
Carte de consensualité

Quirin Arnaud

*European Centre for Soft Computing - Edificio Científico-Tecnológico, 3ª Planta - CL.
Gonzalo Gutierrez Quiros S/N - 33600 Mieres, Spain
arnaud.quirin@softcomputing.es*

RÉSUMÉ. Aujourd'hui de nombreuses méthodes de segmentation et de classification existent, aussi bien supervisées que non supervisées. Dans ce papier, nous nous sommes intéressés à une technique simple permettant de proposer un résultat unique fondé sur une implication homogène de chaque méthode dans le résultat final sans masquer les défaillances individuelles de ces méthodes. Un mécanisme prenant en compte l'utilisation de plusieurs classifieurs pour obtenir un consensus au niveau du résultat final est décrit. Ce mécanisme utilise la notion d'entropie et permet de s'abstenir du choix délicat de la méthode de segmentation ou de classification. La principale qualité de cette technique est le fait qu'elle permet de compenser les défauts de chaque méthode utilisée, sans utiliser une technique d'échantillonnage quelconque pour identifier la pertinence de chaque méthode.

ABSTRACT. Currently, many segmentation and classification methods exist, in a supervised or an unsupervised way. In this paper, a new method allowing the user to solve the problem of unification of the unconsensual results coming up from various classifiers on the same set of data without masking the individual bias of the applied classifiers is presented. The method takes into account the use of several classifiers to reach a consensus for the final result. It will be shown how using the entropy is possible to abstain from the choice of a segmentation or a classification method. The main advantage of this method is that it makes possible to compensate the drawbacks of each used algorithm without using any sampling technique to identify the relevance of each one.

MOTS-CLÉS : carte de consensualité, élection, classification, images de télédétection.

KEYWORDS: consensual maps, election, classification, remote sensing images.

1. Introduction

Actuellement, la plupart des géographes manipulent de nombreux algorithmes de classification, de segmentation, de traitement d'images (pré-traitements, post-traitements des résultats, ...) afin d'extraire de l'information pertinente depuis les images de télédétection. Tous ces algorithmes produisent un ensemble de résultats assez conséquent. Malgré l'augmentation de la qualité des techniques de classification et à cause du nombre important d'outils et même de données à traiter à leur disposition, les experts peuvent se trouver face à un problème de fusion des résultats.

Observant le nombre important de techniques de classification décrites dans la littérature, nous sommes arrivés à un point où nous avons le choix entre de nombreuses méthodes, associées à un nombre tout aussi important de paramètres et donc inévitablement de nombreux résultats. Dans cette étude, nous avons décrit une méthode de fusion de classifieurs. Nous souhaitons une méthode qui permette :

- d'obtenir un résultat unique, fondé si possible sur une implication homogène de chaque méthode dans le résultat final,
- sans masquer les défaillances individuelles, c'est-à-dire, si possible, avoir un retour sur ces méthodes afin de proposer éventuellement des corrections,
- et s'appliquant aussi bien avec des images supervisées que non supervisées (après réaffectation des classes en utilisant une matrice de confusion sur des données d'apprentissage labellisées).

De nombreuses méthodes de fusion de classifieurs ont déjà été proposées dans la littérature, par exemple en utilisant la théorie de l'évidence (Chitroub *et al.*, 2001) ou les modèles de croyances (Aissa *et al.*, 2004). La plupart de ces approches nécessitent des informations supplémentaires comme l'erreur de classification ou un poids (ou *masse*) attribué par un expert. Il existe aussi des approches hybrides (Gańczarski *et al.*, 2005), dans lesquelles les algorithmes sont paramétrés de telle sorte à les faire collaborer entre eux (*hybridation*), ce qui rend l'utilisation de n'importe quelle méthode de classification difficile et proscrit l'utilisation de classifications déjà prêtes. Pour résoudre le problème des pixels qui sont classifiés différemment par les multiples méthodes de classification existantes, nous avons développé une solution à base d'élection consensuelle. L'élection consensuelle utilise une mesure d'information pour noter la qualité du vote et pour décider de la classe d'appartenance de chaque pixel. De nombreuses mesures d'information existent, parmi lesquelles l'entropie de Shannon, l'entropie de Rényi, l'entropie multivariée, les mesures d'information mutuelle et la divergence de Kullback-Leibler (Rényi, 1961; Taneja, 2001). A priori toute mesure d'information peut être utilisée, mais nous avons utilisée celle de Shannon car elle était la plus simple et la plus répandue. Nous présentons ici la solution obtenue de manière abstraite avec quelques résultats qui nous semblent intéressants. Cependant, cette solution a été mise en oeuvre et implémentée à l'aide de plusieurs outils informatiques, notamment ENVI et ICU (ICU, 2005).

Dans la partie suivante, nous commençons par décrire de manière abstraite la notion de *mesure consensuelle*. Ensuite nous présentons deux méthodes de visualisation basée sur cette mesure, permettant de produire des *cartes de vote* et des *cartes de consensualité*. Ensuite, nous discuterons un exemple combinant les résultats de huit algorithmes, dont deux non supervisés, à partir d'une image traitée dans le cadre d'un projet européen (Tidal Inlets Dynamics and Environment, TIDE, (TIDE, 2005)).

2. Mesure consensuelle

Une décision de classification d pour un échantillon P est la dénomination de la classe de cet échantillon par une méthode donnée. D'un point de vue statistique, d peut être modélisée comme la probabilité p_c qu'a une classe donnée c d'être affectée à P .

Quand l'étude concerne n classes, d est alors modélisée comme la distribution de probabilité comprenant l'ensemble des probabilités $(p_1, \dots, p_i, \dots, p_n)$, avec $\sum p_i = 1$, pour toutes les n classes. Pour les méthodes correspondant aux problèmes de type *hard*, où chaque échantillon est classé uniquement dans une classe, la valeur d'un élément p_i est 1 si l'échantillon est affecté à la classe i et 0 sinon. Pour les méthodes correspondant aux problèmes de type *soft* ou flous, la probabilité p_i dépend de l'estimation de la proportion de la classe i pour cet échantillon.

Lorsque différentes méthodes (M_1, \dots, M_m) s'affrontent sur le même échantillon, la distribution de probabilité peut être vue comme un modèle probabiliste indiquant la probabilité qu'a une classe donnée d'être sélectionnée pour un échantillon donné. Dans ce cas, chaque élément p_i est une moyenne des probabilités p_i^M assignées par chaque méthode dans leurs distributions respectives (Ghali *et al.*, 1997).

Un échantillon est dit avoir une bonne qualité de classification quand les décisions concernant sa classe sont consensuelles, c'est-à-dire qu'une forte homogénéité est observée dans les distributions de probabilité.

Nous allons maintenant introduire les concepts théoriques nécessaires pour mesurer l'homogénéité d'une telle distribution. Cette homogénéité repose sur le concept d'entropie.

Si $P = (p_1, p_2, \dots, p_n)$ est une distribution de probabilité, alors l'information exprimée par cette distribution, appelée l'*entropie* de P (Shannon, 1948), est définie par l'équation suivante :

$$I(P) = - \sum_{i=1}^n p_i \log(p_i) \quad [1]$$

Moins la distribution est uniforme, moins elle délivre d'information. Si nous considérons que $P = (p_1, p_2, \dots, p_n)$ est la probabilité de distribution de la classification

d'un échantillon dans n classes par tous les classifieurs, alors la mesure de l'entropie peut être utilisée comme une mesure de l'homogénéité de la décision. Trivialement, d'après l'équation 2, $I(1, 0, \dots, 0) = 0$ et l'entropie est comprise entre 0 et $\log(n)$, où n est le nombre de classes.

$$\lim_{p \rightarrow 0} p \log(p) = 0 \quad [2]$$

Nous définissons l'entropie normalisée d'une distribution, aussi appelée son homogénéité, par l'équation :

$$H(P) = \frac{I(P)}{\log(n)} \quad [3]$$

Dans ce cas, la consensualité peut être définie par l'équation :

$$C(P) = 1 - H(P) \quad [4]$$

Cette mesure s'interprète très facilement. La consensualité est maximale quand tous les classifieurs classifient un échantillon donné dans la même classe, et minimale quand tous les classifieurs le classifient dans des classes différentes. Cette mesure est illustrée sur la figure 1.

Il est possible de calculer de manière théorique dans quel cas la mesure obtenue décrit une majorité simple au niveau de la décision des classifieurs. Dans quels cas la majorité n'est pas atteinte ? Il suffit pour cela d'avoir au moins deux classes, a et b pour lesquelles le vote des classifieurs est identique et majoritaire, soit $p_a = p_b$ avec $p_i \leq p_a \forall i \neq a, b$. Par exemple, et comme expliqué ci-dessus, lorsque les probabilités sont toutes identiques ($p_i = p \forall i$), $C(P) = 0$. Il est facile de voir que la consensualité maximale est atteinte, sans qu'il y ait majorité, lorsque tous les classifieurs votent à égalité pour deux classes et deux seules, et ceci quel que soit le nombre de classes. Dans ce cas on a $p_a = p_b = 0.5$ avec $p_i = 0 \forall i \neq a, b$. La valeur correspondante de $C(P)$, nommée $C_{maj}(n)$ est donnée par l'équation suivante :

$$C_{maj}(n) = 1 + \frac{\sum_{i=1}^n p_i \log(p_i)}{\log(n)} = 1 + \frac{\log(0.5)}{\log(n)} \quad [5]$$

Le but de C_{maj} est de proposer un seuil au delà duquel la consensualité présente une majorité simple de manière certaine. Ce seuil ne dépend que du nombre de classes, et non du nombre de classifieurs. Il faut cependant prendre cette mesure avec précaution car il peut y avoir une majorité en dessous de C_{maj} (voir la figure 1b où

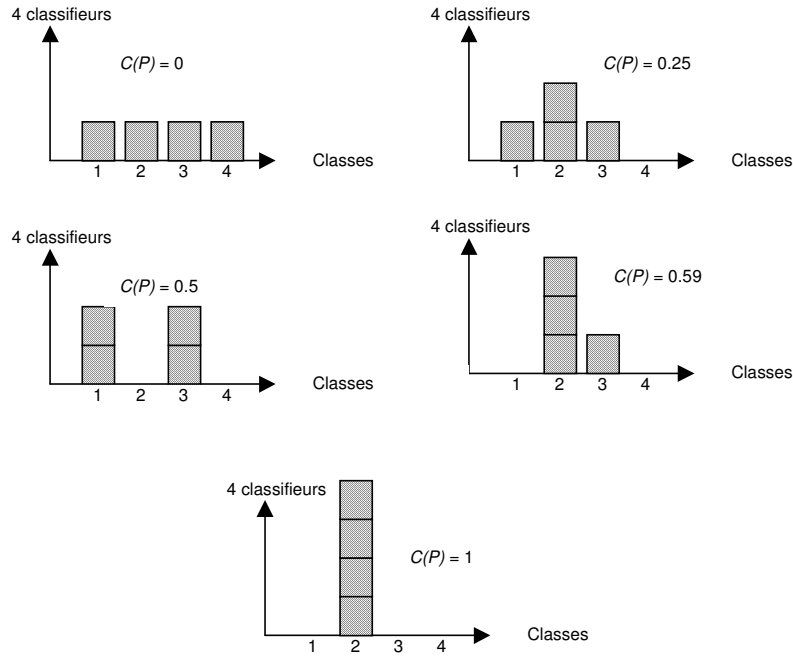


Figure 1. Comparaison de différentes valeurs de consensualité sur un échantillon donné. Un carré gris représente un vote d'un classifieur pour une classe. En abscisse : index des 4 classes, en ordonnée : nombre de classifieurs votant pour chaque classe

$C_{maj}^{(4)} = 0.5$). Il n'y a d'ailleurs pas de raison particulière de considérer ce cas comme peu probable en télédétection.

3. Cartes de consensualité

Dans cette étude nous avons utilisé la mesure $C(P)$ pour produire deux types de cartes distinctes :

La carte de consensualité est une carte dans laquelle chaque pixel P représente la consensualité $C(P)$, calculée à partir des décisions de classification de chaque classifieur pour ce pixel. Les valeurs sont ensuite affichées en niveaux de gris, de telle sorte qu'une valeur importante apparaisse plus blanche qu'une valeur plus faible. Par exemple, si tous les classifieurs sont d'accord pour classer un échantillon dans la même classe, le pixel correspondant de la carte sera affiché

en blanc ($C(P) = 1$). Si la classification de cet échantillon est homogène (par exemple, deux classifieurs indiquent la classe 'A', deux autres la classe 'B' et les deux derniers la classe 'C'), le pixel correspondant aura une consensualité faible et sera imprimé en noir ($C(P) = 0$).

La carte du vote est une classification dans laquelle chaque pixel est affecté à la classe dominante, d'après tous les classifieurs. Chaque pixel est imprimé dans la couleur correspondante à la classe en fonction de la légende proposée par l'utilisateur pour les classifications d'origine. Si la consensualité d'un échantillon est inférieure ou égale à un certain seuil C_{min} , alors le pixel est imprimé en noir, signifiant « non classifié ».

Une autre statistique a été proposée à partir de la carte de consensualité. Il s'agit d'un histogramme qui indique pour chaque catégorie de valeurs dans la carte, le pourcentage correspondant de pixels. Les valeurs en abscisse de cet histogramme sont généralement prises par pas de 0.1 points, c'est-à-dire $[0; 0.1[; [0.1; 0.2[; \dots; [0.9; 1.0]$, mais d'autres échantillonnages peuvent être employés. Cet histogramme peut être utilisé comme post-validation de la carte de consensualité et constitue une preuve importante de sa qualité. Le nombre de pixels le plus élevé doit se situer dans le dernier intervalle. L'analyse de cet histogramme peut permettre de faire des conclusions intéressantes. Par exemple, si l'intervalle $[0.9; 1.0]$ est maximal et qu'un maximal local est observé dans l'intervalle $[0.5; 0.6[$, nous pouvons conclure qu'il y a principalement deux types d'échantillon : ceux qui sont consensuels, et ceux qui ne le sont systématiquement pas en raison d'un choix ambigu entre deux classes. L'identification de ces classes peuvent conduire à l'exclusion de certains classifieurs ou au rajout de suffisamment d'informations expertes pour améliorer la qualité du vote. L'analyse des résultats à la lumière d'un seul classifieur, dans une démarche non collaborative, ne permettrait pas de telles identifications.

La plupart des traitements de cette étude ont été effectués à l'aide du logiciel ENVI, sauf le traitement des résultats non supervisés qui ont été effectués à l'aide d'un logiciel de traitement d'images classique. La production des cartes décrites ci-dessous peut se faire en IDL (dans le logiciel ENVI) ou à l'aide d'un programme en C. Nous avons préféré cette dernière solution pour des raisons de temps de calcul (environ 5 secondes sur un CPU de 3 GHz).

4. Etude de cas

4.1. Présentation

Notre étude de cas se base sur une image de la région de San Felice (Italie), d'une taille de 754x293 pixels. Cette région est une région rurale située dans la périphérie de Venise, qui a fait l'objet d'une étude spécifique visant à protéger et étudier l'évolution de la végétation face à la pollution humaine dans le cadre du projet européen TIDE (TIDE, 2005). Les six classes d'études sont différentes classes de végétation et

d'algues marines, ainsi que le sol et l'eau, dans le but de générer des masques appropriés. Le capteur considéré est le capteur hyperspectral aéroporté CASI, possédant 288 canaux avec une résolution spatiale de 1.3 m au sol. La date de l'image est le 8/2/2003. Les longueurs d'onde observées vont du visible au proche infrarouge (0.4329 - 0.8741 μm). L'expertise utilisée est une expertise de terrain (*ground truthing*), renvoyée par des experts après étude d'échantillons au sol (spectrométrie). Nous disposions à la fois d'expertise *hard* (affectation des échantillons à une et une seule classe) et *floue* (affectation probabiliste des échantillons à toutes les classes), ce qui nous a permis d'employer des résultats provenant de méthodes de classification dures et floues, en plus de méthodes supervisées et non supervisées. En supervisé, 50% du jeu d'exemples validés ont été utilisés en apprentissage, le reste en test.

En plus de la carte consensuelle, la technique permet de comparer entre eux les différents résultats que nous avons pu obtenir, issus d'un travail collaboratif induit par le projet TIDE. Le tableau 1 indique la liste des classifieurs sélectionnés pour cette étude.

Nom	Supervisée ?	Floue ?	Paramètres	Forme des résultats	Durée CPU (3 GHz)
C4.5	Oui	Non	Confiance (0.25), feuilles (3)	Arbre de décision	Courte (5 min)
ICU	Oui	Non	Population (50 à 200), générations (100 à 500)	Règles de classification	Moyenne (30 min)
ICUX	Oui	Oui	Population (100 à 300), générations (200 à 1000)	Règles de classification	Longue (1h)
XCS-R	Oui	Non	Population (2000 à 3000), générations (200000)	Règles de classification	Longue (1h)
MLP ^a	Oui	Oui	Topologie (5 à 15 neurones cachés), itérations	Réseau de neurones	Moyenne (30 min)
SAM ^b	Oui	Non	Seuil (0.1 <i>rad</i>)	Angles	Courte (2 min)
K-Means	Non		Noyaux (15 à 20)	Classif. réaffectée	Courte (10 min)
SOM ^c	Non		Taille de la fenêtre (3x3 à 5x5)	Classif. réaffectée	Courte (10 min)

0. (a) *Multi-Layer Perceptron*

0. (b) *Spectral Angle Mapper* (Yuhua *et al.*, 1992)

0. (c) *Self-Organizing Map* (Kohonen, 1982)

Tableau 1. *Classifieurs utilisés pour l'élection d'un résultat consensuel*

Les résultats des classifieurs sont présentés de manière indépendante sur la figure 2. Bien que ce ne soit pas l'objet de cette étude, nous présentons de manière brève les différents classifieurs utilisés pour montrer qu'ils s'appuient sur des techniques très différentes :

C4.5 est une méthode inductive permettant de générer des arbres de décision développée par Quinlan (Quinlan, 1986).

ICU (*I See You*) est une méthode évolutive permettant de générer des classifieurs sous la forme de règles lisibles et intelligibles par l'expert, en vue de pouvoir comprendre la connaissance extraite. Ces classifieurs sont ensuite utilisés pour classifier l'image (Korczak *et al.*, 2003). ICU est aussi un logiciel de pré-traitement d'images spectrales et hyperspectrales, et de classification supervisée d'images de télédétection (ICU, 2005).

ICUX (eXtended ICU) est une méthode évolutive basée sur ICU permettant d'intégrer de l'expertise floue (c'est-à-dire une approximation probabiliste sur l'af-

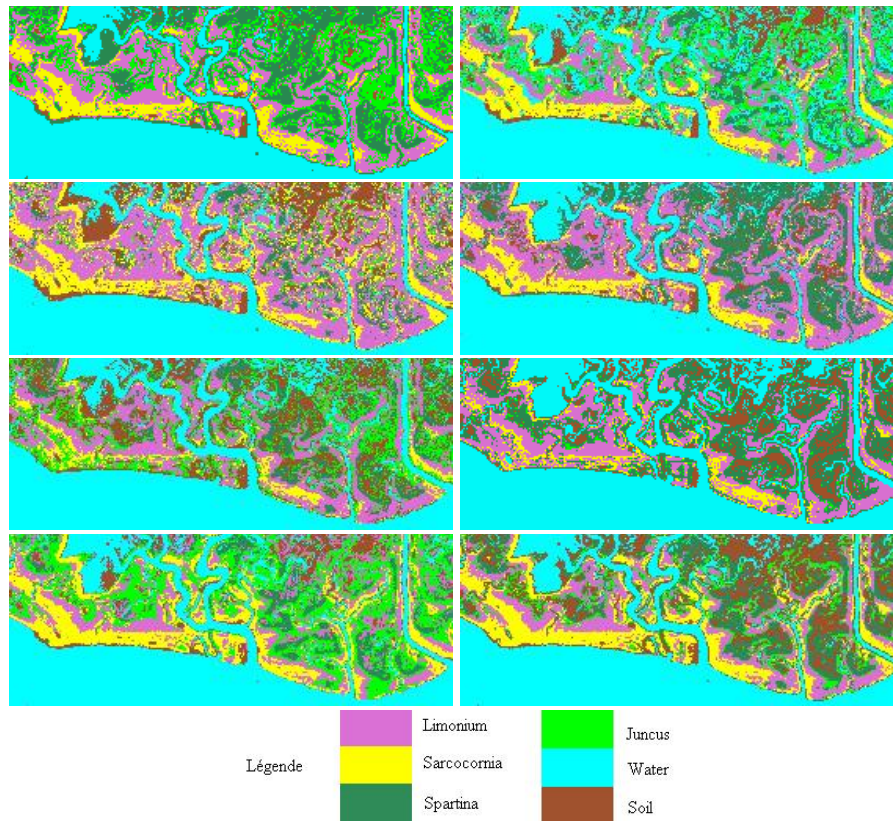


Figure 2. Classifications de l'image de CASI par différents classifieurs. De gauche à droite, puis de haut en bas : (a) C4.5, (b) ICU, (c) ICUX, (d) XCS-R, (e) MLP, (f) SAM, (g) K-Means (15 classes), (h) SOM (fenêtre 4x4)

fection d'un échantillon d'image aux différentes classes d'étude (Quirin *et al.*, 2005)).

XCS-R (eXtended Classifier System with Real values) est un système de classifieurs (apprentissage par renforcement) générant une population complète de classifieurs pour chaque classe d'étude (Wilson, 2000).

MLP (Multi Layer Perceptron) est une méthode permettant la construction d'un réseau de neurones. Le nombre de couches et la topologie du réseau ont été choisis en fonction des meilleurs résultats obtenus sur des échantillons utilisés en test (non appris).

SAM (Spectral Angle Mapper) est une méthode statistique couramment utilisée en télédétection, basée sur la minimisation de l'angle entre un échantillon et la position de la classe idéale dans l'espace d'étude (Yuhas *et al.*, 1992).

K-Means est une méthode de clustering non supervisée affectant à chaque échantillon le noyau dont il est le plus proche. En non supervisé, nous avons réaffecté et fusionné les classes des échantillons en optimisant la qualité (*accuracy*) de la matrice de confusion obtenue sur les données d'apprentissage utilisées pour les autres méthodes.

SOM (Self Organizing Maps) est une méthode de clustering non supervisée basée sur une grille de neurones artificiels dont les poids sont adaptés pour correspondre aux échantillons d'entrée (Kohonen, 1982). Pour cette méthode, nous avons aussi réaffecté et fusionné les classes des échantillons de la même manière que pour K-Means.

Après visite du terrain, pour comparer les différentes classifications obtenues et pour juger de leur qualité, nous avons pu tirer un certain nombre de critiques :

- La partie située dans le quart Nord-Est de la zone étudiée est principalement constituée de petits bras de mer, d'eau stagnante, de plantes très fines, de sable et de sol. La moitié des classifications ne donnent pas le résultat souhaité. Par exemple C4.5 indique une présence de *Juncus* et de *Spartina* trop importante, ICUX et XCS-R indiquent une présence de *Limonium* trop importante et la classification de K-Means indique une présence de *Juncus* trop importante. Seules les classifications de ICU, MLP, SAM et SOM donnent le résultat attendu par les experts dans cette zone. Il faut noter qu'aucun exemple d'apprentissage situé dans le quart Nord-Est n'a été disponible pour les algorithmes. Cependant, les échantillons choisis pour le sol et l'eau situés ailleurs sont représentatifs de leurs classes respectives.

- La partie qui se trouve en bordure de mer, dans le Sud de la zone étudiée est principalement constituée de *Sarcocornia* (*Fruticosa*), une végétation typique de marais et de sols salés. La plupart des classifications donnent le résultat attendu, sauf MLP et SAM qui présentent une zone de *Sarcocornia* moins dense, constituée de nombreuses autres classes, ce qui ne correspond pas à la réalité du terrain. Là encore, aucun échantillon d'apprentissage ne couvrait le Sud-Ouest de la zone étudiée, ce qui peut expliquer les erreurs des algorithmes.

- Le lac salé qui se trouve au Nord-Ouest de l'image est constitué principalement d'eau et de sable humide. Aucune végétation ne peut se développer clairement à l'intérieur de ce lac. La classification qui se rapproche le plus du résultat attendu est celle de l'algorithme MLP car la zone de sable (ici représenté par la classe "Sol") n'est pas trop étendue et que les zones de bordure sont clairement délimitées. Les autres classifications présentent une zone de sable trop étendue (algorithme ICUX), une végétation différente (C4.5) ou une zone d'eau trop importante.

- D'une manière générale, les cours d'eau ne sont pas clairement délimités, ce qui rend difficile l'analyse de l'évolution de la position de ces cours d'eau en zone inondable au cours du temps, qui est l'un des objectifs de notre étude.

- Enfin, d'une manière générale, la quantité de *Juncus* trouvée sur le terrain lors de nos visites successives n'est pas aussi importante, en surface, que celle indiquée par la plupart des classifications obtenues.

On peut donc remarquer qu'aucune des méthodes originales ne peut résoudre complètement le problème de classification proposé. Cependant, il faut noter que de nombreuses autres zones (la majorité), non incluses dans les échantillons d'apprentissage, correspondent à la réalité observée sur le terrain. L'eau en général se comporte correctement, et c'est le cas aussi pour la petite zone de sable située au centre-Sud de l'image (au bord de l'embouchure du cours d'eau principal), pour le Limonium situé dans la moitié Ouest et dans le coin Sud-Est de l'image, pour les bordures de sable au Sud et pour les surfaces moyennes de *Sarcocornia* observées dans l'image entière.

4.2. Cartes de vote et de consensualité

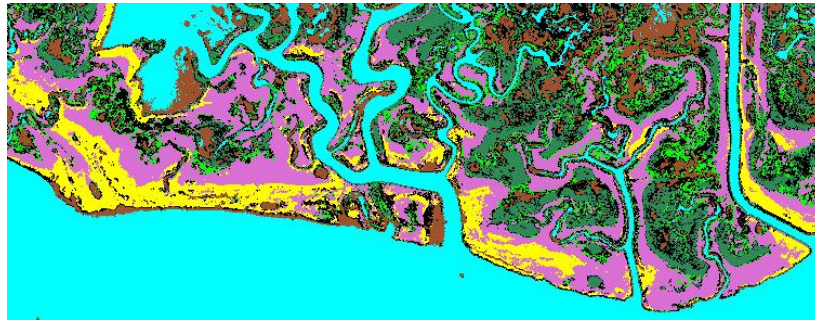


Figure 3. Carte de vote pour CASI

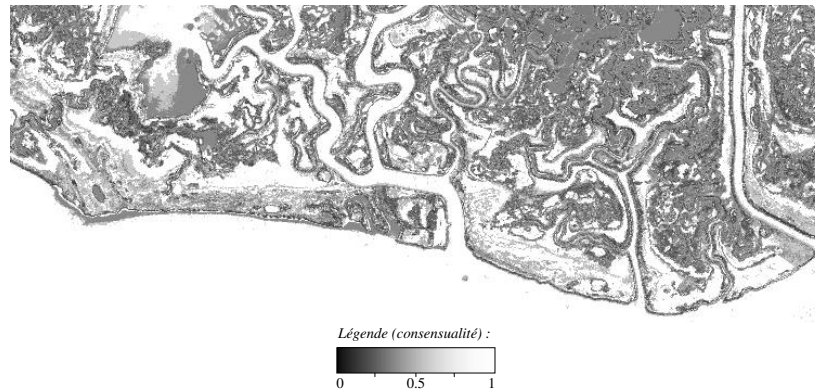


Figure 4. Carte de consensualité pour CASI

Après étude des différents classifieurs, nous avons observé que de nombreuses zones ne sont pas toutes classifiées à l'identique, notamment pour les zones de végétation situées au Nord-Est. La figure 3 montre la carte du vote. La valeur théorique du seuil pour être certain de la majorité du vote avec 6 classes est de $C_{maj}^{(6)} = 0.613$,

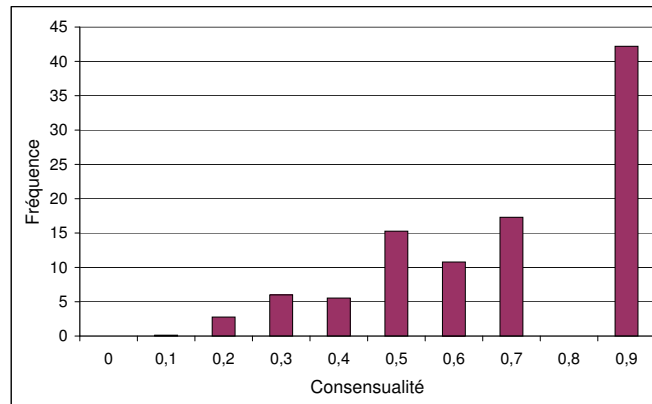


Figure 5. Histogramme de la consensualité pour CASI

mais il est plus prudent de choisir une valeur inférieure correspondante à 4 classes (d'une part, l'eau et le sol ne sont pas des classes d'intérêt pour l'expert et d'autre part, une majorité peut être acquise en-dessous du seuil C_{maj} comme expliqué plus haut). Pour cette carte, nous avons donc fixé le seuil C_{min} à $C_{maj}^{(4)} = 0.5$.

De nombreux pixels ont pu être votés, le nombre de pixels noirs étant très faible. Ces pixels non consensuels se situent surtout en bord de mer, la végétation y est, en effet, extrêmement mélangée. Par rapport aux critiques énoncées plus haut, sur les classifications indépendantes, nous observons les résultats suivants :

- Le sol est beaucoup plus apparent dans la partie située dans le quart Nord-Est de la zone étudiée. Les bras de mer sont beaucoup plus détaillés et précis, même si une classification optimale en laisserait apparaître d'avantage. La quantité de *Juncus* dans cette zone a été réduite, et celle de *Limonium* aussi.

- La partie au Sud de la zone étudiée, en bordure de mer, démontre la présence de *Sarcocornia* en quantité importante, sur une large partie du littoral. Ce résultat fait partie des résultats attendus par les experts du terrain.

- La zone au Nord-Ouest de l'image montre un lac correctement délimité et possédant une partie relativement importante de sable au Sud (détectée par la classe "Soil"). Ceci correspond tout à fait à la réalité, même si un peu plus de sable devrait être présent dans cette zone. Pourtant, seulement une classification (l'algorithme MLP) présentait un résultat proche du résultat optimal souhaité par les experts. Il semble que chaque classification n'était pas consensuelle pour les pixels situés au Sud de ce lac : chaque classifieur votait pour une classe de végétation différente, ou de l'eau, ou du sol. Peu de classifieurs ont voté pour une large surface de sol au même endroit, mais chacun a voté pour du sol à des endroits différents. Ceci explique pourquoi la mesure consensuelle fonctionne bien dans cette partie de l'image : les décisions de plusieurs classifieurs, même si leur nombre est faible par rapport au nombre total de classifieurs,

a été dominante concernant l'eau et le sol, même si les classifieurs concernés n'ont pas été les mêmes à chaque fois.

- Les cours d'eau apparaissent de manière plus précise, parce que les défauts individuels de chaque classification ont été lissés.

- Enfin, la quantité globale de *Juncus* trouvée par la carte de vote correspond bien aux observations de terrain, c'est-à-dire une présence très faible mais répartie sur toute la zone. Ceci permet d'avoir une bonne idée sur la qualité globale de la classification obtenue, même si nous présentons une approche locale.

La figure 4 montre la carte de consensualité. Les pixels grisés sont les plus difficiles à classifier. On peut remarquer que les zones problématiques détaillées ci-dessus sont les zones qui apparaissent de manière moins consensuelles que les autres zones. Par exemple, la région du Nord-Est apparaît en gris sombre, ainsi que la partie du lac la plus problématique. De même la partie au Sud, en bordure de mer, laisse apparaître des régions plus sombres. Par contre les régions les moins problématiques, comme la zone de sable située au centre-Sud de l'image apparaît en blanc pur. Ceci démontre que la carte de consensualité est cohérente avec les observations des experts. Nous envisageons de l'utiliser comme support pour déterminer à l'avance les zones problématiques afin de mener une validation plus poussée dans ces zones.

L'histogramme de la figure 5 nous apprend qu'une grande majorité de pixels (plus de 40%) sont pleinement consensuels et les dix classifieurs ont pu fournir une expertise efficace sur plus de 70% de l'image ($C(P) > 0.5$). Ceci correspond parfaitement à nos observations, puisque les classifications sont globalement correctes. L'histogramme permet de faire une remarque intéressante : les pixels les moins consensuels (avec une valeur inférieure à 0.2) sont les moins fréquents : en fait, ils correspondent principalement aux bordures des cours d'eau ou de la mer car, à cause de la présence d'eau, la végétation apparaît moins clairement pour les détecteurs spatiaux. Globalement, la forme de l'histogramme suit une forme que nous avons souvent observée, en tout cas sur ce type d'image (rurales) : la fréquence des pixels possédant une consensualité donnée augmente lentement avec la valeur de la consensualité, et les pixels possédant une forte consensualité sont toujours majoritaires. Cette forme est clairement déterminée par la présence d'eau dans ce type d'image, qui est généralement plus facile à discriminer que la végétation.

4.3. Mesures statistiques

La figure 6 présente les matrices de confusion obtenues pour les méthodes supervisées (C4.5, ICU, ICUX, XCS-R, MLP et SAM) sur des données validées par des experts géographes au sol. Les intitulés des lignes et des colonnes représentent les différentes classes de l'image et les valeurs correspondent aux pourcentages de pixels observés. Globalement la qualité des résultats individuels supervisés est plutôt bonne (l'indice de qualité (*accuracy*) obtenu sur la matrice est de 0.99 pour C4.5, 0.99 pour ICU, 0.99 pour ICUX, 0.94 pour XCS-R, 0.98 pour MLP et 0.96 pour SAM), mais

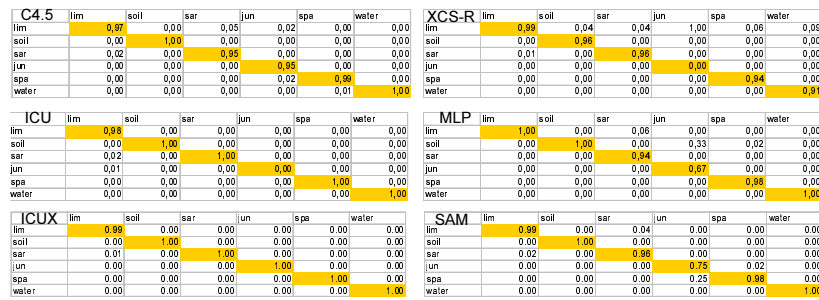


Figure 6. Matrices de confusion obtenues pour chacune des méthodes supervisées, uniquement sur les données de test

elle est moins bonne pour les méthodes non supervisées (0.36 pour K-Means, 0.84 pour SOM). Ces matrices de confusion sont cohérentes avec nos observations. Par exemple, l’algorithme C4.5 a quelques difficultés pour proposer une quantité faible de Juncus dans le Nord-Est de l’image. C’est cette classe qui a obtenu le moins bon score dans la matrice. Pour ICU et XCS-R les classes les plus problématiques sont le Limonium (dans le Nord-Est) et le Juncus (dans toute l’image), ce qui est confirmé par les observations. Concernant MLP, la classe Sarcocornia est l’une des moins bien classées d’après la matrice, ce qui se retrouve dans l’observation du bord de mer.

La figure 7 du haut présente la position des zones validées par les experts géographes au sol qui ont été choisies pour valider la méthode de la mesure consensuelle *a posteriori*. Ces zones ont été utilisées comme zones de test pour les résultats supervisés et non supervisés. Concernant les méthodes non supervisées, ces zones n’ont été utilisées *a priori* que pour déterminer le nombre optimal de classes dans les données, et pour réaffecter les bonnes couleurs aux classes obtenues, à l’aide d’un logiciel classique de traitement d’images. La figure 7 du bas présente la position des zones validées sur la carte de vote obtenue avec la méthode de la mesure consensuelle. Sur cet exemple, nous notons que les zones de test correspondent aux classes votées par notre méthode. Certaines zones de test comme l’eau (classe *Water*) sont incluses sur la carte obtenue, et d’autres zones, correspondant à des régions spécifiquement délimitées par l’expert, comme les classes Sarcocornia (*Sar*) et Spartina (*Spa*), sont délimitées de la même façon sur la carte obtenue, entourées d’une autre classe. Même si l’objectif principal de la carte de vote est de visualiser l’efficacité d’un ensemble de classifieurs, et non pas sa qualité, l’exemple montré ici permet aussi de juger de la justesse des décisions prises par la mesure consensuelle.

Comme mentionné plus haut, les méthodes supervisées ont appris à l’aide d’une expertise de terrain (*ground truthing* ou ROI) dont nous présentons les échantillons dans la figure 7. La figure 8 montre la matrice de confusion en comparant la carte de vote finale aux expertises de terrain. La ligne nommée "unclassified" correspond au nombre d’échantillons qui n’ont pas pu être classifiés par la méthode consensuelle

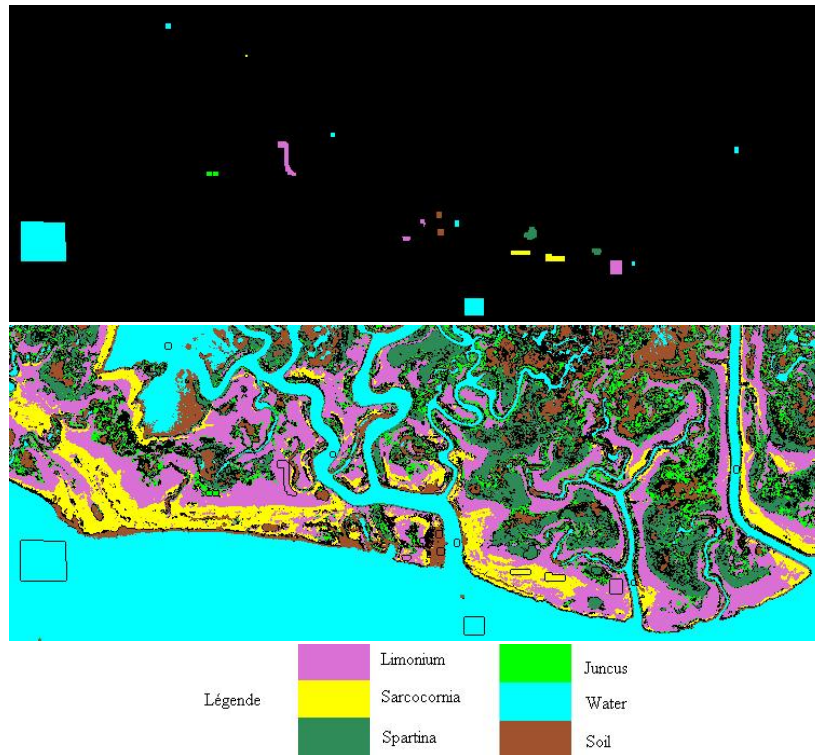


Figure 7. En haut : Regions Of Interest (ROI) experts utilisés pour valider les résultats supervisés et pour affiner les résultats non supervisés. En bas : ROI plaqués sur la carte de vote

vote \ roi	unclassified	water	spa	jun	soil	lim	sar
unclassified	35826	0	6	3	0	2	0
water	70376	1904	0	0	0	0	0
spa	22576	0	142	0	0	0	0
jun	12647	0	2	31	0	0	0
soil	17355	0	0	0	66	0	0
lim	42911	0	0	6	0	367	12
sar	16522	0	0	0	0	2	166

Figure 8. Matrice de confusion obtenue entre la carte de vote (classification finale) et le ROI de validation, uniquement sur les données de test

(Le seuil C_{min} a été fixé à 0.5). La colonne nommée "unclassified" correspond au nombre d'échantillons non classés par l'expert. La qualité (*accuracy*) obtenue sur cette matrice est de 0.99. D'autres mesures statistiques effectuées sur cette matrice donnent des résultats similaires (par exemple, le κ -index vaut 0.99 et l'erreur quadratique par

rapport aux classes attendues est de 0.01). Nous constatons que malgré les défauts de certaines méthodes, la qualité finale obtenue sur la carte de vote est identique à la meilleure qualité obtenue sur les résultats individuels (la moyenne des qualités étant de 0.88). On peut donc constater que les défauts individuels, observés d'une méthode à l'autre, sont correctement corrigés et n'influencent pas négativement sur la qualité de la classification finale. Bien entendu, d'autres travaux comprenant d'autres expertises au sol (très coûteuses) devront être menés pour confirmer ou infirmer ce résultat.

5. Conclusion

Nous avons présenté dans ce papier la méthode consensuelle. Cette méthode permet de lisser et de compenser les défauts de chaque classifieur utilisé et est utile lorsqu'on ne peut pas choisir directement la méthode à utiliser (soit par manque de connaissances sur cette méthode, soit par manque de critères de qualité, soit parce que le nombre de méthodes à analyser manuellement est trop grand), ou lorsqu'on ne peut pas évaluer la pertinence de celle-ci. Nous avons aussi présenté deux visualisations scientifiques découlant de cette mesure. Les résultats obtenus sur l'image de CASI sont encourageants. Ils sont aussi relativement surprenants car ils ont été générés par des méthodes très diverses, avec des paramètres et des jeux d'apprentissages différents (par exemple, certaines méthodes nécessitaient seulement quatre pixels pour l'apprentissage, alors que pour d'autres, il en fallait quelques centaines). La technique que nous avons développée, basée sur l'entropie, permet donc d'unifier efficacement une dizaine de classifieurs différents, en présentant un résultat final plutôt homogène et fidèle à l'observation de terrain. Cependant, l'histogramme est une méta-mesure des résultats qui peut se révéler insuffisante pour certains experts. Nous projetons de développer d'autres mesures statistiques permettant d'affiner la compréhension des cartes de consensualité et de vote, afin d'avoir une idée plus locale des défaillances individuelles de chaque méthode. Une mesure statistique intéressante pourrait être les courbes ROC (Zhou, 1998). Nous pensons les appliquer pour l'analyse de la sensibilité et la spécificité de la classification finale en fonction d'un paramètre donné, par exemple le seuil de consensualité. Nous nous intéressons aussi à d'autres opérateurs d'agrégation pour remplacer la mesure entropique. De nombreux opérateurs ont été proposés à cette fin parmi lesquels la moyenne arithmétique, la moyenne pondérée, le médian, la moyenne géométrique, la somme symétrique et l'opérateur *Ordered Weighted Averaging* (OWA) (Detyniecki, 2001). Parmi ceux-ci, l'opérateur OWA est très intéressant dans le sens où il intègre des pondérations qui peuvent être utilisées pour noter la pertinence de chaque classifieur, mais surtout l'identification de majorité simple est simplifiée. L'utilisation de ces nouvelles mesures permettra de savoir si la classification obtenue est juste, et non plus seulement efficace, comme c'est le cas actuellement.

Remerciements

Ce travail a été financé par le projet européen *Tidal Inlets Dynamics and Environment* (projet TIDE, contrat n°EVK3-CT-2001-00064 sous le cinquième programme

cadre, (TIDE, 2005)). L'auteur tient à remercier Massimo Menenti et Jerzy Korczak, de l'Université Louis Pasteur (Strasbourg), ainsi que Marco Marani, Enrica Bellico, Monica Camuffo et Sergio Ferrari de l'Université UNPADU (Padoue) pour leurs idées, la mise à disposition des données multispectrales utilisées (images de San Felice et expertises dures et floues acquises durant le projet TIDE) et leur aide sur ces données. Nous remercions enfin les relecteurs anonymes pour toutes leurs remarques (très) pertinentes.

6. Bibliographie

- Aissa A. B., Faouzi N. E. E., Lefevre E., « Classification multisource via la théorie des croyances », *Workshop Fouilles de données complexes dans un processus d'extraction de connaissances, 4ème journée d'Extraction et Gestion des Connaissances, EGC'2004*, p. 31-44, 2004.
- Chitroub S., Houacine A., Sansal B., « Fusion de classifieurs en utilisant la théorie de l'évidence pour l'amélioration de la classification d'image », *18e colloque sur le traitement du signal et des images*, 2001.
- Detyniecki M., « Numerical Aggregation Operators : State of the Art », *International Summer School on Aggregation Operators and their Applications*, Asturias, Spain, 2001.
- Gançarski P., Wemmert C., « Collaborative Multi-strategy Classification : Application to per-pixel Analysis of Images », *Sixth International Workshop on Multimedia Data Mining (MDM/KDD2005)*, Chicago, 2005.
- Ghali A., Daemi M. F., « A probability distribution for shape description », *International conference on image processing and its applications*, p. 121-125, 1997.
- ICU, Document en ligne, 2005. Discovery of classification rules : evolutionnary classifiers, logiciel disponible en ligne sur <http://lsiit.u-strasbg.fr/afd/logiciels/icu>, accessible le 22/6/2007.
- Kohonen T., « Self-organized formation of topologically correct feature maps », *Biological Cybernetics*, vol. 43, p. 59-69, 1982.
- Korczak J. J., Quirin A., « Découverte de règles de classification par approche évolutive : application aux images de télédétection », *Journées francophones d'Extraction et de Gestion de Connaissances (EGC2003)*, Lyon, 2003.
- Quinlan J. R., « Induction of Decision Trees », *Machine Learning*, vol. 1, p. 81-106, 1986.
- Quirin A., Korczak J. J., « Representation of Genetic Individuals for Unmixing Multispectral Data », *2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, Edinburg, 2005.
- Rényi A., « On measures of information and entropy », *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, p. 547-561, 1961.
- Shannon C. E., « A mathematical theory of communication », *Bell System Technical Journal*, vol. 27, p. 379-423 and 623-656, 1948.
- Taneja I. J., Document en ligne, 2001. Generalized Information Measures and Their Applications, livre en ligne sur www.mtm.ufsc.br/~taneja/book/book.html, accessible le 22/6/2007.
- TIDE, Document en ligne, 2005. Tidal Inlets Dynamics and Environment, Research Project Supported by the European Commission under the Fifth Framework Programme 2001-

2005, contract n° EVK3-CT-2001-00064, <http://www.istitutoveneto.it/tide>, accessible le 22/6/2007.

Wilson S. W., *Learning Classifier Systems : From Foundations to Applications, LNAI 1813*, chapter Get real ! XCS with continuous-valued inputs, p. 209-220, 2000.

Yugas R. H., Goetz A. F. H., Boardman J. W., « Discrimination among semiarid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm », *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, Pasadena, Jet Propulsion Laboratory, p. 147-149, 1992.

Zhou X. H., « Comparing the correlated areas under the ROC curves of two diagnostic tests in the presence of verification bias », *Biometrics*, vol. 54, p. 453-470, 1998.