Développement et automatisation de méthodes de classification à partir de séries temporelles d'images de télédétection - Application aux changements d'occupation des sols et à l'estimation du bilan carbone

Antoine Masse

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Traitement d'image appliqué à la télédétection

Présentée et soutenue par :
Antoine Masse

le : vendredi 11 octobre 2013

Titre :

Développement et automatisation de méthodes de classification à partir de séries temporelles d'images de télédétection - Application aux changements d'occupation des sols et à l'estimation du bilan carbone

---

Ecole doctorale :
Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E)

Unité de recherche :
CESBIO UMR5126

Directeur(s) de Thèse :

Danielle Ducrot et Philippe Marthon

Rapporteurs :
Henri Maitre, Professeur Emerite, Telecom ParisTech
Gregoire Mercier, Professeur, Telecom Bretagne
Emmanuel Trouvé, Professeur, Université de Savoie

Membre(s) du jury :

Yannick Deville, Professeur, IRAP Toulouse
Florence Tupin, Professeur, Telecom ParisTech

Résumé

La quantité de données de télédétection archivées est de plus en plus importante et grâce aux nouveaux et futurs satellites, ces données offriront une plus grande diversité de caractéristiques : spectrale, temporelle, résolution spatiale et superficie de l'emprise du satellite. Les données de télédétection sont aujourd'hui abondantes et facilement accessibles, ce qui ouvre de nouvelles problématiques quant à la manière d'utiliser ces données multidimensionnelles de manière optimale.

Cependant, il n'existe pas de méthode universelle qui maximise la performance des traitements pour tous les types de caractéristiques citées précédemment; chaque méthode ayant ses avantages et ses inconvénients.

Les travaux de cette thèse se sont articulés autour de deux grands axes que sont l'amélioration et l'automatisation de la classification d'images de télédétection, dans le but d'obtenir une carte d'occupation des sols la plus fiable possible. Lors d'un suivi de plusieurs années, ce processus permet d'obtenir une carte des changements d'occupation des sols sur de grandes zones et de longues périodes.

Au cours des différents chapitres du manuscrit, les méthodes seront testées sur un panel de données très variées de :

- capteurs : optique (Formosat-2, Spot 2/4/5, Landsat 5/7, Worldview-2, Pleiades) et radar (Radarsat, Terrasar-X).
- résolutions spatiales : de haute à très haute résolution (de 30 mètres à 0.5 mètre).
- répétitivités temporelles (jusqu'à 46 images par an).
- zones d'étude : agricoles (Toulouse, Marne), montagneuses (Pyrénées), arides (Maroc, Algérie).

Le manuscrit de thèse comprendra le développement et l'application des idées suivantes :

- un état de l'art sur les méthodes de classification appliquées aux données de télédétection.
- l'évaluation et l'extraction d'indice de performance permettant de juger automatiquement des différentes caractéristiques d'une classification supervisée
- la sélection automatique de données pour la classification supervisée
- la fusion automatique d'images issues de classifications supervisées afin de tirer avantage de la complémentarité des données multi-sources et multi-temporelles
- la classification automatique basée sur des séries temporelles et spectrales de référence, ce qui permettra la classification de larges zones sans référence spatiale

A l'exception des applications directes de chaque chapitre, deux applications majeures seront détaillées :

- l'obtention d'un bilan carbone à partir des rotations culturales obtenues sur plusieurs années
- la cartographie de la trame verte (espaces écologiques): étude de l'impact du choix du capteur sur la détection de ces éléments

Abstract

As acquisition technology progresses, remote sensing data contains an ever increasing amount of information. Future projects in remote sensing like Copernicus will give a high temporal repeatability of acquisitions and will cover large geographical areas. As part of the Copernicus project, Sentinel-2 (European project satellite from ESA) will carry an optical payload with visible, near infrared and shortwave infrared sensors comprising 13 spectral bands: 4 bands at 10 m, 6 bands at 20 m and 3 bands at 60 m spatial resolution (the last one is dedicated to atmospheric corrections and cloud screening), with a swath width of 290 km. Sentinel-2 combines a large swath, frequent revisit (5 days), and systematic acquisition of all land surfaces at high-spatial resolution and with a large number of spectral bands. All of these characteristics make a unique mission to serve Global Monitoring for Environment and Security.

The context of my research activities has involved the automation and improvement of classification processes for land use and land cover mapping in application with new satellite characteristics. This research has been focused on four main axes:

- Selection of the input data for the classification processes
- Improvement of classification systems with introduction of ancillary data
- Fusion of multi-sensors, multi-temporal and multi-spectral classification image results.
- Classification without ground truth data

These new methodologies have been validated on a wide range of images available:

- Various sensors (optical: Landsat 5/7, Worldview-2, Formosat-2, Spot 2/4/5, Pleiades; and radar: Radarsat, Terrasar-X)
- Various spatial resolutions (30 meters to 0.5 meters)
- Various time repeatability (up to 46 images per year)
- Various geographical areas (agricultural area in Toulouse, France, Pyrenean mountains and arid areas in Morocco and Algeria).

These methodologies are applicable to a wide range of thematic applications like Land Cover mapping, carbon flux estimation and greenbelt mapping.

Remerciements

Ce manuscrit est le fruit de 3 ans de travail au sein du Cesbio et de l'IRIT.

Je tiens tout d'abord à remercier les deux personnes sans qui tout cela n'aurait pas été possible, Danielle Ducrot et Philippe Marthon, mes deux tuteurs qui m'ont encadré et appris le métier extraordinaire d'enseignant-chercheur. Il y a également toutes les personnes de ces deux laboratoires qui m'ont apporté soutien et bonne humeur durant ces trois ans. Ce fut réellement une chance et un plaisir de travailler avec eux.

Je remercie également Henri Maitre, Grégoire Mercier et Emmanuel Trouvé qui ont accepté d'être mes rapporteurs, et cela, parfois même durant leurs vacances. Je remercie également Yannick Deville et Florence Tupin pour leur participation au jury de thèse.

Je n'entrerai pas dans le détail des noms de toutes les personnes ayant participé de près ou de loin à cette grande aventure, car qu'ils soient amis, familles ou collègues, je les remercie tous infiniment.

Introduction générale

1. La télédétection d'hier à demain

La télédétection est l'ensemble des techniques qui permettent d'obtenir de l'information sur la surface de la Terre sans contact direct avec celle-ci (y compris l'atmosphère et les océans) notamment par l'acquisition d'images. La télédétection englobe tous les processus qui consistent à capter et enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchi, à traiter et analyser l'information qu'il représente, pour ensuite mettre en application cette information.

La télédétection commença en 1856 lorsqu'un appareil photographique fut installé de façon fixe à bord d'un ballon. La photographie aérienne verticale fut ensuite largement utilisée pour la cartographie, notamment lors de la première guerre mondiale. Pendant l'entre-deux guerres, la photographie aérienne devient un outil opérationnel pour la cartographie, la recherche pétrolière et la surveillance de la végétation. La période de 1957 à 1972 marque les débuts de l'exploration de l'Espace et prépare l'avènement de la télédétection actuelle. Le lancement des premiers satellites, puis de vaisseaux spatiaux habités à bord desquels sont embarqués des caméras, révèle l'intérêt de la télédétection depuis l'espace. La première application opérationnelle de la télédétection spatiale apparaît dans les années 1960 avec les satellites météorologiques de la série ESSA.

Le lancement en 1972 du satellite Landsat-1, premier satellite de télédétection des ressources terrestres, ouvre l'époque de la télédétection moderne. Le développement constant des capteurs et des méthodes de traitement des données numériques ouvre de plus en plus le champ des applications de la télédétection et en fait un instrument indispensable de gestion de la planète et un outil économique. En 1986, le premier satellite de la famille Spot est lancé et ouvre l'aire de la diffusion et de la commercialisation des images de télédétection.

Dans le domaine du rayonnement visible et infrarouge, les capteurs à très haute résolution spectrale sont d'utilisation courante. Les satellites sont de plus en plus petits et permettent des applications ciblées. Les satellites à capteur radar, actif ou non, se généralise pour une utilisation civile (SMOS, Terrasar, Sentinel-1, etc.). La diffusion accélérée et l'augmentation de la puissance des ordinateurs contribuent à promouvoir de nouvelles méthodes d'utilisation des données toujours plus abondantes fournies par la télédétection spatiale.

Demain, la diversité des satellites d'observation va s'accroitre. L'apparition de constellation de satellites permettra l'acquisition régulière d'une même scène, à différentes résolutions et dans diverses longueurs d'ondes.

Ces futures constellations telle que Sentinel (Copernicus, ex-GMES, lancement prévu en 2014 par ESA) mettront gratuitement et rapidement leurs données à disposition après les dates d'observation. La démocratisation de la télédétection et de l'imagerie satellite entrainera la nécessité de mise à disposition d'outils intuitifs et automatique pour le grand public.

2. L'utilisation de données multidimensionnelles de télédétection pour la cartographie des sols : enjeux et problématiques

Aujourd'hui, la quantité d'images de télédétection est abondante, variée et accessible. Divers problèmes se posent pour l'utilisation de ces données, notamment pour la cartographie de l'occupation des terres. Parmi ces problèmes, nous citerons les suivants :

- Quelles sont les meilleures images à utiliser pour une application thématique ?
- Comment peut-on utiliser la diversité des informations pour améliorer les performances des systèmes de classification ?
- Quelles sont les limites intrinsèques d'une méthode automatique en fonction des données utilisées ? Peut-on se libérer de ces limites ? Si oui comment ?

Cette thèse se proposera de donner une réponse à ces questions en présentant une approche originale, permettant le traitement automatique et optimal des images, afin d'obtenir une carte d'occupation des sols.

Cependant, il n'existe pas de méthode universelle qui maximise la qualité des traitements pour tous les types de caractéristiques. Chaque méthode a ses avantages et ses inconvénients. Le but de cette thèse ne sera donc pas de mettre en avant une méthode plutôt qu'une autre, mais de montrer et de valider une méthodologie permettant d'utiliser la diversité des données afin de fournir une approche la plus robuste, performante et automatique.

3. L'approche proposée et l'organisation du manuscrit

Ce manuscrit est composé de sept chapitres. Le premier chapitre présentera un état de l'art des méthodes de classification d'images, applicables à la télédétection. L'accent sera mis sur les particularités de ces méthodes : supervisées ou non supervisées (avec ou sans vérité terrain), paramétriques ou non paramétriques, ponctuelles, objet ou globales.

Au chapitre 2, dans le but de caractériser les méthodes et les données utilisées pour la classification, j'introduirai de nouveaux indices de performance afin d'évaluer les résultats de classifications supervisées. Un point sera également fait sur l'interprétation des classifications non supervisées, et justifiera l'introduction au chapitre 5 d'une nouvelle méthode de classification sans vérité terrain de la zone à étudier.

Aux chapitres 3 et 4, j'introduirai une méthodologie alliant automatisation et amélioration des classifications, appelée Sélection-Classification-Fusion (figure 0.1).

La première partie de cette méthode présentée dans le chapitre 3 consistera à introduire un algorithme de sélection robuste et flexible, afin de sélectionner les meilleurs jeux de données d'entrée du processus de classification, en fonction de critères de recherche précis introduits au chapitre 2. Cette sélection pourra se faire d'un point de vue temporel, spectral mais également sur le choix du classifieur ou de toute autre caractéristique pouvant avoir une influence sur le processus de classification.

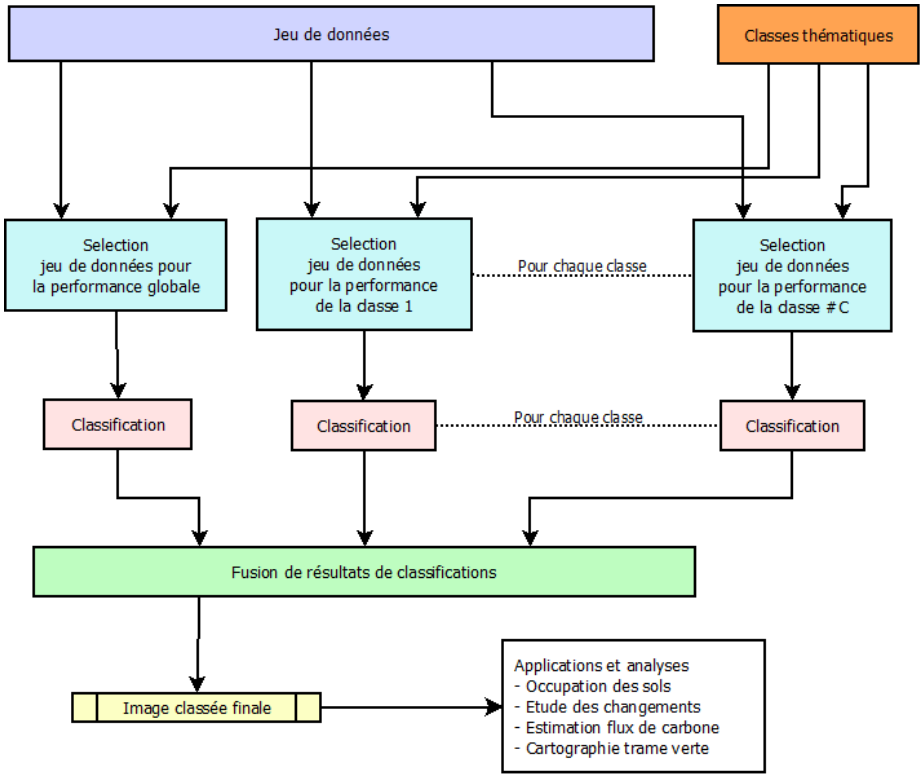


figure 0.1 : Illustration de l'organisation de la méthodologie présentée

La seconde partie de la méthode (chapitre 4) consistera à améliorer le résultat final par fusion de résultats de classification en utilisant la complémentarité de plusieurs résultats obtenus dans des conditions différentes (temporelles, sources, classifieur). Cette méthode a l'avantage de pouvoir fusionner automatiquement n'importe quel résultat de classification sans hypothèse particulière sur les données. Plusieurs applications de fusion de résultats de classification seront présentées (optique/radar, multi-classifieur,…).

Au chapitre 5, une méthode de classification guidée par une collection de données statistiques de valeurs radiométriques sera proposée. Cette méthode automatique permet d'obtenir une occupation du sol sans vérité terrain de la zone à étudier.

Les méthodes de ces 5 premiers chapitres seront validées sur de multiples jeux de données : optique (Formosat-2, Spot 2/4/5, Landsat 5/7, Worldview-2, Pleiades) et radar (Radarsat, ERS, Terrasar-X), résolutions spatiales variées de haute à très haute résolution (de 30 mètres à 0.5 mètre), répétitivités temporelles variées (jusqu'à 46 images par an) et sur des zones d'étude différentes (agricoles (Toulouse, Marne), montagneuses (Pyrénées), arides (Maroc, Algérie)).

Les résultats des méthodes présentées seront utilisés pour deux projets majeurs : la détection des changements et l'estimation des flux de carbone (chapitre 6) et la cartographie des trames vertes et bleues (chapitre 7). La détection des changements et l'estimation des flux de carbone font partie du programme Copernicus de l'ESA et nécessite la création d'une carte de rotation des occupations du sol. Les outils développés ont également servi à la cartographie des trames vertes (engagement national pour l'environnement, loi Grenelle 2 de l'environnement). Cette application fait notamment partie du projet Copernicus mais à une échelle européenne.

Chapitre I. La classification

Résumé : En télédétection, le problème de la classification multidimensionnelle est complexe et dépend de nombreux paramètres. Il n'existe pas une mais plusieurs façons de classer les pixels d'une image et chacune est à utiliser dans des contextes différents, d'où la nécessité pour la compréhension des chapitres suivants de la présentation détaillée de ces méthodes de classification qui peuvent être supervisées ou non supervisées (avec ou sans vérité terrain), paramétriques ou non paramétriques, ponctuelles, objet ou globales. Ces méthodes sont ensuite appliquées et comparées sur un exemple concret d'images de télédétection.

Mots clés : classification multidimensionnelle, surface de décision, ICM, données exogènes

« Si les nuances infinies du langage ne s'accommodent point des classifications rigides qu'on veut faire, tant pis pour les classifications. La science doit s'accommoder à la nature. La nature ne peut s'accommoder à la science. »

Ferdinand Brunot, 1922

Sommaire du chapitre I

1. Introduction

1.1. Définitions et existence

« La classification proprement dite est une opération de l'esprit qui, pour la commodité des recherches ou de la nomenclature, pour le secours de la mémoire, pour les besoins de l'enseignement, ou dans tout autre but relatif à l'homme, groupe artificiellement des objets auxquels il trouve quelques caractères communs, et donne au groupe artificiel ainsi formé une étiquette ou un nom générique. » Telle est la définition de Cournot (Cournot, 1851) concernant la classification d'un point de vue général.

Il y a deux points de vue philosophiques qui permettent de séparer la classification en deux catégories (Russel & Whitehead, 1925) :

- La démarche platonicienne qui suppose la préexistence des catégories à observer sur l'observateur, ici l'humain. L'humain ne fait donc que découvrir, plus ou moins imparfaitement, le monde qui l'entoure. Cette manière de voir le monde est aussi appelée « réalisme » et n'est plus majoritaire depuis le Moyen-âge.
- Le nominalisme qui suppose que l'humain donne une sémantique de la scène en fonction de son observation. Cette observation conduit donc au classement. L'humain fait ainsi parti du problème de classification et le processus de classification sera corrélé au facteur humain.

Il y a également deux points de vue mathématiques :

- La classification supervisée qui utilise la connaissance de la zone à étudier (vérités terrain) pour décider des catégories à attribuer.
- La classification non supervisée qui ne présuppose aucune connaissance de la zone à étudier, donc sans vérité terrain.

La classification non supervisée découle naturellement du réalisme et la classification supervisée du nominalisme.

Dans le cadre de la classification supervisée, la significativité et la représentativité des échantillons sont alors primordiales, le processus de classification ne fournissant qu'une fonction discriminante initialisée à partir de cette référence. Certaines méthodes se révéleront ainsi plus robustes que d'autres aux erreurs contenues dans les échantillons de références, la fiabilité des vérités terrain n'étant pas absolue.

Dans le cadre de la classification non supervisée, un compromis sur le degré de représentation de la scène doit être pris en compte. Ce compromis se fait dans le choix du nombre de classes et de la quantité de pixels à explorer pour initialiser le processus. Mis à part cette étape de découverte, le principe des méthodes de classification non supervisée reste identique à celui des méthodes de classification supervisée.

Commençons par définir quelques termes de classification qui nous serviront dans la suite du manuscrit.

Définition : Nous appellerons *classifieur* toute fonction ou ensemble de règles de décision permettant de transformer un ensemble de données appartenant à $\mathbb{R}^n$ en un ensemble de classes discrètes appartenant à $\mathbb{N}$.

Nous ne définirons pas le terme *classe* au sens mathématique du terme mais de la façon suivante :

Définition : Une *classe* est définie par un ensemble d'éléments la composant dont les caractéristiques ont été jugées similaires (humainement ou mathématiquement). Les éléments composant une classe peuvent ainsi être les pixels d'une image ou des ensembles de pixels alors appelés *objets*.

1.2. La classification appliquée à la télédétection pour la cartographie de l'occupation des sols

En télédétection, la classification consiste à effectuer la correspondance entre les éléments d'une scène de l'image matérialisés par des pixels, représentés par des valeurs radiométriques multidimensionnelles, et un ensemble de classes thématiques connues *a priori* ou non. Cette correspondance est réalisée par des fonctions discriminantes sous forme de règles de décision, qui sont définies et apprises à partir des informations *a priori*. En résumé, la classification vise à attribuer aux pixels des étiquettes dont l'origine est thématique (Ducrot, 2005). Les jeux d'images de télédétection peuvent être composés d'images acquises à plusieurs dates (dimension temporelle) et suivant différentes longueurs d'ondes (dimension spectrale).

1.3. Les catégories de méthodes de classification

Dans le paragraphe 1.1, nous avons abordé une première partition concernant le type de classification : supervisée et non supervisée.

Un deuxième partitionnement concerne la modélisation de la fonction discriminante. Si cette fonction suit une distribution statistique connue, on parle de méthode de classification paramétrique. Cette association offre la possibilité d'affecter à chaque site (ou pixel) une probabilité d'appartenance à une classe donnée. Tout classifieur probabiliste est donc paramétrique. Dans le cas contraire, on parle de méthode de classification non paramétrique, car aucune distribution n'est utilisée pour modéliser les données et seule la relation entre individus est considérée. Cette catégorie comprend notamment les méthodes basées sur le seuillage, la minimisation de distance et les méthodes à noyau.

Un troisième partitionnement concerne le mode opératoire de la méthode de classification. Il existe trois modes opératoires : ponctuel, objet et global. La méthode ponctuelle suppose l'indépendance des pixels de l'image et les traite indépendamment les uns des autres. La méthode objet suppose l'indépendance des objets dans l'image mais la dépendance des éléments d'un même objet. La méthode globale suppose la dépendance de tout élément de l'image avec l'ensemble de ses voisins. Comme nous le verrons plus tard, cette dépendance se limite à un nombre restreint de voisins lorsque la méthode est sous hypothèses markoviennes.

notation	définition
IM	un ensemble d'images
IM_{dk}	l'image de la date d au canal k
s	un site (ou pixel)
s_{ijdk}	le pixel s de coordonnées (i, j, d, k)
x_s	la valeur observée au pixel s
IM_s	un site s de l'image
D	un ensemble de dates
d	une date d'acquisition de l'image
K	un ensemble de canaux spectraux
k	un canal spectral d'acquisition de l'image
zone	un ensemble de pixels à étudier
C	un ensemble de classes (ou étiquettes)
c_s	la classe attribuée au pixel s
V	un système de voisinage
V_s	un point s du voisinage V
U	une fonction énergie
μ_c	le vecteur des moyennes de la classe c
σ_c	le vecteur des écarts types de la classe c
Σ_c	la matrice des covariances de la classe c
S	un ensemble de sites
M	l'ensemble des matrices de confusions
m_{c_i,c_j}	la confusion de la matrice m pour les points classés en tant que c_j mais appartenant à la classe c_j dans l'image de validation
CLA	l'ensemble des images des résultats de classification
#E	le cardinal ou nombre d'élément d'un ensemble fini E

tableau I.1 : Termes et définitions

2. Les méthodes de classification multidimensionnelles applicables à la télédétection

Dans la suite de ce chapitre, je présenterai les différentes méthodes de classification appartenant aux trois partitions détaillées précédemment. Chaque méthode sera accompagnée d'illustrations de son fonctionnement.

Commençons par introduire les notations qui seront utilisées au cours des différents chapitres et regroupées au tableau I.1. Dans le cadre de la classification d'images de télédétection multidimensionnelles, nous noterons **IM** l'ensemble des images disponibles à classer et $\mathbf{s_{ijdk}}$ les éléments ou pixels appartenant à **IM** avec (**i**, **j**, **d**, **k**) les coordonnées du point **s** dans l'univers **IM** (**i** pour la ligne, **j** pour la colonne, **d** pour la date et **k** pour la bande spectrale ; voir la figure I.1 pour une illustration de cette représentation en quatre dimensions). La valeur observée au pixel **s** est notée $\mathbf{x_s}$.

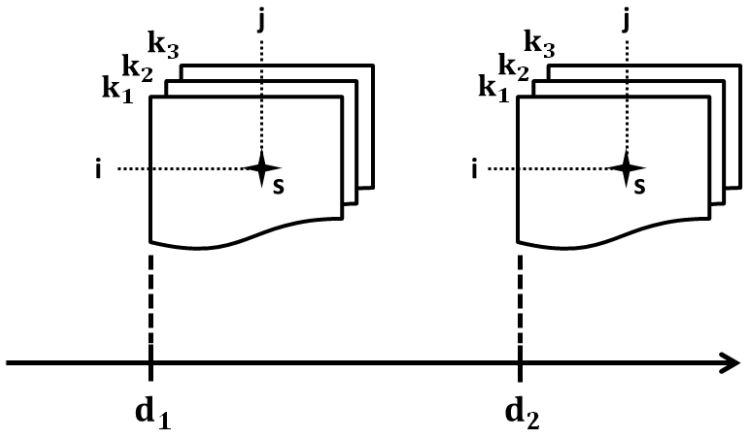


figure I.1 : Position du pixel s suivant les 4 dimensions utilisées en télédétection: i pour la ligne, j pour la colonne, d pour la date et k pour la bande spectrale)

Chaque élément thématique de référence est appelé classe et noté **c**. Pour chaque classe c, on peut calculer la moyenne, l'écart-type et la covariance des pixels pour chaque date **d** et bande spectrale **k** à partir des données de références (ou apprentissage). Ces statistiques sont notées respectivement $\boldsymbol{\mu_c}$, $\boldsymbol{\sigma_c}$ et $\boldsymbol{\Sigma_c}$.

Dans un but d'illustration des méthodes de classification présentées dans ce chapitre, nous prendrons un exemple volontairement simpliste d'une image RVB (Rouge-Vert-Bleu). Seules les bandes spectrales rouge et verte seront utilisées pour la classification, ce qui facilitera la visualisation des nuages de points et des surfaces de décisions (figure I.2). Des échantillons de 3 classes ont été acquis pour l'apprentissage et la vérification (figure I.3). L'image représente une fleur rouge disposant d'un pistil jaune avec un arrière-plan vert. Le nuage de point et les références choisies sont présentés en figure I.4. Cet exemple permettra de visualiser et de comprendre les mécanismes des différentes méthodes de classifications.

2.1. Les méthodes de classification ponctuelle

figure I.2 : Image utilisée pour l'illustration des méthodes de classification. Seules les bandes Rouge et Verte seront utilisées pour la classification

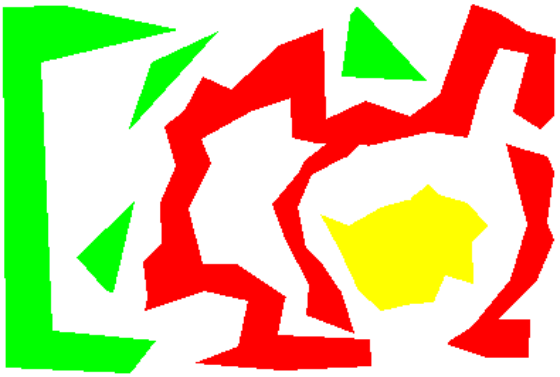


figure I.3 : Echantillons d'apprentissage pour les classifications supervisées : en Rouge les pétales de la fleur, en Jaune le pistil et en Vert la végétation en arrière-plan

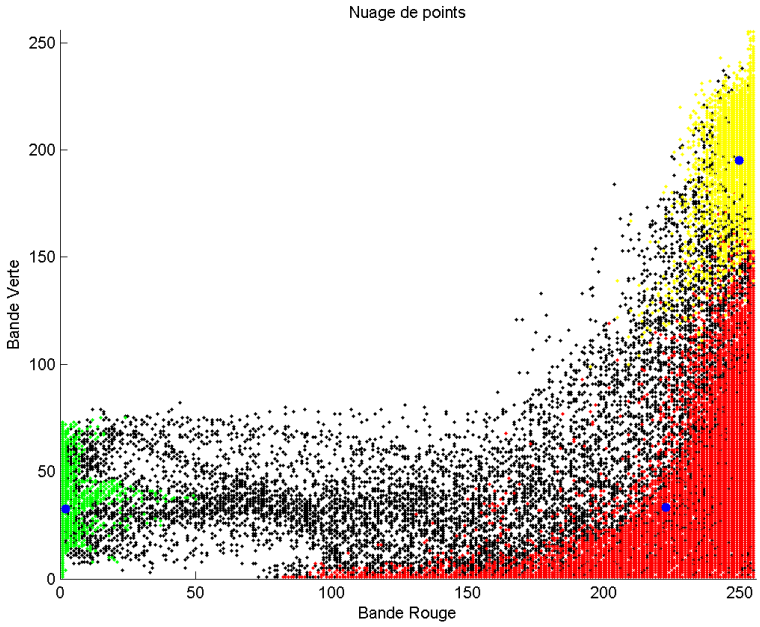


figure I.4 : Nuage de points associé à la figure I.2 et aux échantillons de la figure I.3. Le centre des classes est indiqué par un point bleu

figure I.5 : Surface de décision de la méthode de classification par seuillage : parallélépipède. La superposition du rectangle rouge et du jaune rend la décision impossible.

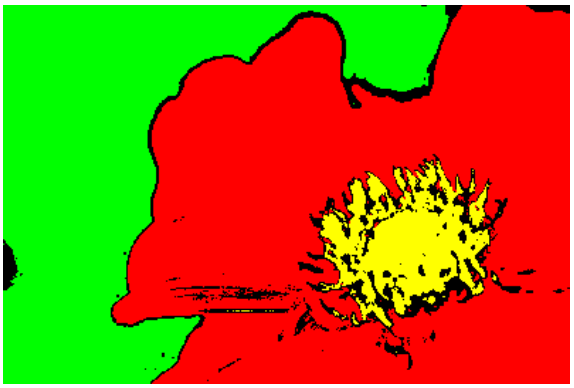


figure I.6 : Image résultant de la classification de la figure I.2 par la méthode du parallélépipède (α=3)

2.1.1. La méthode de classification par seuillage

La première façon de classer un individu dans une partition est issue de la logique booléenne. Cette méthode assez simple consiste à vérifier la présence ou non de l'individu dans une partition afin de décider de son appartenance. Cette méthode est non paramétrique.

En pratique, on choisira de classer un individu en fonction d'un seuil définissant les différentes partitions C de l'univers IM. La méthode la plus utilisée pour ce seuillage s'appelle la méthode de classification du « parallélépipède ».

Un ensemble de paramètres de régularisation noté **α** permet de régler l'étendue (l'aire) du parallélépipède. La règle de décision permettant de déterminer la classe optimale c_s^{opt} pour chaque pixel s de l'image est la suivante :

$$c_s^{opt} = c \text{ si et seulement si } x_s \in \left[\mu_{c,d,k} - \alpha_{d,k} * \sigma_{c,d,k}\ , \mu_{c,d,k} + \alpha_{d,k} * \sigma_{c,d,k}\right], \forall d \in D, k \in K$$

Cette règle de décision forme un parallélépipède multidimensionnel de centre $\mu_{c,d,k}$ et de largeur $2 * \alpha_{d,k} * \sigma_{c,d,k}$ (figure I.5, page précedente). Le principal inconvénient de cette méthode est son incapacité à décider lors de la superposition de plusieurs surfaces de décision (exemple de la figure I.6 avec la superposition de la surface rouge et jaune). De ce fait, cette méthode génère en général beaucoup de pixels non-classés c'est-à-dire dont la décision n'a pu être prise.

2.1.2. Les méthodes de classification par minimisation de distance

Une seconde façon de classer un individu consiste à caractériser les interactions et relations de l'individu avec les partitions de manière « réelle », et non plus booléenne. En choisissant un certain type de relation ou d'affinité, il sera facile de quantifier la proximité d'un élément à une partition.

Une des manières de caractériser cette interaction est d'utiliser la notion de distance qui permet de mesurer précisément la longueur entre deux éléments (et par extension entre un élément et une partition ou encore entre deux partitions).

Commençons par définir la notion de distance entre deux points.

Définition : Dans un espace vectoriel normé $(E, \|.\|)$, on définit une *distance* **d** à partir de la norme $\|.\|$ entre deux points x et y tel que $\forall\ (x,y) \in E \times E : d(x,y) = \|y - x\|, d :\ E \times E \rightarrow \mathbb{R}^+$ et qui vérifie les propriétés suivantes :

i. $\forall\ x,y \in E :\ d(x,y) = d(y,x)$ (propriété de symétrie)
ii. $\forall\ x,y \in E :\ d(x,y) = 0 \Leftrightarrow x = y$ (propriété de séparation)
iii. $\forall\ x,y,z \in E :\ d(x,z) < d(x,y) + d(y,z)$ (propriété de l'inégalité triangulaire)

L'espace **E** ainsi muni d'une distance **d** est appelé *espace métrique*.

En particulier dans $\mathbb{R}^n$, la distance entre deux points s'écrit généralement à l'aide de la formulation de Minkowski, que l'on appelle *p-distance* (équation I.1).

$$\begin{aligned} \forall\ (x,y) \in \mathbb{R}^n \times \mathbb{R}^n, d_p(x,y) &= \|x - y\|_p \\ &= \sqrt[p]{\sum_{i=1..n} |x_i - y_i|^p} \end{aligned} \qquad \text{(I.1)}$$

Les trois formulations dérivées de la p-distance les plus utilisées sont :

– la 1-distance ou distance de Manhattan :

$$\forall\,(x,y)\in\mathbb{R}^n\times\mathbb{R}^n, d_1(x,y)=\sum_{i=1}^{n}|x_i-y_i| \tag{I.2}$$

– la 2-distance ou distance euclidienne :

$$\forall\,(x,y)\in\mathbb{R}^n\times\mathbb{R}^n, d_2(x,y)=\sqrt{\sum_{i=1}^{n}|x_i-y_i|^2} \tag{I.3}$$

– la ∞-distance ou distance de Tchebychev :

$$\forall\,(x,y)\in\mathbb{R}^n\times\mathbb{R}^n, d_\infty(x,y)=\sup_{1\le i\le n}|x_i-y_i| \tag{I.4}$$

Rappelons pour information qu'il y a équivalence des normes et donc des distances dans $\mathbb{R}^n$.

On définit ensuite les distances d'ordre 2 en ajoutant des statistiques d'ordre 2 (écart-type et covariance) dont la plus utilisée pour mesurer la distance entre un point et un ensemble est la distance de Mahalanobis définie par l'équation I.5:

$$\forall\, x\in\mathbb{R}^n, c\in C,\qquad d_{maha}(x,c)=\sqrt{(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)} \tag{I.5}$$

Afin de comparer deux ensembles, il convient d'utiliser la distance de Bhattacharyya (Bhattacharyya, 1943) ou la divergence de Kullbach-Leibler (respectivement les équations I.6 et I.7). Nous présentons une version améliorée de la divergence de Kullbach-Leibler qui ajoute la propriété de symétrie à la divergence originale de (Kullback & Leibler, 1951). Néanmoins, nous préfèrerons la distance de Bhattacharyya pour la comparaison de deux ensembles du fait du manque de la propriété de l'inégalité triangulaire pour la divergence de Kullbach-Leibler.

$$\begin{aligned}\forall\,(c_1,c_2)\in C\times C, d_{bhatta}(c_1,c_2) &= \frac{1}{8}\left(\mu_{c_1}-\mu_{c_2}\right)^T\left(\frac{\Sigma_{c_1}+\Sigma_{c_2}}{2}\right)^{-1}\left(\mu_{c_1}-\mu_{c_2}\right)\\ &\quad+\frac{1}{2}\ln\left(\frac{\left|\frac{\Sigma_{c_1}+\Sigma_{c_2}}{2}\right|}{\sqrt{|\Sigma_{c_1}|}\sqrt{|\Sigma_{c_2}|}}\right)\end{aligned} \tag{I.6}$$

$$\begin{aligned}\forall\,(c_1,c_2)\in C\times C,\qquad div_{kl,sym}(c_1,c_2) &= \frac{1}{2}\left(tr\left(\Sigma_{c_1}\Sigma_{c_2}^{-1}\right)+tr\left(\Sigma_{c_2}\Sigma_{c_1}^{-1}\right)\right.\\ &\quad\left.+\left(\mu_{c_1}-\mu_{c_2}\right)^T\left(\Sigma_{c_1}^{-1}+\Sigma_{c_2}^{-1}\right)\left(\mu_{c_1}-\mu_{c_2}\right)\right)\end{aligned} \tag{I.7}$$

Les distances d'ordre supérieur ou égal à 3 ne sont pas utilisées du fait de leur complexité de calcul, notamment pour l'inversion de matrices de dimension 3 et plus.

Introduisons maintenant ces mesures entre les pixels à classer et les pixels de référence aux algorithmes de classification.

L'algorithme de classification par minimisation de distance (ou encore K-plus-proches-voisins) introduit par Cover (Cover & Hart, 1967) est très utilisé en télédétection (Lee, Weger, Sengupta, & Welch, 1990). Cet algorithme attribue au pixel s la classe la plus proche c_s^{opt} (équation I.8).

$$c_s^{opt} = \arg\min_{c_s \in C} \; d(x_s, c_s) \tag{I.8}$$

Avec $d(x_s, c_s)$ une distance entre le pixel s et la classe c_s.

La classe c_s peut alors être restreinte à sa moyenne μ_c pour les distances d'ordre 1 ou à sa moyenne μ_c et sa matrice de covariance Σ_c pour les distances d'ordre 2.

Voyons au travers de l'exemple introduit en début de chapitre l'illustration des différences entre les distances introduites précédemment.

Dans le cas des 3 distances d'ordre 1, les frontières de décision de l'algorithme de classification sont définies par des polygones de Voronoï (Voronoï, 1908). Les surfaces de décision sont illustrées pour chacune des distances présentées précédemment : la distance d-1 (Manhattan, figure I.7), la distance d-2 (euclidienne, figure I.9) et la distance d-∞ (Tchebychev, figure I.11). Les classifications associées sont présentées aux figures I.8, I.10 et I.12.

Pour les distances d'ordre 2, la surface de décision d'un classifieur utilisant la distance de Mahalanobis forme des hyperboloïdes (figure I.13) dont le centre est la moyenne de la classe et les rayons définis par les valeurs propres de la matrice de covariance (explication en figure I.15). Le résultat de la classification par minimisation de distance de Mahalanobis est présenté en figure I.14. L'apport de cette nouvelle distance est visible, les confusions jaune/rouge et vert/rouge sont très nettement diminuées.

La minimisation de distance conduit à des décisions incorrectes lorsque les dispersions des classes sont très différentes entre elles, *id est* des déterminants de matrice de covariance différents et donc des volumes d'hyperboloïdes différents. La figure I.16 illustre ce phénomène. Bien que la distance euclidienne de **s** à c_1 soit plus petite que celle de **s** à c_2, le point **s** est plus proche du nuage de point de c_2, d'où l'intérêt de la distance de Mahalanobis.

Rien ne justifie alors l'utilisation de la même métrique pour les différentes classes (Saporta, 2006), d'où la nécessité de normaliser les mesures et donc d'introduire la notion de probabilité.

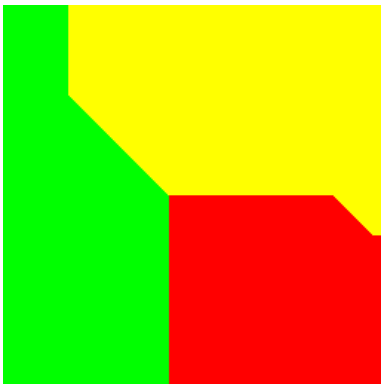


figure I.7 : Surface de décisions de la méthode par minimisation de distance de Manhattan

figure I.8 : Image résultant de la classification de la figure I.2 par minimisation de distance de Manhattan

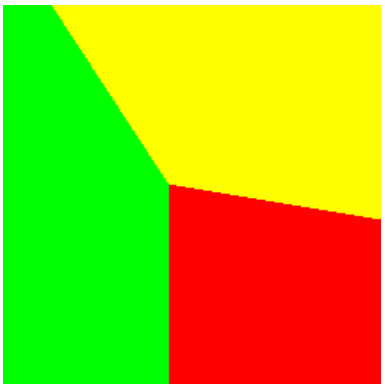


figure I.9 : Surface de décision de la méthode par minimisation de distance euclidienne

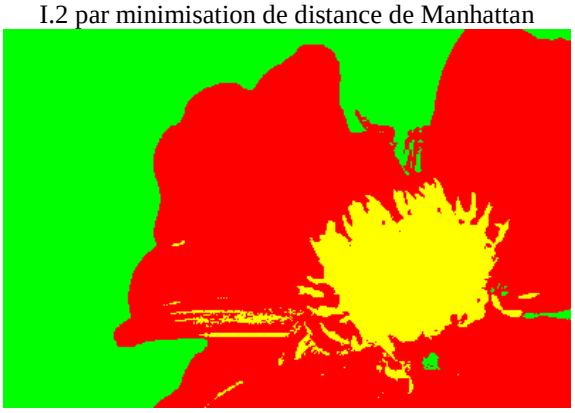


figure I.10 : Image résultant de la classification de la figure I.2 par minimisation de distance euclidienne

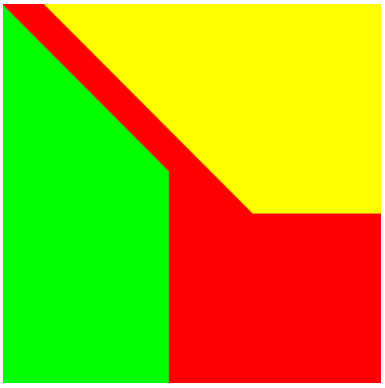


figure I.11 : Surface de décision de la méthode de par minimisation de distance de Tchebychev

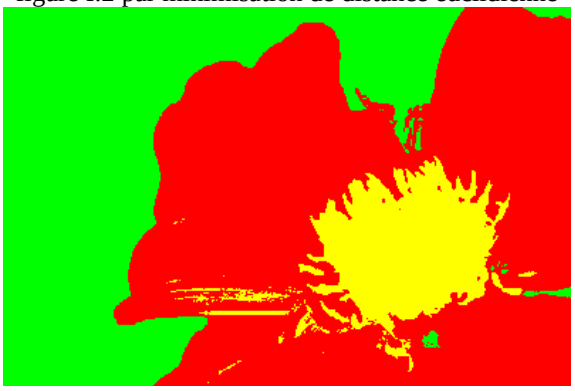


figure I.12 : Image résultant de la classification de la figure I.2 par minimisation de distance de Tchebychev

figure I.13 : Surface de décision de la méthode par minimisation de distance de Mahalanobis

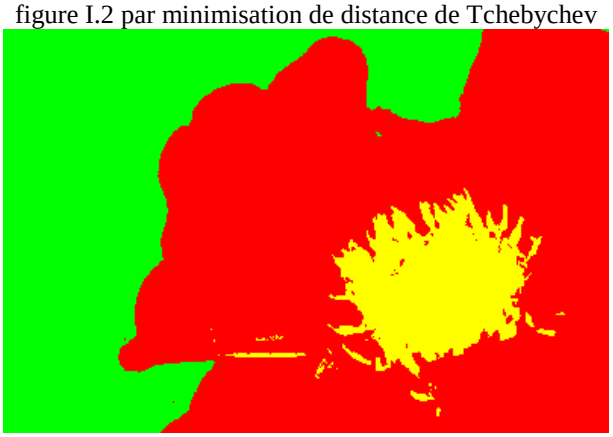


figure I.14 : Image résultant de la classification de la figure I.2 par minimisation de distance de Mahalanobis

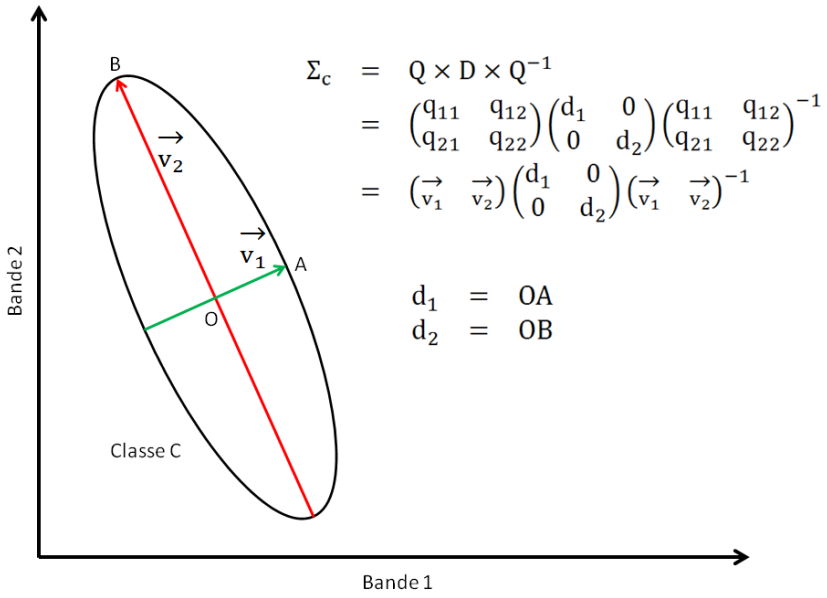


figure I.15 : Construction d'un ellipsoïde représentant une classe C suivant deux dimensions

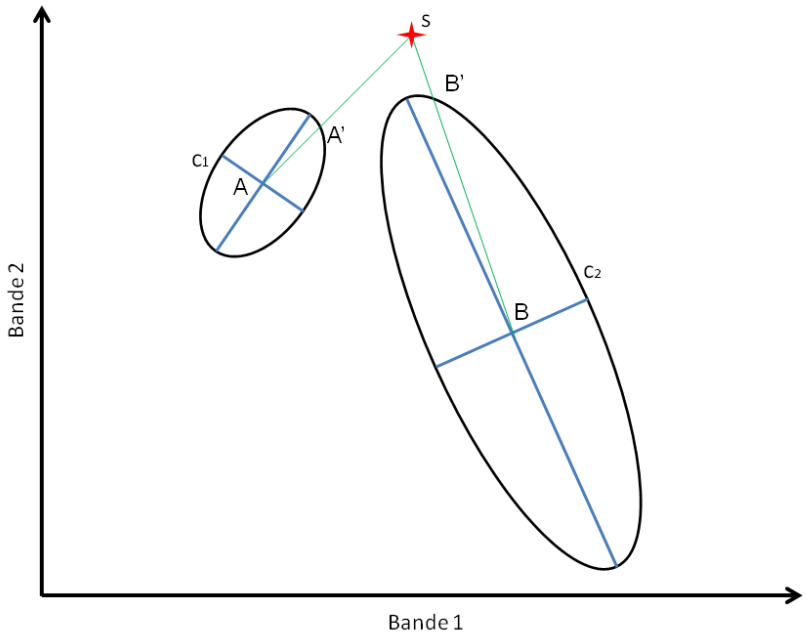


figure I.16 : Illustration de la nécessité de pondérer la distance euclidienne sA et sB par l'étendue de chaque classe (volume des ellipsoïdes). Le point s est plus proche de la classe c_2, d'où l'intérêt de la distance de Manahalobis (sA' et sB')

2.1.3. La méthode de classification avec notion de probabilité d'appartenance

2.1.3.1. La notion de probabilité et de loi de probabilité

Le premier usage du mot probabilité apparaît en 1370 avec la traduction de l'éthique à Nicomaque d'Aristote par Oresme et désigne alors « le caractère de ce qui est probable » (Oresme, 1940). Le concept de probable chez Aristote est défini dans les Topiques (de Pater, 1965) par : « Sont probables les opinions qui sont reçues par tous les hommes, ou par la plupart d'entre eux, ou par les sages, et parmi ces derniers, soit par tous, soit par la plupart, soit enfin par les plus notables et les plus illustres ».

La notion d'opinion collective introduite par Aristote illustre bien la probabilité de prise de décision. Chaque décision est prise en fonction de tous les observateurs, quelle que soit leur importance, ce qui implique une notion de normalisation et de collectivité.

La probabilité objective d'un évènement n'existe pas et n'est donc pas une grandeur mesurable, mais tout simplement une mesure d'incertitude, pouvant varier avec les circonstances et l'observateur, donc subjective. La seule exigence étant qu'elle satisfasse aux axiomes du calcul des probabilités (Saporta, 2006).

La théorie mathématique des probabilités ne dit pas quelle loi de probabilité mettre sur un ensemble IM parmi toutes les lois possibles. Ce problème concerne les applications du calcul des probabilités, et renvoie à la nature « physique » du concept de probabilité qui formalise et quantifie le sentiment d'incertitude vis-à-vis d'un évènement (Saporta, 2006).

Définissons la notion de probabilité dans un cadre mathématique.

Définition : On appelle *probabilité* ou *mesure de probabilité* sur un espace mesurable $(\mathrm{IM}, \mathcal{A})$ une application P de $\mathcal{A} \rightarrow [0,1]$ telle que :

i. $\forall\, \mathrm{A} \in \mathcal{A}, 0 \leq \mathrm{P}(\mathrm{A}) \leq 1$
ii. $\mathrm{P}(\emptyset) = 0$ et $\mathrm{P}(\mathrm{IM}) = 1$
iii. $\forall\, \mathrm{A_n}$ une suite d'ensembles disjoints de $\mathcal{A}, \mathrm{P}(\bigcup_{n=1}^{\infty} \mathrm{A_n}) = \sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A_n})$

La loi de probabilité la plus utilisée en classification probabiliste est la loi gaussienne car ses hypothèses sur les données permettent de réduire la complexité de calcul et le risque d'erreur. Rappelons que dans le cas de la classification, la loi gaussienne suppose qu'une classe peut être modélisée par son vecteur des moyennes et sa matrice de covariance (De Moivre, 1756)(figure I.15).

2.1.3.2. La classification par Maximum de Vraisemblance

La méthode de classification par Maximum de Vraisemblance fut développée par Fisher en 1922 (Fisher, 1922). Le principe de cette méthode est le suivant : chaque pixel de l'image sera associé à la classe dont il maximise la probabilité d'appartenance.

figure I.17 : Surface de décision de la méthode de classification par maximum de vraisemblance

figure I.18 : Image résultant de la classification de la figure I.2 par la méthode de classification par maximum de vraisemblance

La méthode de classification par maximum de vraisemblance revient à chercher c_s^{opt}, *id est* la classe à attribuer au pixel s, qui maximise sa probabilité d'appartenance *a posteriori* (équation I.9).

$$c_s^{opt} = \arg\max_{c_s} P(C = c_s | X = x_s) \tag{I.9}$$

Par application de la règle de Bayes, nous obtenons :

$$c_s^{opt} = \arg\max_{c_s} \frac{P(X = x_s | C = c_s) P(C = c_s)}{P(X = x_s)} \tag{I.10}$$

$$c_s^{opt} = \arg\max_{c_s} P(X = x_s | C = c_s) P(C = c_s) \tag{I.11}$$

En supposant que l'on ne dispose d'aucune information *a priori* sur les probabilités relatives d'occurrence des classes, $P(C = c_s)$ est constant pour tout s.

$$c_s^{opt} = \arg\max_{c_s} P(X = x_s | C = c_s) \tag{I.12}$$

En appliquant les hypothèses gaussiennes sur la structure de chaque partition c_s, nous obtenons ainsi l'équation I.13.

$$c_s^{opt} = \arg\max_{c_s} \frac{e^{-\frac{1}{2}(x-\mu_{c_s})^T \Sigma_{c_s}^{-1}(x-\mu_{c_s})}}{\left|\Sigma_{c_s}\right|^{1/2} 2\pi^{n/2}} \tag{I.13}$$

Avec n la dimension de l'univers des données IM (c'est à dire le nombre de plans de l'image, temporels ou spectraux).

Comme pour le classifieur par minimisation de distance de Mahalanobis, les surfaces de décision du maximum de vraisemblance sont représentées par des hyper-ellipsoïdes (figure I.17). Les aires des ellipsoïdes sont différentes pour la méthode de minimisation de distance de Mahalanobis et celle du maximum de vraisemblance, même si cette différence est assez faible dans notre exemple.

2.1.4. Les méthodes de classification non paramétriques à noyaux

A la différence des méthodes présentées précédemment, il n'y a ici aucune estimation de statistique à partir des échantillons. Ces méthodes sont donc particulièrement intéressantes lorsque le nombre de points de certains échantillons est limité. Dans certaines conditions, la nature des échantillons ne permet pas un échantillonnage suffisant, exemple des zones urbaines très hétérogènes et riches en diversité de classe. Inversement, lorsque le nombre d'échantillons est trop important, la complexité de calcul des méthodes de classifications non paramétriques à noyau devient trop élevée et les résultats peuvent être détériorés du fait de la considération de certains pixels non significatifs.

Parmi ces méthodes de classification non paramétriques à noyau, nous introduirons les méthodes Support Vector Machines et Réseaux neuronaux.

Nous n'entrerons pas ici dans les détails des méthodes et nous ne fournirons que les éléments nécessaires à la compréhension globale des méthodes et de leurs avantages et inconvénients vis-à-vis des autres méthodes citées précédemment, laissant ainsi le lecteur intéressé se diriger vers la bibliographie fournie.

2.1.4.1. La méthode des Support Vectors Machines (SVM)

Les méthodes d'apprentissage à noyau et plus particulièrement les Support Vectors Machines (SVM) ont été introduites depuis plusieurs années en théorie d'apprentissage pour la classification et la régression (Vapnik, 1998). L'application des SVM a commencé avec la reconnaissance de texte (Joachims, 1998) puis la reconnaissance de visage (Osuna, Freund, & Girosit, 1997). Plus récemment, cette méthode est appliquée à la classification d'images de télédétection (Huang, Davis, & Townshend, 2002; Melgani & Bruzzone, 2004; Roli & Fumera, 2001; J. Zhang, Zhang, & Zhou, 2001).

Le principe original de la méthode est simple, il consiste à rechercher une surface de décisions ou hyperplan afin de séparer deux classes. La surface de décisions sera déterminée par un sous-ensemble des échantillons d'apprentissage afin de maximiser la discrimination des deux classes. Les éléments de ce sous ensemble sont appelés *vecteurs de support* (figure I.19).

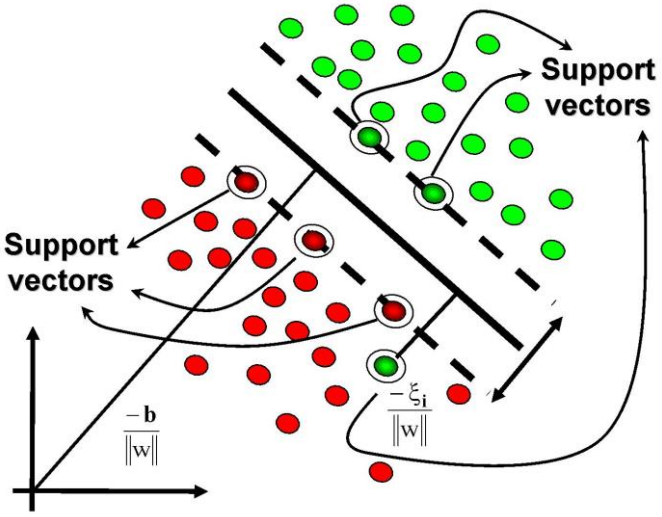


figure I.19 : Illustration de la notion de vecteurs de support ou « support vectors », source anonyme

Cette approche est intéressante puisque l'optimisation est supposée maximiser directement la performance de la classification. Le noyau (ou système d'équations de l'hyperplan) peut prendre plusieurs formes et nous citerons les plus utilisées : noyau linéaire (figure I.20 et figure I.21), noyau polynomial (figure I.22 et figure I.23) et noyau gaussien (figure I.24 et figure I.25).

Lorsque les deux classes ne sont pas linéairement séparables, la méthode consiste à projeter les données dans des dimensions supérieures dans lesquelles les classes seront linéairement séparables.

L'algorithme de classification SVM a été initialement prévu pour la classification binaire, néanmoins deux méthodes existent pour étendre cet algorithme à la classification multi-classes.

La première méthode est appelée « un contre les autres » et consistent à déterminer pour chaque classe une surface de décisions entre la classe et toutes les autres, ce qui donne #C surfaces de décision. Cette méthode est illustrée aux figures I.20, I.22 et I.24.

La deuxième méthode appelée « un contre un » consiste quant à elle à calculer la surface de décision entre chaque classe et chacune des autres classes, d'où l'obtention de $\#C \times (\#C - 1)$ surfaces de décision. Cette méthode est illustrée aux figures I.21, I.23 et I.25. La seconde méthode est la plus performante mais aussi la plus couteuse en calcul.

Afin de combiner les surfaces de décision, des méthodes de fusion sont appliquées. Certaines de ces méthodes seront citées au chapitre 4.

Les figures I.27 à I.29 illustrent les résultats de ces trois classifications par méthode SVM sur l'exemple de la fleur.

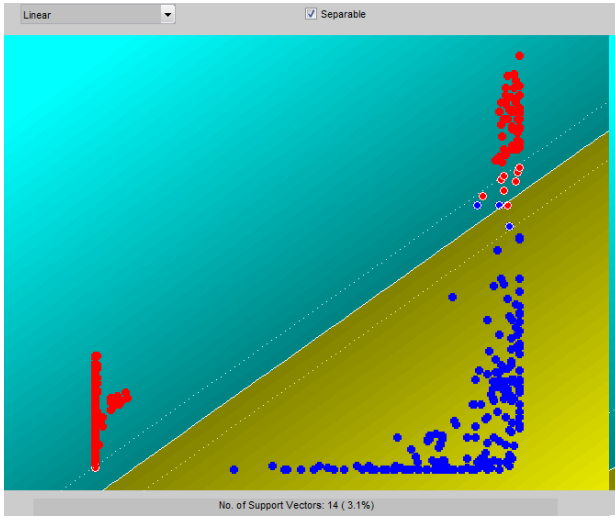


figure I.20 : Illustration hyperplan separateur pour le cas « 1 contre les autres » et un SVM à noyau linéaire

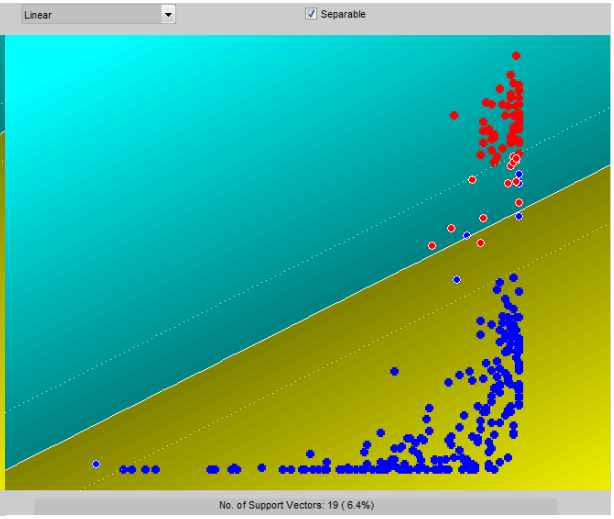


figure I.21 : Illustration hyperplan separateur pour le cas « 1 contre 1 » et un SVM à noyau linéaire

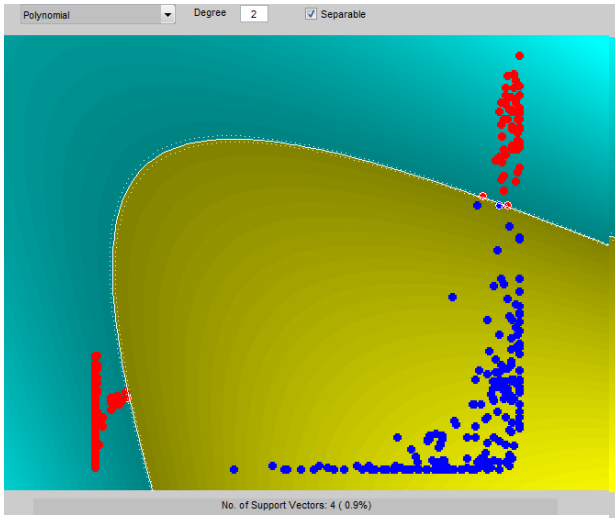


figure I.22 : Illustration hyperplan separateur pour le cas « 1 contre les autres » et un SVM à noyau polynomial d'ordre 2

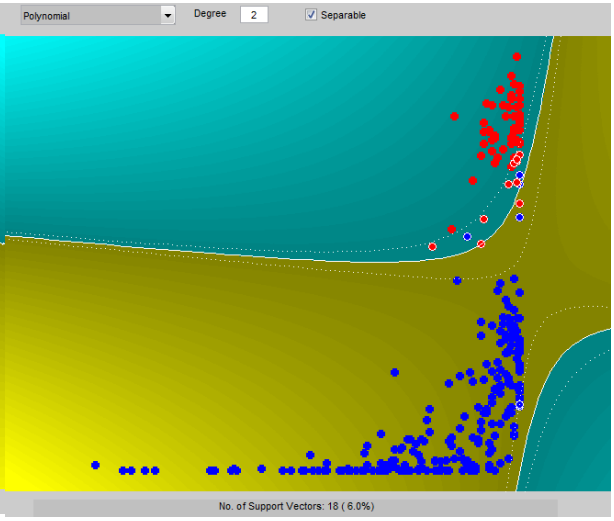


figure I.23 : Illustration hyperplan separateur pour le cas « 1 contre 1 » et un SVM à noyau polynomial d'ordre 2

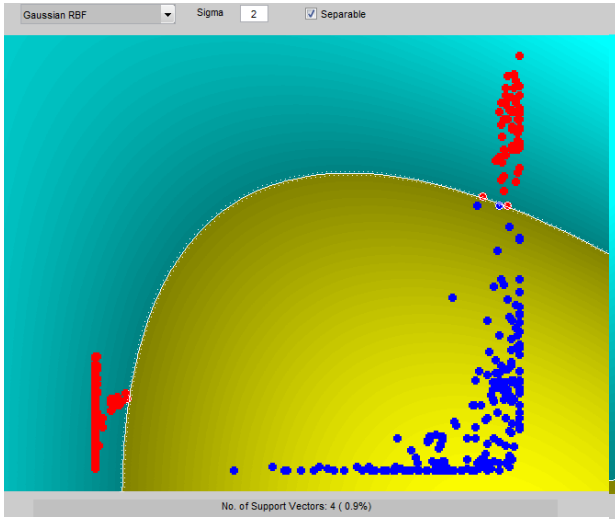


figure I.24 : Illustration hyperplan separateur pour le cas « 1 contre les autres » et un SVM à noyau gaussien

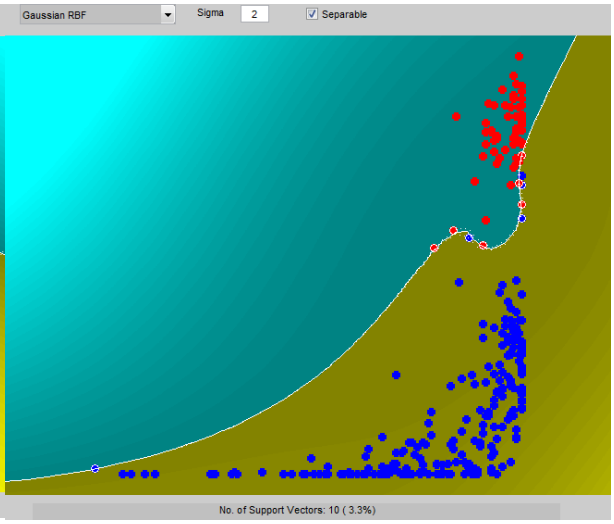


figure I.25 : Illustration hyperplan separateur pour le cas « 1 contre 1 » et un SVM à noyau gaussien

2.1.4.2. La méthode des Réseaux Neuronaux

Le modèle des réseaux de neurones le plus utilisé pour la classification d'images de télédétection est le perceptron multicouche avec entrainement par algorithme de retro propagation (Rumelhart, Hintont, & Williams, 1986).

Le perceptron multicouche comprend une couche d'entrée (les images à classer) et une couche de sortie (les classes à attribuer). Entre ces deux couches, une ou plusieurs couches cachées peuvent être positionnées. Après la phase d'apprentissage, le système agit en « boite noire » : à un pixel d'entrée correspond une classe de sortie.

La correspondance attendue entre la couche d'entrée et la couche de sortie ne pouvant être absolue, le perceptron se comporte comme un approximateur. Un compromis devra donc être trouvé entre le nombre de couches cachées (et de neurones les composant) et la qualité d'approximation du système (figure I.26).

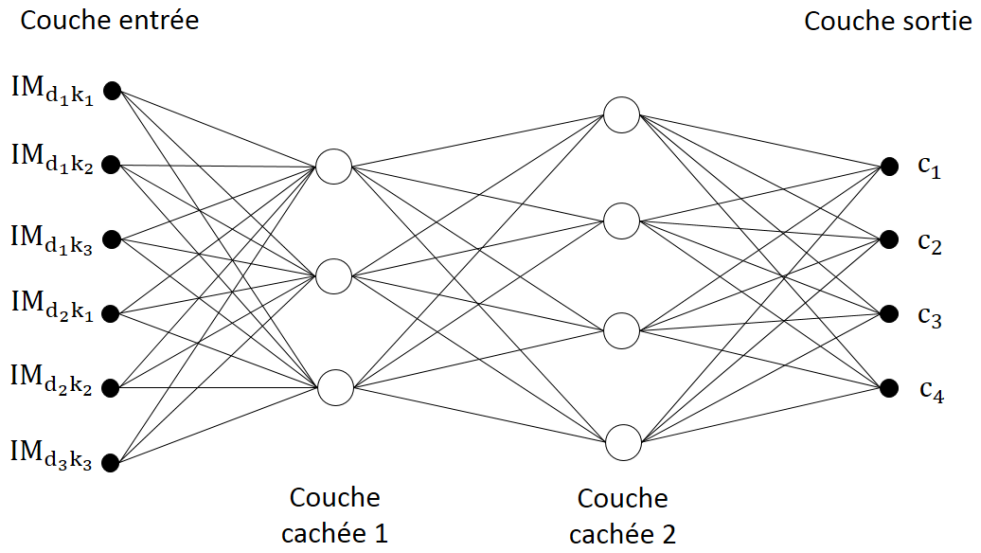


figure I.26 : Exemple de perceptron à deux couches cachées avec en entrée les images par date et par canal et en sortie une décision sous forme d'attribution de classe de l'ensemble C.

Les détails des algorithmes et les démonstrations associées sont disponibles dans (Collet, 2001).

La figure I.30 illustre la classification par méthode des réseaux de neurones sur l'exemple de la fleur.

figure I.27 : Image résultant de la classification de la figure I.2 par la méthode de classification SVM à noyau linéaire

figure I.28 : Image résultant de la classification de la figure I.2 par la méthode de classification SVM à noyau polynomial de degré 2

figure I.29 : Image résultant de la classification de la figure I.2 par la méthode de classification SVM à noyau gaussien

figure I.30 : Image résultant de la classification de la figure I.2 par la méthode de classification par réseaux de neurones avec noyau polynomial d'ordre 2

2.2. La méthode de classification dit « objet »

Le principe des méthodes de classification dites « objet » est de ne plus considérer l'élément à classer comme un pixel mais plutôt comme un ensemble de pixels que l'on nommera *objet*. Cet objet peut être défini de plusieurs façons, l'approche qui sera développée par la suite est la définition d'objet par segmentation.

2.2.1. La notion d'objet obtenu par segmentation

Afin d'obtenir ces objets, il convient de regrouper les pixels par affinité, le plus généralement par homogénéité des valeurs.

Dans le cas d'un regroupement par homogénéité de valeurs, la formation de ces objets commence par le calcul d'une image de gradient (ou image de force de contour) ce qui permet de donner un poids à chaque contour suivant la force de l'hétérogénéité détectée. Celle-ci peut avoir été obtenue par tout type de filtre (Sobel, etc…).

La détection des contours se fait de la manière suivante : à partir d'un premier point du contour présumé, on effectue par connexité un suivi de contour grâce aux renseignements obtenus durant la recherche tels que sa direction ou sa longueur. Avec ces éléments, le contour est reconstruit pour en faire un contour fermé en regroupant les points détectés (figure I.31).

Une fois que les contours sont déterminés et fermés, on peut segmenter l'image par la méthode des bassins versants (Beucher & Lantuéjoul, 1979). La segmentation se base sur l'image de force de contours qui s'interprète comme une surface dont les lignes de crête correspondent aux contours de l'image. Pour détecter les lignes de crête on simule une « inondation » de la surface. Ses « vallées » vont être noyées et des bassins vont se former. Dans un premier temps on « remplit ces bassins » jusqu'à une hauteur prédéterminée S1 (le seuil de détection), telle que seuls les pics les plus hauts émergent, séparant l'eau en plusieurs « lacs » distincts. L'inondation jusqu'au niveau S1 fait disparaître les sommets les plus bas, c'est-à-dire ceux qui ne représentent pas de contours significatifs, puis un identificateur est attribué, unique à chaque bassin formé (figure I.32.a).

Une fois que les bassins sont formés et étiquetés, on fait monter l'eau d'un niveau supplémentaire. En grossissant, des bassins qui étaient jusqu'alors isolés vont se toucher. Ainsi si un point est le lieu de réunion de deux bassins portant des étiquettes distinctes, ce point peut être marqué définitivement comme frontière (figure I.32.b). L'exécution se termine lorsque tous les points de la surface sont immergés. A ce stade les points marqués comme frontières constituent les contours de l'image.

L'image segmentée par les bassins versants est étiquetée et les caractéristiques (moyenne et variance par exemple) sont calculées sur chacune des régions. Les régions voisines sont ensuite comparées par l'intermédiaire d'un critère. Les couples de régions qui ne satisfont pas les limites imposées seront alors fusionnés. Ce principe permet ainsi de regrouper les régions homogènes. On peut ainsi régler la précision de la segmentation obtenue. Les figures I.34 à I.36 illustrent le principe de la segmentation et de la fermeture des contours sur l'image de la fleur.

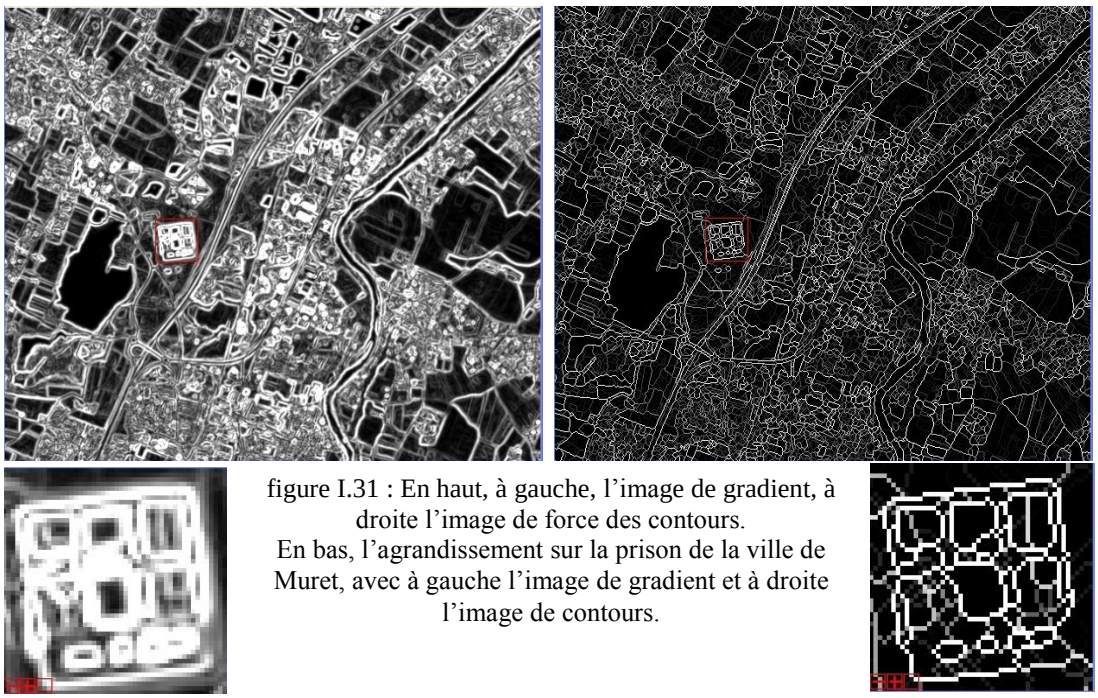


figure I.31 : En haut, à gauche, l'image de gradient, à droite l'image de force des contours.
En bas, l'agrandissement sur la prison de la ville de Muret, avec à gauche l'image de gradient et à droite l'image de contours.

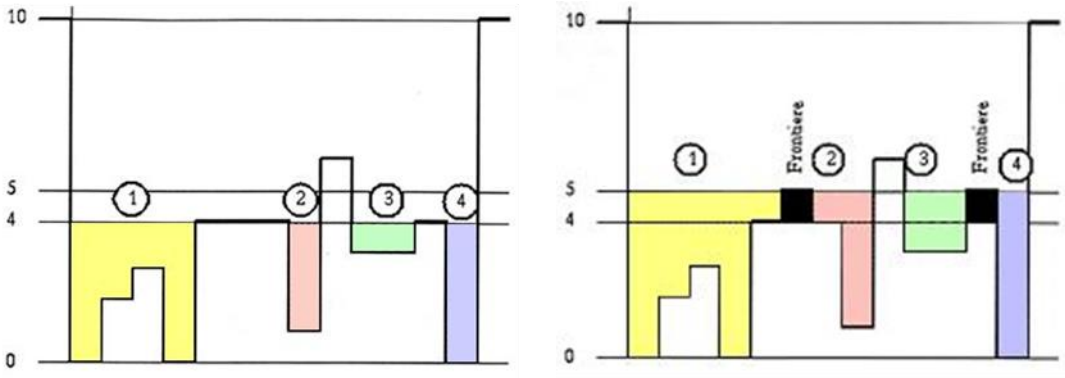


figure I.32 : (a) à gauche l'inondation au niveau 4
(b) à droite l'inondation au niveau 5 pour déterminer les frontières du niveau 4

figure I.33 : Image originale de la fleur

figure I.34 : Image des gradients obtenus à partir des écarts-type calculés sur l'image de la figure I.33

figure I.35 : Image des forces de contour

figure I.36 : Image obtenu à partir de l'image de la figure I.35, d'un choix de force et d'une fermeture du contour.

2.2.2. La classification des objets

Afin de classer les objets, il convient de calculer une distance entre deux ensembles. Pour cela, une distance et une divergence permettent de décider de cet appariement : la distance de Bhattacharyya (équation I.6) et la divergence de Kullbach-Leibler symétrique (équation I.7).

La distance de Bhattacharyya est la plus souvent employée car elle dispose de propriétés supplémentaires vis-à-vis de la divergence de Kullbach-Leibler (paragraphe 2.1.2).

Cet algorithme attribue à l'objet **o** la classe la plus proche c_o^{opt} (équation I.14).

$$c_o^{opt} = \arg\min_{c_o \in C} \; d_{bhata}(o, c_o) \qquad \text{(I.14)}$$

Cette méthode est par définition très dépendante du choix lors du seuillage des objets. Prenons l'exemple de deux seuillages différents et de leur étiquetage par la méthode des bassins versants (figure I.37 et figure I.38). Les images classées résultantes sont respectivement présentées aux figures I.39 et I.40. Dans les deux cas, nous observons que le choix de la taille de l'objet est bien adapté à la classe verte ; dans le cas de l'image de la figure I.40, la taille des objets est bien adaptée à la classe jaune mais dans aucun des cas à la classe rouge. A chaque classe devrait donc correspondre un choix de seuil différent, ce qui compliquerait l'algorithme et le choix du seuil. Ceci est une perspective de recherche intéressante dans le domaine de la classification d'objets.

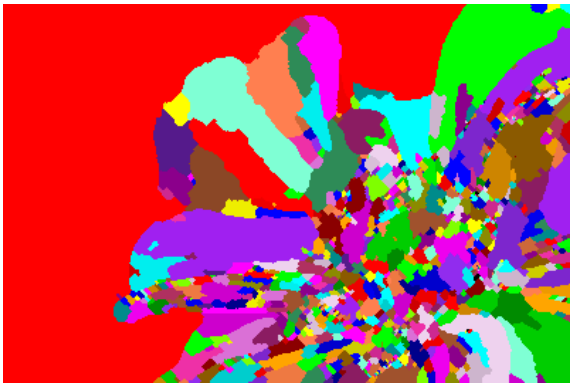


figure I.37 : Etiquetage de l'image fleur avec un seuil de valeur 10 sur une échelle de 1 à 50

figure I.38 : Etiquetage de l'image fleur avec un seuil de valeur 20 sur une échelle de 1 à 50

figure I.39 : Classification des objets de l'image de la figure I.37

figure I.40 : Classification des objets de l'image de la figure I.38

2.3. Les méthodes de classification dites « globales »

L'intérêt de ces méthodes se situe dans l'aspect spatial du traitement qui permet de donner une probabilité a priori d'une classe pour toute l'image, c'est-à-dire dans sa globalité. En pratique, les hypothèses markoviennes permettent de réduire cette dépendance globale aux voisins les plus proches de chaque point à classer. Cet outil simple et efficace s'appelle les champs de Markov.

Ces méthodes ont été introduites par Geman & Geman (Geman & Geman, 1984) pour la classification d'images bruitées. Ces méthodes ont été ensuite appliquées aux images radar mais également à d'autres types d'images (Kato, Zerubia, & Berthod, 1999; Kato, Zerubia, & Berthod, 1992). L'idée fondamentale des méthodes de classification globale repose sur les Champs de Markov (CAM) (Cocquerez & Philipp-Foliguet, 1995; Pony, Descombes, & Zerubia, 2000).

Commençons par définir les notions de voisinage et de champs de Markov.

2.3.1. La notion de voisinage

Rappelons que les notations sont disponibles au tableau I.1.

Définition : Un champ aléatoire V défini sur un ensemble de sites S est *markovien* si et seulement si la probabilité d'observer une valeur x_s en un site s ne dépend que d'un nombre fini de sites voisins V_s (avec V le système de voisinage).

Considérons une famille arbitraire **V** de « voisinages locaux » de S :

$$V = \{V_s, s \in S \backslash \{s\}\} \tag{I.15}$$

A cette famille de voisinage est associée une classe de distribution de probabilités appelée *champ de Markov* et caractérisée par la propriété I.16.

Propriété : Soit $p(x_s)$ la probabilité d'occurrence de x_s, $p(x_s) > 0$ pour toute configuration de x_s alors :

$$p(x_s | x_r, r \in S \backslash \{s\}) = p(x_s | x_r, r \in V(s) \backslash \{s\}) \tag{I.16}$$

La valeur des autres sites r est supposée connue.

A un système de voisinage donné correspond un ensemble de cliques (figure I.41).

Définitions : Une *clique* est un ensemble de points du treillis mutuellement voisins, l'*ordre* d'une clique correspond au nombre de sites qui composent la clique.

On notera **CL** l'ensemble des cliques et **cl** une clique de CL.

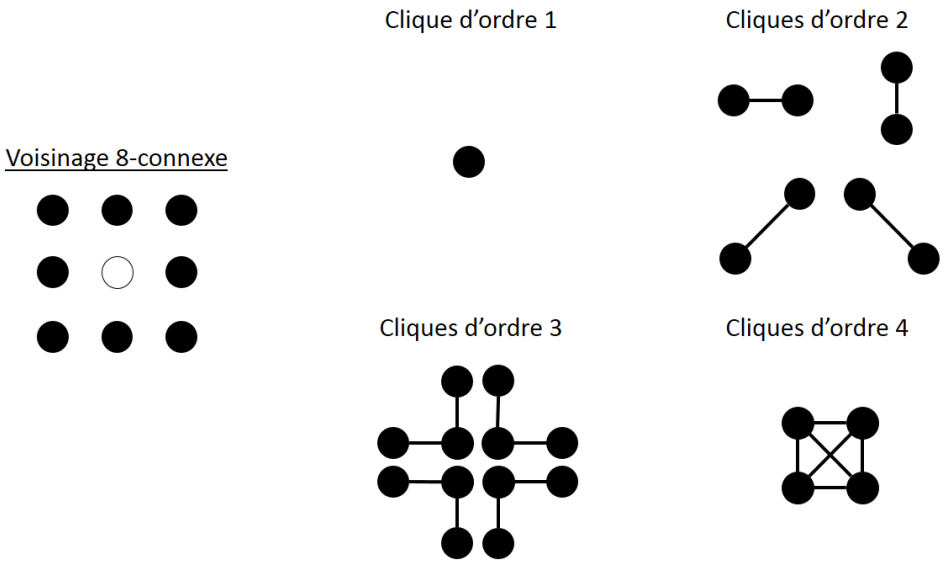


figure I.41 : Schéma représentant les cliques possibles pour un voisinage 8-connexe.

Définition : On appelle *fonction énergie* toute fonction $\mathbf{U}(\mathbf{c_x})$ que l'on peut décomposer sur l'ensemble des cliques CL de l'image (σ-additive) et qui peut s'exprimer sous la forme d'une somme de potentiels associés à ces cliques :

$$U(c_s) = \sum_{cl \in CL} V_{cl}(c_s) \text{ avec s le pixel central et } c_s \in C \qquad (I.17)$$

En général, les champs de Markov sont stationnaires (*id est* indépendant de la position du site s) et la fonction potentielle est indépendante de la position de la clique (Ducrot, 2005).

Le théorème de Hammersley-Clifford (Besag, 1974; Hammersley & Clifford, 1968) permet de caractériser les champs de Markov en termes globaux à partir de la probabilité *a priori* d'une configuration des classes p(C=c). Il permet d'établir une correspondance entre un champ de Markov et un champ de Gibbs lorsqu'aucune réalisation de C n'est de probabilité nulle.

Théorème Hammersley-Clifford: Un champ aléatoire C défini sur un réseau de site S fini et dénombrable est un champ de Markov relativement à V un système de voisinage borné si et seulement si C est un champ de Gibbs de potentiel associé à V.

La distribution de probabilité a priori p(c) est donc une distribution de Gibbs défini par :

$$p(c) = \frac{1}{Z} e^{-\frac{U(c)}{T}}, \text{avec Z constante de normalisation et T variable de "temperature"}$$

La constante de normalisation Z est aussi appelée fonction de partition et assure que p(c) définisse bien une mesure de probabilité. Z s'exprime par l'équation I.18.

$$Z = \sum_{c \in C} e^{-\frac{U(c)}{T}} \qquad (I.18)$$

La variable de « température » T a été ajoutée au processus ICM afin de pondérer la fonction énergie. En pratique, la variation de ce paramètre de température permet de régler la vitesse convergence de l'algorithme. Ce principe est issu de la méthode du recuit simulé (Kirkpatrick, Jr, & Vecchi, 1983)).

Nous voyons donc que la probabilité d'une configuration ne dépend que d'un ensemble d'interactions locales. En outre, plus l'énergie totale U(c) est grande, moins la configuration est possible.

Plusieurs modèles existent pour définir la fonction potentielle $V_{cl}(x)$. Nous citerons en exemple le modèle de Potts (Potts, 1952) ou encore le « chien-modèle » (Descombes, Mangin, Sigelle, & Pechersky, 1995) qui permet de conserver les structures fines et linéaires à partir d'une image de contour (paragraphe 2.2.1).

2.3.2. Les méthodes de classification contextuelle

Pour des estimateurs de Maximisation *a posteriori* (MAP), plusieurs méthodes de classification contextuelle basée sur un CAM existent telles que la méthode du recuit simulé (algorithme stochastique) ou l'Iterated Conditional Mode (ICM, algorithme déterministe). Les modèles stochastiques sont beaucoup plus performants et permettent une convergence vers un maximum global alors que les algorithmes déterministes sont plus rapides mais peuvent converger vers un maximum local (Ducrot, 2005).

Nous présenterons dans la suite du manuscrit plusieurs méthodes basées sur l'utilisation de l'algorithme ICM du fait de sa performance, de sa rapidité et de sa robustesse.

2.3.3. Minimisation de l'énergie a posteriori par la méthode ICM

La méthode Iterative Conditional Mode (ICM) est une méthode de recherche déterministe proposée par Besag (Besag, 1974). Cette méthode itérative permet d'approcher la solution optimale du MAP en affectant à chaque pixel s la classe qui maximise la probabilité conditionnelle d'appartenance.

Dans un cadre bayésien, le problème peut être modélisé de la façon suivante : soit C une variable aléatoire représentant le champ des étiquettes possibles, X étant une variable aléatoire représentant les valeurs des pixels s (en général $C \subset \mathbb{N}$ et $X \subset \mathbb{R}^n$). Par application du Maximum A Posteriori (MAP), la règle de Bayes permet d'écrire :

$$P(C = c|X = x) = \frac{P(X = x|C = c)P(C = c)}{P(X = x)} \qquad (I.19)$$

$P(X = x)$ est une constante et $P(X = x|C = c)$ décrit le processus d'observation des données.

Plusieurs hypothèses sur les données sont posées pour la résolution de ce problème :

Hypothèse 1 : indépendance conditionnelle des pixels :

$$P(X = x|C = c) = \prod_{s \in IM} P(X = x_s|C = c_s) \tag{I.20}$$

Hypothèse 2 : hypothèse markovienne :

$$P(C = c) = \frac{1}{Z} e^{-\frac{U(c)}{T}} \tag{I.21}$$

Rappelons que Z est la constante de normalisation et T la variable de « température ».

Nous pouvons ainsi écrire :

$$\begin{aligned} P(C = c_s|X = x_s) &\approx P(X = x_s|C = c_s)P(C = c_s) \\ &\approx e^{\ln P(X = x_s|C = c_s) - U(c_s)} \\ &\approx e^{-W(c_s, x_s)} \end{aligned} \tag{I.22}$$

Avec

$$\begin{aligned} W(c_s, x_s) &= -\ln p(x_s|c_s) + U(c_s) \\ &= -\ln p(x_s|c_s) + \sum_{cl \in CL} V_{cl}(c_s) \end{aligned} \tag{I.23}$$

On constate bien que la distribution *a posteriori* est une distribution de Gibbs et donc que le champ C conditionnellement à X est un champ de Markov d'après le théorème de Hammersley-Clifford. La configuration c_s qui nous intéresse est donc celle qui maximise la probabilité *a posteriori* $P(C = c_s|X = x_s)$ ou encore celle qui minimise l'énergie $W(c_s, x_s)$.

2.3.4. La classification par méthode ICM

Rappelons que la classification par méthode ICM est une méthode itérative basée en règle générale sur un classifieur de type Maximum de Vraisemblance classique comme présenté au paragraphe 2.1.3.2. Puis à chaque itération de l'algorithme ICM, une nouvelle probabilité est calculée en fonction du contexte nominal du pixel *id est* la classe attribuée au voisinage du pixel à classer. L'écriture de la probabilité *a posteriori* à maximiser a été introduite à l'équation I.22

A chaque pixel s de l'image, le but est de trouver c_s^{opt} *id est* la classe à attribuer au pixel s qui maximise la probabilité *a posteriori*.

$$c_s^{opt} = \arg\max_{c_s} P(C = c_s|X = x_s) \tag{I.24}$$

Nous avons montré que ce problème était équivalent à :

$$c_s^{opt} = \arg\max_{c_s} P(X = x_s | C = c_s) P(C = c_s) \quad \text{(I.25)}$$

Dans le cadre de la modélisation gaussienne des données, on suppose que chaque classe c peut être modélisée par une loi gaussienne. Pour chaque classe c, le vecteur des moyennes et la matrice de covariance des pixels pour chaque date d et bande spectrale k sont calculés et notés respectivement μ_c et Σ_c. On peut donc réécrire l'équation I.25 de la manière suivante :

$$c_s^{opt} = \arg\max_{c_s \in C} \left(\frac{e^{-\frac{1}{2}(x-\mu_{c_s})^T \Sigma_{c_s}^{-1}(x-\mu_{c_s})}}{\left|\Sigma_{c_s}\right|^{1/2} 2\pi^{d/2}} \times e^{-\frac{U(c_s)}{T}} \right) \quad \text{(I.26)}$$

Ce qui revient à minimiser l'énergie $W(c_s, x_s)$, d'où

$$c_s^{opt} = \arg\min_{c_s \in C} \left(\frac{1}{2} (x-\mu_{c_s})^T \Sigma_{c_s}^{-1} (x-\mu_s) + \ln\left(\left|\Sigma_{c_s}\right|^{1/2} 2\pi^{d/2} \right) + \frac{U(c_s)}{T} \right) \quad \text{(I.27)}$$

Rappelons que T est la variable de température qui nous permet d'accélérer la convergence de c_s^{opt} vers son optimum au fur et à mesure des itérations. Ce paramètre permettra également d'empêcher la convergence de l'algorithme vers un maximum local.

L'algorithme est composé de deux types d'itération : les itérations intérieures ou itération markoviennes qui permettent la gestion de la convergence et les itérations extérieures basées sur le principe des nuées dynamiques et qui permet la convergence des statistiques. En pratique pour la classification supervisée, il n'y a pas d'itérations extérieures.

L'algorithme ICM est présenté ci-après et illustré par la figure I.42. La figure I.43 illustre le résultat de la classification de l'image de la fleur par la méthode de classification ICM.

Algorithme de classification Iterated Conditional Mode (ICM) :

Entrée IM : ensemble d'images à classer
C : l'ensemble des classes de référence

Initialisation Estimations des statistiques
Classification par Maximum de Vraisemblance
(ou tout autre classifieur)

Tant que le nombre d'itérations extérieures n'est pas atteint ou que le critère d'arrêt n'est pas vérifié
Ré-estimation des statistiques de C
Pour chaque itération intérieure
Calcul des fonctions énergie en chaque pixel
Classification de chaque point avec utilisation de la fonction énergie
Fin Pour
Décroissance de la température de convergence

Fin Tant que

Sortie	Image classée

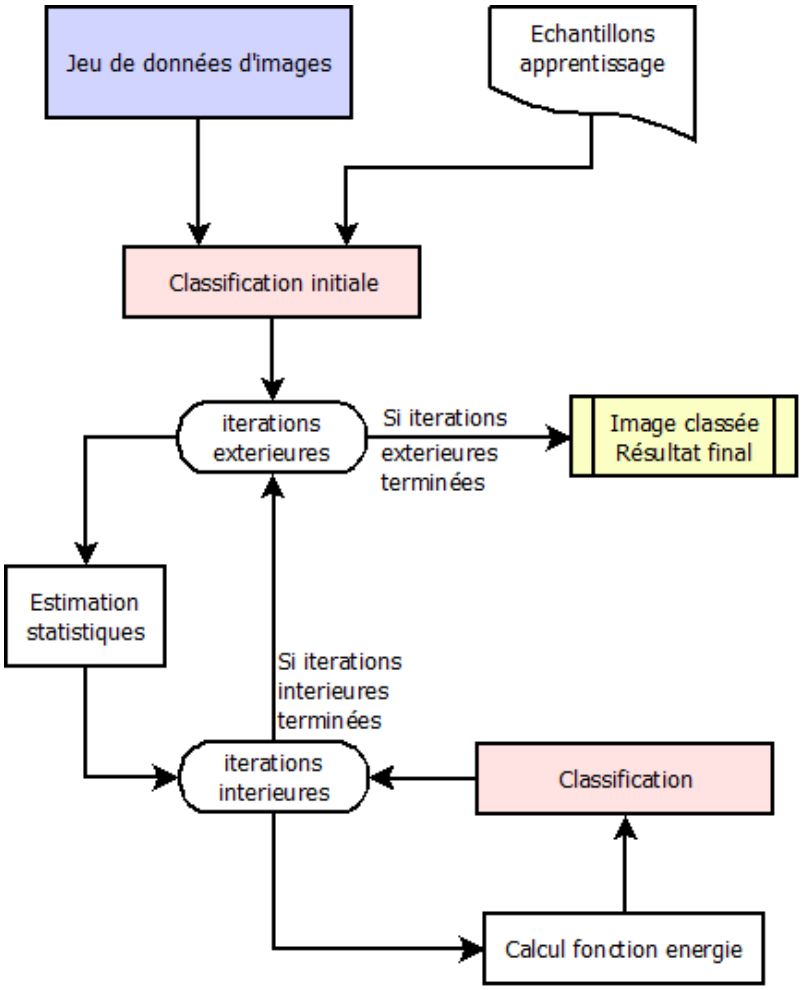


figure I.42 : Schéma illustrant le déroulement de l'algorithme ICM, valable pour le mode supervisée et non supervisée

figure I.43 : Image résultat de la classification par méthode ICM

2.4. Amélioration du processus ICM : ajout d'informations exogènes

Une amélioration importante a été apportée à la méthode ICM pour qu'elle soit utilisable dans toutes les conditions (type de données) et sur de larges zones d'étude (taille d'image, gestion des imperfections). J'ai ainsi introduit l'utilisation de données exogènes, qualitatives ou quantitatives, au processus de classification afin de l'améliorer. Ces données peuvent être, par exemple, l'information de l'altitude, la restriction de classes ou encore le masquage des données altérés ou manquantes.

La première information apportée est une restriction des classes possibles lors du processus de décision en fonction du contexte dans lequel se trouve le pixel. Cette méthode se base sur la théorie des possibilités. Prenons l'exemple concret présenté à la figure I.44. Nous avons ajouté une information de type qualitative sur la possibilité de présence de certaines classes en fonction de l'altitude dans le massif pyrénéen. Nous avons, par exemple, interdit la présence de cultures au-dessus de 1300 mètres d'altitude. Cette méthode permet d'améliorer la rapidité du processus et de réduire les confusions liées à la nature des échantillons, et donc aux erreurs *a priori*.

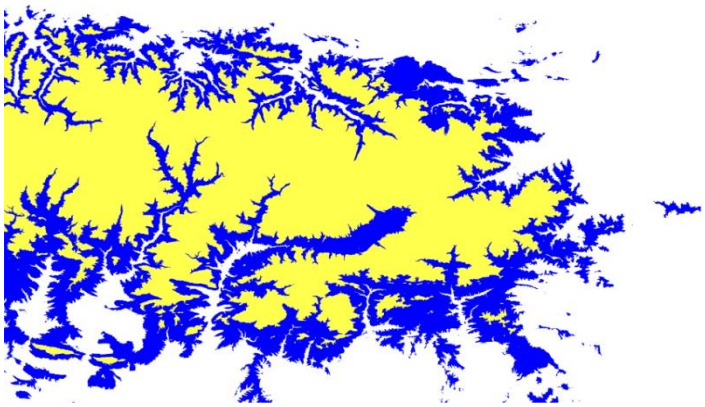


figure I.44 : Exemple de restriction de classes possibles en fonction des régions prédéfinies : ici les Pyrénées orientales avec une restriction par palier d'altitude. En blanc de 0 à 800m, en bleu de 800 à 1300m et en bleu plus de 1300m.

La seconde information apportée est celle du marquage du manque ou de l'altération de données. En certains points, l'information acquise peut être altérée ou masquée (défauts de capteur, nuages, etc.). Il convient donc de masquer ces informations à l'algorithme de classification afin que l'initialisation et le déroulement de la classification se fasse correctement. Cette méthode est particulièrement délicate avec l'utilisation de la méthode paramétrique ICM du fait de la présence d'un nombre important de combinaisons d'apprentissage.

3. La classification non supervisée et son interprétation

Comme nous l'avons vu en introduction de ce chapitre, la limitation des méthodes de classification supervisée, quel que soit le classifieur, réside dans la nécessité d'utiliser des échantillons et donc une vue restrictive et non objective de la réalité. Il est même parfois impossible d'effectuer une classification supervisée faute de disponibilité de vérités terrain (année passée, zone inconnue ou inaccessible).

Dans ces deux cas, il faut alors appliquer des méthodes de classifications non supervisées. Le principe de ces méthodes est similaire aux méthodes supervisées présentées précédemment, à la différence près qu'il faut ajouter une étape d'apprentissage non dirigée ou d'exploration, qui permettra d'initialiser le processus.

Parmi les méthodes de classification non supervisées, nous citerons la méthode du K-means (Duda, Hart, & Stork, 2012) et celle de l'ISODATA (Ball & Hall, 1965). Nous avons également associé à l'algorithme ICM un processus d'apprentissage non supervisé basé sur les nuées dynamiques (Diday, 1971). Deux exemples de cette méthode de classification sont présentés aux figures I.45 et I.46.

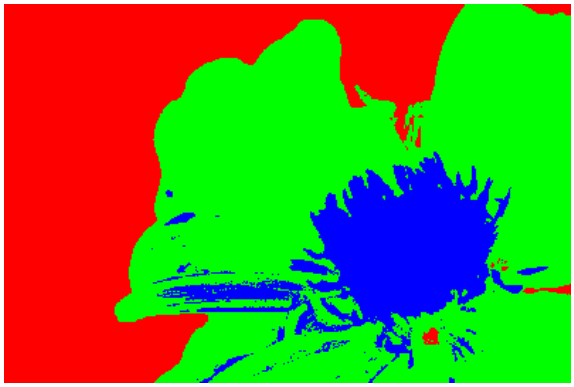


figure I.45 : Image résultante d'une classification non supervisée de type ICM avec 3 classes.

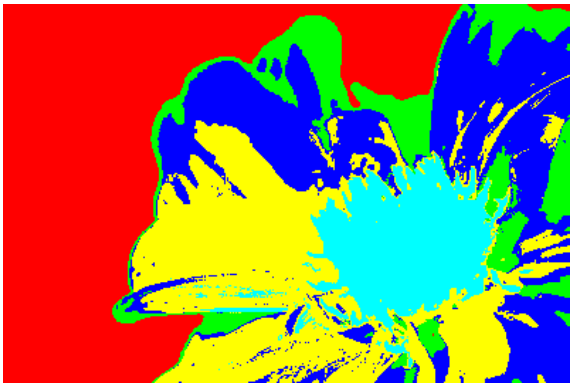


figure I.46 : Image résultante d'une classification non supervisée de type ICM avec 5 classes.

Ces méthodes ne sont toutefois pas entièrement automatique et peuvent nécessiter la définition arbitraire du nombre de classes (ou intervalle de valeurs) ou encore de la manière de regrouper et séparer des classes.

Pour une utilisation thématique des résultats, une interprétation de l'image obtenue par la classification non supervisée est nécessaire, ce qui revient à trouver une étiquette adéquate à chaque classe obtenue par la classification non supervisée.

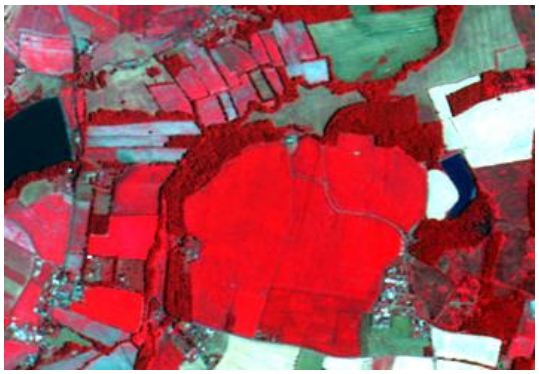
Il convient de préciser que l'ensemble des classes dépend généralement de la zone à classer, de la résolution du capteur ou encore de la nature du capteur lui-même. Une approche nouvelle visant à répondre à ce problème sera présentée au chapitre 5.

4. Application et comparaison des différents classifieurs avec des données de télédétection multidimensionnelles.

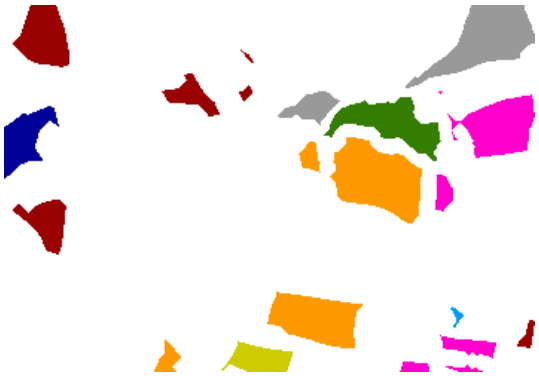
Nous présenterons ici quelques résultats afin d'illustrer la comparaison des classifieurs introduits dans ce chapitre. L'application est la suivante : nous allons effectuer la cartographie de l'occupation du sol à partir d'images Formosat-2 (optique, 8 mètres résolution, 4 bandes spectrales : Bleu, Vert, Rouge et Proche-Infrarouge) sur l'année 2009 et d'un ensemble de classes thématiques acquises par des vérités terrain. Les caractéristiques de ce capteur, des images et des classes utilisées sont présentés en annexe 1.3. Les images utilisées possèdent chacune 16 dates et 4 bandes spectrales et une emprise de 24x24 km². L'évaluation ne sera ici que visuelle, la validation numérique à l'aide d'un second jeu d'échantillons de vérification sera présentée en application du chapitre 2.

Des extraits de résultats sont présentés aux figures I.47 et I.48. La méthode de classification par seuillage est la plus mauvaise, peu de points sont correctement classés. Les différences entre les autres méthodes sont mineures, nous citerons notamment les confusions entre classes de cultures pour la minimisation de distance euclidienne (e), la bonne classification des cultures pour l'algorithme ICM (l) et l'amélioration par rapport au maximum de vraisemblance (g), la bonne classification des classes de bâti et d'eau pour l'algorithme SVM notamment pour le lac de droite (h et i) et enfin les confusions de la classification objet pour les parcelles hétérogènes (k).

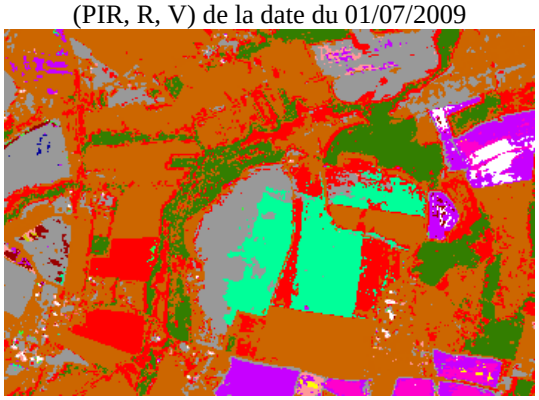
(a) Composition colorée RVB en couleur réelle de la date du 01/07/2009



(b) Composition colorée RVB en fausse couleur (PIR, R, V) de la date du 01/07/2009

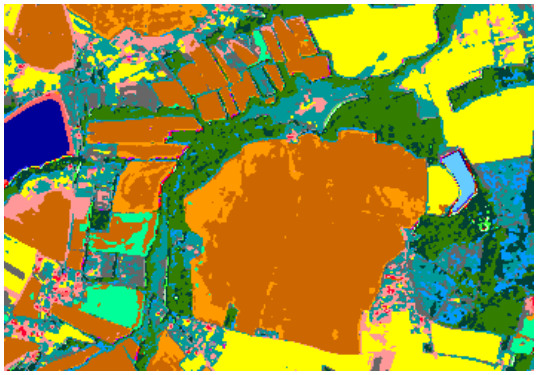


(c) Echantillons utilisés (apprentissage et vérification)

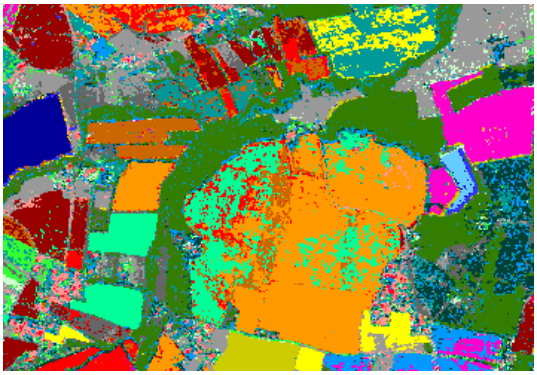


(d) Méthode de seuillage des parallélépipèdes

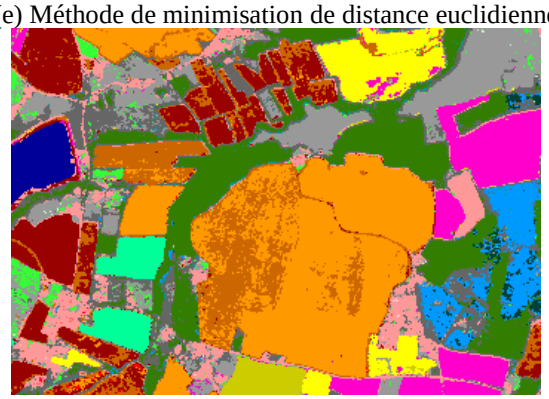
figure I.47 : Applications aux images Formosat-2 année 2009, (a) et (b) exemples de compositions colorées, (c) les échantillons utilisés pour la classification et la validation et (d) le résultat de la classification pour la méthode de seuillage des parallélépipèdes.



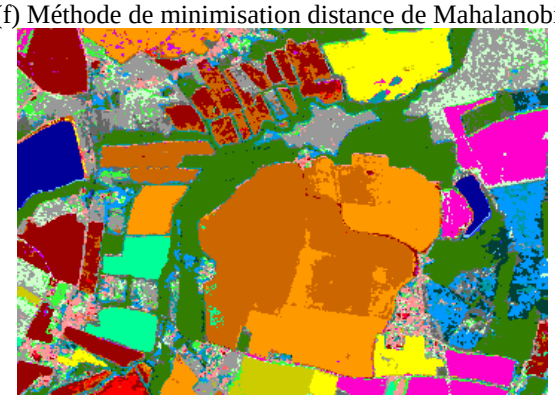
(e) Méthode de minimisation de distance euclidienne



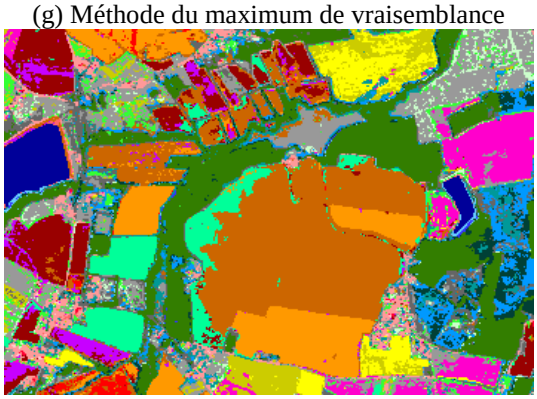
(f) Méthode de minimisation distance de Mahalanobis



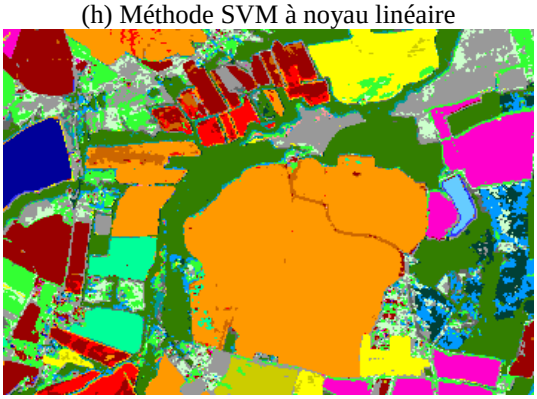
(g) Méthode du maximum de vraisemblance



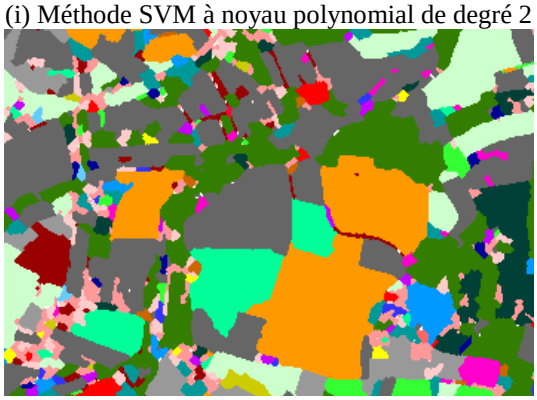
(h) Méthode SVM à noyau linéaire



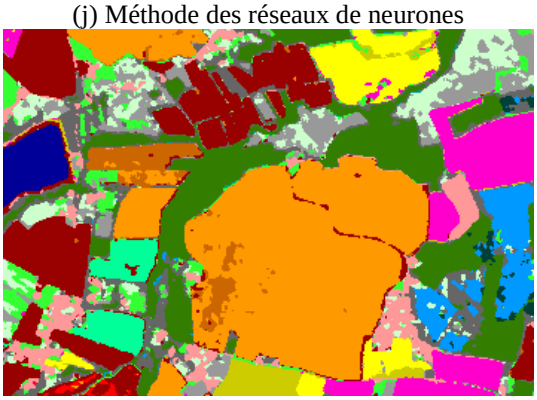
(i) Méthode SVM à noyau polynomial de degré 2



(j) Méthode des réseaux de neurones



(k) Classification objet métrique de Bhattacharyya



(l) Méthode ICM

figure I.48 : Extraits des images résultats des méthodes de classifications (e) à (l)

5. Conclusion

La citation de Brunot résume le problème fondamental de la classification : « Si les nuances infinies du langage ne s'accommodent point des classifications rigides qu'on veut faire, tant pis pour les classifications. La science doit s'accommoder à la nature. La nature ne peut s'accommoder à la science. » (Ferdinand Brunot, 1922). Quelles que soient les méthodes présentées dans ce chapitre, qu'elles soient paramétriques ou non, avec une approche ponctuelle, objet ou globale ou encore supervisée ou non supervisée interprétée, toutes ces méthodes dépendent du point de vue subjectif de la définition des classes thématiques composant la scène observée.

Chacune de ces méthodes possède ses caractéristiques propres. Les méthodes paramétriques sont très efficaces mais nécessitent un nombre minimum de points d'échantillonnage contrairement aux méthodes non paramétriques à noyau qui peuvent se contenter de peu de points, mais qui, en contrepartie, demandent beaucoup de temps de calcul lorsque le nombre de classes et d'images est important.

Pour ce qui est de la considération de l'objet à classer, la méthode globale ICM apporte l'information du contexte au principe de la classification, ce qui permet une amélioration de la qualité (visuelle) et des performances (voir chapitre 2). En plus, nous avons ajouté la considération d'informations exogènes dans le processus de classification, ce qui a permis la gestion de données telles que l'information de l'altitude, la restriction de classes ou encore le masquage des données altérés ou manquantes (défauts, nuages,…).

Dans ce chapitre, nous n'avons pas parlé de l'évaluation de la classification, ceci fera l'objet du chapitre suivant. Afin d'améliorer le système de classification, les chapitre 3 et 4 seront consacrés à la sélection de données en entrée du processus de classification afin de réduire la redondance et la complexité des calculs ainsi qu'à l'utilisation de la complémentarité des classifications et de leur fusion.

Chapitre II. L'évaluation de la classification

Résumé : La validation du résultat de classification est une étape importante du processus car elle permet de s'assurer de la performance des résultats obtenus et ainsi de la significativité des analyses leur succédant. L'intérêt de ce chapitre sera de déterminer un indice unique permettant de mesurer simultanément plusieurs caractéristiques de la classification. Ce critère de performance sera un outil important de recherche et d'optimisation de la performance des classifications. Ce nouvel indice sera comparé aux indices préexistants sur des exemples simples ainsi que sur des exemples concrets. J'introduirai également une nouvelle matrice de confusion qui permettra de tenir compte de l'incertitude de la prise de décision pour le calcul de la performance d'une classification supervisée.

Mots clés : matrice de confusion, performance, précision, rappel, incertitude, similitude.

«Le vrai génie réside dans l'aptitude à évaluer l'incertain, le hasardeux, les informations conflictuelles.»

Winston Churchill

Sommaire du chapitre II

1. L'évaluation de classification : définitions et contexte

1.1. Contexte

La validation du résultat de classification est une étape importante du processus car elle permet de s'assurer de la performance et de la significativité des résultats obtenus.

Dans le cadre d'un processus supervisé, deux jeux d'échantillons sont acquis, la partie apprentissage est utilisée pour initialiser le processus de classification tandis que la partie vérification est utilisée pour valider le résultat de la classification à travers le calcul d'une matrice de confusion.

Dans ce chapitre, je présenterai deux extensions de la matrice de confusion, la première avec l'intégration de la notion de probabilité, la seconde avec l'intégration de similitude de classe. Je présenterai ensuite deux nouveaux indices extraits des matrices de confusion qui permettront de synthétiser l'information à travers un indice unique.

L'intérêt d'obtenir un indice unique permettant de mesurer simultanément plusieurs caractéristiques de la classification sera un outil important pour les chapitres suivants qui l'utiliseront comme critère de recherche et d'optimisation.

1.2. Définitions

Commençons par préciser quelques termes qui seront largement utilisés dans la suite du manuscrit. En anglais, deux termes sont employés pour décrire la performance d'une classification : « precision » et « recall ». Ces deux termes sont respectivement traduits en français par « précision » et « rappel » (Davis & Goadrich, 2006; Grossman & Frieder, 2004).

Pour le terme précision, il existe en anglais deux mots : « precision » et « accuracy ». L'« accuracy » correspond à la justesse, l'exactitude ou encore la décision correcte. La « precision » correspond au niveau de détails. La définition d'une classe peut devenir de plus en plus précise. On peut mieux définir la classe « cultures » en « cultures d'été » et « cultures d'hiver », puis les cultures d'hiver en classes de « blé », « orge », …. Enfin la classe de « blé » peut également être décrite en « blé tendre » et « blé dur ». La description des classes est donc devenue plus précise et la classification sera plus détaillée.

Dans le reste du document, nous emploierons les termes de performance, rappel et précision définis de la manière suivante :

Définition : On appelle *performance* d'une classification un résultat chiffré permettant de mesurer l'exactitude du résultat d'une classification.

Définition : Le *rappel d'une classe* c est défini comme étant la proportion d'éléments bien classés par rapport au nombre d'éléments appartenant à la classe c dans les données de référence (échantillons de validation). L'indice évaluant le rappel est appelé Producer Accuracy.

Définition : La *précision d'une classe* c est défini comme étant la proportion d'éléments bien classé par rapport au nombre d'éléments attribués à la classe c par le processus de décision de la classification. La précision est relative à l'erreur d'excédent d'une classe. L'indice évaluant la précision est appelé User Accuracy.

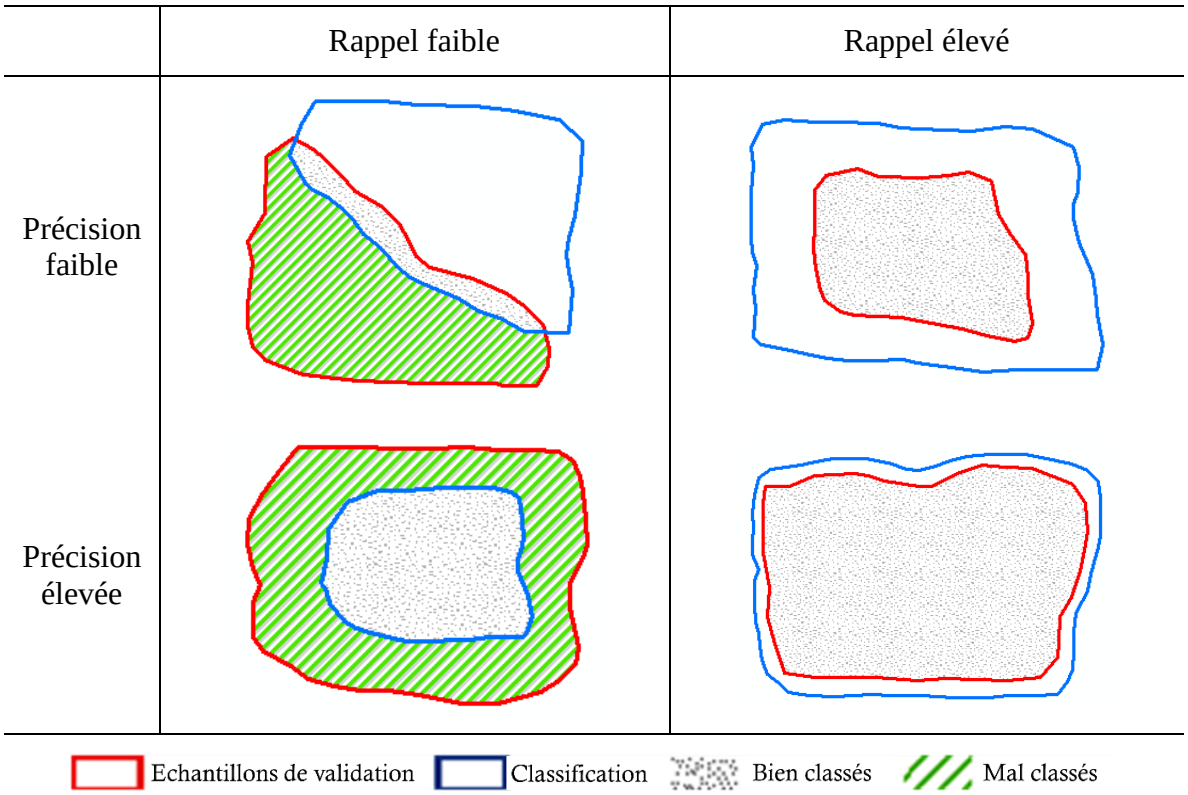


figure II.1 : Illustration des notions de rappel (proportion de l'aire des bien-classés par rapport à l'aire des échantillons de validation) et de précision (proportion de l'aire des bien-classés par rapport à l'aire de la classification)

La différence entre précision et rappel est illustrée en figure II.1

2. L'évaluation à partir de références spatiales

Classiquement, une classification s'évalue à l'aide de références spatiales c'est-à-dire qu'il y aura comparaison entre les pixels de l'image issue de la classification et ceux de l'image servant de référence pour la validation.

2.1. La matrice de confusion

Dans le cadre d'une classification supervisée disposant de références spatiales, la matrice de confusion est un outil permettant de mesurer la qualité du système de classification utilisé.

Définition : La *matrice de confusion* est un outil permettant de mesurer la concordance entre un ensemble d'éléments observés et un ensemble d'éléments de référence. Ici les éléments observés correspondent aux pixels issus de la classification et les éléments de référence à nos échantillons de vérification.

Notations : Soit la matrice de confusion **m** qui pour chaque couple de classes ($\mathbf{c_i}$, $\mathbf{c_j}$) associe le nombre de pixels **s** correspondant à la classe $\mathbf{c_i}$ dans l'image de référence mais classés en tant de $\mathbf{c_j}$ par l'algorithme de classification. Chaque élément de cette matrice de confusion **m** sera noté $\mathbf{m}_{c_i,c_j}$ avec $c_i \in C,\ i \in [1, \#C]$ et #C le nombre de classes présentes.

$$m_{c_i,c_j} = \sum_{s \in S_{verif}} \mathbb{1}_{\{c_s^{opt}=c_j\} \cap \{c_s^{verif}=c_i\}} \quad \text{(II.1)}$$

Rappelons que c_s^{opt} et c_s^{verif} correspondent à la classe attribuée au pixel s respectivement par la classification et par l'image des échantillons de vérification. La figure II.2 illustre la représentation de la matrice de confusion.

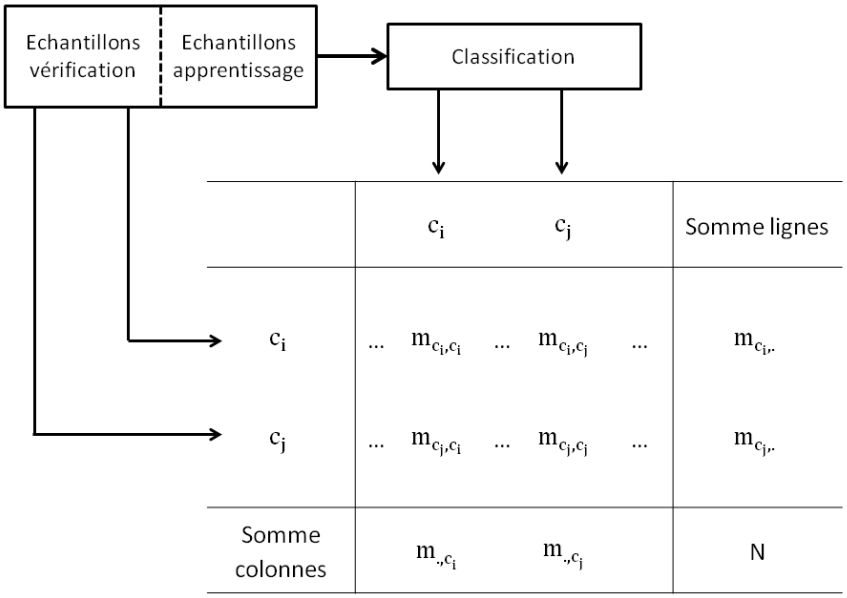


figure II.2 : Représentation de la matrice de confusion

Notons $\mathbf{m_{c_i,.}}$ la somme des pixels de référence appartenant à la classe c_i et $\mathbf{m_{.,c_j}}$ la somme des pixels de référence classés en tant que c_j par l'algorithme de classification. Le nombre total de pixels de vérification est noté **N** avec :

$$N = \sum_{i \in [1,\#C]} m_{c_i,.} = \sum_{j \in [1,\#C]} m_{.,c_j} \quad \text{(II.2)}$$

Nous noterons que la matrice de confusion ne tient compte que de deux états : 1 si le pixel appartient aux 2 catégories et 0 sinon. Aucune information sur la certitude de la prise de décision, lors de la classification n'intervient dans le calcul de la matrice de confusion.

Pour cela, j'introduis maintenant une nouvelle matrice de confusion avec intégration de la notion de probabilité d'appartenance afin de prendre en compte l'incertitude de la prise de décisions.

2.2. La matrice de confusion par pondération de probabilité

Rappelons que chaque élément de la matrice de confusion **m** est noté $\mathbf{m_{c_i,c_j}}$ avec $c_i \in C,\ i \in [1, \#C]$ et #C le nombre de classes présentes

Contrairement à la matrice de confusion classiquement utilisée, nous définissons maintenant chaque élément m_{c_i,c_j} comme étant la somme des pixels pondérés par leur probabilité d'appartenance $p(c_s^{opt} = c_j | c_s^{verif} = c_i)$.

La formulation de m est présentée en équation II.3.

$$\begin{aligned} m_{c_i,c_j} &= \sum_{s \in S_{verif}} p(c_s^{opt} = c_j | c_s^{verif} = c_i) \\ &= \sum_{s \in S_{verif} \text{ tq } c_s^{verif} = c_i} p(c_s^{opt} = c_j) \end{aligned} \quad \text{(II.3)}$$

Un exemple de cette matrice est présenté en annexe 1.3.

2.3. Indices pour l'évaluation à partir des matrices de confusion

A partir de la matrice de confusion présentée aux paragraphes 2.1 et 2.2, plusieurs indices peuvent être extraits afin d'évaluer la performance de la classification.

Nous séparerons les indices en 2 catégories distinctes :

- les indices dits « locaux », « par classe » ou encore « Un contre Tous », qui permettent de juger de la performance d'une classe par rapport aux autres.
- les indices dits « globaux » ou « Tous contre Tous », qui permettent de juger de la performance globale de la classification.

Afin d'illustrer les différents indices qui seront présentés ci-après, nous prendrons un exemple simple à 2 classes défini par le tableau II.1.

	Classe 1	Classe 2	Somme pixels classes références
Classe 1	A	B	A+B
Classe 2	C	D	C+D
Somme pixels classes attribuées	A+C	B+D	N = A+B+C+D

tableau II.1 : Exemple pris d'une matrice de confusion à 2 classes

Nous fixerons les valeurs de la diagonale (A et D) et ferons varier les valeurs hors diagonales (B et C) afin d'observer l'implication de ces changements sur les indices présentés ci-après.

2.3.1. Les indices « Un contre Tous »

Les deux indices couramment utilisés pour évaluer la performance d'une classe sont le Producer Accuracy (PA) et le User Accuracy (UA) :

- Le « Producer Accuracy » (PA) est l'indice qui évalue le rappel d'une classe.

$$PA_i = \sum_{j=1}^{n} \frac{m_{c_i,c_i}}{m_{c_i,c_j}} = \frac{m_{c_i,c_i}}{m_{c_i,.}} \in [0,1] \quad \text{(II.4)}$$

- Le « User Accuracy » (UA) est l'indice qui évalue la précision d'une classe.

$$UA_i = \sum_{j=1}^{n} \frac{m_{c_i,c_i}}{m_{c_j,c_i}} = \frac{m_{c_i,c_i}}{m_{.,c_i}} \in [0,1] \quad \text{(II.5)}$$

De ces deux indices de performance découlent deux indices d'erreurs, respectivement erreur d'omission et erreur de commission :

- L'erreur d'omission représente le déficit d'une classe ou encore la perte de pixels d'une classe lors de la classification.

$$Erreur\_omission_i = Err\_om_i = 1 - PA_i = 1 - \frac{m_{c_i,c_i}}{m_{c_i,.}} \in [0,1] \quad \text{(II.6)}$$

- L'erreur de commission représente l'excès d'une classe ou encore la surreprésentation des pixels d'une classe lors de la classification.

$$Erreur\_commission_i = Err\_com_i = 1 - UA_i = 1 - \frac{m_{c_i,c_i}}{m_{.,c_i}} \in [0,1] \quad \text{(II.7)}$$

Nous définirons deux types de comportement de classes à partir de ces deux erreurs :

- Les classes dispersibles, qui ont un taux important d'erreurs d'omission et donc un PA faible.
- Les classes absorbantes, qui ont un taux important d'erreurs de commission et donc un UA faible.

2.3.2. Les indices globaux (« Tous contre Tous »)

Dans la littérature, la diversité d'indices permettant d'évaluer une classification est grande. Chaque indice possède une signification et un objectif d'évaluation précis. Nous décrirons trois catégories d'indices constituées comme suit.

La première catégorie est constituée des indices orientés « rappel » *id est* qui mesurent la performance en fonction de la référence. Dans cette catégorie, nous citerons le « Average Accuracy » (AA) qui est une moyenne des « Producer Accuracy » de chaque classe.

$$AA = \frac{1}{n}\sum_{i=1}^{n} \frac{m_{c_i,c_i}}{m_{c_i,.}} = \frac{1}{n}\sum_{i=1}^{n} PA_i \in [0,1] \quad \text{(II.8)}$$

Nous voyons ici un exemple de confusion de terminologie. Le AA est appelé « Average Accuracy » alors qu'il ne tient compte que du rappel des classes. Cet indice devrait donc se nommer « Average Recall ». Dans un souci de cohésion, nous conserverons la terminologie AA largement utilisée dans notre domaine d'application.

La seconde catégorie est constituée des indices orientés « précision » *id est* qui mesurent la performance en fonction de la classification. Nous introduirons ici le « Average Precision » qui est à la précision ce que l'Average Accuracy est au rappel.

$$AP = \frac{1}{n}\sum_{i=1}^{n}\frac{m_{c_i,c_i}}{m_{.,c_i}} = \frac{1}{n}\sum_{i=1}^{n} UA_i \in [0,1] \quad \text{(II.9)}$$

La dernière catégorie regroupe les indices dits globaux *id est* qui mesurent la performance « sans orientation ». Cette catégorie est sans conteste la plus utilisée car elle est par définition globale, c'est-à-dire non orientée. Nous citerons tout d'abord le « Overall Accuracy » définit par l'équation II.10, appelé également Pourcentage de pixels Correctement Classés (PCC).

$$OA = \frac{1}{N}\sum_{i=1}^{n} m_{c_i,c_i} \in [0,1] \quad \text{(II.10)}$$

Cet indice représente la proportion de pixels présents sur la diagonale de la matrice de confusion. L'OA est un indice de performance global au sens où il permet de mesurer la performance quel que soit la taille des classes contrairement aux indices AA et AP qui pondèrent cette performance globale en attribuant un poids à chaque classe (le poids étant défini par l'effectif de la classe).

Nous citerons également la F-mesure (ou F-score ou encore « F-measure » en anglais, (Bahadur & Rao, 1960; Rijsbergen, 1979)) qui est un indice de performance basé sur la pondération des notions de précision globale et de rappel global. Le paramètre de pondération est noté β et l'indice est défini comme suit :

$$F_\beta = \frac{(1+\beta^2)\times AA \times AP}{(\beta^2 \times AP) + AA}, \beta \in \mathbb{R}^+ \quad \text{(II.11)}$$

La valeur β=1 est très largement utilisée car elle permet de donner une moyenne harmonique entre le rappel global et la précision globale. Elle est notée $\mathbf{F_1}$.

$$F_1 = \frac{2 \times AA \times AP}{AP + AA}, \beta \in \mathbb{R}^+ \quad \text{(II.12)}$$

Un autre indice utilisé est le Kappa de Cohen qui permet d'estimer le taux de concordance entre deux observateurs (Cohen, 1960, 1968; Fleiss, Cohen, & Everitt, 1969). Néanmoins la première trace de cet indice remonte à 1892 avec (Galton, 1892) mais le Kappa est majoritairement attribué à Cohen. Sans être réellement un indice de performance, le Kappa est utilisé pour juger de la qualité de la classification multi-classes et cela à tort selon Pontius (Pontius, 2000). Néanmoins, cet indice est communément utilisé en télédétection (Congalton, 1991; Montserud & Leamans, 1962; Smits, Dellepiane, & Schowengerdt, 1999; Wilkinson, 2005) mais également dans d'autres domaines dont notamment la psychologie (Fleiss et al., 1969; Landis & Koch, 1977; Tinsley & Weiss, 1975).

La formulation générale du coefficient de Kappa de Cohen se pose sous la forme :

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \in [-1,1] \tag{II.13}$$

Avec P_o : la proportion d'accord observée et P_e : la proportion d'accord aléatoire.

Dans le cadre de la comparaison entre une classification et une référence, le Kappa de Cohen s'écrit alors :

$$\begin{aligned} \text{kappa} &= \frac{N\sum_{i=1}^{n} m_{c_i,c_i} - \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{\cdot,c_i}}{N^2 - \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{c_i,i}} \\ &= \frac{OA - {}^{1}/_{N^2} \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{\cdot,c_i}}{1 - {}^{1}/_{N^2} \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{\cdot,c_i}} \in [-1,1] \end{aligned} \tag{II.14}$$

D'après Cicchetti (Cicchetti & Fleiss, 1977), le test du Kappa est significatif lorsque $N > 2 \times (\#C)^2$, $\#C$ le nombre de classes.

Les limites du Kappa sont grandes. Feinstein et Cicchetti (Feinstein & Cicchetti, 1990) décrivent le paradoxe du Kappa lorsque les effectifs marginaux sont non équilibrés. Pontius (Pontius Jr & Millones, 2011; Pontius, 2000) décrit le Kappa comme étant un indice monodimensionnel incapable de justifier la concordance (donc la pseudo-performance dans le cas de la classification). Les auteurs justifient cette limite par le fait que l'indice confond la quantité et la position d'une classe.

Le deuxième paradoxe concerne les erreurs systématiques. Ces erreurs entre observateurs entrainent un meilleur résultat du test (à tort) car la concordance aléatoire est plus faible. Par hypothèse, les observateurs doivent être indépendants dans le cadre de l'évaluation d'une classification donc les erreurs systématiques n'ont pas lieu d'exister en théorie.

Plusieurs extensions et améliorations du Kappa sont proposées par Pontius (Pontius Jr & Millones, 2011) afin d'en réduire les paradoxes. Nous en citerons deux, le « Kappa location » et le « Kappa quantity ».

$$\text{K\_location} = \frac{OA - {}^{1}/_{N^2} \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{\cdot,c_i}}{\sum_{i=1}^{n} \min(\frac{1}{n}, \frac{m_{c_i,\cdot}}{N}) - {}^{1}/_{N^2} \sum_{i=1}^{n} m_{c_i,\cdot} \times m_{\cdot,c_i}} \tag{II.15}$$

$$\text{K\_quantity} = \frac{OA - \left(\frac{1}{n} + \text{Klocation}\left(\sum_{i=1}^{n} \min\left(\frac{1}{n}, \frac{m_{c_i,\cdot}}{N}\right) - \frac{1}{n}\right)\right)}{\sum_{i=1}^{n} \frac{m_{c_i,c_i}{}^2}{N^2} + \text{K\_location}(\mathbf{X})} \tag{II.16}$$

$$\text{Avec : } \mathbf{X} = \left((1 - \sum_{i=1}^{n} \frac{m_{c_i,c_i}{}^2}{N^2}) - \left(\frac{1}{n} + \text{K\_location}\left(\sum_{i=1}^{n} \min\left(\frac{1}{n}, \frac{m_{c_i,\cdot}}{N}\right) - \frac{1}{n}\right)\right)\right) \tag{II.17}$$

En opposition au Kappa et à sa notion de concordance, il existe des indices statistiques d'association qui ne tiennent pas compte de la liaison et qui n'exigent pas des variables de même nature. Nous citerons en exemple le coefficient de Kendall (Kendall, 1948) et celui de Cramer (Siegel & Castellan, 1988) sans entrer dans les détails.

2.4. Introduction et validation de nouveaux indices

Dans l'optique d'automatiser les processus liés à la classification, nous devons disposer de critères permettant de juger de la performance de la classification. Afin de limiter et d'unifier les critères présentés dans les paragraphes précédents, nous nous proposons d'introduire 2 nouveaux indices afin de synthétiser l'information (Masse, Ducrot, & Marthon, 2012).

Le premier est un indice de performance par classe qui jugera à la fois de la précision et du rappel de la classe. Nous appellerons cet indice le « Omission and Commission Index » (OCI) car il tient compte à la fois du rappel (l'inverse de l'erreur d'omission) et de la précision (l'inverse de l'erreur de commission) de la classe à évaluer.

$$\begin{aligned} \mathrm{OCI_i} &= \mathrm{PA_i} \times \mathrm{UA_i} \\ &= (1 - \mathrm{Err\_om_i})(1 - \mathrm{Err\_com_i}) \\ &= 1 - \mathrm{Err\_om_i} - \mathrm{Err\_com_i} + \mathrm{Err\_om_i} \times \mathrm{Err\_com_i} \end{aligned} \quad \text{(II.18)}$$

L'originalité de l'indice OCI repose sur sa capacité à observer à la fois les caractéristiques de dispersion et d'absorption des classes. Nous introduisons maintenant une version « Tous contre Tous" de cet indice nommé « Average Omission and Commission Index » (AOCI).

$$\mathrm{AOCI} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{OCI_i} \quad \text{(II.19)}$$

L'AOCI et la F1-mesure sont des indices proches par définition et qui permettent de regrouper les deux caractéristiques importantes que sont la précision et le rappel.

L'OCI et l'AOCI permettent ainsi de synthétiser l'information à travers un critère unique, ce qui permettra de simplifier les algorithmes de recherche et d'optimisation de performance de classification qui seront présentés dans les chapitres suivants.

2.5. Illustrations des indices présentés

Afin d'illustrer les différents indices qui seront présentés ci-après, nous prendrons un exemple simple à 2 classes défini par le tableau II.1. Nous fixerons les valeurs de la diagonale (A et D) et ferons varier les valeurs hors diagonales (B et C) afin d'observer l'implication de ces changements sur les indices présentés ci-après. Posons A = 50 et D = 150 et faisons varier les termes non diagonaux de la matrice de confusion qui sont B et C entre 0 et 300.

Les figures II.3 à II.8 illustrent les simulations des indices par classe et les figures II.9 à II.14 illustrent les simulations des indices globaux.

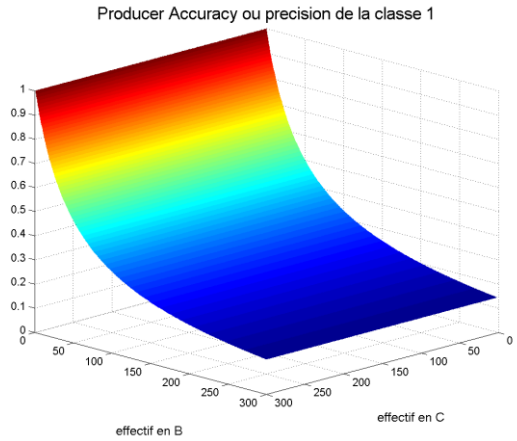


figure II.3 : Simulation de l'indice PA pour la classe 1

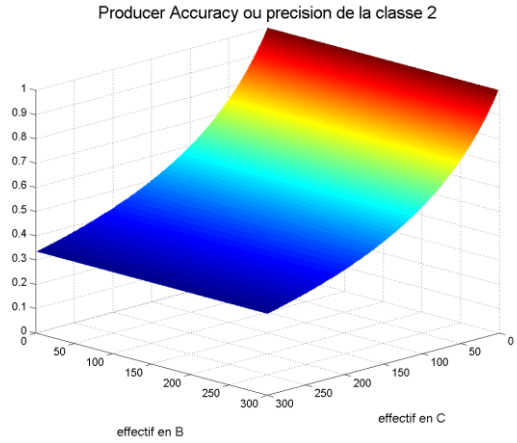


figure II.4 : Simulation de l'indice PA pour la classe 2

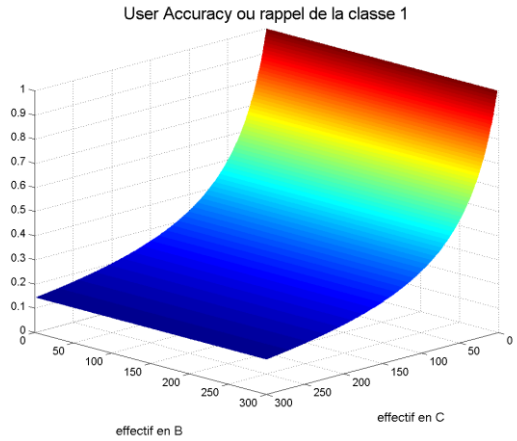


figure II.5 : Simulation de l'indice UA pour la classe 1

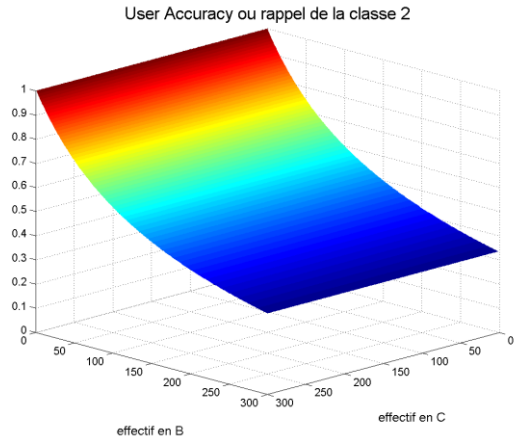


figure II.6 : Simulation de l'indice UA pour la classe 2

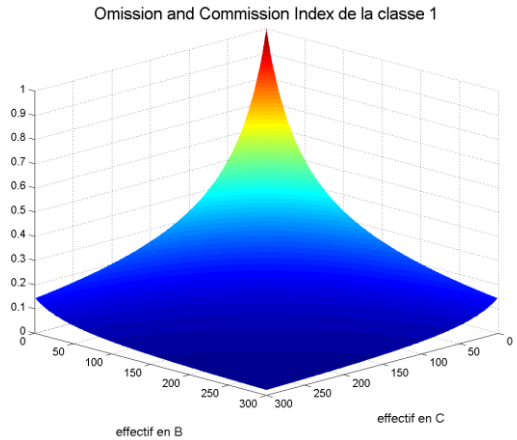


figure II.7 : Simulation de l'indice OCI pour la classe 1

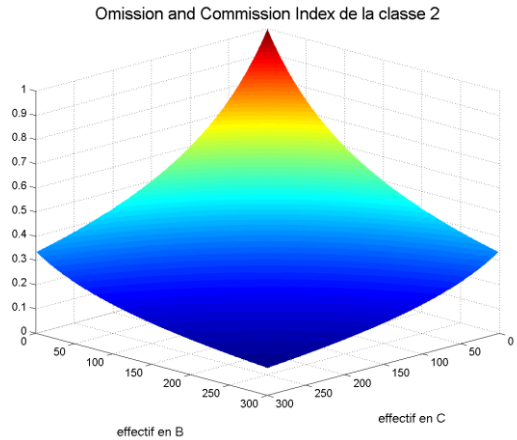


figure II.8 : Simulation de l'indice OCI pour la classe 2

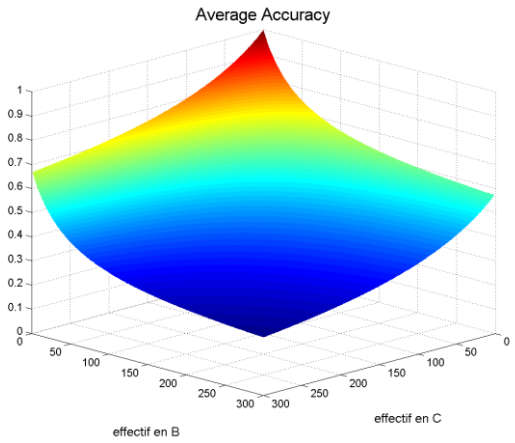


figure II.9 : Simulation de l'indice AA

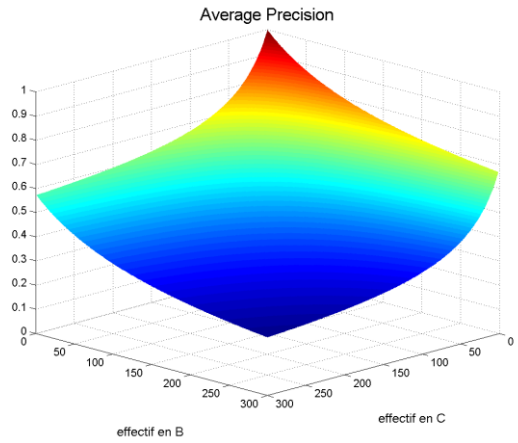


figure II.10 : Simulation de l'indice AP

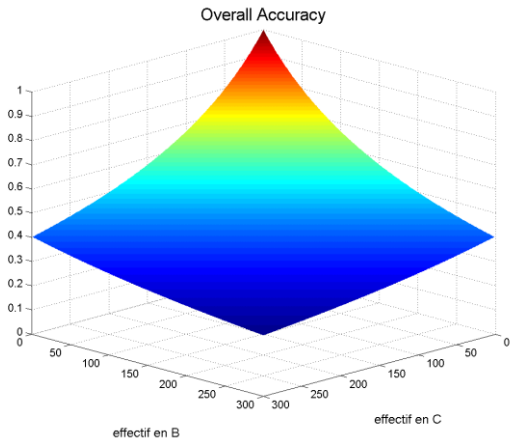


figure II.11 : Simulation de l'indice OA

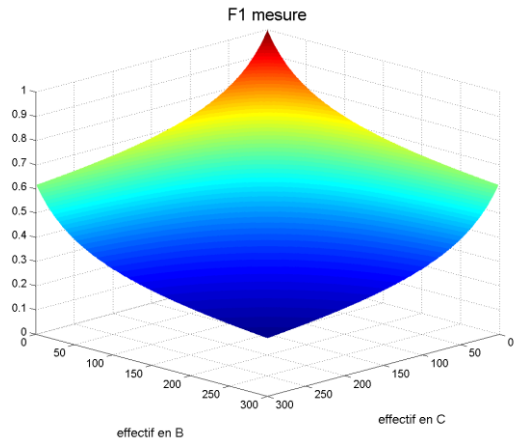


figure II.12 : Simulation de l'indice F1 mesure

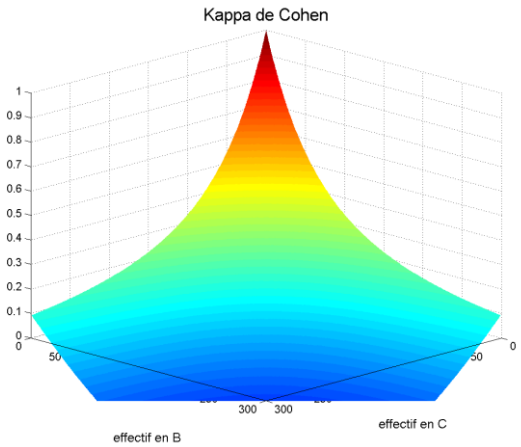


figure II.13 : Simulation de l'indice Kappa

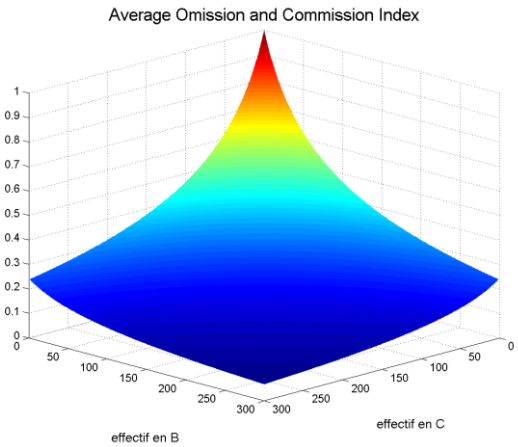


figure II.14 : Simulation de l'indice AOCI

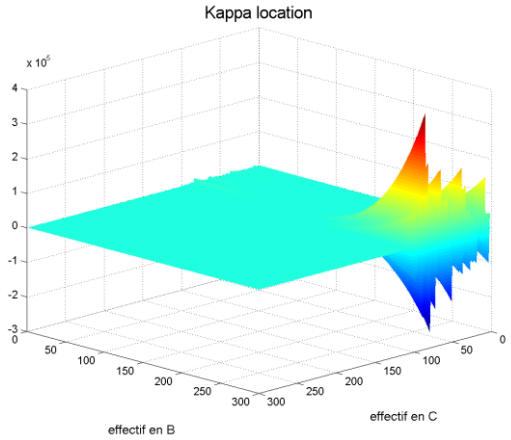


figure II.15 : Simulation de l'indice Kappa location

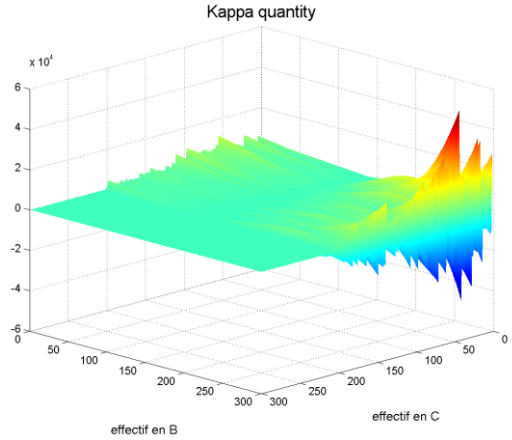


figure II.16 : Simulation de l'indice Kappa quantity

L'effectif en A étant plus faible que l'effectif en D, la performance de la classe 1 doit « chuter » plus rapidement que celle de la classe 2 lors de l'augmentation des effectifs hors-diagonal. Ce phénomène est clairement observé aux figures II.3 à II.6. Les figures II.7 et II.8 montrent la simulation de l'indice OCI, combinaison des UA et PA.

Concernant les indices globaux et du fait de la différence d'effectifs entre A (50) et D (150), nous observons clairement l'orientation de « l'Average Accuracy » vers le « rappel » et l'orientation de « l'Average Precision » vers la « précision » (figures II.9 et II.10).

L'OA, le Kappa, la F1 mesure et le OACI présente une forme symétrique, ils ne sont donc pas orientés (figures II.11 à II.14). Des effets de bords sont observables sur les indices F1 et Kappa. Les simulations des indices OA et AOCI sont similaires, l'AOCI étant cependant plus sévère, indice plus faible, lorsque l'un des deux effectifs hors-diagonal est proche de 0.

Les simulations des Kappa location et Kappa quantity parlent d'elles-mêmes (figure II.15 et figure II.16) et illustrent l'instabilité de ces indices. Par exemple, lorsque l'effectif en C est proche de 0 et l'effectif en B proche de 300, les deux indices sont très instables avec des valeurs comprises entre -2.10^5 et 3.10^5.

3. L'évaluation de classification à partir de valeurs radiométriques spectrales et temporelles de référence

Lorsque qu'il n'est pas possible de comparer une classification à un jeu de données spatiales de validation, la notion de matrice de confusion disparait. Sans possibilité de comparaison spatiale, il faut alors faire une comparaison spectrale et temporelle afin de vérifier que la classe obtenue en sortie de la classification est similaire à la classe d'apprentissage. Cette comparaison est aussi utilisée pour l'interprétation des classifications non-supervisées à partir de séries temporelles et spectrales de valeurs radiométriques.

3.1. Indice de similitude

Dans le Chapitre I, nous avons introduit la notion de distance qu'il convient maintenant d'adapter pour obtenir un indice de similitude entre classe de référence et classe attribuée par la classification. Voyons maintenant comment transformer ces distances en indices de similitude.

$$\mathrm{Ind}_{\mathrm{simi}}(c_i, c_j) = e^{-d(c_i,c_j)} \quad \text{(II.20)}$$

Les distances utilisées sont celles qui permettent de comparer deux ensembles. Parmi celles introduites au chapitre I, seule la distance de Bhattacharyya et la divergence de Kullbach-Leibler peuvent convenir. Nous avons conclu au chapitre I que la distance Bhattacharyya était celle a utiliser en priorité.

3.2. Matrice de similitude

Le principe est identique aux matrices de confusion présentées dans les paragraphes 2.1 et 2.2 avec la nouvelle définition de chaque élément de la matrice tel que :

$$m_{c_i,c_j} = \mathrm{Ind}_{\mathrm{simi}}(c_i, c_j) \quad \text{(II.21)}$$

3.3. Evaluation d'une matrice de similitude

L'évaluation d'une matrice de similitude se fait de la même manière qu'une matrice de confusion classique.

Un exemple de matrice de similitude est présenté en annexe 1.3. Les applications du chapitre V utiliseront les indices extraits de la matrice de similitude pour la validation de la méthode.

4. Applications

4.1. Comparaison des classifieurs sur des données de télédétection

La première application se focalisera sur la comparaison des méthodes de classification présentées dans le chapitre 1 à l'aide d'images Formosat-2 pour l'année 2009 (voir annexe 1.3 pour rappel des caractéristiques, des dates et des classes utilisées). Notons que nous comparons les indices pour des applications utilisant un grand nombre d'images, une définition précise des classes et un nombre important d'échantillons de vérification.

Les méthodes utilisées sont rappelées dans le tableau II.2. Plusieurs logiciels ont été utilisés dont Envi 5.0, Orfeo Tool Box (OTB) et ceux développés et améliorés durant ma thèse (CESBIO).

Parmi les méthodes ponctuelles, le maximum de vraisemblance (MV) est le classifieur le plus performant que ce soit par classe (16 fois meilleur sur 28 classes pour l'OCI) ou d'un point de vu global (4 fois sur 6). Viennent ensuite les classifieurs par Réseaux Neuronaux (NN) et Support Vector Machines (SVM) avec noyau polynomial (tableaux II.3 et II.4). Ces résultats vont dans le sens d'études faites précédemment sur l'intérêt du Maximum de Vraisemblance pour la classification d'un grand nombre de classes et dont le nombre d'échantillons par classe est suffisamment important pour assurer leur significativité (Ducrot, 2005).

La méthode globale ICM est la plus performante grâce à la considération spatiale du processus de classification. Ses indices globaux sont en moyenne 5% plus élevés que ceux de la méthode du Maximum de Vraisemblance (tableau II.3) et meilleurs pour 24 des 28 classes pour l'indice OCI. Les seules classes dont la performance est plus faible pour la méthode ICM sont les classes d'eau (Eau, Lac, Gravières) et la classe de Bâti/Surface minérale (tableau II.4). Cette baisse de performance s'explique par l'utilisation du contexte par la méthode ICM sur de petites classes sensibles aux problèmes de décalage entre images prises à des dates différentes.

Nous noterons également la faible performance de la classification objet. L'explication de cette faible performance est assez simple et provient du choix du seuil d'homogénéité lors de la définition des objets par la méthode des bassins versants. Le seuil choisi peut ne pas convenir à certaines classes.

Au vu des différences observées entre les différents classifieurs, il parait intéressant de combiner les informations de chacun des classifieurs afin d'obtenir un résultat optimal, notamment entre méthodes paramétriques et non paramétriques. Cette fusion d'information sera l'objet du Chapitre IV.

Méthode	Paramétrique	Considération spatiale	Logiciel
Parallélépipède	Non	Ponctuelle	Envi
Minimum distance euclidienne	Non	Ponctuelle	Envi/Cesbio
Minimum distance Mahalanobis	Non	Ponctuelle	Envi/Cesbio
Maximum de Vraisemblance (MV)	Oui	Ponctuelle	Envi/Cesbio
Réseaux neuronaux (NN)	Non	Ponctuelle	Envi
SVM à noyau Linéaire	Non	Ponctuelle	OTB
SVM à noyau Polynomial	Non	Ponctuelle	OTB
Iterated Conditional Mode (ICM)	Oui	Contextuelle	Cesbio
Objet Bhattacharyya	Non	Objet	Cesbio

tableau II.2: Récapitulatif des méthodes présentées au chapitre 1.

Méthode de classification	AA	AP	OA	Kappa	AOCI	F1 mesure
Parallélépipède	20,16	NaN	13,97	0,10	9,31	NaN
Minimum Distance Euclidienne	50,51	33,89	28,75	0,25	24,16	0,41
Minimum Distance Mahalanobis	70,78	57,76	67,82	0,65	44,14	0,64
Maximum de Vraisemblance	**89,48**	**85,42**	**83,87**	0,82	**76,14**	**0,87**
Réseaux neuronaux	65,90	NaN	86,23	**0,85**	56,55	NaN
SVM à noyau Linéaire	86,76	71,42	80,76	0,79	62,87	0,78
SVM à noyau Polynomial	88,83	74,23	82,93	0,81	66,47	0,81
Iterated Conditional Mode	**94,04**	**90,91**	**91,36**	**0,91**	**85,60**	**0,92**
Objet Bhattacharyya	51,82	48,46	38,33	0,35	26,39	0,50

tableau II.3 : Comparaison des critères d'évaluation globaux pour chacune des méthodes introduites pour les données Formosat-2 de 2009. Les valeurs sont indiquées en pourcentage (sauf pour le Kappa et F1).

OCI de la classe (en %)	Parallélépipède	Minimum Distance Euclidienne	Minimum Distance Mahalanobis	Maximum de Vraisemblance	Réseaux neuronaux	SVM à noyau Linéaire	SVM à noyau Polynomial	ICM	Objet Bhattacharyya
bâti dense	0,73	6,78	9,35	**84,55**	0,00	41,15	58,41	**87,05**	0,62
bâti/surf minérale	**95,06**	72,20	78,42	93,92	0,00	82,71	86,63	90,65	1,96
bâti diffus	0,08	0,16	1,27	**22,08**	0,02	10,68	9,03	**43,10**	10,35
blé dur	1,24	21,28	37,10	61,34	**70,15**	50,93	60,09	**84,36**	23,83
blé tendre	1,62	0,00	22,45	50,65	**72,39**	51,35	55,17	**82,83**	13,47
chanvre	0,05	9,35	42,98	**94,25**	86,30	58,21	76,17	**97,93**	23,77
colza	12,70	0,00	83,50	88,91	**91,74**	88,07	84,75	**95,71**	22,13
eau libre	0,00	98,31	95,76	99,30	89,53	99,48	**99,55**	97,73	87,84
eucalyptus	0,00	51,34	39,60	**93,30**	64,92	83,01	82,19	**96,80**	41,60
eucalyptus 2	0,00	8,59	12,38	**90,56**	45,09	69,89	63,78	**96,31**	12,76
feuillus	86,09	78,32	80,68	**95,24**	91,72	92,45	94,39	**95,75**	70,95
friche	0,00	10,30	20,37	27,32	30,95	39,54	**46,79**	**50,33**	1,74
gravières	0,00	97,75	97,83	**99,72**	95,45	99,11	98,33	99,52	78,77
jachère/surface gel	0,46	4,80	30,37	43,17	**48,57**	42,30	48,27	**57,26**	1,10
lac	0,48	96,81	97,20	**99,92**	94,76	99,96	99,37	99,90	47,61
mais	0,08	32,58	82,27	**96,44**	94,22	92,02	89,95	**98,15**	70,29
mais ensilage	16,61	7,45	69,30	**92,75**	77,17	78,50	71,41	**93,66**	23,26
mais non irrigue	0,08	0,00	2,38	**92,33**	0,00	10,32	7,57	**93,04**	1,02
orge	0,00	0,00	33,42	68,43	**71,17**	50,11	48,56	**83,86**	6,12
peupliers	0,00	19,20	45,69	**92,98**	71,29	79,61	82,28	**96,24**	44,95
pois	0,84	0,00	0,77	**61,69**	4,71	25,06	38,22	**94,25**	4,42
prairie temporaire	0,01	0,00	24,28	**53,60**	50,93	38,96	49,28	**61,82**	0,29
prairie permanente	0,00	0,00	20,19	37,85	30,56	29,27	**44,50**	**48,74**	13,67
résineux	0,00	42,42	36,06	**95,78**	68,37	86,42	89,80	**96,99**	56,29
soja	6,66	0,00	50,51	91,91	**72,21**	80,67	85,01	**97,84**	48,85
sorgho	8,59	14,12	34,92	54,20	**71,35**	57,57	59,13	**80,36**	4,97
surface minérale	15,84	5,29	18,59	**60,04**	0,00	37,03	46,02	**83,66**	15,74
tournesol	13,50	0,01	69,09	86,64	**91,23**	81,64	85,06	**93,03**	10,54

tableau II.4 : Comparaison des valeurs de l'indice OCI (en %) en fonction des différentes méthodes présentées et des classes présentes pour les données Formosat-2 année 2009

Des applications similaires de comparaisons de classifieurs ont été faites sur les données Spot-2/4/5 de l'année 2007 (annexe 1.4) ainsi que pour les données Landsat-5 de l'année 2010 (annexe 1.5). Nous ne présenterons ici que les indices globaux, nous reviendrons sur les détails de ces applications au Chapitre IV pour la fusion des résultats de classification.

Le tableau II.5 présente les performances des résultats de classification obtenus à partir des différents classifieurs pour les données Spot-2/4/5 de l'année 2007 et le tableau II.6 celles des résultats de classification obtenus à partir des différents classifieurs pour les données Landsat 5 de l'année 2010.

Les résultats sont similaires à ceux observés au tableau II.3, la méthode du maximum de vraisemblance est la plus performante des méthodes ponctuelles et la méthode ICM offre les meilleurs résultats.

Méthode de classification	AA	AP	OA	Kappa	AOCI	F1 mesure
Parallélépipède	7,35	NaN	4,81	0,03	3,66	NaN
Minimum Distance Euclidienne	63,13	49,29	53,59	0,49	34,36	0,55
Minimum Distance Mahalanobis	71,64	57,22	67,09	0,63	42,32	0,64
Maximum de Vraisemblance	**83,43**	**81,03**	**81,75**	**0,79**	**69,38**	**0,82**
Réseaux neuronaux	47,94	NaN	68,48	0,63	36,15	NaN
SVM à noyau Linéaire	74,16	58,45	64,74	0,61	45,43	0,65
SVM à noyau Polynomial	73,16	58,62	66,39	0,62	45,59	0,65
Iterated Conditional Mode	**84,84**	**81,74**	**82,91**	**0,81**	**70,67**	**0,83**
Objet Bhattacharyya	45,19	49,58	35,51	0,31	22,60	0,47

tableau II.5 : Comparaison des critères d'évaluation globaux pour chacune des méthodes introduites pour les données Spot 2/4/5 année 2007. Les valeurs sont indiquées en pourcentage (sauf pour le Kappa et F1).

Méthode de classification	AA	AP	OA	Kappa	AOCI	F1 mesure
Parallélépipède	13,09	NaN	29,11	0,22	3,43	NaN
Minimum Distance Euclidienne	41,56	23,86	45,27	0,41	10,01	0,30
Minimum Distance Mahalanobis	48,10	27,57	51,23	0,47	13,18	0,35
Maximum de Vraisemblance	**68,81**	**43,61**	66,69	0,63	**30,41**	**0,53**
Réseaux neuronaux	22,06	NaN	**68,84**	**0,65**	12,98	NaN
SVM à noyau Linéaire	56,52	32,99	58,46	0,54	18,88	0,42
SVM à noyau Polynomial	59,53	35,09	59,83	0,56	21,07	0,44
Iterated Conditional Mode	**75,09**	**54,03**	**70,50**	**0,67**	**41,25**	**0,63**
Objet Bhattacharyya	48,22	29,79	39,60	0,35	13,49	0,37

tableau II.6 : Comparaison des critères d'évaluation globaux pour chacune des méthodes introduites pour les données Landsat-5 de 2010. Les valeurs sont indiquées en pourcentage (sauf pour le Kappa et F1).

4.2. Comparaison des matrice de confusion avec et sans pondération de probabilité

Reprenons l'application des données Formosat-2 de l'année 2009 utilisée précédemment et comparons les matrices de confusion avec et sans pondération de probabilité.

Le calcul de cette matrice de confusion avec pondération par probabilité ne peut se faire que dans le cas d'une classification supervisée paramétrique. Nous avons choisi ici de reprendre l'exemple du paragraphe 4.1 pour le Maximum de Vraisemblance. Les matrices de confusion avec et sans pondération par les probabilités d'appartenance sont présentées en annexe 1.3, le résumé de ces matrices par le biais des indices de performance extraits des matrices sont présentés ci-après (figure II.17 et figure II.18).

L'ajout de l'information de probabilité, en plus de l'information d'imprécision contenue dans la matrice de confusion classique, permet de réajuster les résultats. D'un point de vue global, les résultats sont en moyenne 2% supérieurs pour la matrice de confusion avec probabilités (figure II.17) ce qui signifie que les éléments de la diagonale (les biens classés) ont une certitude plus forte que les éléments hors de la diagonales (les mal-classés).

La figure II.18 nous montre la différence de valeur de l'indice OCI entre la matrice de confusion classique et la matrice de confusion avec probabilités. Les conclusions sont identiques à celles du paragraphe précédent. Cependant deux classes sortent du lot : la classe de surface minérale et le bâti dense. Dans le cas de la classe de surface minérale, l'indice OCI est plus élevé avec la pondération de probabilité ce qui suggère une certitude lors du choix du classifieur alors que, pour la classe de bâti dense, l'effet est inverse.

Pour conclure sur cette comparaison entre les deux types de matrices de confusion, il est important de retenir que la matrice avec pondération de probabilité permet de mettre en exergue des classes qui sont par définition sujettes à confusion (signatures spectrales et temporelles proches de celles d'autres classes). Cette nouvelle matrice permet donc d'améliorer la mesure de la confusion pour l'utilisation d'un classifieur paramétrique.

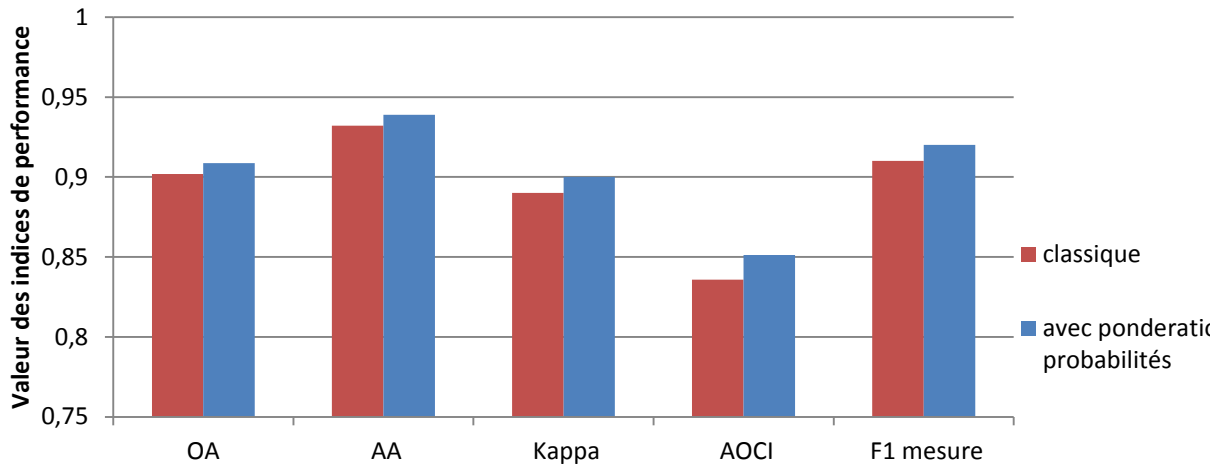


figure II.17 : Comparaison des indices globaux avec et sans la pondération par probabilité

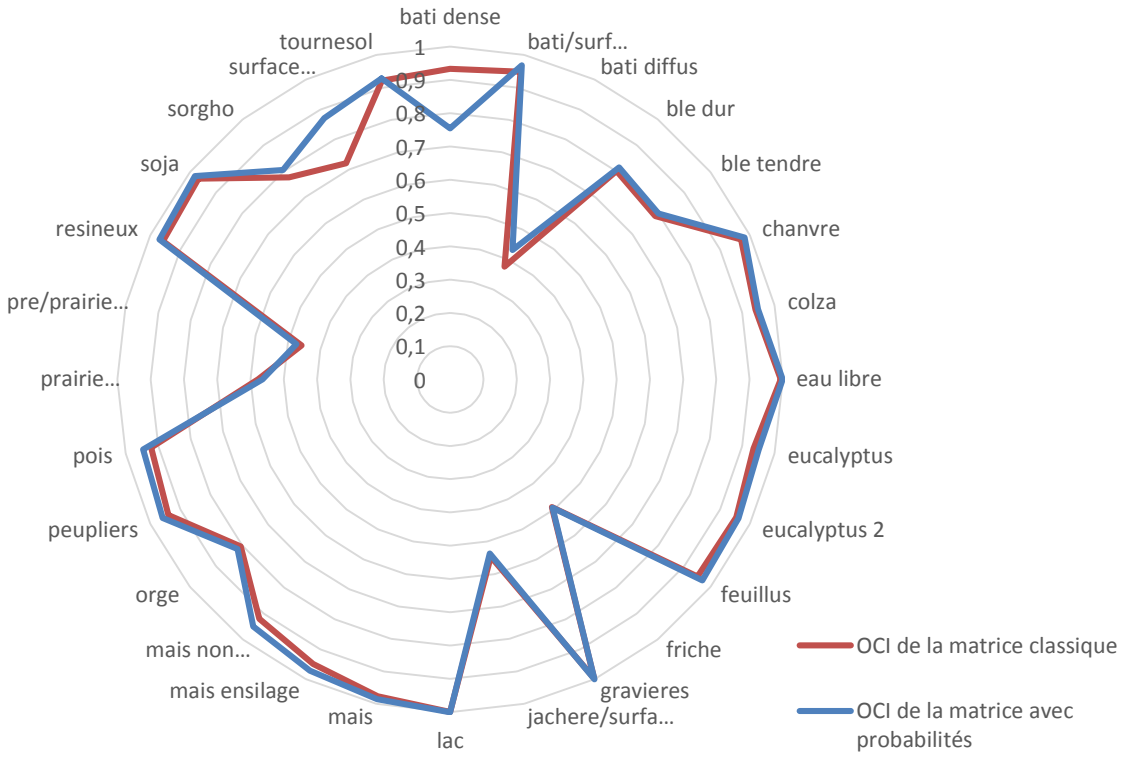


figure II.18 : Comparaison des valeurs de l'indice OCI entre la matrice de confusion classique et la matrice de confusion avec probabilités.

5. Conclusion

L'objectif de ce chapitre était d'introduire des indices de performance permettant de synthétiser l'information d'une matrice de confusion de manière unique afin de limiter les critères d'optimisation qui seront utilisés dans les chapitres 3 et 4. Cet objectif est rempli par l'introduction de deux nouveaux indices : l'OCI et l'AOCI.

Le premier permet l'évaluation de la performance d'une classe par rapport aux autres en tenant compte des notions de précision et de rappel. Le second est une extension du premier et mesure la performance globale de la classification.

Nous disposons désormais d'indices multi-caractéristiques pouvant servir de critères d'optimisation pour des algorithmes de sélection et d'optimisation.

Le second objectif concernait l'intégration de la probabilité d'appartenance au calcul de la confusion entre les classes. Cette pondération a permis de mieux mesurer la confusion avec l'ajout de la similitude de classe, c'est-à-dire la correspondance spectrale et temporelle entre classes de référence et classes issues de la classification.

Chapitre III. La sélection automatique de données de grande dimension pour la classification

Résumé : Devant la grande quantité d'images de télédétection aujourd'hui disponible, il est utile de pouvoir sélectionner les données d'entrée de la classification, quelle que soit leur nature : temporelle, spectrale ou type de fonction discriminante. Ce chapitre visera donc à introduire un système de sélection basé sur le principe des Algorithmes Génétiques et les critères de recherche introduit au chapitre 2, dans le but de maximiser la performance de la classification.

Mots clés : sélection, recherche séquentielle, algorithmes génétiques, grande dimension, système de Sélection-Classification-Fusion.

« C'est la sélection des détails et non pas leur nombre qui donne à un portrait sa ressemblance »

Alexis Carrel, 1999.

Sommaire du chapitre III

1. Introduction de la sélection de données pour la classification

Il a été montré que la classification d'un grand nombre d'images multi-temporelles fournit une performance globale élevée de la classification mais pas nécessairement la meilleure performance pour chaque classe (Ducrot, Masse, Marais-Sicre, & Dejoux, 2010; Ducrot, Sassier, Mombo, Goze, & Planes, 1998).

Plusieurs explications peuvent être données à ce phénomène. Premièrement, la redondance d'informations peut causer des problèmes de convergence lors de la classification (par exemple des dates de prise de vue très proches) (Benediktsson, Sveinsson, & Amason, 1995; Hughes, 1968). Deuxièmement, un nombre important d'images prises à des dates différentes impliquera des erreurs d'ortho-rectification et de correction, ce qui peut provoquer des décalages jusqu'à plusieurs pixels dans certaines conditions et d'où l'introduction d'erreurs lors de la classification. La qualité de la classification de classes à géométrie fine, tel que les routes ou les haies, peut alors être dégradée par l'utilisation de l'ensemble de ces données. Troisièmement, certaines classes ont des signatures spectrales identiques à certaines dates (culture et sol nu en hiver par exemple) ce qui peut nuire à la discrimination des classes par l'algorithme de classification.

L'idée développée dans ce chapitre sera donc de sélectionner automatiquement la meilleure collection d'images fournie en entrée du processus de classification. Cette opération aura pour but de maximiser à la fois la performance globale de la classification ainsi que la performance de chaque classe à l'aide des indices introduits au chapitre 2.

Je présenterai, dans un premier temps, les divers algorithmes de sélection existants et leurs applications à la télédétection. J'introduirai ensuite une nouvelle méthode de sélection basée sur la méthode des Algorithmes Génétiques (paragraphe 2) et appliquée au problème de la sélection de données pour la classification d'images de télédétection (paragraphe 3). Une extension originale du problème de sélection sera ensuite détaillée au paragraphe 4 afin de pouvoir sélectionner plus de caractéristiques (type de classifieur par exemple). Enfin, je présenterai le système automatique de classification dont la première partie utilise le nouvel algorithme de sélection et qui sera nommé processus de Sélection-Classification-Fusion (paragraphe 5).

2. Les algorithmes de sélection

Le problème de la recherche performante pour la sélection de caractéristiques a été intensivement investigué dans la littérature en reconnaissance de forme (Fukunaga, 1990; Jain & Zongker, 1997; Kittler, 1978) et plusieurs stratégies optimales et sous-optimales y ont été introduites. Dash (Dash & Liu, 1997) nous propose une hiérarchisation intéressante des méthodes de sélection. L'ensemble des méthodes de sélection y est séparé en trois parties : les méthodes optimales, sous-optimales et aléatoires (figure III.1).

La première famille de méthode de sélection regroupe les méthodes dites optimales ou complètes. Cette famille est elle-même composée de 2 sous-familles, les méthodes optimales exhaustives et les méthodes optimales non exhaustives.

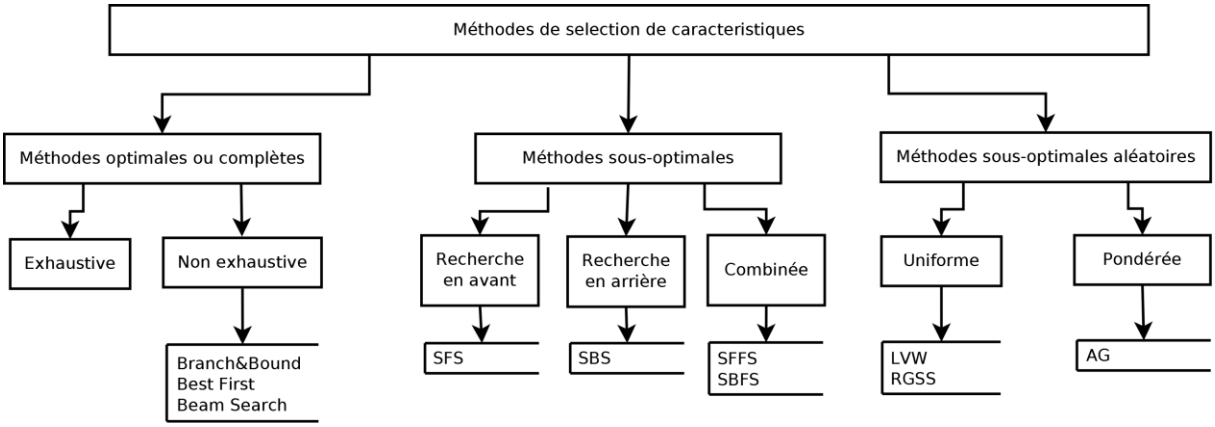


figure III.1 : Hiérarchisation des méthodes de sélection

Les méthodes optimales exhaustives parcourent l'ensemble de l'univers des possibilités D (D étant le nombre de dates disponibles par exemple) qui est exprimé par $2^{\#D}$. On notera que ces méthodes ne sont pas utilisables pour #D grand (Fukunaga, 1990; Jain & Zongker, 1997).

Les méthodes optimales non exhaustives permettent de pallier à ce problème de dimension de recherche. Nous citerons notamment la méthode du « Branch & Bound » (Fukunaga, 1990; Narendra & Fukunaga, 1977) ainsi que la méthode du « Best first » (Pearl, 1984) et sa version améliorée « Beam-Search » (Reddy, 1976).

La seconde famille regroupe les méthodes dites sous-optimales, *id est* qui ne convergent pas forcément vers la solution globale mais vers une solution approchée. Ces méthodes ont l'avantage d'être rapides et simples à manipuler. Stearns (Stearns, 1976) introduit la méthode du «inclusion de l – exclusion de r » ou (l,r)-search.

Les deux cas particuliers issus de cette méthode sont la méthode de recherche en avant et son homologue, la méthode de recherche en arrière, appelées respectivement Sequential Forward Selection (SFS ou (1,0)-search) et Sequential Backward Sélection (SBS ou (0,1)-search).

L'inconvénient majeur de ces méthodes réside dans le choix statique des paramètres d'exploration l et r. Pour pallier à ce problème, les méthodes Sequential Floating Forward Selection (SFFS) et Sequential Floating Backward Selection (SFBS) ont été introduites et permettent l'introduction de paramètres d'inclusion et d'exclusion dynamiques (Ferri, Pudil, Hatef, & Kittler, 1994; Pudil, Novovicová, & Kittler, 1994). Nous verrons dans la suite du chapitre les détails de ces quatre méthodes.

La troisième et dernière famille pourrait être considérée comme faisant partie des méthodes dites sous-optimales mais dont le principe repose sur le parcours aléatoire de l'univers de recherche contrairement aux méthodes Séquentielles présentées précédemment. On la nommera « aléatoire » afin de la distinguer des autres méthodes présentées au paragraphe précédent. Cette famille peut elle aussi être séparée en deux sous-familles que nous nommerons pondérée et uniforme (non-pondérée). Les méthodes aléatoires uniformes comprennent les algorithmes qui génèrent des sous-ensembles de recherche de manière entièrement aléatoire. Nous citerons les algorithmes Las Vegas Wrapper (LVW) et l'algorithme Rapid Gravity Survey Systems (RGSS) qui introduisent l'aléatoire dans les méthodes SFS et SBS. La deuxième sous-famille est également aléatoire, mais désormais chaque sous-ensemble est pondéré suivant son importance,

ce qui nécessite un indice de pondération pour caractériser cette importance. Les Algorithmes Génétiques (GA) (Goldberg, 1989; Holland, 1962) sont la méthode la plus connue de cette sous famille.

Dans le cadre de la sélection de données de télédétection, nous citerons les travaux de Ferri et Serpico (Ferri et al., 1994; Serpico & Bruzzone, 2001) pour l'extraction de caractéristiques et les travaux de Pu (Bajcsy & Groves, 2004; Pu & Gong, 2000; X. Zhang & Niu, 2009) pour la sélection de bandes hyper-spectrales.

Parmi les algorithmes cités précédemment, nous utiliserons ceux de la famille sous-optimale séquentielle (SFS, SBS et leur amélioration SFFS et SFBS) et celui de la famille sous-optimale aléatoire pondérée (AG). Ces algorithmes sont les plus usités dans la littérature citée précédemment.

2.1. Le problème à résoudre

Commençons par poser le problème que nous souhaiterions résoudre : à partir d'un jeu temporel d'images, nous souhaiterions extraire automatiquement le meilleur sous-ensemble de données afin de maximiser une performance choisie de la classification. L'équation III.1 formalise ce problème d'optimisation.

$$D^{opt} = \arg\max_{D_j \subseteq D}(\text{performance de la classification des données } D_j) \qquad \text{(III.1)}$$

Nous cherchons donc le sous-ensemble d'images D^{opt} parmi les dates disponibles D qui maximise la performance du processus de classification.

Notons que cette recherche sera dépendante du classifieur et des classes utilisées. Nous proposerons une extension du problème posé en fin de chapitre afin d'y ajouter des caractéristiques à sélectionner (bandes spectrales et classifieur, paragraphe 4).

2.2. Les méthodes SFS et SBS

Le SBS et SFS sont des méthodes de sélection sous optimale mais très rapide, leur complexité de calcul est estimée à $O((\#D)^2)$ (Jain & Zongker, 1997; Kittler, 1978). Le principe est simple : l'algorithme est initialisé soit à l'aide de l'ensemble entier (SBS) soit de l'ensemble vide (SFS). Puis, tant que le nouveau sous-ensemble choisi est plus performant que le précédent, on enlève un élément de l'ensemble (SBS) ou en ajoute un (SFS) (Roli & Fumera, 2001). Le risque d'obtenir un maximum local non global est très important avec cette méthode car il n'y a pas de procédure de retour en arrière.

Les méthodes SFS et SBS sont en réalité des cas particuliers de l'algorithme (l,r)-search introduit par Stearns (Stearns, 1976). Cet algorithme est aussi appelé « inclusion de l – exclusion de r » et consiste à ajouter (inclusion) et retirer (exclusion) à chaque itération un nombre fixé de dates. Le choix arbitraire de ces deux paramètres l et r sont le plus gros facteur limitant de cette méthode. Nous verrons par la suite des méthodes dérivées de celle-ci avec un choix dynamique pour l et r.

Ecrivons l'algorithme général du (l,r)-search

Entrée : $D = \{d_i \text{ tq } \forall i \in [1, \#D]\}$
Initialisation : Si l>r Alors k :=0 ;$D^{opt} = D^{prec} := \emptyset$
Sinon k :=#D ; $D^{opt} = D^{prec} := D$
Tant que $\left((\text{performance}(D^{opt}) \geq \text{performance}(D^{prec}) et (k > 0) et (k \leq \#D)\right)$
$D^{prec} = D^{opt}$
Etape 1 : Inclusion
Répéter l fois
$d^{+} := \arg \max_{d \in D \setminus D^{opt}} \text{performance}(D \cup \{d\})$
$D^{opt} = D^{opt} \cup \{d^{+}\}$; k=k+1;
Fin Répéter
Etape 2 : Exclusion
Répéter r fois
$d^{-} := \arg \max_{d \in D^{opt}} \text{performance}(D \setminus \{d\})$
$D^{opt} = D^{opt} \cup \{d^{+}\}$; k=k-1;
Fin Répéter
Fin Tant que
Sortie : D^{opt}

2.3. Les méthodes SFFS et SFBS

Pour pallier aux limitations évoquées précédemment pour les méthodes SFS et SBS, de nouvelles méthodes dérivées de celles-ci ont été développées (Pudil et al., 1994) et permettent de choisir dynamiquement le sens et l'intensité de déplacement de l'algorithme dans l'univers de recherche.

Ecrivons l'algorithme du SFFS

Entrée : $D = \{d_i \text{ tq } \forall i \in [1, \#D]\}$
Initialisation : k :=0 ;$D^{opt} = D^{prec} := \emptyset$
Tant que $\left((k > 0) et (k \leq \#D)\right)$
Etape 1 : Inclusion
$D^{prec} = D^{opt}$
$d^{+} := \arg \max_{d \in D \setminus D^{opt}} \text{performance}(D \cup \{d\})$
$D^{opt} = D^{opt} \cup \{d^{+}\}$; k=k+1;
Etape 2 : Exclusion conditionnelle
$d^{-} := \arg \max_{d \in D^{opt}} \text{performance}(D \setminus \{d\})$
Si $(\text{performance}(D^{opt} \setminus \{d^{-}\} m) \geq \text{performance}(D^{prec})$
Alors $D^{opt} = D^{opt} \setminus \{d^{-}\}$; k=k-1;
Retour Etape 2
Sinon **Retour** Etape 1
Fin
Sortie : D^{opt}

L'algorithme du SBFS s'écrit facilement par interversion de l'inclusion et de l'exclusion ainsi que de la condition de départ de l'algorithme précédent.

2.4. La méthode des Algorithmes Génétiques

La méthode des Algorithmes Génétiques (AG) est une méthode sous optimale, aléatoire et pondérée. La partie aléatoire permet l'exploration et empêche l'arrêt sur des maxima locaux tandis que la pondération permet d'accélérer la vitesse de convergence de l'algorithme en favorisant les meilleurs individus.

Le principe des Algorithmes Génétiques est dérivé de la génétique et de l'évolution naturelle (Dawkins & Sigaud, 1989). La méthode remonte à 1962 avec les premiers travaux de Holland sur les systèmes adaptatifs (Holland, 1962). Mais il faut attendre 1989 et le livre de David Goldberg pour obtenir la forme définitive et complète de cette méthode telle que nous l'utiliserons ici (Goldberg, 1989).

La méthode repose sur la définition des cinq éléments suivants (Alliot & Durand, 2005) :

- Le codage de chaque individu. Ce codage permet de modéliser chaque individu comme ensemble de caractéristiques qu'il conviendra de sélectionner. Dans notre cas, un individu sera composé des dates à utiliser (et plus tard de bandes spectrales et d'un classifieur, paragraphe 4.1). Le codage est pour le moment considéré comme binaire.
- Initialisation de la population d'individus. Cette initialisation se fait de manière aléatoire.
- Une fonction permettant de sélectionner les meilleurs individus appelée « fitness measure » qui permet de juger de la performance de chaque individu.
- Des opérateurs génétiques permettant de faire varier les individus et donc d'explorer l'univers de recherche au cours des générations.
- Des paramètres de dimensionnement : la taille de la population, le critère d'arrêt, les probabilités des différents opérateurs génétiques.

Nous voyons immédiatement que cette méthode permet de modéliser intuitivement le problème de recherche de meilleures dates pour la classification. Le point négatif sera de déterminer les bons paramètres afin d'améliorer la vitesse de convergence de la méthode.

Redéfinissons les termes constituant les AG dans le cadre de notre problème de sélection de dates.

Définition : Un *individu* de la population (ou « chromosome » en génétique) sera noté $\mathbf{D_i^{iter}}$ et definira les dates utilisées par l'individu **i** à la génération **iter**.

Notation : Le codage d'un individu se fera de la manière suivante :

$$D_{i,d}^{iter} = \begin{cases} 1 \text{ si la date d } \in D_i^{iter} \\ 0 \text{ sinon} \end{cases} \quad \text{(III.2)}$$

Dans l'exemple suivant, D_i^{iter} est composé des images aux dates 3, 5 et 6. La notation sera donc la suivante :

$$D_i^{iter} = \{IM_3, IM_5, IM_6\} ; \quad \text{(III.3)}$$

$$D_{i,3}^{iter} = 1 ; \qquad D_{i,4}^{iter} = 0 \quad \text{(III.4)}$$

Durant chaque génération, quatre opérations sont appliquées par les AG : la mutation, le croisement, la sélection et enfin la reproduction (figure III.2). Décrivons chacune de ces quatre opérations.

La première opération est la mutation (figure III.3). Cette opération permet d'opérer une permutation intra-individu sur les dates. Elle est définie de la manière suivante pour la mutation de la date d_{mut} pour l'individu D_i^{iter}:

$$\text{mutation}\left(D_i^{iter}, d_{mut}\right) = \begin{cases} D_{i,d}^{iter} = 1 - D_{i,d}^{iter} & \text{si } d = d_{mut} \\ D_{i,d}^{iter} = D_{i,d}^{iter} & \text{sinon} \end{cases} \quad \text{(III.5)}$$

Le choix de d_{mut} se fait par tirage aléatoire. La mutation est l'opération la plus importante des AG car elle permet l'exploration de l'univers de recherche et empêche ainsi la convergence vers un maximum local.

La seconde opération est le croisement qui permet d'effectuer une permutation entre deux individus (figure III.4). Cet échange inter-individu peut s'appliquer sur plusieurs éléments à la fois. Cette opération est définie de la manière suivante pour deux individus i et j et un ensemble D^{crois} de dates à permuter :

$$\text{croisement}\left(D_i^{iter}, D_j^{iter}, D^{crois}\right) = \begin{cases} D_{i,d}^{iter} = D_{j,d}^{iter} \text{ si } d \in D^{crois} \\ D_{i,d}^{iter} = D_{i,d}^{iter} \text{ sinon} \end{cases} \quad \text{(III.6)}$$

Une probabilité d'occurrence est appliquée sur les deux opérations précédentes et notée respectivement $\mathbf{p_{mutation}}$ et $\mathbf{p_{croisement}}$.

Les troisième et quatrième opérations consistent à évaluer les individus et à les reproduire pour la génération suivante, suivant le score de leur évaluation.

Dans le cadre d'une application supervisée, nous utiliserons les indices de performance extraits de la matrice de confusion comme fonction d'évaluation. Cette matrice de confusion est obtenue après classification des données multidimensionnelles de l'ensemble D_i^{iter}. La performance est notée $\mathbf{performance}\left(\mathbf{D_i^{iter}}\right)$.

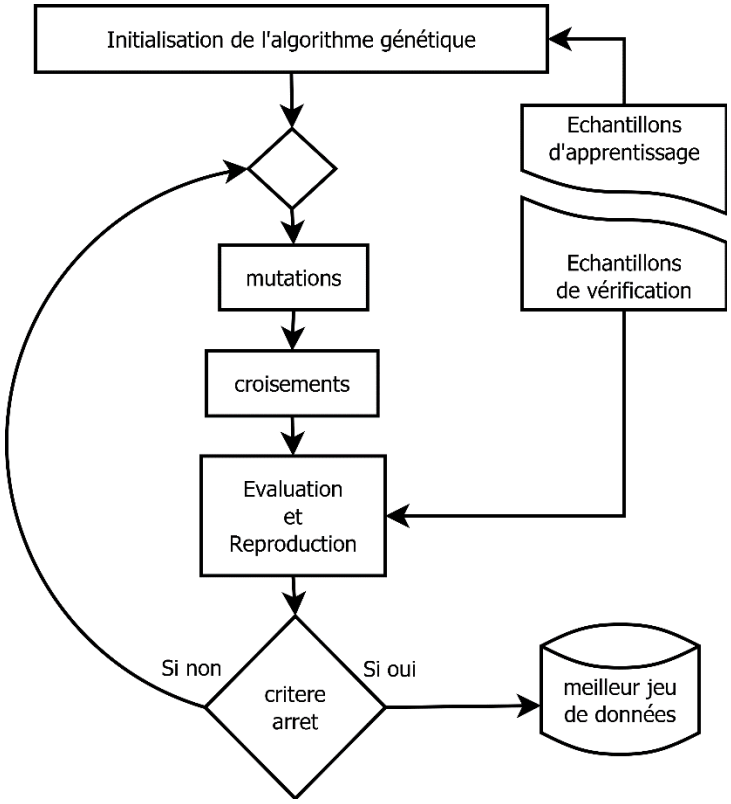


figure III.2 : Schéma d'exécution des Algorithmes Génétiques

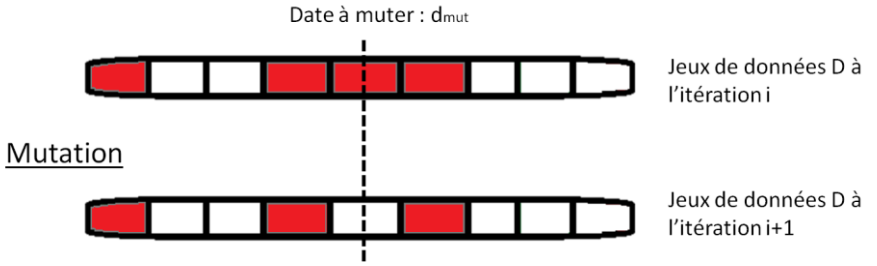


figure III.3 : Illustration du principe de mutation

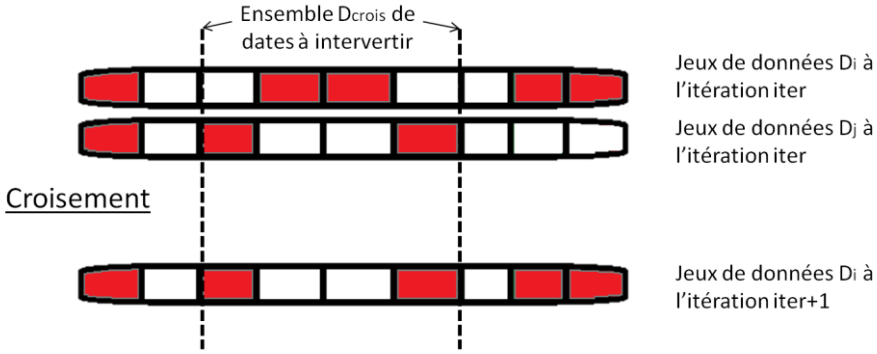


figure III.4 : Illustration du principe de croisement entre deux individus

La reproduction se fait à partir de la performance de chaque individu. Meilleur sera l'individu, plus grande sera sa présence à la génération suivante. Le choix le plus fréquent pour la fonction de reproduction est la suivante :

$$D_j^{iter+1} = D_i^{iter}, \forall\, j \in \left[1, \frac{\text{performance}\left(D_i^{iter}\right) \times \text{nb\_individus}}{\sum_{k=1}^{\text{nb\_individus}} \text{performance}\left(D_k^{iter}\right)}\right] \qquad \text{(III.7)}$$

Ce type de reproduction permet de conserver les meilleurs individus sans leur donner d'omniprésence, ce qui pourrait causer la convergence précoce vers un maximum local.

Ecrivons l'algorithme des Algorithmes Génétiques :

Entrée : $D = \{d_i \text{ tq } \forall i \in [1, \#D]\}$
Initialisation : Tous les individus sont initialisés aléatoirement
Tant que (si critère_arrêt faux)
Pour chaque individu i
Etape 1 : Mutation
$D_i^{iter} = X_{mutation} \times \text{mutation}\left(D_i^{iter}, d_{mut}\right)$
avec $X_{mutation} \sim \text{Bernouilli}(1, p_{mutation})$, d_{mut} tiré aleatoirement
Etape 2 : Croisement
$D_i^{iter} = X_{croisement} \times \text{croisement}\left(D_i^{iter}, D_j^{iter}, D^{crois}\right)$

avec $X_{croisement} \sim \text{Bernouilli}(1, p_{croisement})$, D^{crois} tiré aleatoirement
Etape 3 et 4 : Evaluation et Reproduction
$D_j^{iter+1} = D_i^{iter}\ \forall\, j \in \left[1, \frac{\text{performance}\left(D_i^{iter}\right) \times \text{nb\_individus}}{\sum_{k=1}^{\text{nb\_individus}} \text{performance}\left(D_k^{iter}\right)}\right]$
Fin Pour
$D^{opt} = \underset{k}{\text{argmax}}\ \text{performance}\left(D_k^{iter}\right)$
Si $\left(\overline{\text{performance}}\left(D^{iter}\right) - \overline{\text{performance}}\left(D^{iter-1}\right) < \varepsilon\right)$ **alors** critère_arrêt = 1
Fin Tant que
Sortie: D^{opt}

3. Applications des méthodes de sélection de données

Nous appliquerons les cinq méthodes détaillées précédemment aux applications suivantes :

- sélection temporelle sur les données Formosat-2 année 2009 (16 dates, annexe 1.3).
- sélection temporelle sur les données Spot-2/4/5 année 2007 (6 dates, annexe 1.4).

3.1. Application sélection temporelle sur Formosat-2 année 2009

Rappelons que nous disposons de 16 dates acquises par le satellite Formosat-2 pour l'année 2009 sur la zone Sud-Ouest de Toulouse et que ces 16 dates ont été géo-référencées, ortho-rectifiées et corrigées d'effets atmosphériques, ce qui rend la comparaison temporelle possible. La liste des classes est disponible en annexe 1.3.

Parmi ces 16 dates, nous disposons de 3 dates de Mars, 3 dates de Juillet et 4 dates d'Août, ce qui peut impliquer de la redondance d'information pour certaines classes (voir histogramme des dates utilisées en annexe 1.3).

Appliquons les méthodes de sélection présentées aux paragraphes précédents pour le problème de la sélection du meilleur jeu temporel d'images. Le tableau III.1 regroupe les paramètres utilisés par la méthode de sélection par Algorithmes Génétiques. Pour les cinq méthodes, la méthode de classification utilisée lors de l'étape d'évaluation est l'algorithme ICM.

La comparaison des résultats de sélection des meilleures dates est présentée au tableau III.2. D'un point de vue global, les quatre méthodes séquentielles (SFS, SBS, SFFS et SBFS) obtiennent des résultats similaires. Les Algorithmes Génétiques fournissent des résultats identiques voire supérieurs.

Le tableau III.3 fournit des détails supplémentaires sur la nature des jeux de données sélectionnés par la méthode AG. D'un point de vue général, les jeux de données possèdent un nombre élevé de dates (entre 8 et 16 sur les 16 dates disponibles) mais ne possèdent que très rarement l'ensemble des 16 dates.

En matière de convergence et de temps d'exécution (figure III.6), l'avantage revient aux méthodes dites Backward car les meilleurs jeux de données sont souvent ceux qui disposent du plus grand nombre de dates. Les AG convergent eux aussi rapidement vers leur optimum (entre 2 et 6 générations suivant les cas de figures).

Les figures III.6 à III.8 illustrent parfaitement le genre de problème que la sélection de données permet de corriger. Nous présentons ici l'exemple de la classe « Tournesol », qui est très largement surclassée dans l'extrait de la classification obtenue à partir du meilleur jeu de données en termes de performance globale (figure III.7). Lorsque l'on sélectionne ce jeu de données en fonction de l'indice de performance de la classe « Tournesol », la surreprésentation de cette classe disparait (figure III.8). Cette amélioration s'explique par l'absence d'une date de mars dans le nouveau jeu de données (tableau III.3). Rappelons qu'au mois de mars, la culture du Tournesol n'a pas encore commencé, et que le champ est en sol nu (figure III.6).

En conclusion, les AG semblent être un bon compromis entre rapidité et convergence optimale même si l'avantage de la rapidité revient à la méthode SBFS.

Paramètres AG	Valeur
P_mutation	1/100
P_croisement	6/10
Taille de la population	50
Nombre de générations maximum	10
Critère évaluation	AOCI ou OCI

tableau III.1 : Paramètres utilisés pour la méthode des Algorithmes Génétiques

Méthodes	Valeur optimale AOCI	Valeur optimale OCI classe Feuillus	Valeur optimale OCI classe Blé dur	Valeur optimale OCI classe Tournesol	Valeur optimale OCI classe Eau libre
SFS	**85,60**	95,23	**84,36**	**93,03**	98,28
SBS	**85,60**	95,24	**84,36**	92,11	99,37
SFFS	**85,60**	95,24	84,01	**93,03**	97,11
SBFS	**85,60**	95,24	**84,36**	92,03	97,11
AG	**85,60**	**95,75**	**84,36**	**93,03**	**99,73**

tableau III.2 : Comparaison des méthodes de sélection par évaluation de leur performance respective (en %)

	Global	Classe Feuillus	Classe Blé dur	Classe Tournesol	Classe Eau libre
Performance	85,60	95,75	84,36	93,03	99,73
Jeu de données	16 dates	13 dates	15 dates	15 dates	8 dates
Février	15	15	15	15	
Mars	17; 21; 30	21; 30	17; 21	17; 21	
Mai	3	3	3	3	3
Juin	23	23	23	23	
Juillet	1; 12; 26	1; 12; 26	1; 12; 26	1; 12; 26	12; 26
Août	5; 14; 22; 30	22; 30	5; 14; 22; 30	5; 14; 22; 30	5; 22; 30
Septembre	6; 24	6; 24	6; 24	6; 24	6; 24
Octobre	16	16	16	16	

tableau III.3 : Extrait de performance et détails des dates des meilleurs jeux de données sélectionnés par l'AG

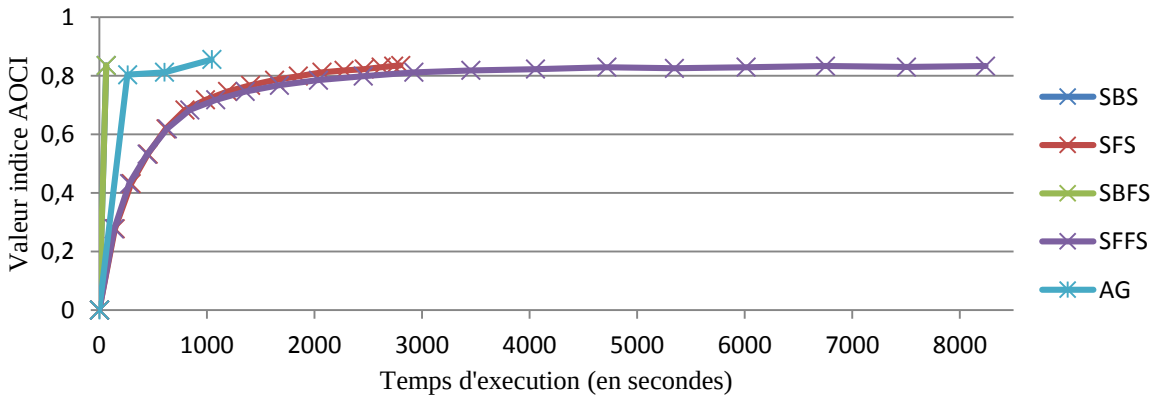


figure III.5 : Temps d'exécution des méthodes de sélection pour la recherche du jeu de données Formosat-2 qui maximise la performance globale (AOCI).

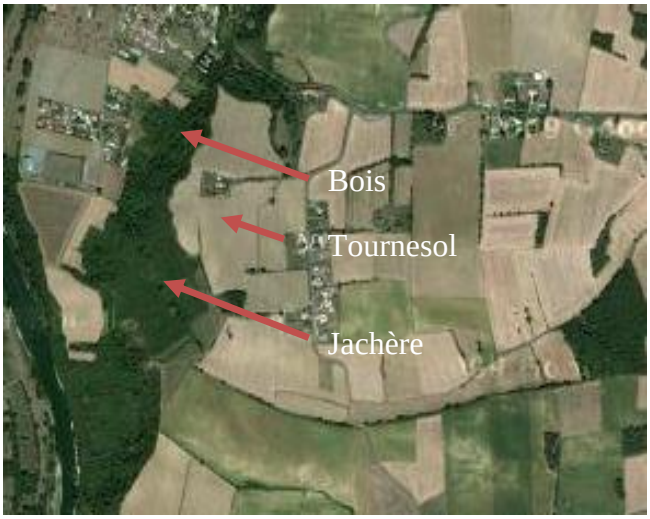


figure III.6 : Extrait Google Map ® représentant la réalité de la scène

figure III.7 : Extrait de la classification du jeu de données sélectionné pour la meilleure performance globale

figure III.8 : Extrait de la classification du jeu de données sélectionné pour la meilleure performance de la classe Tournesol

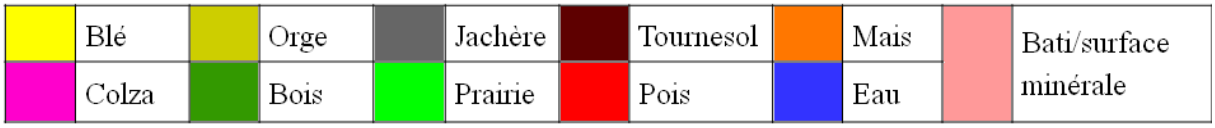


figure III.9 : Légende

3.2. Application sélection temporelle sur Spot 2/4/5 année 2007

Rappelons que nous disposons de 6 dates acquises par les satellites Spot pour l'année 2007 sur la zone Sud-Ouest de Toulouse. Elles ont été géo-référencées, ortho-rectifiées et corrigées d'effets atmosphériques, ce qui rend la comparaison temporelle possible. La liste des classes est disponible en annexe 1.4.

Notons ici que les différences de vitesse de convergence des différentes méthodes sont comparables à celle obtenue avec l'application Formosat-2 (figure III.10).

Le tableau III.4 fournit les détails sur la sélection des meilleurs jeux de données par la méthode des Algorithmes Génétiques. Ces résultats de sélection sont logiques et similaires à ceux qu'un expert aurait pu effectuer manuellement. Par exemple, pour les classes de cultures, la sélection a été faite en choisissant les dates suivant leur stade phénologique, donc discriminantes.

	Global	Classe Feuillus	Classe Blé dur	Classe Tournesol	Classe Eau libre
Performance	70,67	96,53	58,87	88,61	96,9
Jeu de données	6 dates	5 dates	5 dates	4 dates	5 dates
Jours du mois de :					
Février	15	15	15	15	15
Avril	8	8	8		8
Juin	20	20	20		20
Juillet	25	25	25	25	
Septembre	15	15		15	15
Novembre	16		16	16	16

tableau III.4 : Extrait de performance et détails des meilleurs jeux de données sélectionnés par l'AG

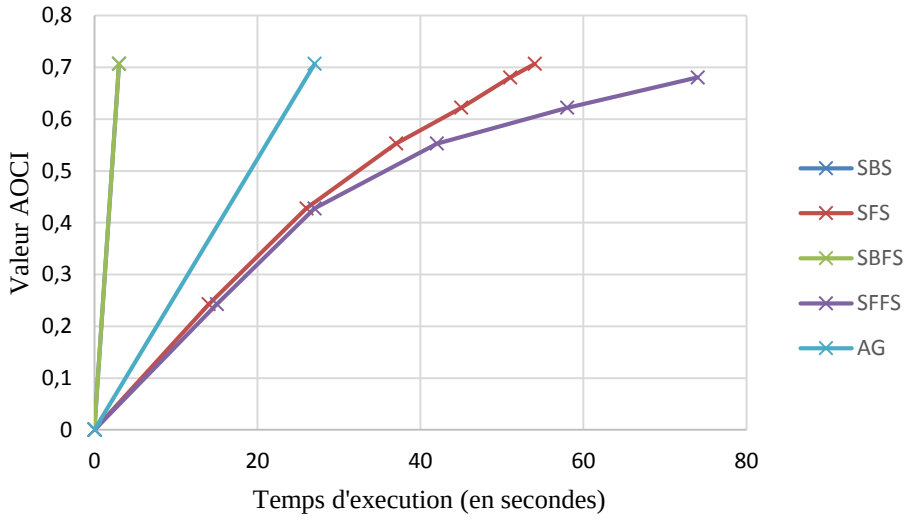


figure III.10 : Temps d'exécution des méthodes de sélection pour la recherche du jeu de données Spot qui maximise la performance globale (AOCI)

4. Extension du problème de sélection

En plus de la sélection temporelle, nous pouvons ajouter une sélection sur les canaux spectraux, sur le classifieur utilisé ou sur toute information *a priori* pouvant être incorporée à une méthode de classification.

Les méthodes présentées de type Séquentiel (SBS, SFS, SBFS et SFFS) ne peuvent fonctionner qu'avec des éléments à sélectionner dont les états sont binaires. Or le choix du classifieur n'est plus binaire s'il y en a plus de deux parmi lesquels choisir. Pour utiliser ces méthodes, il faudrait alors décomposer ce nombre en base 2, ce qui aurait pour conséquence l'augmentation importante des caractéristiques à choisir. La flexibilité des AG permet d'utiliser une formulation plus flexible.

4.1. Adaptation de la méthode des AG

Commençons par réécrire le problème à résoudre :

$$\{\mathrm{IM}^{\mathrm{opt}}, \mathrm{classifieur}^{\mathrm{opt}}\} = \arg\max(\text{performance classification } \mathrm{IM_j} \text{ avec } \mathrm{classifieur_i}) \quad \text{(III.8)}$$

Le nouveau problème revient ainsi à rechercher les meilleures images dans l'espace IM suivant les dimensions temporelles et spectrales, au sens de la performance de la classification de ces données et du meilleur classifieur choisi parmi les méthodes disponibles et applicables.

Voyons maintenant les modifications à apporter à la formulation des Algorithmes Génétiques présentés au paragraphe 2.4 afin de résoudre le nouveau problème de l'équation III.8.

Les changements se situent au niveau du processus de mutation et de croisement et notamment pour la nouvelle caractéristique du choix de classifieur. Ce choix, contrairement aux autres caractéristiques, n'est pas binaire et nécessite donc une modification des processus de mutation (équation III.9) et de croisement (équation III.10).

$$\mathrm{mutation}\left(\left\{\mathrm{IM_i^{iter}}, \mathrm{classifieur_i^{iter}}\right\}, \mathrm{element_{mut}}\right) = \begin{cases} \mathrm{IM_{i,d,k}^{iter}} = 1 - \mathrm{IM_{i,d,k}^{iter}} & \text{si } \mathrm{element_{mut}} = \{\mathrm{d_{mut}}, \mathrm{k_{mut}}\} \\ \mathrm{IM_{i,d,k}^{iter}} = \mathrm{IM_{i,d,k}^{iter}} & \text{sinon} \\ \mathrm{classifieur_i^{iter}} = \text{choix aleatoire parmi} \\ \quad \text{l'ensemble des classifieurs possibles} \\ \quad \text{privé de } \mathrm{classifieur_i^{iter}} & \text{si } \mathrm{element_{mut}} = \mathrm{classifieur} \\ \mathrm{classifieur_i^{iter}} = \mathrm{classifieur_i^{iter}} & \text{sinon} \end{cases} \quad \text{(III.9)}$$

$$\mathrm{croisement}\left(\left\{\mathrm{IM_i^{iter}}, \mathrm{classifieur_i^{iter}}\right\}, \left\{\mathrm{IM_j^{iter}}, \mathrm{classifieur_j^{iter}}\right\}, \left\{\mathrm{IM_i^{a\_croiser}}, \mathrm{classifieur_i^{a\_croiser}}\right\}\right) = \begin{cases} \mathrm{IM_{i,d,k}^{iter}} = \mathrm{IM_{j,d,k}^{iter}} & \text{si } \mathrm{element_{crois}} = \{\mathrm{d_{a\_croiser}}, \mathrm{k_{a\_croiser}}\} \\ \mathrm{IM_{i,d,k}^{iter}} = \mathrm{IM_{i,d,k}^{iter}} & \text{sinon} \\ \mathrm{classifieur_i^{iter}} = \mathrm{classifieur_j^{iter}} & \text{si } \mathrm{element_{crois}} = \mathrm{classifieur_{a\_croiser}} \\ \mathrm{classifieur_i^{iter}} = \mathrm{classifieur_i^{iter}} & \text{sinon} \end{cases} \quad \text{(III.10)}$$

4.2. Applications du problème complet

Reprenons les applications précédentes en incluant maintenant la sélection des canaux et du classifieur, et uniquement avec la méthode de sélection par Algorithmes Génétiques.

Pour des questions de complexité de codage, seuls cinq classifieurs ont été intégrés dans les applications qui suivent : méthode de minimisation de distance de Manhattan, euclidienne, de Tchebychev, de Mahalanobis et méthode ICM.

4.2.1. Applications aux données Formosat-2 année 2009

Le tableau III.5 montre quelques résultats de la sélection de données spectrales, temporelles et du classifieur, appliquée aux données Formosat-2 année 2009 disposant de 64 plans temporels et spectraux. Le jeu de données qui obtient la meilleure performance globale dispose de 63 canaux sur 64 et utilise la méthode ICM. Prenons l'exemple de la classe « Eau libre » dont le meilleur jeu de données contient moins de canaux « proche infrarouge », ce qui s'explique par le fait que cette classe n'a, par définition, que peu de confusion avec des classes de végétation, qui sont très absorbantes dans cette longueur d'onde. L'algorithme ICM est très souvent choisi, sauf pour les classes d'eau et de bâti.

4.2.2. Applications aux données Spot 2/4/5 année 2007

Le tableau III.6 montre quelques résultats de la sélection de données spectrales, temporelles et du classifieur, appliquée aux données Spot-2/4/5 année 2007 disposant de 19 plans temporels et spectraux. Le jeu de données qui obtient la meilleure performance globale dispose de 18 canaux sur 19 et utilise la méthode ICM. Prenons l'exemple de la classe « Tournesol » dont le meilleur jeu de données ne contient que peu de canaux « proche infrarouge », canal n'apportant pas d'information de dissociation entre cultures proches. L'algorithme ICM est très souvent utilisé, sauf pour les classes d'eau et de bâti.

4.2.3. Applications aux données Landsat-5 année 2010

L'application liée à l'utilisation des données Landsat est très intéressante car elle dispose d'une richesse de classes très importante (79 classes, voir annexe 1.5) avec des classes de culture mais surtout de végétation naturelle.

Le tableau III.7 montre quelques résultats de la sélection de données spectrales, temporelles et du classifieur, appliquée aux données Landsat-5 année 2010 disposant de 15 plans temporels et spectraux (annexe 1.5). Le jeu de données qui obtient la meilleure performance globale dispose de 15 canaux sur 15 et utilise la méthode ICM. Prenons l'exemple de la classe « Feuillus » dont le meilleur jeu de données ne contient que peu de canaux de la date de Juillet.

	Global				Feuillus				Tournesol				Eau libre			
Indice AOCI	83,64				95,45				92,44				100			
Nombre de plans	63 sur 64				54 sur 64				56 sur 64				60 sur 64			
Dates / canaux	B	V	R	PIR	B	V	R	PIR	B	V	R	PIR	B	V	R	PIR
15 Février	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
17 Mars	x	x	x	x	x	x	x	x		x	x	x	x	x	x	
21 Mars	x	x	x	x	x	x		x	x	x		x	x	x	x	x
30 Mars	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x
3 Mai	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
23 Juin	x	x	x	x	x	x		x	x	x	x		x		x	
1 Juillet	x	x	x	x	x	x	x	x		x	x		x	x	x	x
12 Juillet	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
26 Juillet	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5 Août	x	x	x	x		x		x	x	x	x		x	x	x	
14 Août	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
22 Août	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
30 Août	x		x	x	x	x		x		x	x	x	x	x	x	x
6 Septembre	x	x	x	x	x	x			x	x	x	x	x	x	x	x
24 Septembre	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
16 Octobre	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Classifieur	ICM				ICM				ICM				Min dist Eucl			

tableau III.5 : Extraits des résultats de la sélection de données Formosat-2 année 2009

	Global				Blé dur				Tournesol				Eau libre			
Indice AOCI	77,55				96,96				60,75				97,89			
Nombre de plans	18 sur 19				16 sur 19				13 sur 19				13 sur 19			
Dates / canaux	V	R	PIR	MIR	V	R	PIR	MIR	V	R	PIR	MIR	V	R	PIR	MIR
15 Fév. 2007	x	x	x	ND	x	x	x	ND		x		ND		x		ND
8 Avril 2007	x	x	x	ND	x	x	x	ND	x	x		ND	x	x	x	ND
20 Juin 2007	x	x	x	ND	x	x	x	ND	x	x		ND	x	x		ND
25 Juillet 2007	x	x	x	ND	x	x		ND	x	x	x	ND			x	ND
15 Sept. 2007	x		x	x	x	x	x	x	x	x	x	x	x	x	x	
16 Nov. 2007	x	x	x	ND			x	ND	x			ND	x	x	x	ND
Classifieur	ICM				ICM				ICM				Min Dist Eucl			

tableau III.6 : Extraits des résultats de la sélection de données Spot 2/5/5 année 2007, ND pour canal non disponible

	Global					Classe Feuillus					Classe Tournesol				
Indice AOCI	41,25					57,93					95,17				
Nombre de plans	15 sur 15					12 sur 15					12 sur 15				
Dates / canaux	B	V	R	PIR	MIR	B	V	R	PIR	MIR	B	V	R	PIR	MIR
20 Juillet 2010	x	x	x	x	x				x	x	x	x	x	x	x
21 Août 2010	x	x	x	x	x	x	x	x	x	x	x	x	x		x
22 Sept. 2010	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Classifieur	ICM					ICM					ICM				

tableau III.7 : Extraits des résultats de la sélection de données Landsat-5 année 2010

5. Introduction du système de Sélection-Classification-Fusion

Dans le cadre de l'automatisation et de l'amélioration des processus de classification supervisée, j'ai introduit un nouveau système de Sélection-Classification-Fusion (SCF) (Masse, Ducrot, & Marthon, 2011) décomposé en trois parties distinctes : la sélection des données, la classification et la fusion des résultats (figure III.11). En appliquant la nouvelle méthode de sélection de données présentée au paragraphe 4, le but est d'obtenir de manière automatique les meilleures images et le meilleur classifieur en fonction d'un critère unique, une optimisation multicritères étant difficile à résoudre pour un nombre de classes important et un univers de recherche de grande dimension.

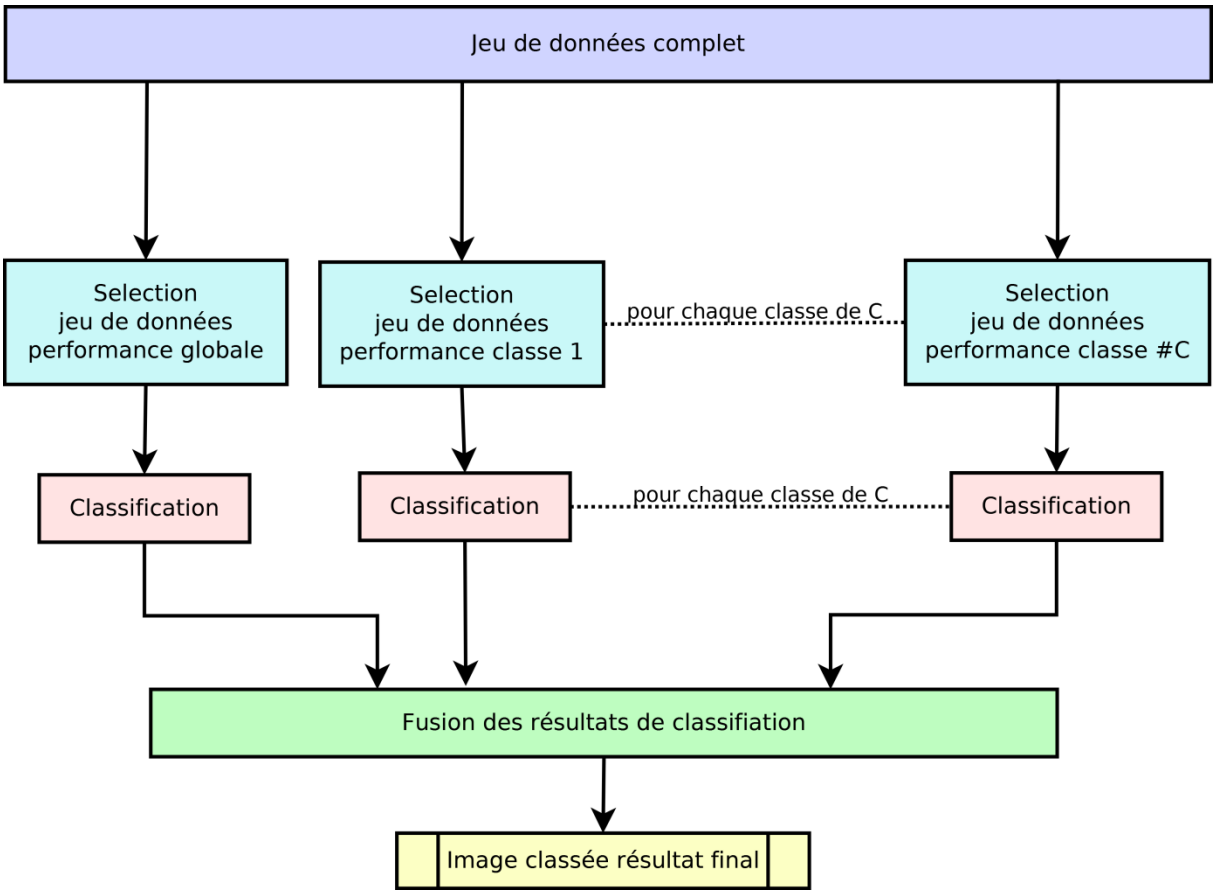


figure III.11 : Schéma illustratif de la méthode de Sélection-Classification-Fusion (SCF)

Nous reviendrons plus en détail sur la partie fusion du processus au cours du chapitre suivant, mais pour le moment considérons qu'un tel algorithme existe et permette de fusionner plusieurs résultats de classification de manière automatique et optimale (partie verte de la figure III.11).

Le but du système SCF est donc de décomposer un système d'optimisation multicritères qui consisterait à trouver la classification qui maximise la performance globale et la performance de chaque classe, en plusieurs processus parallèles d'optimisation monocritère. Ces opérations indépendantes améliorent la rapidité de calcul pour les étapes de sélection et de classification.

6. Conclusion

L'objectif de ce chapitre était de développer une méthode permettant la sélection automatique d'un jeu de données d'entrée d'un processus de classification supervisée, afin de maximiser le critère de performance choisi parmi ceux introduits au chapitre 2.

Parmi les méthodes présentées, la méthode des Algorithmes Génétiques s'est montrée la plus flexible et la plus performante. Cette méthode permet la sélection automatique des données temporelles, spectrales et du choix du classifieur à appliquer.

Cette méthode a été validée sur plusieurs applications différentes et comparée à d'autres approches généralement utilisées en télédétection et sélection de caractéristiques. Les résultats de sélection présentés sont logiques et similaires à ceux qu'un expert aurait pu effectuer manuellement. Rappelons l'exemple des classes de cultures où la sélection correspond aux choix des dates des stades phénologiques. La méthode pourra être utilisée dans le cadre du système automatique d'amélioration du résultat de classification appelé Sélection-Classification-Fusion.

Chapitre IV. La fusion d'images issues de classifications supervisées

Résumé : Les données de télédétection sont de plus en plus variées et abondantes : images optiques et radar, basse, haute, très haute résolution, multi-spectrales, images dérivées (images texturales, images masques) et données exogènes (Modèles Numériques de Terrain, données physiques). Afin d'utiliser cette diversité d'information pour la classification, j'introduirais une méthode originale de fusion de résultats de classification. Cette méthode se voudra automatique, performante et robuste afin d'utiliser au mieux la diversité et la complémentarité des données multidimensionnelles pour la classification. Les applications porteront principalement sur la fusion optique/radar, multi-temporelle et multi-classifieur. Ces applications se serviront des meilleurs jeux de données sélectionnés au chapitre précédent.

Mots clés : fusion résultat classification, confusion, applications optique/radar, multi-classifieurs, multi-temporelles

«Aucun modèle générique n'existe pour tous les problèmes et aucune technique n'est applicable à tous les problèmes. Cependant, nous disposons d'un ensemble varié d'outils performants pour un ensemble tout aussi varié de problèmes à résoudre»

Kanal, 1974

Sommaire du chapitre IV

1. Introduction

Grâce à l'évolution des techniques d'acquisition, les données de télédétection contiennent de plus en plus d'informations. Pour bien comprendre l'intérêt fondamental de la fusion de données, rappelons la diversité des données utilisées en traitement d'images de télédétection : images optiques et radar, basse, haute, très haute résolution, multi-spectrales, images dérivées (images texturales, images masques) et données exogènes (Modèles Numériques de Terrain, données physiques).

L'objectif de ce chapitre est d'introduire une méthode de fusion de données automatique, performante et robuste afin d'utiliser au mieux la diversité et la complémentarité des données multidimensionnelles pour la classification.

Commençons par introduire la notion de fusion. Bloch et Maitre (Bloch, 2005; Bloch & Maître, 2002) définissent la *fusion* comme étant la combinaison d'images hétérogènes pour aider le processus de décision. Une autre définition plus générale est fournie par Wald (Wald, 1999), qui décrit la fusion de données comme étant une structure formelle qui regroupe les moyens et outils pour l'union des données d'une même scène, mais provenant de sources différentes. Nous retrouvons la même notion d'hétérogénéité et de complémentarité.

Chaque source de données fournit une part d'information imparfaite au sens incertain, imprécis, incomplet, contradictoire voire non performant (Smets, 1999). Une méthode efficace de fusion sera celle qui utilisera la complémentarité de chaque caractéristique d'une donnée (dans notre cas temporelle, spectrale,…) afin d'en extraire le meilleur de chacune et d'obtenir un résultat unique. On dit que deux sources d'information sont complémentaires lorsque leur performance diffère en fonction des conditions d'observation et de la nature des propriétés observées. Plus les données sont complémentaires, meilleur est le résultat de leur fusion.

J'introduirai dans un premier temps les différentes méthodes de fusion existantes, qu'elles soient *a priori*, pendant ou *a posteriori* du processus de classification (paragraphe 2). Je présenterai alors une nouvelle méthode de fusion de résultats de classification (paragraphe 3). Cette méthode s'exécutera *a posteriori* du processus de classification et possède de nombreux avantages par rapport aux méthodes de fusion existantes. Cette méthode fera partie intégrante du processus automatique de Sélection-Classification-Fusion présenté au chapitre précédent (paragraphe 4). Enfin au paragraphe 5, nous verrons un large panel d'applications pour valider l'utilisation de la nouvelle méthode de fusion présentée avec notamment des fusions optique/radar, multi-classifieurs et multi-temporelles.

2. La fusion de données

En traitement d'images de télédétection, la fusion peut se décomposer en trois catégories : la fusion de données *a priori*, la fusion de décisions de classifieurs et la fusion de résultats de classifications *a posteriori*, que nous allons détailler dans la suite de ce chapitre. La différence fondamentale entre ces différentes catégories se situe dans la position du processus de fusion au sein de la chaine de traitement (figures IV.1 à IV.3).

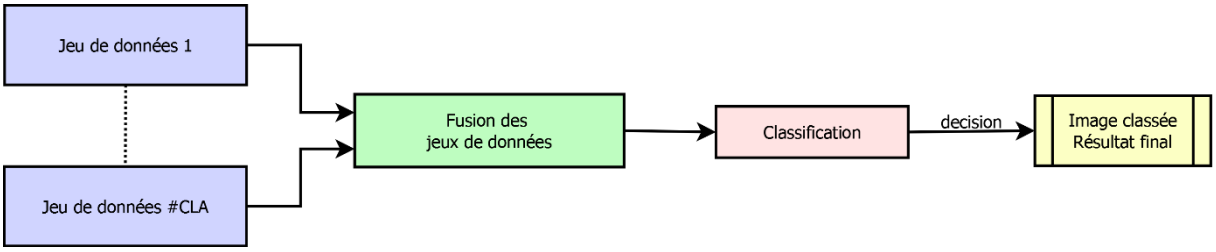


figure IV.1 : Illustration du fonctionnement de la fusion de données a priori

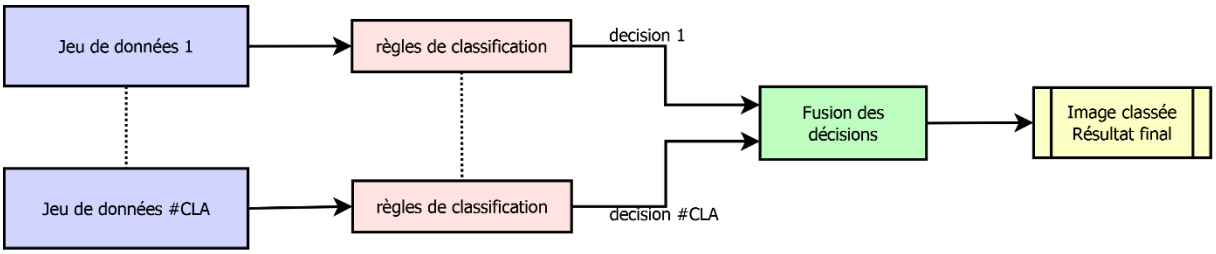


figure IV.2 : Illustration du fonctionnement de la fusion de décisions

La fusion induisant une comparabilité des données à traiter, sa position est importante et peut ne pas être compatible avec toutes les applications. Nous verrons par la suite que la catégorie qui offre le plus de souplesse, i.e. dont les hypothèses sont les moins contraignantes, regroupe les méthodes de fusion de résultats de classification a posteriori.

2.1. La fusion de données a priori

Le premier ensemble regroupe les méthodes de fusion de données a priori. Ces méthodes consistent à fusionner plusieurs images avant le processus de classification (figure IV.1).

Sans entrer dans le détail des méthodes, nous citerons différents travaux concernant le développement et l'application de méthodes de fusion de données. Pour la fusion de données multi sources, nous citerons les travaux de (Agostinelli, Dambra, Serpico, & Vernazza, 1991; Solberg, Jain, & Taxt, 1994). Pour la fusion de données multi résolutions, nous citerons par exemple les travaux de Chavez et ceux de Numez sur la combinaison d'images panchromatiques haute résolution avec des images multi spectrales basse résolution (Chavez, Sides, & Anderson, 1991; Nunez et al., 1999). Enfin, pour la fusion de données multi temporelles, nous citerons les travaux sur les approches simplifiées avec classification en cascade de Swain (Swain, 1978) et ceux de Kalayeh sur l'utilisation du contexte de la corrélation inter pixel (Kalayeh & Landgrebe, 1986).

2.2. La fusion de décisions de classification

Le second ensemble de méthodes regroupe les méthodes de fusion de décision de classifieur au cours du processus de classification. La fusion de décision fournit une décision finale sur la décision du choix de la classe à attribuer au pixel par combinaison des décisions obtenues sur chaque donnée (figure IV.2). Ces méthodes sont très utilisées pour les classifications SVM qui nécessitent de telles méthodes pour une utilisation multi-classes (Chapitre I, paragraphe 2.1.4.1).

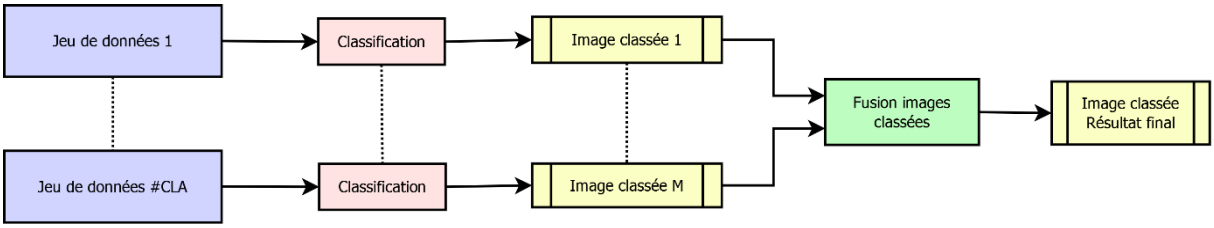


figure IV.3 : Illustration du fonctionnement de la fusion de résultats de classification

Nous citerons ensuite les méthodes basées sur la théorie des possibilités (Dubois & Prade, 1988, 1993; L. A. Zadeh, 1965; L. A. Zadeh, Anvari, Bloch, & Beasse, 1998; Lotfi A. Zadeh, 1968) et la logique floue (Tupin, Bloch, & Maître, 1999; Wang, Hall, & Subaryono, 1990).

Enfin, nous citerons les méthodes basées sur la théorie de l'évidence introduite par Dempster-Shafer (DST) (Dempster, 1960; Shafer, 1976) et qui sont largement utilisées aujourd'hui dans le domaine du traitement d'image (Bloch, 1996; Foucher, Germain, Boucher, & Bénié, 2002; Le Hégarat-Mascle, Bloch, & Vidal-Madjar, 1998) mais aussi dans le domaine du traitement du signal (X. Zhang & Niu, 2009). Plus récemment, la théorie de Dezert-Smarandache (DSmT) (Smarandache, 1991; Smarandache & Dezert, 2006) et son amélioration avec intégration des informations contextuelles (Elhassouny, Idbraim, Bekkari, Mammass, & Ducrot, 2012) ont été introduites et permettent une amélioration des performances de la DST.

Nous citerons les applications de ces méthodes pour la fusion de données multi-sources (Chair & Varshney, 1986; Reibman & Nolte, 1987) et l'étude comparative des méthodes de fusion de Laanaya (Laanaya, Martin, Aboutajdine, & Khenchaf, 2008). Ces études ont permis de montrer deux limitations importantes des méthodes de fusion de décisions :

- Premièrement, ces méthodes nécessitent d'avoir des classifieurs et des fonctions de décisions comparables, ce qui n'est pas le cas, par exemple, pour la fusion de décisions de classifieurs paramétriques pour des données optique et radar.
- Deuxièmement, lorsque le nombre de classes devient important, les processus de fusions deviennent complexes. Dans le cadre de la théorie DSmT, Elhassouny montre qu'à partir de 6 classes, la complexité de calcul devient trop importante (Elhassouny et al., 2012).

2.3. La fusion de résultats a posteriori

Un résumé intéressant des méthodes de fusions de résultats de classification, aussi appelé fusion d'images de classifications, est proposé par Kuncheva (Kuncheva, Bezdek, & Duin, 2001). Cette famille de méthodes de fusion a posteriori regroupe les méthodes de fusions par vote majoritaire (Ramasso, Valet, & Teyssier, 2005) et les méthodes de fusion bayésienne « naïves ». Nous citerons également les applications sur la fusion multi-classifieurs (Kittler, Hatef, Duin, & Matas, 1998) et multi-sources (Briem, Benediktsson, & Sveinsson, 2002). Le fonctionnement de ces méthodes de fusion de résultats a posteriori est illustré en figure IV.3.

Ces méthodes présentent d'emblée l'avantage évident d'une normalisation de toutes les données à fusionner car les produits utilisés sont des images classées avec une structure de classes identique.

3. La fusion d'images de classification supervisées

Je présenterai ici trois méthodes de fusion d'images de classification supervisées. Les deux premières seront basées sur un choix par vote majoritaire et la troisième sera une méthode originale consistant à minimiser la confusion du produit final. Commençons par définir certaines notations.

Notations : Nous notons **CLA** l'ensemble des images résultantes de classification et **cla** un élément de cet ensemble, $\mathbf{performance_{cla}}$ la performance globale de la classification cla et $\mathbf{performance\_classe^{c}_{cla}}$ la performance de la classification cla pour la classe c de l'ensemble des classes C.

3.1. Méthode de fusion par vote majoritaire

Le premier algorithme utilisé pour la fusion d'images issues de classification est la méthode la plus simple : pour chaque pixel, on lui attribue la classe qui apparait le plus souvent dans les images classées fournies en entrée.

Algorithme de la méthode de fusion par vote majoritaire :

Entrée : CLA : ensemble des images issues de diverses classifications
C : l'ensemble des classes à considérer

Pour chaque pixel s

$$c_s^* = \underset{c \in C}{\operatorname{argmax}}(\#\{cla_s == c \text{ tq } cla \in CLA\})$$

Fin pour

3.2. Méthode de fusion par vote majoritaire pondéré

Le second algorithme est basé sur le principe du premier mais avec l'ajout un système de pondération de l'information. La classe finale ne sera plus attribuée à la classe majoritaire mais à la classe possédant la somme des scores de performance maximale. Chaque classification aura comme poids sa performance globale.

Algorithme de la méthode de fusion par vote majoritaire pondéré :

Entrée : CLA : ensemble des images cla issues de diverses classifications
C : l'ensemble des classes à considérer
$performance_{cla}$: la performance de la classification cla

Pour chaque pixel s

$$c_s^* = \underset{c \in C}{\operatorname{argmax}}\left(\#\{cla_s == s, cla \in CLA\} \times \sum_{cla_s == c} performance_{cla}\right)$$

Fin pour

3.3. Méthode de fusion par minimisation de confusion

Nous allons désormais ajouter une restriction supplémentaire : la performance de chaque classe. Le principe est le suivant : nous déterminons une classification de référence (la meilleure en matière de performance globale) et pour chaque classe, une classification de référence pour cette classe (en matière de performance de la classe). Chaque pixel O de la classification de référence est ensuite comparé au pixel P de la classification de référence pour la classe trouvée en O (figure IV.4). Parmi ces deux classes trouvées, on sélectionne celle qui minimisera la confusion finale.

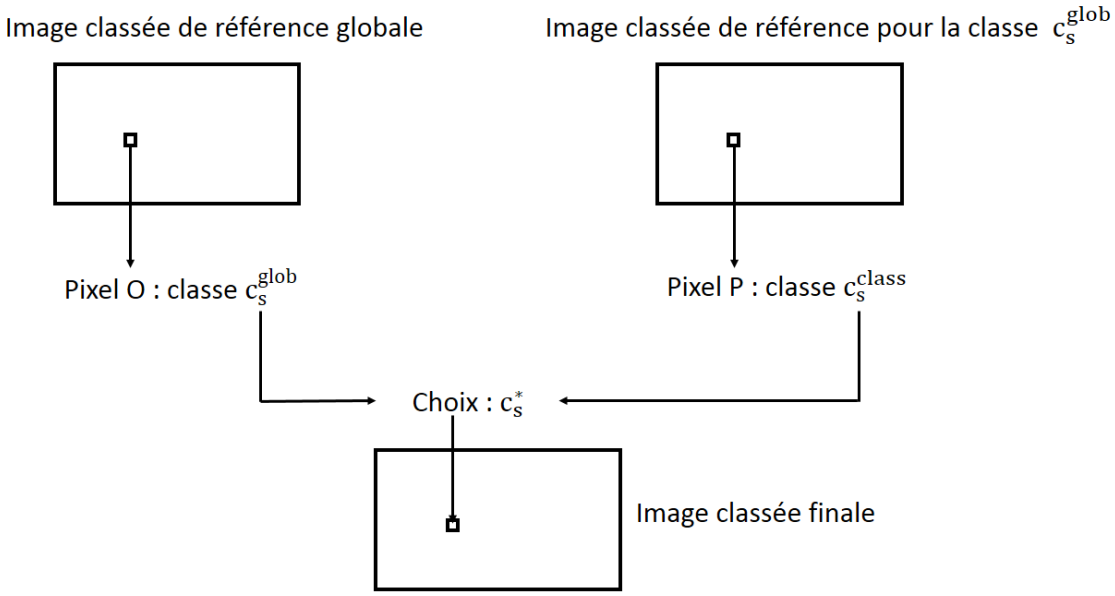


figure IV.4 : Illustration de l'étape 2 de l'algorithme de fusion par minimisation de confusion

Cette méthode applique donc une restriction de l'espace des possibilités d'attribution de la classe finale car contrairement aux méthodes précédentes qui utilisent #CLA possibilités (*id est* le nombre d'images donné en entrée de la méthode), la nouvelle méthode restreint l'univers des possibilités à 2 éléments.

L'algorithme sera donc décomposé en deux parties. La première consistera à déterminer les images de référence suivant #C + 1 critères : la référence qui maximisera la performance globale et pour chaque classe une référence qui maximise la performance de la classe. La seconde consistera à parcourir l'image et à comparer les éléments de référence en fonction des classes observées, et à choisir celle qui minimisera les confusions de l'image finale (et donc en maximisera la performance).

L'algorithme est présenté ci-après et schématisé à la figure IV.5. Le principe de la méthode est le suivant : on parcourt l'image classée de référence qui maximise la performance globale. La classe ainsi trouvée est comparée à celle de l'image classée qui maximise la performance de cette classe. Si les classes observées sont identiques, cette classe est conservée. Sinon, on choisit celle qui minimise la confusion du résultat final en comparant les matrices de confusion des deux références.

Algorithme de la méthode de fusion par minimisation de confusion

Entrée : CLA : ensemble des images cla issues de diverses classifications
M : ensemble des matrices de confusions m associées à CLA
C : l'ensemble des classes à considérer
performance_{cla} : la performance de la classification cla
$\text{performance\_classe}^{c}_{cla}$: la performance de la classification cla pour la classe c

Etape 1.a : Extraction de la meilleure classification au sens performance globale

$$cla^{ref\_globale} = \underset{cla}{\text{argmax}}\, \text{performance}_{cla}$$

Etape 1.b : Extraction des meilleures classifications au sens performance de chaque classe

Pour chaque classe $c \in C$

$$cla^{ref\_classe_c} = \underset{cla}{\text{argmax}}\, \text{performance\_classe}^{c}_{cla}$$

Fin Pour

Etape 2 : Parcours de l'image et attribution de la classe (figure IV.4)

Pour chaque pixel s

$$c_s^{glob} := c_s^{cla^{ref\_globale}} \quad ; \quad m^{glob} := m^{cla^{ref\_globale}}$$

$$c_s^{class} := c_s^{cla^{ref\_classe_{c_s^{glob}}}} \quad ; \quad m^{class} := m^{cla^{ref\_classe_{c_s^{glob}}}}$$

$$A := m^{glob}_{c_s^{glob}, c_s^{class}} \quad ; \quad C := m^{class}_{c_s^{glob}, c_s^{class}}$$

$$B := m^{glob}_{c_s^{class}, c_s^{glob}} \quad ; \quad D := m^{class}_{c_s^{class}, c_s^{glob}}$$

$$c_s^* = \begin{cases} c_s^{glob} & \text{si } (A+B)-(C+D) \leq 0 \\ c_s^{class} & \text{sinon} \end{cases}$$

Fin pour

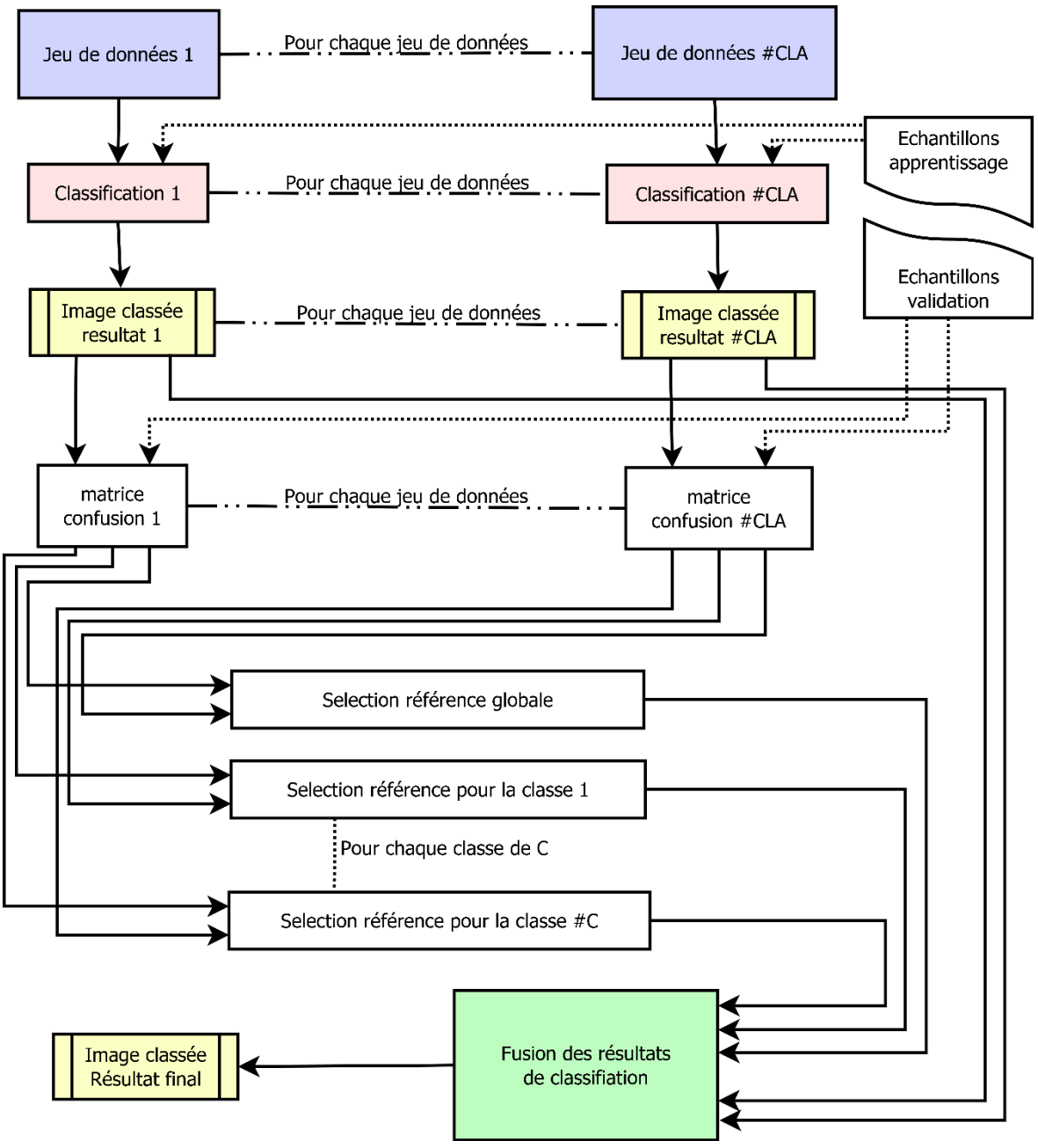


figure IV.5 : Illustration de la méthode de fusion par minimisation de confusion

4. Application au processus de Sélection-Classification-Fusion

A partir des jeux de données sélectionnés par les algorithmes du chapitre 3, nous appliquons maintenant le processus de fusion par minimisation de confusion afin de combiner les résultats des classifications de ces jeux (figure IV.6, partie verte).

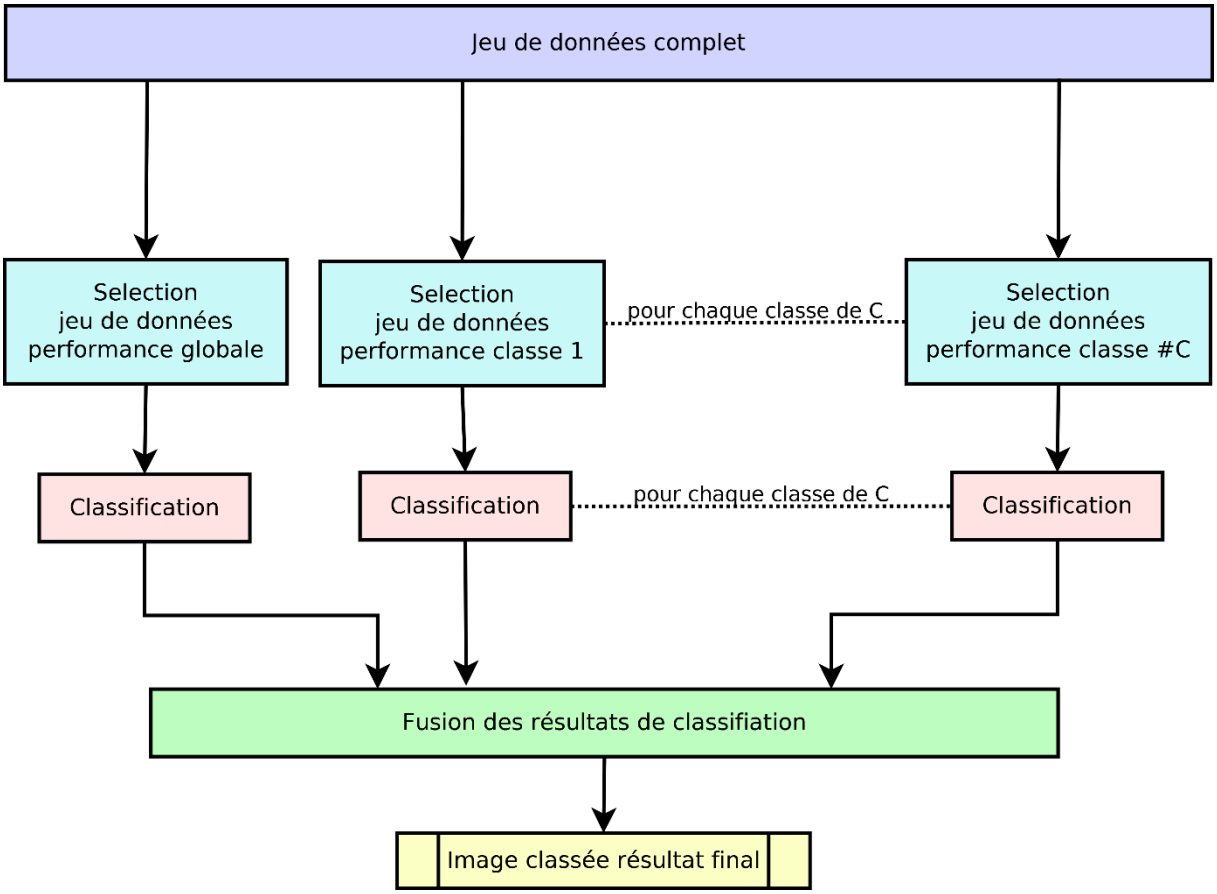


figure IV.6 : Illustration du processus de Sélection-Classification-Fusion

Notons que ce processus, excepté la prise d'échantillons, est automatique et permet d'obtenir un résultat optimal. Notons également sa structure parallèle qui lui permet des exécutions rapides sur des architectures multiprocesseurs.

En résumé, ce système permet d'améliorer de manière entièrement automatique la performance de n'importe quelle classification d'un jeu de données multidimensionnelles et ceux sans surcout important du temps de calcul.

La suite de l'application de ce système sera présentée au paragraphe 5.2.

Critères d'évaluation (en % sauf Kappa)	Parallélépipède	Minimum Distance Euclidienne	Minimum Distance Mahalanobis	Maximum de vraisemblance	Réseaux neuronaux	SVM à noyau linéaire	SVM à noyau Polynomial	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	20,16	50,51	70,78	**89,48**	65,90	86,76	88,83	88,61	**91,32**	91,18
AP	NaN	33,89	57,76	**85,42**	NaN	71,42	74,23	80,57	82,07	**86,50**
OA	13,97	28,75	67,82	**83,87**	86,23	80,76	82,93	85,44	87,84	**88,76**
Kappa	0,10	0,25	0,65	0,82	**0,85**	0,79	0,81	0,84	0,87	**0,88**
AOCI	9,31	24,16	44,14	**76,14**	56,55	62,87	66,47	72,01	75,33	**79,19**
F1 mesure	NaN	0,41	0,64	**0,87**	NaN	0,78	0,81	0,84	0,86	**0,89**

tableau IV.1 : Comparaison des indices de performance globale en fonction des différents classifieurs utilisés et de leur fusion pour l'application Formosat-2 année 2009.

5. Applications des méthodes de fusion a posteriori de résultats de classifications

5.1. La fusion de résultats de classifications obtenus à partir de différents classifieurs

Nous avons observé au cours des chapitres I et II la complémentarité des classifieurs ponctuels. Afin d'exploiter cette complémentarité, nous avons appliqué les méthodes de fusion *a posteriori* introduites dans le paragraphe 3 sur les résultats de classifications du Chapitre II.

Reprenons l'application des images Formosat-2 année 2009 (rappel en annexe 1.3) et les classifications effectuées au chapitre 2 sur l'évaluation des différents classifieurs ponctuels. Le Maximum de Vraisemblance l'emportait majoritairement sur tous les tableaux (tableaux IV.1 et IV.2) suivi de deux autres classifieurs (NN et SVM à noyau Polynomial).

Comparons maintenant les trois méthodes de fusion a posteriori introduites : vote majoritaire, vote majoritaire pondéré et minimisation de confusion.

La fusion des classifieurs a permis d'obtenir une image classée dont les indices de performance globale sont en moyenne 2,5% plus élevés (tableau IV.1). Les résultats de la méthode de fusion par minimisation de confusion sont proches de ceux de la méthode par vote majoritaire pondéré mais néanmoins supérieur de 1%. La diversité des résultats est plus marquée concernant l'indice de performance par classe OCI (tableau IV.2). La méthode par minimisation de confusion obtient des résultats proches voire supérieurs à ceux des deux autres méthodes. L'exemple de la classe « blé dur » est illustré aux figures IV.7 à IV.13, seule la méthode de fusion par minimisation de confusion a permis d'améliorer la distinction entre « orge » et « blé dur ».

OCI (en %) pour la classe	Paralléléppède	Minimum Distance Euclidienne	Minimum Distance Mahalanobis	Maximum de vraisemblance	Réseaux neuronaux	SVM à noyau linéaire	SVM à noyau Polynomial	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de
bâti dense	0,73	6,78	9,35	**84,55**	0,00	41,15	58,41	48,93	60,48	**84,55**
bâti/surf min.	**95,06**	72,20	78,42	93,92	0,00	82,71	86,63	90,47	92,09	**93,92**
bâti diffus	0,08	0,16	1,27	**22,08**	0,02	10,68	9,03	14,78	19,29	**20,58**
blé dur	1,24	21,28	37,10	61,34	**70,15**	50,93	60,09	58,21	66,61	**70,23**
blé tendre	1,62	0,00	22,45	50,65	**72,39**	51,35	55,17	49,25	62,65	**72,36**
chanvre	0,05	9,35	42,98	**94,25**	86,30	58,21	76,17	80,93	87,00	**93,99**
colza	12,70	0,00	83,50	88,91	**91,74**	88,07	84,75	92,03	**93,38**	91,42
eau libre	0,00	98,31	95,76	99,30	89,53	99,48	**99,55**	**99,69**	99,67	99,48
eucalyptus	0,00	51,34	39,60	**93,30**	64,92	83,01	82,19	84,92	90,81	**91,28**
eucalyptus 2	0,00	8,59	12,38	**90,56**	45,09	69,89	63,78	70,91	78,13	**90,51**
feuillus	86,09	78,32	80,68	**95,24**	91,72	92,45	94,39	91,13	94,19	**95,35**
friche	0,00	10,30	20,37	27,32	30,95	39,54	**46,79**	35,88	42,97	**48,24**
gravières	0,00	97,75	97,83	**99,72**	95,45	99,11	98,33	97,84	**99,95**	99,72
jachère/gel	0,46	4,80	30,37	43,17	**48,57**	42,30	48,27	52,93	**55,77**	54,15
lac	0,48	96,81	97,20	99,92	94,76	**99,96**	99,37	97,21	**100,0**	99,97
mais	0,08	32,58	82,27	**96,44**	94,22	92,02	89,95	93,72	95,94	**96,57**
mais ensilage	16,61	7,45	69,30	**92,75**	77,17	78,50	71,41	85,24	88,45	**92,75**
mais non irrig.	0,08	0,00	2,38	**92,33**	0,00	10,32	7,57	15,97	30,43	**92,33**
orge	0,00	0,00	33,42	68,43	**71,17**	50,11	48,56	62,34	**69,87**	69,65
peupliers	0,00	19,20	45,69	**92,98**	71,29	79,61	82,28	76,85	84,92	**92,54**
pois	0,84	0,00	0,77	**61,69**	4,71	25,06	38,22	34,56	**60,51**	56,04
prairie temp.	0,01	0,00	24,28	**53,60**	50,93	38,96	49,28	51,89	**56,07**	55,56
prairie perm.	0,00	0,00	20,19	37,85	30,56	29,27	**44,50**	39,19	**43,24**	38,77
résineux	0,00	42,42	36,06	**95,78**	68,37	86,42	89,80	84,70	91,12	**95,78**
soja	6,66	0,00	50,51	**91,91**	72,21	80,67	85,01	88,11	**91,21**	89,09
sorgho	8,59	14,12	34,92	54,20	**71,35**	57,57	59,13	51,83	62,57	**74,28**
surface min.	15,84	5,29	18,59	**60,04**	0,00	37,03	46,02	47,19	49,59	**60,04**
tournesol	13,50	0,01	69,09	86,64	**91,23**	81,64	85,06	86,53	88,57	**93,03**

tableau IV.2 : Comparaison des valeurs de l'indice de performance par classe OCI, en fonction des différents classifieurs utilisés et de leur fusion, suivant les 28 classes présentes pour l'application Formosat-2 année 2009

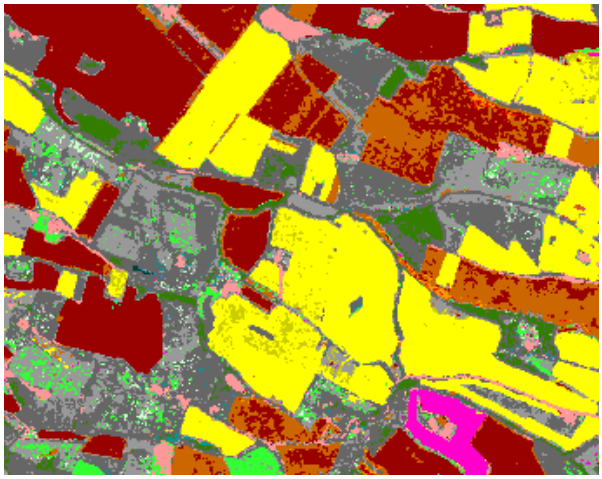


figure IV.7 : (a) Méthode de classification MV

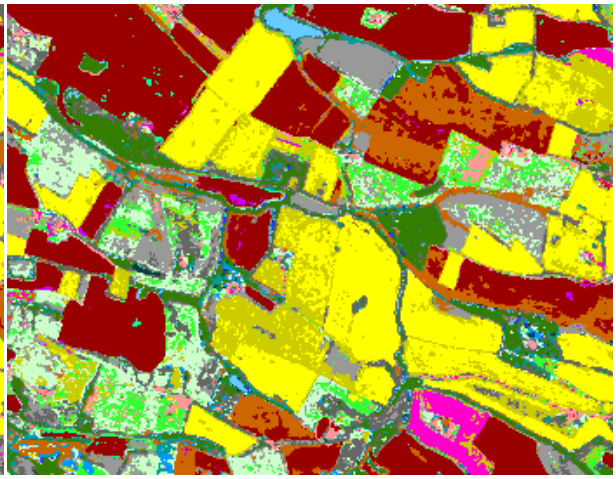


figure IV.8 : (b) Méthode SVM polynomial (degré 2)

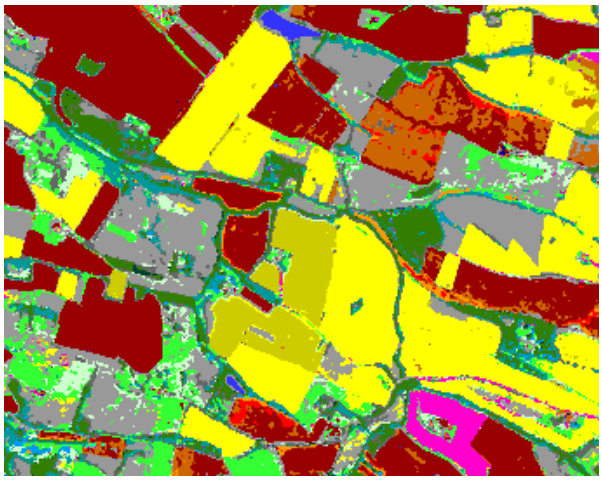


figure IV.9 : (c) Méthode des réseaux de neurones

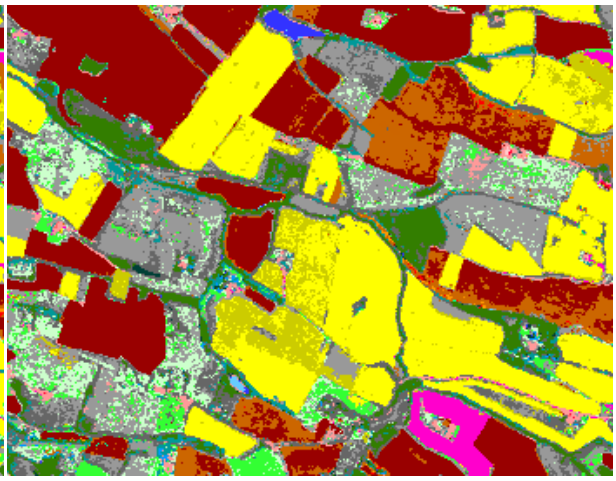


figure IV.10 : (d) Fusion par la méthode de vote majoritaire

figure IV.11 : (e) Fusion par la méthode de vote majoritaire pondéré

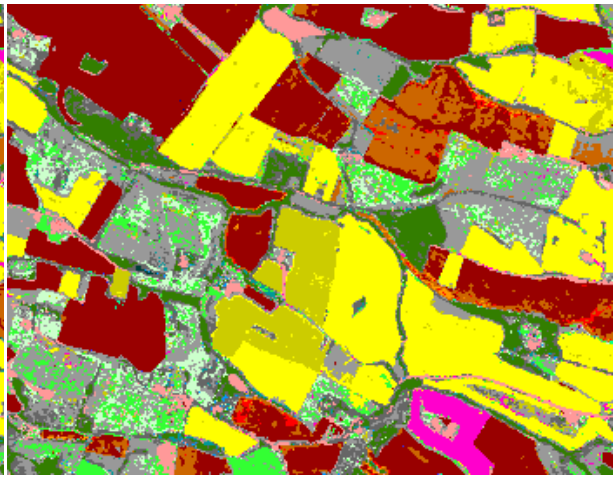


figure IV.12 : (f) Fusion par la méthode de minimisation de confusion

figure IV.13 : Données Formosat-2 année 2009, extrait des résultats de méthodes de classification (a), (b) et (c), et extrait des images de fusion des résultats obtenus à partir des différents classifieurs (d), (e) et (f).

Critères d'évaluation (en % sauf Kappa)	Parallélépipède	Minimum Distance Euclidienne	Minimum Distance Mahalanobis	Maximum de vraisemblance	Réseaux neuronaux	SVM à noyau linéaire	SVM à noyau Polynomial	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de
AA	7,35	63,13	71,64	**83,43**	47,94	74,16	73,16	82.09	83.78	**84.83**
AP	NaN	49,29	57,22	**81,03**	NaN	58,45	58,62	69.75	71.73	**81.77**
OA	4,81	53,59	67,09	**81,75**	68,48	64,74	66,39	77.11	79.35	**81.76**
Kappa	0,03	0,49	0,63	**0,79**	0,63	0,61	0,62	0.74	0.77	**0.79**
AOCI	3,66	34,36	42,32	**69,38**	36,15	45,43	45,59	58.61	61.35	**70.67**
F1 mesure	NaN	0,55	0,64	**0,82**	NaN	0,65	0,65	0.75	0.77	**0.83**

tableau IV.3 : Comparaison des indices de performance globale en fonction des différents classifieurs utilisés et de leur fusion pour l'application Spot-2/4/5 année 2007

Critères d'évaluation (en % sauf Kappa)	Parallélépipède	Minimum Distance Euclidienne	Minimum Distance Mahalanobis	Maximum de vraisemblance	Réseaux neuronaux	SVM à noyau linéaire	SVM à noyau Polynomial	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de
AA	13,09	41,56	48,10	**68,81**	22,06	56,52	59,53	62,29	68,05	**69,08**
AP	NaN	23,86	27,57	**43,61**	NaN	32,99	35,09	40,59	42,06	**44,58**
OA	29,11	45,27	51,23	66,69	**68,84**	58,46	59,83	67,77	**68,75**	67,16
Kappa	0,22	0,41	0,47	0,63	**0,65**	0,54	0,56	0,64	**0,65**	0,64
AOCI	3,43	10,01	13,18	**30,41**	12,98	18,88	21,07	25,81	28,87	**31,28**
F1 mesure	NaN	0,30	0,35	**0,53**	NaN	0,42	0,44	0,49	0,52	**0,54**

tableau IV.4 : Comparaison des indices de performance globale en fonction des différents classifieurs utilisés et de leur fusion pour l'application Landsat-5 année 2010

Nous observons des résultats similaires dans le cadre des applications aux images Spot 2/4/5 de l'année 2007 (annexe 1.4) et Landsat 5 de l'année 2010 (annexe 1.5). Ces résultats sont respectivement résumés dans les tableaux IV.3 et IV.4.

5.2. La fusion de classifications obtenues à partir de jeux de dates différents

Dans le Chapitre III et dans le cadre du système de Sélection-Classification-Fusion, plusieurs sélections de données ont été faites afin de maximiser la performance suivant plusieurs critères : une classification qui maximise la performance globale et pour chaque classe, une classification qui maximise l'indice de performance de cette classe.

Afin d'exploiter cette complémentarité, nous avons appliqué les méthodes de fusion *a posteriori* introduites dans le paragraphe 3 sur ces applications de sélection.

5.2.1. L'application aux données Formosat-2 de l'année 2009

Dans le cadre du processus de Sélection-Classification-Fusion, un jeu de données de référence globale et un jeu de données de référence pour chaque classe ont été sélectionnés. Leur fusion par méthode de minimisation de confusion permet l'augmentation des indices globaux (en moyenne 1,5%, tableau IV.5) et pour chaque classe (en moyenne 3%, tableau IV.6). Les méthodes par vote majoritaire sont ici peu performantes du fait du grand nombre de données à fusionner (nombre de classes +1) et donc de la redondance d'erreurs. La nouvelle méthode est donc plus robuste et performante que les méthodes basées sur les votes majoritaires, même pondérés.

5.2.2. L'application aux données Spot-2/4/5 de l'année 2007

Le tableau IV.7 indique une nette amélioration des performances pour la méthode de fusion par minimisation de confusion. Les conclusions concernant la robustesse et la performance de cette méthode se confirment de nouveau pour cette application.

5.2.3. L'application aux données Landsat-5 de l'année 2010

Les résultats obtenus avec les 3 images Landsat-5 sont très intéressants car ils nous montrent que même avec peu de dates et un grand nombre de classes (79 classes, détails en annexe 1.5), il est possible d'extraire des informations complémentaires et ainsi améliorer la performance finale (tableau IV.8).

L'amélioration des performances est très faible mais néanmoins significative. Nous retrouvons le problème de redondance d'erreurs pour les méthodes de fusion par vote majoritaire.

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**94,04**	93,88	93,88	**94,40**
AP	**90,91**	90,09	90,08	**91,48**
OA	**91,36**	90,91	90,90	**92,36**
Kappa	**0,91**	0,90	0,90	**0,92**
AOCI	**85,60**	84,74	84,74	**86,52**
F1	**0,92**	0,92	0,92	**0,93**

tableau IV.5 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Formosat-2 année 2009

OCI (en %) pour la classe	Référence globale	Référence pour la classe	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
bâti dense	**87,05**	**87,05**	**89,93**	**89,93**	86,33
bâti/surf minérale	90,65	**99,29**	**92,81**	**92,81**	92,09
bâti diffus	**43,10**	**43,10**	42,67	42,59	**43,77**
blé dur	**84,36**	**84,36**	82,12	82,13	**84,27**
blé tendre	**82,83**	**82,83**	80,54	80,53	**82,97**
chanvre	**97,93**	**97,93**	96,40	96,47	**97,13**
colza	**95,71**	**95,71**	**96,04**	96,03	95,91
eau libre	97,73	**99,73**	97,73	97,73	**97,82**
eucalyptus	**96,80**	**96,80**	95,03	95,09	**96,34**
eucalyptus 2	**96,31**	**96,31**	95,85	95,85	**97,20**
feuillus	**95,75**	**95,75**	95,81	95,79	**95,97**
friche	50,33	**52,20**	50,04	50,13	**56,97**
gravières	99,52	**99,98**	**99,54**	**99,54**	**99,54**
jachère/surface gel	57,26	**59,45**	56,79	56,74	**63,25**
lac	99,90	**100,00**	**99,92**	**99,92**	99,90
mais	**98,15**	**98,15**	97,63	97,63	**98,25**
mais ensilage	93,66	**95,23**	92,88	92,86	**93,92**
mais non irrigue	**93,04**	**93,04**	91,40	91,40	**92,63**
orge	**83,86**	**83,86**	83,19	83,18	**83,27**
peupliers	**96,24**	**96,24**	95,96	95,96	**96,15**
pois	**94,25**	**94,25**	90,32	90,52	**94,25**
prairie temporaire	61,82	**62,21**	60,58	60,57	**65,80**
pré/prairie permanente	48,74	**52,93**	48,70	48,55	**54,87**
résineux	**96,99**	**96,99**	96,05	96,11	**97,32**
soja	**97,84**	**97,84**	97,60	97,59	**97,92**
sorgho	**80,36**	**80,36**	77,17	77,16	**82,51**
surface minérale	83,66	**85,62**	77,51	77,51	**82,82**
tournesol	**93,03**	**93,03**	92,46	92,45	**93,24**

tableau IV.6 : Comparaison des valeurs de l'indice de performance par classe OCI de la référence globale et des références par classe et de leur fusion suivant les 28 classes présentes pour l'application Formosat-2 année 2009

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**84,84**	83,54	83,51	**85,04**
AP	**81,74**	77,97	77,96	**81,38**
OA	**81,75**	82,14	82,13	**83,05**
Kappa	**0,79**	0,80	0,80	**0,81**
AOCI	**70,67**	66,86	66,83	**70,72**
F1	**0,83**	0,81	0,81	**0,83**

tableau IV.7 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Spot 2/4/5 année 2007

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**75,09**	72,94	74,26	**75,11**
AP	**54,03**	50,22	51,35	**54,11**
OA	**70,50**	70,66	**70,91**	70,67
Kappa	**0,67**	0,67	**0,68**	**0,68**
AOCI	**41,25**	37,01	38,42	**41,55**
F1	**0,63**	0,59	0,61	**0,63**

tableau IV.8 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Landsat 5 année 2010

5.3. La fusion de classifications obtenues à partir de sources différentes

5.3.1. La fusion de classifications de données Spot et Radarsat

Dans le cadre d'un programme de recherche pour la promotion des données Radarsat (Application Developement and Research Opportinity, ADRO), l'objectif était de définir un indicateur d'environnement pour évaluer la vulnérabilité du milieu naturel face à des pollutions engendrées par l'activité humaine. L'outil nécessaire à cet objectif était donc l'établissement d'une carte d'occupation des sols.

Ces cartes d'occupation des sols ont été établies par le biais de plusieurs classifications à partir des données disponibles :

- 4 images Radarsat, Single Look Complexe (SLC) en mode F1, prises aux dates du 28 Juin 1997, 22 Juillet 1997, 08 Septembre 1997 et 19 Novembre 1997.
- 1 image Spot en mode XS (multi-spectral, 3 bandes) du 13 Août 1997.

La zone d'étude se situe dans un bassin du Grand Morin (le Vannetin affluent du Grand Morin) en Seine et Marne (France) qui fournit en eau potable une partie de la région parisienne. Le site est essentiellement constitué de parcellaires agricoles, de bois et de forêts.

Nous pouvons exploiter la complémentarité des données optiques et radar afin d'extraire une information plus riche, ce qui a permis une meilleure distinction des différentes classes. Seules les méthodes de fusion a posteriori n'imposent pas d'hypothèse sur la normalisation des données et la comparabilité des classifieurs.

Le processus de fusion avec les images Radar non filtrées, classées avec la loi de Gauss Wishart (GW, loi contextuelle), améliore les résultats de l'image SPOT (loi de Gauss) pour la majorité des classes : environ 2% en moyenne (tableau IV.9).

L'apport de la classification ne transparait que sur les classes « orge » et « urbain » (tableau IV.10), néanmoins la complémentarité des données permet d'obtenir d'autres améliorations notamment pour la classe de blé tendre (+13%).

Les figures IV.14 à IV.21 présentent les différents résultats et données utilisés lors de cette application.

La visualisation de cette classification (figure IV.20) montre les améliorations pour la Forêt, le Maïs et le Pois. La confusion des pixels de contours des champs de maïs avec l'herbe est réduite, ce qui améliore les structures des champs.

Critères d'évaluation (en % sauf Kappa et F1)	Classification données Spot	Classification données RadarSat	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**81,30**	72,54	75,59	80,96	**82,71**
AP	**84,45**	74,35	84,58	84,16	**86,09**
OA	**83,59**	77,32	81,48	83,13	**85,70**
Kappa	**0,81**	0,73	0,78	0,80	**0,83**
AOCI	**70,02**	55,73	65,06	69,56	**72,74**
F1	**0,83**	0,73	0,80	0,83	**0,84**

tableau IV.9 : Comparaison des indices des résultats de classification des données optique et radar et de leur fusion en fonction d'indices de performance globaux.

OCI (en %) pour la classe	Classification données Spot	Classification données RadarSat	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
betterave	**97,84**	51,09	80,33	**97,84**	**97,84**
blé dur	**58,53**	38,05	49,29	58,12	**58,53**
blé tendre	**36,95**	36,78	42,06	36,23	**49,61**
colza	**44,77**	27,77	40,48	**44,64**	42,80
forêt	**94,45**	82,13	85,46	93,37	**94,45**
herbe	**89,99**	80,18	85,30	89,95	**89,99**
maïs	**80,67**	58,77	67,42	79,52	**80,67**
orge	16,93	**19,31**	**27,09**	16,95	20,90
pois	**69,88**	61,86	64,99	68,50	**72,37**
sol nu	**98,12**	72,61	90,23	97,89	**98,12**
urbain	82,13	**84,53**	83,01	82,13	**94,81**

tableau IV.10 : Comparaison des indices des résultats de classification des données optique et radar et de leur fusion en fonction d'indices de performance par classe

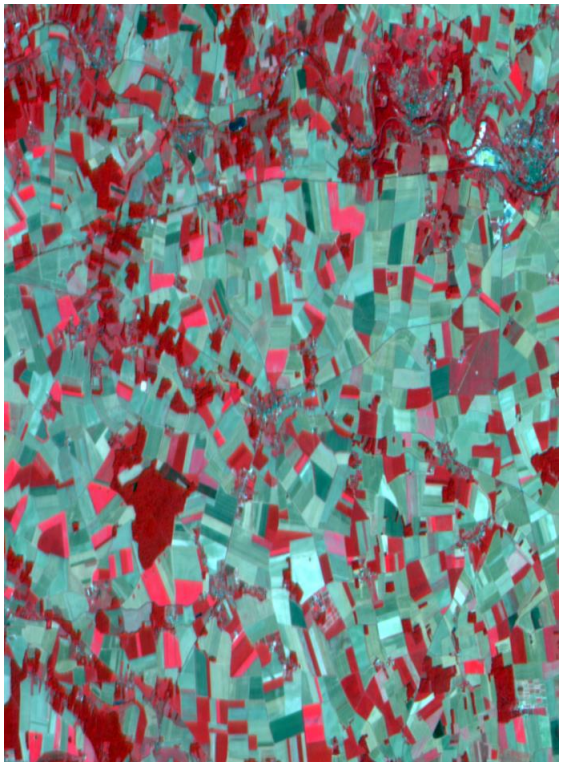


figure IV.14 : Composition colorée de l'image Spot

figure IV.15 : Composition colorée d'une image Radarsat non filtrée utilisées pour la classification

figure IV.16 : Classification des données Spot par la méthode ICM

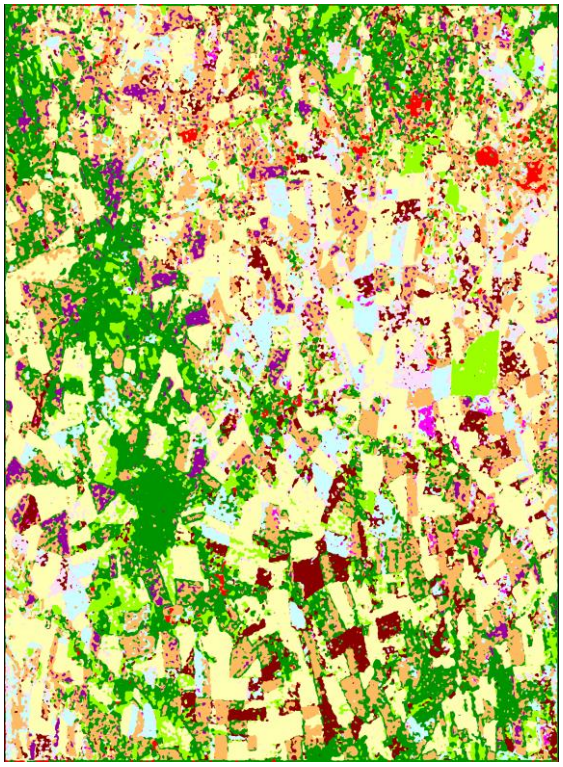


figure IV.17 : Classification des données RadarSat par la méthode ICM et une loi de Gauss-Wishart

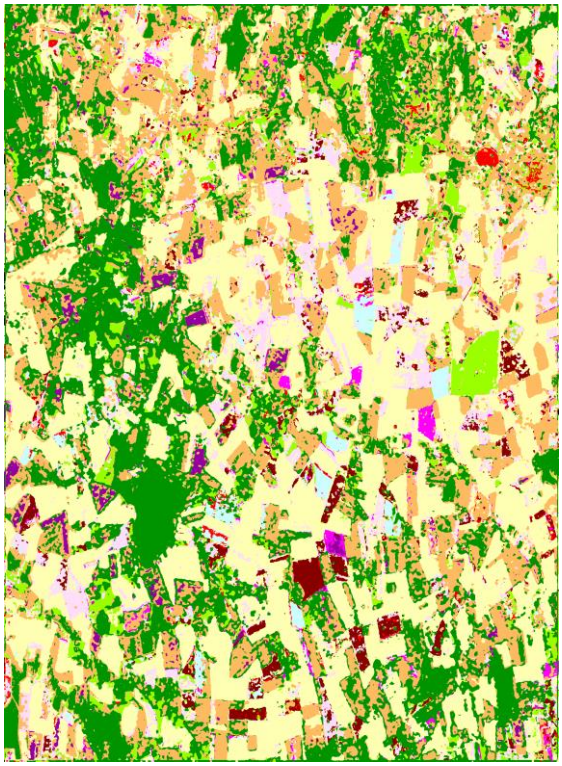


figure IV.18 : Fusion vote majoritaire

figure IV.19 : Fusion vote majoritaire pondéré

figure IV.20 : Fusion par minimisation de confusion

figure IV.21 : Echantillons de vérification

5.3.2. La fusion de classifications de données Spot et ERS

En région au climat tempéré, les données optiques sont généralement bien adaptées pour l'établissement de cartes de surfaces agricoles. Néanmoins, la précision des cartes obtenues par classification de données images est fortement dépendante des dates d'acquisition de ces images. En effet, les stades phénologiques des cultures doivent pouvoir être repérés dans la série temporelle. La précision temporelle des prises de vue optiques étant dépendante de la clémence des conditions météorologiques, les possibilités offertes par les capteurs radar spatiaux intéressent les organismes chargés d'effectuer ces suivis.

Dans le cadre d'un contrat CNES, la classification était la dernière étape d'un processus complet de traitement de données ERS-1 pour les rendre utilisables à des fins de suivi agricole. Ce processus inclut une pré-segmentation des images, un filtrage et une classification.

Les données disponibles sont :

- 6 images ERS-1, Single Look Complex (SLC), prises aux dates : 26 Mars 1993, 30 Avril 1993, 04 Juin 1993, 09 Juillet 1993, 13 Aout 1993 et 17 Septembre 1993.
- 2 données SPOT du 09 Juin 1993 et du 26 Août 1993.
- 2 jeux de vérités terrain (apprentissage et vérification).

La zone sur laquelle nous avons travaillé est formée de 2900 lignes de 1100 pixels, correspondant à une surface au sol d'environ 36x14 km2 (le pixel correspond à une surface au sol de 12.5x12.5 m²).

Une méthode de classification ICM avec loi de Gauss est appliquée aux images Spot. La même méthode est également appliquée aux images Radarsat non filtrées avec une loi de Gauss-Wishart.

L'apport de la complémentarité optique/radar lors de la fusion est ici plus important que pour l'application précédente. La méthode de fusion par minimisation de confusion améliore les résultats de performance globale de 2% en moyenne (tableau IV.11). Le tableau IV.12 montre les résultats des performances par classe. L'apport de la fusion sur les performances par classe est ici assez faible.

Les figures IV.22 à IV.29 présentent les différents résultats et données utilisées lors de cette application. Visuellement, nous pouvons constater l'amélioration des classes de bâti et de forets.

Critères d'évaluation (en % sauf Kappa)	Classification données Spot	Classification données ERS	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**86,60**	74,79	79,15	86,60	**88,05**
AP	**79,21**	70,35	**81,29**	79,22	81,16
OA	**88,36**	78,16	85,18	88,37	**90,79**
Kappa	**0,87**	0,75	0,83	0,87	**0,90**
AOCI	**70,11**	54,90	64,71	70,12	**72,92**
F1	**0,83**	0,73	0,80	0,83	**0,84**

tableau IV.11 : Comparaison des classifications de données Spot et ERS et de leur fusion selon les 3 méthodes introduites précédemment suivant les indices de performance globaux

OCI (en %) pour la classe :	Classification données Spot	Classification données ERS	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
blé dur	**70,54**	62,75	69,95	70,63	**70,91**
blé tendre	**63,21**	49,41	46,70	63,21	**63,75**
colza	65,61	**80,15**	76,21	65,67	**85,04**
eau	**98,59**	93,22	94,37	98,59	**98,64**
forêt	**99,06**	83,78	86,04	99,06	**99,20**
maïs	**90,78**	52,55	86,27	90,78	**90,78**
orge	**29,20**	10,19	**37,07**	29,20	22,06
pois	50,14	**58,76**	**66,58**	50,14	53,52
prairie	**74,07**	25,76	44,48	**74,07**	73,28
sol nu	**90,19**	28,17	49,59	90,19	**90,27**
tournesol	61,58	**74,80**	69,37	61,58	**81,93**
trèfle	**21,40**	13,18	**28,11**	21,40	21,40
urbain	**96,99**	80,97	86,50	96,99	**97,15**

tableau IV.12 : Comparaison des classifications de données Spot et ERS et de leur fusion selon les 3 méthodes introduites précédemment suivant l'indice de performance OCI pour chaque classe

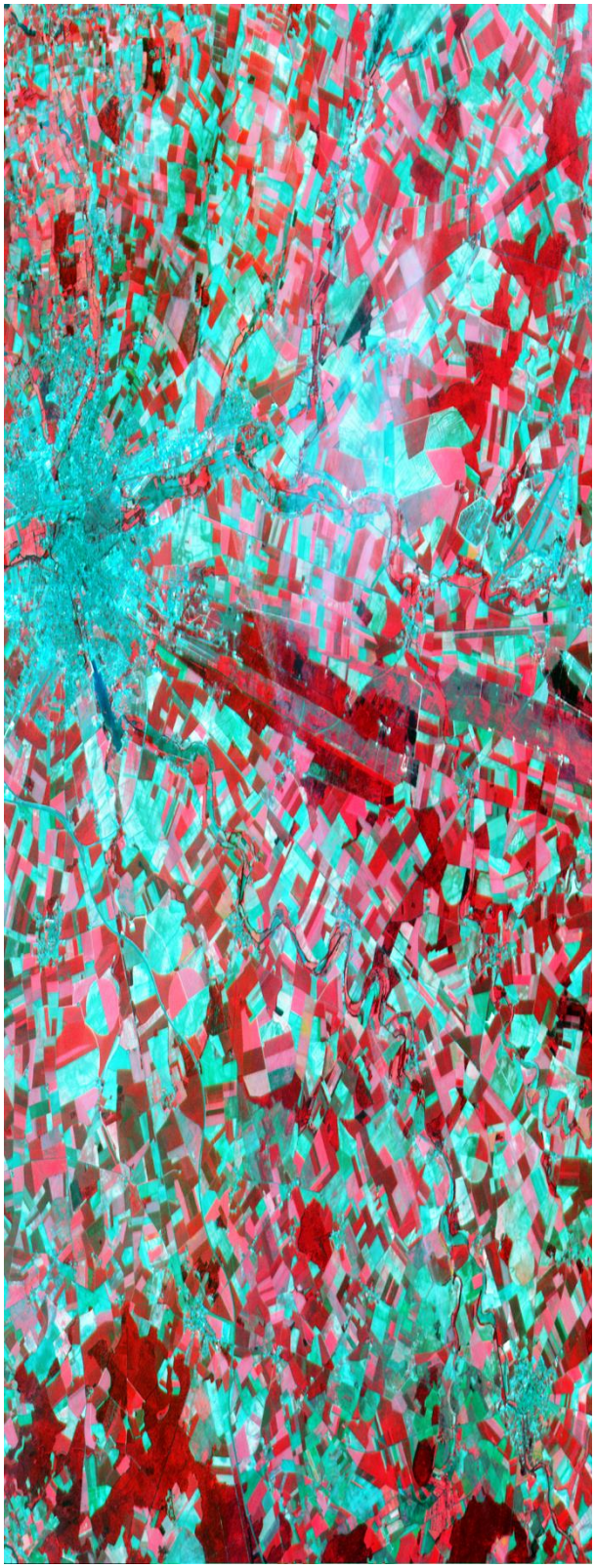


figure IV.22 : Composition colorée des images Spot-2

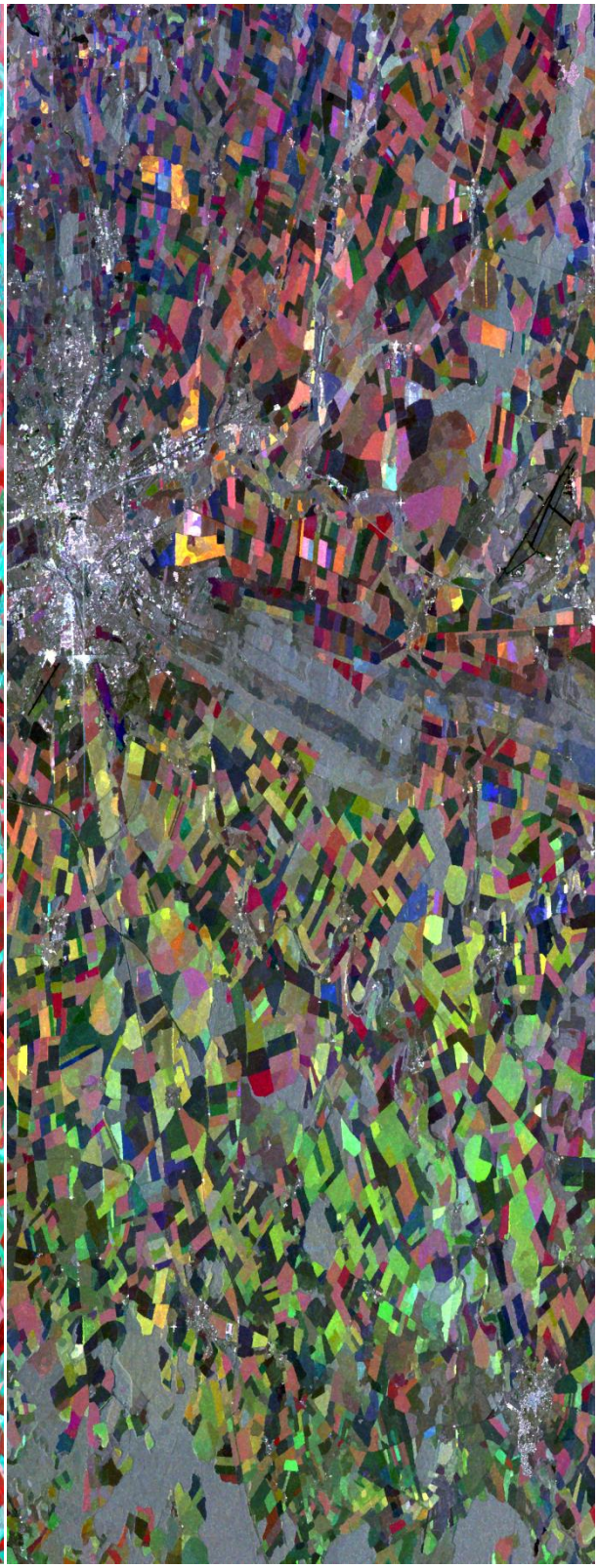


figure IV.23 : Composition colorée des images ERS filtrées

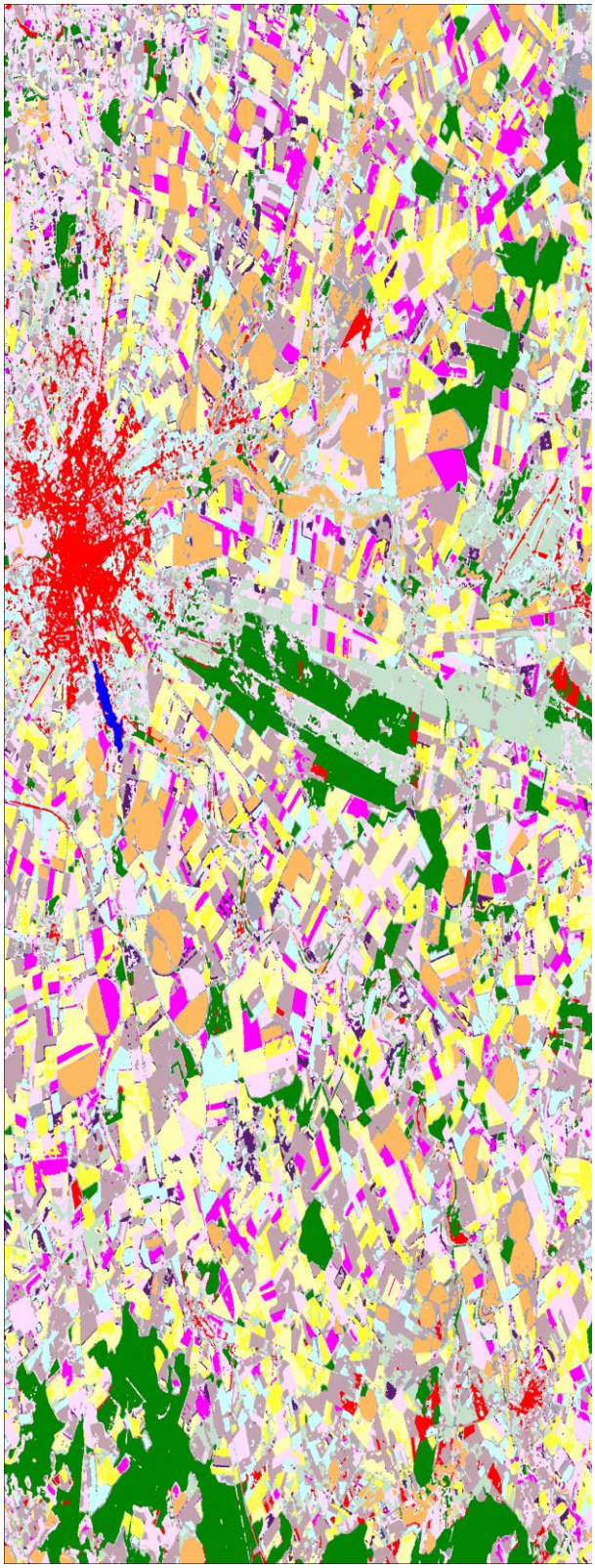


figure IV.24 : Classification des données Spot par classification ICM

figure IV.25 : Classification des images Radarsat non filtrées par classification ICM

figure IV.26 : Fusion par vote majoritaire des classifications Spot et Radarsat

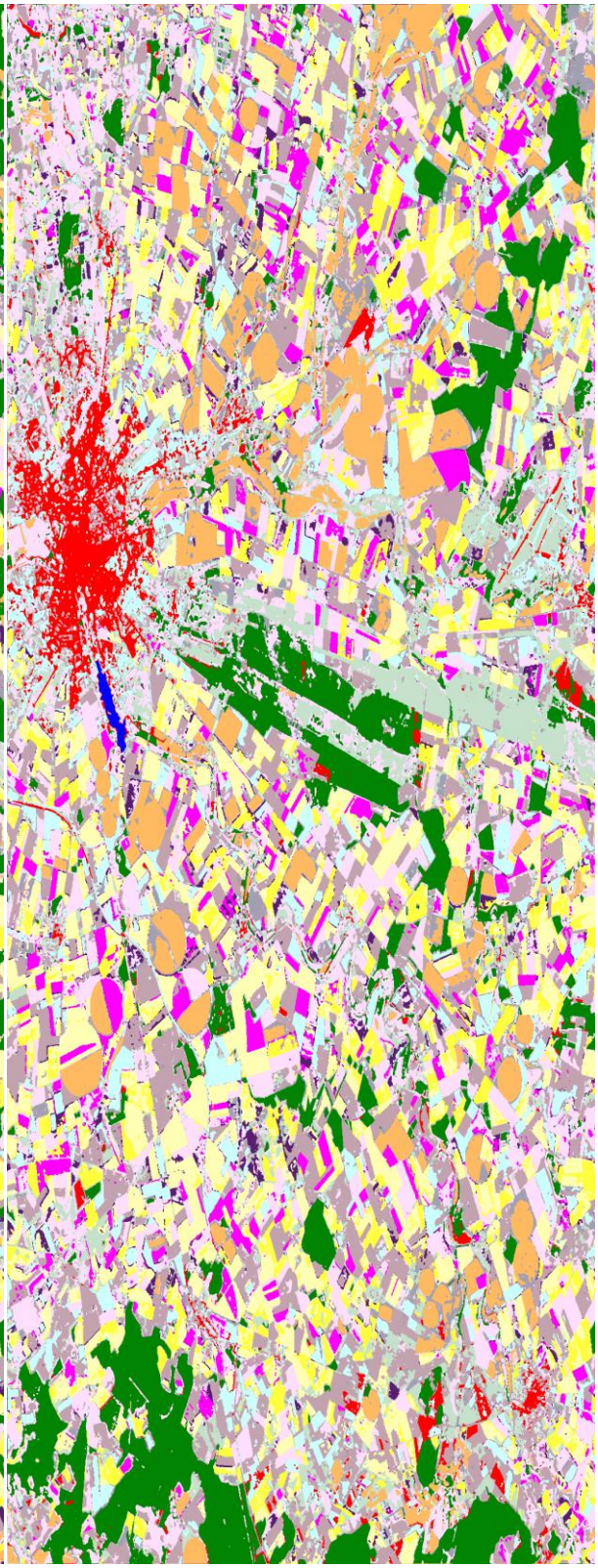


figure IV.27 : Fusion par vote majoritaire pondéré des classifications Spot et Radarsat

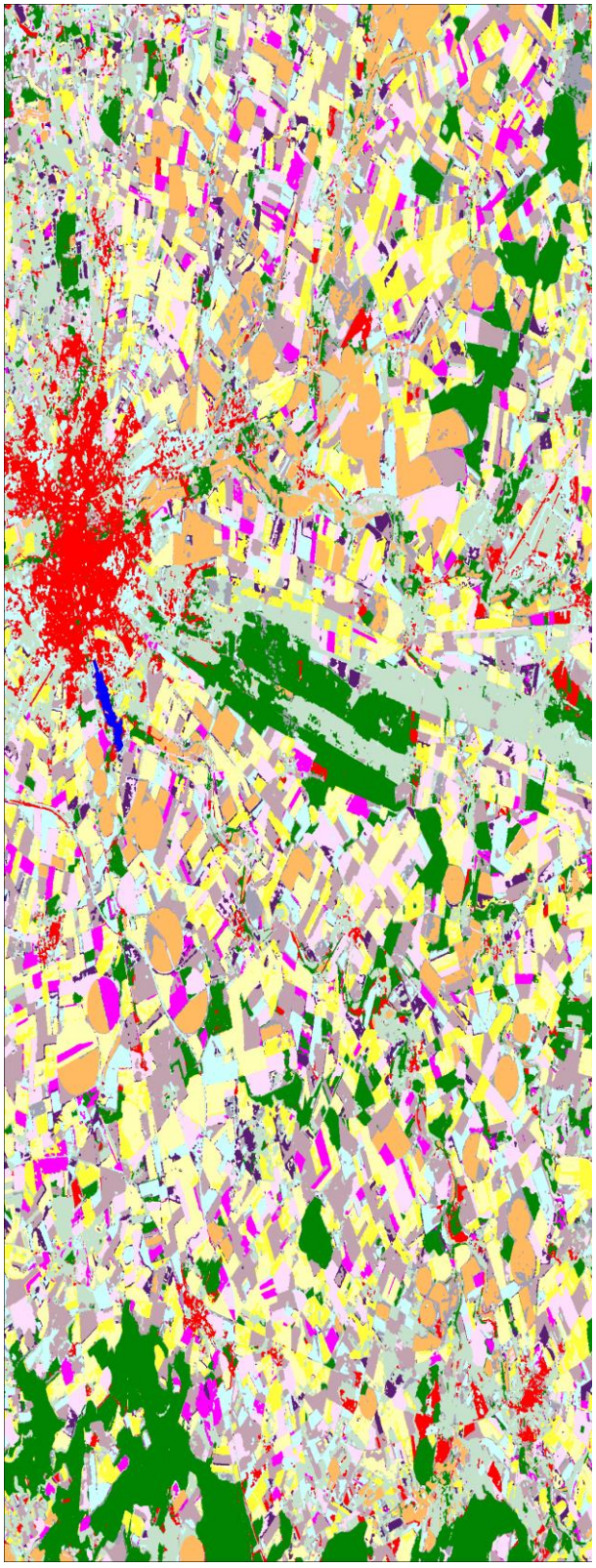


figure IV.28 : Fusion par minimisation de confusion des classifications Spot et Radarsat

figure IV.29 : Echantillons de vérification

5.4. La fusion des meilleurs jeux de données temporelles, spectrales et classifieurs, application du système SCF

Dans le cadre du système de Sélection-Classification-Fusion (SCF), plusieurs sélections de données ont été faites afin de maximiser la performance suivant plusieurs critères : une classification qui maximise la performance globale et pour chaque classe, une classification qui maximise l'indice de performance de cette classe. Cette sélection s'est effectuée sur les dates, les canaux spectraux et le classifieur appliqué (paragraphe 4.2). Afin d'exploiter cette complémentarité, nous avons appliqué les méthodes de fusion a posteriori introduites dans le paragraphe 3 sur ces applications de sélection.

Cette méthode est également appliquée au cours du chapitre 6 sur plusieurs années : données Spot 2/4/5 de 2006 à 2011 et Formosat-2 de 2006 à 2010. Les résultats sont présentés en annexe 1.

5.4.1. L'application aux données Formosat-2 de l'année 2009

Un jeu de données de référence globale et un jeu de données de référence pour chaque classe ont été sélectionnés. Leur fusion par méthode de minimisation de confusion permet l'augmentation des indices globaux (en moyenne 1 %, tableau IV.13). Les méthodes par vote majoritaire sont ici peu performantes du fait du grand nombre de données à fusionner. Les conclusions sont similaires à celles faites au paragraphe 5.2.1.

5.4.2. L'application aux données Spot-2/4/5 de l'année 2007

Le tableau IV.14 indique une faible amélioration des performances pour la méthode de fusion par minimisation de confusion, les résultats de la classification de référence étant déjà élevés. Les conclusions concernant la robustesse de cette méthode se confirment de nouveau pour cette application.

5.4.3. L'application aux données Landsat-5 de l'année 2010

L'amélioration des performances est plus importante que lors de la fusion des meilleurs jeux de données temporels (tableau IV.15). Le fait de sélectionner les canaux spectraux et le classifieur a donc permis de mettre en avant une meilleure complémentarité des résultats de classification. Nous retrouvons le problème de redondance d'erreurs pour les méthodes de fusion par vote majoritaire.

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**94,08**	93,92	93,92	**94,44**
AP	**90,95**	90,13	90,12	**91,52**
OA	**91,40**	90,95	90,94	**92,40**
Kappa	**0,91**	0,90	0,90	**0,92**
AOCI	**85,64**	84,78	84,78	**86,56**
F1	**0,92**	0,92	0,92	**0,93**

tableau IV.13 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Formosat-2 année 2009

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**93,10**	91,67	91,64	**93,32**
AP	**89,70**	85,56	85,55	89,30
OA	**89,71**	90,14	90,13	**91,14**
Kappa	**0,87**	0,88	0,88	**0,89**
AOCI	**77,55**	73,37	73,34	**77,60**
F1	**0,91**	0,89	0,89	**0,91**

tableau IV.14 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Spot 2/4/5 année 2007

Critères d'évaluation (en % sauf Kappa)	Référence globale	Fusion vote majoritaire	Fusion vote majoritaire pondéré	Fusion par minimisation de confusion
AA	**75,09**	73,51	74,84	**75,70**
AP	**54,03**	51,24	52,40	**55,21**
OA	**70,50**	72,66	72,92	**72,67**
Kappa	**0,67**	0,69	0,70	**0,70**
AOCI	**41,25**	39,70	41,21	**44,57**
F1	**0,63**	0,60	0,62	**0,64**

tableau IV.15 : Comparaison des indices de performance globale de la référence globale et des fusions pour l'application Landsat 5 année 2010

6. Conclusion

L'objectif de ce chapitre était d'introduire une méthode de fusion de données automatique, performante et robuste afin d'utiliser au mieux la diversité et la complémentarité des données multidimensionnelles pour la classification.

La méthode de fusion originale introduite lors de ce chapitre a montré de multiples avantages : elle permet la combinaison de résultats obtenus avec des lois différentes et des jeux de données variés (multi-temporelles et multi-sources). Ces méthodes n'ont pas de limites du nombre de classes et de tailles d'images, contrairement aux méthodes telles que DST ou DSmT. De plus, l'exécution de cette méthode de fusion originale est très rapide, de l'ordre de la minute.

Comme toute autre méthode de fusion, la méthode présentée n'a d'intérêt que lorsque les informations fournies sont complémentaires et donc non redondantes. Néanmoins elle permet la fusion automatique de tous types de données sans perte de performance, ce qui lui procure robustesse et performance.

Cette méthode de fusion est donc choisie dans le cadre du système automatique d'amélioration du résultat de classification appelé Sélection-Classification-Fusion.

Une perspective de ce travail serait l'intégration de la matrice de similitude en lieu et place de la matrice de confusion afin d'étendre cette méthode au monde des classifications non supervisées.

Chapitre V. La classification à partir d'une collection de données statistiques de valeurs radiométriques

Résumé :

Par manque de connaissance et absence de vérités terrain, on ne peut pas appliquer de méthode de classification supervisée. Dans ce cas, une classification non supervisée est utilisée, mais se pose alors le problème de son interprétation (vu au chapitre I). Pour faciliter la connaissance des thèmes de l'image étudiée, j'ai développé une méthode permettant la classification sans vérité terrain. Cette méthode de classification utilise des statistiques de valeurs radiométriques, interpolées par méthode de Krigeage à partir de données d'années différentes et/ou de zones différentes. Cette méthode a été testée sur deux applications distinctes, l'une pour la classification à l'aide de références d'une zone différente mais comparable, l'autre à l'aide de références d'années précédentes. Les résultats sont satisfaisants et indicatifs pour l'occupation du sol.

Mots clés : interpolation, krigeage, classification sans vérité terrain

"Les mathématiques ont des inventions très subtiles et qui peuvent beaucoup servir, tant à contenter les curieux qu'à faciliter tous les arts et à diminuer le travail des hommes."

René Descartes

Sommaire du chapitre V

1. Introduction

Lorsqu'un processus d'acquisition de vérités terrains est impossible (zone éloignée, année passée, année en cours), les seules méthodes de classifications utilisables sont les méthodes non supervisées. Ces classifications nécessitent une interprétation empirique dans la majorité des applications. Ce processus d'interprétation est généralement fastidieux. Il peut être purement visuel. Lorsque l'on dispose de valeurs radiométriques du type de thèmes présents dans la zone à étudier, une autre solution consiste à l'interprétation à l'aide des mesures de similitudes (voir Chapitre II).

Nous prendrons ici le problème inverse à la seconde solution : au lieu d'interpréter une classification non supervisée, nous dirigerons le processus de classification à partir de données radiométriques spectrales et temporelles connues et obtenues sur des années ou des zones différentes. Ce processus se rapprochera du mode supervisé.

Ce processus s'avère particulièrement adapté au traitement de larges zones d'étude qu'offrent les données des programmes Landsat et Copernicus. Dans le cas de Copernicus, le satellite "Sentinel-2" fournira gratuitement des images avec une couverture systématique de toutes les terres pour une répétitivité de 10 jours avec un seul satellite, et une répétitivité de 5 jours avec deux satellites, le tout à une résolution de 10 mètres.

2. Interpolation des données statistiques

2.1. Méthodes d'interpolation par Krigeage

A partir de données disposées de manière irrégulière dans l'espace temporel et spectral, la méthode d'interpolation la plus efficace est la méthode du Krigeage.

Le Krigeage porte le nom de son précurseur, l'ingénieur minier sud-africain D.G. Krige. Dans les années 1950, Krige (Krige, 1951) a développé une série de méthodes statistiques empiriques afin de déterminer la distribution spatiale de minerais à partir d'un ensemble de forages. La formalisation et la nomination de la méthode sont cependant dues à Matheron (Matheron, 1963). Depuis, le domaine de ses applications a largement été étendu, touchant notamment la météorologie, les sciences de l'environnement et l'électromagnétisme.

Commençons par définir cette méthode.

Définition : le *Krigeage* est une méthode d'estimation linéaire, sans biais, minimisant la variance d'estimation telle que calculée à l'aide du variogramme [Marcotte]. Cette méthode tient compte non seulement de la distance entre les données et le point à estimer, mais également de la distance des données entre elles.

Dans le cadre d'un modèle d'estimation stationnaire, il existe deux formes de Krigeage pour estimer la variable Y^* en fonction des observations Y_i et des λ_i, variables à optimiser.

- le Krigeage simple qui suppose que la moyenne μ du processus est connue

$$Y^* = \mu + \sum_{i=1}^{nb\_obs} \lambda_i (Y_i - \mu) \tag{V.1}$$

- le Krigeage ordinaire, le plus utilisé, qui suppose une moyenne inconnue

$$Y^* = \sum_{i=1}^{nb\_obs} \lambda_i Y_i \tag{V.2}$$

Considérons la méthode d'interpolation par Krigeage ordinaire.

Hypothèse : pour un estimateur sans biais, l'équation V.3 doit être vérifiée.

$$\sum_{i=1}^{nb\_obs} \lambda_i = 1 \tag{V.3}$$

Le problème à résoudre est donc un problème de minimisation sous contrainte d'une fonction quadratique, que l'on peut résoudre à l'aide du Lagrangien de l'équation V.4.

$$\begin{aligned} L(\lambda, \mu) &= \sigma^2 + 2\mu \left(\sum_{i=1}^{nb\_obs} \lambda_i - 1 \right) \\ &= \mathrm{var}(Y) + \sum_{i=1}^{nb\_obs} \sum_{j=1}^{nb\_obs} \lambda_i \lambda_j \mathrm{cov}(Y_i, Y_j) - 2 \sum_{i=1}^{nb\_obs} \lambda_i \mathrm{cov}(Y, Y_i) + 2\mu \sum_{i=1}^{nb\_obs} \lambda_i - 1 \end{aligned} \tag{V.4}$$

On obtient un système linéaire à résoudre qui contient n+1 équations à n+1 inconnues.

$$\sum_{j=1}^{nb\_obs} \lambda_j \mathrm{cov}(Y_i, Y_j) + \mu = \mathrm{cov}(Y, Y_i), \qquad \forall i \in [1, nb\_obs] \tag{V.5}$$

La variance du Krigeage s'écrit de la manière suivante.

$$\sigma_{ko}^2 = \mathrm{var}(Y) - \sum_{i=1}^{nb\_obs} \lambda_i \mathrm{cov}(Y, Y_i) - \mu \tag{V.6}$$

Cette variance de Krigeage ne dépend pas des valeurs observées, mais uniquement du variogramme et de la configuration des points servant à l'estimation par rapport aux points à estimer.

En conclusion, le Krigeage permet de minimiser la variance d'estimation théorique calculée à partir du variogramme. Le Krigeage est en moyenne plus juste que tout autre estimateur linéaire lorsque l'hypothèse de stationnarité est valide et que le modèle de variogramme est adapté, en particulier lorsque les données observées ne forment pas une grille régulière.

Les principales propriétés et caractéristiques associées au Krigeage sont fournies par Marcotte (Marcotte, 1991) :

- Le modèle est linéaire, sans biais et de variance minimale.
- L'interpolateur est exact : si l'on estime un point connu, on retrouve la valeur connue.
- L'interpolation présente un effet d'écran: les points les plus près reçoivent les poids les plus importants. Cet effet d'écran varie selon la configuration et selon le modèle de variogramme utilisé pour le Krigeage. Plus l'effet de pépite est important, moins il y a d'effet d'écran.
- L'interpolateur tient compte de la taille du champ à estimer et de la position des points entre eux.
- Le modèle est transitif. Si l'on observe en un point une valeur coïncidant avec la valeur « krigée » pour ce point, alors les valeurs « krigées » en d'autres points ne sont pas modifiées par l'inclusion de ce nouveau point dans les « krigeages ». Par contre les variances de Krigeage sont diminuées.

2.2. Evaluation de l'interpolation temporelle des classes

Afin d'évaluer l'estimation des nouvelles classes (moyennes, covariances), il convient de calculer les matrices de similitudes (Chapitre II paragraphe 3) et d'en extraire les indices de performance de l'interpolation.

3. Applications

3.1. Application à la classification de données d'années différentes et d'une emprise différente

Prenons l'exemple de la classification d'une zone de 1,2 millions d'hectares à l'ouest de Toulouse, couverte par des images Landsat-5 pour l'année 2011. Ne disposant pas de vérité terrain pour effectuer une classification supervisée, nous avons appliqué la méthode d'interpolation de statistiques. Ces statistiques ont été obtenues à partir de la zone Est de Toulouse (Tarn, Aude, Haute-Garonne) couverte par des images Landsat-5/7 de l'année 2010. A partir des vérités terrain de cette zone connue, nous avons interpolé les statistiques de moyenne et de covariances radiométriques de 2010 sur l'axe temporel de 2011. Nous disposons de 11 dates en 2010 (1 Février ; 24 et 25 Mai ; 11, 19, 20 Juillet ; 20, 21 et 29 Août ; 22 et 29 Septembre) et de 4 dates en 2011 (24 Mars, 9 Avril, 11 Mai et 2 Octobre).

Une acquisition récente de vérités terrain pour la zone à classer nous a permis d'évaluer précisément la qualité de l'interpolation et de la classification. L'interpolation est assez performante avec une racine carrée d'erreurs moyennes des classes de 5,14 et une similitude moyenne des classes interpolées de 91,21%. Le tableau V.1 permet la comparaison des deux méthodes, à l'aide des indices de performance obtenus à partir de vérités terrain pour la validation. Nous observons une différence moyenne de 15% pour les scores des indices de performance globaux. Certaines classes sont très bien classées par la nouvelle méthode et obtiennent des scores proches de ceux de la classification supervisée (Blé, Eau, Maïs,…).

La zone extraite présentée est la commune de Lannemezan dans le département des Haute Pyrénées. La figure V.1 illustre le résultat de la classification à partir des statistiques interpolées, et la figure V.2 le résultat de la classification à partir des vérités terrains.

Indices globaux (en % sauf Kappa et F1)	Classification à partir des vérités terrains	Classification à partir des statistiques interpolées
AA	76,76	63,97
AP	43,66	36,99
OA	77,54	64,68
Kappa	0,60	0,52
AOCI	30,66	28,55
F1	0,552	0,47
Extrait des indices par classe Valeur OCI de la classe :		
Bâti dense	73,19	67,56
Blé	97,77	90,25
Bocages	54,43	50,24
Colza	88,56	81,74
Eau	96,32	94,43
Feuillus	88,75	81,92
Jachères	60,06	50,10
Maïs	89,70	82,80
Prairie	71,07	60,60
Tournesol	84,89	78,36

tableau V.1 : Comparaison des indices de performance de la méthode avec et sans vérités terrain de la zone à étudier

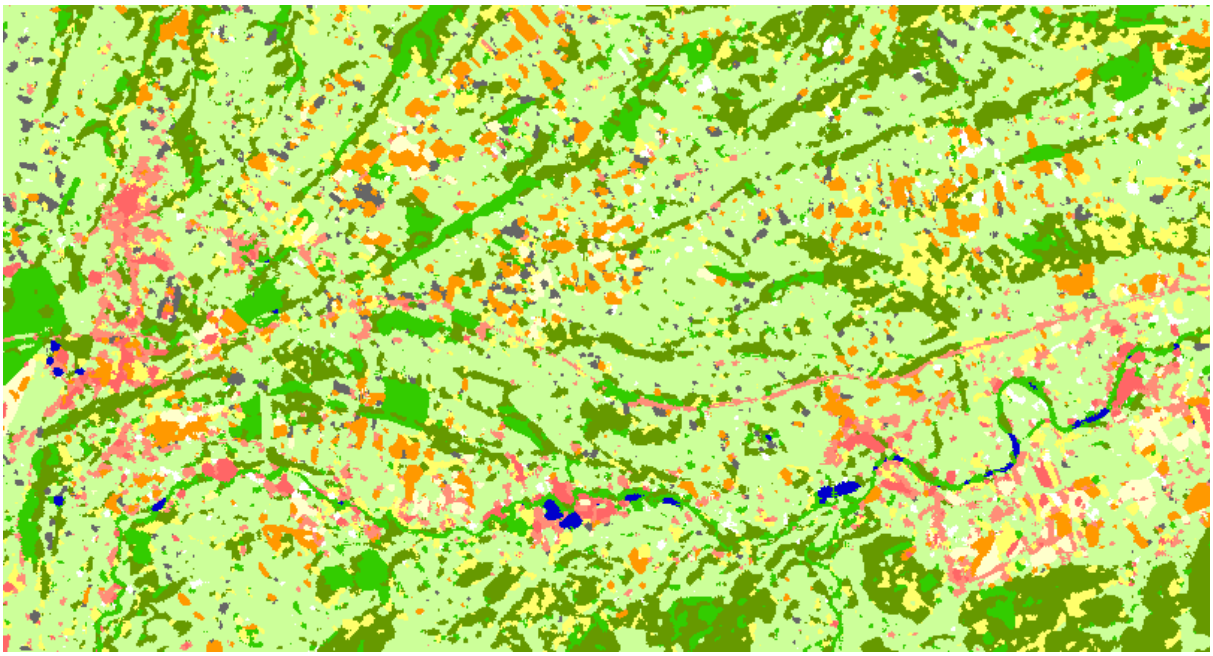


figure V.1 : Extrait de la classification des données Landsat-5 année 2011 à partir des statistiques interpolées, zone de la commune de Lannemezan

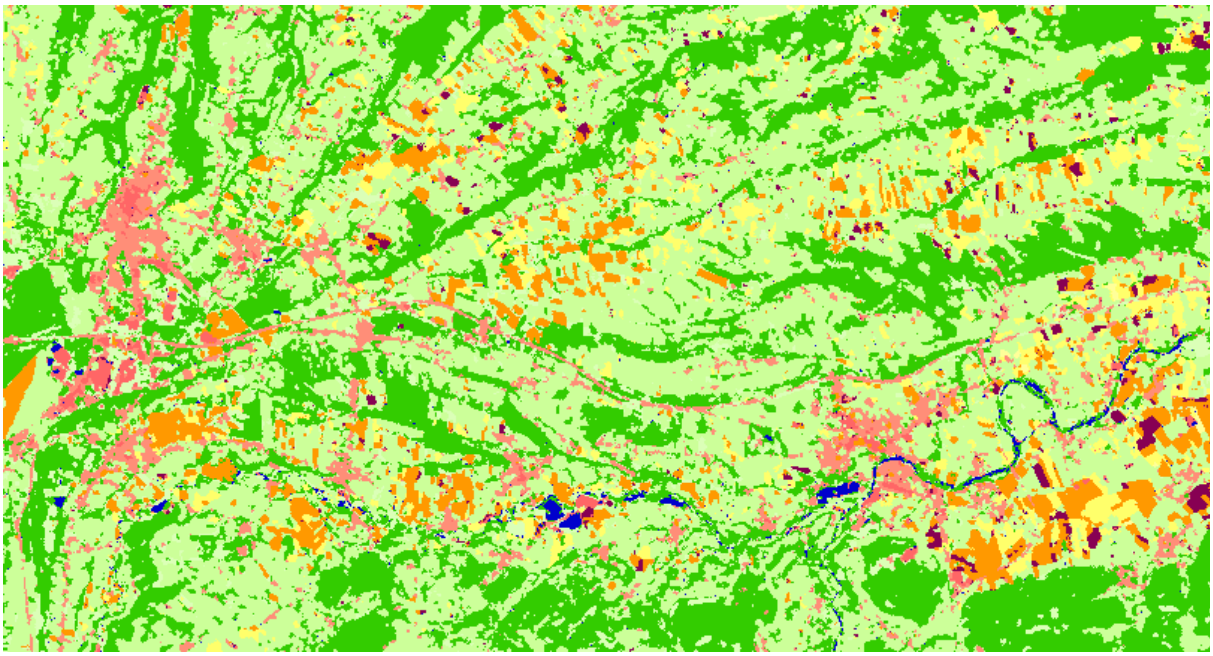


figure V.2 : Extrait de la classification des données Landsat-5 année 2011 à partir des vérités terrain, zone de la commune de Lannemezan

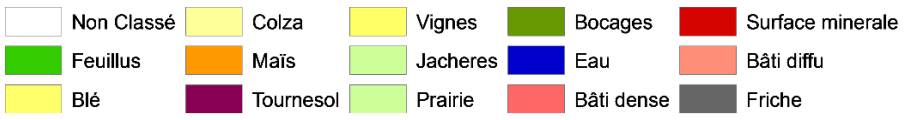


figure V.3 : Légende

3.2. Application à la classification de données d'une année différente

Disposant une série temporelle de 5 années de données Formosat-2 couvrant notre zone d'étude du sud-ouest toulousain, nous nous proposons de tester la méthode d'interpolation et de classification en utilisant les statistiques des années 2006 à 2009 qui serviront à l'apprentissage pour l'interpolation des statistiques sur l'année 2010.

L'interpolation des statistiques de moyennes, écarts-types et extrema pour la classe « Blé » est illustrée en figure V.4. Mis à part certains points extrêmes qui représentent des fluctuations annuelles des stades phénologiques, l'interpolation est correcte avec une racine carrée d'erreurs moyennes des classes de 4,34 et une similitude moyenne des classes interpolées de 96,32%. L'interpolation de la matrice de covariance est plus difficile car elle nécessite un nombre plus important de dates pour une estimation correcte du fait de la bi-dimensionnalité de la covariance. La figure V.5 illustre l'interpolation de la matrice de covariance pour le canal PIR de la classe Feuillus.

En comparant la classification effectuée avec les statistiques interpolées (figure V.6) et celle effectuée avec des vérités terrain acquises sur la zone à classer (figure V.7), nous observons une différence au niveau de certaines cultures (Colza/Blé et Maïs/Tournesol, voir encadrés rouge des figures V.6 et V.7). Nous observons également une différence au niveau des performances globales, de l'ordre de 20% (tableau V.2). Ces différences sont dues aux erreurs d'interpolation des classes de cultures du fait des fluctuations annuelles des stades phénologiques, ce qui peut entrainer des confusions entre cultures d'une même saison.

Indices globaux (en % sauf Kappa et F1)	Classification à partir des vérités terrains	Classification à partir des statistiques interpolées
AA	98,10	80,14
AP	98,00	72,10
OA	98,45	75,22
Kappa	0,97	0,73
AOCI	93,92	69,19
F1	0,98	0,76
Valeur OCI de la classe :		
Bâti (2)	91,30	72,01
Blé dur	94,37	47,70
Chanvre	99,63	75,98
Colza	97,57	53,65
Eau (3)	100,00	76,88
Eucalyptus	100,00	85,14
Feuillus	99,96	86,02
Jachère	64,47	41,21
Maïs (2)	99,84	92,35
Orge	87,85	61,32
Peupliers	100,00	77,66
Prairie	74,97	45,39
Résineux	100,00	84,13
Soja	99,83	48,05
Sorgho	99,47	84,36
Tournesol	93,49	75,21

tableau V.2 : Comparaison des indices de performance de la méthode avec et sans référence spatiale de la zone

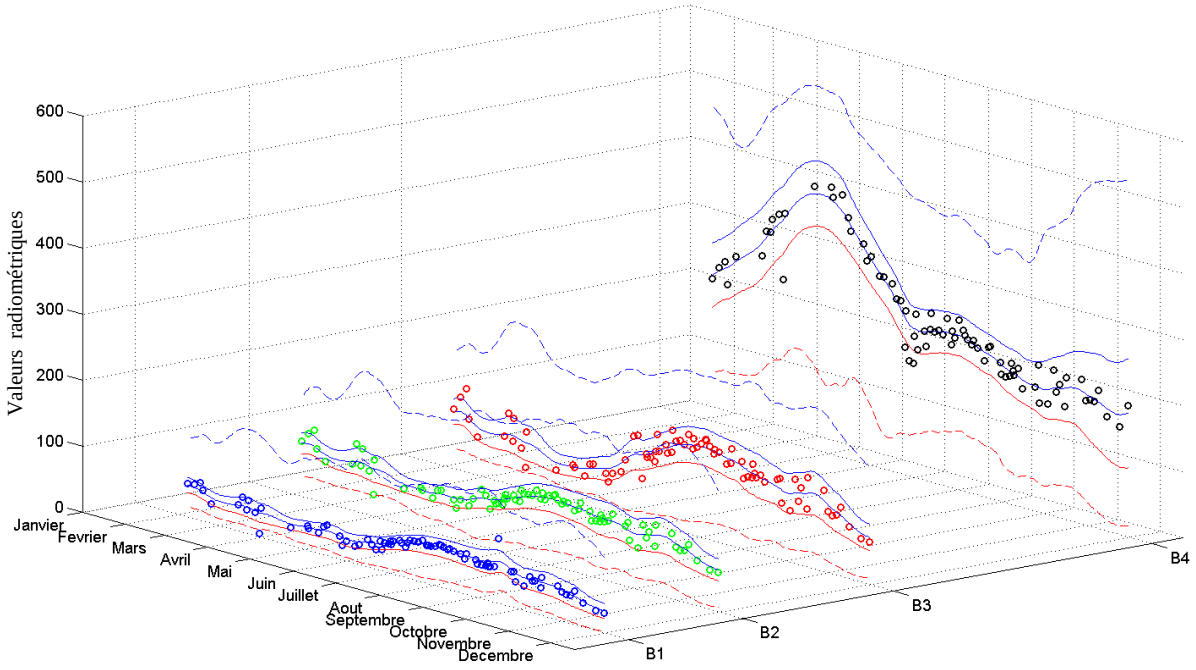


figure V.4 : Interpolation de la moyenne-écart-type-minimum-maximum des valeurs radiométriques de la classe Blé pour les canaux B1 : Bleu, B2 : Vert, B3 : Rouge et B4 : Proche IR à partir de 4 années : 2006 à 2009. Pour chaque canal, le trait central en noir représente la moyenne interpolée, les traits pleins bleu et rouge les écarts-types relatifs à la moyenne et en traits discontinus, les extrema interpolés.

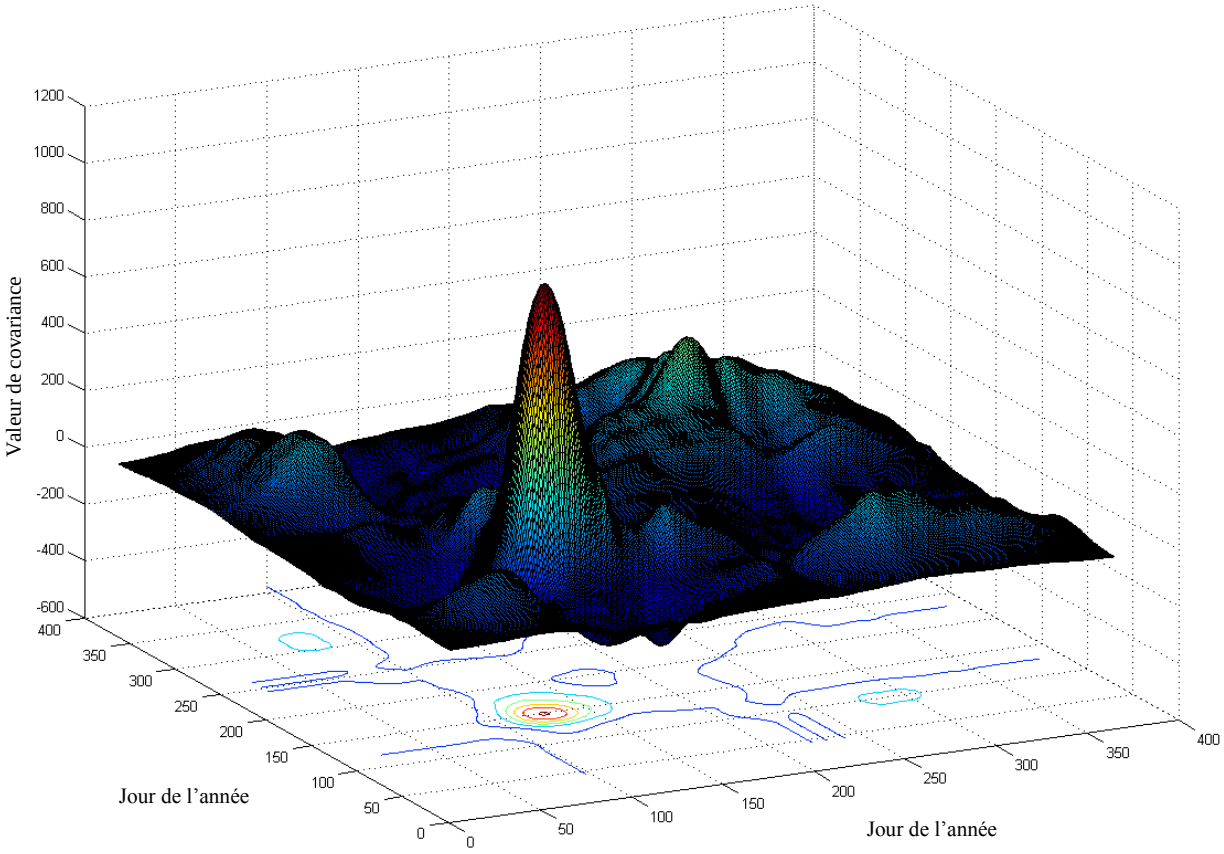


figure V.5 : Interpolation de la matrice de covariance pour le canal PIR de la classe Feuillus. Les échelles sont en jour de l'année et en valeur de covariance.

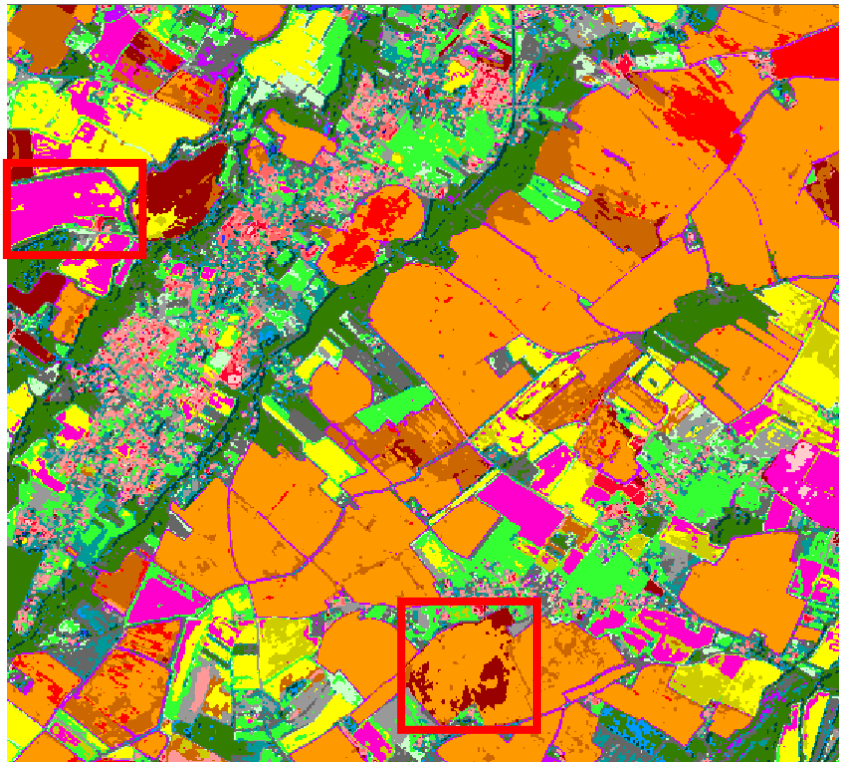


figure V.6 : Extrait de la classification des données Formosat-2 année 2010 à partir des statistiques interpolées

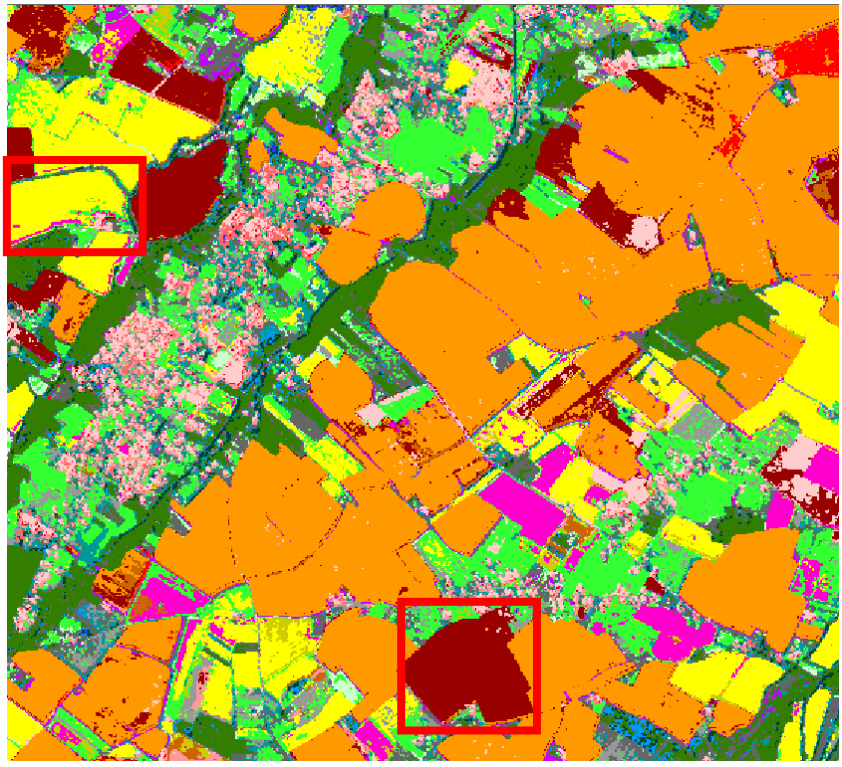


figure V.7 : Extrait de la classification des données Formosat-2 année 2010 à partir des vérités terrain

4. Conclusion

L'objectif de ce chapitre était d'obtenir l'occupation du sol sans vérité terrain de la zone à étudier, mais en disposant d'une collection de références statistiques sur d'autres zones ou d'autres années.

Concernant la méthodologie développée et appliquée, l'interpolation par Krigeage a fourni de très bons résultats. Cette méthode a permis d'interpoler les différentes statistiques utilisables par le système de classification. Comme toute autre méthode d'interpolation, elle nécessite un minimum de points significatifs d'observation, ce qui est difficile à obtenir pour l'interpolation des statistiques bidimensionnelles telles que les matrices de covariance.

Concernant la classification à partir de ces statistiques interpolées, nous remarquons que les performances sont inferieures à celles des méthodes supervisées avec vérités terrain. Cependant, les résultats obtenus fournissent une indication, parfois indispensable, de l'occupation du sol sans vérité terrain. Des différences se situent essentiellement au niveau des classes dont le cycle phénologique varie d'une année à l'autre, ce qui engendre des confusions entre classes de cultures très proches en matière de pratique culturale.

Les résultats sont donc satisfaisants et fournissent une indication pour des modèles de prévision. Les résultats de l'application Landsat année 2011 ont notamment servi au projet BLLAST (Boundary Layer Late Afternoon and Sunset Turbulence , site de Lannemezan) dont l'objectif est l'étude de la diffusion de la vapeur d'eau. Ce type de modèle nécessite des cartes d'occupations du sol de l'année en cours.

Chapitre VI. Analyse de changements annuels et modélisation de flux de carbone

Résumé : Un changement de mode d'occupation du sol affecte directement les interactions entre la biosphère et l'atmosphère. A partir d'images de télédétection et de la création de cartes des changements de l'occupation du sol, il est ainsi possible de quantifier les flux de carbone entre le sol et l'atmosphère. J'ai utilisé les méthodes présentées au cours des chapitres précédents et nous avons développé un modèle d'estimation de ces flux afin d'estimer les flux de carbone. Cette méthodologie a été appliquée à deux jeux de données différents et validée par croisement de résultats sur leur zone commune. Un bilan de carbone a ainsi pu être calculé à partir de l'estimation des flux.

Mots clés : changement d'occupation du sol, rotation de cultures, estimation de flux de carbone, SCF

Sommaire du chapitre VI

1. Introduction

Grace au nouveau système de classification et à l'amélioration des cartes d'occupation du sol, de nouvelles applications sont aujourd'hui réalisables. Une première application rendue possible par les nouveaux outils présentés aux chapitres précédents est l'estimation des flux de carbone à l'aide d'images de télédétection à haute résolution (Ducrot, Masse, Ceschia, Marais-Sicre, & Krystof, 2010). Elle permet d'appliquer les modèles d'estimation de flux de carbone sur des zones plus vastes.

Les objectifs du chantier Sud-Ouest du CESBIO (Ceschia) visent à comprendre et prévoir l'influence de l'homme et du climat sur les ressources en eau et les écosystèmes. Les activités humaines, parmi lesquelles la production par combustion de carbone fossile, la déforestation et l'agriculture, sont fortement liées à d'importantes concentrations de gaz à effet de serre. Dans le cas du CO_2, une alternative consiste à accroître le rôle de « puits » temporaires que joue la végétation, et principalement en augmentant la durée de stockage du carbone de cette biomasse. Un changement de mode d'occupation du sol affecte directement les interactions entre la biosphère et l'atmosphère. Une action anthropique par le reboisement, le changement d'utilisation des terres et de bonnes pratiques culturales peuvent augmenter la séquestration de carbone dans la biomasse et dans les sols pour des durées de plusieurs décennies, ce qui peut constituer un apport non négligeable à la lutte contre l'effet de serre.

Plusieurs modèles ont été développés pour reproduire les échanges de quantité de carbone entre le sol et l'atmosphère ainsi que pour illustrer différents contextes de changement d'occupation du sol. L'étude de l'évolution du stockage et du déstockage de carbone dans le sol est complexe. Cette évolution est le résultat de réactions chimiques et physiques à plus ou moins long terme intégrant le contexte de la zone observée et les divers échanges avec l'environnement.

2. La rotation des classes, les enjeux

Après un changement d'occupation des terres, les vitesses de stockage dans le sol varient au cours du temps. Elles sont très fortes après un changement d'occupation du sol puis se stabilisent au bout d'une vingtaine d'année après avoir atteint un seuil de saturation du sol. Ainsi dans le cas de forêts et sur des périodes de temps choisies, nous supposons que la teneur en carbone organique du sol est stable.

Le premier enjeu consistera à identifier si une parcelle est cultivée ou s'il y a eu un changement d'occupation du sol (prairie vers cultures, …) et à savoir si le système est stable ou non (changement de grandes classes) afin d'y appliquer un modèle de changement. Ce modèle permettrait de décrire l'évolution du carbone du sol lors d'un changement de grandes classes en prenant en compte la date de changement.

Le second enjeu concerne l'enchainement des cultures et la capacité d'identification de l'enchainement de différentes cultures pour chaque parcelle.

La connaissance de la rotation des cultures nous permet d'identifier à partir des statistiques de rendements (Agreste, base de données statistiques agricoles) la quantité de carbone exporté à la récolte, la quantité de biomasse produite sur la parcelle pour une année donnée, et d'estimer les pertes liées à la décomposition des résidus de cultures des années précédentes.

Le troisième enjeu porte sur la durée de décomposition des résidus de cultures qui peut être plus ou moins longue. Les conséquences de cette décomposition tardive ont un impact sur l'année observée et se répercutent sur l'année suivante. Donc, l'étude des rotations de cultures sur plusieurs années successives permet de prendre en considération ces décalages possibles. Dans cette étude, nous avons considéré que l'essentiel des résidus se décomposait l'année suivante.

3. Le modèle de flux de carbone

A partir de cartes de changement d'utilisation des terres pour les classes principales (forêts, terres cultivées, pâturages), nous souhaitons estimer les changements du Carbone Organique du Sol (COS) en appliquant un modèle basé sur les observations régionales (Arrouays et al., 2012).

Pour réaliser cette carte, nous avons choisi un modèle permettant d'estimer les flux de carbone en fonction des changements de classes (ou du contexte de plantations/constructions, etc…).

3.1. Les modèles existants

Les modèles exponentiels existant (Hénin & Dupuis, 1945) permettent d'avoir une vision annuelle de l'évolution de carbone et se basent sur un bilan entre le taux d'humidification de la matière organique, le taux de minéralisation de la matière organique humidifiée du sol et la prise en compte du temps écoulé depuis le dernier changement d'occupation.

3.1.1. Le modèle de Hénin et Dupuis

Le modèle de Hénin et Dupuis répartit le carbone des sols en un seul compartiment. Trois paramètres décrivent le fonctionnement :

- la quantité d'apport de carbone au sol m (en $tC.ha^{-1}$);
- la proportion K1 de cet apport qui entre dans le compartiment sol, le reste étant supposé être minéralisé instantanément ; K1 est dit coefficient isohumique.
- le coefficient de destruction de la matière organique K2 (en an^{-1}). La minéralisation suit une loi d'ordre 1.

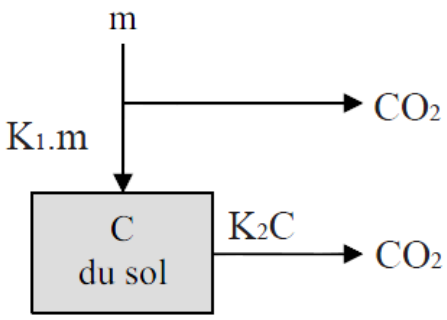


figure VI.1 : Modèle de Henin-Dupuis

En régime stationnaire, la quantité de carbone s'établit à $C_s = \frac{K_1 \times m}{K_2}$

En régime transitoire, le carbone évolue selon la formule de l'équation VI.1.

$$C(t) = C_s - (C_s - C_0)e^{-K_2 \times t}, \text{avec } C_0 \text{ carbone initial} \qquad \text{(VI.1)}$$

Les valeurs des paramètres sont décrites par Rémy et Marin-Laflèche (Rémy & Marin-Laflèche, 1976). Des tables de valeurs indiquent par type de culture les restitutions moyennes, aériennes ou souterraines (m), le coefficient isohumique (K1) propre à chaque production végétale ou à chaque amendement (fumiers, etc.). Le coefficient K2 est déterminé par la teneur en argile et en calcaire du sol.

Notation	Sémantique	Définition
BC	bilan carbone	variation du stock de carbone
NPP_tot	Net Primary Production total	Taux carbone pour la formation de biomasse totale
NPP_aerien	Net Primary Production aérien	Taux carbone pour la formation de biomasse aérienne
NPP_racinaire	Net Primary Production racinaire	Taux carbone pour la formation de biomasse racinaire
RS	Root/Shoot	rapport en fonction de NNP_aerien et NPP_racinaire
HR	Harvest Résidu	Pourcentage de matière NON sèche
HI	Harvest Index	rapport entre la quantité de matière exportée et le NPP_aerien
Rend	rendement	Quantité produite (source : Agreste)
Ten	teneur	Teneur en carbone par matière sèche
StockCulture	Stock culture	Quantité de carbone encore présente dans le sol de l'année précédente
CC	Constante Correctrice	Constante permettant de corriger l'estimation des exports en carbone

tableau VI.1 : Notations, sémantiques et définitions

3.1.2. Modèle développé au CesBIO par Eric Ceschia

La quantité de carbone stockée pendant une période dans un sol dépend fortement de ce qui repose dessus. Cela peut être des sols nus, mais aussi des sols pour la culture, ou des sols de forêts. Ainsi, une meilleure compréhension des échanges entre l'air, les plantes/arbres/constructions et le sol permet de développer un modèle nouveau et différent des autres approches, plus proche des réalités des aléas des sols étudiés.

Dans le cas où on reste sur les cultures, nous avons développé une méthode plus précise et spécifique aux cultures. Cette approche est fondée sur le NPP, des indices de récoltes (harvest index), les exports de carbone sous diverses formes, les résidus et la respiration de ces cultures.

Le calcul des paramètres non observables se fait à partir de données disponibles (par exemple la base de données Agreste).

Soit **a** une classe et **n** une année.

$$\begin{aligned} \text{NPP\_racinaire}(a,n) &= \text{RS}(a) * \text{NPP\_aerien}(a,n) \\ \text{HI}(a,n) &= \text{export}/\text{NPP\_aerien} \\ \text{export}(a,n) &= \text{rend}(a,n) * (1-\text{HR}) * \text{ten}(a,n) \end{aligned}$$

Ainsi on en déduit le NPP_tot(a,n) :

$$\begin{aligned} \text{NPP\_tot}(a,n) &= \text{NPP\_aerien}(a,n) + \text{NPP\_racine}(a,n) \\ &= \text{NPP\_aerien} \times (1 + \text{R}/\text{S}) \\ &= (\text{export}/\text{HI}) \times (1 + \text{R}/\text{S}) \\ &= \frac{\text{rend}(a,n) \times (1-\text{HR}) \times \text{ten}(a,n) + \text{CC}}{\text{HI}} \times \left(1 + \frac{\text{R}}{\text{S}}\right) \end{aligned}$$

Ainsi la formule de la variation du stock de Carbone du sol est :

$$\begin{aligned} \text{BC}(a,n) &= \text{NPP}_{\text{tot}}(a,n) - \text{Respiration}(a,n-1) - \text{export}(a,n) \\ \text{Respiration}(a,n-1) &= \text{taux} * (\text{NPP\_tot}(a,n-1) - \text{export}(a,n)) \end{aligned}$$

'taux' est un paramètre représentant le rapport entre la quantité de carbone libérée par la respiration et celle stockée dans le sol les années précédentes.

Ce modèle n'a pas fait l'objet d'une publication pour le moment.

3.2. Le modèle choisi

Nous avons choisi d'inclure deux modèles pour illustrer la fluctuation de carbone dans le sol sur les zones étudiées. Dans le cas des changements de classes principales, forêts et pâturages, le bilan carbone est estimé d'après les modèles exponentiels. Dans le cas des cultures, l'approche faisant intervenir la NPP a été utilisée (modèle CesBIO).

Le calcul du bilan carbone par la première approche n'influe pas sur le calcul par la deuxième, ce qui explique notre approche combinée.

La première prend en considération les transitions de classes principales alors que la seconde est strictement axée sur les classes de cultures.

La méthodologie générale est résumée à la figure VI.2.

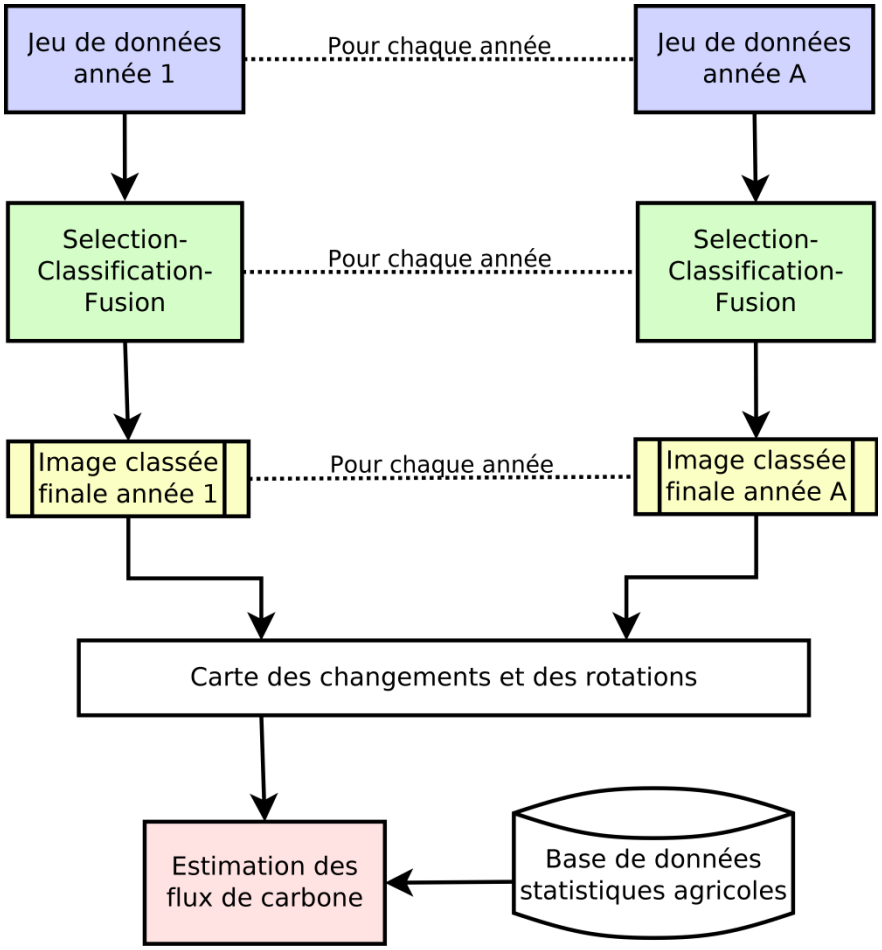


figure VI.2 : Méthodologie proposée pour l'estimation des flux de carbone à partir de la carte de rotation et des statistiques agricoles

4. Applications : carte des changements et estimation des flux de Carbone

4.1. A partir d'images Formosat-2 des années 2006 à 2009

La zone d'étude est l'intersection des zones étudiées précédemment dans les années 2006, 2007, 2008 et 2009. Celle-ci a une surface totale de 50802 Ha.

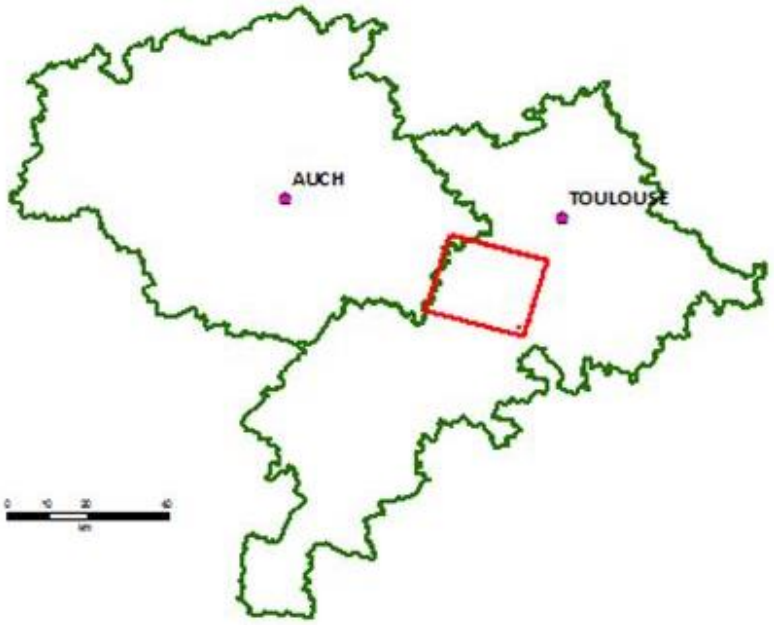


figure VI.3 : Zone d'étude, emprise commune des différentes images Formosat-2 utilisées

Dans un premier temps, le système de classification automatique SCF a été appliqué sur les données de chaque année. Les classes étudiées et les performances des classifications sont présentées en annexe 1.3. Un regroupement de certaines classes a été opéré suivant la nomenclature suivante : **Bois** (Feuillus, Résineux, Eucalyptus, Peupliers), **Eau** (Lac, Gravière), **Bâti et surfaces minérales (**Bâti dense, Bâti industriel, Bâti lâche, Réseau routier, sols nus) et **Surfaces enherbées** (Prairie temporaire, Pré/prairie permanente, Jachère, Friche). Les classes de cultures sont restées indépendantes : **Blé, Colza, Orge, Maïs, Tournesol, Sorgho, Soja, Pois, Chanvre** (non regroupés).

Le tableau VI.2 montre que le Maïs est la seule culture qui n'a pas beaucoup changée. Elle est la plus représentative des monocultures dans la zone d'étude. Les autres monocultures sont peu présentes : Blé, Soja, Tournesol (tableau VI.3).

La rotation Blé-Tournesol est la rotation la plus importante avec un pourcentage de représentativité dans la zone de 8%. La rotation annuelle entre le Blé et le Tournesol atteint un pourcentage du 5% par rapport au total de la zone étudiée (tableau VI.4).

A partir des rotations de cultures (figure VI.4 à VI.5) et des statistiques agricoles de la zone, on calcule le bilan carbone à partir du modèle présenté précédemment (tableau VI.5). Sur la zone étudiée de 50000 ha, ce bilan est de 51409 tonnes de Carbone stocké en 4 ans soit 0,257 tonne de carbone stocké dans le sol par hectare et par an.

La figure VI.10 illustre le bilan carbone par pixels de l'image.

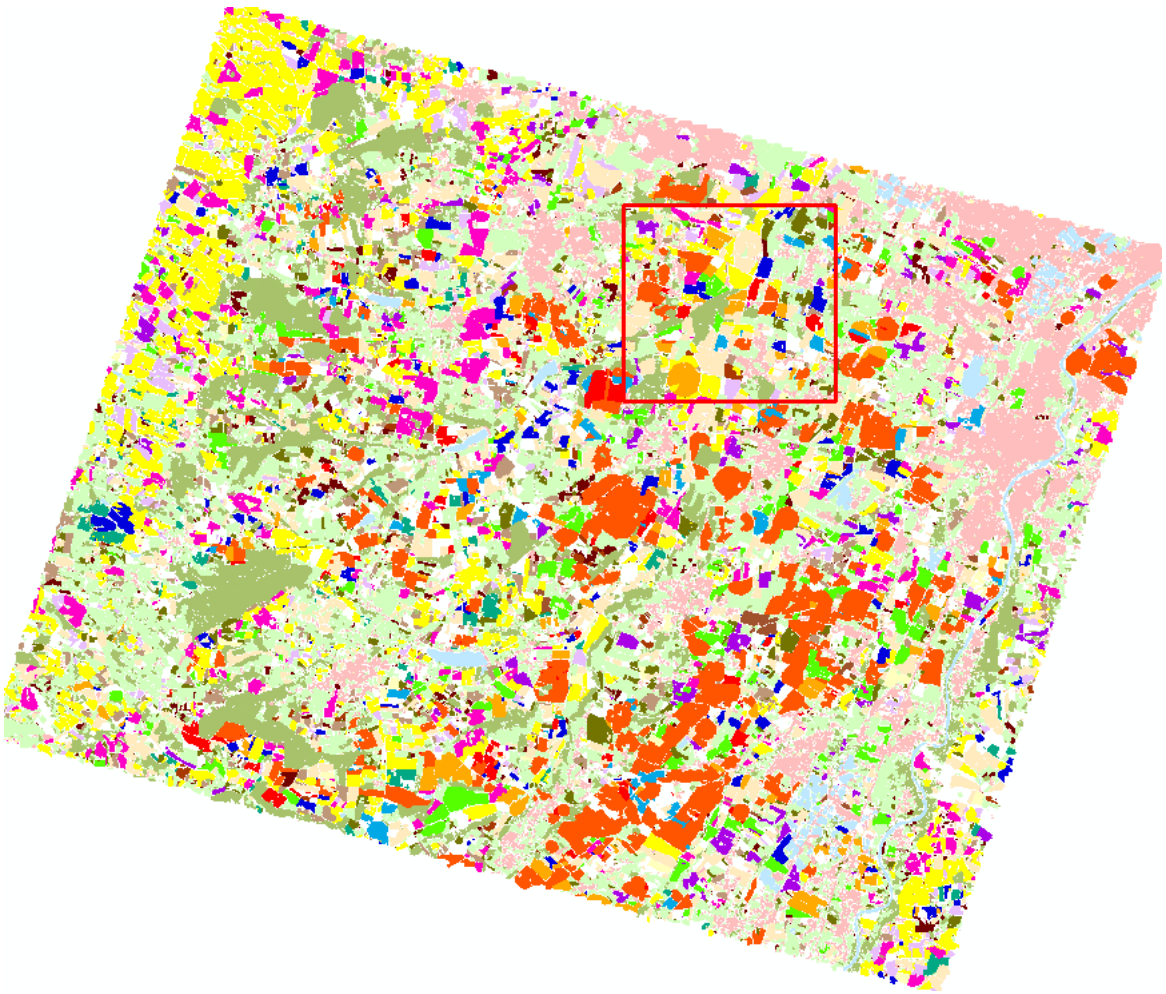


figure VI.4 : Image des rotations majoritaires obtenues à partir des images Formosat-2 sur les années 2006 à 2010

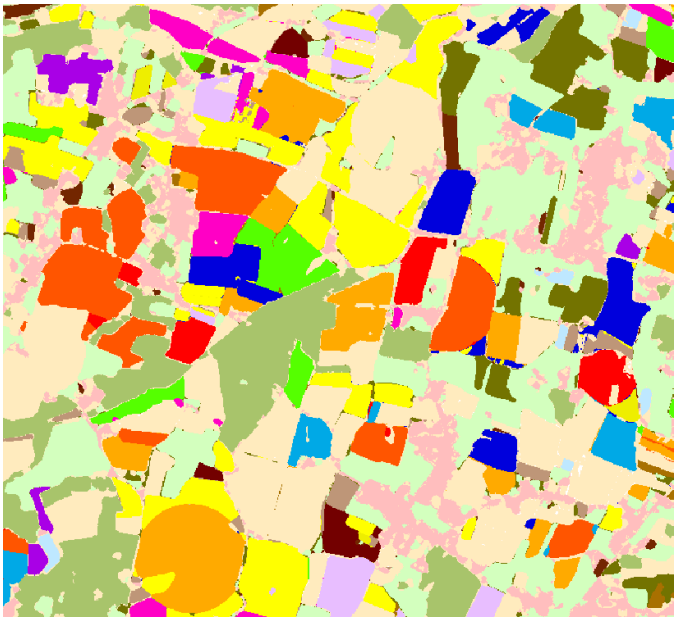


figure VI.5 : Extrait de la figure VI.4 correspondant à l'encadré rouge

figure VI.6 : Légende des rotations majoritaires

		Image de 2009										
		Bois	Blé	Colza	Orge	Maïs	Tournesol	Sorgho	Soja	Surf. enherbées	Eau	Bâti/Surf min
Image de 2006	Bois	**89,16**	0,03	0,02	0,06	0,04	0,09	0,07	0,00	10,22	0,03	0,25
	Blé	0,12	**19,31**	11,53	6,76	5,63	34,23	5,19	1,17	13,36	0,00	1,08
	Colza	0,12	38,71	**13,13**	5,42	1,03	22,54	4,85	0,43	8,02	0,00	1,01
	Orge	0,14	20,30	12,03	**19,71**	2,66	18,67	3,40	0,03	19,28	0,05	1,69
	Maïs	0,10	4,75	1,51	2,29	**78,02**	5,22	2,29	1,93	3,34	0,00	0,10
	Tournesol	0,25	35,74	6,18	3,86	6,03	**26,77**	5,50	0,32	12,12	0,00	1,77
	Sorgho	0,11	24,55	3,76	0,51	22,73	20,47	**14,97**	0,23	12,38	0,00	0,09
	Soja	0,07	18,28	6,31	2,16	28,77	11,11	3,96	**22,26**	5,68	0,00	0,15
	Surf. enherbées	6,60	3,03	1,44	1,43	1,00	3,52	0,91	0,07	**77,27**	0,02	4,54
	Eau	1,04	0,02	0,00	0,01	0,02	1,35	0,04	0,00	1,32	**85,04**	11,14
	Bâti/Surf min	0,85	0,35	0,15	0,15	0,11	0,62	0,04	0,00	13,69	0,14	**83,89**

tableau VI.2 : Sur la diagonale : pourcentage des classes qui n'ont pas changé de 2006 à 2009, les autres éléments signifient un changement entre 2006 et 2009

Non-changements/Monoculture 2006-2007-2008-2009	Surface (ha)	Fréquence (%)
Maïs-Maïs-Maïs-Maïs	3338,55	7,799
Blé-Blé-Blé-Blé	89,34	0,209
Tournesol-Tournesol-Tournesol-Tournesol	23,20	0,054
Soja-Soja-Soja-Soja	0,92	0,002
Bois-Bois-Bois-Bois	3019,05	7,052
Surf. enherbées-Surf. enherbées-Surf. enherbées-Surf. enherbées	7185,22	16,785

tableau VI.3 : Monocultures entre 2006 et 2009

	Surface (ha)	Fréquence (%)
Blé-Tournesol- Blé-Tournesol	2191,92	5,120
Blé-Colza- Blé-Colza	453,95	1,060
Blé-Maïs- Blé-Maïs	159,6	0,373
Blé-Orge- Blé-Orge	32,74	0,076
Maïs-Soja- Maïs-Soja	31,56	0,074
Tournesol-Surf. enherbées-Tournesol-Surf. enherbées	29,98	0,070
Blé-Soja-Blé-Soja	25,74	0,060
Blé-Surf. enherbées-Blé-Surf. enherbées	25,18	0,059
Maïs-Surf. enherbées-Maïs-Surf. enherbées	18,48	0,043
Orge-Maïs-Orge-Maïs	8,84	0,021
Orge-Tournesol-Orge-Tournesol	8,53	0,020
Maïs-Tournesol-Maïs-Tournesol	5,12	0,012
Colza-Orge-Colza-Orge	0,62	0,001
Surf. enherbées-Orge-Surf. enherbées-Orge	0,52	0,001

tableau VI.4 : Bicultures entre 2006 et 2009 : rotations annuelles sur ces 4 années

Rotations	Bilan carbone (en tC/ha)	Quantité de Carbone (en tC)	Surface (en ha)
mais-mais-mais-mais	4,009	15566,938	3883,469
tournesol-blé-tournesol-blé	4,374	5696,373	1302,298
blé-tournesol-blé-tournesol	2,079	3219,666	1548,986
tournesol-blé-blé-tournesol	3,315	1752,335	528,621
tournesol-blé-colza-blé	4,060	1267,442	312,192
colza-blé-colza-blé	4,310	1108,521	257,171
blé-blé-tournesol-blé	3,215	1012,628	314,925
blé-blé-colza-blé	2,901	775,902	267,443
blé-colza-blé-tournesol	1,810	745,225	411,757
mais-blé-mais-mais	3,949	703,616	178,176
mais-mais-tournesol-mais	3,937	633,545	160,941
tournesol-blé-blé-colza	2,502	560,208	223,891
colza-blé-tournesol-blé	4,625	549,254	118,765
tournesol-blé-tournesol-tournesol	3,120	549,003	175,936
blé-colza-blé-blé	3,063	519,352	169,530
…	…	…	…
…	…	…	…
blé-surf. enherbées-blé-tournesol	-1,073	-100,362	93,542
blé-surf. enherbées-blé-colza	-1,073	-102,635	95,661
surf. enherbées - surf. enherbées -blé-tournesol	-1,690	-103,562	61,274
surf. enherbées - surf. enherbées -blé-blé	-1,690	-113,600	67,213
surf. enherbées -blé-blé-tournesol	-1,690	-115,893	68,570
surf. enherbées - surf. enherbées -blé-colza	-1,690	-162,655	96,237
surf. enherbées - surf. enherbées -blé- surf. enherbées	-1,073	-168,547	157,094
surf. enherbées - surf. enherbées -tournesol- surf. enherbées	-1,073	-200,408	186,790
surf. enherbées - surf. enherbées -tournesol-blé	-1,690	-263,740	156,045
surf. enherbées - surf. enherbées - surf. enherbées -tournesol	-1,690	-267,407	158,214

tableau VI.5 : Rotations des classes et bilan carbone associé pour les données Formosat-2 des années 2006 à 2009

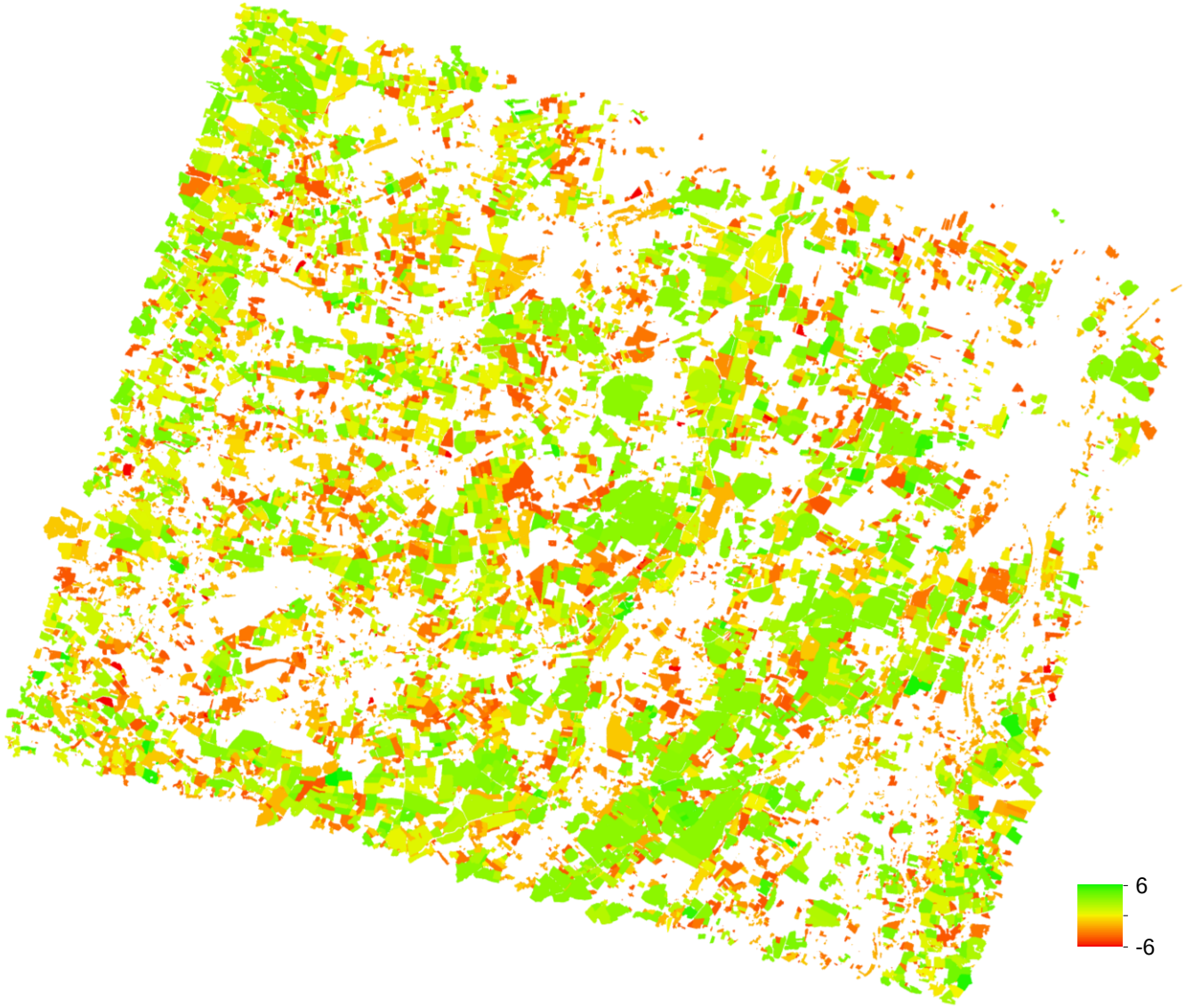


figure VI.7 : Estimation du flux de Carbone pour l'emprise Formosat-2 années 2006 à 2009 : unité en tonnes de Carbone par hectare, agrandissement en figure VI.10

4.2. A partir d'images Spot-2/4/5 des années 2006 à 2010

La zone d'étude est l'intersection des zones étudiées pour les années 2006, 2007, 2008, 2009 et 2010. Celle-ci a une surface totale de 46146 ha.

Comme pour l'application précédente, le système de classification automatique SCF a été appliqué aux images Spot de chacune des cinq années. Les performances de ces classifications sont présentées en annexe 1.4.

Les rassemblements de classes sont similaires à ceux de l'application précédente. Le tableau VI.6 montre les classes n'ayant pas changé sur la période de 5 années.

Non-changements/Monoculture 2006-2007-2008-2009-2010	Surface (ha)	Fréquence (%)
Bois-Bois-Bois-Bois-Bois	3583,48	7,931
Blé-Blé-Blé-Blé-Blé	82,88	0,183
Mais-Mais-Mais-Mais-Mais	2545,04	5,633
Tournesol-Tournesol-Tournesol-Tournesol-Tournesol	2,36	0,005
Surf. Enherb-Surf. Enherb.-Surf. Enherb-Surf. Enherb.-Surf. Enherb.	11867,88	26,266
Eau-Eau-Eau-Eau-Eau	320,28	0,709

tableau VI.6 : Non changements et monoculture de l'année 2006 à 2010

Les rotations de cultures et des statistiques agricoles de la zone ont permis le calcul du bilan de carbone à partir du modèle présenté précédemment (tableau VI.7).

Le bilan carbone de la zone étudiée de 46146 ha est de 90386 tonnes de Carbone stocké en 5 ans soit 0,392 tonne de carbone par hectare par an stocké dans le sol.

La figure VI.8 illustre le bilan carbone par pixels de l'image.

L'application suivante concernera la validation croisée de ces deux applications sur leur zone commune.

Rotations	Bilan carbone (en tC/ha)	Quantité de Carbone (en tC)	Surface (en ha)
blé-tournesol-blé-tournesol-blé	4,284	15166,956	3540,560
mais-mais-mais-mais-mais	4,103	10442,559	2545,040
tournesol-blé-tournesol-blé-tournesol	2,641	5632,628	2132,520
blé-colza-blé-tournesol-blé	3,524	2479,432	703,560
blé-tournesol-blé-colza-blé	4,807	2460,058	511,800
tournesol-blé-blé-tournesol-blé	3,339	1862,828	557,840
blé-blé-blé-tournesol-blé	3,462	1637,878	473,120
blé-tournesol-blé-mais-blé	4,939	1340,653	271,440
blé-colza-blé-colza-blé	4,047	1317,551	325,560
blé-mais-blé-mais-blé	4,698	1300,344	276,800
blé-tournesol-blé-blé-tournesol	4,534	1147,307	253,040
blé-blé-tournesol-blé-tournesol	2,764	1131,392	409,360
blé-blé-blé-colza-blé	3,985	960,330	241,000
mais-mais-mais-mais-blé	4,045	892,243	220,560
mais-blé-mais-mais-mais	3,522	882,587	250,560
…	…	…	…
…	…	…	…
surf. min.-surf. enherbées-surf. enherbées-tournesol- surf. enherbées	-0,371	-0,238	0,640
bois-surf. enherbées-bois-mais-bois	-0,228	-0,264	1,160
surf. min.-surf. min.-surf. min.-tournesol- surf. enherbées	-0,371	-0,327	0,880
surf. enherbées-surf. enherbées-surf. min.-blé- surf. enherbées	-0,384	-0,384	1,000
bois-bois-bois-mais- surf. enherbées	-0,748	-0,419	0,560
surf. enherbées-blé-surf. enherbées-blé- surf. enherbées	-0,295	-0,532	1,800
surf. enherbées-surf. enherbées-bois-mais- surf. enherbées	-0,748	-3,890	5,200
bois-bois-bois-mais-bois	-0,228	-5,217	22,920
surf. enherbées-surf. enherbées-surf. enherbées-blé- surf. enherbées	-0,384	-6,859	17,840
surf. enherbées-surf. enherbées-surf. enherbées-tournesol- surf. enherbées	-0,371	-10,432	28,080

tableau VI.7 : Rotations des classes et bilan carbone associé pour les données Spot des années 2006 à 2010

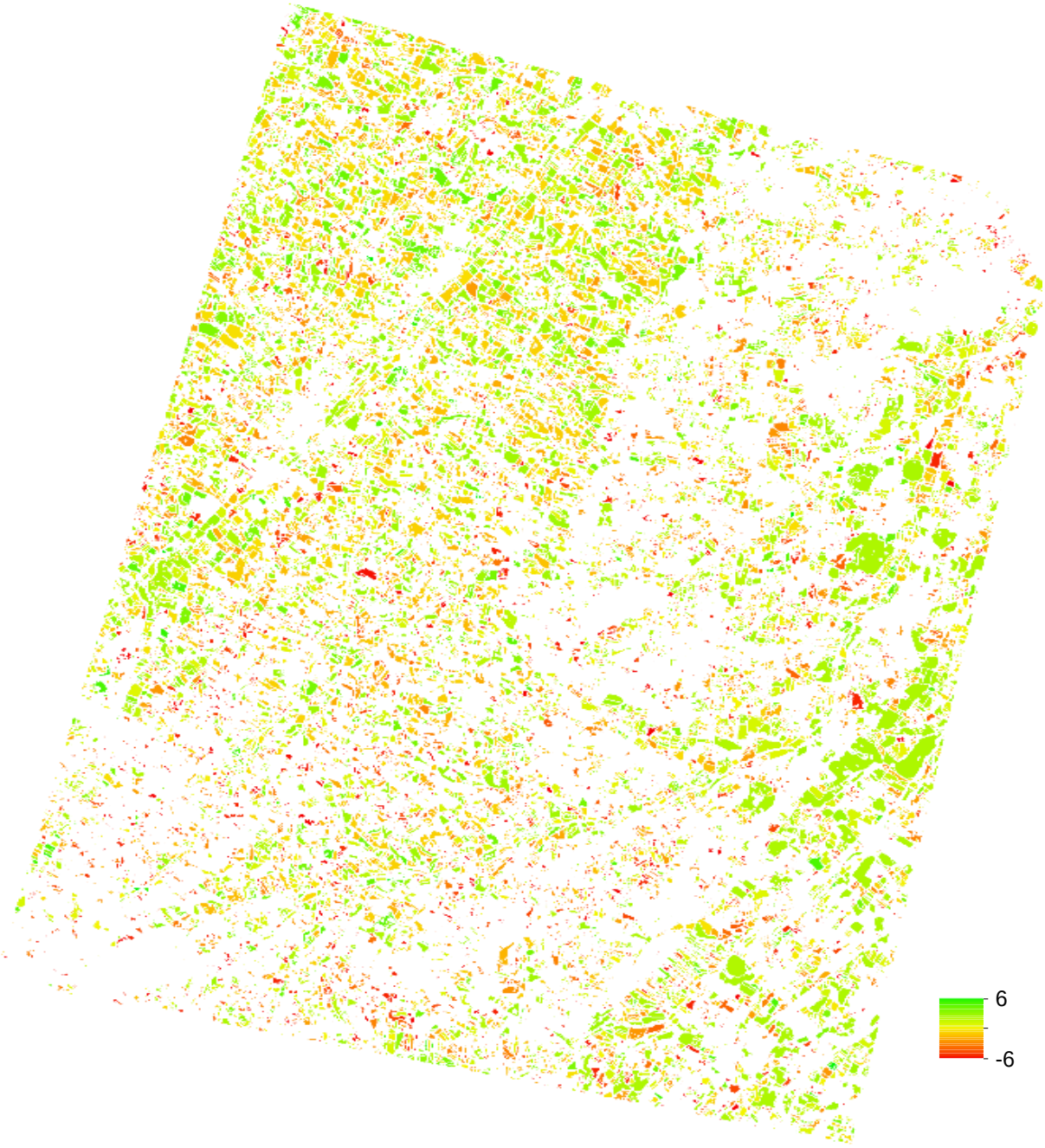


figure VI.8 : Estimation du flux de Carbone de l'emprise Spot-2/4/5 années 2006 à 2010 : unité en tonnes de Carbone par hectare, agrandissement en figure VI.11

4.3. Validation croisée de la zone commune des 2 satellites

Une comparaison de résultats entre les deux études précédentes a été réalisée afin d'évaluer la stabilité des modèles d'estimation de flux de Carbone. En pratique, la zone commune à toutes les années des deux applications est fortement réduite. La surface de superposition des deux applications est de 12002 ha.

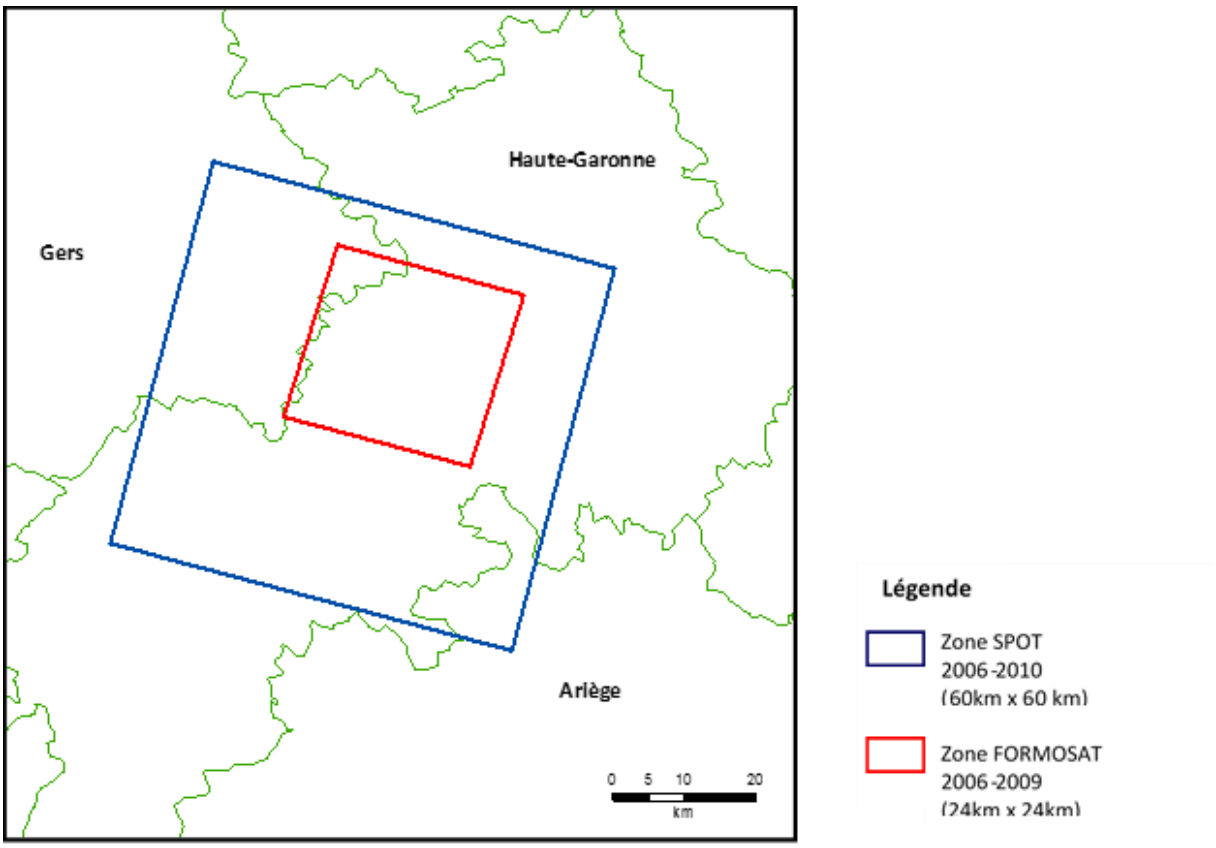


figure VI.9 : Zones couvertes par les satellites SPOT et FORMOSAT

Du fait d'une résolution différente ente les images Formosat-2 (8m) et Spot (20m), l'érosion a un effet plus important sur le résultat de l'application des données Spot. Ce processus d'érosion est nécessaire car il permet de réduire l'influence des mixels et des erreurs de décalages multi-temporels des images sur le modèle d'estimation. (figure VI.10 et figure VI.11).

Les résultats obtenus à partir des deux jeux de données indépendants sont similaires, les seules différences sont dues à une année supplémentaire pour les images Spot.

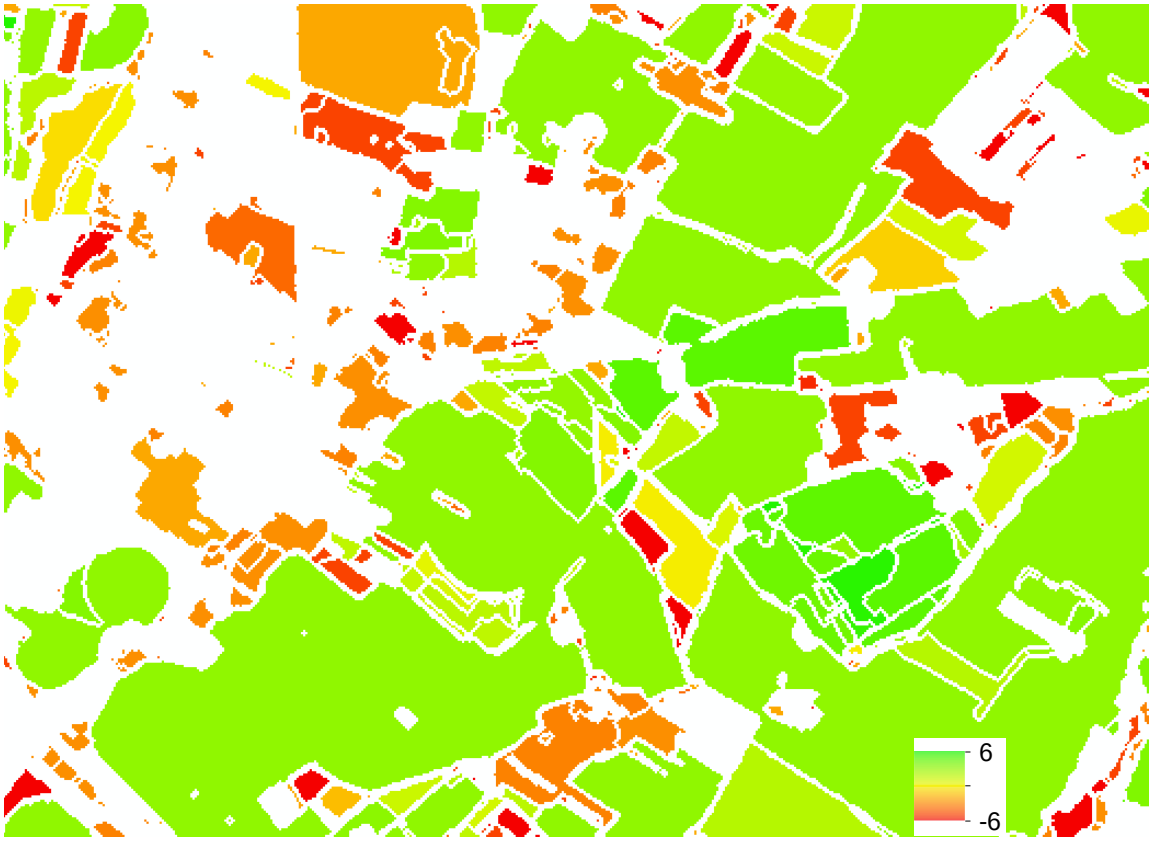


figure VI.10 : Flux de carbone obtenu à partir des images Formosat-2 des années 2006 à 2009

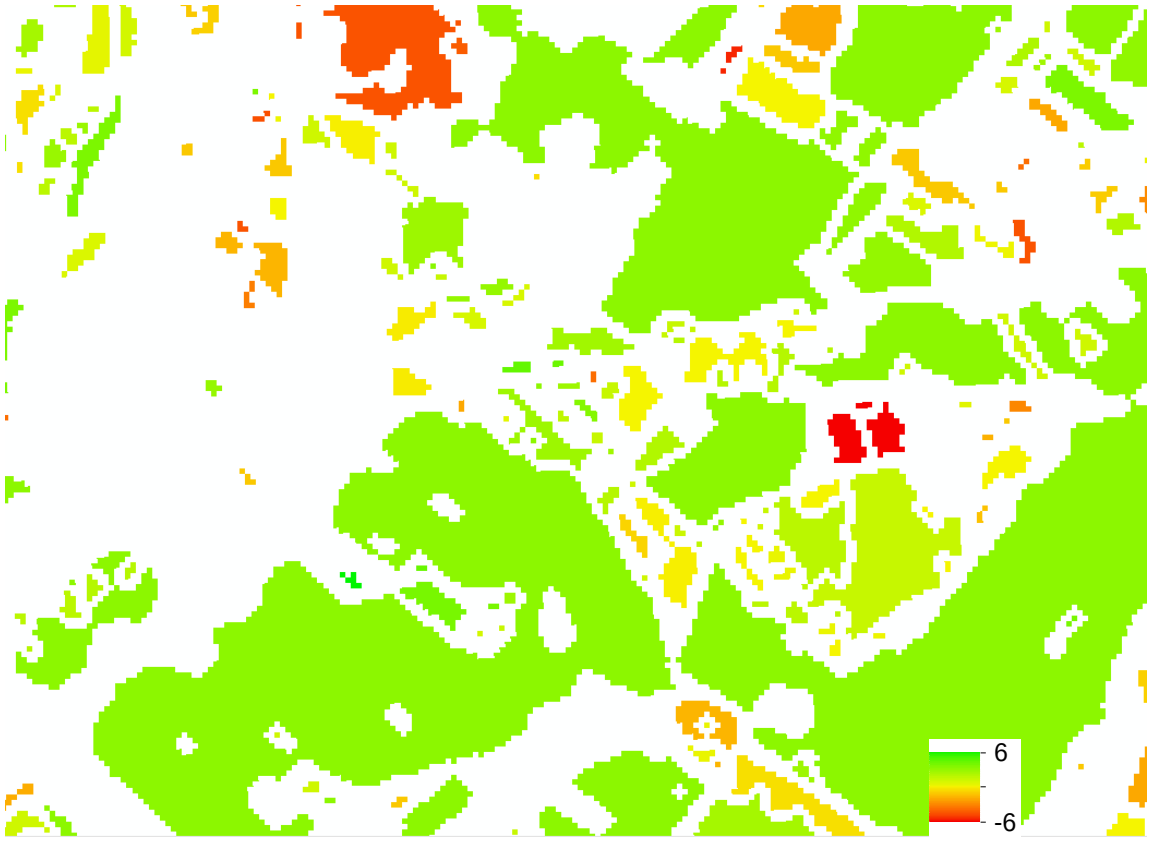


figure VI.11 : Flux de carbone obtenu à partir des images SPOT des années 2006 à 2010

5. Conclusion

L'objectif de ce chapitre n'était pas d'évaluer la qualité des modèles d'estimation de flux de carbone mais plutôt l'évaluer la robustesse des méthodes de classification et des cartes de rotations qui en sont déduites. Les modèles d'estimation de flux de carbone nécessitent des classifications les plus justes possibles afin d'éviter l'accumulation d'erreurs et la présence de faux changements d'occupation des terres.

Les applications ont permis de conclure qu'il était possible de quantifier les flux de carbone entre le sol et l'atmosphère à partir d'images de télédétection à haute résolution. Ces applications ont également permis de sélectionner les pratiques culturales qui permettent le meilleur stockage de carbone dans le sol et donc de limiter les émissions de gaz à effets de serre.

La validation croisée effectuée à partir des deux applications présentées (capteurs différents) a montré que le résultat du bilan carbone est constant et donc que la méthodologie appliquée (le système de classification automatique SCF) est robuste.

Chapitre VII. La classification de la Trame Verte et Bleue en milieu agricole, entre enjeux et réalités.

Résumé : L'objectif principal de cette application est d'explorer les possibilités offertes par les données de télédétection et leur classification pour cartographier les éléments de la trame verte (haies, corridors écologiques, espaces boisés). La méthodologie proposée est testée sur plusieurs capteurs : radar avec TerraSar-X, optique avec un satellite à Très Haute Résolution Spatiale (THRS) SPOT 5 (2,5m et 10m), Formosat (8m) ainsi que Worldview-2 (1,7m) et Pleiades (0,7m). Nous verrons ainsi que la multi-temporalité et la haute résolution des images sont un atout pour la cartographie des trames vertes.

Mots clés : Trame verte et bleue, haies, Copernicus, Grenelle-2

Sommaire du chapitre VII

1. Introduction

La Trame Verte et Bleue, qui est l'un des engagements du Grenelle Environnement (en 2007), est une démarche qui vise à maintenir et à reconstituer un réseau d'échanges sur le territoire national pour que les espèces animales et végétales puissent, de la même manière que l'Homme, communiquer, circuler, s'alimenter, se reproduire, se reposer. En d'autres termes, assurer leur survie.

Elle contribue ainsi au maintien des services que nous rend la biodiversité : qualité des eaux, pollinisation, prévention des inondations, amélioration du cadre de vie, etc.

La biodiversité, fruit d'un travail de la nature depuis quelques 4 milliards d'années, est aujourd'hui menacée du fait de l'activité incessante des hommes. En France, environ 165 ha de milieux naturels et terrains agricoles (soit un peu plus de quatre terrains de football) sont détruits chaque jour, remplacés par des routes, habitations, zones d'activités. Cela équivaut à plus de 60000 ha par an, soit un département comme la Savoie tous les 10 ans.

L'enjeu de la constitution d'une trame verte et bleue s'inscrit bien au-delà de la simple préservation d'espaces naturels isolés et de la protection d'espèces en danger. Il est de (re)constituer un réseau écologique cohérent qui permette aux espèces de circuler et d'interagir, et aux écosystèmes de continuer à rendre à l'homme leurs services.

Une espèce a besoin d'un effectif minimal d'individus pour survivre aux différents obstacles naturels tels une épidémie ou une prédation, et donc d'un territoire assez grand. Le projet Trame Verte et Bleue consiste directement à constituer un territoire pouvant accueillir ces espèces et assurer ainsi leur survie. Une ville entière sans arbre ni surface herbée est synonyme d'une discontinuité entre plusieurs espaces naturels aux alentours, et donc synonyme de l'extinction de certaines de ces espèces abritées par ces milieux. En créant ce maillage et donc une continuité entre différents espaces abritant de nombreuses espèces, on contribue à leur survie et au maintien de la biodiversité.

L'objectif principal de ce chapitre est d'explorer les possibilités offertes par les méthodes de classification appliquées aux données de télédétection pour la cartographie des haies et des espaces boisés. Une méthodologie est proposée et testée sur des images à caractéristiques diverses (optique/radar, haute et très haute résolution spatiale, répétitivité temporelle) dont les capteurs suivants : TerraSar-X (radar, 3m), Spot-5 (2,5m et 10m), Formosat (8m), Worldview-2 (1,7m) et Pléiades (0,7m).

L'étape la plus complexe est la validation des résultats de classification, la trame verte étant un « objet » particulier, qui est en général est fin, hétérogène et multiforme. La prise d'échantillons est alors délicate et fastidieuse, mais indispensable pour valider la méthodologie. Pour avoir une validation de la détection des haies la plus précise possible, nous utiliserons la photo-interprétation à partir de la BD-Ortho ® (source IGN) et d'images de satellites.

2. Etudes de l'utilisation de la télédétection pour la cartographie de la trame verte en milieu agricole

L'utilisation d'images de télédétection pour la cartographie de la trame verte en milieu agricole est une problématique jusqu'à lors peu étudié. Pour cela, nous avons mis en pratique les outils présentés aux chapitres précédents.

Nous présenterons plusieurs études concernant la classification de la Trame verte, l'une sur l'intérêt de la résolution spatiale et l'autre sur le choix du classifieur.

Les détails de ces études sont publiés et les articles disponibles en annexe 3.1 et 3.2.

2.1. Etude de l'impact de la résolution spatiale pour la cartographie de la trame verte en milieu agricole

Pour cette étude, nous avons tenté de déterminer les capteurs qui permettent la meilleure cartographie des éléments de la trame verte (haies, arbres isolés, …). La liste des capteurs utilisés est la suivante :

- Spot-5, année 2010, 10 mètres de résolution
- Formosat-2, année 2007, 8 mètres de résolution
- Terrasar-X, année 2009, 3 mètres de résolution
- Spot-5, année 2010, 2.5 mètres de résolution
- Pléiades-1, année 2012, 0.70 mètre de résolution

2.1.1. Comparaison des capteurs à haute résolution pour la cartographie des haies

La comparaison des capteurs Formosat-2, Spot-5 (2.5m) et TerraSAR-X a donné lieu à un article de conférence à comité de lecture (Ducrot, Masse, & Ncibi, 2012) disponible en annexe 3.2. L'objectif principal de cette étude est d'explorer le potentiel des méthodes de classification pour la cartographie des haies, des corridors et des forêts. Nous présentons une méthodologie basée sur la classification, supervisée et non supervisée, la segmentation et la fusion de classification. Nous avons conclu que l'emplacement des haies peut être détecté aux échelles de 2.5 à 8 mètres, les meilleurs résultats de classification sont cependant obtenus avec une résolution de 2,5 mètres. La multi temporalité est aussi importante pour la détection des haies, notamment avec des dates de printemps et d'automne.

Cette étude a été complétée avec le capteur Spot-5 à 10 mètres de résolution. La classification obtenue fournit de très bons résultats. Les haies sont bien discernables sauf certaines plus étroites qui se distinguent plus difficilement. Néanmoins, le résultat est très satisfaisant pour cette résolution (figures VII.1 et VII.2). Les lisières de forêts sont souvent classées en Haie car elles possèdent une signature spectrale différente de celle des forêts de feuillus ou de résineux, mais proche de celle des haies. Par des méthodes de post-classifications, elle a pu être extraite en tant que classe Lisières, et donc séparée de la classe Haie.

Figure VII.1 : Composition colorée d'une image Spot-5 (10m) en fausse couleur (PIR, R, V)

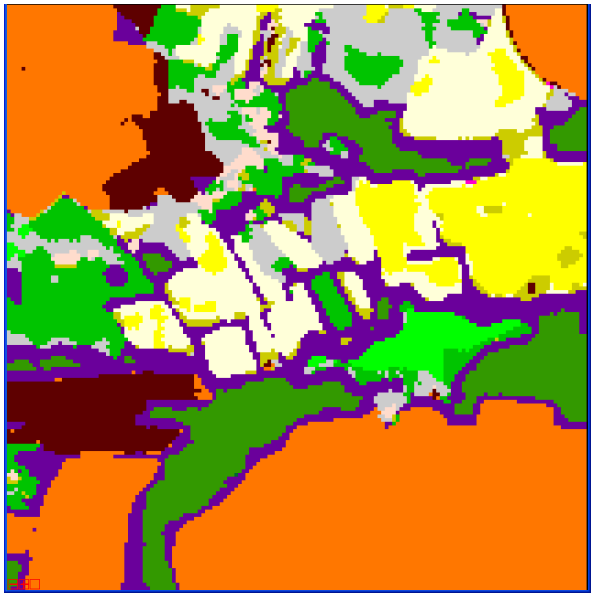


Figure VII.2 : Résultat de la classification des données Spot-5. Légende : **Haies en violet**, Feuillus en vert, Jachère en gris, Prairie en vert clair

2.1.2. Utilisation de données à Très Haute Résolution Spatiale (THRS)

Dans l'article (Sheeren et al., 2012) présenté en annexe 3.1 et dont je suis co-auteur, nous proposons d'évaluer les potentialités de l'imagerie satellitaire à très haute résolution spatiale pour cartographier la composante arborée de la trame verte en milieu agricole. L'identification des formations ligneuses, incluant les haies et les arbres isolés, est réalisée à partir de données multi-angulaires WorldView-2 en comparant deux approches de classification supervisée : modèle de mélange gaussien et méthode contextuelle ICM. Les résultats montrent de bonnes performances en exploitant les deux vues du couple stéréoscopique (kappa de 0,88) avec un gain de 13 à 22 % par rapport à l'utilisation d'une seule vue. La qualité d'extraction des haies varie selon le type de haie, le voisinage des objets, les ombres et les effets de la visée oblique. Les différences de performance entre les méthodes de classification sont quant à elle négligeables.

En complément de cette étude, le nouveau satellite Pléiades (annexe 1.6) a permis très récemment l'acquisition d'images à Très Haute Résolution Spatiale (0,7 mètre). La classification supervisée de la zone commune aux deux images Pléiades présentées en (annexe 1.6, dates de Mars et Novembre 2012) permet une cartographie précise des haies du fait de la très haute résolution du capteur. Les classes « feuillus, résineux, peupliers, haie haute et haie basse » se côtoie au sein d'une seule et même haie car les haies champêtres sont naturellement composées de ces espèces (figures VII.3 à VII.5).

figure VII.3 : Composition colorée (fausses couleurs) de l'image Pléiades de mars 2012

figure VII.4 : Composition colorée (fausses couleurs) de l'image Pléiades de novembre 2012

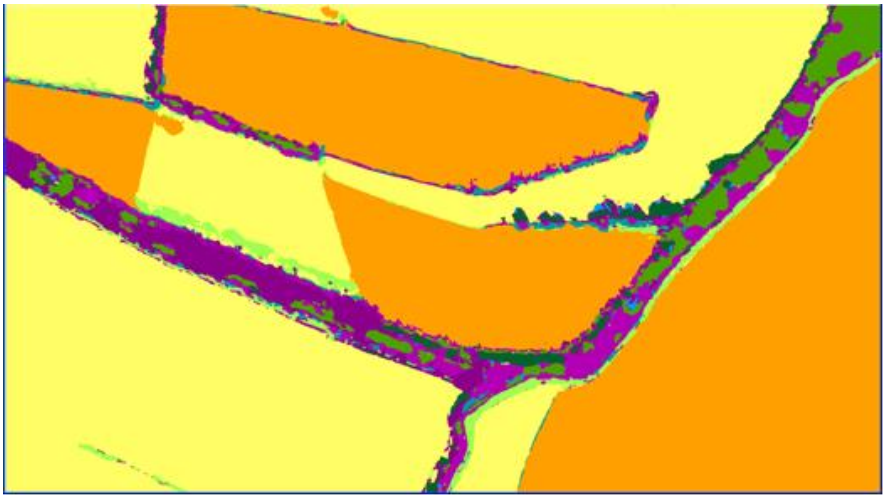


figure VII.5 : Extrait de l'image classée par la méthode ICM. Légende : Orange = cultures été, Jaune = cultures hiver, Violet clair = Haie haute, Violet foncé = Haie basse, Vert = Feuillus, Vert foncé = Résineux

2.2. Etude de l'impact du choix de la méthode de classification pour la cartographie des haies

A partir des données Spot-5 de l'année 2010 (10 mètres de résolution spatiale), nous avons comparé les méthodes ICM et SVM présentées au chapitre 1 (en mode supervisé).

Le tableau VII.1 montre que l'algorithme ICM offre de meilleurs résultats pour la cartographie de l'occupation du sol à partir des 10 dates de 2010 et de 31 classes (celles présentées en annexe 1.4 ainsi que 5 classes spécifiques à la trame verte dont la classe haie, bosquets et ligneux). Les résultats sont en moyenne 10% supérieur pour la méthode ICM en matière d'indices de performance globale. Pour les classes de trames vertes, les différences de résultats sont beaucoup plus marquées avec des écarts de plus de 60% en faveur de la méthode ICM.

Indices de performance (en % sauf Kappa et F1)	Classification ICM	Classification SVM
AA	**95,24**	90,52
AP	**91,42**	70,11
OA	**90,11**	81,28
Kappa	**0,89**	0,79
OACI	**87,38**	63,41
F1	**0,93**	0,79
Valeur OCI classe Haie	**49,34**	37,11
Valeur OCI classe Bosquets	**94,61**	79,61
Valeur OCI classe ligneux	**84,88**	19,13

tableau VII.1 : Comparaison des performances des classifications ICM et SVM, images Spot-5 année 2010

Cette application confirme les résultats présentés précédemment en faveur de la méthode ICM pour des classes à majorité agricoles.

Les figures VII.6 à VII.8 montrent un extrait des résultats de ces classifications.

figure VII.6 : Référence : BD Ortho ®, IGN

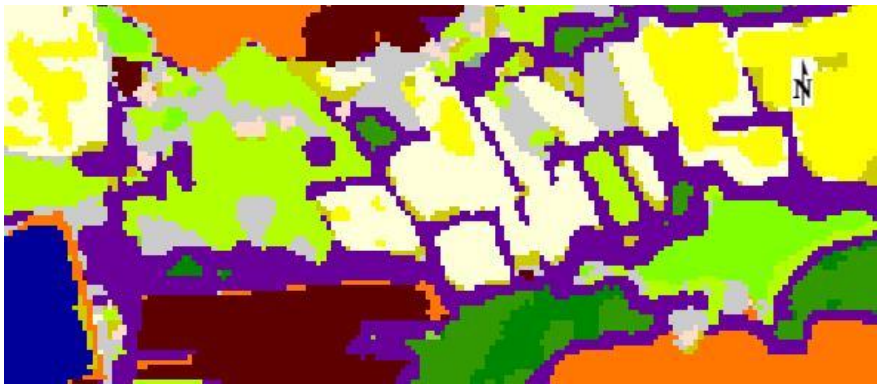


figure VII.7 : Classification par méthode ICM des données Spot (2,5m, année 2010), rappel de légende : **haies en violet**, Feuillus en vert foncé, Prairie en vert clair, Friche en gris

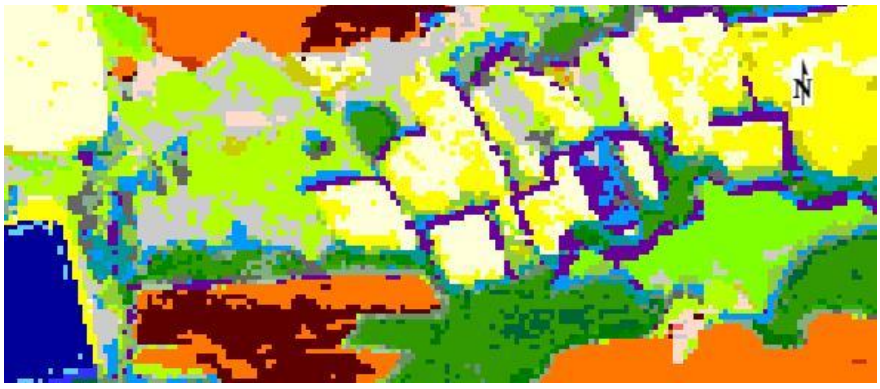


figure VII.8 : Classification par méthode SVM, rappel de légende : **haies en violet**, Feuillus en vert foncé, Prairie en vert clair, Friche en gris, Eucalyptus en bleu clair

3. Conclusion

Les récents efforts des élus à promouvoir la préservation des habitats naturels et à inscrire cet effort dans les pages du Grenelle de l'Environnement ont mis en avant l'importance des haies. De nouvelles techniques de reconnaissance, de recensement et de cartographie se développent et l'automatisation de celles-ci permettront d'être plus efficace. Les résultats présentés montrent des cartes indicatives, nécessaires aux différents acteurs du maintien des corridors écologiques, qui peuvent disparaitre du fait des activités anthropiques grandissantes. Du choix de l'image satellite à la validation de la classification non supervisée ou supervisée par l'utilisateur, une attention particulière est requise, tant au niveau du choix des dates que lors de l'échantillonnage qu'il soit de vérification ou d'apprentissage.

En général, la qualité globale des résultats est satisfaisante avec une bonne moyenne de pixels correctement classés pour la classe Haie. Des confusions avec les bordures des bois et avec des Bois parsemés (proches des haies qui sont constituées de beaucoup de feuillus et résineux) peuvent apparaitre, mais sont corrigés par traitements de post-classification. Toutefois, pour obtenir une plus grande précision des résultats, le mode supervisée s'impose.

On peut conclure en soulignant que l'emplacement des haies peut être détecté aux échelles de 2.5 à 10m, que les dates de juin et de préférence celles d'automne, et les bandes spectrales MIR et/ou rouge sont importantes pour détecter les haies. Une résolution inférieure à 2,5m fournit beaucoup de détails sur les haies et la prise d'échantillons est plus difficile du fait de l'hétérogénéité de cette classe.

La méthode de classification globale ICM est la méthode ayant fournie les meilleurs résultats, même pour des objets fins et hétérogènes, mais le nombre d'échantillons d'apprentissage doit être suffisant.

Conclusion générale

Les nouveaux et futurs satellites offriront une plus grande diversité de caractéristiques : spectrale, temporelle, résolution spatiale et superficie de l'emprise du satellite. Les données de télédétection sont aujourd'hui abondantes et ouvertes au grand public, ce qui ouvre de nouvelles problématiques quant à la manière d'utiliser ces données.

Tout au long de cette thèse, je vous ai présenté plusieurs méthodes permettant de traiter de manière automatique et performante des images de télédétection dans le but d'obtenir une carte d'occupation des sols. Ces méthodes forment un système qui maximise la performance de la classification en fonction des données fournies (images de télédétection et ensemble de classes thématiques). Il produit une image classée prête à une étude thématique.

Les problématiques introduites et l'approche proposée

Au cours de ce manuscrit, j'ai été en mesure de répondre aux questions posées en introduction.

- Quelles sont les meilleures images à utiliser pour une application thématique ?

Le problème de sélection de données est un problème important en télédétection lorsque le nombre de données à traiter devient trop élevé. L'approche proposée lors de cette thèse s'est basée sur le principe des algorithmes génétiques que j'ai adapté au domaine de la télédétection. Le système de sélection présenté permet désormais la sélection automatique du meilleur jeu de données temporelles, spectrales ainsi que du classifieur à utiliser, et maximise ainsi la performance de la classification.

- Comment peut-on utiliser la diversité des informations pour améliorer les performances des systèmes de classification ?

La fusion de données fait partie des problématiques de la télédétection. L'approche proposée est ici différente des méthodes habituellement utilisées et permet de fusionner de manière originale des résultats de classifications par minimisation des confusions entre classes et donc la maximisation des critères de performance.

Ces deux processus sont regroupés dans une méthode générale appelée Sélection-Classification-Fusion qui permet d'obtenir automatiquement la meilleure classification en fonction d'un jeu de données et d'un ensemble de classes.

- Quelles sont les limites intrinsèques d'une application en fonction des données utilisées ? Peut-on se libérer de ces limites ? Si oui comment ?

Pour une application thématique, il n'est pas raisonnable de s'abstraire d'une définition *a priori* de classes. Pour cela, j'ai présenté une méthode permettant la classification sans vérité terrain de la zone à étudier, mais à l'aide de statistiques de valeurs radiométriques, obtenues à l'aide de référence de zones et/ou d'années différentes, et interpolées par méthode de krigeage.

Les résultats

Le système de Sélection-Classification-Fusion a été appliqué à un grand nombre de jeux de données : 8 capteurs (Formosat, Spot, Landsat, Pleiades, Worldview, Terrasar, Ers et Radarsat) et plus de 500 images. Pour toutes ces applications, les résultats montrent que ce système a apporté une amélioration significative des performances de la classification.

Les applications de la méthode de classification sans vérité terrain de la zone à étudier ont également montré des résultats satisfaisants, relativement proches de ceux obtenus avec des vérités terrain. Ces résultats, même indicatif, sont indispensables pour certains modèles de prédiction.

Ce système de classification a également été appliqué pour obtenir une carte des changements et des rotations de cultures, dans le but d'estimer les échanges de carbone entre le sol et l'atmosphère. Les applications (même zone, capteurs différents) ont montré la robustesse du système et la constance des résultats.

Enfin les études faites sur la cartographie des objets fins tels que les haies ou les éléments naturels ligneux ont montré l'importance de la télédétection pour des applications écologiques à grande échelle et la justesse des méthodes de classification utilisées.

Les perspectives

L'approche proposée durant cette thèse est un système de classification. Sa structure permet l'ajout de nouvelles méthodes, de nouvelles lois de classification pour un nouveau type de données,...

Une perspective intéressante de ce travail concerne l'intégration de la notion de logique floue. L'apport de cette notion dans le système de Sélection-Classification-Fusion permettra de réduire les erreurs de prise de décision notamment lorsque cette décision est incertaine.

Concernant l'interpolation de signatures spectrales et temporelles, il sera intéressant d'ajouter la possibilité d'inclure de nouvelles informations sur les variations interannuelles, afin de réduire les erreurs d'interpolation de certaines classes telles que les cultures dont les dates des stades phénologiques varient chaque année.

La grande quantité et la diversité des données de télédétection permettent aujourd'hui la validation des nouvelles méthodes proposées. Les données Sentinel-2 ne seront réellement disponibles qu'à partir de 2014 et c'est à ce moment-là que ces méthodes seront appliquées à grande échelle.

Références

Agostinelli, A., Dambra, C., Serpico, S., & Vernazza, G. (1991). Multisensor data fusion by a region-based approach *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on* (pp. 2617–2620).

Alliot, Jean-Marc, & Durand, Nicolas. (2005). Algorithmes génétiques. *Centre d'Etudes de la Navigation Aérienne*.

Arrouays, D., Antony, V., Bardy, M., Bispo, A., Brossard, Michel, Jolivet, C., . . . Schnebelen, N. (2012). Fertilité des sols: conclusions du rapport sur l'état des sols de France. *Innovations Agronomiques, 21*, 1733–1744.

Bahadur, Raghu Raj, & Rao, R. Ranga. (1960). On deviations of the sample mean. *The Annals of Mathematical Statistics, 31*(4), 1015–1027.

Bajcsy, Peter, & Groves, Peter. (2004). Methodology for hyperspectral band selection. *Photogrammetric engineering and remote sensing, 70*, 793–802.

Ball, Geoffrey H., & Hall, David J. (1965). ISODATA, a novel method of data analysis and pattern classification: DTIC Document.

Benediktsson, Jon A., Sveinsson, Johannes R., & Amason, K. (1995). Classification and feature extraction of AVIRIS data. *Geoscience and Remote Sensing, IEEE Transactions on, 33*(5), 1194–1205.

Besag, Julian. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.

Beucher, Serge, & Lantuéjoul, Christian. (1979). Use of watersheds in contour detection.

Bhattacharyya, Anil. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc, 35*(99-109), 4.

Bloch, I. (1996). 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.

Bloch, I. (2005). Fuzzy Spatial Relationships for Image Processing and Interpretation: A Review. *Image and Vision Computing, 23*(2), 89-110.

Bloch, I., & Maître, H. (2002). Fusion d'informations en traitement d'images : spécificités, modélisation et combinaison par des méthodes numériques. *Techniques de l'Ingénieur, TE 5 230*, 1-26.

Briem, G. J., Benediktsson, J. A., & Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remote sensing data. *Geoscience and Remote Sensing, IEEE Transactions on, 40*(10), 2291–2299.

Ceschia, Eric. (2013). Fonctionnement des surfaces continentales et gestion durable des territoires : le chantier Sud-Ouest. Retrieved 21/06/2013, from http://www.cesbio.ups-tlse.fr/fr/sud_ouest.html

Chair, Z., & Varshney, P. K. (1986). Optimal Data Fusion in Multiple Sensor Detection Systems. *IEEE Transactions on Aerospace and Electronic Systems, AES-22*(1), 98-101.

Chavez, Pats, Sides, Stuart C., & Anderson, Jeffrey A. (1991). Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and remote sensing, 57*(3), 295–303.

Cicchetti, Dominic V., & Fleiss, Joseph L. (1977). Comparison of the null distributions of weighted kappa and the C ordinal statistic. *Applied Psychological Measurement, 1*(2), 195–201.

Cocquerez, Jean Pierre, & Philipp-Foliguet, Sylvie. (1995). Analyse d'images: filtrage et segmentation.

Cohen, Jacob. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37–46.
Cohen, Jacob. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin, 70*(4), 213.
Collet, Claude. (2001). *Précis de télédétection: traitements numériques d'images de télédétection* (Vol. 3): Puq.
Congalton, Russell G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35-46.
Cournot, Antoine-Augustin. (1851). *Essai sur les fondements de nos connaissances et sur les caractères de la critique philosophique* (Vol. 1): Librairie de L. Hachette.
Cover, Thomas, & Hart, Peter. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on, 13*(1), 21–27.
Dash, Manoranjan, & Liu, Huan. (1997). Feature selection for classification. *Intelligent data analysis, 1*(1-4), 131–156.
Davis, Jesse, & Goadrich, Mark. (2006). The relationship between Precision-Recall and ROC curves *Proceedings of the 23rd international conference on Machine learning* (pp. 233–240).
Dawkins, Richard, & Sigaud, Bernard. (1989). *L'horloger aveugle*: R. Laffont.
De Moivre, Abraham. (1756). *The doctrine of chances: or, A method of calculating the probabilities of events in play*: Chelsea Pub. Co.
de Pater, Wilhelmus Antonius. (1965). *Les Topiques d'Aristote et la dialectique platonicienne: la méthodologie de la définition* (Vol. 10): Éditions St. Paul.
Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics, 16*(1), 41–50.
Descombes, Xavier, Mangin, Jean-Fran ccois, Sigelle, M., & Pechersky, E. (1995). Fine structures preserving Markov model for image processing *Proceedings of the scandinavian conference on image analysis* (Vol. 1, pp. 349–356).
Diday, Edwin. (1971). Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques. *Revue de statistique appliquée, 19*(2), 19–33.
Dubois, Didier, & Prade, Henri. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty (traduction revue et augmentée de "Théorie des Possibilités")*. New York: Plenum Press.
Dubois, Didier, & Prade, Henri. (1993). *Foreword of 'Fuzzy Decision Procedures with Binary Relations — Towards a Unified Theory' by L. Kitainik*: Kluwer.
Ducrot, D. (2005). Méthodes d'analyse et d'interprétation d'images de télédétection multisources et extraction des caractéristiques du paysage, HDR.
Ducrot, D., Masse, A., Ceschia, E., Marais-Sicre, C., & Krystof, D. (2010). *A methodology for the detection of land cover changes: application to the Toulouse southwestern region.* Paper presented at the Remote Sensing.
Ducrot, D., Masse, A., Marais-Sicre, C., & Dejoux, J. F. (2010). *Multisensor and multitemporal image fusion methods to improve remote sensing image classification*: RAQRS'III.
Ducrot, D., Masse, A., & Ncibi, A. (2012). Hedgerow detection in HRS and VHRS images from different source (optical, radar) *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* (pp. 6348 -6351).
Ducrot, D., Sassier, H., Mombo, J., Goze, S., & Planes, J.G. (1998). Contextual methods for multisource land cover classification with application to Radarsat and SPOT data *Remote Sensing* (pp. 225–236).
Duda, Richard O., Hart, Peter E., & Stork, David G. (2012). *Pattern classification*: Wiley-interscience.

Elhassouny, Azeddine, Idbraim, Soufiane, Bekkari, Aissam, Mammass, Driss, & Ducrot, Danielle. (2012). Multisource Fusion/Classification Using ICM and DSmT with New Decision Rule. In A. Elmoataz, D. Mammass, O. Lezoray, F. Nouboud & D. Aboutajdine (Eds.), *ICISP* (Vol. 7340, pp. 56-64): Springer.

Feinstein, A. R., & Cicchetti, Dominic V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

Ferri, F. J., Pudil, P., Hatef, Mohamad, & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. *Machine Intelligence and Pattern Recognition, 16*, 403–403.

Fisher, Ronald A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*, 309–368.

Fleiss, Joseph L., Cohen, Jacob, & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin, 72*(5), 323–327.

Foucher, S., Germain, M., Boucher, J. M., & Bénié, G. B. (2002). Multisource classification using ICM and Dempster-Shafer theory. *Instrumentation and Measurement, IEEE Transactions on, 51*(2), 277–281.

Fukunaga, Keinosuke. (1990). *Introduction to statistical pattern recognition*: Academic press.

Galton, Francis. (1892). *Finger prints*: Macmillan and Company.

Geman, Stuart, & Geman, Donald. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721–741.

Goldberg, David E. (1989). Genetic algorithms in search, optimization, and machine learning.

Grossman, David A., & Frieder, Ophir. (2004). *Information retrieval: Algorithms and heuristics* (Vol. 15): Springer.

Hammersley, John M., & Clifford, Peter. (1968). Markov fields on finite graphs and lattices.

Hénin, S., & Dupuis, M. (1945). Essai de bilan de la matière organique du sol *Annales agronomiques* (Vol. 15, pp. 17–29).

Holland, John H. (1962). Outline for a Logical Theory of Adaptive Systems. *J. ACM, 9*(3), 297–314.

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*(4), 725–749.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on, 14*(1), 55–63.

Jain, Anil, & Zongker, Douglas. (1997). Feature selection: Evaluation, application, and small sample performance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19*(2), 153–158.

Joachims, Thorsten. (1998). *Text categorization with support vector machines: Learning with many relevant features*: Springer.

Kalayeh, Hooshmand M., & Landgrebe, David A. (1986). Utilizing multitemporal data by a stochastic model. *Geoscience and Remote Sensing, IEEE Transactions on*(5), 792–795.

Kato, Zoltan, Zerubia, Josiane, & Berthod, Marc. (1999). Unsupervised parallel image classification using Markovian models. *Pattern Recognition, 32*(4), 591–604.

Kato, Zoltan, Zerubia, Josiane, & Berthod, Mark. (1992). Satellite image classification using a modified Metropolis dynamics *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on* (Vol. 3, pp. 573–576).

Kendall, Maurice George. (1948). Rank correlation methods. *Rank correlation methods.*

Kirkpatrick, Scott, Jr, D. Gelatt, & Vecchi, Mario P. (1983). Optimization by simulated annealing. *science, 220*(4598), 671–680.

Kittler, Josef. (1978). Feature set search algorithms. *Pattern recognition and signal processing*, 41–60.

Kittler, Josef, Hatef, Mohamad, Duin, Robert P. W., & Matas, Jiri. (1998). On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20*(3), 226–239.

Krige, Daniel G. (1951). *A statistical approach to some mine valuation and allied problems on the Witwatersrand*. University of the Witwatersrand.

Kullback, Solomon, & Leibler, Richard A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Kuncheva, Ludmila I., Bezdek, James C., & Duin, Robert P. W. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, *34*(2), 299–314.

Laanaya, H., Martin, A., Aboutajdine, D., & Khenchaf, A. (2008). Classifier fusion for post-classification of textured images *Information Fusion, 2008 11th International Conference on* (pp. 1–7).

Landis, J. Richard, & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

Le Hégarat-Mascle, S., Bloch, I., & Vidal-Madjar, D. (1998). Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern recognition*, *31*(11), 1811–1823.

Lee, Jonathan, Weger, Ronald C., Sengupta, Sailes K., & Welch, Ronald M. (1990). A neural network approach to cloud classification. *Geoscience and Remote Sensing, IEEE Transactions on*, *28*(5), 846–855.

Marcotte, Denis. (1991). Cokriging with MATLAB. *Computers & Geosciences*, *17*(9), 1265–1280.

Masse, Antoine, Ducrot, Danielle, & Marthon, Philippe. (2011). Tools for multitemporal analysis and classification of multisource satellite imagery *Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the* (pp. 209–212).

Masse, Antoine, Ducrot, Danielle, & Marthon, Philippe. (2012). Evaluation of supervised classification by class and classification characteristics. *Proceedings of SPIE*, *8390*, 83902R.

Matheron, Georges. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246–1266.

Melgani, Farid, & Bruzzone, Lorenzo. (2004). Classification of hyperspectral remote sensing images with support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on*, *42*(8), 1778–1790.

Montserud, R. A., & Leamans, R. (1962). Comparing global vegetation maps with the kappa statistic. *Ecological Modelling*, 275-293.

Narendra, Patrenahalli M., & Fukunaga, Keinosuke. (1977). A branch and bound algorithm for feature subset selection. *Computers, IEEE Transactions on*, *100*(9), 917–922.

Nunez, Jorge, Otazu, Xavier, Fors, Octavi, Prades, Albert, Pala, Vicen cc, & Arbiol, Roman. (1999). Multiresolution-based image fusion with additive wavelet decomposition. *Geoscience and Remote Sensing, IEEE Transactions on*, *37*(3), 1204–1211.

Oresme, Nicole. (1940). *Le Livre de Éthiques d'Aristote*. New York: AD Menut.

Osuna, Edgar, Freund, Robert, & Girosit, Federico. (1997). Training support vector machines: an application to face detection *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* (pp. 130–136).

Pearl, Judea. (1984). Heuristics: intelligent search strategies for computer problem solving

Pontius Jr, Robert Gilmore, & Millones, Marco. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407–4429.

Pontius, Robert G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing, 66*(8), 1011–1016.

Pony, Olivier, Descombes, Xavier, & Zerubia, Josiane. (2000). Classification d'images satellitaires hyperspectrales en zone rurale et périurbaine.

Potts, Renfrey Burnard. (1952). Some generalized order-disorder transformations *Proceedings of the Cambridge Philosophical Society* (Vol. 48, pp. 106–109).

Pu, Ruiliang, & Gong, Peng. (2000). Band selection from hyperspectral data for conifer species identification. *Geographic Information Sciences, 6*(2), 137–142.

Pudil, Pavel, Novovicová, Jana, & Kittler, Josef. (1994). Floating search methods in feature selection. *Pattern recognition letters, 15*(11), 1119–1125.

Ramasso, E., Valet, L., & Teyssier, S. (2005). Classification fusion by decision templates for insulating part control *Information Fusion, 2005 8th International Conference on* (Vol. 1, pp. 6).

Reddy, D. Raj. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE, 64*(4), 501–531.

Reibman, A. R., & Nolte, L. W. (1987). Optimal Detection and Performance of Distributed Sensor Systems. *IEEE Transactions on Aerospace and Electronic Systems, AES-23*(1), 24-30.

Rémy, J. C., & Marin-Laflèche, A. (1976). L'entretien organique des terres. Coût d'une politique de l'humus. *Entreprises Agricoles, 4*, 63–67.

Rijsbergen, C. J. (1979). *Information Retrieval (Second ed.)* (Vol. 2). London, UK %7 Butterworths.

Roli, Fabio, & Fumera, Giorgio. (2001). Support vector machines for remote sensing image classification *Europto Remote Sensing* (pp. 160–166).

Rumelhart, David E., Hintont, Geoffrey E., & Williams, Ronald J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Russel, Bertrand, & Whitehead, Alfred North. (1925). Principia mathematica. *Volume, 1*(2), 3.

Saporta, Gilbert. (2006). *Probabilités, analyse des données et statistique*: Editions Technip.

Serpico, Sebastiano B., & Bruzzone, Lorenzo. (2001). A new search algorithm for feature selection in hyperspectral remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on, 39*(7), 1360–1367.

Shafer, Glenn. (1976). *A mathematical theory of evidence* (Vol. 1): Princeton university press Princeton.

Sheeren, David, Masse, Antoine, Ducrot, Danielle, Fauvel, Mathieu, Collard, Fanny, & May, Stéphane. (2012). Remote sensing of the forest and tree network in agricultural landscapes. Evaluation of the potentials of very high resolution multi-angular imagery. *International Journal of Geomatics and Spatial Analysis, 22*(4), 539-533.

Siegel, Sidney, & Castellan, N. John. (1988). Non parametric statistics for the behavioral sciences.

Smarandache, Florentin. (1991). *Only problems, not solutions!* : American Research Press.

Smarandache, Florentin, & Dezert, Jean. (2006). *Advances and applications of DSmT for information fusion: collected works* (Vol. 2): American Research Press.

Smets, Philippe. (1999). Practical Uses of Belief Functions. In K. B. Laskey & H. Prade (Eds.), *UAI* (pp. 612-621): Morgan Kaufmann.

Smits, P. C., Dellepiane, S. G., & Schowengerdt, R. A. (1999). Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. *International Journal of Remote Sensing*, *20*(8), 1461–1486.

Solberg, A. H. Schistad, Jain, Anil K., & Taxt, Torfinn. (1994). Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. *Geoscience and Remote Sensing, IEEE Transactions on*, *32*(4), 768–778.

Stearns, Stephen D. (1976). On selecting features for pattern classifiers *Proceedings of the 3rd International Joint Conference on Pattern Recognition* (pp. 71–75).

Swain, Philip H. (1978). Ba scan Classification.

Tinsley, Howard E., & Weiss, David J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*(4), 358.

Tupin, Florence, Bloch, Isabelle, & Maître, Henri. (1999). A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors. *IEEE T. Geoscience and Remote Sensing*, *37*(3), 1327-1343.

Vapnik, Vladimir. (1998). *Statistical learning theory. 1998*: Wiley, New York.

Voronoï, Georges. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik, 134*, 198–287.

Wald, Lucien. (1999). Some terms of reference in data fusion. *IEEE T. Geoscience and Remote Sensing, 37*(3), 1190-1193.

Wang, Fangju, Hall, G. Brent, & Subaryono. (1990). Fuzzy information representation and processing in conventional GIS software: database design and application. *International Journal of Geographical Information Science*, *4*(3), 261-283.

Wilkinson, Graeme G. (2005). Results and implications of a study of fifteen years of satellite image classification experiments. *Geoscience and Remote Sensing, IEEE Transactions on*, *43*(3), 433–440.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338 - 353.

Zadeh, L. A., Anvari, M., Bloch, I., & Beasse, L. (1998). Applications of Fuzzy Logic in Automatic Target Recognition. Final report, grant N00014-96-1-0556, Office of Naval Research: University of California, Berkeley, and Ecole Nationale Supérieure des Télécommunications.

Zadeh, Lotfi A. (1968). Fuzzy Algorithms. *Information and Control, 12*(2), 94-102.

Zhang, Junping, Zhang, Ye, & Zhou, Tingxian. (2001). Classification of hyperspectral data using support vector machine *Image Processing, 2001. Proceedings. 2001 International Conference on* (Vol. 1, pp. 882–885).

Zhang, X., & Niu, Z. (2009). Study on Decision Classification Fusion Model *Hybrid Intelligent Systems, 2009. HIS'09. Ninth International Conference on* (Vol. 3, pp. 107–109).

Références de l'auteur

Ducrot, D., Masse, A., Marais-Sicre, C., & Dejoux, J. F. (2010). Multisensor and multitemporal image fusion methods to improve remote sensing image classification: RAQRS'III.

Ducrot, D., Masse, A., & Ncibi, A. (2012). Hedgerow detection in HRS and VHRS images from different source (optical, radar) Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International (pp. 6348 -6351).

Ducrot, D., Masse, A., Ceschia, E., Marais-Sicre, C., & Krystof, D. (2010). A methodology for the detection of land cover changes: application to the Toulouse southwestern region. Paper presented at the Remote Sensing.

Masse, Antoine, Ducrot, Danielle, & Marthon, Philippe. (2011b). Tools for multitemporal analysis and classification of multisource satellite imagery Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the (pp. 209–212).

Masse, Antoine, Ducrot, Danielle, & Marthon, Philippe. (2012). Evaluation of supervised classification by class and classification characteristics. Proceedings of SPIE, 8390, 83902R.

Sheeren, David, Masse, Antoine, Ducrot, Danielle, Fauvel, Mathieu, Collard, Fanny, & May, Stéphane. (2012). Remote sensing of the forest and tree network in agricultural landscapes. Evaluation of the potentials of very high resolution multi-angular imagery. International Journal of Geomatics and Spatial Analysis, 22(4), 539-533.

Liste des figures

Liste des tableaux

Annexes

Sommaire de l'annexe

Annexes

1. Données utilisées pour les applications

1.1. Récapitulatif des données utilisées

Le tableau 0.1 résume toutes les images utilisées au cours de ce manuscrit.

Satellite	Année	Nombre de dates	Taille de chaque image (en pixels)	Total nombre d'images utilisées
Formosat	2006	17	18 millions	68
	2007	10	18 millions	40
	2008	9	18 millions	36
	2009	16	18 millions	64
	2010	10	18 millions	40
Spot	1993	2	16 millions	6
	1997	1	16 millions	3
	2006	3	16 millions	9
	2007	6	16 millions	19
	2008	6	16 millions	20
	2009	5	16 millions	18
	2010	10	65 millions	36
	2011	11	16 millions	39
Landsat	2010	3	98 millions	15
Pléiades	2012	2	1,6 milliards	8
Worldview	2010	1 (stereo)	100 millions	8
Terrasar	2009	5	20 millions	5
ERS	1993	6	16 millions	18
Radarsat	1997	4	16 millions	4

tableau 0.1 : Récapitulatif des données utilisées lors des applications de ce manuscrit

1.2. Récapitulatif des caractéristiques des satellites utilisés

Les caractéristiques des satellites dont les images ont été utilisées lors des applications sont résumées aux tableaux suivants (tableau 0.2 et tableau 0.3).

Nom satellite	Landsat 5	Landsat 7	Spot 2	Spot 4	Spot 5
Famille	Landsat	Landsat	Spot	Spot	Spot
Nationalité	USA	USA	France	France	France
Début mission	1984	1999	1990	1998	2002
Fin mission	2013		2009	2013	
statut	dégradé	actif	achevé	spot4 take 5	actif
Masse	1961	2200			3000
Instruments	MSS et TM : radiomètre	ETM+ : radiomètre	HRV	HRVIR, Pastel, Vegetation-1	HRG, Vegetation-2, HRS
Bandes	0,45-0,52	0,45-0,52			
Longueur	0,52-0,6	0,53-0,61	0,50-0,59	0,50-0,59	0,50-0,59
d'ondes	0,63-0,69	0,63-0,69	0,61-0,68	0,61-0,68	0,61-0,68
(µm)	0,76-0,9	0,78-0,9	0,78-0,89	0,78-0,89	0,78-0,89
	1,55-1,75	1,55-1,75		1,58-1,75	1,58-1,75
	2,08-2,35	2,09-2,35			
Résolution (m)	30	30	20	20	10 (20 MIR)
IR	10,4-12,5	10,4-12,5			
Résolution (m)	120	100			
Pan		0,52-0,9	0,50-0,73	0,61-0,68	0,48-0,71
Résolution (m)		15	10	10	5
Super Pan					0,48-0,71
Résolution (m)					2,5
Altitude (km)	705	705	832	832	832
Cycle (jours)	16	16	26	26	26
Heure	9h30-10h	10h-10h15			
Emprise (km x km)	180x172	180x172	60x60	60x60	60x60

tableau 0.2 : Caracteristiques des familles de satellites Landsat et Spot

Nom satellite Famille	Formosat-2	Sentinel-2 Sentinel	Pleiades-1ab Pléiades	Worldview-2	Terrasar-X
Nationalité	Taiwan	Europe	France	USA	Allemagne
Début mission fin mission statut	2004 actif	 prévu	2011 actif	2009 actif	2007 actif
Masse	764	1200	980	2800	1230
Bandes (µm)	0,45-0,52 0,52-0,60 0,63-0,69 0,76-0,90	0,433-0,453** 0,4575-0,5225 0,5425-0,5775 0,65-0,68 0,6975-0,7125* 0,7325-0,7475* 0,773-0,793* 0,7845-0,8995 0,855-0,875* 0,935-0,955** 1,360-1,39** 1,565-1,655* 2,1-2,28*	0,43-0,55 0,49-0,61 0,6-0,72 0,75-0,95	0,4-0,45 0,45-0,51 0,51-0,58 0,585-0,625 0,63-0,69 0,705-0,745 0,77-0,895 0,86-1,04	
Résolution (m)	8	10 / 20*/ 60**	2,8	1,85	1 (3 / 18)
Pan Résolution (m)	0,45-0,90 2		0,48-0,83 0,7	0,45-0,8 0,46	
Altitude (km)		786	694	770	514
Cycle (jours)	5	5	26	1,1	11
Heure				10h30	
Emprise (km x km)	24x24	290x290	20x20	96x110 48x110 stéréo	10x5 / 30x50 / 100x150

tableau 0.3 : Caracteristiques des satellites Formosat-2, Sentinel-2, Pleiades, Worldview-2 et Terrasar-X

1.3. Les données Formosat-2

La figure 0.1 illustre les histogrammes des dates utilisées pour les applications Formosat-2 des années 2006 à 2010.

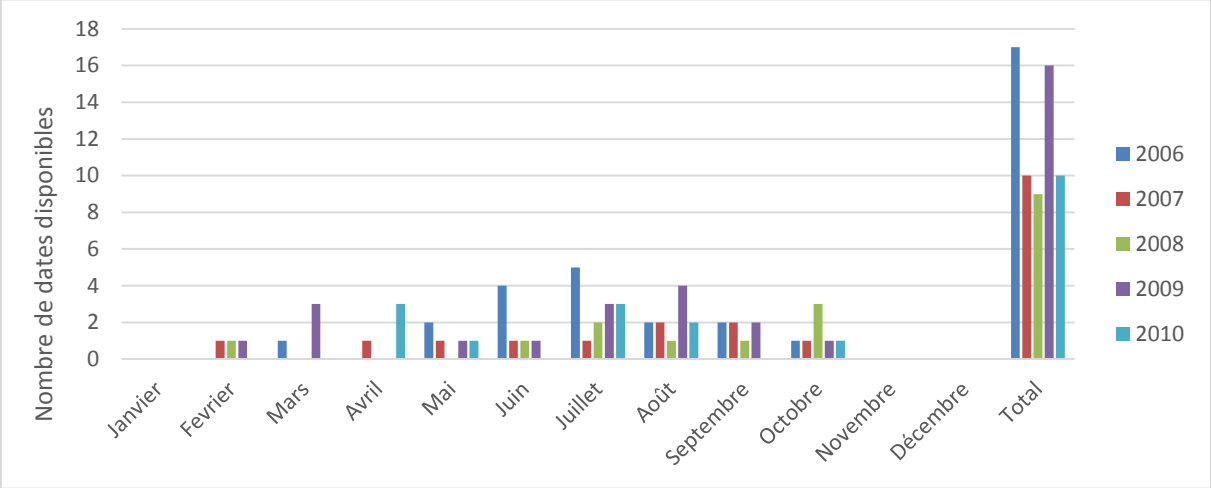


figure 0.1 : Histogramme des images Formosat-2 utilisées

Les images des résultats finaux de classification sont présentés (figure 0.2 à figure 0.6).

Le tableau 0.4 fournit les classes thématiques pour chaque année ainsi que la superficie des vérités terrain considérée pour l'apprentissage et pour la validation.

Le tableau 0.5 et le tableau 0.6 fournissent les performances de chaque résultat final de classification par année (résultat du système de Sélection-Classification-Fusion).

Les tableaux (0.7 à 0.9) fournissent respectivement la matrice de confusion classique, la matrice de confusion avec pondération de probabilités de la méthode Maximum de Vraisemblance et la matrice de similitude.

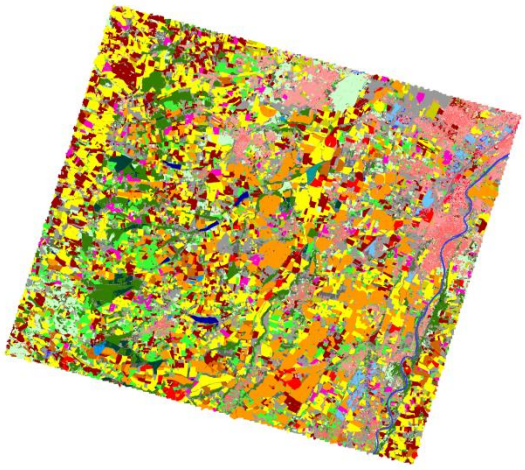


figure 0.2 : Classification finale année 2006

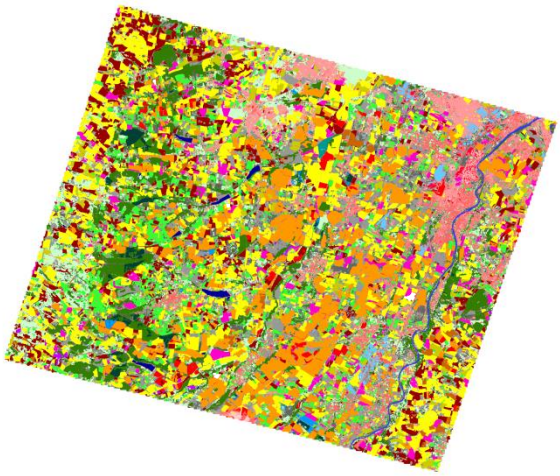


figure 0.3 : Classification finale année 2007

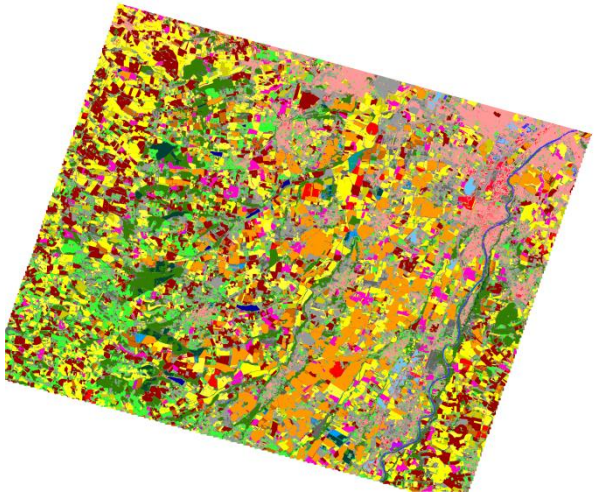


figure 0.4 : Classification finale année 2008

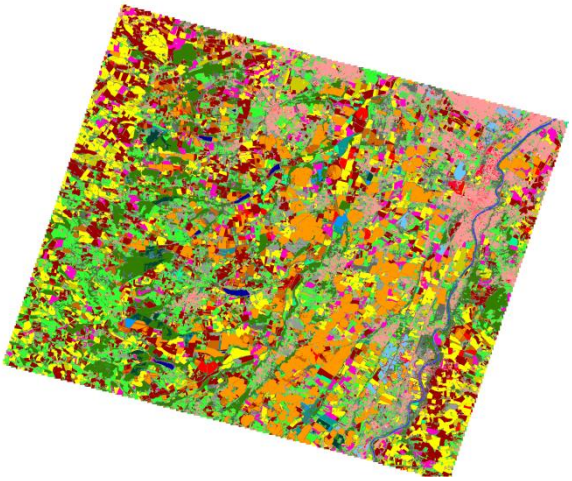


figure 0.5 : Classification finale année 2009

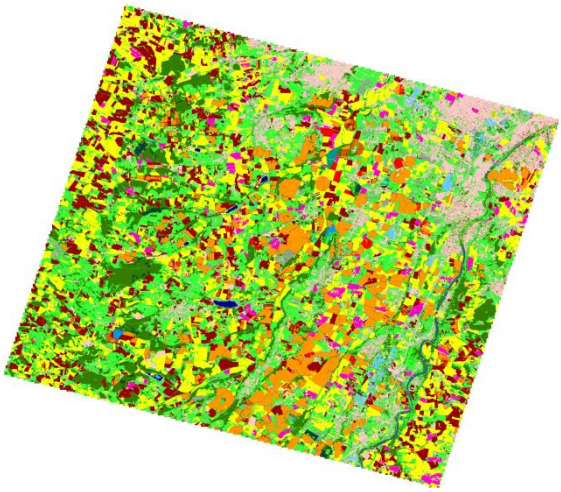


figure 0.6 : Classification finale année 2010

Surface des classes (Ha)	2006		2007		2008		2009		2010	
	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.
Bâti dense	3,90	3,90	3,55	3,55	0,90	0,89	0,90	0,89	1,25	1,25
Bâti diffus	3,36	3,36	2,75	2,75	1,77	1,77	1,77	1,77		
Bâti ind./surf min.	2,22	2,21	1,58	1,58	0,89	0,89	0,89	0,89	6,98	6,98
Blé dur	961,97	961,97	212,27	212,27	461,34	461,34	206,13	206,13	128,49	128,49
Blé tendre					542,06	542,05	463,16	463,16		
Chanvre							42,46	42,46	6,85	6,85
Colza	176,04	176,04	88,73	88,72	295,56	295,56	291,94	291,94	39,44	39,44
Eau libre	30,57	30,57	18,58	18,57	29,18	29,17	21,89	21,89	7,58	7,58
Eucalyptus	15,27	15,26	26,28	26,27	27,16	27,16	20,23	20,23	2,07	2,07
Eucalyptus 2							17,79	17,79		
Feuillus	108,13	108,13	94,59	94,59	108,37	108,37	107,96	107,96	14,98	14,98
Friche	5,83	5,83	3,53	3,52	4,21	4,21	32,02	32,01	47,42	47,41
Gravières	66,41	66,41	46,99	46,98	41,39	41,38	41,39	41,38	2,11	2,10
Jachères	541,41	541,41	51,89	51,89	242,13	242,12	278,69	278,69	13,07	13,07
Lac	43,71	43,71	45,89	45,89	31,99	31,99	32,30	32,29	0,23	0,23
Mais	237,20	237,20	187,58	187,58	542,11	542,11	518,65	518,65	78,36	78,35
Mais ensilage	61,25	61,24			53,48	53,48	29,15	29,15	8,81	8,81
Mais non irrigue					13,91	13,90	5,27	5,26		
Orge	147,09	147,09	6,22	6,22	204,34	204,33	214,75	214,74	5,34	5,34
Peupliers	15,21	15,21	13,20	13,20	15,21	15,21	23,11	23,10		
Pois	45,82	45,82	2,27	2,27			33,18	33,18		
Prairie perm.	267,58	267,57					261,75	261,74		
Prairie temp.	215,73	215,73	137,06	137,06	306,13	306,13	401,43	401,43	0,42	0,41
Résineux	41,22	41,22	32,66	32,66	42,11	42,11	20,81	20,81	1,22	1,22
Soja	56,97	56,97	2,43	2,43	88,60	88,60	136,76	136,76	7,42	7,42
Sorgho	72,73	72,73	6,55	6,55	59,32	59,32	91,57	91,57	2,44	2,43
Surface min.	1,46	1,45			1,75	1,75	1,75	1,75		
Tournesol	476,84	476,84	18,78	18,77	573,88	573,88	796,26	796,26	28,01	28,01

tableau 0.4 : Classes utilisées pour les différentes années avec la superficie des vérités terrain associées en Ha

	2006	2007	2008	2009	2010
AA	94,02	90,43	91,71	94,4	96,01
AP	87,04	84,29	85,58	91,48	96,08
OA	92,23	92,82	88,46	92,36	97,06
Kappa	0,91	0,92	0,87	0,92	0,96
AOCI	82,53	78,12	78,58	86,52	92,19
F1	0,9	0,87	0,89	0,93	0,96

tableau 0.5 : Performances globales des résultats finaux des différentes années

Valeur OCI pour la classe :	2006	2007	2008	2009	2010
Bâti dense	54,74	51,04	85,27	86,33	
Bâti diffus	49,43	58,68	25,96	43,77	
Bâti industriel/surf minérale	52,14	91,30	82,07	92,09	76,77
Blé dur	94,40	88,83	64,20	84,27	96,69
Blé tendre			60,81	82,97	
Chanvre				97,13	
Colza	95,25	93,95	95,05	95,91	98,73
Eau libre	98,03	95,11	92,67	97,82	100,00
Eucalyptus	92,67	94,28	72,62	96,34	100,00
Eucalyptus 2				97,20	
Feuillus	95,61	96,69	95,80	95,97	99,96
Friche	8,42	33,85	13,51	56,97	
Gravières	99,55	98,15	99,64	99,54	100,00
Jachères	69,24	57,48	58,63	63,25	66,85
Lac	99,34	98,41	99,76	99,90	100,00
Mais	97,99	97,53	96,88	98,25	99,67
Mais ensilage	98,65		96,58	93,92	99,71
Mais non irrigue			80,91	92,63	
Orge	83,77	92,74	78,56	83,27	87,85
Peupliers	98,15	97,48	97,26	96,15	100,00
Pois	97,55	98,03		94,25	
Prairie permanente	66,66			54,87	99,81
Prairie temporaire	63,31	71,29	59,00	65,80	80,37
Résineux	97,58	97,26	83,96	97,32	100,00
Soja	94,11	96,05	97,43	97,92	99,83
Sorgho	95,74	97,85	64,69	82,51	98,95
Surface minérale	83,77		88,82	82,82	44,44
Tournesol	94,57	79,24	95,91	93,24	94,14

tableau 0.6 : Performances par classe des résultats finaux des différentes années

tableau 0.7 : Matrice de confusion en effectif de la classification des données Formosat-2 année 2009 par la méthode ICM

	bâti dense	bâti/surf minérale	bâti diffus	blé dur	blé tendre	chanvre	colza	eau libre	eucalyptus	eucalyptus 2	feuillus	friche	gravières	jachère/gel	lac	mais	mais ensilage	mais non irrigue	orge	peupliers	pois	prairie temporaire	prairie permanente	pré/prairie	résineux	soja	sorgho	surface minérale	tournesol	Somme vérification
bâti dense	**139**																													139
bâti/surf min		**139**																												139
bâti diffus	3		**25**																											28
blé dur			6	**3123**	412		65					9		145					111			15					86	31	9	4012
blé tendre			14	1419	**19129**		541					39		723					1159			11					41		25	23101
chanvre						**4418**						1		28		1											37		27	4512
colza			32	34	22		**243**					15		776					43			4					16		2	1187
eau libre			4					**3323**						2		2												2	12	3345
eucalyptus									**378**	2	5	42		2						4					25					458
eucalyptus 2										**2738**	2	23		11																2774
feuillus			3						15		**16395**	342		1						79		1			2				1	16839
friche			34	2			3		24		66	**4813**		21						11		4	1				7		13	4999
gravières			18										**6448**																	6466
jachère/gel			117	186	6	42	198		63	141	12	6787		**2256**		15			72	93		948	168				687	15	219	12025
lac							3								**542**															545
mais														13		**6589**	242	6								26	1743		181	8800
mais ensilage			1													1	**454**										4		4	464
mais non irri																12		**426**									13		2	453
orge			19	2149	212		24					39		562					**12611**			6	1				111		78	139
peupliers											17	2		8						**3557**		1					3		1	3589
pois			1	61			2		1			26		85					18		**626**	2					127		3	952
prairie temp			152	39			415		12	79	137	2646		525					183	12		**19654**	49				99	6	265	24273
prairie perm	21	7	378	3			29		4	11	5	186		2996					4	11		2754	**5852**				28	63	16	12368
résineux			8						15		41	14		8		1									**316**					403
soja														4		76										**877**	552		97	1606
sorgho			4	96	78						1	14		19		61		1				1				1	**13843**		85	14204
surface min																												**272**		272
tournesol			64	155	12	122	25			6		38		723		6	96				6	19				6	621	58	**5767**	64613
Somme classif	163	146	880	7267	19871	4582	1548	3323	512	2977	16681	15036	6448	8908	542	6764	792	433	14201	3767	632	23420	6071	343	910	18018	447	6807	171489

tableau 0.8 : Matrice de confusion avec pondération de probabilité de la classification des données Formosat-2 année 2009 par la méthode ICM

	bâti dense	bâti/surf minérale	bâti diffus	blé dur	blé tendre	chanvre	colza	eau libre	eucalyptus	eucalyptus 2	feuillus	friche	gravières	jachère/gel	lac	mais	mais ensilage	mais non irrigue	orge	peupliers	pois	prairie temporaire	pré/prairie permanent	résineux	soja	sorgho	surface minérale	tournesol	Somme vérification
bâti dense	4,0																												4,0
bâti/surf min		2,0																											2,0
bâti diffus	0,1		5,8																										5,9
blé dur				1214	15,6		1,5					0,4		4,8					3,5			0,5				2,1	0,6	0,2	1243,2
blé tendre			0,4	558,2	756		12,6					1,5		25,2					37,9			2,8				8,8		0,4	1403,8
chanvre						187,1						0,0		0,9												1,0		0,6	189,6
colza			0,4	10,3	0,7		825,7					0,4		27,0					1,4			1,3				2,5		0,3	870,0
eau libre			0,1					242,0																				0,1	242,2
eucalyptus									198,7	0,1	0,2	1,6		0,1						0,2				1,7					202,6
eucalyptus 2										166,1	0,1	0,8		0,4															167,4
feuillus			0,1						0,6		1124,5	14,0		0,4						3,5				1,0					1144,1
friche			0,7				0,1		1,3		3,3	216,7		0,8						0,5		0,1	0,0			0,2		0,2	223,8
gravières			0,1										612,2																612,4
jachère/gel			2,7	4,3	0,2	0,5	5,1		3,4	7,3	0,5	326,2		837,0		0,5			2,0	4,8		36,2	8,1			19,3	0,5	4,8	1263,5
lac															398,8														398,8
mais																2783,3	5,6	0,2							0,5	47,2		3,7	2840,4
mais ensilage																0,0	185,3									0,1		0,1	185,5
mais non irri																0,4		19								0,3		0,1	19,8
orge			0,6	61,9	7,3		6,0					13,5		22,9					470,9			2,1	0,0			2,7		1,4	589,4
peupliers											0,7	0,7		0,3						220,0		0,0				0,1		0,0	221,8
pois			0,0	1,4			0,6					1,2		3,9					0,2		31,2					3,6		0,8	43,0
prairie temp			3,4	10,7			11,2		0,5	4,6	6,6	110,6		220,4					5,8	0,8		781,6	19,6			24,8	0,3	5,7	1206,5
prairie perm		0,1	8,1	1,0	0,0		0,5		0,3		2,0	68,4		104,4					0,1	0,7		100,8	278,3			0,5	1,4	0,4	567,1
résineux			0,2						0,8		2,1	0,4		0,2		0,0								232,3					236,2
soja														0,1		2,8									340,7	13,7		2,1	359,5
sorgho				4,4	3,0						0,0	0,3		0,6		2,2		0,1				0,0			0,0	498,6		2,4	511,6
surface min																											9,5		9,5
tournesol			1,1	6,1	0,4	3,2	0,6					1,2		25,6			1,7					0,5			0,2	185,6	1,4	1691,9	1919,5
Somme clas.	4,1	2,1	23,7	1872,2	783,3	190,8	863,9	242,0	205,6	178,0	1140,1	758,1	612,2	1275,1	398,8	2789,2	192,6	19,2	521,7	230,5	31,2	926,1	306,1	235,1	341,4	810,8	13,7	1715,1	16682,8

	bâti dense	bâti/surf minérale	bâti diffus	blé dur	blé tendre	chanvre	colza	eau libre	eucalyptus	eucalyptus 2	feuillus	friche	gravières	jachère/gel	lac	mais	mais ensilage	mais non irrigue	orge	peupliers	pois	prairie temporaire	pré/prairie permanente	résineux	soja	sorgho	surface minérale	tournesol
bâti dense	100,0																											
bâti/surf min		100,0																										
bâti diffus			100,0																									
blé dur				40,3	50,7									0,4				0,1	8,6									
blé tendre				32,9	62,5									0,6				0,1	4,0									
chanvre						51,0										12,2	0,1	0,2			0,6				12,6	11,5		11,7
colza				0,3	0,4		98,5												0,6			0,1						
eau libre								100,0																				
eucalyptus									99,4																			
eucalyptus 2										94,6										5,3								
feuillus											99,7	0,3																
friche												99,0								1,0								
gravières													100,0															
jachère/gel			0,3	1,4	4,6									84,1				5,3	1,4			1,8	0,4		0,5			0,2
lac															100,0													
mais			0,2			8,2										78,6	2,5	0,4							6,5	0,5		3,3
mais ensilage						4,5								0,1		39,0	13,0	7,6			0,2				1,8	2,6		31,1
mais non irri														2,0			0,3	97,3							0,4			
orge				13,9	25,5									0,5				0,4	59,2						0,3			0,1
peupliers												15,2		1,8				0,2		82,1		0,3	0,4					
pois						4,2										0,1					27,6				4,0	30,3		33,9
prairie temp														0,1						0,1		50,1	49,6					
prairie perm	0,0	0,0						0,0					0,0		0,0					28,1		8,0	63,9					
résineux																								100,0				
soja						0,5								0,7			1,4	11,4	0,2		0,4				61,4	1,8		22,2
sorgho			0,3			5,3								0,1		0,4	0,8	3,1			0,5				35,2	25,0		29,5
surface min																											100,0	
tournesol			0,3			2,6										0,3	1,0	0,2			2,2				27,2	10,9		55,4

tableau 0.9 : Matrice de similitude de la classification des données Formosat-2 année 2009 par la méthode ICM

1.4. Les données Spot-2/4/5

La figure 0.7 illustre les histogrammes des dates utilisées pour les applications Formosat-2 des années 2006 à 2010.

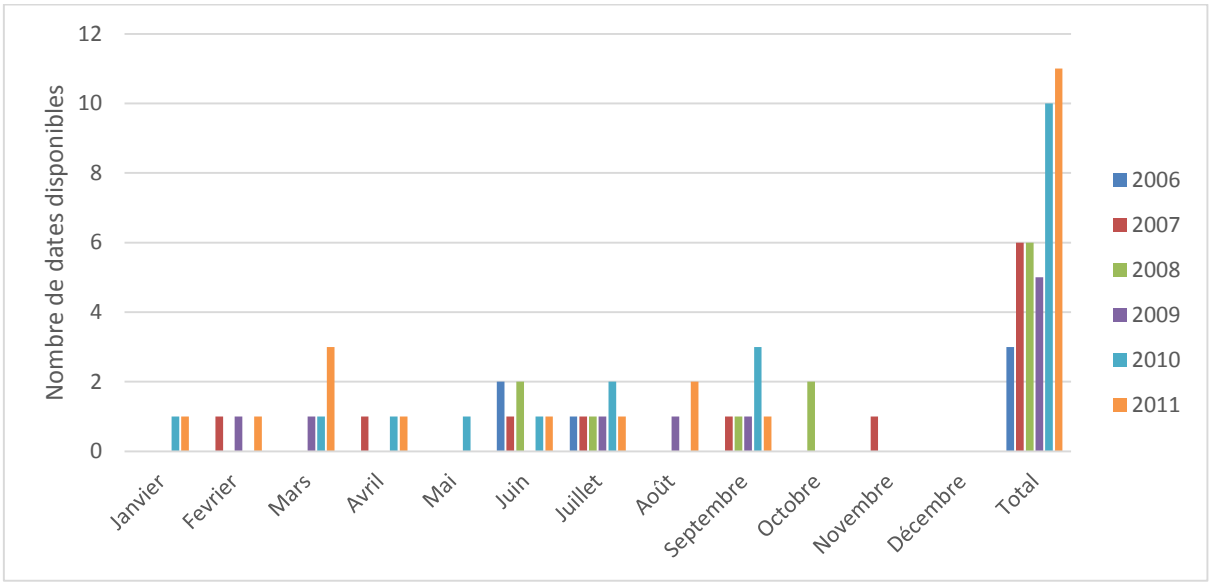


figure 0.7 : Histogramme des images Spot utilisée

Les résultats finaux sont présentés (figure 0.8 à figure 0.13)

Le tableau 0.10 fournit les classes thématiques pour chaque année ainsi que la superficie des vérités terrain considérée pour l'apprentissage et pour la validation.

Les 0.11 et 0.12 fournissent les performances de chaque résultat final de classification par année (résultat du système de Sélection-Classification-Fusion).

Superficie des classes (Ha)	2006		2007		2008		2009		2010		2011	
	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.	Appr.	Valid.
autres gels			2,32	2,28					205,87	205,87		
bâti dense	4,60	4,56	2,48	2,44	18,68	18,68	0,76	0,72	3,36	3,36	18,68	18,68
bâti diffus	9,36	9,32	7,96	7,92	19,64	19,60	0,88	0,84	1,58	1,57	19,64	19,60
bâti industriel	6,40	6,36	20,60	20,60	32,48	32,44	10,32	10,32	2,52	2,52	25,64	25,64
blé dur	1688,24	1688,20	521,36	521,32	1354,64	1354,60	78,80	78,76	766,52	766,52	488,00	488,00
blé tendre			735,68	735,64	1116,48	1116,48			1291,06	1291,06		
chanvre									3,93	3,92	7,72	7,72
colza	188,76	188,76	228,92	228,92	433,60	433,56	68,80	68,80	322,21	322,21	143,64	143,60
eau libre	74,24	74,20	36,40	36,40	82,60	82,60	61,56	61,52	38,07	38,06	39,96	39,96
eucalyptus	28,28	28,24	27,24	27,24	27,32	27,28	22,48	22,44	11,56	11,56	23,76	23,72
feuillus	449,96	449,92	130,36	130,32	164,12	164,12	98,12	98,08	122,01	122,00	194,92	194,88
friche	5,80	5,80	6,76	6,72			5,76	5,72	1,40	1,40		
gravière	40,28	40,24	86,88	86,84	50,36	50,36	41,12	41,08	25,21	25,20	62,96	62,92
gravière ancienne									2,41	2,41		
jachère/surf. gel	79,00	78,96	98,08	98,04	16,28	16,24	27,60	27,60				
lac	86,60	86,56	65,68	65,64	44,56	44,52	80,32	80,28	48,89	48,89	44,56	44,52
mais	578,68	578,68	127,44	127,40	722,04	722,04	87,32	87,32	466,90	466,90	368,12	368,12
orge	160,08	160,08	146,16	146,16	263,80	263,76	2,36	2,32	211,83	211,82	50,80	50,80
peuplier	15,16	15,16	15,20	15,20	15,20	15,20	18,08	18,04	14,62	14,61	32,48	32,44
pois			27,48	27,48	52,12	52,08	4,04	4,00			24,28	24,28
prairie perm.	77,20	77,20	58,56	58,52	24,84	24,80	13,92	13,92	401,56	401,55	136,32	136,32
prairie temp.	138,68	138,64	86,72	86,68	17,12	17,08	16,80	16,76	48,52	48,52	259,40	259,36
prairie temp. 2			207,32	207,32			48,44	48,40				
résineux	92,52	92,48	74,32	74,32	84,72	84,68	43,00	43,00	42,21	42,20	41,72	41,72
soja	30,00	29,96	40,68	40,64	88,24	88,20	9,08	9,08	20,82	20,82	34,40	34,36
sorgho	34,72	34,72	12,72	12,68	102,24	102,20	2,60	2,60	46,54	46,54		
surface minérale	14,28	14,28	25,84	25,84	1,72	1,68	1,32	1,32	0,87	0,86	1,72	1,68
tournesol	670,60	670,56	86,04	86,00	1618,80	1618,80	52,08	52,08	684,18	684,18	569,24	569,24

tableau 0.10 : Classes utilisées pour les différentes années avec la superficie des vérités terrain associées (en Ha)

Indices (en % sauf Kappa et F1)	2006	2007	2008	2009	2010	2011
AA	81,67	89,17	89,90	82,90	95,46	92,47
AP	69,78	84,24	84,03	83,55	91,96	92,95
OA	92,26	87,32	88,68	88,22	89,52	93,47
Kappa	0,90	0,85	0,87	0,87	0,88	0,93
AOCI	61,93	76,02	75,57	69,77	88,00	86,07
F1	0,75	0,87	0,87	0,83	0,94	0,93

tableau 0.11 : Performances globales des résultats finaux des différentes années

Valeur OCI pour la classe :	2006	2007	2008	2009	2010	2011
Bâti dense	12,14	43,16	72,65	22,22	96,53	69,75
Bâti diffus	57,18	67,15	51,40	22,32	53,40	60,63
Bâti industriel	22,24	30,25	92,15	90,08	99,21	95,82
Blé dur	95,53	64,82	73,98	85,41	67,47	94,47
Blé tendre		75,20	65,23		72,52	
Chanvre					95,92	98,96
Colza	97,40	91,68	84,98	74,63	97,41	97,64
Eau libre	95,91	90,61	87,90	86,01	99,24	92,41
Eucalyptus	77,81	83,68	82,35	85,09	100,00	88,17
Feuillus	94,51	92,08	89,60	90,27	99,49	93,07
Friche	5,24	29,62		50,62	86,66	
Gravière ancienne					98,34	
Gravières	94,63	90,89	92,69	89,47	98,65	92,95
Jachère	41,07	89,47	77,90	48,34	72,00	
Lac	95,94	95,80	96,14	95,27	98,28	93,71
Mais	90,62	86,91	87,76	93,86	92,93	95,80
Orge	60,95	62,56	68,17	43,10	67,05	84,19
Peupliers	86,06	92,13	89,77	69,52	96,51	90,63
Pois		71,06	48,93	65,19		89,55
Prairie permanente	42,03	56,56	86,84	20,47	84,59	56,33
Prairie temp.	52,31	59,93	39,39	69,23	53,17	59,12
Prairie temp. 2				53,06		
Résineux	87,99	93,45	91,22	87,48	98,32	95,24
Soja	34,46	95,87	65,36	96,57	98,70	94,76
Sorgho	26,73	77,62	37,40	89,80	86,35	
Surface minérale	0,00	96,79	61,22	56,57	91,92	69,05
Tournesol	91,80	87,16	95,04	90,01	95,25	95,18

tableau 0.12 : Performances par classe des résultats finaux des différentes années

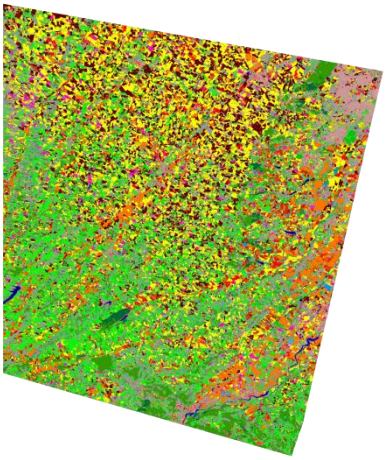


figure 0.8 : Classification finale année 2006

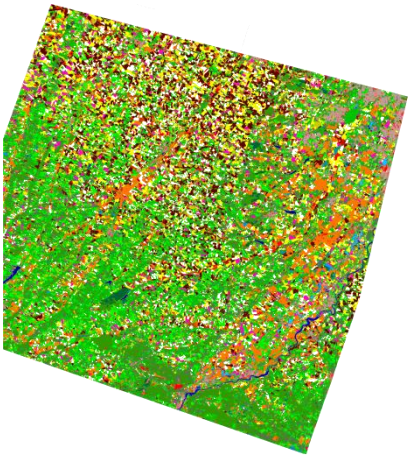


figure 0.9 : Classification finale année 2007

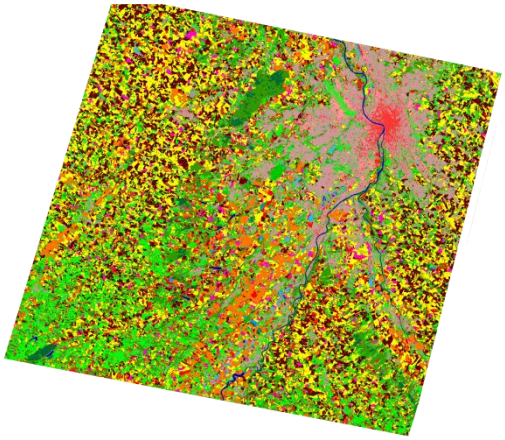


figure 0.10 : Classification finale année 2008

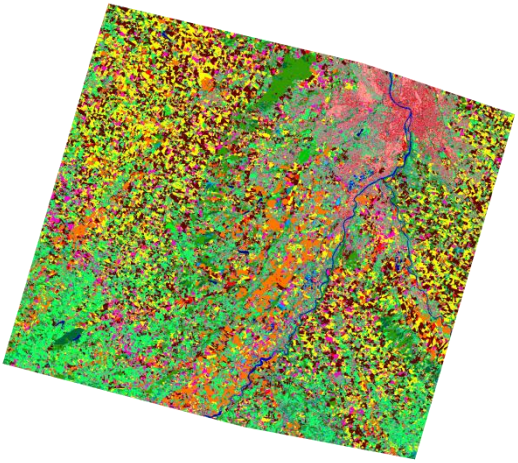


figure 0.11 : Classification finale année 2009

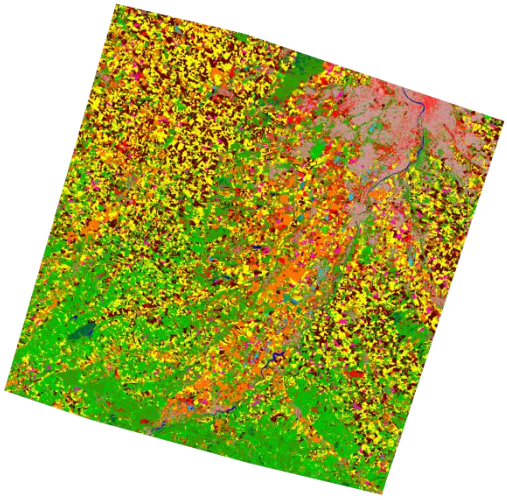


figure 0.12 : Classification finale année 2010

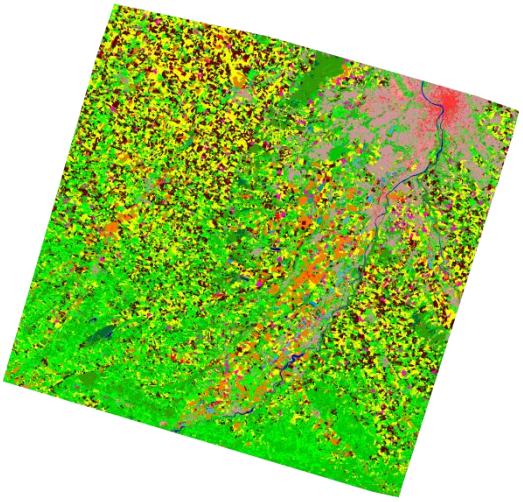


figure 0.13 : Classification finale année 2011

1.5. Les données Landsat-5/7

L'emprise des images Landsat utilisées est résumée à la figure 0.14

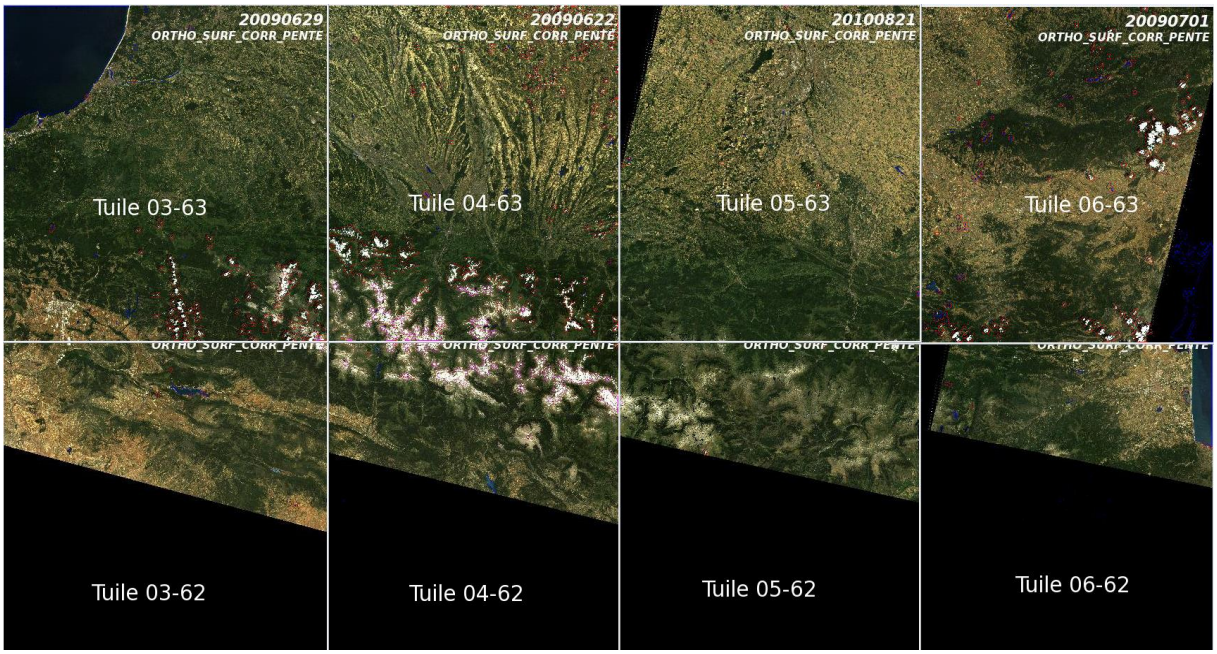


figure 0.14 : Emprise des données Landsat utilisées

La figure 0.15 présente le résultat final de l'occupation du sol des Pyrénées. Un regroupement de classes a été effectué afin de pouvoir visualiser facilement les différentes classes thématiques étudiées.

La liste des classes est présentée ci-dessous :

Classe thématique	Classes utilisées
Feuillus	Feuillus, Eucalyptus, Peupliers, Chêne, Châtaigner, Robinier
Résineux	Résineux, Pin Sylvestre, Pin maritime, Pin Douglas, Mélèze, Sapin pectiné, Pin noir
Mixte bois	Mixte feuillus/résineux, Foret ouverte
Végétation naturelle	Pelouse brune, Pelouse chlorophyllienne, Mixte pelouse (4), Landes à fougères, Landes buissonnante, Landes mixte, Landes et arbres dispersés, Landes et pelouse, Lande et roche, Friche montagne, Friche buissonnante, Garrigue parsemée, Garigue dense
Cultures	Blé dur, Blé tendre, Colza, Orge, Mais, Tournesol, Vigne, Jachères
Surfaces herbées	Friche, Prairie temporaire, Prairie permanente, Bocages
Eau	Eau libre, Lac, Océan, Gravières
Surface minérale	Bâti dense, Bâti diffu, Surface minérale, Sable, Roche nue, Calcaire nu, Grés rouge
Divers	Ombre, Neige

Les Pyrénées

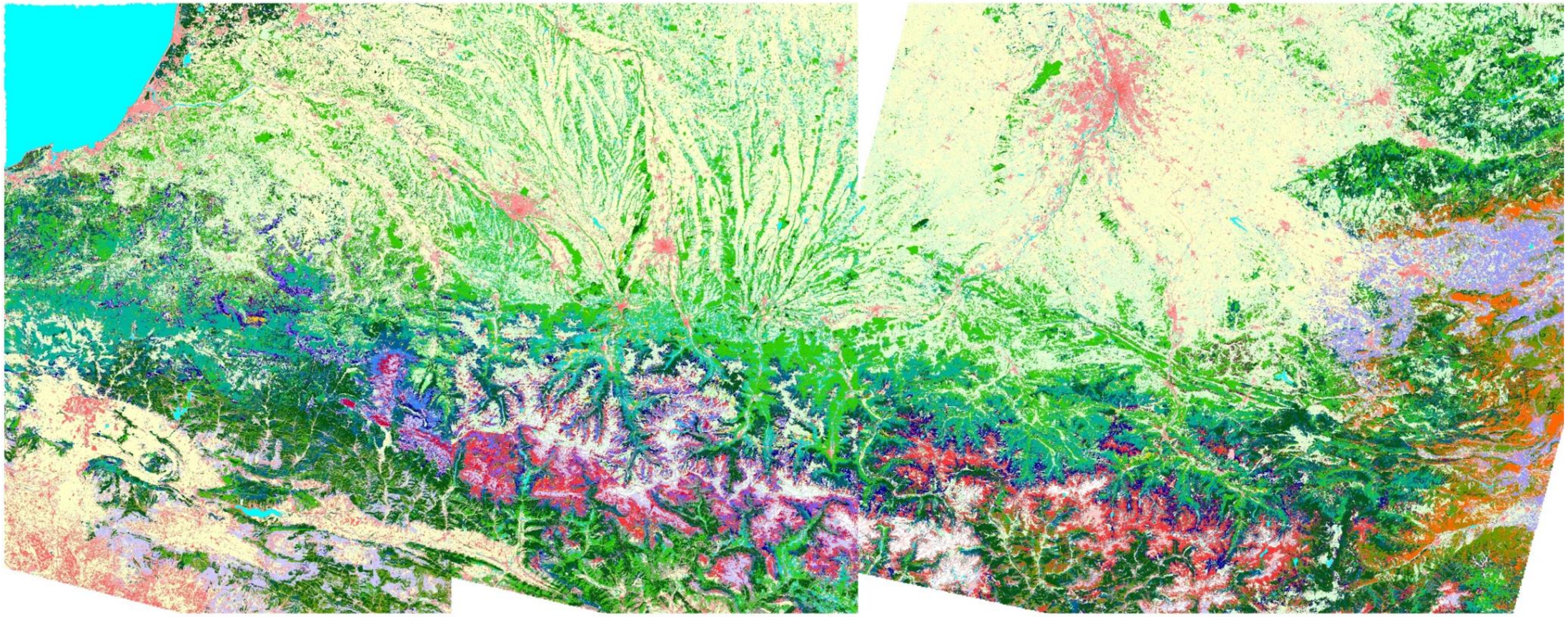


figure 0.15 : Occupation du sol des Pyrénées à partir d'images Landsat 5 et 7

1.6. Les données Pléiades-1

Des classifications supervisées ont pour la première fois été réalisées sur la zone Sud-Ouest à partir d'images Pléiades de résolution 70cm. Seulement deux dates sont disponibles : une image du 23/03/2012 (figure 0.16) et une autre du 09/11/2012 (figure 0.17). Les classifications ont été effectuées sur la zone commune aux deux images. Cette zone commune a pour taille 32786 lignes sur 14802 colonnes soit une superficie de 23780 hectares.

figure 0.16 : Composition colorée (fausses couleurs) de la zone commune : date de Mars 2012

figure 0.17 : Composition colorée (fausses couleurs) de la zone commune : date de Novembre 2012

2. Applications et collaborations

2.1. Apport de la télédétection en géologie

Conclusion des auteurs de ce travail réalisé en partenariat avec le Cesbio, auteurs Mme Hammad Nabila et Mme Boussada Nawel.

L'évolution de la géologie a largement bénéficié des techniques de télédétection et de traitement d'image. L'étendue des terrains d'étude, les conditions climatiques très rudes, et l'accès limité aux affleurements (reliefs escarpés, masqués par des recouvrements) et les durées de missions limitées, sont des problèmes courants lors des travaux de recherches et rendent souvent l'information géologique et structurale inaccessible, hétérogène et discontinue.

Pour pallier à ces problèmes, l'utilisation de données de télédétection aérienne et spatiale peut constituer une source d'information très appréciable qui apporte aux géologues, de toutes spécialités confondues, une information nouvelle et différente, qui élargit l'angle de vue autrefois limité par ces contraintes multiples.

En géologie, les images de télédétection montrent le terrain d'une manière globale, ce qui facilite l'identification de grandes structures (failles, limites lithologiques).

De ce fait, elle permet d'enrichir la cartographie géologique et notamment de détecter de nombreux linéaments, et de caractériser certaines lithologies.

Cependant, l'analyste doit toujours tenir compte des limites de cette technique, en relations avec les propriétés spectrales et spatiales des données dont il dispose, et de la nature du terrain sur lequel il travaille (présence de patine et recouvrement en domaine désertique ou de couvert végétal en contexte tropical et humide).

Il est à noter que, malgré l'apport non négligeable de cette technique, la télédétection reste une source d'information qui doit être vérifiée, validée et complétée par des missions de terrain.

Le résultat de cette étude est présenté en figure 0.18.

2.2. Cartographie d'une palmeraie aux environs d'Agadir, Maroc

Travaux réalisé dans le cadre d'un partenariat avec l'université Abno Zohr, Agadir.

Le résultat de cette étude est présenté en figure 0.19.

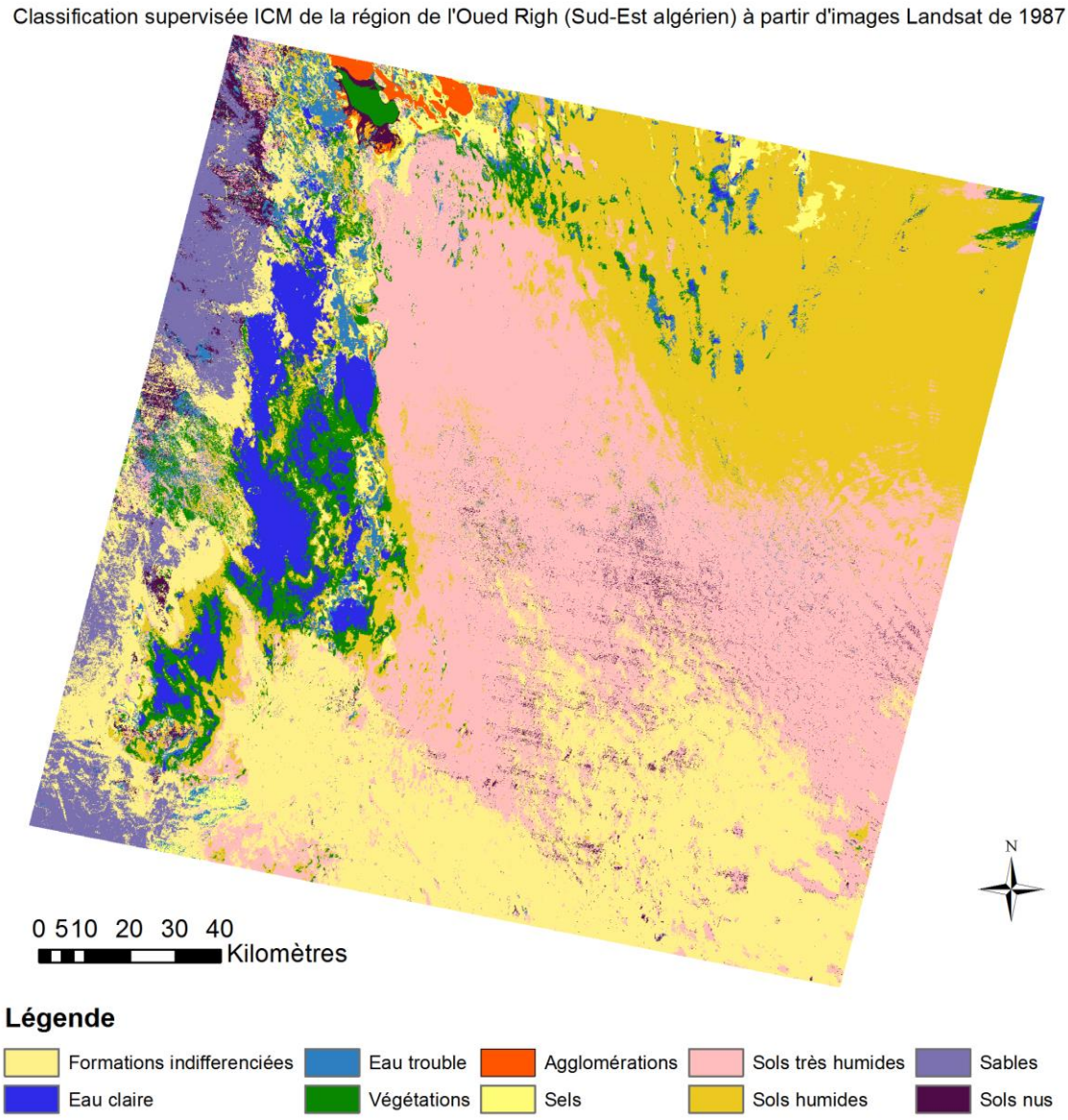


figure 0.18 : Résultat de l'application à la géologie de la classification d'images Landsat par la méthode ICM

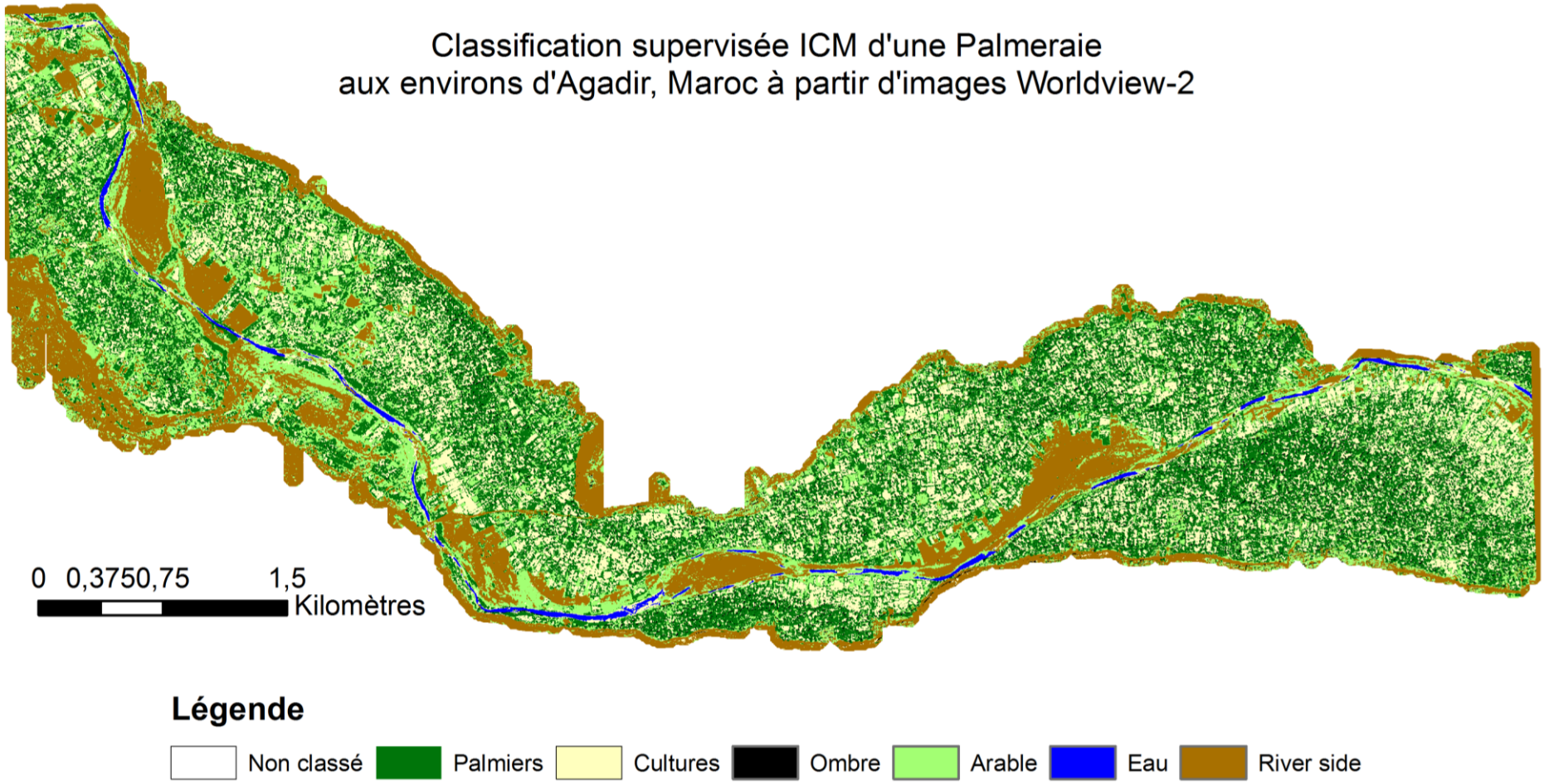


figure 0.19 : Résultat de la classification supervisée par méthode ICM d'une palmeraie aux environs d'Agadir, Maroc à partir d'images Worldview-2