.38}$$

For the decision on both singletons and unions, we can use the minimum Jousselme distance [48] in the rule of Appriou [9].

The minimum Jousselme distance:

$$x \in A, \quad \text{if } d_J(m_x, m_{[A]}) = \min_{A \in 2^\Omega}\{d_J(m_x, m_{[A]})\} \tag{2.39}$$

where $m_x$is the BBA of an object x, and $d_J(m_x, m_{[A]})$ is Jousselme distance between the m_x and the categorical $m_{[A]}$ with A as the focal element. Jousselme distance [72] between two BBA m_1 and m_2 is defined by:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T J(m_1 - m_2)} \tag{2.40}$$

where J is the Jaccard weighting matrix defined as:

$$J = \frac{|A \cap B|}{|A \cup B|}, \forall A \subseteq \Omega, \forall B \subseteq \Omega \tag{2.41}$$

The rule of Appriou:

The principle is to weight the decision function, such as the pignistic probability, by an utility function relying on the cardinality of the elements. For $L \in 2^\Omega$ is chosen as the label if:

$$L = \underset{X \in 2^\Omega}{argmax}(m_d(X) BetP(X)) \tag{2.42}$$

where $m_d(X)$ is a mass defined by:

$$m_d(X) = \frac{K_d \lambda_X}{|X|^r} \tag{2.43}$$

K_d is a normalization factor and λ_X is applied to integrate the lack of knowledge about one of the elements of 2^Ω. $|X|$ stands for the cardinality of X. The value r ranges from 0 to 1, allowing to choose a decision which varies from a total ignorance when r is 0 and a decision based on a singleton when r is equal to 1.

2.4.3 Operations over the frame of discernment

The frame of discernment Ω represents all possible states for a piece of uncertain and/or imprecision information which is usually modeled by BBAs over 2^Ω. It is sometimes necessary to change the frame of discernment of certain BBAs to better handle the information. In this section, we present some commonly used operations, including *refinement*, *coarsening* and *conditioning*, to change the frame of discernment of BBAs. Besides, we employ a new concept *transformation* to describe some operations to change the frame of discernment of BBAs in more general cases.

2.4.4 Refinement and coarsening

A frame of discernment Ω is a refinement of a frame of discernment Θ if elements of Ω can be obtained by splitting some or all of the elements of Θ. There exists a one-to-one refining mapping $\rho : 2^\Omega \to 2^\Theta$ for all $B \subseteq \Theta$ verifying:

$$m^\Theta(B) = \begin{cases} m^\Omega(A) & \text{if } A \subseteq \Omega,\ B = \rho(A), \\ 0 & \text{otherwise,} \end{cases} \tag{2.44}$$

If two frames of discernment have the same refinement, then the two frames are said compatible. Coarsening is the opposite operation of refinement. If a frame of discernment Θ is a refinement of Ω, then Ω is a coarsening of Θ.

2.4.5 Marginalization and vacuous extension

Marginalization is an operation to transfer a BBA defined on a product space $\Omega \times \Theta$ to Ω as:

$$m^{\Omega\times\Theta\downarrow\Omega}(A) = \sum_{B\subseteq\Omega\times\Theta, proj(B\downarrow\Omega)=A} m^{\Omega\times\Theta}(B), \forall A \subseteq \Omega \tag{2.45}$$

where $proj(B \downarrow \Omega)$ is the projection of B onto Ω. Vacuous extension is the converse operation defined by:

$$m^{\Omega\uparrow\Omega\times\Theta}(B) = \begin{cases} m^{\Omega}(A) & \text{if } B = A \times \Theta, A \subseteq \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{2.46}$$

2.4.6 Transformation with more information

It is usually difficult to directly transfer a BBA in one frame of discernment to another one. The commonly used operations, as presented previously, are often constrained to specific cases.

Due to the lack of information, the transformation of BBAs often assigns uncertainty on local or total ignorance according to *the Least Commitment Principle*. It is possible to take into account more information and to generate a more general transformation of BBAs.

Given two independent frames of discernment, Ω and Θ, a transformation proposed by Karem [50] is to transfer a BBA m^{Θ} on Θ as a BBA m^{Ω} based on the likeness measured by proportion. This transformation is originally proposed for classification and clustering problems, and thus the likeness specifically refers to two groups of objects with different labels (*i.e.* a class and a cluster).

The likeness is modeled by a BBA m_{ns} for an object x, defined by:

$$m_{ns}(A) = \frac{|\{x : x \in A\} \cap \{x : x \in B\}|}{|\{x : x \in B\}|}, A \subseteq \Omega, B \subseteq \Theta, x \in X \tag{2.47}$$

The transformation from Θ to Ω is defined by:

$$m^{\Theta\uparrow\Omega}(A) = m^{\Theta}(B) m^{\Omega}_{ns}(A) \tag{2.48}$$

$$m^{\Theta\uparrow\Omega}(\Omega) = 1 - m^{\Theta}(B)(1 - m^{\Omega}_{ns}(\Omega)) \tag{2.49}$$

The BBA m_{ns} can be considered as a piece of evidence indicating the confidence about the statement that a group of objects $A, \forall A \in \Omega$ and a group of objects $B, \forall B \in \Theta$ is the same. The original BBA m^{Θ} is considered as a discounting factor to modify the reliability of the evidence modeled by m_{ns}. Compared to the traditional operations, such

as ballooning extension, this transformation takes into account the similarity information so that it avoids assigning the most of uncertainty on local or total ignorance. This is more useful in combination problems because the transferred BBAs are still informative even after transformation, thanks to the consideration of additional information-likeness.

Our research interest focus on the fusion of heterogeneous information, which requires a transformation to change all BBAs into the same frame of discernment. Each transferred BBAs are considered as a source to combine with others so that they are supposed to be informative. More detailed propositions and discussions on the transformation with more information are presented in chapter 4.

2.5 Conclusion

In this chapter, we give the introduction of state-of-the-art, including multi-spectral land cover classification, information fusion and ensemble learning, and the basic concepts of the belief functions.

Our research focuses on the combination of supervised and unsupervised methods. From a land cover classification perspective, the objective of our research is to improve accuracy and robustness when training samples are limited and uncertain. Compared to the combination of several supervised methods, merging with supervised and unsupervised methods appear to be more robust and may further reduce the need for training samples.

Due to limited training samples in our study, we pay attention to the semi-supervised ensemble method at the output level (*i.e.* combination of supervised and clustering results). Most of the existing methods considered results from clustering as supplementary constraints for supervised methods, which may make supervised and unsupervised methods in an unequal position. BGCM, C3E, and EC3 take the probability of class distribution as the core of the objective function in which clustering results are used as supplementary constrains. We may doubt that if the final combination results depend more on supervised methods than unsupervised methods. It indicates that on the one hand, sufficient training samples are required to guarantee the performance of supervised methods, and on the other hand, unsupervised methods have not been taken full advantage. Furthermore, these existing methods concentrate on the situation where all supervised and unsupervised sources achieve the same number of classes or clusters and their combination results are limited to the decision merely on singletons, which cause loss of information.

To avoid the drawbacks of the traditional semi-supervised methods above, we consider solving the combination of supervised and clustering results in the framework of belief functions, so that uncertainty and imprecision can be taken into account. To combine heterogeneous information represented by BBAs on a different frame of discernment, transforming all BBAs into the same frame of discernment is of great importance. The existing operations over the frame of discernment cannot satisfy the requirements in our problems, because most of them achieve the transformation with the assumption that no more information is available. Often, many BBAs are possible as the outcomes of the same transformation, and the least informative one should be selected according to the *Least Commitment Principle*. Therefore, the BBAs after transformation usually distribute

uncertainty on local or total ignorance yet have no information on singletons. In the fusion of heterogeneous information, these less informative transferred BBAs are not suitable to use as a source in combination. Furthermore, it is also difficult for these operations to transfer BBAs into no matter which frame of discernment. Therefore, we focus on the transformation that can be used in heterogeneous information fusion. More details related to the transformation of BBAs are discussed in chapter 4.

Chapter 3

Land cover identification based on belief functions

Resume

In this chapter, we manage to apply belief functions in land cover classification methods and present our contribution to land cover identification. As presented in chapter 2, land cover classification based on supervised methods requires a large number of labeled samples to guarantee accuracy. However, labeled samples are unavailable for all data provided by TOTAL. Therefore, we consider to automatically generate land cover maps

by spectral indexes without using labeled samples. We propose a water detection method and an automatic land cover identification method using information from vegetation index instead of labeled data, which will be used as a classifier to generated land cover maps in chapter 5. In this chapter, we also apply object association based on belief functions to combine the information from clustering results.

3.1 Introduction

Collecting sufficient labeled samples is an important yet difficult work in land cover classification, especially for hard-to-access areas. When labeled data is unavailable, some expert knowledge, such as spectral indexes, can be used to provide information on land covers. Different land covers show their specific reflections in different spectral bands, which can be used to identify land covers [133]. Often, spectral indexes/signatures are easily influenced by many factors, such as sensors, resolutions, study areas, *etc.* that related thresholds of land covers requires an amount of labeled data to determine. We thus study the automatic yet less accurate thresholds determined by spectral properties and manage to combine them by belief functions.

In this chapter, we introduce our related contributions to using belief functions in land cover classifications. Unlike traditional researches, we focus on applying belief functions in an automatic context where no labeled data are required yet the results have semantic meanings, by taking into account spectral indexes/signatures. We present the basic spectral indexes based methods for identifying water, bare soil, vegetation, and impervious surface in section 3.2. For each of them, the correspondent BBAs modeling is also presented in section 3.2. In section 3.3, we propose an automatic water detection method based on belief functions, using spectral properties mentioned in section 3.2. Similarly, in section 3.4, we propose another automatic land cover identification. In section 3.5, we manage to apply object association by combining supervised and unsupervised methods. We draw the related conclusions in section 3.6.

3.2 Basic land cover identification methods and modelling of BBAs

Spectral indexes are developed to reflect the spectral properties of certain land cover. The principal idea to automatically identify certain land covers is to determine a threshold in the corresponding spectral indexes that separates the land cover from others. As labeled data is unavailable for RapidEye and WorldView-2 provided by TOTAL, we propose some automatic identification methods based on spectral indexes to generate land cover maps without using labeled data.

In this section, we briefly introduce four automatic methods based on different spectral properties and indexes, which are constructed by spectral bands available in RapidEye and WorldView-2 (*i.e.* Blue, Green, Red, and Red-Edge). In section 3.2.1, we present a water detection method based on Near Infrared Red (NIR) and Red channels. In section 3.2.2, we present a bare soil detection method. In section 3.2.3, we propose a vegetation detection method based on normalized difference vegetation index ($NDVI$). In section 3.2.4, we propose a method to identify water and impervious surface.

3.2.1 Water detection

As water has the strongest absorption in NIR band, its NIR reflection is able to show a great difference compared to other land covers. Based on this property, a threshold

can be detected automatically to identify water and non-water, proposed in [143]. The process to determine the water threshold is detailed as the following steps:

1) Calculate the histogram of NIR band;
2) Find the two first local peaks in the NIR histogram;
3) Use a five-degree polynomial function to approximate the part between the two local peaks;
4) Find the minimal of the five-degree; polynomial approximation and use its corresponding NIR value as the threshold of water.

Water corresponds to values below this threshold in NIR. On the contrary, non-water land covers reflect superior values to the threshold. This threshold noted as T_{NIR}, thus can be directly used to separate water from other land covers as a binary classification method. A pixel labeled as *Water* must have a lower NIR value in comparison to the threshold T_{NIR}. As water has reflection than other land covers in NIR, for a *Water* pixel, the closer to T_{NIR} in NIR, the more uncertain the label. However, this does not mean it is more reliable to label the pixel as *Non-water*, for the reason that this pixel does not have a higher NIR value than T_{NIR}. Therefore, we can use SBBAs on the frame of discernment $\Omega_1 = \{Water, Non\text{-}water\}$, to represent this information, where the unique focal elements are hypothesis on *Water* and the ignorance. For *Non-water* pixels, similarly, the closer to T_{NIR} in NIR , the more uncertain the label. This can also be modeled by a SBBA where *Non-water* and the ignorance are the only focal elements.

For a pixel x, if $n_x \leq T_{NIR}$, then its BBA is defined as:

$$\begin{cases} m^{\Omega_1}(Water) = \frac{\alpha_w}{\mathscr{N}} \left(1 - e^{-\frac{T_{NIR} - n_x}{T_{NIR} - n_{min}}}\right) \\ m^{\Omega_1}(Non\text{-}water) = 0 \\ m^{\Omega_1}(\Omega_1) = 1 - m^{\Omega_1}(Water) \end{cases} \tag{3.1}$$

On the contrary, if $n_x > T_{NIR}$, we have:

$$\begin{cases} m^{\Omega_1}(Water) = 0 \\ m^{\Omega_1}(Non\text{-}water) = \frac{\alpha_{nw}}{\mathscr{N}} \left(1 - e^{-\frac{n_x - T_{NIR}}{n_{max} - T_{NIR}}}\right) \\ m^{\Omega_1}(\Omega_1) = 1 - m^{\Omega_1}(Non\text{-}water) \end{cases} \tag{3.2}$$

where n_x is the NIR value of x and T_{NIR} is the threshold of water. N is a normalization coefficient to make mass value range from 0 to 1, given by:

$$\mathscr{N} = 1 - e^{-1} \tag{3.3}$$

The coefficients α_w and α_{nw} are the two discounting coefficients for *Water* and *Non-water* classes. The values of the two discounting coefficients can be the same or different. Since the threshold is located around the lower end of NIR band, a great difference exists

between the distance from the threshold to the pixels in its left side and its right side, requiring a normalized step for the two types of distances. The coefficients $T_{NIR} - n_{min}$ and $n_{max} - T_{NIR}$ are normalized coefficients for *Water* and *Non-water* individually, defined through the largest distance from each end of NIR axis to the threshold. n_{min} stands for the minimal value in NIR and n_{max} represents the maximum value in NIR.

3.2.2 Bare soil detection

The reflectance of bare soil shows a linear relationship between NIR and *Red* bands, which is known as soil line and first introduced by Richardson and Wiegand [118]. Theoretically, even though soil line can be easily influenced by soil types, it does not vary from soil brightness caused by soil moisture. Soil line concept has been widely accepted to interpret and to analyze remote sensing imaging [12]. Soil line is defined by:

$$NIR = \beta_1 Red + \beta_0 \tag{3.4}$$

where β_1 is called soil line slope and β_0 is the intercept. Many approaches are proposed to determine the two coefficients, most of which require labeled data to analyze soil properties. We use the method proposed in [2] which can automatically estimate β_1 and β_0, by deriving a set of minimum NIR values across the *Red* band based on a user-defined bandwidth. The details to extract soil line are briefly presented as follows:

1) Derive a set of reflectance pairs $(Red_i, NIR^*_{i,n})$ where $NIR^*_{i,n}$ represents the minimum NIR value in the bandwidth n for each Red_i with $i \in [0, 255]$;

2) Calculate several soil lines using a subset of $(Red_i, NIR^*_{i,n})$ based on quarterlies, which correspond to $0\%, 25\%, 50\%, 75\%, 100\%$ of the range along the *Red* band. Take the soil line with the largest coefficient of determination as the initial soil line;

3) Remove the outlier pixels in $(Red_i, NIR^*_{i,n})$ which has the maximum vertical deviation by an iterative process, until the soil line parameters converge.

This automatic soil line indicates that the closer pixels are, the more possible they are bare soil, as shown in Figure 3.1 where pixel A is more likely to be bare soil than pixel B.

This information can be represented by BBAs on the frame of discernment $\Omega_2 = \{Soil, Non\text{-}soil\}$, using the distance from a pixel x to the soil line in the plane of NIR and Red. We label the pixels near to soil line within certain distance as *Soil* and the rest corresponds to *Non-soil*.

For pixels labeled as *Soil*, their BBAs can be modeled by:

$$\begin{cases} m^{\Omega_2}(Soil) = \frac{\alpha_s}{\mathscr{N}} \left(1 - e^{-\frac{max_{x \in X} d(p_x, \mathcal{L}) - d(p_x, \mathcal{L})}{max_{x \in X} d(p_x, \mathcal{L})}}\right) \\ m^{\Omega_2}(Non\text{-}soil) = 0 \\ m^{\Omega_2}(\Omega_2) = 1 - m^{\Omega_2}(Soil) \end{cases} \tag{3.5}$$

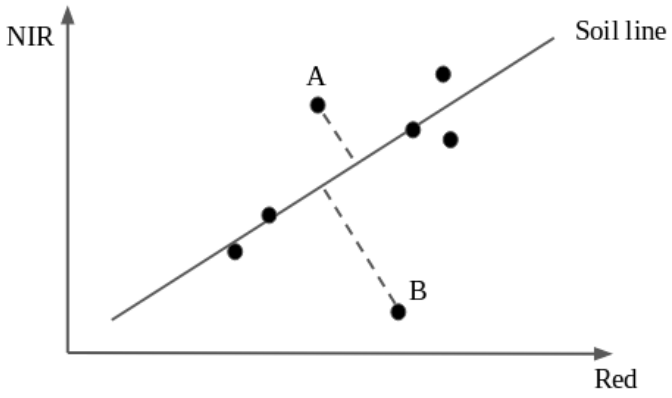


Figure 3.1. Soil line.

For pixels labeled as *Non-soil*, their BBAs can be modeled by:

$$\begin{cases} m^{\Omega_2}(Soil) = 0 \\ m^{\Omega_2}(Non\text{-}soil) = \frac{\alpha_{ns}}{\mathscr{N}} \left(1 - e^{-\frac{d(p_x, \mathcal{L})}{max_{x \in X} d(p_x, \mathcal{L})}}\right) \\ m^{\Omega_2}(\Omega_2) = 1 - m^{\Omega_2}(Non\text{-}soil) \end{cases} \tag{3.6}$$

where p_x represents the reflectance pair (Red, NIR) of pixel x, and $d(p_x, \mathcal{L})$ is Euclidean distance between p_x and soil line $\mathcal{L}$ in $Red - NIR$ plane. $max_{x \in X} d(p_x, \mathcal{L})$ represents the largest distance for all pixels $x \in X$ to soil line. α_s and α_{ns} are two discounting coefficients to modify the corresponding reliability. $\mathscr{N}$ is the same normalized coefficient in equation (3.3).

3.2.3 Vegetation detection

Vegetation has evident characteristics in different wavebands, based on which many spectral indexes are proposed. One of the most widely used ones is the Normalized Difference Vegetation Index (NDVI). NDVI, always ranging from -1 to $+1$ can quantify vegetation by measuring the difference between NIR and Red light, defined by:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{3.7}$$

However, there is no distinct boundary for each land cover type. Negative values of NDVI are often highly likely to be water. On the contrary, NDVI values close to $+1$ are possibly healthy vegetation. When close to 0, NDVI indicates that probably no green leaves exist and it could even be an urbanized area. Overall, NDVI is a standardized way to measure healthy vegetation. A threshold superior yet near to 0 in NDVI can roughly distinguish vegetation from other land covers. Traditionally, an amount of labeled data is a prerequisite to determine the threshold in NDVI. Therefore, we proposed an iterative

approach to find a threshold in NDVI without labeled data. The details of the proposed method are as follows:

1) Find a threshold thr_0 in NDVI and another threshold thr_1 in the NDVI interval [thr_0,1] through Otsu method [108].

2) Find a threshold thr_{i+1} in the NDVI interval [thr_i,1] through Otsu method until thr_{i+1} is equal or very close to thr_i, and keep all the thresholds detected in the threshold list.

3) In the interval [thr_0,thr_1], Otsu method is applied to find a new threshold thr_n. If $thr_n < thr_1$, thr_n is inserted after thr_0 in the list as the new thr_1. Continue to insert threshold until thr_1 is equal or very close to thr_0.

4) Calculate the first order of difference of the threshold list, noted as dif.

5) Use a three-degree polynomial function to approximate dif and the NDVI value corresponds to the maximal of the polynomial function is the final threshold T_{veg} to extract vegetation.

If a pixel has higher NDVI value in comparison to the threshold T_{veg}, it can be labeled as vegetation while a lower NDVI value indicates it is highly to be non-vegetation. Based on the properties of NDVI, the further to T_{veg} a pixel is, the more likely its label to be certain, as shown in Figure 3.2. We can model the information by BBAs on the frame of discernment $\Omega_3 = \{Veg, Non\text{-}veg\}$.

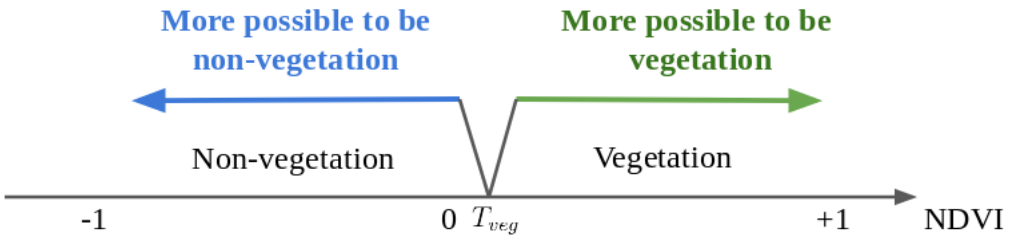


Figure 3.2. Threshold of vegetation in NDVI.

For a pixel x labeled as vegetation (*i.e.* $v_x \leq T_{veg}$), its BBAs are defined by:

$$\begin{cases} m^{\Omega_3}(Veg) = \frac{\alpha_v}{\mathcal{N}} \left(1 - e^{-\frac{v_x - T_{veg}}{T_{veg} - v_{min}}}\right) \\ m^{\Omega_3}(Non\text{-}veg) = 0 \\ m^{\Omega_3}(\Omega_3) = 1 - m^{\Omega_3}(Veg) \end{cases} \tag{3.8}$$

For a pixel labeled as non-vegetation (*i.e.* $v_x > T_{veg}$), its BBAs are defined by:

$$\begin{cases} m^{\Omega_3}(Veg) = 0 \\ m^{\Omega_3}(Non\text{-}veg) = \frac{\alpha_{nv}}{\mathscr{N}} \left(1 - e^{-\frac{T_{veg} - v_x}{v_{max} - T_{veg}}}\right) \\ m^{\Omega_3}(\Omega_3) = 1 - m^{\Omega_3}(Non\text{-}veg) \end{cases} \tag{3.9}$$

where v_x represents the NDVI value of pixel x, v_{min} stands for the minimal value in NDVI and v_{max} represents the maximum value in NDVI of all pixels. α_v and α_{nv} are two discounting coefficients to modify the correspondent reliability. $\mathscr{N}$ is the same normalized coefficient as previous BBAs.

3.2.4 Water and impervious surfaces detection

Impervious surfaces refer to artificial structures, such as roads, pavements, *etc.* as well as industrial areas. Several spectral indexes are proposed to extract impervious surfaces, including Normalized Difference Built-up Index (NDBI), Normalized Difference Impervious Surface Index (NDISI). Biological Complex Index (BCI) is a new proposed spectral index and is highly sensitive to impervious surfaces [29]. Ranging form -1 to 1, BCI indicates vegetation by negative values, impervious surfaces by positive values, and bare soil by near-to-zero values. BCI also performs well in quantifying brightness of vegetation and impervious surfaces: the further from zero a BCI value, the brighter the objects, as shown in Figure 3.3.

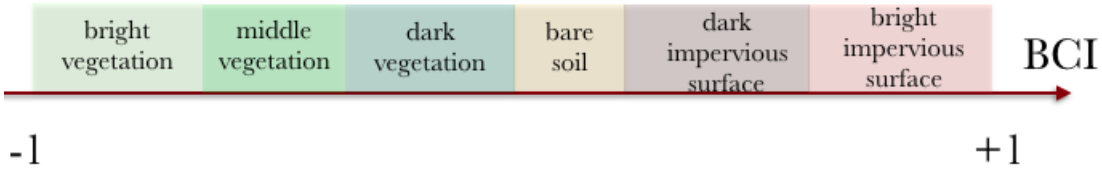


Figure 3.3. The scheme of BCI.

BCI is a normalized difference approach that compresses three components: *Brightness*, *Wetness* and *Greenness*, defined by:

$$BCI = \frac{0.5(\mathscr{H} + \mathscr{L}) - \mathscr{V}}{0.5(\mathscr{H} + \mathscr{L}) + \mathscr{V}} \tag{3.10}$$

where $\mathscr{H}$, $\mathscr{V}$, $\mathscr{L}$ refer to the normalized *Brightness*, *Wetness* and *Greenness* such that:

$$\mathscr{H} = \frac{Brightness - Brightness_{min}}{Brightness_{max} - Brightness_{min}} \tag{3.11}$$

$$\mathscr{V} = \frac{Greenness - Greenness_{min}}{Greenness_{max} - Greenness_{min}} \tag{3.12}$$

$$\mathscr{L} = \frac{Wetness - Wetness_{min}}{Wetness_{max} - Wetness_{min}} \tag{3.13}$$

The three components: *Brightness*, *Wetness* and *Greenness*, are generated by the Tasseled-Cap Transformation (TCT) which converts the original bands of an image into a new set of bands by linear combination. The linear coefficients used to create the tasseled-cap bands are derived statistically from images and empirical observations and are specific to each imaging sensor. We use the TCT coefficients originally derived from Quick Bird-2[157] yet also suitable for World View-2 [116], to calculate the three components by four visible wavebands: *Blue*, *Green*, *Red* as follows:

$$Brightness = 0.319 \times Blue + 0.542 \times Green + 0.490 \times Red + 0.604 \times NIR \tag{3.14}$$

$$Greenness = -0.121 \times Blue - 0.331 \times Green - 0.517 \times Red + 0.780 \times NIR \tag{3.15}$$

$$Wetness = 0.625 \times Blue + 0.375 \times Green - 0.639 \times Red - 0.163 \times NIR \tag{3.16}$$

Even though BCI is sensitive to the brightness of impervious surfaces and shows evident separability among different land covers, it cannot indicate water, which sometimes limits its applications. Often, a water mask is a prerequisite to use BCI in the detection of impervious surfaces. To make each one-against-all method independent, we regard impervious surfaces and water as the same land cover. Instead of the traditional sequential process that identifies water before impervious surfaces, we combine them in a parallel way by belief functions, later detailed in section 3.5. Therefore, in this section, we keep the water with impervious surfaces as the same class and propose a method to separate them from other land covers. The proposed water and impervious detection are displayed as follows:

1) For all pixels with positive BCI values, Otsu method is applied to find a threshold, noted as thr_t.

2) Apply Otsu method again in the BCI interval (0,thr_t] to find the final threshold T_{BCI}.

3) Pixels with BCI values superior to T_{BCI} are labeled as *Water-and-Impervious surface*, and BCI values inferior to T_{BCI} labeled as *Non Water-and-Impervious surface*.

Similar to the vegetation detection method, for a pixel x, the further to T_{BCI} x, the more likely its label, which can be represent by BBAs on the frame of discernment $\Omega_4 = \{WIS, Non\text{-}WIS\}$. *WIS* indicates the union of water and impervious surfaces, while *Non-WIS* represents other land covers neither water nor impervious surfaces.

For pixel x labeled as *WIS* (*i.e.* $BCI_x \geq T_{BCI}$), its BBAs are defined by:

$$\begin{cases} m^{\Omega_4}(WIS) = \frac{\alpha_{wis}}{\mathscr{N}} \left(1 - e^{-\frac{(BCI_x - T_{BCI})}{T_{BCI} - BCI_{min}}}\right) \\ m^{\Omega_4}(Non\text{-}WIS) = 0 \\ m^{\Omega_4}(\Omega_4) = 1 - m^{\Omega_4}(WIS) \end{cases} \tag{3.17}$$

For pixel x labeled as *Non-WIS* (*i.e.* $BCI_x < T_{BCI}$), its BBAs are defined by:

$$\begin{cases} m^{\Omega_4}(WIS) = 0 \\ m^{\Omega_4}(Non\text{-}WIS) = \frac{\alpha_{nwis}}{\mathscr{N}} \left(1 - e^{-\frac{(T_{BCI} - BCI_x)}{BCI_{max} - T_{BCI}}}\right) \\ m^{\Omega_4}(\Omega_4) = 1 - m^{\Omega_4}(Non\text{-}WIS) \end{cases} \tag{3.18}$$

where BCI_x represents the BCI value of pixel x, BCI_{min} stands for the minimal value in BCI and BCI_{max} represents the maximum value in BCI of all pixels. α_{wis} and α_{nwis} are two discounting coefficients to modify the correspondent reliability. $\mathscr{N}$ is the same normalized coefficient as previous BBAs.

3.3 Automatic water detection method based on belief functions

Detection of surface water in the natural environment is an important application of remote sensing. Traditional methods require the manual labeling process which is time-consuming for large and complex areas. We propose an automatic approach to detect water bodies based on belief functions. This approach uses a spectral model (NIR model), presented in section 3.2.1, to roughly identify water and to label samples that are used to training a supervised model. The results of the two models are combined by belief functions in a sequential process. The proposed method performs well in mapping principle water bodies, including little streams and branches. It also labels all objects usually confused with water as ignorance, including half-dry watery areas, built-up areas, and semi-transparent clouds and shadows. Ignorance indicates the limitations of the spectral properties of water and insufficiency of information, which can provide valuable information for further land cover classification.

3.3.1 Methodology

As water has evident properties in NIR band, we apply the water detection method presented previously, to roughly identify water from other land covers. However, very thin clouds and shadows on vegetation could be confused with small and shallow water bodies sometimes. With information from NIR band, these confusing objects may be difficult to identify from each other while it is still meaningful to gather them as a group for further study. The confusing objects always have NIR values approaching the threshold, while principle water pixels and other land covers with obvious distinction preserve some distances from the threshold.

To improve the efficiency of water detection, we combine the NIR model with a supervised model, trained by samples that are automatically selected by the NIR model. The supervised model includes two steps: the learning step and the classification step. In the learning step, labeled data is used to train parameters in the model. The classification step allows predicting labels of new data based on the learning function. However, the lack of learning data or the availability of inappropriate samples often leads to wrong

classification and low accuracy. This problem is more pronounced in water detection since reference data is often chosen from a satellite image, which results easily in a lack of enough information on surface water. Large water bodies, such as rivers or lakes, are often easy to be detected, while smaller ones like streams tend to be confused with other land covers since not sufficient information is supported to represent their spectral reflections. In this case, lacking information could increase imprecision and uncertainty of classification, causing unsatisfying results. On account of the unavailability of the ground truth, training samples are automatically labeled by the NIR model.

Considering the original spectrum (*e.g.* *Red*, *Green*, *Blue*) is not sensitive enough to identify land covers, the supervised methods could achieve better performances in a transformed feature space. The feature space is composed of specific indexes extracted from the original bands. We utilize three indexes: NDVI, the Normalized Difference Water Index (NDWI), and the Red Edge Normalized Difference Water Index (RE_NDWI) [75] to construct the new feature space. NDVI, calculated in equation (3.7), is strongly related to vegetation and reflects the difference of basic land covers. NDWI improves the separability of water from vegetation and soil while sometimes it is not able to distinguish water from built-up areas efficiently. To overcome this drawback, we use another index RE_NDWI based on *Red-edge* band, to increase the separability of water from built-up areas. NDWI and RE_NDWI are calculated as:

$$NDWI = \frac{NIR - Green}{NIR + Green} \tag{3.19}$$

$$RE\_NDWI = \frac{Green - Red\text{-}edge}{Green + Red\text{-}edge} \tag{3.20}$$

Due to the lack of Shortwave Infrared (SWIR) band in our study, we calculate $NDWI$ through *Green* and NIR bands [143]. In our proposed approach, the supervised method is conducted in the three-dimensional space composed of $NDVI$, $NDWI$, and RE_NDWI. As we have explained before, the NIR model and the supervised model are not independent since the former provides training samples to the latter, signifying the supervised model has more reliable results in comparison to the spectral model. Therefore, it is necessary to modify BBAs of the NIR model in terms of supervised classification before fusion. The two discounting coefficients α_w and α_{nw} used in the spectral model are calculated by the confusion matrix, considering results from supervised learning as true labels, and are given by:

$$\alpha_w = p(\psi_w|\vartheta_{nw}) \tag{3.21}$$

$$\alpha_{nw} = p(\psi_{nw}|\vartheta_w) \tag{3.22}$$

We note by ϑ_w the label of pixel x assigned by supervised model as *Water* and ϑ_{nw} as *Non-water*, and ψ_w and ψ_{nw} for the counterparts in the NIR model.

For a supervised learning model, instead of directly choosing training data from a satellite image, the spectral model is applied first to generate training samples. Only data with a high belief degree of its attributed label can be utilized to train the supervised learning model. Training samples are chosen separately for *Water* and *Non-water*. BBAs are defined based on the distance to the center of class in the feature space. For instance, a pixel x, more being away from the center of *Water* signifies that it is less credible to make x

pertain to this class. Nevertheless, this does not mean into *Non-water* is more reasonable for x, leading to an augmentation of the BBA value of ignorance. Since the supervised model is generated using results from the spectral model, it should be considered to precede the spectral model. This brings these two models into the dependent position instead of traditionally independent classifiers in belief functions framework. Therefore, the NIR model can regard results from the supervised model as ground truth to calculate its discounting coefficients.

The two models employed here are not independent for the reason that training samples used in the supervised model are generated from the spectral model. And the discounting coefficient of the spectral model was updated based on the supervised model. Traditional combination rules in belief functions framework only deal with the situation in which all the sources are independent. Thus, we applied the average rule of combination which allows us to consider multiple perspectives from different dependent sources.

For the supervised model, we note $m_{sup}^{\Omega_1}(Water)$ for BBA of *Water* , $m_{sup}^{\Omega_1}(Non\text{-}water)$ for *Non-water*. c_w is the center of *Water* and c_{nw} represents the center of *Non-water*. $d(c_w, x)$ and $d(c_{nw}, x)$ are the Euclidean distance from pixel x the the centers c_w and c_{nw}. The BBAs are defined by:

if $d(c_w, x) \leq d(c_{nw}, x)$

$$\begin{cases} m_{sup}^{\Omega_1}(Water) = \dfrac{\alpha_{sup}}{\mathscr{N}} \left(e^{-\dfrac{d(c_w, x)}{maxd(c_w, x_w)}} \right) \\ m_{sup}^{\Omega_1}(Non\text{-}water) = 0 \\ m_{sup}^{\Omega_1}(\Omega_1) = 1 - m_{sup}^{\Omega_1}(Water) \end{cases} \tag{3.23}$$

if $d(c_w, x) > d(c_{nw}, x)$

$$\begin{cases} m_{sup}^{\Omega_1}(Water) = 0 \\ m_{sup}^{\Omega_1}(Non\text{-}water) = \dfrac{\alpha_{sup}}{\mathscr{N}} \left(e^{-\dfrac{d(c_{nw}, x)}{maxd(c_{nw}, x_{nw})}} \right) \\ m_{sup}^{\Omega_1}(\Omega_1) = 1 - m_{sup}^{\Omega_1}(Non\text{-}water) \end{cases} \tag{3.24}$$

where x_w refers to the pixels closer to c_w than c_{nw}, and x_{nw} is the pixels more approaching to c_{nw} instead of c_w. α_{sup} is a discounting coefficient and $\mathscr{N}$ is the same normalized factor used in the spectral model.

To decide for both singletons and ignorance, the Appriou's rule is applied in our approach, as displayed previously in equation (2.42).

Principle steps of this proposed approach are shown in Figure 3.4 with the following details:

1) Use the NIR model to roughly identify water, and calculate BBAs of *Water*, *Non-water* by equations (3.1), (3.2), where discounting coefficients are initialized as 1;

2) Choose pixels randomly from each class with a relatively high mass value as the training data for the supervised model, such as SVM;

3) Use the supervised model to predict data, and calculate BBAs of the supervised results by equations (3.23), (3.24);

4) Update discounting coefficients in the spectral model by results of the supervised model by equations (3.21), (3.22);

5) Utilize the average rule to combine the two classification results by equation (2.34);

6) Calculate pignistic probability, and apply the decision rule of Appriou to attribute labels for pixels by equation (2.42).

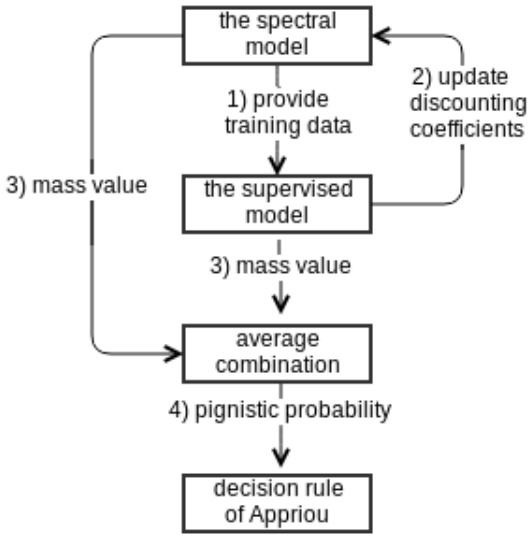


Figure 3.4. Principle steps of the proposed automatic water detection method.

As explained previously, the fusion is not directly conducted between both models since one is trained relying on the other. It is crucial to update the mass values of the spectral model before the fusion step. Since both models are dependent, we choose the average combination rule during fusion. The ignorance is also considered as a label in the final result through the decision rule of Appriou, which presents indistinguishable pixels due to restriction of the model itself and lack of specific spectral information. The results of the proposed method are presented in the next section.

3.3.2 Experiments and results

We verify the proposed method on RapidEye data with a resolution of 5 m, provided by TOTAL yet without any ground truth. The study area locates in Papua New Guinea, and there are five available wavebands, as shown in Table 3.1. We compare the water mask from the NIR model, the results of SVM, and that after fusion by belief functions. For the supervised model, SVM is applied in our study while it is also flexible to choose other supervised learning methods.

The original NIR image of RapidEye is shown in Figure 3.5, and the original image in the three-dimensional space composed of NDVI, NDWI, RE_NDWI is shown in Figure 3.6. The NIR model generates a water threshold equal to 5427.18. For all the data, it labels 8.44% as *Water* and 91.56% as *Non-water*. As shown in Figure 3.7(a), it identifies the principal water bodies and small branches, also including some parts of thin clouds and shadows. The pixels approaching the threshold in NIR bands often consist of confusing objects, which shows similar reflectance in NIR. As we explained previously, training samples for SVM are provided by the NIR model. We take labeled data with a BBA superior to 0.7 from the NIR model to feed the training process of SVM. This value can be selected flexibly according to the study area and expected accuracy of the approach.

Table 3.1 – Spectral bands of RapidEye.

channel	spectral band name	spectral coverage (nm)
1	Blue	440-510
2	Green	520-590
3	Red	630-685
4	Red-edge	690-730
5	NIR(Near-Infrared)	760-850

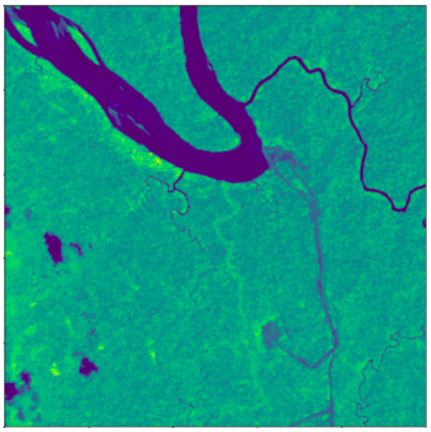


Figure 3.5. Original NIR image.

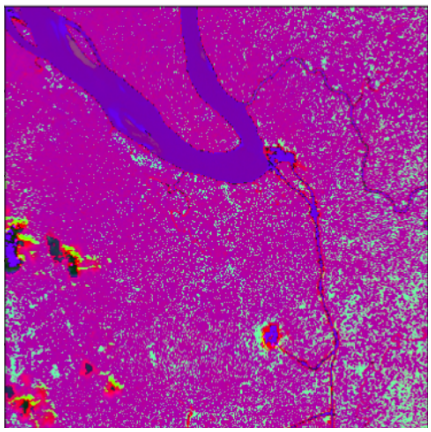
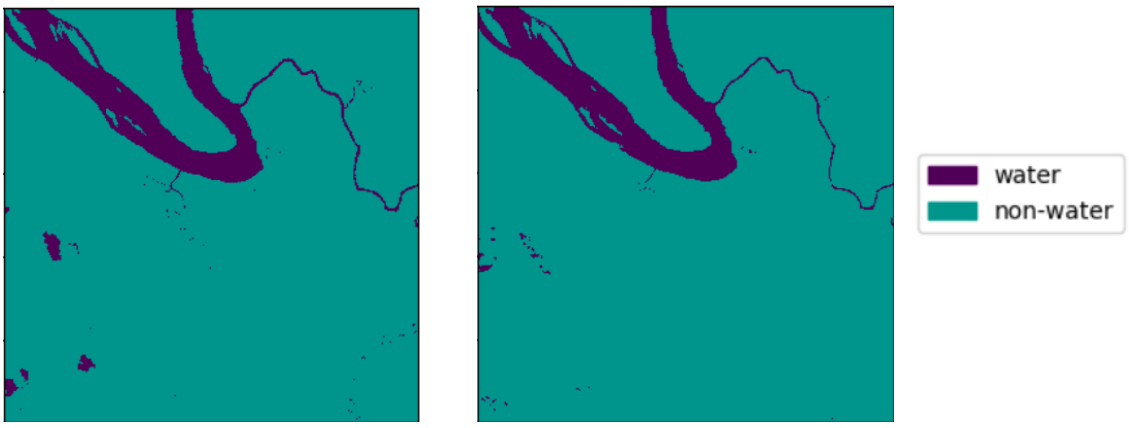


Figure 3.6. Original image in the three dimensional space.

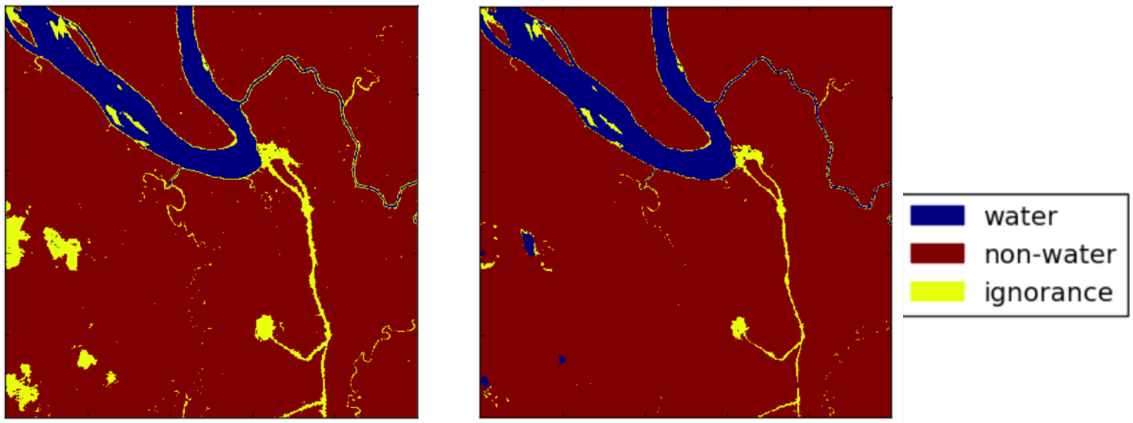
The supervised model separates clouds and shadows from surface water owing to the high belief degree of training samples, as shown in Figure 3.7(b). For the data, it classifies 7.64% as *Water*, 92.36% as *Non-water*. Water can be more effectively detected compared to the result from the spectral model, especially in extracting water from very thin clouds.

For the decision rule of Appriou, λ_X in equation (2.43) is set to 1 since we systematically announce the whole frame of discernment when there are only two singletons: *Water* and *Non-water*. For the parameter r within $[0, 1]$, it enables us to make decision ranging from the choice of a singleton to total ignorance, controlling the region of ignorance. Various values of r were tested in the experiment and the relation between r and the region of ignorance is shown in Figure 3.9. The region of ignorance gradually reduces when r



(a) NIR model. (b) Supervised model.

Figure 3.7. Classification results of NIR model and supervised model.



(a) Decision with $r = 0.1$ (b) Decision with $r = 0.9$

Figure 3.8. Decisions with $r = 0.1$ and $r = 0.9$.

changes from 0 to 1, approximating to the linear relationship. The larger r is, the less consideration is taken in ignorance. We display the results with $r = 0.1$, and $r = 0.9$ are shown here in Figure 3.8(a), and Figure 3.8(b).

Compared to results in Figure 3.7(a) and Figure 3.7(b), the proposed approach not only efficiently identifies principal river bodies but also shows clearly ignorance which includes all confusing objects that cannot be directly distinguished. Comparison between the results and original satellite data indicates the ignorance here is composed of thin clouds and its shadows, a route and half-dry watery areas, such as the edge of rivers and extremely small and shallow stream-way on the ground. The proposed method with $r = 0.1$ identifies 7.41% as *Water*, *87.92%* as *Non-water* and 4.67% as ignorance. BBAs values after fusion for each class are shown in Figures 3.10(a), 3.10(b), and 3.10(c). Although the distance for confusing objects to the NIR threshold in the spectral model is hard to

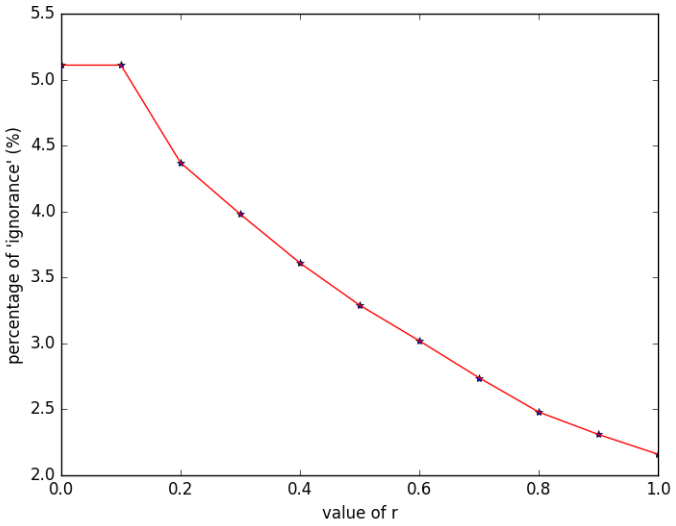
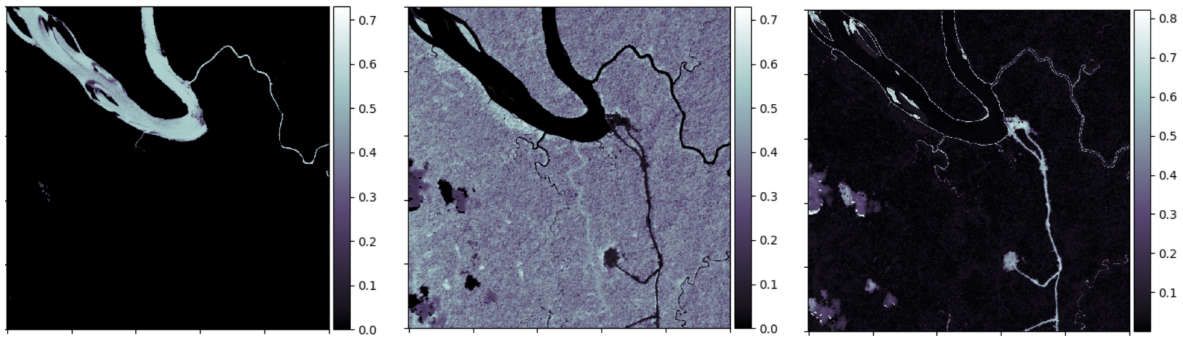


Figure 3.9. The relation between r and the percentage of ignorance.

measure, the proposed approach allows signifying all confusing objects through ignorance, clearly improving the accuracy of water detection.

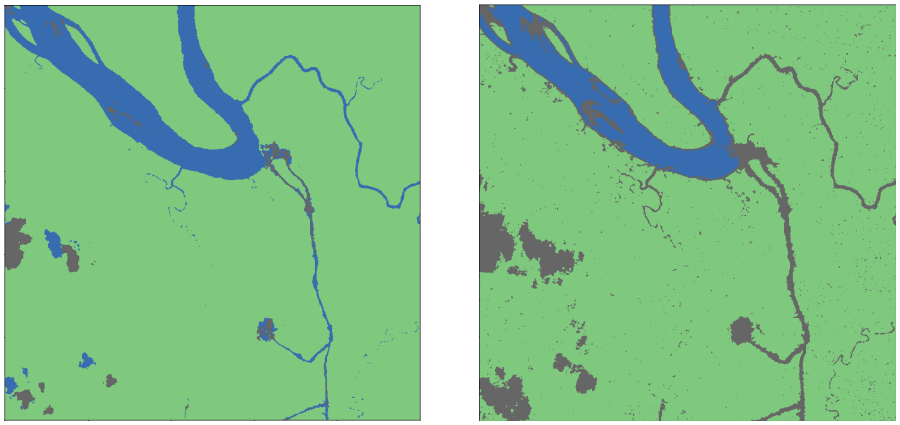


(a) BBAs on *Water*. (b) BBAs on *Non-water*. (c) BBAs on ignorance.

Figure 3.10. BBAs after fusion on each class.

Due to the lack of ground truth, we verify our approach through the visual comparison of the original multi-spectral images with the classification results. The proposed approach signifies a satisfying ability to find nearly all the principal rivers bodies and its little branches, separating water bodies from clouds and shadows. For the *Non-water* class, the proposed approach also displays very satisfying results in our manual verification in ENVI, in which nearly all the non-water areas were correctly identified. Furthermore, in the class labeled as ignorance, almost all the extremely small and half-dry stream-way were detected, also including the objects often confused lightly with water in the land cover classification. The ignorance here signifies the limitation of the spectral model and the supervised model and also represents the insufficiency of spectral information.

We also evaluate the proposed method from the aspect of clustering methods by silhouette score [123]. The silhouette value is a measure of clustering quality using the mean intra-cluster distance and the mean nearest-cluster distance for each sample. The value ranges from -1 to $+1$, where a high value indicates that samples are well clustered. The comparison results with two clustering methods: k-means and Gaussian Mixture Model (GMM) [117] are shown in Figure 3.11 and silhouette scores of different methods are displayed in Table 3.2.



(a) k-means with k=3 (b) GMM with k=3

Figure 3.11. Results of k-means and GMM.

Table 3.2 – Silhouette score of different methods.

	k-means	GMM	Decision with $r = 0.1$	Decision with $r = 0.9$
Silhouette score	0.781	0.640	0.784	0.709

The experiment results show that the proposed approach performs well on the detection of water bodies in the natural environment. Not only can large water bodies as rivers be detected efficiently, but also small water bodies are identified through disturbing objects such as clouds. In comparison with clustering results, the proposed method can generate semantic labels and decision with $r = 0.1$ has the highest silhouette score. The proposed approach overcomes drawbacks of the spectral properties of water as well as the unavailability of a certain spectral band in our data, such as Shortwave Infrared (SWIR) band. Moreover, the ignorance gathers all similar objects to water in our classification system, signifying restriction in the two basic models and multi-spectral information themselves.

Besides, the ignorance shown in the final results can provide valuable information for further land cover classification, especially in a fully automatic and unsupervised context because it helps to specify supplementary information or technology that should be applied to its inner objects. We could use supplementary information, for instance, Middle-infrared (MIR) band or some methods specific in identifying built-up or clouds from data labeled as ignorance to separate the objects we are interested in.

3.4 Automatic land cover identification method based on belief functions

Different land covers signify different spectral reflectance, some of which can be represented by spectral indexes. Often, to determine appropriate thresholds to identify land covers in spectral indexes, the amount of ground truth is required. Significant progress has been made recently in developing more powerful classifiers based on spectral properties to extract land covers [104], [137]. Complementary information of identified objects in land cover classification potentially allows a higher classification accuracy. In this section, we propose an automatic land cover identification methods based on belief functions, combining the four one-against-all methods presented in section 3.2. Different from the automatic water detection method, this approach is based on a parallel combination structure, which also performs well in the identification of land covers without ground truth.

3.4.1 Methodology

The proposed method uses belief functions to combine the four one-against-all methods, which separately and independently identify water, bare soil, vegetation and the union of water and impervious surfaces from other land covers. The related methods and their correspondent BBAs, constructed on different *frames of discernment*, are previously present in section 3.2. The four *frames of discernment* Ω_1, Ω_2, Ω_3 and Ω_4 have the same *refinement* Ω, as detailed in Table 3.3. The four methods are therefore combined on $\Omega = \{Water, Soil, Veg, IS\}$ where IS represents impervious surfaces.

Table 3.3 – Relationship among the frames of discernment Ω_1, Ω_2, Ω_3, Ω_4 and Ω.

Frame of discernment	Elements	
Ω_1	*Water*	*Non-water (Soil* $\cup$ *Veg* $\cup$ *IS)*
Ω_2	*Soil*	*Non-soil (Water* $\cup$ *Veg* $\cup$ *IS)*
Ω_3	*Veg*	*Non-veg (Water* $\cup$ *Soil* $\cup$ *IS)*
Ω_4	WIS (*Water* $\cup$ *IS*)	NWIS (*Soil* $\cup$ *Veg*)
Ω	*Water, Soil, Veg, IS*	

As the four detection methods are independent, we can use the conjunctive rule to combine their BBAs on the frame of discernment Ω. The maximum of pignistic probability is applied as the decision rule to obtain the basic land covers: *Vegetation*, *Bare soil*, *Water*, and *Impervious surfaces*. Based on the classification results on the basic land covers, BCI can be further used to indicate the brightness of vegetation and impervious surfaces. Their colors become darker when BCI is approaching to 0. For the pixels that are labeled as vegetation after fusion, we merely focus on their correspondent BCI values and use k-means with three partitions to separate *Vegetation* into three refined classes: *Bright vegetation*, *Middle vegetation*, *Dark vegetation*. We apply a similar process on the *Impervious surfaces* to refine them by their brightness. The proposed method contains

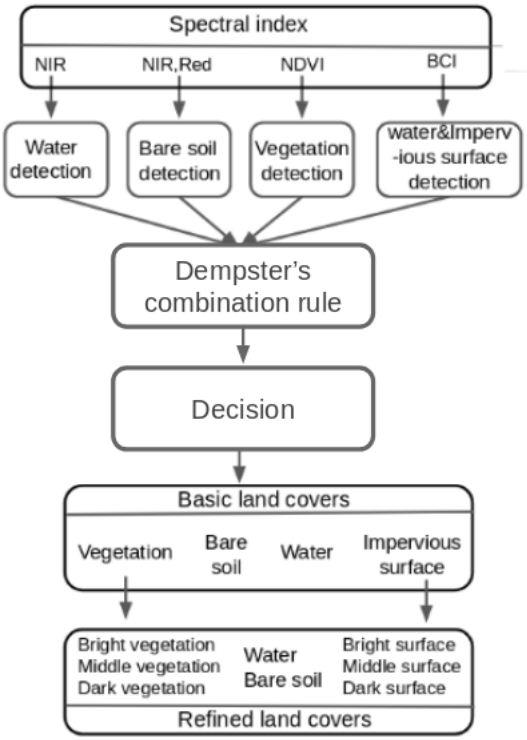


Figure 3.12. Work-flow of the automatic land cover identification method.

two principal steps: combining the four one-against-all methods to identify basic land covers; refining *Vegetation* and *Impervious surfaces* by BCI, as shown in Figure 3.12. The reason that we refine the basic land covers into refined ones after their identification by belief functions, is to avoid possible noises or influences from other land covers effect the results of k-means.

3.4.2 Experiments and results

In this section, we verify the propose automatic land cover identification method based on WorldView-2 data with the resolution of $0.5m$, also provided by TOTAL. Similar to the RapidEye data used in section 3.4, the coresponding area also locates in Papua New Guinea and no ground truth is available. Four wavebands are available for WorldView-2 in our study, as shown in Table 3.4.

Table 3.4 – Spectral bands of WorldView-2

channel	spectral band name	spectral coverage (nm)
1	Blue	450-510
2	Green	510-580
3	Red	630-690
4	NIR(Near-Infrared)	770-895

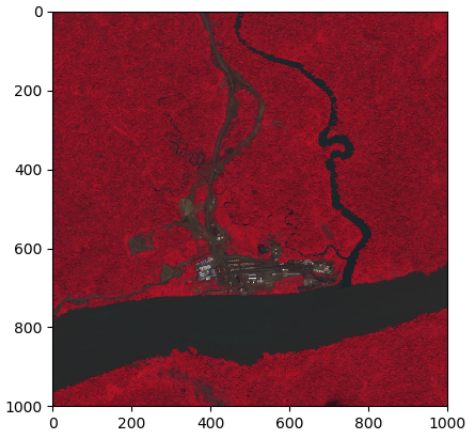
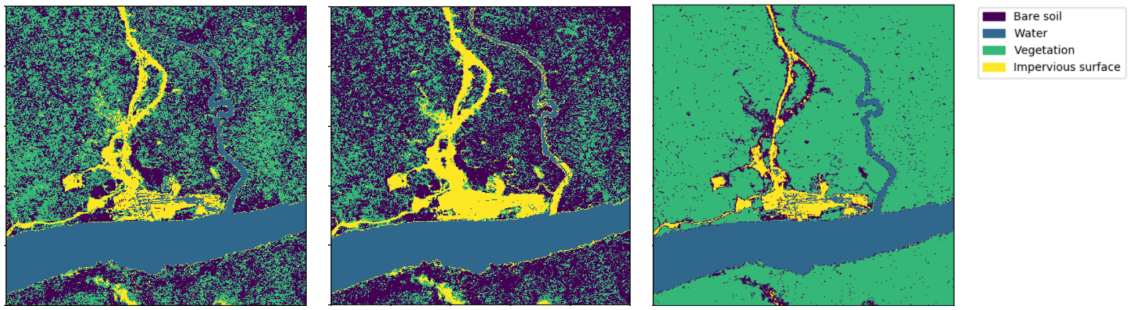


Figure 3.13. False color composite of WorldView-2 image.

We display the original false-color composite image of the study area in Figure 3.13, where vegetation appears with red color, water corresponds to dark brown or black color. Bare soils, roads, and buildings are shown in light brown color compared to water.

In this experiment, we first generate a basic land cover map identifying *Water*, *Bare soil*, *Vegetation* and *Impervious surfaces* based on belief functions. We also compare this result with two clustering methods: k-means and Gaussian Mixture Model (GMM). The number of partitions is set to 4 correspondent to four basic land covers in the frame of discernment Ω. The comparisons are displayed in Figure 3.14. Identifying vegetation as an entire cluster is difficult for k-means and GMM because spectral dissimilarities are not sufficient to make vegetation distinguishable from bare soil, water, and impervious surface. Another limitation is that clustering methods cannot provide semantic labels. On the contrary, the proposed identification method for basic land covers can automatically generate semantic labels. After the identification of basic land covers, we refine vegeta-



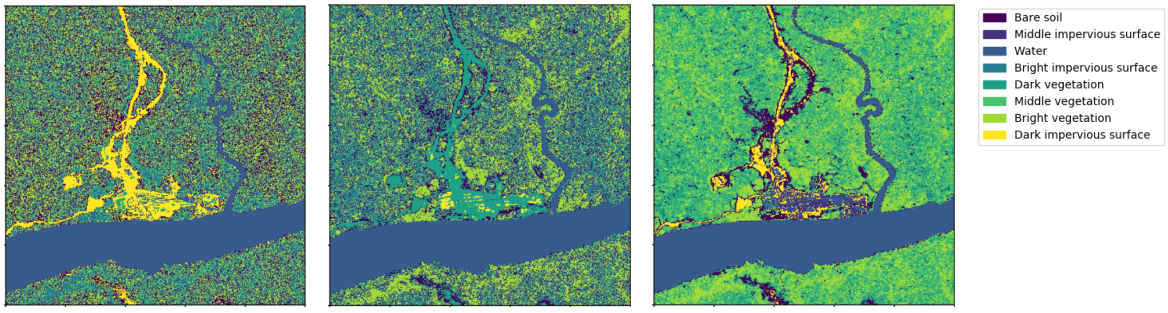
(a) k-means with k=4 (b) GMM with k=4 (c) The proposed method for basic land covers

Figure 3.14. Results of k-means, GMM and the proposed automatic method for basic land covers.

tion and impervious surface in terms of their brightness, thus generating more specific semantic labels. Vegetation and impervious surface are refined to more detailed classes

according to the brightness. Since the surface water has various spectral reluctance, GMM has less satisfying performance for identifying the water. Taking advantage of spectral features, the proposed method can efficiently distinguish surface water and also shows better performance in mapping meaningful sub-groups of vegetation and impervious surface. Compared to k-means and GMM with $k = 8$, our method still has better performance in regrouping original data and generates semantic labels for each cluster, as shown in Figure 3.15.

Due to the lack of ground truth, we evaluate the proposed methods and other two clustering methods by silhouette score. This score focuses on the measure of compactness, various and density of clusters, and thus is often applied to evaluate clustering methods. It is bounded between -1 for incorrect clustering and $+1$ for highly dense and well-separated clustering. We display silhouette scores of both basic land cover and refined land cover identification, Compared to k-means and GMM with the number of clusters set to 4 and 8. The results are shown in Table 3.5, the proposed method has a higher silhouette score in comparison to the others, which indicates the four land covers brought by our method are also better defined as clusters.



(a) k-means with k=8 (b) GMM with k=8 (c) The proposed method for refined land covers

Figure 3.15. Results of k-means, GMM and the proposed automatic method for refined land covers.

Table 3.5 – Silhouette score of different methods.

	k-means	GMM	The proposed method
Basic land covers	0.544	0.494	0.551
Refined land covers	0.458	0.406	0.476

We also verify the proposed method by visual comparison. The proposed method satisfyingly distinguishes the basic and also refined land covers. Water including a river and even some small streams are well identified. Vegetation and impervious surface with various brightness are identified. Bare soil near the road is satisfyingly detected, while it is hard to observe when merging with vegetation.

3.5 Combination based on object association

Conventional supervised classification approaches have significant limitations in the land cover classification from remote sensing data because a large amount of high quality labeled samples are difficult to guarantee. To overcome this limitation, a combination with the unsupervised approach is considered as one promising candidate. In this section, we propose a novel framework to achieve the combination through object association based on belief functions. Inspired by object association, the framework can label the unsupervised clusters according to the supervised classes even though they have different numbers.

The proposed framework is tested on the different combinations of commonly used supervised and unsupervised methods. To achieve the numerical analysis and evaluation, we use the open data of LandSat-8 OLI with ground truth instead of the data provided by TOTAL. This section is organized as follows: we present the related training and validation data in section 3.5.1, and the object-association based combination method in section 3.5.2; in section 3.5.3, we show the related experiment results.

3.5.1 Training and validation data

The study area used in this experiment locates in Colorado in United States, within latitudes 38°30′53.44″N - 36°22′23.05″N and longitudes 108°0′10.24″W - 105°18′35.60″W. It contained San Juan National Forest and Rio Grande National Forest, as shown in Figure 3.16. The two national forest parks are mainly composed of deciduous forest, evergreen forest, and mixed forest. The remaining natural vegetation still occurs as shrub, grassland, and herbaceous. A large area of pasture is also contained near Monte Vista. This selected area contains two national forest parks and various developed areas, thus can be used to mimic a complex environment with a wide range of land covers. We separate the study area into two subareas of equal size: validation area and training area.

The satellite data used in this study can have LandSat-8 OLI consisting of eight spectral bands data with a spatial resolution of 30 meters, a panchromatic band with a resolution of 15 meters, and two thermal bands with a resolution of 100 meters. The image acquired on the 11 June 2018 was obtained from USGS Earth Explorer [1]. Geometric correction of the image has already been done through UTM map projection by NASA. The bands used for our study include band 1 to 7.

Training and validation data applied in our study were obtained from National Land Cover Database 2011 (NLCD 2011) [66], which was created by the Multi-Resolution Land Characteristics (MRLC) Consortium. NLCD 2011 is a 16-class land cover classification that has been applied consistently across the United States at a spatial resolution of 30 meters. The codes for the land cover classification of NLCD 2011 are shown in Table 3.6.

In our study, we select half of the study area as the training area which includes all 16 types of land covers mentioned in Table 3.6 with the size of 2000×4000 pixels to generate training samples, as shown in Figure 3.16. The other half of the study area with the same size is chosen as the validation area as shown in Figure 3.16. To simulate the situation where training samples are insufficient for a supervised classifier, only a small amount of

Table 3.6 – Land cover classification scheme of NLCD 2011.

Code	Legend	Code	Legend
11	Open Water	42	Evergreen Forest
12	Perennial Ice/Snow	43	Mixed Forest
21	Developed, Open Space	52	Shrub/Scrub
22	Developed, Low Intensity	71	Grassland/Herbaceous
23	Developed, Medium Intensity	81	Pasture/Hay
24	Developed, High Intensity	82	Clutivated Crops
31	Barren Land (Rock/Sand/Clay)	90	Woody Wetlands
41	Deciduous Forest	95	Emergent Herbaceous Wetlands

pixels per class are randomly selected as training samples.

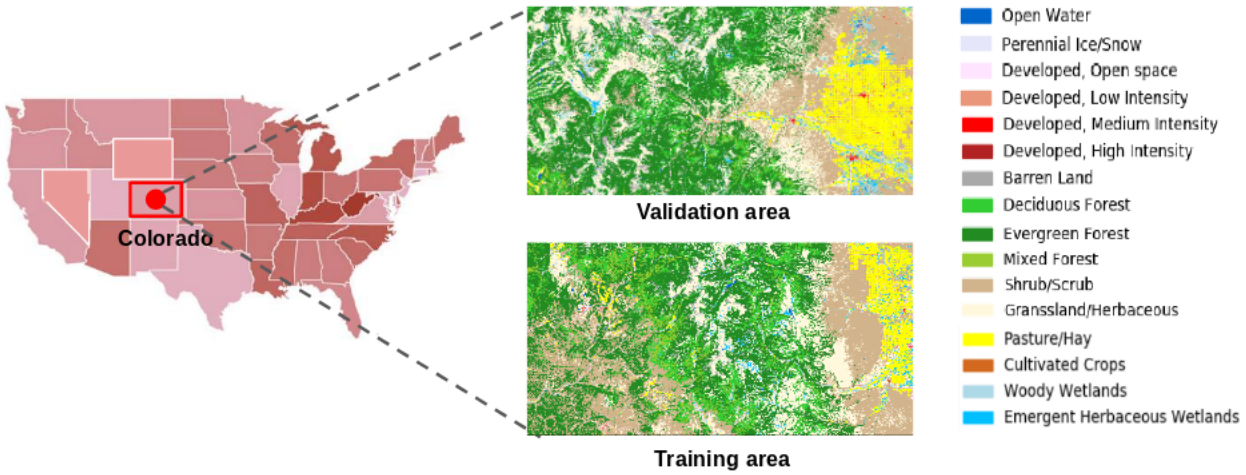


Figure 3.16. Map of the study area.

3.5.2 Methodology

Object association

Object association was originally derived from the tracking problem to estimate the status of mobile objects detected by multiple sensors. Mathematically speaking, object association refers to match two finite sets of objects $\Theta = \{\theta_1, ...\theta_n\}$ and $\Omega = \{\omega_1, ..., \omega_p\}$ with possibly different combinations. It is usually assumed that each object in one set should be matched with at most one object in the other set. An object can possibly disappear between two successive time frames, which leads to no corresponding counterpart for this object in the other set.

The problem can be considered in the framework of belief functions. Therefore, a piece of evidence presenting the possible association of two objects θ_i and ω_j can be modeled by

a mass function m_{ij} on the discernment $\mathcal{H}_{ij} = \{\mathcal{Y}_{ij}, \mathcal{N}_{ij}\}$. Let f_{ij} represent the relation between the object θ_i and ω_j. The mass function $m_{ij}(\{\mathcal{Y}_{ij}\})$ refers to the uncertainty that θ_i and ω_j are the same (*i.e.* $f_{ij} = 1$). $m_{ij}(\{\mathcal{N}_{ij}\})$ indicates that θ_i and ω_j are different (*i.e.* $f_{ij} = 0$) and $m_{ij}(\{\mathcal{H}_{ij}\})$ represents we know nothing about the relation f_{ij} .

Several solutions about choosing the best relation f^* based on pairwise mass functions m_{ij} have been proposed. In [96], the authors combines the BBA m_{ij} for each i and then find the best relation f^* through maximum pignistic probability. In [37], the problem is transferred into an equivalent linear assignment problem and then searches for the most plausible relation as the best one, which thus can be solved in polynomial time. Since the number of classes and clusters is large in our study, we apply the approach proposed in [37] on the combination of supervised and unsupervised methods. It can be stated as the following linear program:

$$\max_f (\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij} f_{ij}) \tag{3.25}$$

subject to

$$\sum_{j=1}^{p} f_{ij} \leq 1 \quad \forall i \in \{1, ..., n\} \tag{3.26}$$

$$\sum_{i=1}^{n} f_{ij} \leq 1 \quad \forall j \in \{1, ..., p\} \tag{3.27}$$

$$f_{ij} \in \{0, 1\} \quad \forall i \in \{1, ..., n\}, \forall j \in \{1, ..., p\} \tag{3.28}$$

in which e_{ij} is a coefficient generated by m_{ij} presenting the uncertainty as well as imprecision on the evidence of relation between θ_i and ω_j.

Combination of supervised and unsupervised methods

Let us denote Ω as the set of classes generated from the supervised method and Θ as the set of clusters derived from the unsupervised method. Based on the principle of object association, we have to match elements from two sets of objects objects $\Theta = \{\theta_1, ...\theta_n\}$ and $\Omega = \{\omega_1, ..., \omega_p\}$ with the most possible combination.

BBAs m_{ij} with $1 \leq i \leq n$, $1 \leq j \leq p$, indicates all available evidences about the association between Ω and Θ. Each m_{ij} encodes a piece of evidence about a binary variable f_{ij}. This variable indicates θ_i is the same type as ω_j when it equals to 1, and 0 otherwise. Each m_{ij} is thus on the discernment $\mathcal{H}_{ij} = \{\mathcal{Y}_{ij}, \mathcal{N}_{ij}\}$ having:

$$m_{ij}(\{\mathcal{Y}_{ij}\}) = \eta_{ij} \tag{3.29}$$

$$m_{ij}(\{\mathcal{N}_{ij}\}) = \beta_{ij} \tag{3.30}$$

$$m_{ij}(\{\mathcal{Y}_{ij}, \mathcal{N}_{ij}\}) = 1 - \eta_{ij} - \beta_{ij} \tag{3.31}$$

where η_{ij} and β_{ij} satisfy $\eta_{ij} + \beta_{ij} \leq 1$.

We assume that $n < p$ because the number of classes, a kind of prior knowledge, can be considered as the lower bound of the number of clusters which is an unknown variable in real situation. Hence, a cluster usually contains less pixels in comparison to a class. To measure the likeness between ω_j and θ_i, we consider the percentage of pixels in θ_i is labeled as ω_j. We thus use the ratio of pixels in intersection of θ_i labeled as ω_j to pixels in θ_i as the likeness δ_{ij}. For all pixels x, we have:

$$\delta_{ij} = \frac{|\{x : x \in \omega_j\} \cap \{x : x \in \theta_i\}|}{|\{x : x \in \theta_i\}|}, \tag{3.32}$$

For η_{ij} and β_{ij}, we have:

$$\eta_{ij} = \alpha\delta_{ij} \tag{3.33}$$

$$\beta_{ij} = \gamma(1 - \delta_{ij}) \tag{3.34}$$

in which α represents a discounting coefficient ranging from 0 to 1 to measure the reliability of classifiers.

To assign label of classes to clusters, we consider the following linear optimization problem:

$$\max_f(\sum_{i,j} e_{ij}f_{ij}) \tag{3.35}$$

subject to

$$\sum_{j=1}^{p} f_{ij} \geq 1 \quad \forall i \in \{1, ..., n\} \tag{3.36}$$

$$\sum_{i=1}^{n} f_{ij} = 1 \quad \forall j \in \{1, ..., p\} \tag{3.37}$$

$$f_{ij} \in \{0, 1\} \quad \forall i \in \{1, ..., n\}, \forall j \in \{1, ..., p\} \tag{3.38}$$

with

$$e_{ij} = ln\frac{1 - \beta_{ij}}{1 - \eta_{ij}} \tag{3.39}$$

The most plausible relation f^*, which indicates assignments between classes and clusters, can thus be found by solving the above optimization problem.

Unlike classical object association problem where each object in one set has at most one counterpart in the other set, in our problem, each class ω_j can be assigned to multiple clusters while each cluster θ_i is merely matched with one class as we always have $n < p$.

Generation of a land cover map for a large area

The proposed framework can produce a semantic land cover map on a large area merely by deploying a supervised approach on a small slice of the validation region. We select a slice of an area containing all types of land covers to employ a supervised learning approach. On the contrary, unsupervised methods are applied to the whole area. The principle behind the proposed framework is to assign the most plausible label to each

cluster in the same area where semantic labels are offered by the supervised approach; later we can enlarge these labeled clusters to the whole area.

The main steps of the proposed framework are shown in Figure 3.17, in which the region marked with orange rectangular denoted as A is the selected slice for conducting a supervised learning approach. The set of classes produced by the supervised approach deployed on the region A is denoted Ω^A. The set of clusters in the region A denoted as Θ^A. The object association is firstly applied between Ω^A and Θ^A. Therefore clusters in Θ^A can be all assigned by a semantic label from the supervised approach. The set of clusters from the unsupervised method on a large area including area A is denoted as Θ^L. The numbers of clusters of Θ^A and Θ^L are the same. To enlarge semantic labels to the whole area represented by the set of clusters Θ^L, we assign each cluster $\theta_i \in \Theta^L$ with the same label as its largest subset $\theta_*^A \in \Theta^A$.

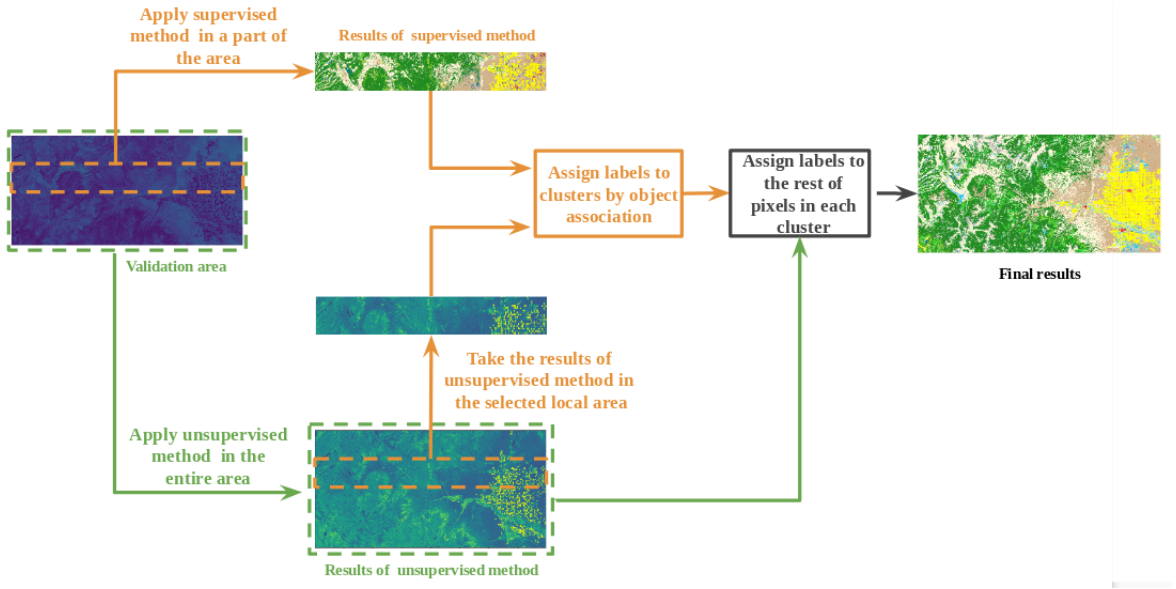


Figure 3.17. The proposed framework based on object association.

3.5.3 Experiments and results

In this section, we test the proposed framework with different combinations of supervised and unsupervised methods. To verify the possibility to apply the proposed framework in a large area, we only carry out the supervised method on a quarter of the validation area. Based on the labels it generates, a semantic land cover map of the entire validation area can be produced by the proposed framework, as displayed in Figure 3.17. We also apply the supervised method on the entire validation area to compared to the results from the proposed framework. For supervised methods, we apply the most commonly used classifiers: Support Vector Machine (SVM) and Maximum likelihood Classification (MLC). For unsupervised methods, we use the Iterative Self-Organizing Data Analysis Techniques (ISODATA), which are also widely applied in remote sensing applications [97].

Apart from the experiments on different combination of methods, we also evaluate the

proposed framework with different sizes of training data and made the comparison with supervised methods. We repeat all the experiments in our study 10 times to take their average results.

Combination of MLC and ISODATA

We start with the combination of MLC and ISODATA in the proposed framework with the number of clusters ranging from 20 to 500. To mimic the situation where training samples are severely insufficient, we first select a very small amount of training samples to train MLC. Table 3.7 summarizes the comparison between the proposed framework and MLC on the validation area with training samples accounting for 0.2% in the training area. The overall accuracy and the accuracy of each land cover category in Table 3.7 are measured by confusion matrix. We use NLCD classification code mentioned in Table 3.6 to represent each land cover category and marked the best accuracy in bold. The overall accuracy of the proposed framework improves as the number of clusters k increasing and the best performance is reached when $k = 100$.

We display the results of MLC in Figure 3.18(a) to compare to the semantic land cover map generated through the combination of MLC and ISODATA, as shown in Figure 3.18(b). MLC misclassified a part of the shrub as medium intensity developed area. However, after the combination with ISODATA in the proposed framework, the accuracy of the shrub is enhanced to 0.684%. This improvement is presented by the reduction of the red area in Figure 3.18.

Besides shrub, other terrestrial vegetation such as deciduous forest, evergreen forest, and grassland also have good improvements in classification accuracy after the combination. However, some land cover categories such as developed areas, barren land, and emergent herbaceous wetlands, show unsatisfying classification accuracy and thus are difficult to enhance in the proposed framework.

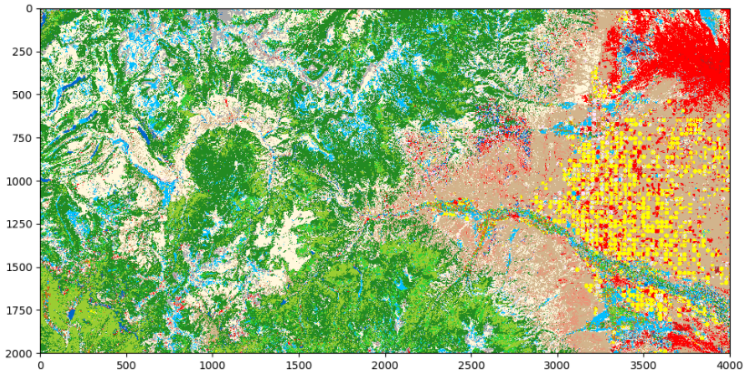
We also test the proposed framework and MLC with the different quantities of training samples fixing the number of clusters $k = 100$. We begin by testing on the severely insufficient training samples where only use 0.05% of the training area. Table 3.8 reports the overall accuracy of MLC and the proposed framework with the combination of MLC and ISODATA on the validation area. Figure 3.19 gives a more visible display of the improvement of their accuracies with the increasing of training samples. The proposed framework always performs better than MLC when the percentage of training samples ranging from 0.05% to 100% and the largest improvement of accuracy occurs at 0.2%.

Table 3.7 – Comparison of MLC and MLC + ISODATA in classification accuracy.

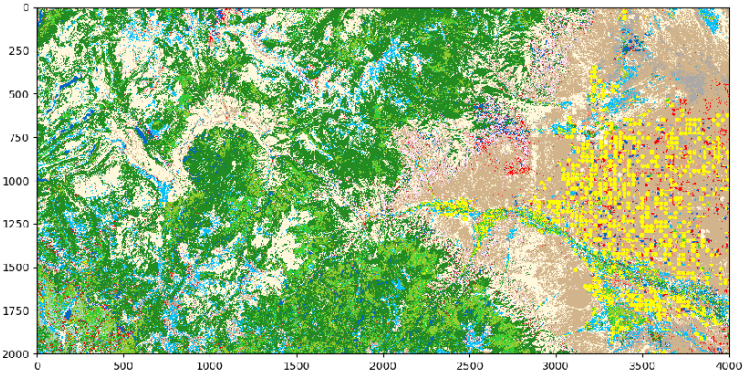
Number of clusters	Accuracy of each land cover category																Overall accuracy
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95	
20	0.316	0.051	**0.078**	0.077	0.133	0.019	0.064	0.212	0.494	0.464	0.352	0.205	0.255	0.142	0.116	0.199	0.325
40	0.590	0.031	0.041	0.057	0.064	0.008	0.034	0.189	0.672	0.574	0.571	0.452	0.257	0.163	0.047	0.235	0.475
60	0.541	0.001	0.026	0.079	0.044	0.034	0.037	0.179	0.694	0.572	0.634	0.451	0.306	0.075	0.039	0.267	0.496
80	0.463	0.043	0.027	0.063	0.025	0.022	0.059	0.198	**0.711**	0.521	**0.684**	0.429	0.308	0.071	0.027	0.294	0.505
100	0.579	0.026	0.019	0.069	0.033	0.032	0.046	0.206	0.698	0.606	0.676	**0.479**	0.313	0.048	0.014	0.339	**0.514**
200	**0.746**	0.006	0.027	0.088	0.098	0.012	0.079	0.164	0.682	0.595	0.659	0.453	0.319	0.065	0.040	0.347	0.501
300	0.697	0.019	0.020	0.088	0.077	0.022	0.0941	0.232	0.684	0.580	0.672	0.441	0.318	0.057	0.046	0.363	0.504
400	0.689	0.005	0.021	0.092	0.115	0.0115	0.084	0.219	0.678	0.600	0.657	0.441	**0.324**	0.091	0.047	0.352	0.500
500	0.739	0.012	0.035	**0.110**	0.106	0.057	0.101	**0.239**	0.664	0.551	0.624	0.469	0.307	0.119	0.062	**0.367**	0.497
MLC	0.771	**0.064**	0.065	0.109	**0.375**	**0.294**	**0.591**	0.231	0.572	**0.764**	0.449	0.377	0.278	**0.262**	**0.180**	0.327	0.432

Table 3.8 – Overall accuracy of MLC and MLC+ISODATA with different sizes of training samples.

Percentage of the training samples	0.05%	0.1%	0.2%	0.4%	0.8%	1.2%	1.6%	2.0%	2.4%	5.0%	10.0%	20.0%	40.0%	80.0%	100.0%
MLC	0.418	0.437	0.446	0.447	0.451	0.453	0.456	0.457	0.456	0.460	0.470	0.484	0.499	0.528	0.539
MLC+ISODATA	0.475	0.504	0.513	0.510	0.505	0.508	0.507	0.509	0.509	0.511	0.514	0.518	0.529	0.554	0.562



(a) Classification results of MLC.



(b) Classification results of MLC+ISODADA.

Figure 3.18. Comparison of MLC and MLC+ISODADA.

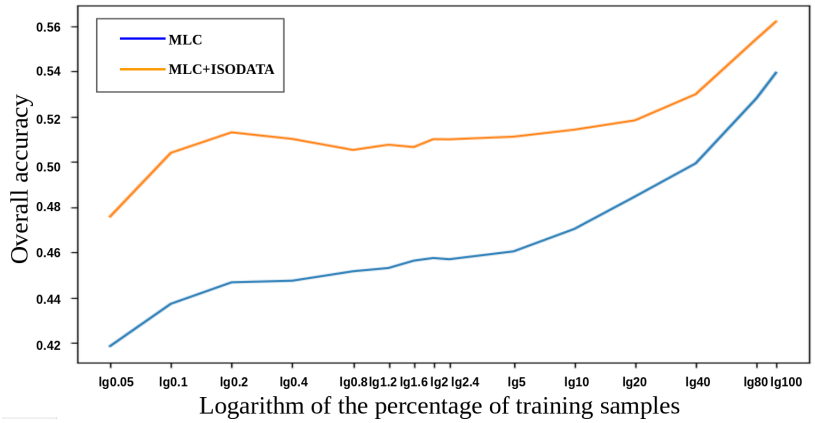


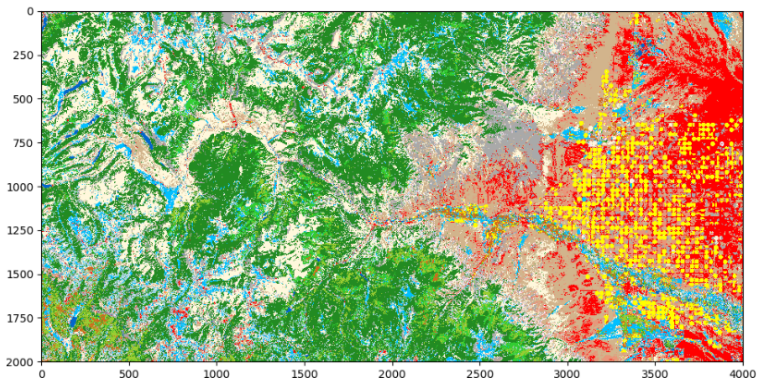
Figure 3.19. Accuracy with different sizes of training samples.

Combination of SVM and ISODATA

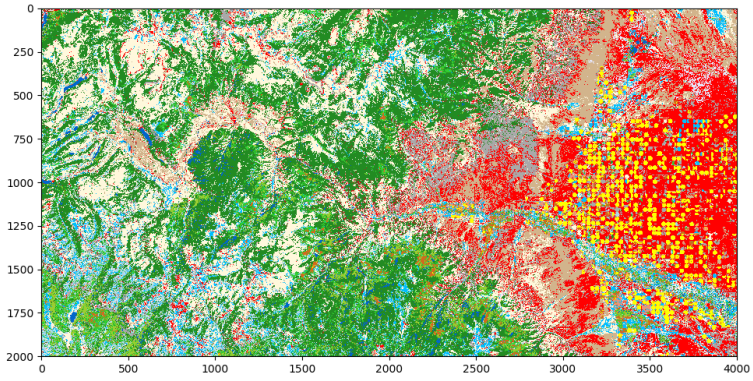
A combination of Support Vector Machine (SVM) and ISODATA was also considered in the proposed framework. In our study, we selected Radial Basis Function (RBF) kernel as it is proved to have better results for LandSat data with 7 bands [67]. The error weight and Gamma value in SVM were set separately to 100 and 0.167 as is proposed in [155].

Table 3.9 reports the overall accuracy as well as the accuracy of each land cover category when training samples account for 0.2%. The performance of SVM on the entire validation area is also displayed at the bottom of the Table 3.9 for comparison. The number of clusters k ranged from 20 to 500 with the best performance of each land cover category marked in bold. For this combination, the best overall accuracy is achieved when k is 60 and the correspondent results are shown in Figure 3.20(a) and Figure 3.20(b).

Compared to MLC, SVM is less competitive to distinguish shrub and medium intensity developed areas when training data is severely insufficient. The combination with ISODATA shown a good improvement in the accuracy of shrub and grassland. However, the improvement of overall accuracy from 0.441 to 0.467 is less apparent in comparison to the combination of MLC and ISODATA.



(a) Classification results of SVM.



(b) Classification results of SVM+ISODADA.

Figure 3.20. Comparison of SVM and SVM+ISODADA.

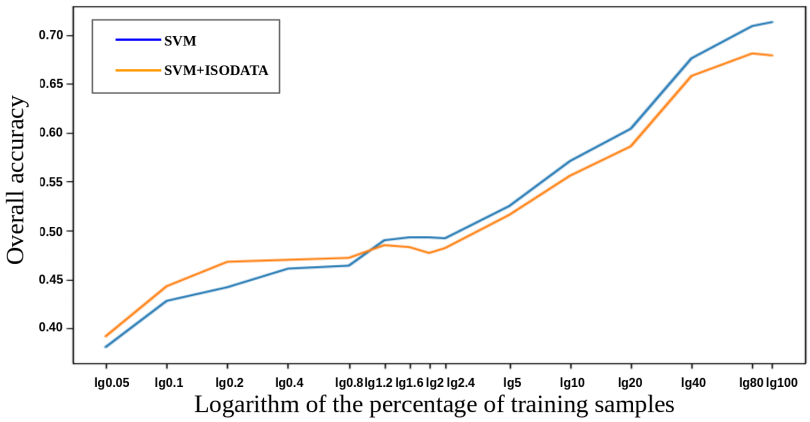


Figure 3.21. Accuracy with different sizes of training samples.

Another experiment to evaluate the performance of the proposed framework with different sizes of training data was also conducted. The number of clusters k in ISODATA is fixed at 60 at which this combination achieves the best performance in the previous experiments. The results are summarized in Table 3.10 and Figure 3.21. The overall accuracy is improved when the percentage of training data is from 0.05% to 0.8%, which is less satisfying in comparison to the combination of MLC with ISODATA.

We apply two combinations in the proposed framework: MLC with ISODATA, and SVM with ISODATA. For the experiments with the different number of clusters k, we all started when k is 20, which is approximately equal to the number of classes set by MLC and SVM. However, the two combinations are both less accurate than MLC or SVM in this situation because of the limitation of unsupervised methods. As a clustering method, ISODATA separates data merely based on their spectral dissimilarities in the seven optical bands used in the study. In the complex natural environment, however, different types of land cover may have similar spectral properties and the same type may reflect differently in spectral bands due to varieties of temperature and illumination. A direct application of ISODATA on our study data is thus less satisfying. The proposed framework based on object association achieves the combination in the decision level. It can be considered as an average result of supervised and unsupervised methods when the number of clusters is approximate to the number of classes. Therefore, the combination of MLC and ISODATA or SVM and ISODATA is less accurate in comparison to the original supervised approach.

However, when the number of clusters is largely superior to the number of classes, the overall accuracy is gradually promoted with an increment of k. For the combination of MLC and ISODATA, the overall accuracy achieves the optimal value at $k = 100$ and gradually converges when k is closing to 500. The combination of SVM and ISODATA reaches its best performance at $k = 60$. However, the increment of k after 60 doesn't bring more improvement in the accuracy.

A relatively large value of k in ISODATA can separate original data into multiple homogeneous clusters which thus can be considered as over-segmentation or superpixels in spectral space. In this situation, pixels in each pure cluster are gathered as an entire object whose label is associated with the proposed framework. With the increment of k, the combination of SVM and ISODATA is less efficient than MLC and ISODATA probably

Table 3.9 – Comparison of SVM and SVM + ISODATA in classification accuracy.

Number of clusters	Accuracy of each land cover category																Overall accuracy
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95	
20	0.010	0.070	**0.202**	0.104	0.245	0.001	0.219	0.285	0.569	0.436	0.193	0.126	0.263	0.161	**0.202**	0.232	0.321
40	0.014	0.011	0.056	0.046	0.562	0.007	0.274	0.243	0.661	0.270	0.126	**0.584**	0.268	0.149	0.112	0.084	0.445
60	0.737	0.032	0.006	0.097	0.437	0.001	0.320	**0.411**	0.629	0.003	**0.503**	0.424	0.344	0.029	0.045	0.214	**0.467**
80	0.756	**0.187**	0.026	0.092	0.491	0.006	0.351	0.406	0.668	0.373	0.335	0.427	0.278	0.073	0.058	0.328	0.457
100	**0.772**	0.027	0.041	0.124	0.565	0.008	0.383	0.336	0.670	0.227	0.223	0.464	0.312	0.053	0.092	0.287	0.450
200	0.739	0.025	0.015	0.153	0.545	0.001	0.339	0.342	0.665	0.262	0.214	0.447	**0.347**	0.014	0.149	0.299	0.448
300	0.739	0.008	0.015	0.153	0.545	0.001	0.339	0.342	0.665	0.262	0.214	0.447	0.346	0.014	0.149	0.299	0.443
400	0.681	0.010	0.033	0.169	0.649	0.002	0.368	0.331	**0.684**	0.401	0.232	0.422	0.331	0.113	0.125	0.312	0.450
500	0.716	0.014	0.013	**0.207**	0.541	0.001	0.410	0.318	0.668	0.404	0.258	0.423	0.338	0.094	0.146	0.335	0.451
SVM	0.765	0.058	0.057	0.202	**0.686**	**0.106**	**0.503**	0.313	0.683	**0.522**	0.202	0.439	0.321	**0.209**	0.158	**0.359**	0.441

Table 3.10 – Overall accuracy of SVM and SVM+ISODATA with different sizes of training samples.

Percentage of the training samples	0.05%	0.1%	0.2%	0.4%	0.8%	1.2%	1.6%	2.0%	2.4%	5.0%	10.0%	20.0%	40.0%	80.0%	100.0%
SVM	0.381	0.428	0.442	0.461	0.464	0.490	0.493	0.493	0.492	0.525	0.571	0.604	0.676	0.709	0.713
SVM+ISODATA	0.392	0.423	0.468	0.470	0.472	0.485	0.483	0.477	0.482	0.516	0.556	0.586	0.658	0.681	0.679

because the spectral properties of land covers in our study area are more distinguishable for MLC than SVM when training data is insufficient.

With different sizes of training data, the combination of MLC and ISODATA always brings better performances than MLC alone and this improvement of accuracy is most satisfying when the training samples account for 0.2% in the training area. However, for SVM and ISODATA, the improvement of overall accuracy only occurs when training samples are from 0.1% to 0.8 possibly because SVM cannot find appropriate hyperplanes for classification through such limited training data. Therefore, this combination is less effective in comparison to the first one in our study area.

The uncertainty and imprecision are measured through mass functions in the object association. We applied the ratio of pixels in the intersection of a class s_i in a cluster c_j to pixels in c_j as the similarity between a class and a cluster. The uncertainty thus reduces when the number of clusters increases, which explains why results become better with large k. However, when k becomes too large, the results are less competitive due to the loss of information in unsupervised methods.

3.6 Conclusion

In this chapter, we have introduced our major contributions to land cover classification methods. We have managed to handle the difficulties caused by limited training samples or the situation where no ground truth is available. The two automatic methods presented in sections 3.2 and 3.3 focus on the spectral properties and information on land covers signified by different spectral indexes. The methods only based on spectral indexes without labeled data are often less accurate. We thus apply multiple methods and combine their results by belief functions. The automatic water detection presented in section 3.2 combines information in a sequential process. On the contrary, the automatic land cover identification presented in section 3.3 achieves the combination in a parallel way. Based on belief functions, both two methods well handle different classification results and show a satisfying performance.

It is important to notice that sequential combinations are generally less cautious than parallel combinations. If the sources are not reliable enough, sequential combinations cannot handle the mistakes that occurred in the first place, and mistakes are possible to accumulate during the combination process. In the application of land cover classification, water usually has the most evident spectral properties and is hardly confused with other land covers. On the contrary, water can affect other land covers and is often removed before the land cover classification. Therefore, we use a sequential combination process based on belief functions to identify water, yet apply a parallel fusion for land cover classification.

In section 3.5, we manage to combine supervised and unsupervised methods by object association using belief functions. With limited training samples, this combination can improve the accuracy of classification. The principle of this framework is to associate each cluster with the most similar class so that the semantic label of the class is assigned to the cluster. Major limitations of this framework include the difficulties to combine several supervised and unsupervised methods, and to take into account sufficient information

from supervised and unsupervised results themselves. To preserve information from the original sources, we propose a new combination framework that is presented in the next two chapters.

Chapter 4

Evidential fusion of supervised and clustering methods (EFSC)

Resume

In real-life applications, heterogeneous information at output level is difficult to combine, as explained in Chapter 1. In this chapter, we propose a solution: an evidential fusion framework, to combine heterogeneous information at the output level. We first propose a transformation method that transfers both uncertainty and imprecision in a BBA from its original frame of discernment to any other target frame. Based on this

transformation, multiple supervised and unsupervised classification results can be combined by an iterative fusion process.

4.1 Introduction

In real-life applications, supervised and clustering results are often available. However, raw data, such as remote sensing data, used as inputs for supervised/clustering methods is sometimes inaccessible due to confidentiality and privacy constrains of corporations. This makes supervised ensemble methods arduous to work. Moreover, even though raw data is accessible, limited labeled samples can also seriously affect supervised methods. As explained in Chapter 1, in the METIS project, land cover maps from multiple sensors is usually heterogeneous and maybe sometimes conflict, and can be seriously affected by quality of labeled samples. We model this real problem as the combination of multiple supervised and unsupervised classification results. This is because semantic labels from multiple sensors can be sometimes conflicting or meaningless to each other, and thus some semantic labels can be removed and only the ways to group pixels are taken into account.

In this chapter, we focus on the combination of multiple classification and clustering results at the output level based on belief functions for three purposes: (1) to improve the accuracy of classification when raw data is inaccessible or training samples are highly limited; (2) to reduce uncertain and imprecise information in the supervised results; and (3) to study how the predictions from classification and clustering results affect the combination at the output level.

We propose a transformation method to transfer information to any target frame of discernment, therefore making the combination of multiple classification and clustering results possible at the output level. Another benefit of this transformation is that the method can provide an interface for user-defined land use map. The target frame where multiple information is combined can be defined by users, and different land cover maps generated by sensors can be transferred to the user-defined frame of land use map, which are later explained in chapter 6. In a nutshell, the proposed transformation method can provide a theoretical foundation to solve the real-world difficulties proposed in sections 6.2.4 and 1.2.3.

Based on the transformation method, we propose an evidential fusion method to combine supervised and clustering results (EFSC) at the output level. The proposed EFSC can effectively combine information from clustering with classification and thus decrease the dependence on raw data and labeled samples. The EFSC method attempts to deal directly with uncertainty and imprecision at the output level of supervised results by combining several unsupervised results. In this way, the most reliable information from different sources can be extracted and then be combined to improve the overall accuracy. The iterative fusion strategy in EFSC can help to take as much information as possible from clustering results and thus reduce the reply on raw data information. This framework can solve the difficulty explained in section 1.2.1, caused by limited labeled samples.

The rest of the chapter is organized as follows: In section 4.2, we propose a transformation method to transfer heterogeneous information into the same frame. In section 4.3, we propose an iterative fusion process to retain the most trustful information when combining multiple supervised and unsupervised predictions. A numerical example of EFSC is presented in section 4.4. In section 4.5, we compare the proposed method in framework of belief functions with current semi-ensemble methods on synthetic data. In section 4.6, we give the conclusion of this chapter.

4.2 Transformation of heterogeneous information

In this section, we present the proposed transformation method in the framework of belief functions. This method can transform both uncertainty and imprecision in a BBA from its original frame of discernment to any other target frame, based on the similarity at the output level. Section 4.2.1 explains how we measure and model the similarity information between classification or clustering results in different frames of discernment. We display the proposed transformation for heterogeneous information in section 4.2.2.

4.2.1 Likeness/similarity measures

To transfer a BBA in one discernment to another one, it is critical to synthesize some additional information, otherwise, the transferred BBA becomes vacuous BBA. As we consider this problem in the context of the combination of supervised and unsupervised classification results, we take into account of likeness/similarity between classes and clusters as additional information during the transformation. Many measures have been proposed to measure the likeness [50] or similarity [69], [38] between two finite sample sets, which thus can be used for a class and a cluster. The proposed transformation can use different likeness/similarity measures according to users' needs. In the following, we present some widely used likeness/similarity measures, including similarity measures and evaluate them for our research.

Jaccard similarity

Jaccard similarity index [69] originally is used to measure the similarity between two finite sample sets. In land cover classification problems, a class and a cluster can be considered as two different finite sets which may contain some common pixels. Therefore, we can use Jaccard similarity index to measure the similarity between a class and a cluster. The more pixels they have in common, the more similar they are. For a group of objects X, suppose we have two methods c and s such that c can separate X in the frame of discernment Θ and s in Ω. In the context of belief functions, Jaccard similarity index is measured between 2^{Ω} and 2^{Θ}, defined as:

$$Jac(T_i, O_j) = \frac{|\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in O_j\} \cup \{x : x \in T_i\}|}, O_j \subseteq \Omega, T_i \subseteq \Theta, x \in X \tag{4.1}$$

Proportion likeness

Proportion likeness is used in the transformation proposed by Karem [50]. Given a group of objects X, Ω and Θ are frames of discernment from different results. In the context of belief functions, Proportion likeness is measured between 2^{Ω} and 2^{Θ}, defined as:

$$Pro(T_i, O_j) = \frac{|\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in T_i\}|}, O_j \subseteq \Omega, T_i \subseteq \Theta, x \in X \tag{4.2}$$

Dice similarity

Dice similarity coefficient, also known as the Sørensen–Dice index or simply Dice coefficient, is also a widely-used similarity measure between two sets of data [38]. Given a group of objects X, Ω and Θ are frames of discernment from different results. In the context of belief functions, Dice similarity is measured between 2^{Ω} and 2^{Θ}, defined as:

$$Dice(T_i, O_j) = \frac{2 \times |\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in O_j\}| + |\{x : x \in T_i\}|}, O_j \subseteq \Omega, T_i \subseteq \Theta, x \in X \tag{4.3}$$

Recovery similarity

The recovery similarity [150] is based on the relationship between the size of the intersection and the size of class and cluster themselves. Given a group of objects X, Ω and Θ are frames of discernment from different results. In the context of belief functions, Recovery similarity is measured between 2^{Ω} and 2^{Θ}, defined as:

$$Rec(T_i, O_j) = \frac{|\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in T_i\}|} \times \frac{|\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in O_j\}|} \tag{4.4}$$

Measures such as Jaccard, Proportion, Dice, and Recovery only apply information at the output level. In other words, these measures only require supervised/clustering results to measure the likeness/similarity, whereas do not need to access the raw data which is used to achieve supervised/clustering methods. Therefore, this sort of measures is more suitable for the combination when only supervised/clustering results are available.

BBAs of likeness/similarity information

To synthesis the likeness/similarity information during transformation, we construct BBAs to represent the similarity information. For two supervised/clustering results in frames of discernment Ω and Θ, the likeness/similarity can be measured for all $O_j \subseteq \Omega$, $T_i \subseteq \Theta$. Θ is the original frame of discernment where BBAs are constructed and Ω is the target frame of discernment where BBAs are supposed to transfer.

We use $Sim(T_i, O_j)$ to represent the likeness/similarity between cluster T_i and class O_j. $Sim(T_i, O_j)$ gives uncertainty on the information that T_i and O_j are the same and nothing more, which can be modeled by a SBBA $m_{ij}^{\mathcal{H}_{ij}}$ in the frame of discernment $\mathcal{H}_{ij} = \{\mathcal{Y}_{ij}, \mathcal{N}_{ij}\}$. $\mathcal{Y}_{ij}$ represents T_i and O_j are the same while $\mathcal{N}_{ij}$ indicates they are different. The likeness/similarity information between O_j and T_i can be modeled as a SBBA by:

$$\begin{cases} m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) = Sim(T_i, O_j), \\ m_{ij}^{\mathcal{H}_{ij}}(\mathcal{H}_{ij}) = 1 - Sim(T_i, O_j). \end{cases} \tag{4.5}$$

where $m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij})$ indicates the support on the information that O_j and T_i are the same, and $m_{ij}^{\mathcal{H}_{ij}}(\mathcal{H}_{ij})$ represents we know nothing about the relationship between O_j and T_i.

The likeness/similarity can be measured between any subset O_j of Ω and T_i of Θ, and each $Sim(T_i, O_j)$ can be used to construct a SBBA by equation (4.5). Therefore, we can construct a matrix $\mathscr{M}$ with the size $2^{\Theta} \times 2^{\Omega}$, to represent the information given by SBBA

$m_{ij}^{\mathcal{H}_{ij}}$, with $i = 1, ..., 2^{\Theta}, j = 1, ..., 2^{\Omega}$. Each element of this matrix is the focal element besides ignorance of the BBA $m_{ij}^{\mathcal{H}_{ij}}$. Therefore, for all $i = 1, ..., 2^{\Theta}, j = 1, ..., 2^{\Omega}$, $\mathscr{M}$ can be written as:

$$\mathscr{M} = \begin{bmatrix} m_{11}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{1j}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{12^{\Omega}}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{i1}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{i2^{\Omega}}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{2^{\Theta}1}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{2^{\Theta}j}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) & \cdots & m_{2^{\Theta}2^{\Omega}}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) \end{bmatrix} \tag{4.6}$$

In the matrix $\mathscr{M}$, a non-zero element $m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij})$ represents the uncertainty that T_i and O_j are the same. A zero element $m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij})$ represents that we know nothing about the relationship between T_i and O_j.

4.2.2 The proposed transformation of BBA

The information provided by a method, noted as c, in the frame of discernment Θ can be modeled by a BBA m_c^{Θ}. To combine it with BBAs in the frame of discernment Ω, the crucial step is to transfer m_c^{Θ} as m_c^{Ω}. The transformation proposed in [50] takes only uncertainty on singletons from clustering, yet ignoring imprecision. To fully exploit the information in clustering, we decompose the original BBA m_c^{Θ} into a set of SBBAs, so that both uncertainty and imprecision can be well preserved during the transformation. Therefore, the transformation requires the original BBA m_c^{Θ} is non-dogmatic so that can be decomposed. If m_c^{Θ} is dogmatic, we can use a discounting coefficient to change it to a non-dogmatic BBA. The proposed transformation consists of the following four steps:

Step 1: Decompose a BBA on the original frame of discernment to SBBAs

Suppose m_c^{Θ} is a separable mass which accordingly can be decomposed as several SBBAs as:

$$m_c^{\Theta} = \bigoplus_{\emptyset \neq T_i \subseteq \Theta} m_{c,T_i}^{\Theta} \tag{4.7}$$

where $m_{c,\ T_i}^{\Theta}$ indicates the SBBA from the method c and its focal element is T_i. If $T_i \subset \Theta$, we have:

$$\begin{cases} m_{c,\ T_i}^{\Theta}(T_i) = t, \quad t \in [0, 1], \\ m_{c,\ T_i}^{\Theta}(\Theta) = 1 - t \end{cases} \tag{4.8}$$

otherwise when $T_i = \Theta$, $m_{c,\ T_i}^{\Theta}$ is a vacuous BBA:

$$m_{c,\ T_i}^{\Theta}(\Theta) = 1 \tag{4.9}$$

The SBBAs after decomposition can be considered as multiple independent sources that entirely preserve the information from m_c^{Θ}.

Step 2: Transfer a SBBA on the original frame of discernment to the target frame of discernment

We denote the transferred SBBA on T_i to O_j as $m^{\Omega}_{T_i,O_j}$. It indicates the uncertainty about the information that object x clustered as T_i should be labeled as O_j, incorporating the similarity between T_i and O_j.

To calculate $m^{\Omega}_{T_i,O_j}$ in Ω, we can use the uncertainty on T_i or Θ as a discounting coefficient to modify the likeness/similarity BBA $m^{\mathcal{H}_{ij}}_{ij}$. There are two strategies for performing this simple transformation: <1> using the uncertainty on T_i as a discounting coefficient to weaken $m^{\mathcal{H}_{ij}}_{ij}(\mathcal{Y}_{ij})$; <2> using the ignorance on Θ as a discounting coefficient to weaken $m^{\mathcal{H}_{ij}}_{ij}(\mathcal{H}_{ij})$.

In strategy <1>, $\forall T_i \subseteq \Theta$, if $O_j \subset \Omega$, we have:

$$\begin{aligned} m^{1,\Omega}_{T_i,O_j}(O_j) &= m^{\mathcal{H}_{ij}}_{ij}(\mathcal{Y}_{ij}) * m^{\Theta}_{c,T_i}(T_i) \\ &= Sim(T_i, O_j) * t \end{aligned} \tag{4.10}$$

$$\begin{aligned} m^{1,\Omega}_{T_i,O_j}(\Omega) &= 1 - m^{\mathcal{H}_{ij}}_{ij}(\mathcal{Y}_{ij}) * m^{\Theta}_{c,T_i}(T_i) \\ &= 1 - t * Sim(T_i, O_j) \end{aligned} \tag{4.11}$$

otherwise we have a vacuous BBA:

$$m^{1,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{4.12}$$

In strategy <2>, $\forall T_i \subseteq \Theta$, if $O_j \subset \Omega$, we have:

$$\begin{aligned} m^{2,\Omega}_{T_i,O_j}(O_j) &= 1 - m^{\mathcal{H}_{ij}}_{ij}(\mathcal{H}_{ij}) * m^{\Theta}_{c,T_i}(\Theta) \\ &= 1 - (1-t) * (1 - Sim(T_i, O_j)) \\ &= t + Sim(T_i, O_j) - t * Sim(T_i, O_j) \end{aligned} \tag{4.13}$$

$$\begin{aligned} m^{2,\Omega}_{T_i,O_j}(\Omega) &= m^{\mathcal{H}_{ij}}_{ij}(\mathcal{H}_{ij}) * m^{\Theta}_{c,T_i}(\Theta) \\ &= (1-t) * (1 - Sim(T_i, O_j)) \\ &= 1 - t - Sim(T_i, O_j) + t * Sim(T_i, O_j) \end{aligned} \tag{4.14}$$

otherwise we have a vacuous BBA:

$$m^{2,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{4.15}$$

According to the *Least Commitment Principle* (LCP), we should choose the less committed transferred SBBA between strategy <1> and strategy <2>.

Lemma 1. For two given non dogmatic SBBAs $m^{1,\Omega}_{T_i,O_j}$ and $m^{2,\Omega}_{T_i,O_j}$, $m^{1,\Omega}_{T_i,O_j}$ is no more committed than $m^{2,\Omega}_{T_i,O_j}$ with w-ordering.

Proof. According to equations (4.10), (4.11) (4.13), (4.14), $\forall t, Sim(T_i, O_j) \in [0, 1]$, we have:

$$m^{1,\Omega}_{T_i,O_j}(O_j) + m^{1,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{4.16}$$

$$m^{2,\Omega}_{T_i,O_j}(O_j) + m^{2,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{4.17}$$

For all $O_j \subseteq \Omega$, we always have:

$$\begin{aligned}
2 &\leq \frac{1}{t} + \frac{1}{Sim(T_i, O_j)} \\
&\Leftrightarrow 2t * Sim(T_i, O_j) \leq t + Sim(T_i, O_j) \\
&\Leftrightarrow t * Sim(T_i, O_j) \leq t + Sim(T_i, O_j) - t * Sim(T_i, O_j) \\
&\Leftrightarrow m^{1,\Omega}_{T_i,O_j}(O_j) \leq m^{2,\Omega}_{T_i,O_j}(O_j) \\
&\Leftrightarrow -ln(m^{1,\Omega}_{T_i,O_j}(\Omega)) \geq -ln(m^{2,\Omega}_{T_i,O_j}(\Omega)) \\
&\Leftrightarrow w^{1,\Omega}_{T_i,O_j} \geq w^{2,\Omega}_{T_i,O_j} \\
&\Leftrightarrow m^{2,\Omega}_{T_i,O_j} \sqsubseteq_w m^{1,\Omega}_{T_i,O_j}
\end{aligned} \tag{4.18}$$

$m^{1,\Omega}_{T_i,O_j}$ is no more committed than $m^{2,\Omega}_{T_i,O_j}$ and should therefore be selected, simplified as $m^{\Omega}_{T_i,O_j}$.

□

Step 3: Combine all the evidence from transferred SBBAs

The transformed SBBA $m^{\Omega}_{T_i,O_j}$ represents only one piece of evidence on object x in O_j. From the perspective of clustering, x can also be clustered into different singletons or unions in Θ. The imprecision should also be considered when reconstructing the supports for the assertion that x is in O_j. Since all SBBAs transferred from T_i ($\forall T_i \subseteq \Theta$) to O_j are independent, we can combine them by Dempster's rule:

$$m^{\Omega}_{c,O_j} = \bigoplus_{\forall T_i \subseteq \Theta} m^{\Omega}_{T_i,O_j}, \tag{4.19}$$

The SBBA m^{Ω}_{c,O_j} on Ω represents the evidence that object x is labeled as O_j knowing all the evidence, provided by the method c on Θ, that x belongs to each T_i ($\forall T_i \subseteq \Theta$).

Step 4: Combine SBBAs on the target frame of discernment

Since m^{Ω}_{c,O_j} represents only a piece of evidence on the focal element O_j, we have to combine all the evidence on Ω to obtain a normal BBA as:

$$m^{\Omega}_{c} = \bigoplus_{\forall O_j \subseteq \Omega} m^{\Omega}_{c,O_j}, \tag{4.20}$$

m^{Ω}_{c} indicates a normal BBA on Ω with the results of the method c which are initially generated on Θ.

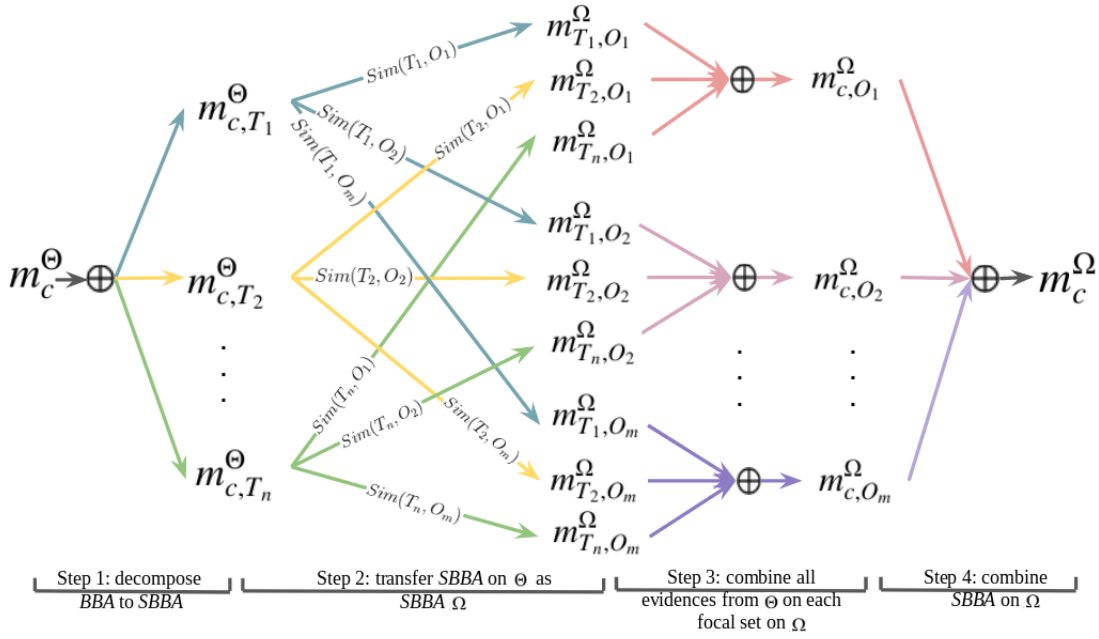


Figure 4.1. Transformation method to change the BBA in the frame of discernment Θ to Ω.

The Figure 4.1 displays the process of the proposed transformation method. m_c^Θ is first decomposed into a set of SBBAs, *e.g.* $m_{c,T_1}^\Theta, ..., m_{c,T_n}^\Theta$. For each SBBA such as m_{c,T_1}^Θ, it can be transferred as a group of SBBAs on Ω using likeness/similarity information. By doing so, each focal element on Ω contains some evidence of Θ, which can be combined by Dempster's rule. To obtain a normal BBA, the SBBA on Ω can also be combined by Dempster's rule. The proposed transformation is more cautious because both uncertainty and imprecision on Θ are preserved and transferred into Ω, whereas the transformation in [50] takes only into account uncertainty.

The transformation is initially proposed for solving the problem that BBAs of supervised/clustering results are in different frames of discernment. In this context, the measures of likeness/similarity are based on the results of classification/clustering, and thus we use several methods (*i.e.* Jaccard, Proportion, Dice, Recovery) that are originally developed for the likeness/similarity between two finite sample sets to generate the matrix $\mathscr{M}$. The essence of the proposed transformation is to transfer a BBA from its original frame Θ of discernment to any-other target frame Ω, with some additional information represented by the matrix $\mathscr{M}$. Each element $m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) \in [0,1]$ in $\mathscr{M}$ indicates the uncertainty that the information related to $T_i \subseteq \Theta$ is the same as the information related to $O_j \subseteq \Omega$.

Lemma 2. For a non-dogmatic BBA m^Θ, when the matrix $\mathscr{M}$ is a zero matrix, the proposed transformation satisfies $m^{\Theta\uparrow\Omega} = m_{[\Omega]}^\Omega$. $m_{[\Omega]}^\Omega$ represents the vacuous BBA on Ω.

Proof: For a non-dogmatic BBA m^{Θ}, according to equation (4.7), we have:

$$m^{\Theta} = \bigoplus_{\emptyset \neq T_i \subseteq \Theta} m^{\Theta}_{T_i} \tag{4.21}$$

If $\mathscr{M}$ is a zero matrix, according to equation (4.5), for all $T_i \subseteq \Theta$, $O_j \subseteq \Omega$, we have:

$$\begin{cases} m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) = 0, \\ m_{ij}^{\mathcal{H}_{ij}}(\mathcal{H}_{ij}) = 1. \end{cases} \tag{4.22}$$

According to equations (4.10), (4.11), (4.12), we have:

$$m^{\Omega}_{T_i,O_j}(O_j) = m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) * m^{\Theta}_{T_i}(T_i) = 0 \tag{4.23}$$

$$m^{\Omega}_{T_i,O_j}(\Omega) = 1 - m_{ij}^{\mathcal{H}_{ij}}(\mathcal{Y}_{ij}) * m^{\Theta}_{T_i}(\Theta) = 1 \tag{4.24}$$

According to equations (4.19), (4.20), the transferred BBA m^{Ω} from m^{Θ} satisfies:

$$\begin{cases} m^{\Omega}(O_j) = 0, \\ m^{\Omega}(\Omega) = 1 \end{cases} \tag{4.25}$$

which can be also written as $m^{\Theta\uparrow\Omega} = m^{\Omega}_{[\Omega]}$.

□

When $\mathscr{M}$ is a zero matrix, it indicates that we know nothing about the relationships between the information on Θ and the information on Ω. Therefore, the transferred BBA of m^{Θ} to the frame of discernment Ω become a vacuous BBA, in accordance with the *Least Commitment Principle*.

4.3 The proposed EFSC method

The proposed transformation allows transferring the heterogeneous information from clustering to the same frame of discernment as classification, based on which combining different classification and clustering methods becomes possible. Clustering with different numbers of clusters affects differently on the fusion with classification. To efficiently synthesize the information derived from clustering, we propose an iterative fusion process that combines several classification and clustering methods.

For a group of objects $X = \{x_1, x_2, ..., x_N\}$ to be classified, suppose we have a group of classification $S = \{s_1, s_2, ..., s_M\}$ and a group of clustering $C = \{c_1, c_2, ..., c_L\}$ with different numbers of clusters. We denote the frame of discernment of a classification as $\Omega = \{\omega_1, \omega_2, ..., \omega_Y\}$ and that of a clustering as $\Theta = \{\theta_1, \theta_2, ..., \theta_Z\}$. To simplify the problem, we can merely measure the likeness/similarity between each class ω_j and each cluster θ_i. Thus the matrix $\mathscr{M}$ can be simplified to the size $|\Theta| \times |\Omega|$.

4.3.1 Selection of partial information in clustering

We propose a criterion to control the iterative process to ensure that only useful information from clustering results is fused. In this case, we are more interested in partial information about clusters rather than in the evaluation of the whole clustering. Therefore, at each iteration, we have to decide whether the information from certain clusters should be fused with the classes.

Let's denote, $\forall \omega_j \in \Omega$, $W_j^k = \{x : x \in \omega_j^k\}$, for all the objects classified as ω_j in iteration k. After combination with a clustering c_l, a new set of objects labeled as ω_j is generated, denoted as $W_{j,c_l} = \{x : x \in \omega_j\}$. This criterion is used to decide whether the information in W_{j,c_l} should be updated to W_j^k. Results updated by the criterion are denoted by W_j^{k+1}. Note that, if the information in W_{j,c_l} is synthesized, W_j^{k+1} becomes the union of W_j^k and W_{j,c_l}. When the uncertainty in W_{j,c_l} is less than that in W_j^k, we update the information in W_j^k by W_{j,c_l}. For each object $x \in W_j^k$, its BBA after decision is transformed to a scalar $R_{W_j^k}$, called by loss of confidence, defined based on Jousselme distance by:

$$R_{W_j^k} = d_J(m(x), m_{[\omega_j]}) \tag{4.26}$$

where $m(x)$ is the BBA of the object x and $m_{[\omega_j]}$ is the categorical BBA such that $m(\omega_j) = 1$. A smaller distance represents less loss of confidence on ω_j for object x. The loss of confidence of class ω_j, denoted $\overline{R_{W_j^k}}$, is the average loss of confidence of all objects x in W_j^k, as follows:

$$\overline{R_{W_j^k}} = \frac{\sum\limits_{x \in W_j^k} R_{W_j^k}}{|W_j^k|} \tag{4.27}$$

We update the information, *i.e.* labels, BBAs and losses of confidence in the iteration k of all objects in each class W_j^k when $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$. In some situations, due to insufficient information on certain classes, no decision may be made on these classes after a combination. We, therefore, stop merging new information with these classes so that they can be retained during the iterative fusion process and has relatively high uncertainty. Details of the criterion are given in the algorithm 1.

Note that due to the cautiousness of this criterion, the loss of confidence only decrease for classes that have less disagreement between classification and clustering. Let's denote $V_{j,c_l} = W_j^k \cap W_{j,c_l}$, $P_{j,c_l} = W_j^k - V_{j,c_l}$, $Q_{j,c_l} = W_{j,c_l} - V_{j,c_l}$. V_{j,c_l} represents a group of objects labeled as ω_j before and after fusion with clustering c_l. P_{j,c_l} indicates a group of objects, originally labeled as ω_j, yet which are not reclassified as ω_j after combining with clustering c_l. Q_{j,c_l} indicates a group of objects, not labeled as ω_j, but reclassified as ω_j after combining with clustering c_l.

For each $x \in P_{j,c_l}$, its information is not changed by W_{j,c_l}. Only for each $x \in V_{j,c_l} \cup Q_{j,c_l}$, its information is modified if the criterion is satisfied. The updated information may result in an increase in the average loss of confidence on W_j^{k+1} compared to W_j^k, even if it satisfies

$\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$, demonstrated in Lemma 3.

Algorithm 1: Criterion for partial information fusion.

Input:

Supervised labels and BBAs on the frame of discernment of objects in a dataset $X = \{x_1, ..., x_N\}$.

Iteration step k.

Output:

Updated labels and BBAs on the dataset $X = \{x_1, ..., x_N\}$.

for $\forall \omega_j \in \Omega$ **do**
  if W_{j,c_l} *is not empty* **then**
    Calculate $\overline{R_{W_j^k}}$, the average loss of confidence on W_j^k in iteration k by equation (4.27).
    Calculate $\overline{R_{W_{j,c_l}}}$, the average loss of confidence on W_{j,c_l} by equation (4.27).
    if $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$ **then**
      Add information in W_{j,c_l} to W_j^k to obtain W_j^{k+1} :
      $m_x^{k+1} \leftarrow m_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
      $L_x^{k+1} \leftarrow L_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
      $R_x^{k+1} \leftarrow R_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
      $m_x^{k+1} \leftarrow m_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
      $L_x^{k+1} \leftarrow L_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
      $R_x^{k+1} \leftarrow R_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
    else
      Still keep the BBAs m^k, labels L^k, and the loss of confidence R^k in iteration $k+1$.
      $m_x^{k+1} \leftarrow m_x^k, \forall x \in W_j^k$
      $L_x^{k+1} \leftarrow L_x^k, \forall x \in W_j^k$
      $R_x^{k+1} \leftarrow R_x^k, \forall x \in W_j^k$
  else
    Still keep the BBAs m^k, labels L^k, and the loss of confidence R^k in iteration $k+1$.
    $m_x^{k+1} \leftarrow m_x^k, \forall x \in W_j^k$
    $L_x^{k+1} \leftarrow L_x^k, \forall x \in W_j^k$
    $R_x^{k+1} \leftarrow R_x^k, \forall x \in W_j^k$

Lemma 3. For each $\omega_j \in \Omega$, the average loss of confidence is increased after the update, when satisfying $\overline{R_{W_j^k}} < \overline{R_{Q_{j,c_l}}}$, *i.e.* $\overline{R_{W_j^k}} < \overline{R_{W_j^{k+1}}}$ iff $\overline{R_{W_j^k}} < \overline{R_{Q_{j,c_l}}}$.

Proof. We have $\overline{R_{W_j^{k+1}}} = \overline{R_{W_j^k \cup W_{j,c_l}}}$ after the update.

$$
\begin{aligned}
&\overline{R_{W_j^k}} < \overline{R_{W_j^{k+1}}} \\
&\quad \Leftrightarrow \overline{R_{W_j^k}} < \overline{R_{W_j^k \cup W_{j,c_l}}} \\
&\quad \Leftrightarrow \frac{|P_j| \cdot \overline{R_{P_j}} + |V_{j,c_l}| \cdot \overline{R_{V_{j,c_l}}}}{|P_j| + |V_{j,c_l}|} < \frac{|P_j| \cdot \overline{R_{P_j}} + |V_{j,c_l}| \cdot \overline{R_{V_{j,c_l}}} + |Q_{j,c_l}| \cdot \overline{R_{Q_{j,c_l}}}}{|P_j| + |V_{j,c_l}| + |Q_{j,c_l}|} \\
&\quad \Leftrightarrow |P_j| \cdot |Q_{j,c_l}| \cdot \overline{R_{P_j}} + |V_{j,c_l}| \cdot |Q_{j,c_l}| \cdot \overline{R_{V_{j,c_l}}} \\
&\quad < |P_j| \cdot |Q_{j,c_l}| \cdot \overline{R_{Q_{j,c_l}}} + |V_{j,c_l}| \cdot |Q_{j,c_l}| \cdot \overline{R_{Q_{j,c_l}}} \\
&\quad \Leftrightarrow \frac{|P_j| \cdot \overline{R_{P_j}} + |V_{j,c_l}| \cdot \overline{R_{V_{j,c_l}}}}{|P_j| + |V_{j,c_l}|} < \overline{R_{Q_{j,c_l}}} \\
&\quad \Leftrightarrow \overline{R_{W_j^k}} < \overline{R_{Q_{j,c_l}}}
\end{aligned}
\tag{4.28}
$$

□

When satisfying $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$, we must have $\overline{R_{V_{j,c_l}}} < \overline{R_{Q_{j,c_l}}}$ so that the average loss of confidence still increase after the update. However, the relationship between $\overline{R_{V_{j,c_l}}}$ and $\overline{R_{Q_{j,c_l}}}$ is not limited. Obviously, if we have $\overline{R_{W_j^k}} > \overline{R_{Q_{j,c_l}}}$, the average loss of confidence of class ω_j is a non-monotonic increasing function and can converge because it has the infinimum.

4.3.2 Iterative fusion process

We repeat the previous step several times to reduce the uncertainty of the supervised results as much as possible until the overall loss of confidence is convergent. At each iteration, the information from clustering results is combined with the previous clustering by the Dempster's rule, thus strengthening the mass values on singletons and reducing the ignorance. As the combination accumulates, the results of the fusion become more and more certain on singletons, so the accuracy can converge for a given pool of clustering methods C. The details of the iterative process are outlined in algorithm 2.

The proposed EFSC fusion strategy includes two principal steps: (1) enchaining the reliable information in each individual supervised method by randomly combining multiple unsupervised methods from C; (2) fusing the reliable information from multiple supervised methods.

After the combination with a clustering method, the updated fusion results are used as the supervised results to calculate the likeness/similarity with a new clustering at the next iteration. In this way, the likeness/similarity can be updated over the iterations and the information in the pool of clustering methods can be fully exploited when the iteration steps are sufficient. In algorithm 2, ε is a user-defined value close to 0 to control the stop condition. The general workflow is detailed in Figure 4.2, where the different initial classifiers are denoted by s_1, s_2,..s_M. For the classifier s_1, for example, we have chosen at random z different clustering methods in C, denoted successively by $c_{s_{11}}$, $c_{s_{12}}$,...$c_{s_{1z}}$, to be combined with s_1 on the basis of the proposed transformation. In

Algorithm 2: Iterative fusion process for one supervised method and multiple unsupervised methods.

Input:

Labels generated from a supervised method c_m on test data $X = \{x_1, ..., x_N\}$: $L_{s_m}(x_1), ...L_{s_m}(x_N)$

Clustering results from a group of unsupervised methods $C = \{c_1, ...c_L\}$ on X: $[Q_{c_1}(x_1), ...Q_{c_1}(x_N)]$, ..., $[Q_{c_L}(x_1), ...Q_{c_L}(x_N)]$

Output:

Labels after combination of the supervised method s_m and multiple clustering methods in C

Loss of confidence based on Jousselme distance on X after fusion: R_{x_1},...,R_{x_N}

1 Calculate the BBAs of the supervised method s_m in the frame of discernment Ω.
2 Begin with the iteration step $k = 0$.
3 Calculate the average loss of confidence $\overline{R^k}$ in step $k = 0$, and initialize the average loss of confidence in step $k = 1$ as $\overline{R^{k+1}} = 0$.
4 **while** $|\overline{R^k} - \overline{R^{k+1}}| > \varepsilon$ **do**
5 Randomly select a clustering c_l in C.
6 Calculate the SBBAs for unsupervised c_l in its original frame of discernment Θ_l.
7 Calculate the matrix JM_{ml} of L_{s_m} and Q_{c_l} by equation (4.1).
8 **for** $\forall x \in X$ **do**
9 Transfer its BBA of c_l in the frame of discernment Θ_l to Ω as the process shown in Figure 4.1.
11 Update the information according to algorithm 1.
12 Update $\overline{R^k}$ and $\overline{R^{k+1}}$.
13 $k \leftarrow k + 1$.

the i_{th} fusion step, new labeled information $L_{s_{1i}}$ (*i.e.* new classification results), can be extracted to combine with the clustering in the next iteration. Reliable information from each individual classification can be extracted as the iterative process converges.

After the transformation, information from clustering can be thus considered as a supplementary source to add information to classification. For a classification method, the combination of clustering by Dempster's rule helps to reinforce the support on the same focal element, which thus increases the corresponding BBA value. A higher BBA on certain focal elements indicates more belief degrees are assigned and the decision on this focal element becomes more certain. Furthermore, a combined BBA after Dempster's combination rule has less uncertainty or imprecision, so that it further approaches the categorical BBA which represents the perfect information with neither uncertainty nor imprecision.

We thus use Jousselme distance between a BBA and the corresponding categorical BBA to represent the loss of confidence. That is to say, the combination reduces uncertainty and imprecision, so that the BBA converge to the perfect information, leading to a smaller Jousselme distance. Therefore, the iterative combination with multiple clustering results by Dempster's rule can gradually reduce uncertainty and imprecision in the former BBA. The criterion based on Jousselme distance to select partial information can control the loss of confidence after combination with each clustering method.

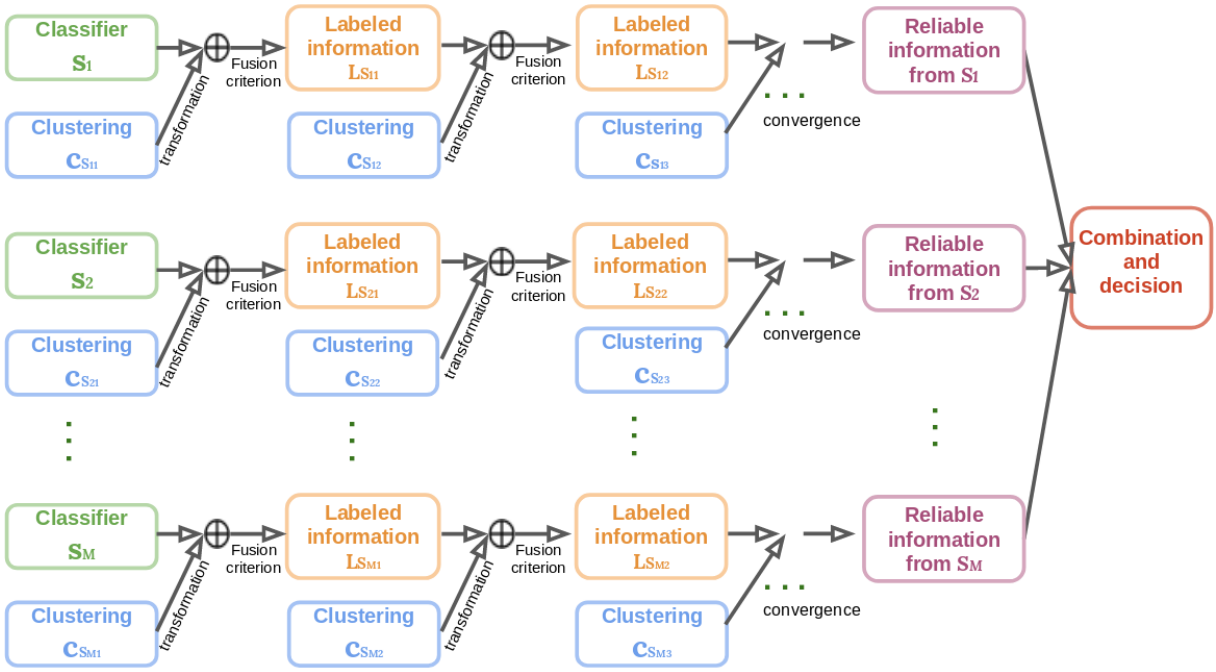


Figure 4.2. Workflow of the proposed EFSC.

4.3.3 Complexity

Dempster's combination rule which is used several times in the proposed EFSC to combine different pieces of information is a P-complete problem [107]. The computational complexity of Dempster's rue can be different according to programming strategies and usually and be solved in polynomial time when only taking accounts into focal elements [147]. The the worst case complexity of Dempster's rule $\mathcal{O}(2^{M|\Omega|})$ where $|\Omega|$ is the size of the frame of discernment Ω and M represents the number of sources to combine [73]. In the following, we discuss the worst computational complexity for a pixel. Meanwhile, the constructions of BBAs of classification or clustering methods are ignored in the calculation of complexity, because these BBAs can be considered as the direct inputs for EFSC framework. We focus on analyzing the complexity generated by Dempster's combination rule.

For the proposed transformation method, the first step to decompose a normal BBA into a group of SBBAs on frame of discernment Θ. The calculation of a SBBA calculated by equations (2.24), (2.25) has the complexity as $\mathcal{O}(2^{|\Theta|})$ and for all subsets except empty set of Θ can be the focal element of a SBBA. Therefore, the first step decomposition has the complexity as $\mathcal{O}(2^{2|\Theta|})$. For the second step to transfer SBBAs on Θ to Ω, the complexity is $\mathcal{O}(2^{|\Theta|+|\Omega|})$ because the number of SBBAs on Θ is 2^{Θ} and each of them has to be transferred to 2^{Ω} SBBAs. Similarly, the third step to combine all SBBAs on Θ on a focal element on Ω is $\mathcal{O}(2^{|\Theta|})$ and we have 2^{Ω} focal elements on Ω, so that the complexity of this step is $\mathcal{O}(2^{|\Theta|+|\Omega|})$. The last step is to combine all SBBAs on Ω to have a normal BBA, which has the complexity as $\mathcal{O}(2^{2|\Omega|})$. Therefore, for the transformation method, the complexity is $\mathcal{O}(2^{2|\Theta|}+2^{2|\Omega|}+2^{|\Theta|+|\Omega|})$. On the contrary, Karem's transformation only discounts a BBA from similarity/likeness by the value of focal element on Θ, so that its complexity is $\mathcal{O}(2^{|\Omega|})$.

When combining one classification and one clustering methods, we use the proposed transformation to a BBA on Θ to Ω and then Dempster's rule to combine two BBAs on Ω, so that the complexity is $\mathcal{O}(2^{2|\Theta|} + 2^{2|\Omega|} + 2^{|\Theta|+|\Omega|} + 2^{2|\Omega|})$ and can be simplified as $\mathcal{O}(2^{2|\Theta|} + 2^{2|\Omega|} + 2^{|\Theta|+|\Omega|})$. When combining one classification and multiple clustering methods, the complexity is $\mathcal{O}(M_U(2^{2|\Theta|}+2^{2|\Omega|}+2^{|\Theta|+|\Omega|}))$ where M_U represents the number of clustering methods. When combining multiple classification and clustering methods, the complexity is $\mathcal{O}(M_S M_U(2^{2|\Theta|} + 2^{2|\Omega|} + 2^{|\Theta|+|\Omega|}) + 2^{M_S|\Omega|})$ where the term $2^{M_S|\Omega|}$ is from the combination of reliable information from M_U different classification results by Dempster's rule. Similarly, when using Karem's transformation method with the proposed iterative fusion process, the complexity is $\mathcal{O}(2^{|\Omega|})$ for the combination of one classification and one clustering methods, $\mathcal{O}(M_U(2^{|\Omega|}))$ for one classification and multiple clustering methods, $\mathcal{O}(M_S M_U(2^{|\Omega|}) + 2^{M_S|\Omega|})$.

Evidently, in the worst case, the proposed EFSC method has exponential complexity because of Dempster's combination rule. To reduce the complexity of Dempster's combination rule, various approximation algorithms have been suggested [13], [109] to solve the combination in polynomial time.

4.4 Numerical example

In this section, we show a numerical example to study the proposed transformation and the iterative fusion process. For a group of objects with eight elements, noted as $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, their corresponding ground truth, classification and clustering results are detailed in Table 4.1. The classification result has the same frame of discernment as the ground truth, *i.e.* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. We have five different clustering results on X, separating objects into either four clusters such as c_1 or three clusters such as c_2, c_3, c_4 and c_5. To combine the classification and clustering results, we have first to transfer the BBAs of clustering methods to the same frame as classification.

Table 4.1 – Labels of ground truth, classification and clustering results on objects in X.

Labels	Objects in X								Frame of discernment
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
Ground truth	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
s_1	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4	$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
c_1	θ_{11}	θ_{11}	θ_{12}	θ_{12}	θ_{12}	θ_{13}	θ_{14}	θ_{14}	$\Theta_1 = \{\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}\}$
c_2	θ_{23}	θ_{21}	θ_{21}	θ_{21}	θ_{21}	θ_{22}	θ_{23}	θ_{23}	$\Theta_2 = \{\theta_{21}, \theta_{22}, \theta_{23}\}$
c_3	θ_{33}	θ_{33}	θ_{31}	θ_{31}	θ_{31}	θ_{32}	θ_{33}	θ_{33}	$\Theta_3 = \{\theta_{31}, \theta_{32}, \theta_{33}\}$
c_4	θ_{41}	θ_{41}	θ_{42}	θ_{42}	θ_{43}	θ_{43}	θ_{42}	θ_{42}	$\Theta_4 = \{\theta_{41}, \theta_{42}, \theta_{43}\}$
c_5	θ_{51}	θ_{51}	θ_{52}	θ_{52}	θ_{52}	θ_{53}	θ_{53}	θ_{53}	$\Theta_5 = \{\theta_{51}, \theta_{52}, \theta_{53}\}$

4.4.1 Construction of BBAs of classification and clustering results

This numerical example simulates a case where raw data of objects is not available, and we only have the results of classification and clustering. The BBAs of classification are therefore generated randomly on 2^Ω without access to the raw data, as shown in Table 4.2. As all possible states of objects labels are included in Ω, the BBAs of classification are constructed under the closed-world assumption satisfying $m(\emptyset) = 0$. The classification results should indicate that decisions are on the singletons corresponding to the current labels. Therefore, all subsets except the empty set in Ω are assigned with random value by the uniform distribution, and the singletons corresponding to the current label have the maximum BBAs, marked in bold in Table 4.2.

The clustering results are on the frames Θ_1, Θ_2, Θ_3, Θ_4 and Θ_5 and their BBAs are successively noted as $m_{c_1}^{\Theta_1}$ for clustering c_1, $m_{c_2}^{\Theta_2}$ for clustering c_2, $m_{c_3}^{\Theta_3}$ for clustering c_3, $m_{c_4}^{\Theta_4}$ for clustering c_4, $m_{c_5}^{\Theta_5}$ for clustering c_5. We use the distance model given in [35] to estimate the BBA of a clustering method by a group of SBBAs. For each object $x_i \in X$, to simplify the calculation, its SBBA related to the cluster to which x_i belongs, is constructed with 0.8 on the singleton and 0.2 on the ignorance. For other clusters that x_i does not belong to, the corresponding SBBAs have 1 on the ignorance, representing we know nothing about them. Therefore, the BBA of x_i can be represented by only one SBBA with the cluster it belongs to as the focal element. For example, for the object x_1 belonging to the cluster θ_{11} in c_1, its SBBA is defined as:

$$\begin{cases} m_{c_1}^{\Theta_1}(\theta_{11})(x_1) = 0.8, \\ m_{c_1}^{\Theta_1}(\Theta_1)(x_1) = 0.2. \end{cases} \tag{4.29}$$

Table 4.2 – BBAs of classification.

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.0578	**0.1271**	0.0616	**0.1212**	0.1274	0.0864	0.0781	0.0183
ω_2	0.0187	0.0737	**0.0941**	0.0258	**0.1679**	0.0135	**0.1291**	0.0221
$\omega_1 \cup \omega_2$	0.0255	0.0650	0.0558	0.1045	0.0336	0.0598	0.0611	0.0591
ω_3	0.0642	0.0502	0.0561	0.1031	0.0650	**0.1164**	0.0700	0.0939
$\omega_1 \cup \omega_3$	0.0588	0.0739	0.0622	0.1198	0.0272	0.0842	0.1019	0.0272
$\omega_2 \cup \omega_3$	0.0702	0.0606	0.0165	0.0792	0.1023	0.0648	0.1258	0.0366
$\omega_1 \cup \omega_2 \cup \omega_3$	0.0936	0.0250	0.0622	0.0138	0.0725	0.0376	0.1001	0.0906
ω_4	**0.1122**	0.0060	0.0501	0.0485	0.0377	0.0078	0.0281	**0.1149**
$\omega_1 \cup \omega_4$	0.1098	0.0072	0.0886	0.0258	0.0207	0.0993	0.0157	0.0515
$\omega_2 \cup \omega_4$	0.0915	0.0630	0.0901	0.1056	0.0468	0.0767	0.0433	0.1017
$\omega_1 \cup \omega_2 \cup \omega_4$	0.1086	0.0626	0.0920	0.0253	0.1212	0.0587	0.0618	0.0322
$\omega_3 \cup \omega_4$	0.0401	0.0865	0.0344	0.0445	0.0139	0.1107	0.0915	0.1085
$\omega_1 \cup \omega_3 \cup \omega_4$	0.0238	0.1065	0.0835	0.0713	0.0496	0.0647	0.0466	0.0900
$\omega_2 \cup \omega_3 \cup \omega_4$	0.0899	0.1042	0.0623	0.0289	0.0321	0.1064	0.0791	0.0936
Ω	0.0345	0.0877	0.0896	0.0819	0.0815	0.0122	0.0389	0.0589

4.4.2 Transformation of BBAs

In this section, we explain how to transfer the BBAs of clustering c_1 defined on Θ_1 to the frame of discernment Ω. We measure the similarity between the results of classification s_1 and clustering c_1 by Jaccard similarity index, as shown in Table 4.3.

Table 4.3 – Similarity of classes and clusters measured by Jaccard similarity index.

Classes	Clusters from clustering c_1			
	θ_{11}	θ_{12}	θ_{13}	θ_{14}
ω_1	0.333	0.25	0	0
ω_2	0	0.5	0	0.25
ω_3	0	0	1	0
ω_4	0.333	0	0	0.333

Let's take the object x_1 again as an example. The clustering label of x_1 is cluster θ_{11} which has intersections with two classes: ω_1 and ω_4. Thus we take the similarities between cluster θ_{11} and class ω_1, and also class ω_4 into account to achieve the transformation of $m_{c_1}^{\Theta_1}(x_1)$. The value of similarity 0.333 between cluster θ_{11} and class ω_1 indicates the belief degree that class ω_1 and cluster θ_{11} are the same, and nothing more. As displayed in equation (4.5), the similarity can be represented by a BBA on the frame of discernment

H, defined as:

$$\begin{cases} m^{\mathcal{H}_{11,1}}_{\theta_{11},\omega_1}(\mathcal{Y}_{11,1}) = 0.333, \\ m^{\mathcal{H}_{11,1}}_{\theta_{11},\omega_1}(\mathcal{H}_{11,1}) = 1 - 0.333 = 0.667. \end{cases} \tag{4.30}$$

The transformation is composed of four major steps as explained in section 4.5.4. The first step is to decompose the BBA of clustering into a group of SBBAs considered as multiple independent sources. For object x_1, its BBA $m^{\Theta_1}_{c_1}(x_1)$ are already in the form of SBBA with $\{\theta_{11}\}$ as the focal element, as shown in equation (4.29), and thus can be directly used. The second step is to transfer the SBBAs to Ω by the corresponding similarity between cluster θ_{11} and all possible classes with non-empty intersections, *i.e.* ω_1 and ω_4. According to equation (4.10) and equation (4.11), the transformed SBBA of x_1 on ω_1 is calculated as:

$$\begin{aligned} m^{\Omega}_{\theta_{11},\omega_1}(\omega_1)(x_1) &= m^{\mathcal{H}_{11,1}}_{\theta_{11},\omega_1}(\mathcal{Y}_{11,1}) * m^{\Theta_1}_{c_1}(\theta_{11})(x_1) \\ &= 0.333 * 0.8 \\ &= 0.2664 \end{aligned} \tag{4.31}$$

$$\begin{aligned} m^{\Omega}_{\theta_{11},\omega_1}(\Omega)(x_1) &= 1 - m^{\mathcal{H}_{11,1}}_{\theta_{11},\omega_1}(\mathcal{Y}_{11,1}) * m^{\Theta_1}_{c_1}(\theta_{11})(x_1) \\ &= 0.7336 \end{aligned} \tag{4.32}$$

The transformed SBBA of x_1 on ω_4 is calculated as:

$$\begin{aligned} m^{\Omega}_{\theta_{11},\omega_4}(\omega_4)(x_1) &= m^{\mathcal{H}_{11,4}}_{\theta_{11},\omega_1}(\mathcal{Y}_{11,4}) * m^{\Theta_1}_{c_1}(\theta_{11})(x_1) \\ &= 0.333 * 0.8 \\ &= 0.2664 \end{aligned} \tag{4.33}$$

$$\begin{aligned} m^{\Omega}_{\theta_{11},\omega_4}(\Omega)(x_1) &= 1 - m^{\mathcal{H}_{11,4}}_{\theta_{11},\omega_1}(\mathcal{Y}_{11,4}) * m^{\Theta_1}_{c_1}(\theta_{11})(x_1) \\ &= 0.7336 \end{aligned} \tag{4.34}$$

The third step is to combine all available SBBAs on each classes, as displayed in equation (4.19). As we only have one SBBA$m^{\Omega}_{\theta_{11},\omega_1}(x_1)$ on ω_1 and one SBBA $m^{\Omega}_{\theta_{11},\omega_4}(x_1)$ on ω_4, thus we have:

$$m^{\Omega}_{c_1,\omega_1}(x_1) = m^{\Omega}_{\theta_{11},\omega_1}(x_1) \tag{4.35}$$

$$m^{\Omega}_{c_1,\omega_4}(x_1) = m^{\Omega}_{\theta_{11},\omega_4}(x_1) \tag{4.36}$$

where $m^{\Omega}_{c_1,\omega_1}(x_1)$ represents that the information of object x_1 from clustering c_1 is transformed on ω_1 and $m^{\Omega}_{c_1,\omega_4}(x_1)$ is transformed on ω_4. The last step is to combine all possible SBBAs on Ω transformed from clustering c_1 by equation (4.20) as:

$$\begin{aligned} m^{\Omega}_{c_1}(x_1) =& m^{\Omega}_{c_1,\omega_1}(x_1) \oplus m^{\Omega}_{c_1,\omega_4}(x_1) \\ \Rightarrow & \begin{cases} m^{\Omega}_{c_1}(\omega_1)(x_1) = 0.2105, \\ m^{\Omega}_{c_1}(\omega_4)(x_1) = 0.2105, \\ m^{\Omega}_{c_1}(\Omega)(x_1) = 0.5789, \\ m^{\Omega}_{c_1}(A)(x_1) = 0, \forall A \subseteq \Omega, A \neq \omega_1, A \neq \omega_4. \end{cases} \end{aligned} \tag{4.37}$$

For other objects, their transformed BBAs on Ω are shown in Table 4.4.

Table 4.4 – Transformed BBAs from clustering c_1 on Ω.

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.2105	0.2105	0.1304	0.1304	0.1304	0.	0.	0.
ω_2	0.	0.	0.3478	0.3478	0.3478	0.	0.1549	0.1549
$\omega_1 \cup \omega_2$	0.	0.	0.	0.	0.	0.	0.	0.
ω_3	0.	0.	0.	0.	0.	0. 8	0.	0.
$\omega_1 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_2 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
ω_4	0.2105	0.2105	0.	0.	0.	0.	0.2253	0.2253
$\omega_1 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_2 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
Ω	0.5789	0.5789	0.5217	0.5217	0.5217	0.2	0.6197	0.6197

4.4.3 Iterative fusion process

After the transformation, the information from clustering c_1 can be combined with the BBAs of classification presented in Table 4.2, and the combined results of all objects are shown in Table 4.5. Decisions are made on the combined BBAs by equation (2.39), and the loss of confidence of each decision is represented by the method from [48] based on Jousselme distance from its BBAs to the corresponding categorical BBAs, as shown in Table 4.6.

The labels after decision should be Compared to the original classification labels according to the proposed criterion in algorithm 1, to decide whether the information from clustering is finally added or not. For each class, we update their labels if the average loss of confidence of this class after the combination is reduced, compared to its counterpart calculated from the BBAs of classification s_1, also shown in Table 4.6.

Let's take the class ω_1 as an example. Its average loss of confidence in s_1, noted as $\overline{R_{\omega_1,s_1}}$, is calculated as:

$$\begin{aligned} \overline{R_{\omega_1,s_1}} &= \frac{R_{x_2,s_1} + R_{x_4,s_1}}{2} \\ &= \frac{0.638 + 0.612}{2} \\ &= 0.625 \end{aligned} \tag{4.38}$$

where R_{x_2,s_1} and R_{x_4,s_1} represent the loss of confidence of objects x_2 and x_4 in s_1. After the combination with clustering c_1, the decisions on the combined BBAs update the label and also the loss of confidence of each object. Therefore, for the class ω_1, its new average loss of confidence $\overline{R_{\omega_1}}$ is calculated as:

$$\overline{R_{\omega_1}} = R_{x_2} = 0.565 \tag{4.39}$$

where R_{x_2} is the loss of confidence of object x_2 after the combination. As we have $\overline{R_{\omega_1}} < \overline{R_{\omega_1,s_1}}$, the information on ω_1 from the original classification s_1 is thus updated by the new one after combination. That is to say, the class ω_1 in s_1 originally contains two objects x_2 and x_4 is updated as only one object x_2 according to the results after combination.

Table 4.5 – BBAs on Ω after combination of clustering c_1 and classification s_1.

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.1735	0.2394	0.1385	0.1824	0.1696	0.0257	0.0624	0.0117
ω_2	0.0132	0.0532	0.3100	0.2295	0.3979	0.0028	0.2353	0.1051
$\omega_1 \cup \omega_2$	0.0181	0.0466	0.0359	0.0731	0.0186	0.0174	0.0473	0.0431
ω_3	0.0456	0.0353	0.0361	0.0721	0.0409	0.7418	0.0552	0.0699
$\omega_1 \cup \omega_3$	0.0417	0.0534	0.0404	0.0840	0.0140	0.0250	0.0833	0.0185
$\omega_2 \cup \omega_3$	0.0498	0.0432	0.0080	0.0550	0.0674	0.0189	0.1044	0.0257
$\omega_1 \cup \omega_2 \cup \omega_3$	0.0664	0.0160	0.0404	0.0081	0.0463	0.0104	0.0817	0.0674
ω_4	0.2373	0.1382	0.0319	0.0329	0.0215	0.0010	0.1063	0.2617
$\omega_1 \cup \omega_4$	0.0780	0.0024	0.0592	0.0168	0.0094	0.0298	0.0074	0.0372
$\omega_2 \cup \omega_4$	0.0650	0.0451	0.0603	0.0738	0.0280	0.0227	0.0317	0.0759
$\omega_1 \cup \omega_2 \cup \omega_4$	0.0771	0.0447	0.0617	0.0164	0.0808	0.0170	0.0479	0.0224
$\omega_3 \cup \omega_4$	0.0284	0.0630	0.02076	0.0301	0.0046	0.0334	0.0107	0.0811
$\omega_1 \cup \omega_3 \cup \omega_4$	0.0169	0.0783	0.0556	0.0493	0.0300	0.0189	0.0346	0.0669
$\omega_2 \cup \omega_3 \cup \omega_4$	0.0638	0.0765	0.0405	0.0189	0.0176	0.0321	0.0632	0.0697
Ω	0.0245	0.0639	0.0600	0.0569	0.0527	0.0024	0.0278	0.0430

Table 4.6 – Decisions on combined BBAs (Compared to the labels and loss of confidence of classification s_1).

Decisions	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Labels	ω_4	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4
Losses of confidence	0.543	0.565	0.508	0.553	0.428	0.178	0.528	0.514
Labels of s_1	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4
Losses of confidence of s_1	0.613	0.638	0.650	0.612	0.573	0.606	0.5969	0.605

The newly updated labels can be considered as a new classification result to combine with another randomly selected clustering result among c_1, c_2, c_3, c_4 and c_5. In Table 4.7, we show the details of the proposed iterative fusion process, including the clustering used for combination, the average loss of confidence on each class (*i.e.* $\overline{R_{\omega_1}}$, $\overline{R_{\omega_2}}$, $\overline{R_{\omega_3}}$, $\overline{R_{\omega_4}}$), the global average loss of confidence $\overline{R}$, and the labels of objects in each iteration. We run ten iterations in total and the average loss of confidence $\overline{R}$ gradually decreases from 0.611 in classification s_1 to 0.002 after the combination with multiple clustering results. The original accuracy of s_1 is 0.625, and it gradually increases to 1 after the fifth iteration.

Table 4.7 – Iterative fusion process.

Combination	Losses of confidence					Labels								Accuracy
	$\overline{R_{\omega_1}}$	$\overline{R_{\omega_2}}$	$\overline{R_{\omega_3}}$	$\overline{R_{\omega_4}}$	$\overline{R}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
s_1	0.625	0.606	0.606	0.608	0.611	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4	0.625
$+c_1$	0.565	0.504	0.178	0.527	0.443	ω_4	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.738
$+c_1$	0.496	0.320	0.039	0.441	0.324	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_2$	0.485	0.255	0.008	0.451	0.299	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_3$	0.421	0.194	0.002	0.436	0.263	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_1$	0.126	0.031	0.000	0.425	0.145	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.028	0.006	0.000	0.138	0.057	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_4$	0.005	0.005	0.000	0.114	0.041	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.001	0.001	0.000	0.025	0.006	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_5$	0.000	0.000	0.000	0.013	0.003	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.000	0.000	0.000	0.009	0.002	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0

4.5 Experiments on synthetic data

In this section, we evaluate the proposed EFSC on synthetic data. We principally discuss the performance of EFSC from the aspect at the output level, by controlling the quality of its direct inputs: classification and clustering results. To make the experiments concise and clear, we configure EFSC into three different ways:

1. **EFSC11**: one classification and one clustering methods.
2. **EFSC1m**: one classification and multiple clustering methods.
3. **EFSCmm**: Multiple classification and multiple clustering methods.

EFSC focuses on the fusion of classification or clustering results rather than raw data, features nor supervised/unsupervised classifiers themselves. This indicates that the direct inputs of EFSC are the classification and clustering results and their corresponding BBAs. Therefore, we verify EFSC at the output level in a controlled environment. In section 4.5.2, we study how the quality of classification and clustering results affect the combination of the EFSC11 configuration. In section 4.5.3, we focus on the EFSC1m configuration and study the same questions as section 4.5.2. In section 4.5.4, we evaluate EFSCmm configuration with multiple classifications and clustering methods and compare the results with other fusion approaches.

We have selected two distinct fusion methods at the production level, based on different principles. EC3 belongs to the group of fusion methods that optimize the agreements between supervised and pooled results, and it was shown to be more efficient in comparison to the other methods in this category [25]. Similar to our work, the method proposed by Karem allows obtaining a combination through belief functions.

4.5.1 Synthetic data at the output level

The synthetic data at the output level includes the outputs of supervised or unsupervised classifiers and their corresponding BBAs, which are used as inputs of EFSC. To study how qualities of classification and clustering results affect the combination, we generate an image as ground truth, and gradually add uniformly distributed mistakes on it to obtain classification and clustering results. In this way, we can directly control the quality of supervised and unsupervised results. The experiments on synthetic labels simulate a case where raw data is inaccessible, and we only have the results of classification and clustering. The experiments are all launched 15 times to obtain the averages.

Table 4.8 – Descriptions of the synthetic ground truth.

	Forest	Shrub/Scrub	Grassland /Herbaceous	Wetlands	Total
Synthetic ground truth	20687	11417	3526	4370	40000

The process to generate an classification result with the proportion of mistakes M_s based on the ground truth is detailed as:

1) Randomly select pixels with proportion M_s per class;

2) For the selected pixels, change their labels randomly into any other class.

The random BBAs of the synthetic classification result on the frame of discernment Ω are constructed as:

1) For a pixel x labeled as ω_i, we have $m(\omega_i) = 1 - M_s$;

2) The rest of BBAs are randomly set to $[0, 1 - M_s)$;

3) Normalize the BBAs to make the sum equal to 1.

Generating a synthetic clustering result includes two major steps: separating or gathering classes into clusters, and add noise with proportion M_c into each cluster to reduce its homogeneity. The quality of clustering indicates whether data are well-separated, which is different from the evaluation of classification. We use the homogeneity in clusters to represent the quality of clustering results, which thus can be controlled by M_c. For example, if a class of the ground truth is separated into two clusters, the homogeneity of each cluster is not reduced, and thus we can still consider it as a perfect clustering. After adding noise with M_c, the homogeneity in each cluster is gradually reduced. We note the number of classes on the ground truth as n, and the number of clusters as k. The process to generate a clustering result is as:

1) If $k = n$, classes on the ground truth is directly taken as clusters;

2) If $k < n$, we randomly select $n - k$ classes on the ground truth and gather them as one cluster. The rest classes are kept as clusters;

3) If $k > n$, we use an iterative process to add clusters. In each iteration, we randomly select a cluster and separate it into two clusters until the clusters are enough. The initial state of this process is the ground truth whose classes are directly taken as clusters;

4) For the clustering result generated previously, we randomly select pixels in each cluster with the proportion M_c and change its labels to any other cluster.

The BBAs of clustering on Θ are constructed in the form of SBBAs. For a pixel x in the cluster θ_j, we have $m(\theta_j) = 1 - M_c$ and $m(\Theta) = M_c$.

4.5.2 Combination of one classification and one clustering methods

In this section, we evaluate EFSC11 configuration by combining one classification and one clustering. We first discuss the effects of different likeness/similarity measures for the proposed transformation used in EFSC11 and Karem's methods. Due to the inaccessibility of raw data in this experiment, we only discuss the similarities using output information.

The Table 4.9 shows the accuracy measured by F1 score for the combination of one classification and one clustering methods with different measures of likeness/similarity. The formula for the F1 score is:

$$\mathrm{F1} = 2 \times \frac{\mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \tag{4.40}$$

We fix the mistakes M_s in classification as 0.5 which represents less satisfying results and the mistakes M_c in clustering as 0.2 to represent a good clustering. The number of clusters are changed as 3, 7, 11, 15. Karem's method originally uses the proportion likeness which surpasses the other three similarities. Compared to Karem's method, EFSC11 performs better with different similarities. For instance, Karem's method cannot improve the accuracy of classification with Dice similarity in most cases, whereas EFSC11 brings evident improvements. In Table 4.9, no evident relations between numbers of clusters and similarities are observed, but EFSC outperforms Karem's method in most cases. In the following experiments, we use Jaccard similarity for the proposed transformation, and proportion likeness for Karem's method.

We also focus on how the quality of classification or clustering affect the combination. For the clustering, we fix its number of clusters to $k = 7$ and gradually add mistakes with proportion $M_c \in [0, 1]$. As we use the homogeneity to represent the quality of a clustering, $M_c = 0$ represents the perfect case where no mistakes are involved. With M_c increasing, the homogeneity of each cluster begins to decrease, which thus can be considered as a decline of the clustering quality.

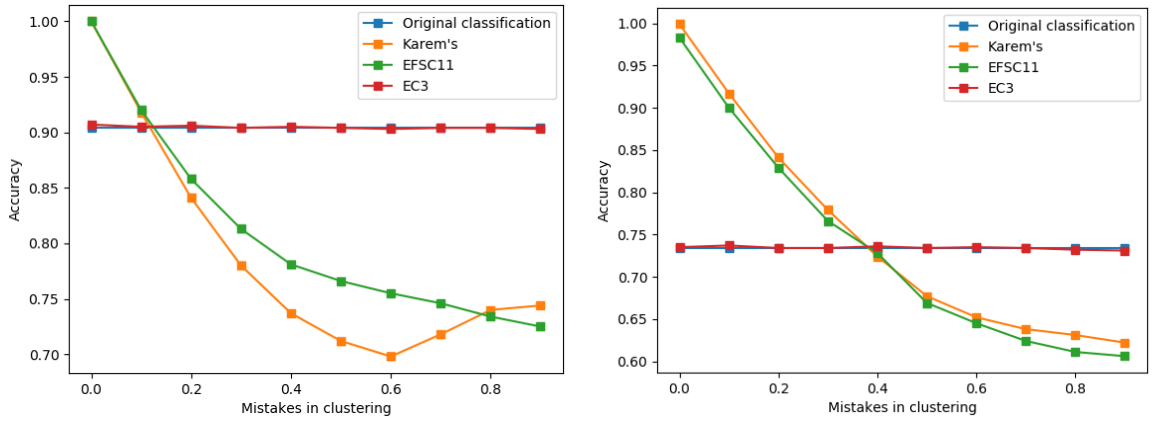
Table 4.9 – Combination of one classification and one clustering methods with different likeness/similarity measures.

Number of clusters	Methods	Likeness/similarity measures			
		Jaccard	Proportion	Dice	Recovery
k=3	Karem's	**0.721**	0.687	0.679	0.694
	EFSC11	0.704	**0.719**	**0.707**	**0.742**
k=7	Karem's	**0.698**	**0.836**	0.834	**0.853**
	EFSC11	0.583	0.821	**0.841**	0.822
k=11	Karem's	0.583	0.587	0.585	0.584
	EFSC11	**0.721**	**0.790**	**0.777**	**0.795**
k=15	Karem's	0.584	0.584	**0.587**	0.584
	EFSC11	**0.588**	**0.589**	0.586	**0.586**
Original classification		0.586			

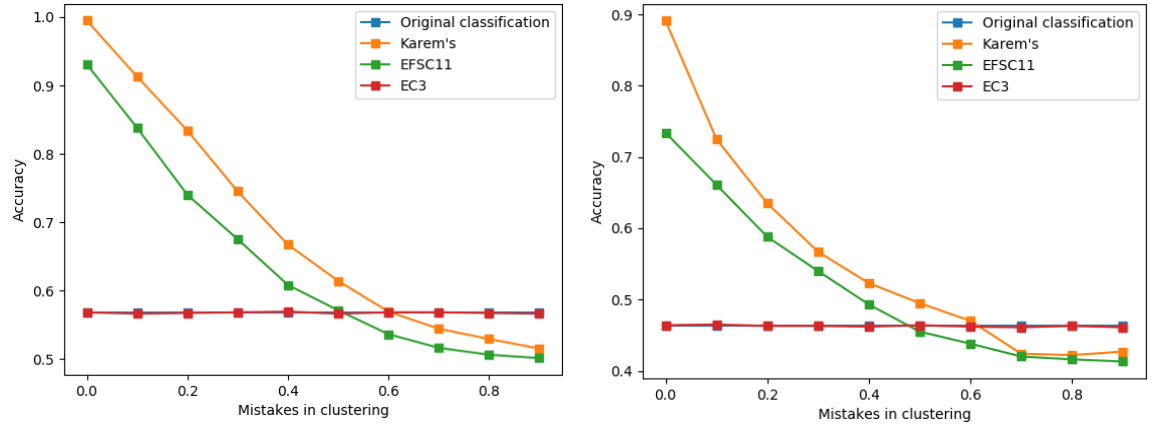
We fix the mistakes M_s in classification as 0.1, 0.3, 0.5 and 0.7 and for each of them, M_c varies from 0 to 0.9, as shown in Figure 4.3. A perfect clustering with $M_c = 0$ increases the accuracy of classification, whereas the accuracy after combinations by EFSC11 and Karem's are not improved when $M_c \geq 0.2$, as shown in Figure 4.3(a). The BBAs of classification are assigned by random values but also guarantee the singletons corresponding to the classification labels have the maximum BBA. Therefore, for the classification with $M_s = 0.1$, even though almost all pixels have true labels, the BBAs are still constructed with uncertainty and imprecision, indicating that the classification is good enough yet not completely reliable. Clustering information can thus be partially taken into account, however, possibly leading to a decrease of the accuracy when the quality of clustering is relatively poorer than the one of the classification. We can also roughly observe this in Figure 4.3(b), 4.3(c) and 4.3(d). When $M_s = 0.3$, the accuracy after combination is improved only if $M_c < 0.4$, and for $M_s = 0.5$ and $M_s = 0.7$, improvement occurs when $M_c < 0.5$.

We also fix M_c for clustering and change the mistakes M_s in classification. To avoid the influence of mistakes in clustering, we select a good enough clustering with $M_c = 0.1$, and study how the quality of classification affects the combination. The results are shown in Figure 4.4. We can observe that combining with a good enough clustering can noticeably enhance the accuracy, even for the low-quality classification (*e.g.* $M_s = 0.7$). When $M_s > 0.8$, it indicates that the classification is almost incorrect and thus the combination is helpless to improve the accuracy, even with a good enough clustering.

In the combination of one classification and one clustering, the improvement of accuracy is still pronounced after combination by EFSC11 and Karem's. Nevertheless, Karem's has better performance than EFSC11 because EFSC11 takes information more prudently so that less information is fused compared to Karem's. Accordingly, Karem's is more pertinent for combining one classification and one clustering, whereas EFSC has the advantages when multiple classification and clustering are available. EC3 has a similar performance as the original classification because it relies more on classification than clustering. EC3 is more pertinent for multiple classification results with high qualities.



(a) $M_s = 0.1$ (accuracy of classification = 0.904). (b) $M_s = 0.3$ (accuracy of classification = 0.734).



(c) $M_s = 0.5$ (accuracy of classification = 0.586). (d) $M_s = 0.7$ (accuracy of classification = 0.463).

Figure 4.3. Accuracy change with mistakes in clustering for the combination of one classification and one clustering.

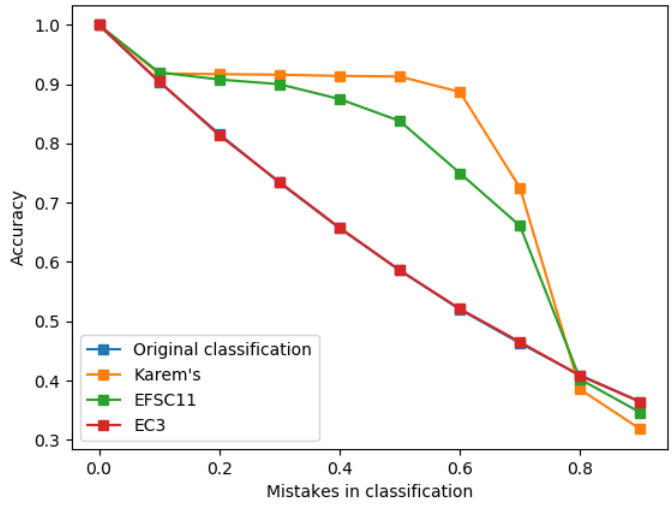


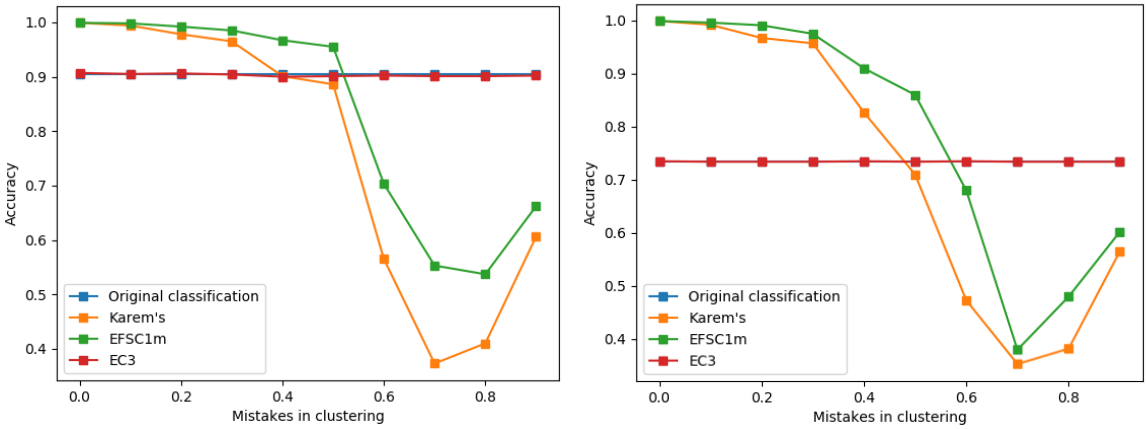
Figure 4.4. Accuracy change with mistakes in classification for the combination of one classification and one clustering.

4.5.3 Combination of one classification and multiple clustering methods

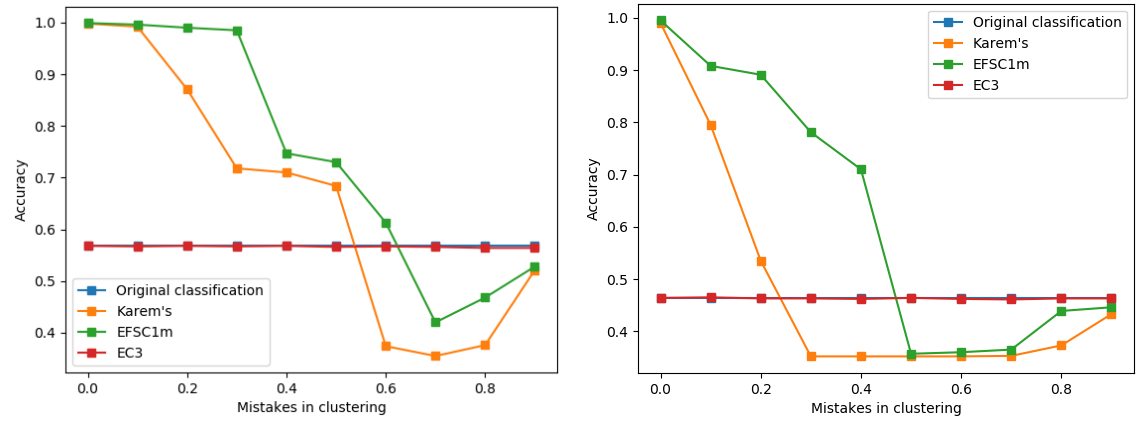
In this section, we evaluate the EFSC1m configuration with one classification and multiple clustering methods. We also fix M_s separately as 0.1, 0.3, 0.5 and 0.7 and gradually change M_c from 0 to 0.9. A group of clustering methods has its number of clusters varying from 3 to 15.

EFSC1m and Karem's methods, combined with multiple clustering methods can improve the accuracy of classification, as shown in Figure 4.5. A high-quality classification boosts the combination results with multiple clustering, as in Figures 4.5(a), 4.5(b).

Even if the mistakes in classification are considerable, *i.e.* $M_s = 0.7$, EFSC1m and Karem's methods can increase the overall accuracy, as shown in Figure 4.5(d).



(a) $M_s = 0.1$ (accuracy of classification = 0.904). (b) $M_s = 0.3$ (accuracy of classification = 0.734).



(c) $M_s = 0.5$ (accuracy of classification = 0.586). (d) $M_s = 0.7$ (accuracy of classification = 0.463).

Figure 4.5. Accuracy change with mistakes in clustering for the combination of one classification and multiple clustering methods.

We can observe that the combination with multiple clustering methods can better decrease the influence of the quality of classification, those in the previous section. When

M_s reaches 0.1, 0.3 and 0.5, the classification accuracy is improved for $M_c < 0.5$. This indicates that when combining with multiple clustering methods, EFSC1m and Karem's methods are more influenced by clustering than classification.

When $M_c > 0.6$, the accuracy after combination slightly increases yet is always worse in comparison to the original classification. This increase occurs when clustering has low quality because the BBAs of clustering, in this case, is mainly distributed on the total ignorance. Therefore, information from clustering with low quality is rarely combined with classification, consequently avoiding further reduction of accuracy.

We also fix the quality of clustering with $M_c = 0.1$ and study the accuracy change with M_s, as shown in Figure 4.6. The accuracy after combination shows an evident decline when $M_c > 0.5$. This result also proved that when the quality of clustering is acceptable, the combination is less influenced by classification.

In the combination of one classification and multiple clustering methods, EFSC1m outperforms the other two methods because it cautiously takes into account the information from clustering. EC3 still shows similar performance as the original classification because when maximizing the consensus of classification and clustering, only one classification is available to provide the semantic labels. This indicates that EC3 relies more on classification than clustering. Compared to EC3, EFSC1m, and Karem's methods with the proposed iterative fusion process can better handle the combination with one classification and multiple clustering methods.

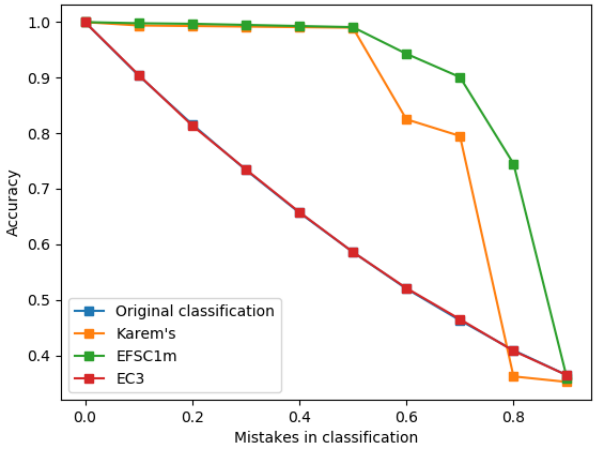


Figure 4.6. Accuracy change with mistakes in classification for the combination of one classification and multiple clustering methods.

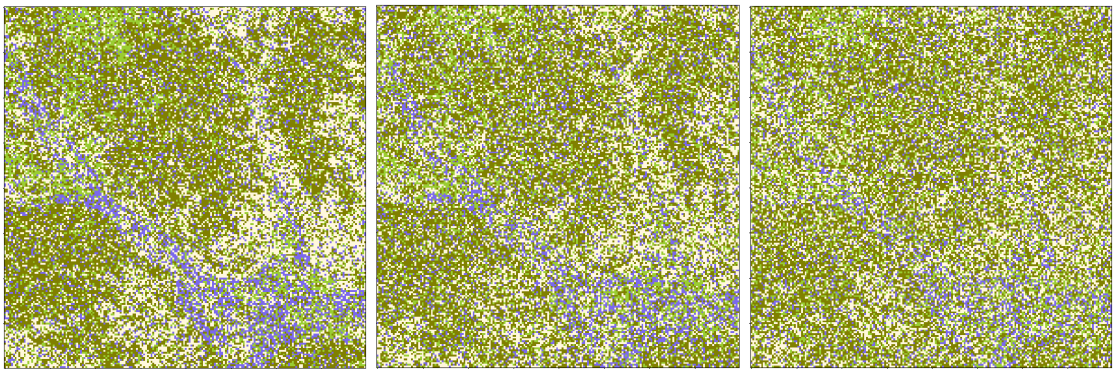
4.5.4 Combination of multiple classification and clustering methods

In this section, we combine multiple classification and clustering results separately by EFSCmm, Karem's and EC3 methods. We also compare the fusion of multiple classifications by Majority Voting (MV). We randomly generate three classification results with

the accuracy of 0.586, 0.522, 0.463. The group of clustering methods has 3 to 15 partitions with $M_c = 0.3$.

Table 4.10 – Accuracy of three original classifications and four fusion methods (Karem's, EFSCmm, EC3, MV) on synthetic data.

Methods	Accuracy per class				Accuracy
	Forest	Shrub /Scrub	Grassland /Herbaceous	Wetlands	
Classification 1	0.667	0.573	0.363	0.421	0.586
Classification 2	0.609	0.497	0.292	0.355	0.522
Classification 3	0.562	0.422	0.230	0.285	0.463
Karem's	0.786	0.828	0.	0.	0.795
EFSCmm	**0.880**	**0.955**	0.026	0.433	**0.889**
EC3	0.682	0.590	**0.389**	0.446	0.604
MV	0.705	0.594	0.388	**0.452**	0.595

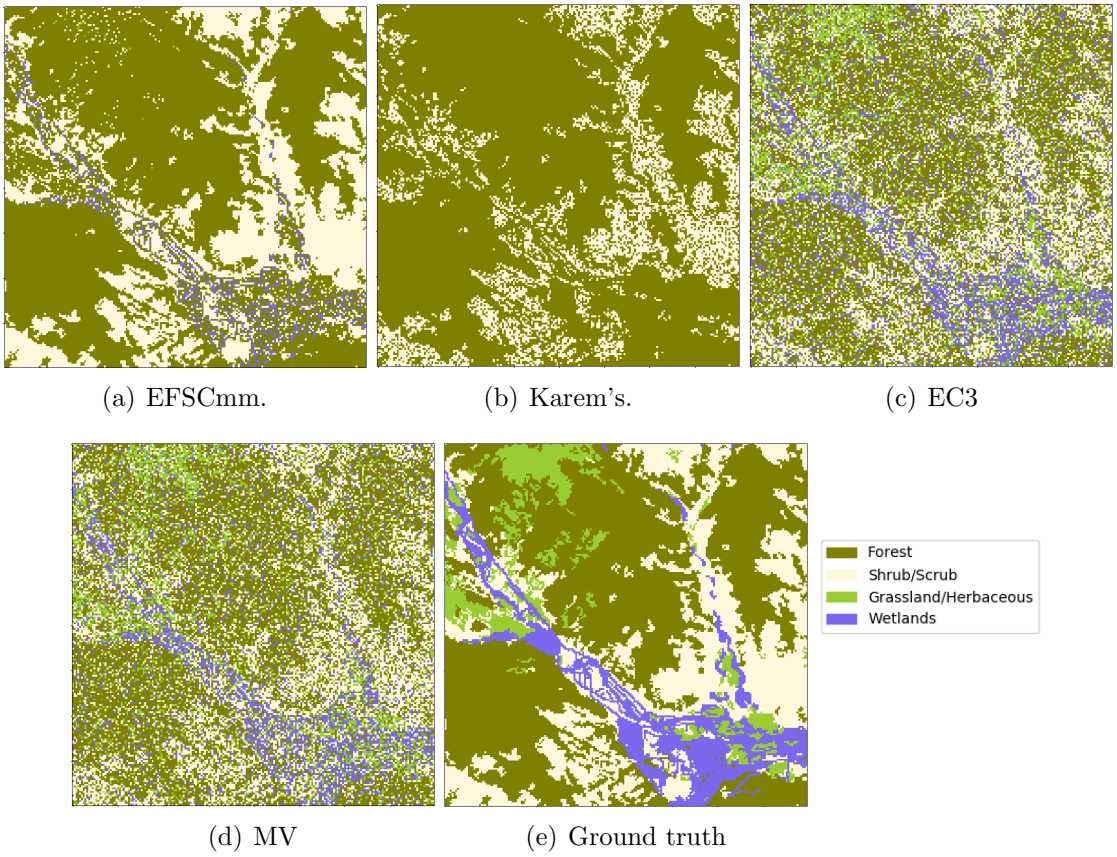


(a) Classification 1. (b) Classification 2. (c) Classification 3.

Figure 4.7. Synthetic classification results.

The Table 4.10 shows the overall accuracy measured by F1 score and for each class. We also show the classification in Figure 4.7 and combination results in Figure 4.8. EFSCmm can noticeably enhance the accuracy by 30% at most compared to the best classification, and it also surpasses other combination methods. Karem's also shows an evident improvement of the overall accuracy, whereas ignores two classes: *Grassland/Herbaceous* and *Wetlands*. Although EFSCmm cannot well classify *Grassland/Herbaceous* either, it improves the accuracy of *Wetlands*, demonstrating it is more prudent than Karem's. EC3 performs better with three classifications compared to its previous performance because it depends more on classification than clustering. For the combination results of the three classifications by MV, we can observe it is less pertinent in comparison to the other three combination methods, indicating the importance of clustering information.

We assume that for each object x, the probability that an object is correctly classified follows Gaussian distribution. Thus for 40000 pixels, each of them can be regarded as



(a) EFSCmm. (b) Karem's. (c) EC3

(d) MV (e) Ground truth

Figure 4.8. Combination results of EFSCmm, Karem's, EC3 and MV on synthetic data.

an independent result. We thus consider approximately that each test is run 40000×15 to calculate the confidence interval. The confidence interval with the confidence level as 95% of EFSCmm is $[0.887, 0.890]$, Karem's $[0.793, 0.796]$, EC3 $[0.602, 0.605]$, and MV $[0.593, 0.596]$. The experiments show that EFSCmm is significantly better than Karem's, EC3 and MV on synthetic data.

4.6 Conclusion

In this chapter, we present the EFSC framework including a transformation method and an iterative fusion process. This framework aims to combine multiple supervised and clustering results at the output level. We evaluate the EFSC framework on synthetic labels at the output level to simulate the case where raw data is inaccessible and only supervised/clustering results are available. In this situation, the BBAs of supervised/clustering results are thus directly constructed by users. In the controlled environment, we study how the qualities of supervised and clustering results affect the combination. Compared to other existing methods, we found that EFSC relies less on supervised results and can take more information from clustering results. Accordingly, EFSC is more appropriate to improve the supervised results with low accuracy.

The combination of heterogeneous information at the output level is always difficult. We will further study how supervised and clustering results affect the combination in the proposed transformation on real remote sensing data in the next chapter.

Chapter 5

Evaluation of EFSC on real remote sensing data

Resume

In this chapter, we evaluate the EFSC method proposed in chapter 4 on real remote sensing datasets to study: (1) how different likeness/similarity and numbers of clusters affect the combination; (3) how mislabeled training samples affect the fusion results; (4) how the EFSC behaves on real remote sensing data.

We also use identification methods based on spectral indexes, proposed in chapter 3, as a classification to combine with multiple clustering methods on the data provided by TOTAL.

5.1 Introduction

In the previous chapter, we propose a solution for combining several supervised and unsupervised classification methods and achieve the evaluation on synthetic data at the output level, to simulate the situation where only supervised and clustering results are available whereas raw data is inaccessible. Besides the situation without raw data, another common situation in real applications is that raw data is available, while labeled samples are severely insufficient or even partially mislabeled. Therefore, in this chapter, we achieve experiments on real remote sensing datasets to study the performance of EFSC on real data and how the quality of raw data (*e.g.* labeled samples) affects the combination of EFSC.

To achieve the numerical comparisons and analysis, we conduct the experiments on the study area located in Colorado, USA, with the data of LandSat-8 OLI because ground truth is available in this area. We use the same configuration of EFSC as the previous chapter:

1. **EFSC11**: one classification and one clustering methods.

2. **EFSC1m**: one classification and multiple clustering methods.

3. **EFSCmm**: multiple classification and multiple clustering methods.

We conduct four experiments to study: (1) how different likeness/similarity measures and numbers of clusters affect the combination; (2) how the EFSC behaves with one classification and multiple clustering methods; (3) how mislabeled training samples affect the fusion results; (4) how the EFSC behaves with multiple classification and multiple clustering methods on real remote sensing data.

In the first experiment in section 5.2.2, we test the EFSC11 configuration and change the numbers of partitions in clustering. In section 5.2.3, the second experiment studies the proposed iterative fusion process. We focus on the EFSC1m configuration to demonstrate that information from multiple clustering methods can progressively reduce the uncertainty and increase the accuracy of the supervised results until convergence. In section 5.2.4, we further investigate the robustness of the EFSC1m configuration for the mislabeled training samples in an artificial case and a real situation. The last experiment in section 5.2.5 focuses on the efficiency of the EFSCmm configuration where multiple supervised results were combined individually with multiple clustering methods. The extracted reliable information then are combined to improve the overall accuracy. EFSCmm is constructed based on EFSC1m whose effectiveness in reducing uncertainty and robustness for mislabeled training samples has already been discussed in sections 5.2.3 and 5.2.4. For EFSCmm, we, therefore, focus only on its performance to improve the overall accuracy.

We also evaluate EFSC in the study area located in Spain with the data from RapidEye and World-View 2, provided by TOTAL. As no ground truth is available, we use the automatic identification methods proposed in chapter 3 as classification results to combine with multiple clustering methods.

5.2 Experiments on LandSat data

5.2.1 Descriptions of study area

The study area is located in Colorado, USA, and contains two national forest parks with a variety of vegetation, as shown in figure 3a. The satellite data used in our experiment are from LandSat-8 OLI. It consists of eight spectral bands with a spatial resolution of 30 meters, a panchromatic band with a resolution of 15 meters, and two thermal bands with a resolution of 100 meters. The remote sensing data acquired 11^{th} June 2018 were obtained from the USGS Earth Explorer. The geometric correction of the image was performed by the transformation of the UTM map by NASA. We used the multi-spectral bands $1-7$ in our test. For ground truth, we use the labels generated by National Land Cover Database (NLCD) 2016 and group its land cover pattern at the basic level: *Water, Developed Area, Forest, Shrub/Scrub, Grassland/Herbaceous, Pasture/Crops, Wetlands*. We randomly select two test areas, one with 2500 pixels and the other with 40000 pixels, detailed in Figure 5.1 and Table 5.1. Most of the experiments are conducted on test area 1, but we also evaluate the proposed EFSC on the test area 2 in section 5.2.5.

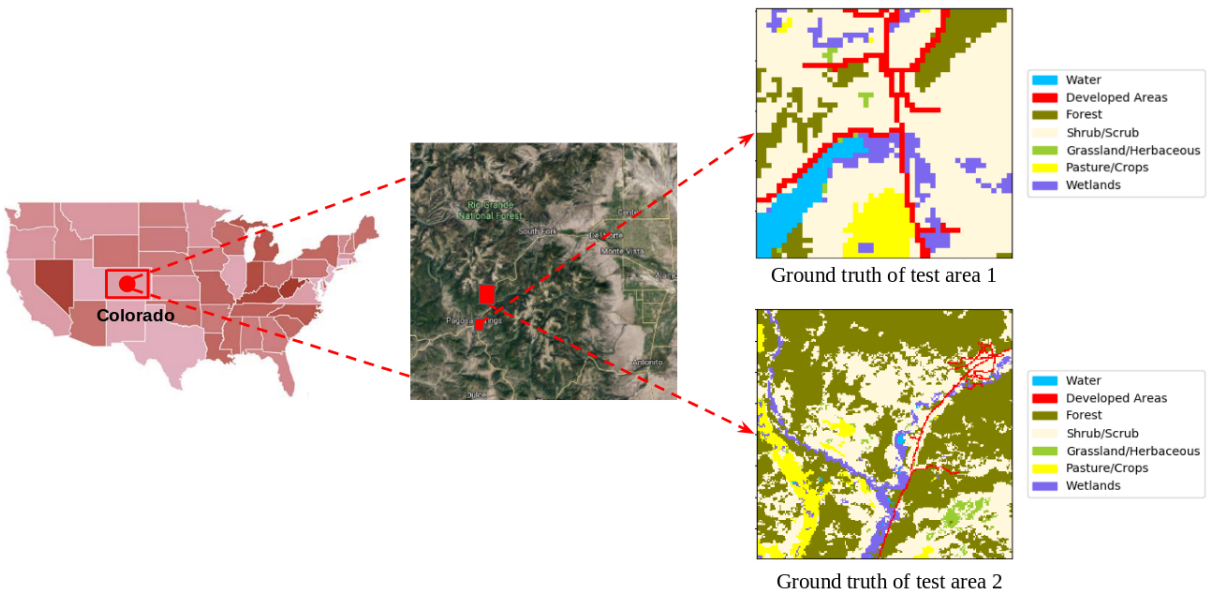


Figure 5.1. Ground truth of test areas.

Table 5.1 – Descriptions of the test areas.

	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Total
Test area 1	146	182	266	1558	15	148	185	2500
Test area 2	77	571	21147	13609	593	2250	1753	40000

5.2.2 Combination of one classification and one clustering methods

In this section, we focus on evaluating the proposed transformation by combining one classification and one clustering. We test nine pairs of combination methods between three classifiers: 5-Nearest Neighbors (5-NN), Random Forest (RF) and Stochastic Gradient Boost (SGB), and three clustering methods: k-means (KM), Spectral Clustering (SC) and Gaussian Mixture Model (GMM). In each combination, we set the number of clusters as 3, 7, 11, 15 in each clustering method to study how this affects the fusion results. We also use four different likeness/similarity measures during fusion to observe the performance. We evaluate the results by F1 score with a weighted average to take into account the imbalance of labels. Each test was launched 15 times to calculate the average.

To study the effects of likeness/similarity, we fix the k-means as the clustering method and change its numbers of clusters when combining with different classification results. Table 5.2 shows the results for EFSC11 and Karem's methods with different likeness/similarity: Jaccard, Proportion, Dice, and Recovery. Compared to Karem's transformation, the EFSC11 has better performance and is more stable with different likeness/similarity measures. Karem's method works better with Proportion likeness and shows less satisfying performance with the other three similarity measures. There is no evident relationship between likeness/similarity measures with the numbers of clusters and classifiers. In all the following experiments, we use Jaccard similarity for the proposed EFSC method. For Karem's method, as the Proportion likeness is evidently better than others, we keep it for the following experiments.

Table 5.2 – Accuracy of Karem's, EFSC11 with combination of one classification (5-NN, RF, SGB) and one clustering (KM) for different likeness/similarity measures on test area 1.

likeness/similarity	Numbers of clusters	5-NN 0.561		RF 0.541		SGB 0.553	
		Karem's	EFSC11	Karem's	EFSC11	Karem's	EFSC11
Jaccard	k=3	0.563	**0.582**	0.561	**0.575**	0.592	**0.604**
	k=7	0.562	**0.568**	0.554	**0.561**	0.574	**0.590**
	k=11	**0.575**	0.572	0.556	**0.560**	**0.597**	0.582
	k=15	0.568	**0.570**	**0.547**	0.534	0.562	**0.570**
Proportion	k=3	0.566	**0.577**	**0.607**	0.570	**0.626**	0.581
	k=7	0.567	**0.579**	**0.584**	0.577	**0.629**	0.598
	k=11	0.570	**0.588**	0.586	**0.589**	**0.629**	0.601
	k=15	0.568	**0.583**	0.587	**0.589**	**0.621**	0.586
Dice	k=3	0.566	**0.602**	0.543	**0.588**	0.556	**0.601**
	k=7	0.563	**0.588**	0.552	**0.576**	**0.573**	0.569
	k=11	0.561	**0.576**	0.548	**0.582**	0.572	**0.613**
	k=15	**0.588**	0.586	0.546	**0.577**	0.564	**0.598**
Recovery	k=3	0.554	**0.587**	0.567	**0.576**	0.567	**0.574**
	k=7	0.562	**0.575**	0.568	**0.577**	**0.579**	0.573
	k=11	**0.573**	0.568	0.575	**0.571**	0.568	**0.573**
	k=15	0.574	**0.579**	**0.591**	0.579	0.576	**0.578**

Table 5.3 – Accuracy of Karem's, EFSC11 and EC3, with combination of one classification (5-NN, RF, SGB) and one clustering (KM, SC,GMM) on test area 1.

Clustering	Numbers of clusters	5-NN 0.561			RF 0.541			SGB 0.553		
		Karem's	EFSC11	EC3	Karem's	EFSC11	EC3	Karem's	EFSC11	EC3
KM	k=3 (0.424)	0.566	**0.582**	0.568	**0.607**	0.575	0.547	**0.626**	0.604	0.558
	k=7 (0.349)	0.567	**0.568**	0.562	**0.584**	0.561	0.541	**0.629**	0.590	0.553
	k=11 (0.315)	0.570	**0.585**	0.562	**0.586**	0.560	0.542	**0.629**	0.582	0.554
	k=15 (0.294)	0.568	**0.570**	0.562	**0.587**	0.534	0.545	**0.621**	0.570	0.553
SC	k=3 (0.401)	**0.587**	0.584	0.568	**0.587**	0.571	0.552	0.561	**0.567**	0.560
	k=7 (0.119)	**0.579**	0.570	0.563	**0.606**	0.558	0.545	**0.603**	0.579	0.553
	k=11 (0.087)	**0.573**	0.571	0.563	**0.593**	0.572	0.541	**0.600**	0.581	0.553
	k=15 (0.009)	0.570	**0.572**	0.550	0.596	**0.572**	0.541	**0.597**	0.580	0.553
GMM	k=3 (0.239)	0.566	**0.613**	0.564	**0.568**	0.546	0.540	**0.593**	0.558	0.554
	k=7 (0.171)	0.565	**0.575**	0.568	**0.603**	0.556	0.543	**0.618**	0.566	0.554
	k=11 (0.111)	0.559	**0.564**	0.554	**0.632**	0.578	0.542	**0.620**	0.562	0.553
	k=15 (0.155)	0.570	**0.581**	0.558	**0.589**	0.551	0.542	0.570	**0.613**	0.553

The Table 5.3 shows the F1 score for the three fusion methods. It is evident that on the test data set, Karem's and EFSC perform better than EC3 which requires high-quality clustering results. The combination results based on Karem's transformation and EFSC have neither evident relation with the silhouette score of clustering nor the number of partitions. Instead of directly taking the maximum agreement of classification and clustering results, these two belief function-based methods can transfer information of clustering results into the frame of discernment of classification with the theory of belief functions. During this process, complementary information that is difficult to measure directly is extracted to improve the accuracy of the classification.

Compared to Karem's transformation, the EFSC11 shows no obvious improvement because the combination with clustering can fuse only limited information with the classification. Our proposed transformation is generally more cautious than Karem's so that it is less possible to change its decisions when the information provided by the clustering is highly limited.

5.2.3 Combination of one classification and multiple clustering methods

In this section, we evaluate the EFSC1m configuration that combines one classification and multiple clustering methods through the proposed iterative fusion process. We select three clustering methods: k-means (KM), Spectral Clustering (SC) and Gaussian Mixture Model (GMM), which are commonly used in land cover classification, to construct the pool of clustering methods. The number of clusters in each clustering method varies from 3 to 15. We perform the iterative fusion process 300 times by randomly selecting a clustering in the pool.

Since k-means and SC are implemented based on distance, we employ the distance model to construct their BBAs. Gaussian Mixture Model separates data by distribution, thus providing a probability for the clustering results. The proposed transformation requires separable BBAs from clustering. Therefore, the probability from GMM should be discounted by a discounting coefficient, so that they can be decomposed by the canonical

decomposition to a group of SBBAs. In this experiment, we assume that the clustering results are highly reliable, so that we set the discounting value as 0.9. The discounting coefficient represents the reliability of the source, which thus can be defined by users according to different situations.

Table 5.4 – Accuracy of the original RF, Karem's EFSC1m and EC3 on test area 1.

Classifiers	States	Accuracy per class							Accuracy
		Water	Developed Areas	Forest	Shrub/ Scrub	Grassland/ Herbacrous	Pasture/ Crops	Wetlands	
RF	Original	**0.922**	0.277	0.518	0.609	0.095	0.214	**0.215**	0.541
	Karem's	0.795	0.012	0.316	**0.816**	0.056	0.079	0.164	0.622
	EFSC1m	0.866	**0.279**	**0.589**	0.762	**0.109**	**0.335**	0.071	**0.633**
	EC3	0.882	0.263	0.567	0.670	0.042	0.264	0.108	0.549

We have verified that the EFSC11 configuration does not depend on the type of classifiers in section 5.2.2. Due to space limitations, we therefore only displayed the results of Random Forest as a supervised method in the EFSC1m configuration. Karem's transformation only works on a supervised method and a classification method. To compare with the EFSC1m configuration, we also used the proposed iterative fusion process to combine several clustering methods with Karem's transformation. As EC3 cannot handle uncertainty and cooperate with the iterative fusion process, we used EC3 to combine directly the classifier with all clustering methods in the clustering pool.

We display the comparison of the accuracy of the three methods in Table 5.4 and mark the best performance on each class by bold text. EC3 improves the accuracy by 0.8% compared to the original classification, which is outperformed by EFSC1m and Karem's methods. The results of the EC3 method were limited by insufficient training samples. Compared to the results in section 5.2.2, the combination with multiple clustering methods does not make a big difference, as the clustering results on our test data have low qualities as measured by the silhouette score. EC3 generally requires sufficient training samples for parameter selection and good clustering results to reach the agreement.

Although Karem's can improve accuracy by 8%, it cannot preserve information on all classes during the iterative fusion process. Some classes, such as *Developed Areas*, are more uncertain so that could be ignored when making decisions based on Karem's transformation. EFSC1m shows a noticeable improvement in accuracy of 9% and the information on all classes is well preserved. Indeed, the Karem's transformation only takes into account the uncertainty of the clustering results while ignoring their imprecision. Moreover, the similarity to achieve their transformation is the proportion of each class in a cluster, which also ignores more information compared to Jaccard similarity index. Therefore, Karem's transformation tends to make a more certain class to become more dominated in the iterative fusion process. On the contrary, our proposed transformation can preserve both uncertainty and imprecision in clustering results, based on which decisions become more cautious. When supervised results are unreliable, it is more appropriate to gradually fuse the clustering information more cautiously. The tests are launched 15 times to calculate the averages. As each pixel can be regarded as an independent result, we consider approximately that each test is run 2500×15 to calculate

the confidence interval. The confidence interval with confidence level of 95% of EFSC1m is $[0.628, 0.637]$, Karem's $[0.617, 0.626]$, EC3 $[0.543, 0.554]$. The experiments show that EFSC1m is significantly better than Karem's and EC3.

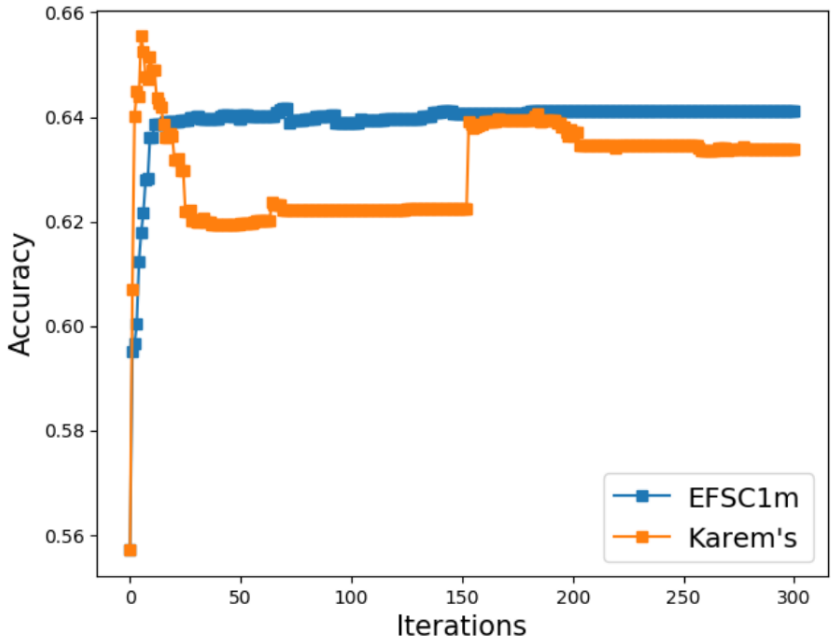


Figure 5.2. Accuracy change in the iterative process of EFSC1m and Karem's on test area 1.

As EC3 neither works in the iterative process nor deals with uncertainty, we only study the EFSC1m and Karem's on the changes of accuracy at each iteration and the reduction of uncertainty. Figure 5.2 shows how accuracy changes with each iteration. The accuracy of Karem's could decrease or heavily changed during the iterative fusion process. This may be caused by the loss of some semantic labels. However, our method becomes more stable and can reach convergence for a fixed clustering pool. Dempster's combination rule can strengthen similar information so that in the proposed iterative fusion process, it can gradually decrease ignorance. Therefore, in our method, the later a grouping is combined, the less important it is until it no longer influences the fusion results.

We also display the change of loss of confidence before and after fusion in Figure 5.3, objects correctly classified in different intervals of loss of confidence are counted. For the original RF, the initial loss of confidence of the correct objects distributes similarly in intervals: $[0., 0.2]$, $(0, 2, 0.4]$, $(0.4, 0.6]$. EFSC1m and Karem's methods can highly reduce the loss of confidence and make the majority of correct objects in the interval $[0, 0.2]$. Compared to the Karem method, EFSC1m has the advantages in making the correct object less uncertain. In this section, we evaluate the performance of EFSC on real remote sensing data with limited training samples. The experiments show that EFSC also outperforms Karem's and EC3 methods, which is in accordance with the conclusion

in the previous experiments on synthetic data. For land cover classification by real remote sensing data, besides the challenge caused by limited labeled samples, the situation with partially mislabeled samples is also highly common in real applications. Therefore, in the next section, we evaluate the performance of EFSC in this situation.

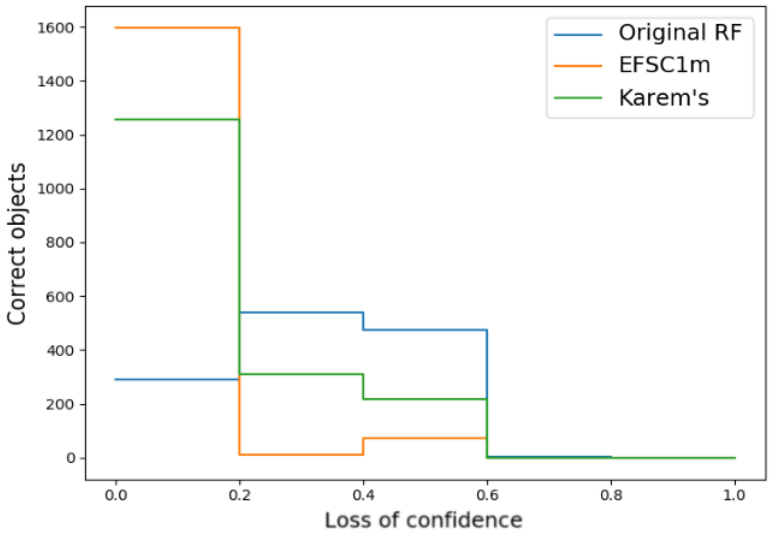


Figure 5.3. Correct objects in different intervals of loss of confidence of EFSC1m and Karem's on test area 1.

5.2.4 Robustness of the EFSC1m on mislabeled training samples

To further evaluate the effectiveness and robustness of the proposed framework, we test it on mislabeled training samples. This is a common problem in land cover classification, as labels and data are generally collected at different times. In this section, we start with a real case where labels were collected in 2011. However, the remote sensing data were collected in 2018 with very different surface circumstances. We evaluate the results on the latest NLCD labels published in 2016, which are quite similar to the reality in the field in 2018. Later, we manually modified the proportions of mislabeled training samples in each class by a uniform distribution to study the evolution of the results. We use the same method configuration as in section 5.2.3. The training sample configuration is detailed in Table 5.5 in which we display the training samples by class and show the proportion of mislabeled samples in the NLCD 2011.

Table 5.5 – Descriptions of training samples in the real case (M: proportion of mislabeled training samples) on test area 1.

	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Total
Training samples	32	257	135	239	47	36	82	828
M in NLCD 2011	0.218	0.377	0.148	0.230	0.979	0.611	0.707	0.368

We employ the NLCD 2011 with 36.8% incorrect labels as training samples for RF and showed the results in Table 5.6 where the best performance of each class is marked in bold. Karem's and EFSC1m still outperform EC3 with mislabeled training samples. EFSC1m can improve the overall accuracy by 13% in comparison to the original RF, which is more satisfying than Karem's and EC3. Due to 97.9% incorrect labels in *Grassland/Herbaceous*, its accuracy after fusion becomes 0 for Karem's and EFSC1m, but for EC3, the accuracy is kept as the original supervised results. For *Pasture/Crops* which has 61.1% incorrect labels, EFSC1m can still improve its accuracy by 6%. The confidence interval with confidence level of 95% of EFSC1m is $[0.597, 0.607]$, Karem's $[0.573, 0.582]$, EC3 $[0.472, 0.483]$. The experiments show that EFSC1m is significantly better than Karem's and EC3.

In the artificial case, we set the proportion of mislabeled training samples per class from 0.0 to 0.9, as shown in Figure 5.4. The accuracy of the original RF decreases slightly when $M \in [0, 0.3]$, where Karem's and EFSC1m have similar performance. When $M > 0.5$, the accuracy of Karem' method has obviously decreased but EFSC1m maintains some improvement. Compared to Karem's, EFSC1m can improve the overall accuracy by 20% at most. The results of EFSC1m decrease when $M > 0.7$, but they are still more satisfactory than Karem's. Apparently, the cautiousness of the EFSC1m makes it more robust in case the training samples are partially incorrect. EC3 has similar results to the original RF, which is less satisfactory in comparison to the other two methods initially. When $M > 0.7$, the accuracy of Karem's and EFSC1m becomes poorer than that of the original RF, whereas EC3 can keep the same. This can be explained by the fact that EC3 is based on consensus maximization, and therefore relies more on classification than clustering. We retrieve classical results obtained in information fusion.

Table 5.6 – Accuracy with the mislabeled training samples in the real case on test area 1.

Methods	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Accuracy
RF	**0.926**	0.234	0.468	0.524	**0.019**	0.301	**0.102**	0.473
Karem's	0.910	0.058	0.184	0.803	0.005	0.012	0.009	0.578
EFSC1m	0.893	**0.273**	**0.539**	**0.713**	0.008	**0.368**	0.079	**0.602**
EC3	**0.926**	0.235	0.493	0.535	**0.019**	0.289	0.100	0.478

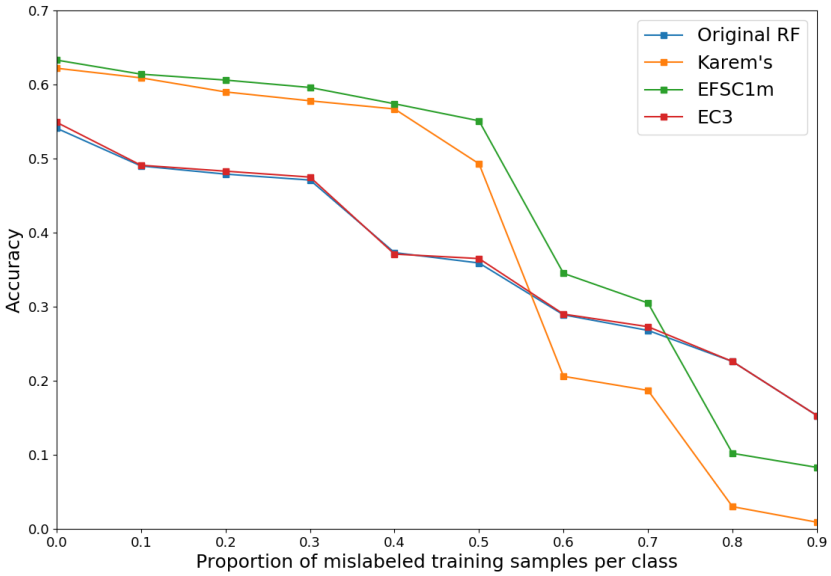


Figure 5.4. Accuracy with different proportions of mislabeled training samples per class in the artificial case on test area 1.

5.2.5 Combination of multiple classification and clustering methods

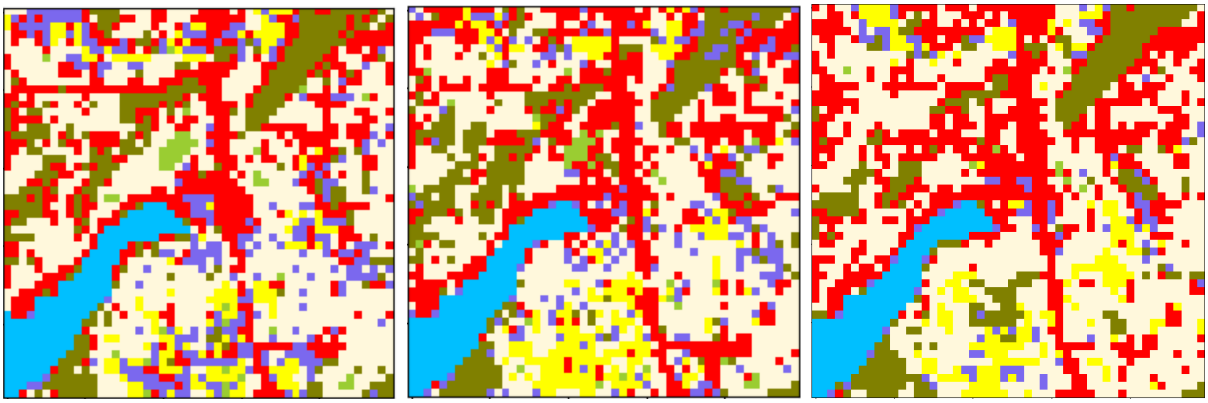
In this section, we evaluate the EFSCmm configuration and compare results with Karem's, EC3, and also the combination of the three classifications by Majority Voting (MV). We retake the three supervised classifiers: 5NN, RF, and SGB in section 5.2.2. For each classifier, we combine them with multiple clustering methods by EFSC1m configuration. In this way, the reliable information from each classifier is thus reinforced. Information extracted by fusion with multiple clustering methods, from different classification results, is discounted because their ignorance in BBAs usually is near to 0. In the experiment, we set this discounting coefficient as 0.8. Note that this coefficient is only used to fuse information by Dempster's rule. We apply the same configuration for Karem's as EFSCmm. As for EC3, we use it directly to combine the three classification methods with all clustering methods. Due to the limitation of space, we do not discuss the details of EFSC1m with 5-NN and SGB. Furthermore, we have already thoroughly studied EFSC1m with RF in the two previous experiments, which can indicate the effectiveness of the EFSC1m using 5-NN or SGB.

Figure 5.5 shows the results of the original classification of 5-NN, RF, and SGB in one experiment. Note that *Shrub/Scrub* are easy to classify as *Developed Areas*, and *Wetlands* is difficult to identify. For each classification results, we use EFSC1m to combine it with multiple clustering methods. Figure 5.6 displays Karem's method with each classification. Land covers such as *Developed Areas, Grassland/Herbaceous* and *Wetlands*, are

almost eliminated in the fusion results, and *Shrub/Scrub* becomes the dominated class. Karem's transformation cannot preserve weak information during an iterative process where the clustering results may be considered several times. Results of EC3 combining each classification with multiple clustering methods are presented in Figure 5.7 and show less improvement compared to the results of each original classification. We display the results of EFSC1m with 5-NN, RF, and SGB in Figure 5.8. The identification of *Developed Areas* is evidently improved and weak information such as *Wetlands* can also be well preserved after fusion with multiple clustering methods.

Table 5.7 – Accuracy of the original classifications (5-NN, RF, SGB) and four fusion methods (Karem's, EFSCmm, EC3, MV on test area 1.

Methods	Accuracy per class							Accuracy
	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	
5-NN	0.903	0.250	0.558	0.639	**0.142**	0.293	0.188	0.561
RF	0.922	0.277	0.518	0.609	0.095	0.214	**0.215**	0.541
SGB	**0.936**	0.240	0.557	0.635	0.017	0.277	0.133	0.553
Karem's	0.893	0.009	0.494	**0.818**	0.025	0.273	0.014	0.631
EFSCmm	0.854	**0.290**	0.525	0.801	0.111	**0.424**	0.030	**0.654**
EC3	0.917	0.267	0.475	0.654	0.112	0.122	0.192	0.563
MV	0.920	0.182	**0.597**	0.613	0.103	0.287	0.173	0.562

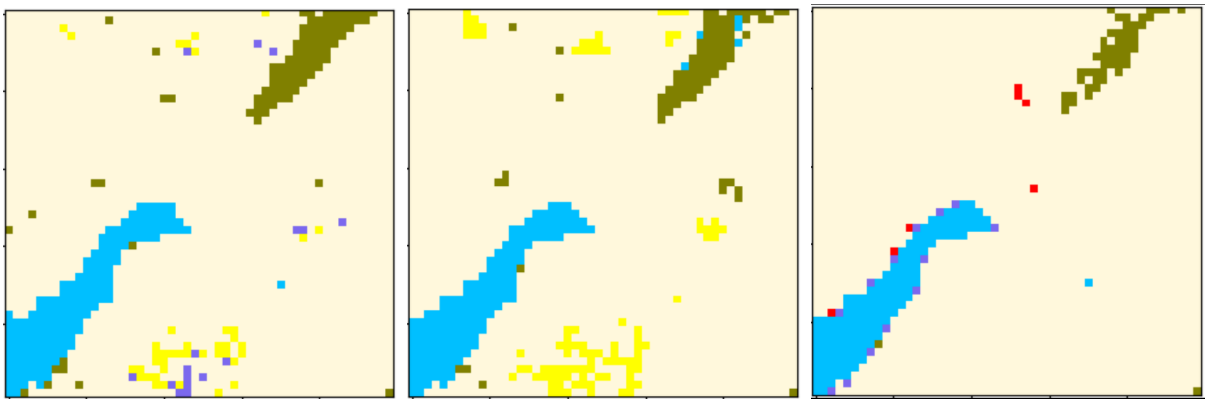


(a) 5-NN (b) RF (c) SGB

Figure 5.5. Original classification results of 5-NN, RF and SGB on test area 1.

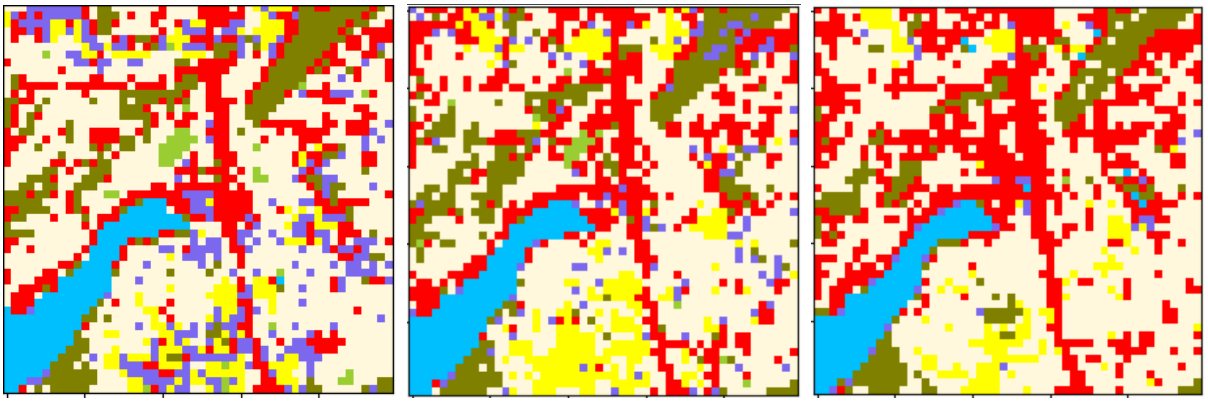
Results of combination with the three classification results and multiple clustering methods are detailed in Table 5.7 and Figure 5.9. The original accuracy of each classifier is also detailed for comparison. EC3 and MV have similar performances because the initial classification results are less reliable and the training samples are not sufficient. Both the EFSCmm and Karem's methods can effectively improve accuracy by combining several supervised and unsupervised methods. Karem's, which is less effective in distinguishing all

possible classes because its transformation of the BBAs only deals with uncertainty while ignoring imprecision. The transformation we propose carefully takes into account the uncertainty and imprecision in clustering. Therefore, compared to the other two methods, the advantage of the EFSC is to strengthen reliable information and also to prudently preserve unreliable information when merging with multiple clustering methods. The cautiousness of the EFSC makes it more effective when training samples are limited or with incorrect labels. We also calculate the confidence interval with confidence level of 95% and obtain $[0.650, 0.657]$ for EFSCmm, $[0.627, 0.634]$ for Karem's, $[0.557, 0.568]$ for EC3, and $[0.556, 0.567]$ for MV. EFSCmm is significantly better than other methods, and EC3 has similar performance as MV when classification and clustering results have poor qualities.



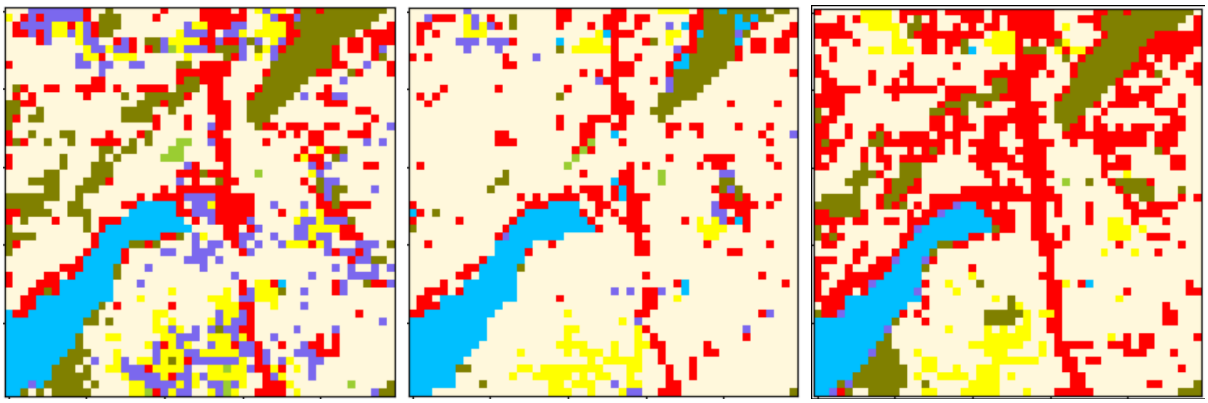
(a) Karem's with 5-NN (b) Karem's with RF (c) Karem's with SGB

Figure 5.6. Karem's with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1 on test area 1.



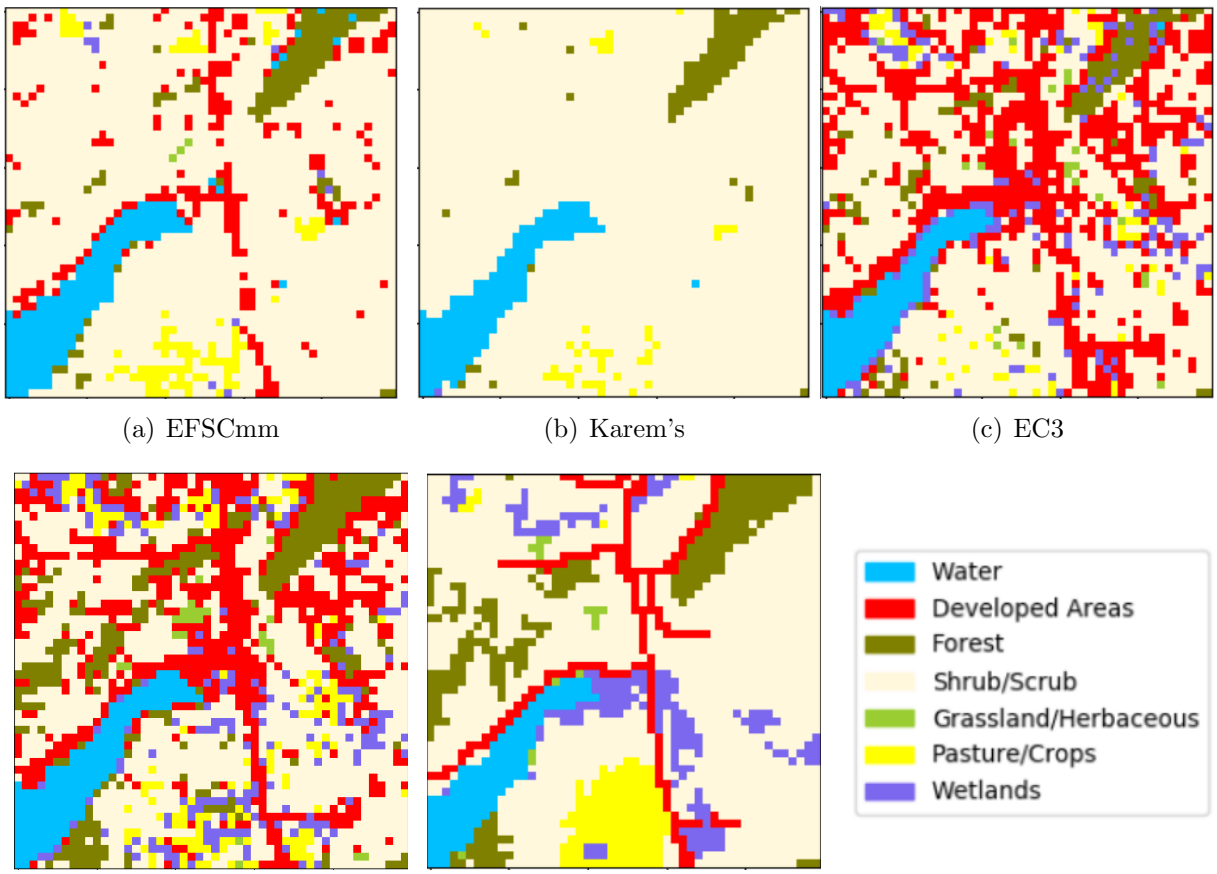
(a) EC3 with 5-NN (b) EC3 with RF (c) EC3 with SGB

Figure 5.7. EC3 with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1.



(a) EFSC1m with 5-NN (b) EFSC1m with RF (c) EFSC1m with SGB

Figure 5.8. EFSC1m with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1.

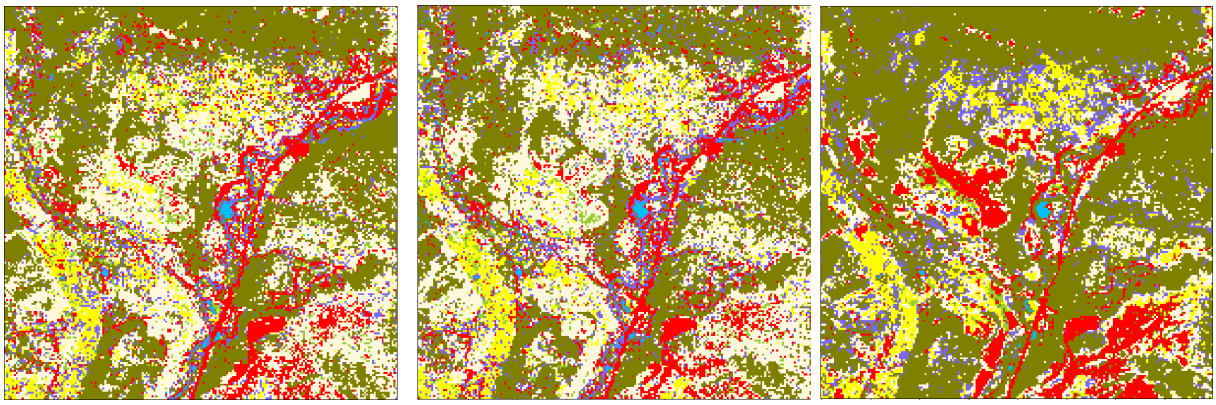


(a) EFSCmm (b) Karem's (c) EC3

(d) MV (e) Ground truth

Figure 5.9. EFSCmm, Karem's and EC3 with combination of multiple classification and clustering methods on test area 1.

We also evaluate the proposed EFSC on test area 2 which is 16 times larger than test area 1. The details and ground truth on test area 2 are shown in Table 5.1 and Figure 5.1. We display the combination results with multiple classification and clustering by EFSCmm, Karem's, and EC3. We also show the combination only for classifications by MV.



(a) 5-NN (b) RF (c) SGB

Figure 5.10. Original classification results of 5-NN, RF and SGB on test area 2.

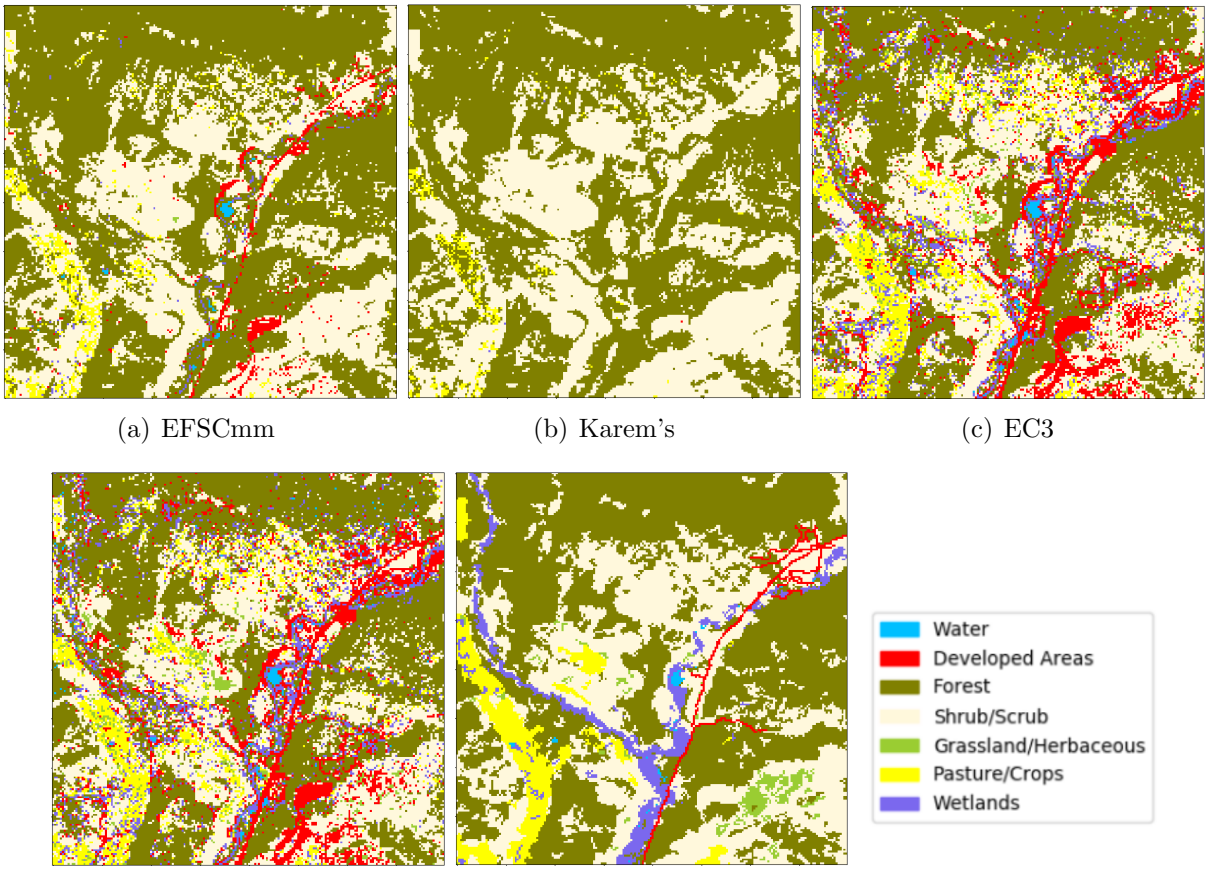
Table 5.8 – Accuracy of the original classifications (5-NN, RF, SGB) and four fusion methods (Karem's, EFSCmm, EC3, MV) on test area 2.

Methods	Accuracy per class							Accuracy
	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	
5-NN	0.281	0.121	0.730	0.523	0.116	0.377	0.162	0.593
RF	0.242	0.139	0.711	0.549	0.126	0.423	0.182	0.598
SGB	0.358	0.143	0.750	0.479	0.082	0.368	0.182	0.594
Karem's	0.	0.	0.783	**0.657**	0.	0.301	0.	0.653
EFSCmm	**0.430**	**0.229**	**0.793**	0.637	0.014	0.327	0.088	**0.664**
EC3	0.331	0.152	0.752	0.571	0.088	**0.438**	0.197	0.631
MV	0.318	0.146	0.754	0.559	**0.126**	0.428	**0.204**	0.624

The accuracy measured by F1 score for each method is detailed in Table 5.8. The classification results are shown in Figure 5.10 and the combination results in Figure 5.11. Test area 2 is dominated by *Forest* and *Shrub/Scrub*, and classes such as *Water* and *Developed Areas* are more unbalanced compared to test area 1. EFSCmm improves the accuracy by 7% at most and also outperforms other methods. The confidence interval with confidence level of 95% of EFSCmm is $[0.662, 0.665]$, Karem's $[0.651, 0.654]$, EC3 $[0.629, 0.632]$ and MV $[0.622, 0.625]$. The experiments show that EFSCmm is also significantly better than Karem's, and EC3 on test area 2.

One of the benefits of belief functions based methods is to correct the information in classifications by clustering. For the three classification results, they are all misclassified

parts of *Shrub/Scrub* as *Pasture/Crops*, which is difficult to correct by EC3 and MV. This is because, for EC3, only classification results can provide semantic labels so that it depends more on classification than clustering. Belief functions based methods, such as EFSC and Karem's, can transform clustering into the same frame as classification, consequently generating more information on semantic labels to combine.



(a) EFSCmm (b) Karem's (c) EC3

(d) MV (e) Ground truth

Figure 5.11. Combination of multiple classification and clustering by EFSCmm, Karem's, and EC3, and combination of classifications by MV on test area 2.

5.3 Evaluation on data provided by TOTAL

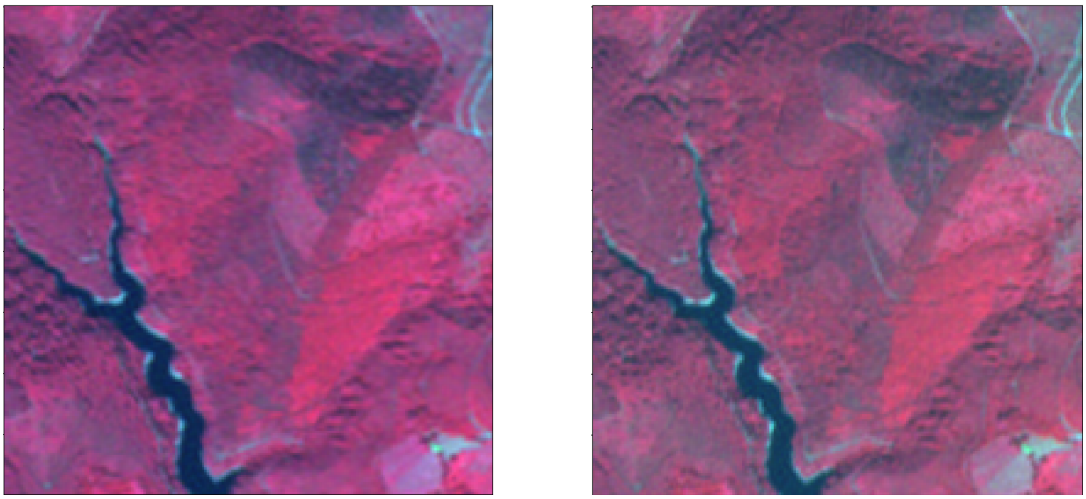
In real applications, besides the situation where limited labeled samples or partially mislabeled samples, as discussed above, it is also possible that no labeled samples are available at all. In this case, traditionally supervised methods cannot be applied to generate land cover maps. In the proposed EFSC framework, supervised results can be generated but not limited by classification methods. When no labeled samples are available, it is also possible to use knowledge-based identification methods, such as the automatic approach proposed in section 3.4, to generate the supervised results used in EFSC. In this section, we evaluate the proposed EFSC in a situation where no labeled samples are available. The

remote sensing data is from RapidEye and WorldView-2, provided by TOTAL. In the previous experiments, we have proved that EC3 is less effective than EFSC and Karem's methods and it is more suitable to combine a large number of classification results than clustering. Therefore, in this section, we merely focus on the performance of EFSC and Karem's methods.

As no ground truth is available in the study area, we use the automatic land cover identification method, proposed in section 3.4, as the classification results to combine with multiple clustering methods. The experiments are achieved both on the data of Rapid Eye and WorldView-2.

5.3.1 Experiments on the data of RapideEye

In this experiment, we use the study area located in north Spain, near to Arano, where two datasets of RapidEye acquired on different dates are available. We use the automatic land cover identification method on each datasets to obtain two land cover maps as different classification results. We select a test area with 200×200 pixels and generate its false composition image by three bands: NIR, *Red* and *Green*, as shown in Figure 5.12.



(a) datasets1: RapidEye data acquired on 22, AUG, 2016

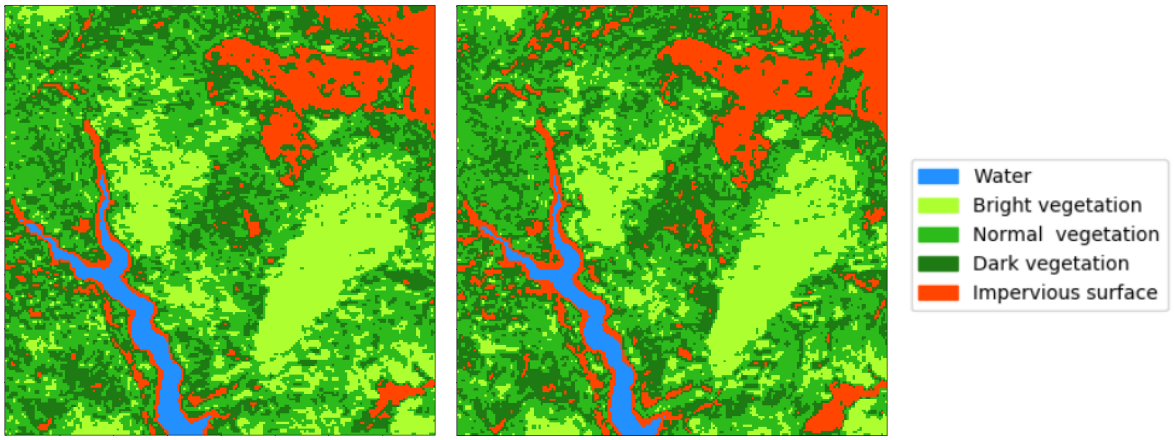
(b) datasets2: RapidEye data acquired on 26, AUG, 2016

Figure 5.12. False composition image of two RapidEye datasets.

We use the proposed automatic land cover identification method to generate the semantic land cover maps on the two RapidEye datasets, as shown in Figures 5.13(a) and 5.13(b). The land cover identification method can generate semantic labels at a basic level and refined level as well. The basic land covers includes *Water*, *Bare soil*, *Vegetation* and *Impervious surface*, which are identified by different spectral properties. For the refined land covers, *Water* and *Bare soil* are kept, whereas *Vegetation* and *Impervious surface* are refined according to their brightness when being necessary. In this test area, only *Vegetation* is refined as three sub-classes: *Bright vegetation*, *Normal vegetation*, and *Dark*

vegetation. Compared to the false composition images in Figure 5.12, we can observe that a river at the left side of the images can be well detected. The riverside is identified as *Impervious surface*. The classification of vegetation with different degrees of brightness is also accordance with the false composition images.

In this experiment, we use the refined land cover maps, as shown in Figure 5.13, as the classification results to combine with multiple clustering results. In the proposed automatic land cover identification method, fusion is achieved only for the identification of the basic land covers, so that its BBAs are on the frame of discernment including *Water*, *Bare soil*, *Vegetation* and *Impervious surface*. To obtain the BBAs on the refined land covers, *i.e*, *Water*, *Bare soil*,*Bright vegetation*, *Normal vegetation*, *Dark vegetation*, and *Impervious surface*, we assign the BBA of *Vegetation* to the union of *Bright vegetation*, *Normal vegetation*, *Dark vegetation* in the refined frame of discernment.



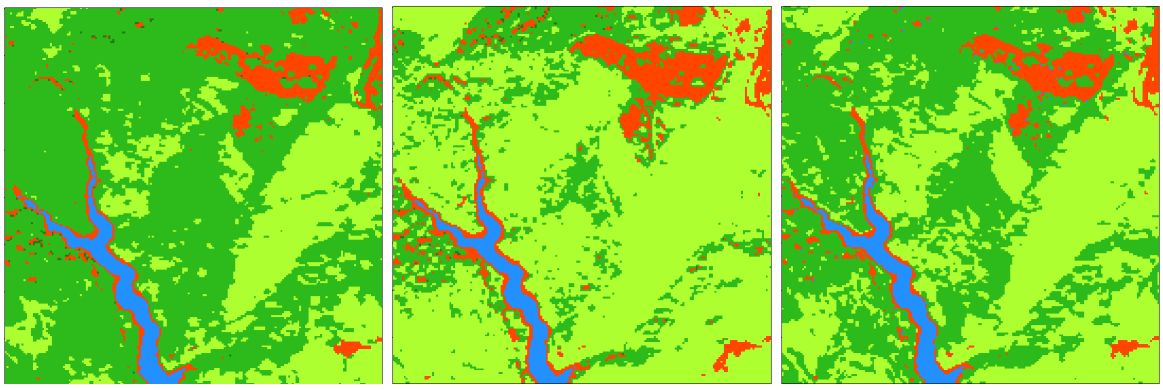
(a) Automatic land cover map on datasets1

(b) Automatic land cover map on datasets2

Figure 5.13. Automatic land cover maps on RapidEye datasets.

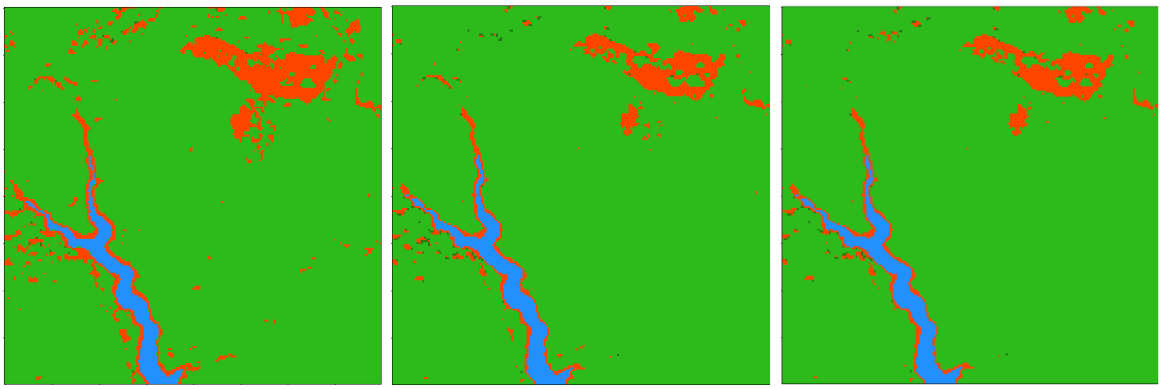
Figures 5.14(a) and 5.14(b) display the combination result of one classification and multiple clustering results by EFSC. We can observe that the class *Dark vegetation* is removed compared to the original land cover maps, because we only have supports on the union of vegetation, and no more evidence from clustering to support decisions on *Dark vegetation* either. The results of EFSC1m cannot preserve the detailed information of brightness detected by the automatic identification method. This is because the information to distinguish *Bright vegetation*, *Normal vegetation*, and *Dark vegetation* is not modeled in the BBAs of the supervised results, and they are only presented as a BBA value on the union of different vegetation. This shows the importance of the initial BBAs of supervised results. In this case, it is more interesting to directly construct BBAs for the final identification at the refined level instead of using the BBAs generated by Dempster's combination at the basic level. We can also observe that in the initial automatic land cover maps, some *Impervious surface* is detected inside vegetation areas. After combining with clustering results by EFSC1m, they are also labeled as vegetation. This may be because even though these small points are labeled as *Impervious surface* by the automatic

identification method, clustering results indicate they are in the same cluster as their neighborhood areas which are labeled as different types of vegetation. As the combination of EFSC1m takes more information on clustering results than classification results, these isolated *Impervious surface* areas are removed. Some areas which are evidently *Impervious surface*, such as the riverside, the areas at the up-right and bottom-right corners, are well preserved after combination. This is because both classification and clustering results support the decision as *Impervious surface* on these areas so that their labels do not change after combination. The combination of EFSC1m results from two different datasets is shown in Figure 5.14(c). Vegetation on the test area is finally separated into two categories by its brightness. EFSC also identifies *Water* and *impervious surface*.



(a) EFSC1m on datasets1 (b) EFSC1m on datasets2 (c) EFSC combination results on RapidEye

Figure 5.14. EFSC1m, and EFSCmm results on RapidEye datasets.



(a) Karem's method on datasets1 (b) Karem's method on datasets2 (c) Karem's combination results on RapidEye

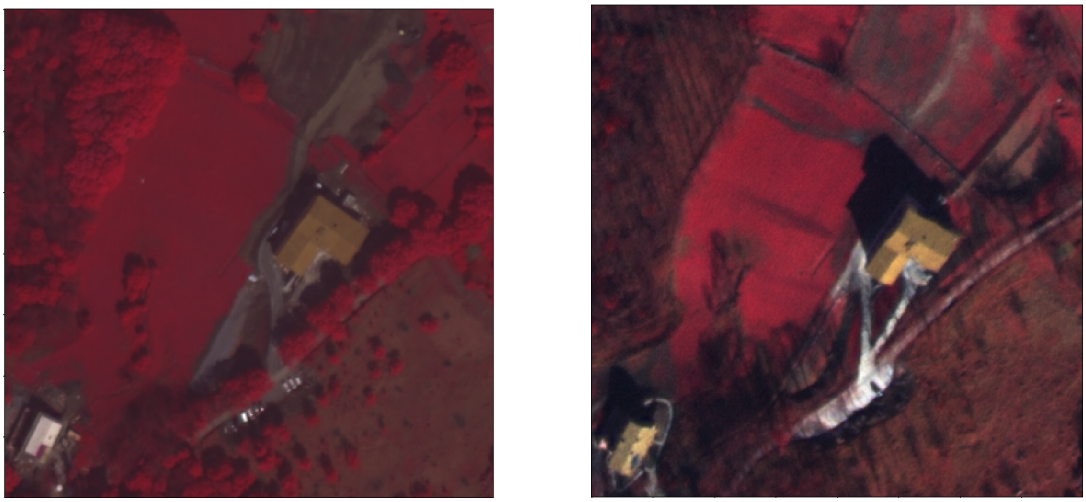
Figure 5.15. Karem's method on RapidEye datasets.

The combination results with one classification and multiple clustering methods using Karem's method are shown in Figures 5.15(a) and 5.15(b). We can observe that the details

on the brightness of vegetation are removed, and one type of vegetation is detected after the combination. For the results in Figures 5.15(a) and 5.15(b), there are still some areas are labeled as *Impervious surface*. We also observe the similar phenomenon of Karem's method in the previous experiments on synthetic data and on LandSat-8 data. This is because Karem's transformation is less cautious in comparison to the proposed one. Therefore, when combing with multiple clustering results, it is possible to only preserve the most agreed information of different baselines in each combination. This actually misleads the direction of the iterative process. Figure 5.15(c) shows the combined results with two land cover maps (one on datasets1, the other on datasets2) with multiple clustering methods using Karem's method. Compared to EFSC, Karem's method removes *Normal vegetation* and *Dark vegetation*, and only keeps *Bright vegetation*.

5.3.2 Experiments on the data of WorldView-2

In this experiment, we focus on the same study area located in north Spain. We also use the automatic land cover identification method on two datasets of WorldView-2 to obtain two land cover maps as different classification results. We select a test area with 400×400 pixels and generate its false composition image by three bands: NIR, Red and $Green$, as shown in Figure 5.16. For the image in Figure 5.16(a), it seems that the up-left corner has an area of trees, and we can also observe that some trees are planted near the building and along the road. For the image in Figure 5.16(b), we do not observe evident trees in those areas. Therefore, these two WorldView-2 datasets are highly collected in different seasons, because there are more trees in Figure 5.16(a), yet less trees in 5.16(b).



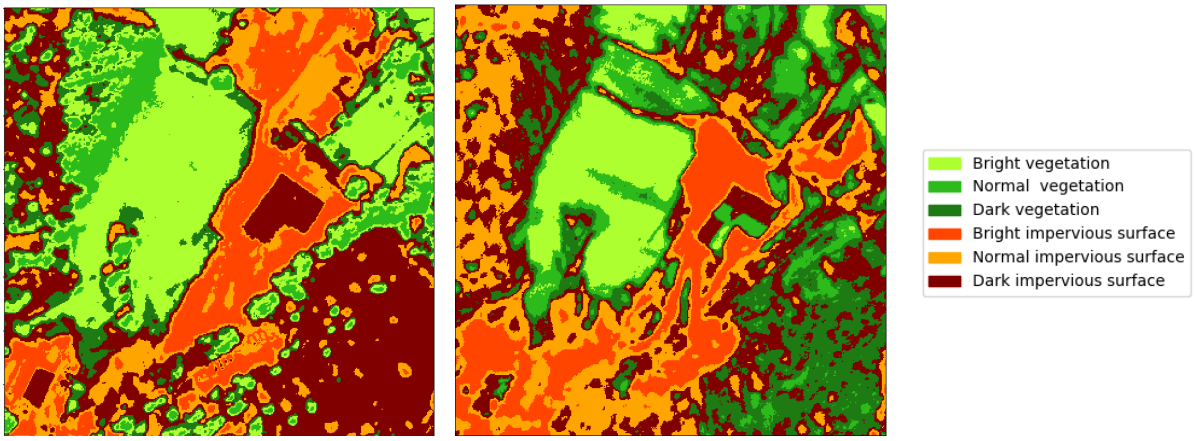
(a) datasets1: WorldView-2 data acquired in 2013

(b) datasets2: WorldView-2 data acquired in 2015

Figure 5.16. False composition image of two WorldView-2 datasets.

We use the proposed automatic land cover identification method to generate the semantic land cover maps on the two WorldView-2 datasets, as shown in Figures 5.17(a)

and 5.17(b). It seems that the automatic identification method has better performance on data acquired in 2013. In Figure 5.17(a), the building and roads are well detected. According to the false composition image in Figure 5.16(a), at the bottom-right corner, some trees can be observed, and there is also a large dark area we cannot well identify by eyes. These dark areas are identified as *Impervious surface* in Figure 5.17(a). In the up-left corner, we can observe that some areas are detected as *Impervious surface* and some are *Bright vegetation* or *Normal vegetation* in Figure 5.17(a). Compared to Figure 5.16(a), these *Impervious surface* areas seem to be the low vegetation and the trees nearby cast shadows on them. In Figure 5.16(b), we can observe that there are more evident shadows on many areas, which are detected as *Impervious surface*, *Normal vegetation* or *Dark vegetation* in Figure 5.17(b). As the automatic identification method is based on spectral indexes from multi-spectral data, shadows have great disturbances of land cover classification.



(a) Automatic land cover map on datasets1

(b) Automatic land cover map on datasets2

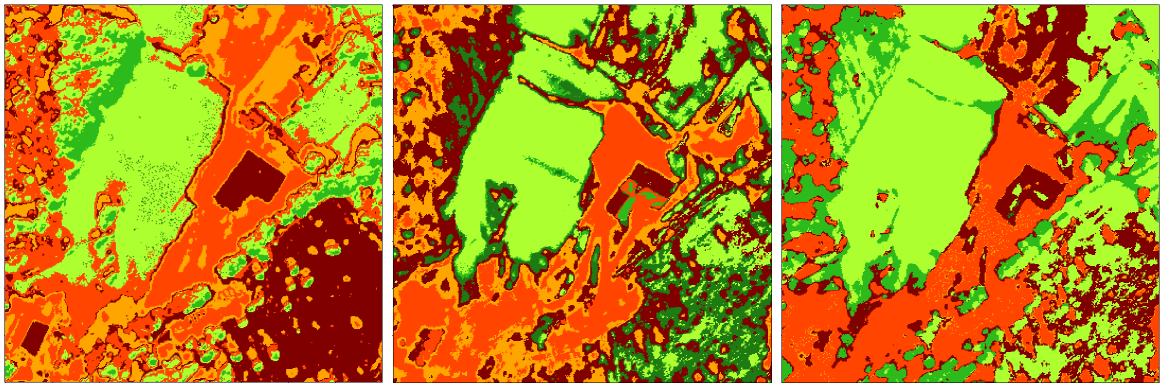
(c) Semantic labels

Figure 5.17. Automatic land cover maps on WorldView-2 datasets.

The process of the experiment is the same as RapidEye. Figures 5.18(a) and 5.18(b) display the combination result of one classification and multiple clustering results by EFSC. We can observe that combination with multiple clustering results can slightly improve the classification of *Impervious surface* in Figures 5.18(a) and 5.18(b), especially for datasets2. This is because removing the disturbances caused by shadows is difficult by the combination of merely multi-spectral information. For the datsets2, neither the automatic identification method nor clustering results can avoid the influences of shadows. Therefore, the combination by EFSC1m cannot evidently improve the classification accuracy. The combination of EFSC1m results from two different datasets is shown in Figure 5.18(c). Figure 5.19(c) shows the combination results with two land cover maps (one on datasets1, the other on datasets2) with multiple clustering methods using Karem's method. Compared to EFSC, Karem's method removes more details.

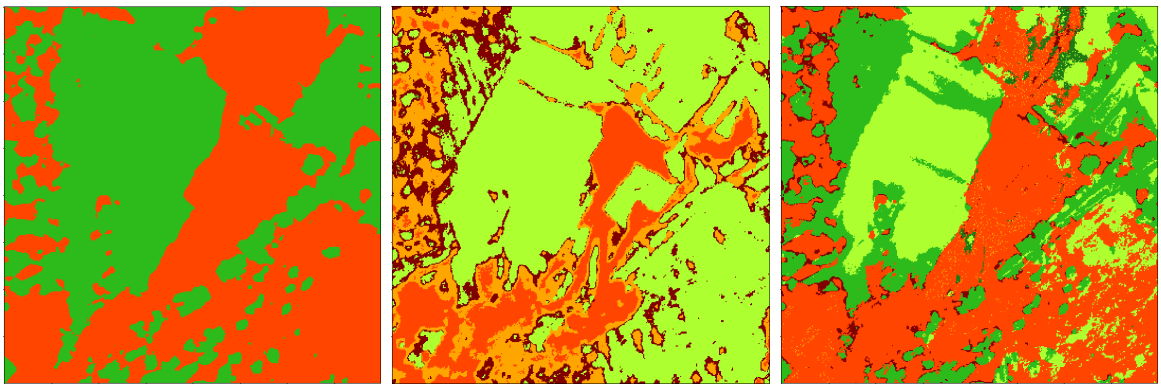
It seems that both EFSC and Karem's methods have a less satisfying performance on WorldView-2 data because of the shadows. For both EFSC and Karem's method, they

rely more on clustering results than supervised results. When some mistakes occur in supervised results, yet not in clustering results, the combination by EFSC and Karem's method is highly possible to correct these mistakes in supervised results. However, if the clustering results also make these mistakes, it is difficult to provide effective information for the combination. Therefore, in this experiment, to efficiently avoid influence of shadows, it is more pertinent to use segmentation/clustering/classification results from other senors, such as Radar, LiDAR, as clustering results in EFSC to combine with multispectral sensors. The combination results with one classification and multiple clustering methods using Karem's method are shown in Figures 5.19(a) and 5.19(b).



(a) EFSC1m on datasets1 (b) EFSC1m on datasets2 (c) EFSC combination results on WorldView-2

Figure 5.18. EFSC1m, and EFSCmm results on WorldView-2 datasets.



(a) Karem's method on datasets1 (b) Karem's method on datasets2 (c) Karem's combination results on WorldView-2

Figure 5.19. Karem's method on WorldView-2 datasets.

5.4 Conclusion

In this chapter, we focus on the performance of the proposed EFSC method on real remote sensing datasets.

The proposed framework EFSC can efficiently improve the classification accuracy when the information at the raw data level is limited (*i.e.* raw data is inaccessible, training samples are insufficient or partially incorrect). However, we can also notice that the combination results in WorldView-2 datasets are less satisfying. This is because the influence of shadows degrades the performance of the automatic land cover classification so that the supervised results used for combination may have extremely low accuracy. Moreover, the shadows change the spectral proprieties of land covers so that also have negative influences on clustering results. Evidently, to efficiently improve supervised results, good enough clustering results are prerequisite.

Combining supervised methods with clustering results helps to reduce the uncertainty of supervised results and to enhance reliable information. EFSC emphasizes the treatment of uncertain information to make prudent decisions. Different configurations of EFSC have been evaluated on LandSaT-8 OLI. In comparison with two other methods: Karem 's and EC3, EFSC has shown significantly better performance compared to Karem's and EC3. EFSC does not rely heavily on training samples. It seems more pertinent to improve the accuracy of the classification at the output level.

Chapter 6

Conclusion and perspectives

Resume

In this chapter, we conclude this thesis and present the perspectives related to our research. This thesis is first inspired by an industrial project of TOTAL, which focuses on the fusion of supervised and unsupervised methods in the land cover classification. As we explained in chapter 1, limited information on the ground usually brings obstacles to the applications of land cover classification. Therefore, in this thesis, we have proposed a fusion system for multiple sources, which can overcome limitations caused by insufficient ground information. The proposed framework, including a transformation and an iterative fusion strategy, has been evaluated on different datasets in the context of a combination of multiple supervised and unsupervised classifications. It can also be used for multiple sensor fusion at the output level and leave an interface to generate land use map by multiple land cover maps.

6.1 Conclusion

In the problems of land cover classification, a combination of multiple classification methods is proved to surpass any single model. However, limited information at the raw data level (*e.g.* inaccessible raw data, insufficient labeled samples, *etc.*) usually makes the low-level combination less effective, especially in the supervised context. Therefore, we have focused on the combination at the output level in the semi-supervised context. The benefits of combining multiple supervised and clustering results at the output level include: (1) the improvement of the overall accuracy when raw data is inaccessible; (2) the reduction of the uncertainty and imprecision caused by limited training samples; (3) clustering information to compensate limited information at the raw data level.

In the state-of-the-art part, we firstly reviewed some important methods in multispectral land cover classification, including pixel-based, subpixel-based, contextual-based, and object-based methods. We also presented some newly developing techniques, such as multiple classifiers systems and deep learning methods. By reviewing the developments of land cover classification techniques, we locate our research interest as multiple classifiers systems, because it has great potential to improve the accuracy of supervised results generated by insufficient labeled data. For multiple classifiers systems, we also reviewed the related researches from the aspects of information fusion and ensemble methods. Considering limited information at the raw data level, we narrowed our topic on the combination at the output level in the semi-supervised context. According to our knowledge, only limited researches have been done for this direction, and most of the existing methods are based on consensus maximization that requires high quality supervised results. However, in many real-life problems, collecting enough labeled samples is difficult, so that supervised methods usually have poor performance. Another limitation of consensus maximization-based methods is that clustering results are usually used as constraints so that the combination method cannot fully exploit the information from clustering. Therefore, we propose a new direction to solve the combination of supervised and clustering results in the framework of belief functions.

In our research, we manage to consider multiple supervised/clustering results as cognitively independent sources. The difficulty is that the results from different methods are usually heterogeneous so that they have different frames of discernment. To solve this problem, we proposed a transformation method to transfer BBAs of different results into the same frame to achieve the combination. Based on this transformation, we proposed an iterative fusion process that can cautiously yet fully take into account the information from clustering results. We evaluate the proposed framework with three different configurations: (1) combination of one supervised and one clustering results; (2) combination of one supervised and multiple clustering results; (3) combination of multiple supervised and clustering results. The experiments are achieved on different datasets, including synthetic data and real remote sensing data. In the experiments on synthetic data, we focus on how the qualities of supervised and clustering results affect the combination and also compare the proposed framework with other existing methods. We found that the proposed framework can take more information from clustering results into account, and can be less affected by the quality of supervised results. In the experiments on real remote sensing data, we evaluate how the quantity and quality of training samples affect the

combination results. We proved that the proposed framework can effectively improve the accuracy of supervised results generated by insufficient training samples. Moreover, the proposed framework is also robust and performs well with some mislabeled training samples. The comparison with other existing methods demonstrates that the proposed framework is more pertinent for the situation where information at the raw data level is limited. In real-life land cover classification problems, raw data information, such as labels, is usually difficult to collect, especially for complex nature areas. The proposed framework can combine heterogeneous information at the output level so that reduces the dependence on the information at the raw data level. Therefore, our method has a great potential for land cover classification of multi-sensors data.

Even though the proposed framework for the combination of supervised and clustering results performs well in the experiments, there are also many limitations and potential areas of improvement. In the transformation method, we only attempted to use some simple measures (*e.g.* Jaccard, Dice) as the likeness/similarity measures between a class and a cluster. It is also interesting to consider adaptive likeness/similarity measures based on limited training samples. In the iterative fusion process, we propose a strategy to cautiously combine the information from new sources. The criterion to select information is based on the loss of confidence, which also needs to be further studied in other applications except for land cover classification. Furthermore, the experiments to evaluate the proposed framework are all based on images/remote sensing data. It is also important to study its performance on other types of data. Our research is originally inspired by the needs of TOTAL, to generate a land use map (accessibility map) by multiple land cover maps. Our proposed transformation method can transfer land cover maps on different frames of discernment into the same one which can be a land use scheme defined by users. However, there are still considerable efforts that need to paid to generate an accessibility map with uncertainty and to optimize the shortest path of displacement. We give more discussion in section 6.2.

6.2 Perspectives

6.2.1 Multiple sensor data fusion at the output level

Multiple sensors are available in METIS, such as multi-spectral sensors, LiDAR, Radar, DEM, and each of them has different formats of information at the raw data level. Moreover, each sensor has its specific processing techniques and usually generates different land cover maps. For example, multi-spectral sensors can take advantage of various spectral indexes, as presented in chapter 3 and usually separate landscape as water, different types of vegetation, and surfaces. LiDAR can be applied to generate vegetation density maps. Radar gives information on the textures and DEM represents the elevation of different objects on the ground.

It is thus neither easy to directly combine them at the raw data level, nor to develop a classification method to fully take advantage of the heterogeneous information. For different sensors, it is more pertinent to select their appropriate data processing techniques and methods to generate different land cover maps. Therefore, a combination at the

output level is more suitable to meet needs in METIS, because it can make full use of information from multiple sensors, as explained in chapter 1.

We have proposed an evidential fusion method to combine several supervised and unsupervised classifications. This method can be applied with or without accessing raw data. When raw data is available, the BBAs of supervised classification can be constructed based on information from training samples. Moreover, clustering results and their BBAs can also be generated by raw data. When raw data is not available, the BBAs of supervised and unsupervised classification can also be defined by users.

When combining multiple heterogeneous land cover maps, we can only keep the land cover maps that we are interested in as supervised results. For the land cover maps whose labels are less interesting, we can remove its semantic labels and consider them as clustering results. Their BBAs can be either constructed by users, as explained in chapter 4 or according to raw data, as presented in chapter 5. The proposed framework includes two major parts: (1) a transformation to transfer different land cover maps in the same frame of discernment; (2) an iterative fusion strategy to fully and cautiously take other land cover maps (or clustering results) into account. When applying the proposed framework in the combination of multiple land cover maps from different sensors, there are many ways to construct BBAs of land cover maps. For example, in the experiments on synthetic data, we use random BBAs on all singletons and unions. We can also use the probability after discounting as BBAs. Our future work will focus on how different constructions of BBAs affects the combination in the proposed framework. There is no universe construction of BBAs in the proposed framework, and the combination results can also be affected by the construction of BBAs in different situations. Therefore, we will consider generating the most appropriate BBAs of supervised classification results by the learning process when using the proposed framework.

Likeness/similarity is another important parameter in the proposed framework. It essentially decides the BBAs after transformation and thus has a significant influence on the final combination results. We have already evaluated different measures of likeness/similarity, including proportion likeness, and also some similarity: Jaccard, Dice, and Recovery. It seems that different measures of likeness are suitable for different cases, and no universal likeness/similarity measure that surpasses others exists. Therefore, in our future work, we also consider to generate the measures of likeness/similarity by active learning process with limited training samples, to make the BBAs after transformation are the most suitable to combine with the given information.

In the proposed framework, the iterative process selects information of each class by making the partial loss of confidence decreasing. However, having more confidence in a decision does not guarantee the decision is closer to the ground truth. This is because when constructing the BBAs of supervised results, a good decision may have large uncertainty, whereas a bad decision may have high confidence. Therefore, it is important to have good BBAs of supervised results which should be well correlated with the ground truth. In our future work, we will also discuss more on different strategies to combine information, and also consider to select information by learning.

6.2.2 Generate land use map by land cover map

For the multi-sensor fusion system used in METIS, the inputs are different land cover maps, whereas the output is supposed to be a land use map (accessibility map). It is difficult to directly generate an accessibility map by the traditional machine learning method because the definition of accessibility is still ambiguous at the current stage. Moreover, the accessibility probably does not correspond to physical or spectral properties of land covers on the ground, as we explained in chapter 1.

In our future work, we will consider defining a frame of discernment of accessibility, so that we can transform other land cover maps by the proposed transformation to combine the information on this target frame of discernment. In our current research, the likeness/similarity is measured by the relationship of two given sets. That is to say, it requires two classification results (supervised or unsupervised) to calculate the likeness/similarity between classes from different results. However, as we explained in the previous section, the measures of likeness/similarity can be considered to generate by the learning process. In this way, the transformation can be achieved when the target frame of discernment is given and does not require supervised or unsupervised classification results in this frame.

For the accessibility map, the measures of likeness/similarity can also be generated by ground surveys. For a specific study area, we can discover whether an evident correlation exists between different land covers and the accessibility, which can be used to generate values of likeness/similarity by active learning. With the help of the likeness/similarity, land cover maps generated by different sensors can finally be transformed and combined on the frame of the accessibility. Another of the benefits of the proposed framework is to generate the corresponding loss of confidence for the final combination results. In other words, the uncertainty and imprecision in the original land cover maps are transferred as the loss of confidence for the land use map. Therefore, the output of this fusion framework is a soft accessibility map that can be used for obtaining the most effective path of displacement.

6.2.3 Accessibility model based on multi-factors

Evaluating accessibility on the ground is also a great challenge. Most evaluations are based on the cost of energy. In the natural environment, vegetation is one of the major factors impeding human movement on the ground. However, unlike slope and surface type, the whole vegetation component is always assigned with the same accessibility. Since the change of species of vegetation may indicate the change of accessibility, it is not practical to ignore the difference of vegetation in the measurement of accessibility.

Furthermore, the density of different vegetation is rarely considered during the evaluation of accessibility, which, however, has a great influence on human movement. Walking in the dense forest can help people avoid too much exposure in the sunlight while the dense shrub is not suggested to cross. Thus, we have to take the density and species of vegetation into consideration in our model.

Instead of measuring accessibility for each pixel, we consider evaluating the cost on the set of pixels, in which pixels are gathered with their similar neighbors. The set of similar pixels can be considered as one component in the evaluation of accessibility, which

could reduce the computation on the one hand and simplify the optimization step on the other hand.

Moreover, if we have information about the occurrence of dangerous animals, such as crocodiles and bears, we can mark the level of danger in different locations and generate warnings if a danger zone is easy to access.

6.2.4 Optimization under uncertainty

The proposed fusion framework can be used to generate a soft accessibility map in which the loss of confidence indicates the quality of decision on the accessibility of a certain area. For example, the accessibility map can display that some area is easy to pass yet this information is highly unreliable (*i.e.* the corresponding loss of confidence is high). To obtain the most effective path of displacement, an optimization process has to be considered, taking into accounting uncertainty about accessibility (*e.g.* loss of confidence). Considering different criteria may generate different accessibility on the same location, the optimization model is supposed to modify the importance of different criteria according to the real situation when generating the strategy of displacement.

The optimization model to generate the most effective path can be considered as a Markov Decision Process (MDP), which consists of a set of locations, a set of actions, and the expected rewards received at the transition from one location to another. An optimal action-selection policy for the Markov decision process can be solved by reinforcement learning. This technique is used for solving problems that can be described as follows: an agent living in an environment is supposed to find an optimal behavioral strategy while perceiving only limited feedback from the environment. The agent receives (not necessarily complete) information on the current state of the environment, can take actions, which may change the state of the environment, and receives reward or punishment signals. The goal of reinforcement learning is to find a policy to maximizes the long-term reward.

Since the initial uncertainty and imprecision in land cover classification are all transferred into the frame of accessibility, it is important to consider how to synthesize uncertainty in the optimization model. For example, the rewards at the transition can be constructed by considering the uncertainty of accessibility and also some ground information. The transition probability in MDP can be enlarged to uncertainty and imprecision. Additionally, the iteration process in reinforcement learning can be also considered through the view of the theory of belief functions.

Publications

Journals

1. Na Li, Arnaud Martin, and Rémi Estival. Heterogeneous information fusion: Combination of multiple supervised and unsupervised classification methods based on belief functions. *Information Sciences*, 544, 238-265, 2020.

International Conferences

1. Na Li, Arnaud Martin, and Rémi Estival. An automatic land covers identification based on Dempster-Shafer theory for multi-spectral images. In *IGARSS 2019*, Yokohama, Japan, July 2019.

2. Na Li, Arnaud Martin, and Rémi Estival. Combination of supervised learning and unsupervised learning based on object association for land cover classification. In *DICTA 2018*, Canberra, Australia, December 2018.

3. Na Li, Arnaud Martin, and Rémi Estival. An automatic water detection approach based on Dempster-Shafer theory for multi spectral images. In *20th International Conference on Information Fusion*, XI'AN, China, July 2017.

Résumé

La classification de la couverture terrestre des sols est essentielle pour la planifcation et la gestion des ressources naturelles (développement, protection, *etc.*), la compréhension de la répartition des habitats ainsi que la modélisation de l'environnement. Souvent, un grand nombre d'échantillons étiquetés est une condition préalable à la classification de la couverture terrestre des sols. Cependant, la collecte d'informations suffisantes sur le terrain, en particulier dans les zones peu accessibles, est généralement difficile et coûteuse. Récemment, la combinaison des résultats de la classification et du clustering a fait l'objet d'une attention croissante pour améliorer la précision de la reconnaissance de la couverture terrestre. Cette problématique s'inscrit dans le cadre du projet industriel METIS (Multiphysics Exploration Technology Integrated System), initié par la société TOTAL. Ce projet vise à développer un système géophysique et logistique pour acquérir, traiter et obtenir des informations structurelles et quantitatives sur la topographie souterraine en temps réel. Un des aspects du projet consiste à estimer l'accessibilité du terrain à partir d'images de télédétection, afin de réduire les coûts et les risques, grâce à une opération de planification qui tient mieux compte du terrain et des impératifs environnementaux. C'est sur cette application qu'ont porté les travaux exposés dans cette thèse.

La combinaison des résultats de la classification et du clustering peut contribuer à réduire l'incertitude des informations de manière à améliorer la qualité de la caractérisation de régions difficiles d'accès et sur lesquelles l'étiquetage d'échantillons d'apprentissage et de vérité terrain est délicate voire impossible. Bien que ce problème ait déjà reçu de nombreuses solutions dans le cas d'images usuelles, notamment pour les classes d'objets utilisées en analyse de scènes, nous choisissons d'exploiter des méthodes robustes aux données imprécises et possiblement mal étiquetées. Nous proposons de nous appuyer sur le formalisme des fonctions de croyance pour construire une méthode capable de combiner les informations issues de différents algorithmes de classification des pixels ou de segmentation sémantique (les labels sémantiques correspondant *a priori* aux classes recherchées) et de classification des pixels (labels correspondant à des classes non supervisées). Ses contributions se situent dans la mise en correspondance des différents ensembles de classes (supervisées ou non-supervisées) et des croyances associées, et dans la mise en œuvre d'un algorithme de fusion pour la cartographie de l'occupation du sol à partir de données de télédétection.

Introduction

L'identification des types de couverture terrestre fournit des informations de base pour la production d'autres cartes sématiques et établit une base de référence pour les activités de surveillance. Par conséquent, la classification de la couverture terrestre à l'aide de données satellitaires est l'une des applications les plus importantes de la télédétection. Une grande quantité d'informations au sol (échantillons étiquetés) est généralement nécessaire pour obtenir une classification de la couverture terrestre de haute qualité. Toutefois, dans les zones naturelles complexes, la collecte d'informations au sol peut être longue et extrêmement coûteuse. De nos jours, les technologies à capteurs multiples font l'objet d'une grande attention dans la caractérisation de la couverture terrestre. Elles apportent des informations différentes et complémentaires des caractéristiques spectrales qui permettent d'aider à surmonter les limitations causées par une information au sol inadéquate.

Un autre problème causé par le manque d'informations au sol est l'ambiguïté des relations entre les cartes de la couverture des terres et les cartes d'utilisation des terres. Les cartes de l'occupation des sols fournissent des informations sur les caractéristiques naturelles qui permettent d'être directement observées à la surface de la Terre. Elle font également référence à la manière dont les gens utilisent les informations sur les paysages à des fins différentes. Sans informations adéquates sur le terrain, il est difficile de produire des cartes d'utilisation des sols à partir des cartes de l'occupation des sols pour des zones complexes. Par conséquent, lorsque l'on combine plusieurs cartes hétérogènes de la couverture des sols, il faut envisager comment permettre aux utilisateurs de synthétiser les cartes d'utilisation des sols.

Dans notre recherche, nous nous concentrons sur la fusion d'informations hétérogènes provenant de différentes sources. Le système de combinaison vise à résoudre les problèmes causés par le nombre limité d'échantillons étiquetés et peut donc être utilisé dans la caractérisation de la couverture des terres pour les zones difficiles d'accès. Les étiquettes sémantiques pour la classification de l'occupation des sols provenant de chaque capteur permettent d'être différentes et permettent de ne pas correspondre au schéma (tous les types de couverture terrestre possibles) final d'étiquettes que les utilisateurs attendent. Par conséquent, un autre objectif de la combinaison est de fournir une interface avec un schéma final probablement différent des cartes de l'occupation des sols d'entrée.

Traitement de l'insuffisance des données labellisées Le problème de l'accessibilité survient généralement dans les zones difficiles d'accès comme les zones de montagne, ce qui rend les relevés au sol extrêmement compliqués. La collecte de données labellisées dans cette situation est généralement si coûteuse, si risquée et si longue qu'il est impossible de recueillir suffisamment de données labellisées. Outre le processus d'étiquetage manuel au sol, les photos aériennes et les images capturées par les drones permettent également d'aider à obtenir des étiquettes sémantiques. Toutefois, ce processus est généralement coûteux dans les applications du monde réel, de sorte que les couvertures terrestres sont simplement étiquetées comme étant des zones de végétation, de l'eau, des sols nus, etc. Les détails plus complexes, liés au type de végétation, à la densité, à l'humidité du sol et à la situation du terrain, sont généralement difficiles à décrire manuellement à partir de simples photos aériennes. Il est également possible d'obtenir des labels erronés au cours

de ce processus.

Il est donc souvent difficile de collecter des données labellisées suffisantes et de haute qualité en raison des limitations techniques ou économiques. Les résultats de classification générés par une quantité limitée de données d'entraînement sont généralement incertains et incorrects. Les limitations causées par l'insuffisance de données étiquetées sont l'un des principaux thèmes de cette thèse. Compte tenu du contexte spécifique de notre recherche, nous proposons quelques orientations pour remédier aux problèmes causés par l'étiquetage des données.

La fusion de plusieurs cartes de la couverture terrestre est un moyen efficace d'améliorer la précision globale de la classification de la couverture terrestre lorsque chaque carte de cette dernière présente de mauvaises performances. Il est important de noter que l'un des avantages des processus de fusion est de surmonter certaines limites dues à l'insuffisance des données étiquetées et d'améliorer la précision dans une certaine mesure.

En outre, nous nous considérons l'incertitude et l'imprécision pour obtenir un cadre de fusion souple. L'incertitude indique qu'un objet appartient à une certaine classe de couverture terrestre avec un certain degré de croyance. L'imprécision décrit que certains objets pourraient appartenir à l'union de deux ou plusieurs classes de couverture terrestre avec un certain degré de croyance. Lorsque les données étiquetées sont insuffisantes, nous parvenons à tirer parti des informations issues de l'incertitude et de l'imprécision. Compte tenu de la faible précision de chaque carte d'occupation des sols, il est plus raisonnable de ne pas décider de l'étiquette d'un objet de manière nette. C'est pourquoi nous nous concentrons sur un cadre de fusion souple pour gérer l'incertitude et l'imprécision.

Génération de cartes des utilisations des terres par les cartes de la couverture des terres L'utilisation et l'occupation des sols sont deux systèmes de classification couramment utilisés pour décrire les circonstances d'un lieu donné. Ces systèmes sont souvent confondus en tant que termes interchangeables dans les applications de télédétection. Cependant, ils décrivent des caractéristiques différentes du paysage et ont des usages différents dans la réalité. Les cartes de l'occupation des sols fournissent des informations sur les caractéristiques de la nature (*e.g.* végétation) ou les constructions artificielles (*e.g.* bâtiments) qui permettent d'être directement observées à la surface de la Terre. Par exemple, une carte de l'occupation des sols peut documenter la façon dont une région est couverte par des types de terres physiques, y compris l'eau, les forêts, les zones humides, les surfaces imperméables et d'autres couvertures terrestres. Les cartes de l'utilisation des terres font référence à la manière dont les gens utilisent les informations sur les paysages à différentes fins, comme l'habitat de la faune, la conservation, etc. Les cartes de l'utilisation des terres ne décrivent pas directement la couverture de surface sur le terrain alors qu'elles sont liées à la couverture des terres dans une certaine mesure.

Les données sur l'occupation des sols sont généralement interprétées à partir d'images satellites de télédétection et/ou de photos aériennes car elles sont simplement liées aux caractéristiques physiques de la surface. En revanche, les données sur l'utilisation des sols permettent de parfois être déterminées à partir d'images satellites ou de photos aériennes, mais ce n'est généralement pas le cas. Les cartes d'occupation des sols intègrent souvent des données auxiliaires, telles que l'altitude, les informations topographiques et les relevés au sol, car les informations utiles sont difficiles à interpréter à partir des images. Les

classes contenues dans une carte d'utilisation des terres/couverture des terres sont appelées schéma dans la classification de la télédétection et elle fait également référence au cadre de discernement dans les fonctions de croyance. Chaque classe du schéma d'utilisation et de couverture des terres ne doit pas être contradictoire ou se chevaucher dans ses définitions. Dans le cadre de notre recherche, le système de fusion proposé devrait pouvoir obtenir un schéma de couverture des terres défini par l'utilisateur qui peut être différent de celui des cartes de couverture des terres d'entrée.

Fusion d'informations hétérogènes Les informations provenant de différents capteurs tels que les satellites multispectraux, le SAR ou le LiDAR sont très différentes. Il est donc essentiel de combiner les informations hétérogènes provenant de plusieurs capteurs. Contrairement au système de fusion traditionnel qui est conçu pour améliorer la précision de la classification de la couverture des terres, nous nous concentrons également sur la génération de cartes d'utilisation des terres qui sont différentes des cartes de la couverture des terres. Toutes les informations fournies par les capteurs et les relevés au sol sont utilisées pour obtenir des cartes de la couverture terrestre. Dans les applications de la vie réelle, il existe un écart évident entre ce que les utilisateurs recherchent et ce que les capteurs permettent de fournir. C'est pourquoi, dans cette thèse, nous nous concentrons sur le développement d'un système de fusion qui non seulement combine des informations hétérogènes, de sorte que les classes d'utilisation des terres puissent être définies par les utilisateurs après la fusion.

Les principales difficultés du système de fusion sont la fusion des cartes de classification de différents schémas d'occupation des sols dans un même schéma. Tous les capteurs n'utilisent pas le même schéma de couverture des terres pour établir leurs cartes de couverture. Il est donc toujours possible d'avoir quelques conflits dans les différents schémas fournis. Dans ce cas, les cartes de couverture du sol sont également des informations hétérogènes. De plus, les classes d'utilisation des terres sont différentes de tous les schémas de couverture des terres des capteurs en raison de la définition des utilisateurs. Le système de fusion doit tenir compte de ces différences, traiter des échantillons de information insuffisants et de mauvaise qualité. La Figure R.1 décrit la vue d'ensemble du système de fusion sur lequel nous nous concentrons. Le système peut réaliser la combinaison de données provenant de différents capteurs et plus important encore, le système de fusion peut obtenir une carte de l'utilisation des terres en fonction des besoins des utilisateurs.

Dans ce travail de thèse, nous étudions un système de fusion qui est censé être indépendant de la source, ce qui indique que nous nous concentrons davantage sur les cartes de la couverture terrestre provenant de différents capteurs plutôt que sur les méthodes de traitement et de classification des données elles-mêmes. Les différentes sources/capteurs (*e.g.* SAR, LiDAR, capteurs multispectraux) capturent des informations distinctes sur le terrain et ont leurs propres techniques de traitement des données. De plus, elles appliquent des méthodes de classification spécifiques pour obtenir des cartes de la couverture terrestre. Les entrées du système de fusion sont des cartes de la couverture terrestre différentes au lieu de données brutes initiales, de caractéristiques intermédiaires ou de méthodes de classification elles-mêmes. Cela signifie que le système de fusion traite directement les informations de haut niveau, les cartes de la couverture

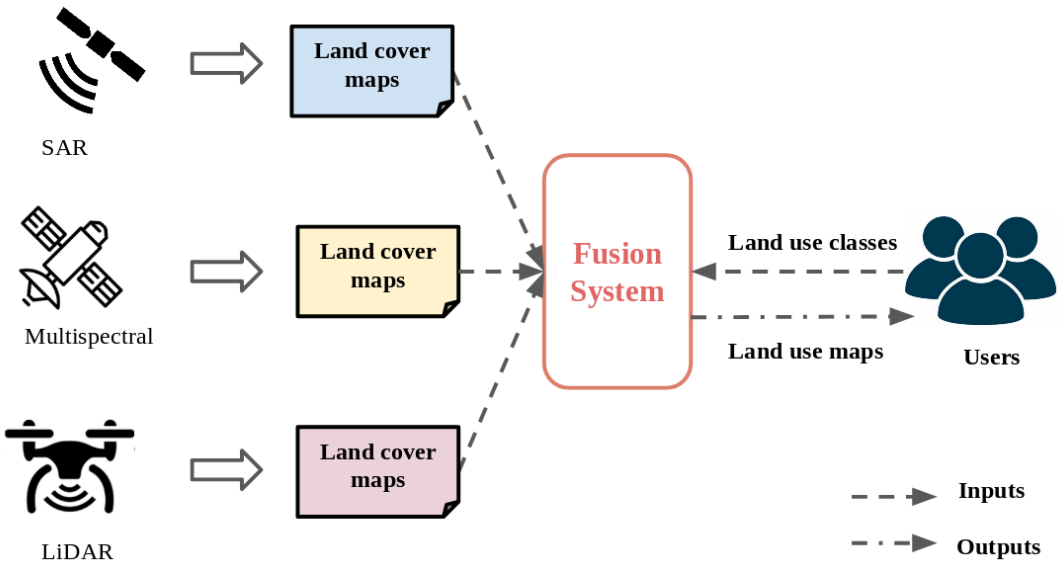


Figure R.1. Vue d'ensemble du système de fusion.

terrestre, et ignore les informations de bas niveau, les types de capteurs, les données, les caractéristiques et les classificateurs. Par conséquent, nous utilisons les cartes de la couverture terrestre générées par différents classificateurs fondés sur des capteurs multispectraux comme sources différentes, comme le montre la Figure R.2.

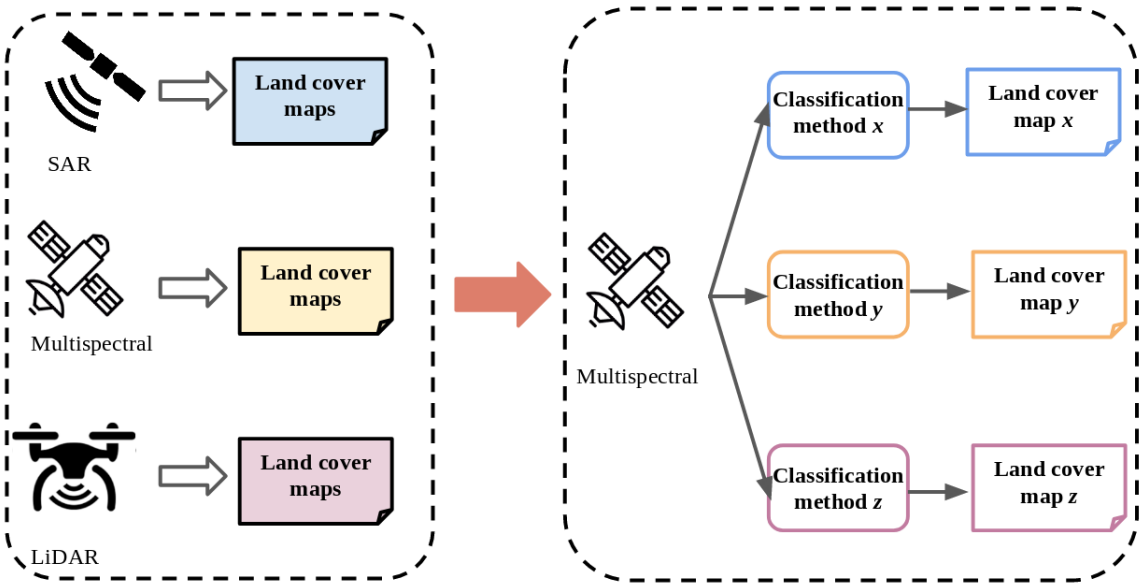


Figure R.2. Fusion des cartes de la couverture terrestre de plusieurs capteurs à la fusion des cartes de la couverture terrestre de plusieurs classificateurs fondés sur des données multispectrales.

Contributions

Dans cette thèse, nous nous intéressons à la fusion d'informations hétérogènes fondée sur la théorie des fonctions de croyance. Cette théorie, également appelée théorie de Dempster-Shafer, est une méthode puissante pour traiter l'incertitude et aussi l'imprécision. Les fonctions de croyance, considérées comme une généralisation de la théorie bayésienne, permettent de combiner les informations existantes avec de nouvelles informations qui pourraient être incertaines.

Dans le domaine des fonctions de croyance, la principale contribution est:

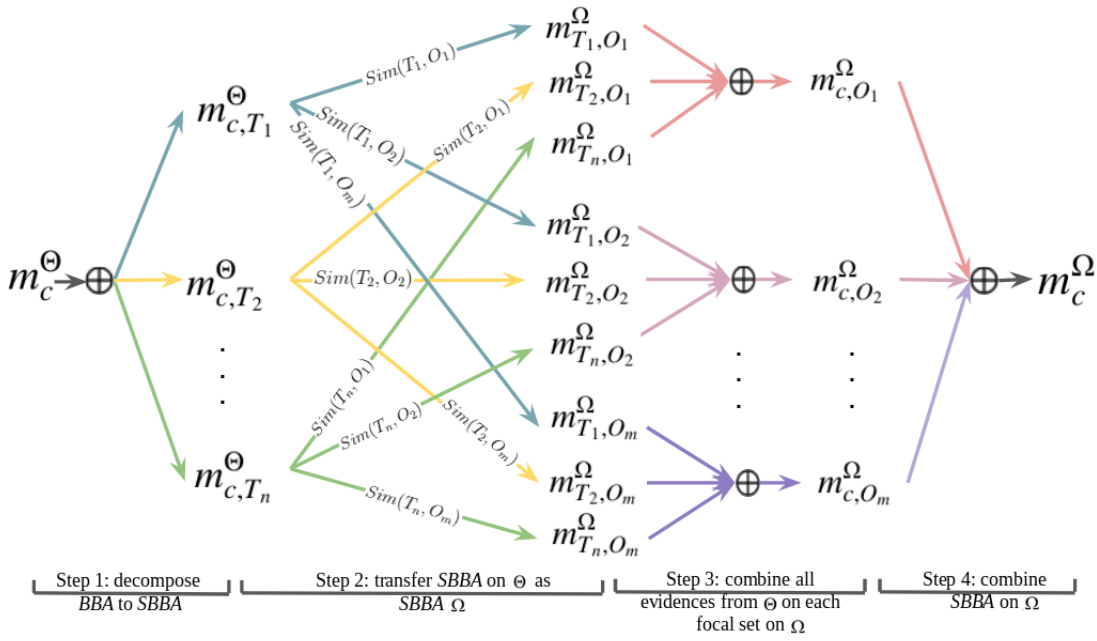


Figure R.3. Méthode de transformation pour changer le BBA dans le cadre du discernement Θ en Ω.

(1) Une nouvelle méthodologie pour transférer les fonctions de masse dans un autre cadre de discernement fondé sur les fonctions de croyance. Les fonctions de masse, également appelées Basic Belief Assignment (BBA), sont l'une des représentations fondamentales de l'information dans les fonctions de croyance. Dans la théorie des fonctions de croyance, les différentes informations sont modélisées par des fonctions de masse et sont ensuite combinées dans le même cadre de discernement. Comme les informations hétérogènes peuvent avoir des cadres de discernement différents, la combinaison est difficile. C'est pourquoi nous proposons une méthode de transformation qui permet de transférer différentes fonctions de masse dans le même cadre de discernement. Les principales propriétés de notre méthode que nous voulons mettre en avant sont les suivantes : (i) la transformation peut préserver à la fois l'incertitude et l'imprécision du cadre de discernement original et les transférer au

cadre cible, (ii) la transformation est fondée sur des mesures de similarité qui ont des choix multiples, (iii) la transformation peut également être étendue à une version d'apprentissage, en utilisant des échantillons d'entraînement pour apprendre les similarités.

La Figure R.3 montre le processus de transfert d'une BBA m_c^{Θ} dans le cadre de discernement Θ vers un autre cadre Ω en fonction des informations de similarité $Sim(T_i, O_j)$ entre deux ensembles d'éléments focaux quelconques $T_i \subseteq \Theta$, et $O_j \subseteq \Omega$. Une BBA m_c^{Θ} est d'abord décomposée en un ensemble de BBA simples (SBBA), *e.g.* $m_{c,T_1}^{\Theta}, ..., m_{c,T_n}^{\Theta}$, dont chacune n'a qu'un seul élément focal en plus de Ω. Pour chaque SBBA telle que m_{c,T_1}^{Θ}, il peut être transféré comme un groupe de SBBA sur Ω en utilisant des informations de ressemblance/similitude. Ce faisant, chaque élément focal sur Ω contient des informations de Θ, qui peuvent être combinées par la règle de Dempster. Pour obtenir une BBA normale, des SBBAs sur Ω peuvent également être combinées par la règle de Dempster. La transformation proposée est plus prudente car l'incertitude et l'imprécision sur Θ sont toutes deux préservées et transférées dans Ω.

Dans le domaine de la fusion de multiples méthodes supervisées et non supervisées, notre contribution est

(2) Un cadre de fusion probatoire pour les méthodes supervisées et non supervisées (EFSC). Sur la base de la méthode de transformation proposée, le cadre EFSC proposé peut efficacement combiner les informations issues du clustering avec la classification et ainsi diminuer la dépendance vis-à-vis des données brutes et des échantillons étiquetés. La méthode EFSC tente de traiter directement l'incertitude et l'imprécision au niveau de la sortie des résultats supervisés en combinant plusieurs résultats non supervisés. De cette façon, les informations les plus fiables provenant de différentes sources peuvent être extraitées et ensuite combinées pour améliorer la précision globale. La stratégie de fusion itérative de l'EFSC peut aider à tirer le plus d'informations possible du clustering des résultats et ainsi réduire la réponse sur les informations de données brutes. Ce cadre permet de tirer pleinement parti des informations au niveau de la sortie, de manière à réduire les influences causées par le nombre limité d'échantillons étiquetés.

Le déroulement général des opérations est détaillé sur la Figure R.4, où les différents classificateurs initiaux sont désignés par s_1, s_2,...,s_M. Pour le classificateur s_1, par exemple, nous avons choisi au hasard z différentes méthodes de clustering en C, désignées successivement par $c_{s_{11}}$, $c_{s_{12}}$,...$c_{s_{1z}}$, à combiner avec s_1 sur la base de la transformation proposée. Dans la ième étape de fusion, de nouvelles informations étiquetées $L_{s_{1i}}$ (c'est-à-dire de nouveaux résultats de classification), sont extraites pour être combinées avec le clustering dans l'itération suivante. Des informations fiables de chaque classification individuelle sont extraites à mesure que le processus itératif converge.

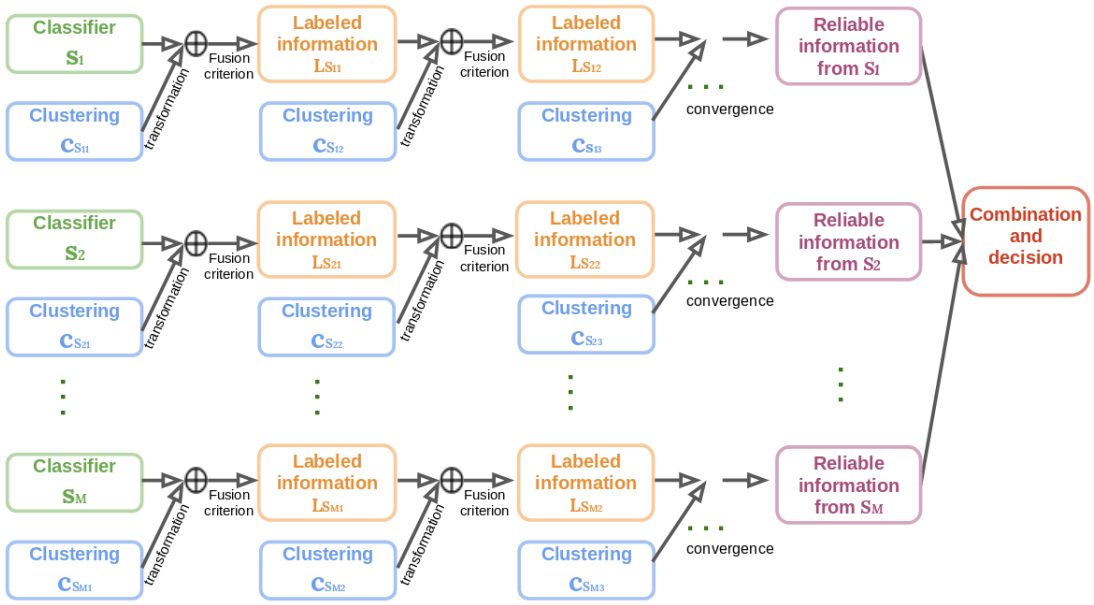


Figure R.4. Flux de travail du EFSC proposé.

La collecte d'un nombre suffisant d'échantillons étiquetés est un travail important mais difficile dans la classification de la couverture terrestre, en particulier pour les zones difficiles d'accès. Lorsque les données étiquetées ne sont pas disponibles, certaines connaissances spécialisées, telles que les indices spectraux, peuvent être utilisées pour fournir des informations sur la couverture des terres. Les différentes couvertures terrestres montrent leurs reflets spécifiques dans différentes bandes spectrales, qui peuvent être utilisées pour identifier les couvertures terrestres. Souvent, les indices/signatures spectraux sont facilement influencés par de nombreux facteurs, tels que les capteurs, les résolutions, les zones d'étude, *etc.*, que les seuils permettant d'identifier la couverture terrestre correspondante nécessitent une quantité de données étiquetées pour être déterminés. Nous étudions donc les seuils automatiques mais moins précis déterminés par les propriétés spectrales et parvenons à les combiner par des fonctions de croyance. Considérant la situation où aucune donnée étiquetée n'est disponible pour les zones difficiles d'accès, nous avons également proposé quelques méthodes de classification de la couverture terrestre qui permettent d'identifier automatiquement les couvertures terrestres sans apprentissage sur la base d'échantillons étiquetés. Nos contributions consistent en

(3) Une nouvelle méthode de détection de l'eau fondée sur les fonctions de croyance. Comme l'eau a la plus forte absorption dans la bande du proche infrarouge (NIR), sa réflexion NIR est capable de montrer une grande différence par rapport aux autres couvertures terrestres. En se fondant sur cette propriété, un seuil peut être détecté automatiquement pour identifier l'eau et les autres éléments et ce processus peut être modélisé par des fonctions de croyance. Les pixels identifiés comme étant de l'eau avec un degré de confiance élevé sont sélectionnés comme

échantillons d'entraînement pour former un classificateur supervisé dont les résultats sont également combinés avec les précédents générés par le seuil NIR. Cette méthode prend en compte non seulement les décisions sur les singletons, mais aussi sur l'ignorance totale. De cette façon, il est également possible de détecter les zones où il pourrait y avoir de l'eau.

(4) Une méthode de classification automatique de l'occupation des sols qui ne prend en compte que les indices spectraux afin de pouvoir obtenir directement des étiquettes sémantiques sans utiliser d'échantillons de formation. La méthode proposée utilise des fonctions de croyance pour combiner les quatre méthodes un-contre-reste, qui identifient séparément et indépendamment l'eau, le sol nu, la végétation et l'union de l'eau et des surfaces imperméables d'autres couvertures terrestres. La méthode proposée comporte deux étapes principales : la combinaison des quatre méthodes un-contre-reste pour identifier les couvertures terrestres de base ; le raffinement de la végétation et des surfaces imperméables en sous-catégories qui indiquent différents niveaux de leur luminosité en fonction de l'indice du complexe biologique (BCI). Cette méthode permet d'identifier différentes couvertures terrestres sans utiliser d'échantillons étiquetés, ce qui convient donc à la classification de la couverture terrestre dans les zones difficiles d'accès.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Arnaud MARTIN, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. He has taught me the methodology to carry out the research. I am extremely grateful for his patience during the discussion I had with him on research work and thesis preparation. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would also like to express my deep gratitude to my co-supervisor Dr. Rémi ESTIVAL of TOTAL, for providing invaluable guidance throughout this research. He gave me the opportunity to do an internship with him and to continue my Ph.D. research in TOTAL. His dynamism, vision, sincerity and motivation have deeply inspired me. He has taught me to present the research works as clearly as possible. It was a great privilege and honor to work and study under his guidance. I am extremely grateful for what he has offered me. I would also like to thank him for his friendship and great sense of humor.

Besides my supervisors, I would like to thank the rest of my thesis committee: Prof. Sylvie LE HEGARAT-MASCLE, Dr. Jean DEZERT, Prof. Cédric WEMMERT, Prof. Sébastien LEFEVRE, and Prof. Zhuga LIU. I would like to express my special thanks to Prof. Sylvie LE HEGARAT-MASCLE, and Dr. Jean DEZERT, the CSI members of my thesis, for their guidance and encouragement. I would like to thank Prof. Cédric WEMMERT, the reporter of my thesis, and Prof. Sébastien LEFEVRE, the president of my thesis, for their encouragement, for their insightful comments and their presence to my thesis defense during the COVID-19.

I would also like to say thanks to all my colleagues and my friends in team DRUID, IRISA and also in TOTAL. In particular, I am grateful to Dr. Salma BEN DHAOU, for her constant encouragement throughout this research work.

Last but not the least, I would like to thank my parents, my finacé Dr. Bing HAN, my best friend Ms. Lina CAO and my cat Mr. Linux, for supporting me spiritually throughout my life. My thanks also go to all the people who have supported me to complete the research work directly or indirectly.

Appendix A

METIS project

In this section, we briefly outline the industrial background of this dissertation. We provide an overview of the METIS (Multiphysics Exploration Technology Integrated System) project launched by TOTAL in the section A.1, then more details on the displacement model are presented in section A.2. In section A.3, we present the remote sensing data available in METIS.

A.1 General overview

Many hard-to-reach places still have abundant but untapped oil beneath the Earth's surface. Because of their incredibly complex surface and subsurface, (that are) composed of folds and thrusts, development and exploration are expensive. At the same time, it is also critical to ensure that their ecosystems are respected and protected, which makes exploration more challenging. For this reason, geophysical imaging of the subsurface in hard-to-reach topography is a major challenge for oil and gas exploration. Seismic techniques are widely used to detect rock structures to reveal the possible presence of oil and natural gas, as shown in Figure A.1. This involves applying a seismic energy source to a discrete location on the surface. The resulting energy is reflected from the interfaces where rock properties change. By recording this reflected energy on an array of geophones placed on the surface of the ground, the results can be processed to produce an image of underground geological structures, as shown in Figure A.2. Surveys can be carried out along lines to produce a vertical profile (2D survey) or over an area to generate a 3D underground volume.

Complex subsurface areas, such as mountainous areas, represent a major challenge: their seismic imagery is particularly difficult and the uncertainty resulting from subsoil mapping is great. Moreover, the often limited accessibility of these areas (the accessibility of these areas is usually limited, which) makes traditional acquisition techniques too costly and risky in terms of HSE (Health, Safety, and Environment). For instance, explorers have to hack out regularly spaced trails in forests to set out seismic sensors by hand. It's a slow-going, hazardous technique for the people involved, requiring huge resources such as helicopters and ultimately yielding poorer quality 2-D images.

To solve major problems related to geophysical and seismic imaging, TOTAL has embarked on an integrated geophysics and logistics project called METIS (Multiphysics

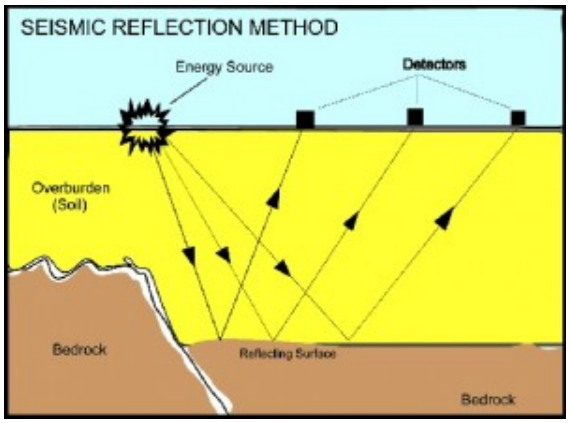


Figure A.1. The principle of seismic acquisition.

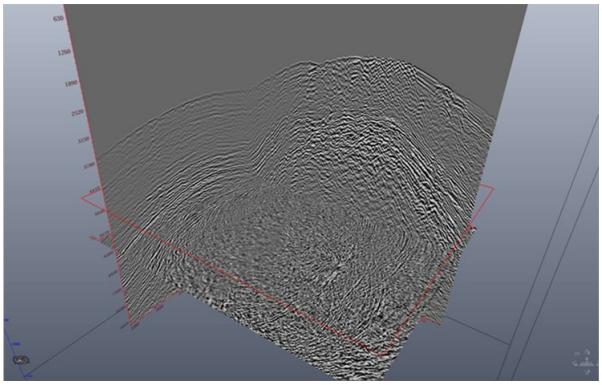


Figure A.2. Seismic imaging.

Exploration Technology Integrated System). METIS project aims at developing an integrated geophysical and logistic system within 10 years to acquire, process, and obtain quality structural and quantitative information in real-time. The ultimate purpose is to be able to complete the exploration and imaging phase virtually without setting foot on the ground.

METIS project involves various sensors dropped by drones, airship carrying and moving heavy loads. Besides this, it is of great importance for METIS to respond to the environmental requirements by reducing the impact of seismic acquisition campaigns in complex areas. Traditionally, openings for seismic acquisitions requires bridging that has a great impact on the natural environments but also expensive and risky, as shown in Figure A.3. Therefore, the estimation of the accessibility from remote sensing data will respond to the need for the project to reduce cost and risk through planning operation that takes better account of the terrain and to environmental imperatives leading to a reduction of bridging demands in the forest area.

Figure A.3. Example of bridging.

A.2 Displacement model

In complex natural areas, impolitic moving to distant locations is always hazardous and exhausting. The displacement model aims to propose the fastest, easiest, or safest way to displace to distant locations, by optimizing accessibility on the ground. The 'accessibility' refers to the ability of humans to reach a particular location in a landscape, varying with the built infrastructure (*e.g.* roads), natural features of the landscape (*e.g.*terrain steepness), barriers to travel (*e.g.* stream and river) and land cover conditions (*e.g.* vegetation density). Evaluating the accessibility is of great significance because it provides a way of reducing cost and increasing the efficiency of traveling in ecological observations. The modeling is challenging for the complicated natural situation on the one hand and lack sufficient information of ground on the other hand.

The accessibility refers mainly to the difficulty of physical displacement on the ground, which is measured based on the identification of land covers.Due to the ambiguity of describing accessibility directly, no universal measure exists in real-world applications. Commonly, accessibility is estimated in different criteria according to specific circumstances and different land covers. For example, for vegetation, its accessibility could be measured for different biological layers with a function of density and height. Vegetation with middle height like shrub or woody plants may impede people more than grass and forest. Thus, an interval of height with a certain density could be considered as a region impossible to reach for humans or vehicles. Besides the measure of accessibility based on densities of different vegetation layers, some articles measure accessibility for specific types of vegetation [39]. The slope is another worthy aspect to be considered. For instance, a slope in $[-60°, 60°]$ is reachable for human and $[-40°, 40°]$ for vehicle [49].

Accessibility could be also quantified by cost values, such as energy, time, money. The more costly it is, the less accessible can be considered. In recent years, several kinds of research to measure energy cost through metabolic rate using human's weight,

load carried, walking speed, slope, and terrain factor have been proposed [119]. Terrain factor, a function of the load carried and velocity reflects how different types of terrain surface affect movement, considered as an important factor when calculating energy cost. According to to [119], vegetation, except heavy vegetation, has the same terrain factor. It is also mentioned that the strength and friction of the surface like bare soil beneath the vegetation can be ignored in the terrain factor of vegetation.

Furthermore, other types of accessibility, such as economical accessibility, weather accessibility can also be taken into consideration together according to needs [8]. This paper focuses on the true accessibility of mountaineering and presents a conceptual framework for assessing it. For mountaineering, true accessibility consists of two factors: (1) destination accessibility (*e.g.* the transport system and in situation services) and (2) real access (*e.g.* social, economic, weather and psychophysical environments). Distance, slope, stream crossings, and vegetation density are incorporated into a least-cost model of energetic expenditure for human access to locations. Accessibility is quantified here by the least-cost model and corresponding methods to calculate energy cost for stream and slope are proposed.

A.3 Remote sensing data

Remote sensing data, including both airborne and spaceborne sensors, provide powerful tools for observing different environment and ground situation. Remote sensing sensors can be divided into two basic categories: passive sensors and active sensors. Passive sensor, including multi-spectral data, and hyperspectral data, refers to a microwave instrument which can receive signals from a natural source, such as a river, vegetation. Active sensors such as SAR and LiDAR, are a radar instrument used to send signals and receive the reflects. Variety in spatial, radiometric, spectral, and temporal resolution makes a certain type of remote sensing sensor cannot cater for all various requirements. The type of sensor, scale, and spatial resolution are supposed to decide based on users' needs.

The definitions of spatial resolution can be found in [103], which refers to the spatial coverage of one pixel and also be called Ground Sampling Distance (GSD). There are generally 4 classification levels: (i) low or coarse resolution refers to GSD of $30m$ or greater, (ii) medium resolution is GSD in the range of 2.0–30 m, (iii) high resolution corresponds to GSD with 0.5–2.0 m, and (iv) very high resolution is pixel sizes < 0.5 m GSD. Since different sensors have different spatial, temporal, spectral, and radiometric characteristics, the selection of appropriate sensors should be done according to the purposes of applications. In general, for continental or global scale classification, remote sensing data with coarse resolution, such as MODIS, SPOT, are usually employed. At a regional scale or large scale, medium resolution data including LandSat TM are preferable. When it comes to fine-scale classification in local areas, high or very high-resolution data are the most appropriate selection [102].

Therefore, it is of great significance to manipulate remote sensing data efficiently and appropriately, grasping specific characteristics of distinct remote sensing data and understanding their advantages and disadvantages.

In METIS, different sensors, such as multi-spectral satellites, SAR, LiDAR are avail-

able to capture information on the ground. Details about different sensors used in METIS are presented in Table A.1. Two satellite sensors Rapideye and WorldView-2 have been employed up till now. RapidEye contains 5 bands: Blue, Green, Red, Near Infrared Red, and Red Edge, with a resolution of 5 m. WorldView-2 belongs to VHR (Very-High-Resolution) sensors with the resolution of 5 cm and four spectral bands: Blue, Green, Red, Near Infrared Red are currently available.

Synthetic Aperture Radar (SAR) is an active sensor different from multi-spectral satellites. It emits electromagnetic radiation that can pass through clouds and atmosphere and then directly reach the ground. SAR data mainly contains information on surface roughness, shape, and soil moisture. SAR can generate cloud-free images and measure while it has a low resolution in METIS, which makes it difficult to map vegetation in species or community level. SAR also has limitations in mountainous terrain due to its low sensitivity to elevated objects. Besides, SAR backscatter is influenced by surface roughness and dielectric constant of the surface material.

LiDAR is a laser profiling and scanning system to measure distances. Similar to SAR, LiDAR emits the coherent light pulse, receives its reflection from targets, and measures the distance by 3D cloud points. LiDAR can create elevation surfaces that are in 3D, and its intensity values can help to identify forest canopy, branches, trunk, ground, or used in urban studies to determine building heights. Nevertheless, unlike SAR penetrating clouds, LiDAR can be restricted by thick clouds and precipitation because of the different wavelengths. Also, LiDAR is a rather costly method and due to its operating platform. Therefore, it is only available for parts of areas, which limits the utilization of LiDAR in METIS. To overcome the limitations of each sensor and to combine their advantages, the fusion of multi-sensors is highly important.

Table A.1 – Available data in METIS.

Sensors in METIS	Characteristic	Utilization	Limitation
Rapideye	*multi-spectral bands* -Blue, Green, Red, Red-Edge, NIR *Resolution*: 5m	Coarser land cover classification; Fine land cover classification; Vegetation mapping.	Difficult to detect ground condition under vegetation canopy; Difficult to remove clouds completely.
WorldView-2	*Mulispectral bands* - Blue, Green, Red, NIR *Resolution* : 0.5m	Coarser land cover classification; Fine land cover classification; Vegetation mapping.	Similar limitation as RapidEye; High resolution improves the classification quality.
SAR	L-band with dual-polarization *Resolution* : 10m	Cloud-free image; Vegetation detection;	cannot map vegetation in species or community level
LiDAR	3D point cloud	Vegetation density; Generation of DEM.	Not available for all study areas.

Contents

Bibliography

[1] *USGS earthexplorer*. https://earthexplorer.usgs.gov/.

[2] G. A. Fox, G. J. Sabbagh, S. W. Searcy, and C. Yang. An automated soil line identification routine for remotely sensed images. *Soil Science Society of America Journal*, 68(4):1326–1331, 2004.

[3] E. Abbasi, M. E. Shiri, and M. Ghatee. A regularized root–quartic mixture of experts for complex classification problems. *Knowledge-Based Systems*, 110:98 – 109, 2016.

[4] A. Acharya, E. R. Hruschka, J. Ghosh, and S. Acharyya. C3e: A framework for combining ensembles of classifiers and clusterers. In C. Sansone, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems*, pages 269–278, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[5] J. B. Adams, D. E. Sabol, V. Kapos, R. A. Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the brazilian amazon. *Remote Sensing of Environment*, 52(2):137 – 154, 1995.

[6] J. Aguirre-Gutiérrez, A. C. Seijmonsbergen, and J. F. Duivenvoorden. Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in mexico. *Applied Geography*, 34:29 – 37, 2012.

[7] X. Ao, P. Luo, X. Ma, F. Zhuang, Q. He, Z. Shi, and Z. Shen. Combining supervised and unsupervised models via unconstrained probabilistic embedding. *Information Sciences*, 257:101 – 114, 2014.

[8] M. Apollo. The true accessibility of mountaineering: The case of the high himalaya. *Journal of Outdoor Recreation and Tourism*, 17:29 – 43, 2017.

[9] A. Appriou. *Uncertainty Theories and Multisensor Data Fusion*. Wiley-IEEE Press, 1st edition, 2014.

[10] N. Audebert, B. Le Saux, and S. Lefevre. Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):159–173, 2019.

[11] C. Baral. Abductive reasoning through filtering. *Artificial Intelligence*, 120(1):1 – 28, 2000.

[12] F. Baret, S. Jacquemoud, and J. F. Hanocq. The soil line concept in remote sensing. *Remote Sensing Reviews*, 7(1):65–82, 1993.

[13] M. Bauer. Approximation algorithms and decision making in the dempster-shafer theory of evidence — an empirical study. *International Journal of Approximate Reasoning*, 17(2):217 – 237, 1997. Uncertainty in AI (UAI'96) Conference.

[14] P. Behnia. Comparison between four methods for data fusion of etm+ multispectral and pan images. *Geo-spatial Information Science*, 8(2):98–103, Jun 2005.

[15] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar. 3-d deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4420–4434, Aug 2018.

[16] J. R. Bertini-Junior and M. do Carmo Nicoletti. An iterative boosting-based ensemble for streaming data classification. *Information Fusion*, 45:66 – 78, 2019.

[17] Y. Bi, J. Guan, and D. Bell. The combination of multiple classifiers using an evidential reasoning approach. *Artificial Intelligence*, 172(15):1731 – 1751, 2008.

[18] S. Bickel and T. Scheffer. Multi-view clustering. pages 19– 26, 12 2004.

[19] E. Binaghi, P. A. Brivio, P. Ghezzi, and A. Rampini. A fuzzy set-based accuracy assessment of soft classification. *Pattern Recognition Letters*, 20(9):935 – 948, 1999.

[20] I. Bloch. Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17(8):905 – 919, 1996.

[21] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pages 92–100, New York, NY, USA, 1998. ACM.

[22] M. J. Canty. Boosting a fast neural network for supervised land cover classification. *Computers Geosciences*, 35(6):1280 – 1295, 2009.

[23] I. Castelli and E. Trentin. Combination of supervised and unsupervised learning for training the activation functions of neural networks. *Pattern Recognition Letters*, 37:178 – 191, 2014. Partially Supervised Learning for Pattern Recognition.

[24] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4):293–307, 2010.

[25] T. Chakraborty. Ec3: Combining clustering and classification for ensemble learning. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 781–786, Nov 2017.

[26] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[27] Y. Chen, P. Dou, and X. Yang. Improving land use/cover classification with a multiple classifier system using adaboost integration technique. *Remote Sensing*, 9(10), 2017.

[28] Y. Chen, X. Zhao, and X. Jia. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, June 2015.

[29] D. Chengbin and W. Changshan. Bci: A biophysical composition index for remote sensing of urban environments. *Remote Sensing of Environment*, 127:247 – 259, 2012.

[30] A. Cornuéjols, C. Wemmert, P. Gançarski, and Y. Bennani. Collaborative clustering: Why, when, what and how. *Information Fusion*, 39:81 – 95, 2018.

[31] P. Corsini, B. Lazzerini, and F. Marcelloni. Combining supervised and unsupervised learning for data clustering. *Neural Computing & Applications*, 15(3):289–297, Jun 2006.

[32] M. Dawelbait and F. Morari. Monitoring desertification in a savannah region in sudan using landsat images and spectral mixture analysis. *Journal of Arid Environments*, 80:45 – 55, 2012.

[33] C. De Stefano, A. Della Cioppa, and A. Marcelli. An adaptive weighted majority vote rule for combining multiple classifiers. In *Object recognition supported by user interaction for service robots*, volume 2, pages 192–195 vol.2, Aug 2002.

[34] A. P. Dempster. *A Generalization of Bayesian Inference*, pages 73–104. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[35] T. Denœux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(5):804–813, May 1995.

[36] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2):234 – 264, 2008.

[37] T. Denœux, N. E. Zoghby, V. Cherfaoui, and A. Jouglet. Optimal object association in the dempster-shafer framework. *IEEE Transactions on Cybernetics*, 44(12):2521–2531, Dec 2014.

[38] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[39] P. N. D.M. Theobald, J.B. Norman. Estimating visitor use of protected areas by modeling accessibility: A case study in rocky mountain national park, colorado. *Journal of Conservation Planning*, 6:1–20, 2010.

[40] H. T. X. Doan and G. M. Foody. Increasing soft classification accuracy through the use of an ensemble of classifiers. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05*, volume 1, pages 4 pp.–, July 2005.

[41] L. Drăguţ, D. Tiede, and S. R. Levick. Esp: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6):859–871, 2010.

[42] P. Du, S. Liu, J. Xia, and Y. Zhao. Information fusion techniques for change detection from multi-temporal remote sensing images. *Information Fusion*, 14(1):19 – 27, 2013.

[43] D. Dubois and H. Prade. A set-theoretic view of belief functions logical operations and approximations by fuzzy sets†. *International Journal of General Systems*, 12(3):193–226, 1986.

[44] D. Dubois and H. Prade. Possibility theory and data fusion in poorly informed environments. *Control Engineering Practice*, 2(5):811 – 823, 1994.

[45] D. Dubois and H. Prade. Possibility theory in information fusion. In *Proceedings of the Third International Conference on Information Fusion*, volume 1, pages PS6–P19 vol.1, July 2000.

[46] P. J. Escamilla-Ambrosio and N. Mort. Hybrid kalman filter-fuzzy logic adaptive multisensor data fusion architectures. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, volume 5, pages 5215–5220 Vol.5, Dec 2003.

[47] G. M. Espindola, G. Camara, I. A. Reis, L. S. Bins, and A. M. Monteiro. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, 27(14):3035–3040, 2006.

[48] A. Essaid, A. Martin, G. Smits, and B. Ben Yaghlane. A Distance-Based Decision in the Credal Level. In *International Conference on Artificial Intelligence and Symbolic Computation (AISC 2014)*, pages 147 – 156, Sevilla, Spain, Dec. 2014.

[49] C. Esteves. Human accessibility modelling applied to protected areas management. *Natureza & Conservacao.*, 9(2):232–239, 2011.

[50] F. Farem, M. Dhibi, and A. Martin. *Combination of Supervised and Unsupervised Classification Using the Theory of Belief Functions*, pages 85–92. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[51] P. Fisher and S. Pathirana. The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sensing of Environment*, 34(2):121 – 132, 1990.

[52] G. M. Foody, D. S. Boyd, and C. Sanchez-Hernandez. Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing*, 28(8):1733–1746, 2007.

[53] G. M. Foody and D. P. Cox. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *International Journal of Remote Sensing*, 15(3):619–631, 1994.

[54] G. M. Foody, M. Pal, D. Rocchini, C. X. Garzon-Lopez, and L. Bastin. The sensitivity of mapping methods to reference data quality: Training supervised image classifications with imperfect reference data. *ISPRS International Journal of Geo-Information*, 5(11), 2016.

[55] D. Frossyniotis, M. Pertselakis, and A. Stafylopatis. A multi-clustering fusion algorithm. In I. P. Vlahavas and C. D. Spyropoulos, editors, *Methods and Applications of Artificial Intelligence*, pages 225–236, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[56] P. Gamba, M. Aldrighi, and M. Stasolla. Robust extraction of urban area extents in hr and vhr sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1):27–34, March 2011.

[57] J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han. Graph-based consensus maximization among multiple supervised and unsupervised models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 585–593. Curran Associates, Inc., 2009.

[58] J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han. A graph-based consensus maximization approach for combining multiple supervised and unsupervised models. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):15–28, Jan 2013.

[59] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, and E. Wolff. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters*, 15(4):607–611, April 2018.

[60] R. Gharbia, A. E. Hassanien, A. H. El-Baz, M. Elhoseny, and M. Gunasekaran. Multi-spectral and panchromatic image fusion approach using stationary wavelet transform and swarm flower pollination optimization for remote sensing applications. *Future Generation Computer Systems*, 88:501 – 511, 2018.

[61] S. Godinho, N. Guiomar, and A. Gil. Using a stochastic gradient boosting algorithm to analyse the effectiveness of landsat 8 data for montado land cover mapping: Application in southern portugal. *International Journal of Applied Earth Observation and Geoinformation*, 49:151 – 162, 2016.

[62] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. Feature extraction: foundations and applications. January 2006.

[63] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, Nov 1973.

[64] S. P. Healey, W. B. Cohen, Z. Yang, C. K. Brewer, E. B. Brooks, N. Gorelick, A. J. Hernandez, C. Huang, M. J. Hughes, R. E. Kennedy, T. R. Loveland, G. G. Moisen, T. A. Schroeder, S. V. Stehman, J. E. Vogelmann, C. E. Woodcock, L. Yang, and Z. Zhu. Mapping forest change using stacked generalization: An ensemble approach. *Remote Sensing of Environment*, 204:717 – 728, 2018.

[65] C. Hett, P. Messerli, A. Heinimann, M. Epprecht, and K. Hurni. A texture-based land cover classification for the delineation of a shifting cultivation landscape in the lao pdr using landscape metrics, 2013.

[66] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown. Completion of the 2011 national land cover database for the conterminous united states - representing a decade of land cover change information. 81:346–354, 05 2015.

[67] C. Huang, L. S. Davis, and J. R. G. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, 2002.

[68] M. H. Ismail, H. Pakhriazad, and M. Shahrin. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. January 2009.

[69] P. Jaccard. The distribution of the flora in the alpine zone.1. *New Phytologist*, 11(2):37–50, 1912.

[70] A. K. Jain, N. K. Ratha, and S. Lakshmanan. Object detection using gabor filters. *Pattern Recognition*, 30(2):295 – 309, 1997.

[71] M. Ji and J. R. Jensen. Effectiveness of subpixel analysis in detecting and quantifying urban imperviousness from landsat thematic mapper imagery. *Geocarto International*, 14(4):33–41, 1999.

[72] A.-L. Jousselme, D. Grenier, and Éloi Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2(2):91 – 101, 2001.

[73] A. Kaltsounidis and I. Karali. Dempster-shafer theory: ηow constraint programming can help. In M.-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, and R. R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 354–367, Cham, 2020. Springer International Publishing.

[74] T. Kavzoglu and P. M. Mather. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23):4907–4938, 2003.

[75] S. Klemenjak, B. Waske, S. Valero, and J. Chanussot. Unsupervised river detection in rapideye data. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 6860–6863, July 2012.

[76] L. Kuncheva, C. Whitaker, C. Shipp, and R. Duin. Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis & Applications*, 6(1):22–31, Apr 2003.

[77] L. I. Kuncheva. Diversity in multiple classifier systems. *Information Fusion*, 6(1):3 – 4, 2005. Diversity in Multiple Classifier Systems.

[78] S. Le-Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar. Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):1018–1031, Jul 1997.

[79] S. Le-Hégarat-Mascle, A. Quesney, D. Vidal-Madjar, O. Taconet, M. Normand, and C. Loumagne. Land cover discrimination from multitemporal ers images and multispectral landsat images: A study case in an agricultural area in france. *International Journal of Remote Sensing*, 21(3):435–456, 2000.

[80] F. Li, S. Li, and T. Denœux. Combining clusterings in the belief function framework. *Array*, 6:100018, 2020.

[81] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu. A review of remote sensing image classification techniques: the role of spatio-contextual information. *European Journal of Remote Sensing*, 47(1):389–411, 2014.

[82] P. Li, P. Ren, X. Zhang, Q. Wang, X. Zhu, and L. Wang. Region-wise deep feature representation for remote sensing images. *Remote Sensing*, 10(6), 2018.

[83] Y. Li, J. Qu, W. Dong, and Y. Zheng. Hyperspectral pansharpening via improved pca approach and optimal weighted fusion strategy. *Neurocomputing*, 315:371 – 380, 2018.

[84] Lijun Dai and Chuang Liu. Multiple classifier combination for land cover classification of remote sensing image. In *The 2nd International Conference on Information Science and Engineering*, pages 3835–3839, 2010.

[85] Z. Liu, Q. Pan, J. Dezert, J. Han, and Y. He. Classifier fusion with contextual reliability evaluation. *IEEE Transactions on Cybernetics*, 48(5):1605–1618, May 2018.

[86] Z. Liu, Q. Pan, J. Dezert, and A. Martin. Combination of classifiers with optimal weight based on evidential reasoning. *IEEE Transactions on Fuzzy Systems*, 26(3):1217–1230, June 2018.

[87] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007.

[88] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277 – 293, 2017.

[89] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657, Feb 2017.

[90] C. D. Man, T. T. Nguyen, H. Q. Bui, K. Lasko, and T. N. T. Nguyen. Improvement of land-cover classification over frequently cloud-covered areas using landsat 8 time-series composites and an ensemble of supervised classifiers. *International Journal of Remote Sensing*, 39(4):1243–1255, 2018.

[91] A. Martin, G. Sevellec, and I. Leblond. Characteristics vs decision fusion for sea-bottom characterization. *Journee d'Acoustique Sous-Marine*, 01 2004.

[92] A. Mathur and G. M. Foody. Land cover classification by support vector machine: towards efficient training. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 742–744, Sep. 2004.

[93] A. E. Maxwell, T. A. Warner, and F. Fang. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.

[94] D. Menaka, L. Padma Suresh, and S. S. Premkumar. Wavelet transform-based land cover classification of satellite images. In L. P. Suresh, S. S. Dash, and B. K. Panigrahi, editors, *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, pages 845–854, New Delhi, 2015. Springer India.

[95] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges. *Information Fusion*, 46:102 – 113, 2019.

[96] D. Mercier, Éric Lefèvre, and D. Jolly. Object association with belief functions, an application with vehicles. *Information Sciences*, 181(24):5485 – 5500, 2011.

[97] M. Merzougui, M. Nasri, and B. Bouali. Isodata classification with parameters estimated by evolutionary approach. In *2013 8th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pages 1–7, 2013.

[98] E. Morvant, A. Habrard, and S. Ayache. Majority vote of diverse classifiers for late fusion. In P. Fränti, G. Brown, M. Loog, F. Escolano, and M. Pelillo, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 153–162, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[99] G. Moser, S. B. Serpico, and J. A. Benediktsson. Land-cover mapping by markov modeling of spatial–contextual information in very-high-resolution remote sensing images. *Proceedings of the IEEE*, 101(3):631–651, March 2013.

[100] S. W. Myint. Urban vegetation mapping using sub-pixel analysis and expert system rules: A critical approach. *International Journal of Remote Sensing*, 27(13):2645–2665, 2006.

[101] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5):1145 – 1161, 2011.

[102] S. S. Nath, G. Mishra, J. Kar, S. Chakraborty, and N. Dey. A survey of image classification methods and techniques. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 554–557, July 2014.

[103] K. Navulur. *Multispectral Image Analysis Using the Object-Oriented Paradigm*. CRC Press, Inc., Boca Raton, FL, USA, 2006.

[104] S. Neil, B. Timothy, and P. Christopher. Classifying the neotropical savannas of belize using remote sensing and ground survey. *Journal of Biogeography*, 33(3):476–490, 2006.

[105] M. M. Nielsen. Remote sensing for urban planning and management: The use of window-independent context segmentation to extract urban features in stockholm. *Computers, Environment and Urban Systems*, 52:1 – 9, 2015.

[106] K. Nigam and R. Ghani. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the Ninth International Conference on Information and Knowledge Management*, CIKM '00, pages 86–93, New York, NY, USA, 2000. ACM.

[107] P. Orponen. Dempster's rule of combination is p-complete. *Artificial Intelligence*, 44(1):245 – 253, 1990.

[108] N. Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[109] M. Oxenham. The effect of finite set representations on the evaluation of dempster's rule of combination. In *2008 11th International Conference on Information Fusion*, pages 1–8, 2008.

[110] M. Pal and P. M. Mather. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4):554 – 565, 2003.

[111] M. Pal and P. M. Mather. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5):1007–1011, 2005.

[112] Z. Pawlak. *Rough Sets: Theoretical Aspects Of Reasoning About Data*, volume 9. 01 1991.

[113] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187:156 – 168, 2016.

[114] R. L. Powell, D. A. Roberts, P. E. Dennison, and L. L. Hess. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, brazil. *Remote Sensing of Environment*, 106(2):253 – 267, 2007.

[115] Qi Chen and Peng Gong. Automatic variogram parameter extraction for textural classification of the panchromatic ikonos imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5):1106–1115, May 2004.

[116] F. Ramdani. Extraction of urban vegetation in highly dense urban environment with application to measure inhabitants' satisfaction of urban green space. In *Journal of Geographic Information System*, volume 5, pages 117–122, 2 2013.

[117] D. Reynolds. *Gaussian Mixture Models*, pages 659–663. Springer US, Boston, MA, 2009.

[118] A. J. Richardson and C. L. Wiegand. Distinguishing vegetation from soil background information. [by gray mapping of landsat mss data]. 1977.

[119] P. Richmond, A. Potter, and W. Santee. Terrain factors for predicting walking and load carriage energy costs: Review and refinement. *Journal of Sport and Human Performance*, 3(3), 2015.

[120] D. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green. Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65(3):267 – 279, 1998.

[121] D. Roberts, M. Smith, and J. Adams. Green vegetation, nonphotosynthetic vegetation, and soils in aviris data. *Remote Sensing of Environment*, 44(2):255 – 269, 1993. Airbone Imaging Spectrometry.

[122] A. Romero, C. Gatta, and G. Camps-Valls. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54:1349–1362, 2016.

[123] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987.

[124] M. Sabzevari, G. Martínez-Muñoz, and A. Suárez. Vote-boosting ensembles. *Pattern Recognition*, 83:119 – 133, 2018.

[125] M. Salah. A survey of modern classification techniques in remote sensing for improved image classification. 11:21, 04 2017.

[126] J. Sander and J. Beyerer. Bayesian fusion: Modeling and application. In *2013 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6, 2013.

[127] T. M. A. Santos, A. Mora, R. A. Ribeiro, and J. M. N. Silva. Fuzzy-fusion approach for land cover classification. In *2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES)*, pages 177–182, June 2016.

[128] F. Sebbak, F. Benhammadi, M. Mataoui, S. Bouznad, and Y. Amirat. An alternative combination rule for evidential reasoning. In *17th International Conference on Information Fusion (FUSION)*, pages 1–8, July 2014.

[129] M. O. Sghaier, M. Hadzagic, and J. Patera. Fusion of sar and multispectral satellite images using multiscale analysis and dempster-shafer theory for flood extent extraction. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8, 2019.

[130] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[131] H. Shen, Y. Lin, Q. Tian, K. Xu, and J. Jiao. A comparison of multiple classifier combinations using different voting-weights for remote sensing image classification. *International Journal of Remote Sensing*, 39(11):3705–3722, 2018.

[132] B. R. Shivakumar and S. V. Rajashekararadhya. Investigation on land cover mapping capability of maximum likelihood classifier: A case study on north canara, india. *Procedia Computer Science*, 143:579 – 586, 2018. 8th International Conference on Advances in Computing Communications (ICACC-2018).

[133] A. Sisay. Remote sensing based water surface extraction and change detection in the central rift valley region of ethiopia. *American Journal of Geographic Information System*, 5(2):33–39, 2016.

[134] P. Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1 – 35, 1993.

[135] P. Smets. *Probability, Possibility, Belief: Which and Where?*, pages 1–24. Springer Netherlands, Dordrecht, 1998.

[136] M. O. Smith, J. B. Adams, and D. E. Sabol. *Spectral Mixture Analysis - New Strategies for the Analysis of Multispectral Data*, pages 125–143. Springer Netherlands, Dordrecht, 1994.

[137] Y. Sohn and N. Rebello. Supervised and unsupervised spectral angle classifiers. *Photogramm Eng Remote Sens*, 68(1):28 – 44, 2013.

[138] B. Solaiman, L. E. Pierce, J. M. Kellndofer, M. C. Mouchot, and F. T. Ulaby. Multisensor fusion through fuzzy reasoning. application to land-cover classification using ers-1/jers-1 sar composites. In *Geoscience and Remote Sensing Symposium*

Proceedings, 1998. IGARSS '98. 1998 IEEE International, volume 3, pages 1320–1322 vol.3, Jul 1998.

[139] V. Solanky and S. K. Katiyar. Pixel-level image fusion techniques in remote sensing: a review. *Spatial Information Research*, 24(4):475–483, Aug 2016.

[140] B. Somers, G. P. Asner, L. Tits, and P. Coppin. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7):1603 – 1616, 2011.

[141] J. Sublime, B. Matei, G. Cabanes, N. Grozavu, Y. Bennani, and A. Cornuéjols. Entropy based probabilistic collaborative clustering. *Pattern Recognition*, 72:144 – 157, 2017.

[142] K. Tarrio, M. A. Friedl, C. E. Woodcock, P. Olofsson, K. Turlej, Z. Zhu, T. R. Loveland, E. Bullock, P. A. Arevalo, C. Holden, V. J. Pasquarella, and Y. Zhang. Global Land Cover mapping and Estimation (GLanCE): a multitemporal Landsat-based data record of 21st century global land cover, land use and land cover change. In *AGU Fall Meeting Abstracts*, volume 2019, pages GC21D–1317, Dec. 2019.

[143] G. Tetteh and M. Schönert. Automatic generation of water masks from rapideye images. *Journal of Geoscience and Environment Protection*, 3(5):17–23, 2015.

[144] B. Tso and R. C. Olsen. A contextual classification scheme based on mrf model with improved parameter estimation and multiscale fuzzy line process. *Remote Sensing of Environment*, 97(1):127 – 136, 2005.

[145] R. Ullah Khan. Deep learning features fusion with classical image features for image access. *International Journal of Advanced Computer Science and Applications*, 9, 01 2018.

[146] M. Usman, W. Wang, and A. Hadid. Feature fusion with deep supervision for remote-sensing image scene classification. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 249–253, Nov 2018.

[147] F. Voorbraak. A computationally efficient approximation of dempster-shafer theory. *International Journal of Man-Machine Studies*, 30(5):525 – 536, 1989.

[148] Q. Wei, N. Dobigeon, and J. Tourneret. Bayesian fusion of multi-band images. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1117–1127, 2015.

[149] W. Wei and J. Liang. Information fusion in rough set theory : An overview. *Information Fusion*, 48:107 – 118, 2019.

[150] C. Wemmert and P. Gancarski. A multi-view voting method to combine unsupervised classifications. pages 447–452, 09 2002.

[151] C. Witharana and D. L. Civco. Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure euclidean distance 2 (ed2). *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:108 – 121, 2014.

[152] D. H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241 – 259, 1992.

[153] Q. Xu, Y. Zhang, and B. Li. Recent advances in pansharpening and key problems in applications. *International Journal of Image and Data Fusion*, 5(3):175–195, 2014.

[154] R. R. Yager. The entailment principle for dempster—shafer granules. *International Journal of Intelligent Systems*, 1(4):247–262, 1986.

[155] X. Yang. Parameterizing support vector machines for land cover classification. 77:27–37, 01 2011.

[156] Y. Yang and S. Newsam. Comparing sift descriptors and gabor texture features for classification of remote sensed imagery. In *2008 15th IEEE International Conference on Image Processing*, pages 1852–1855, Oct 2008.

[157] L. D. Yarbrough, G. Easson, and J. S. Kuszmaul. Quickbird 2 tasseled cap transform coefficients: A comparison of derivation methods.

[158] L. Yong, X. Congfu, and P. Yunhe. A new approach for data fusion: implement rough set theory in dynamic objects distinguishing and tracing. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 4, pages 3318–3322 vol.4, Oct 2004.

[159] W. Yu, W. Zhou, Y. Qian, and J. Yan. A new approach for land cover classification and change analysis: Integrating backdating and an object-based method. *Remote Sensing of Environment*, 177:37 – 47, 2016.

[160] L. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

[161] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100:9 – 34, 1999.

[162] J. Zhang and G. M. Foody. A fuzzy classification of sub-urban land cover from remotely sensed imagery. *International Journal of Remote Sensing*, 19(14):2721–2738, 1998.

[163] J. Zhao, X. Xie, X. Xu, and S. Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43 – 54, 2017.

[164] W. Zhao and S. Du. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 113:155 – 165, 2016.

[165] H. Zhu and O. Basir. A novel fuzzy evidential reasoning paradigm for data fusion with applications in image processing. *Soft Computing*, 10(12):1169–1180, Oct 2006.

[166] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5046–5063, Sep. 2018.

UNIVERSITÉ DE RENNES 1

Titre : Combinaison de classificateurs supervisés et non supervisés fondés sur les fonctions de croyance..... ....

Mots clés : Fonctions de Croyance, Combinaison de Méthodes Supervisées et Non Supervisées, Classification de La Couverture des Terres.

Résumé : L'identification des types de couverture terrestre fournit des informations de base pour la production d'autres cartes thématiques et établit une base de références pour les activités de surveillance. Par conséquent, la classification de la couverture terrestre à l'aide de données satellitaires est l'une des applications les plus importantes de la télédétection. De nos jours, les technologies à capteurs multiples font l'objet d'une grande attention dans la classification de la couverture terrestre.

Dans notre recherche, nous nous concentrons sur la fusion d'informations hétérogènes provenant de différentes sources. Le système de combinaison vise à résoudre les problèmes causés par le nombre limité d'échantillons étiquetés et peut-être donc utilisé dans la classification de la couverture des terres pour les zones difficiles d'accès. Nous proposons un nouveau cadre de fusion sur des fonctions de croyance qui permettent de combiner des informations hétérogènes et gérer l'incertitude. Ce cadre de fusion comprend une méthode de transformation qui peut transférer des informations hétérogènes dans le même cadre et une stratégie de fusion itérative qui prend les informations les plus fiables de chaque résultat de classification et de regroupement de manière prudente.

Title : Combination of supervised and unsupervised classifiers based on belief functions.....

Keywords : Belief functions, Combination of Supervised and Unsupervised Methods, Land Cover Classification

Abstract : Land cover classification using satellite data is one of the most important applications of remote sensing. A great deal of ground information is usually required to generate high-quality land cover classification. However, in complex natural areas, collecting information on the ground can be time-consuming and extremely expensive. Nowadays, multiple sensor technologies have gained great attention in land cover classification. They bring different and complementary information-spectral characteristics that may help to overcome the limitations caused by inadequate ground information.

In our research, we focus on the fusion of heterogeneous information from different sources. The combination system aims to solve the problems caused by limited labeled samples and can thus be used in land cover classification for hard-to-access areas. We propose a new fusion framework based on belief functions that can combine heterogeneous information and manage uncertainty. This fusion framework includes a transformation method that can transfer heterogeneous information in the same framework and an iterative fusion strategy that takes the most reliable information from each classification result and to combine in a cautious way. We evaluate this fusion framework on real remote sensing data and it shows satisfying performance to improve the accuracy of land cover classification when training samples are limited or mislabeled.