

---

# The consensual measure

## Election au niveau des résultats : la mesure consensuelle

Quirin Arnaud, Jerzy Korczak

LSIT/CNRS Laboratory,  
Boulevard Sébastien Brant,  
67400 Illkirch, France.  
{quirin,jjk}@lsiit.u-strasbg.fr

---

*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. Le concept de mesure consensuelle est présenté, ainsi que diverses comparaisons sur une base d'images existante.

*ABSTRACT.* The problem of choice of an adequate, efficient and robust classification method for a given set of data is well-known. In the last two decades, a variety of clustering and classification algorithms have been proposed that have application in analysis of remote sensing images. 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 it 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. The consensual measure is proposed and validated on remote sensing images.

*MOTS-CLÉS :* mesure consensuelle, élection, classification, images de télédétection.

*KEYWORDS:* consensual measure, election, classification, remote sensing images.

---

Version du 31 mai 2006

## 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 télé-détection. 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 importants d'outils et même de données à traiter à leur disposition, les experts se trouvent 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 de nombreuses méthodes, associées à un nombre tout aussi important de paramètres et donc inévitablement de nombreux résultats. Nous cherchions alors un moyen de :

- pouvoir 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 dans le contexte non supervisé que supervisé.

Effectuer une simple moyenne des résultats est donc proscrite. De plus, des approches hybrides existent [GAN 05], mais nous souhaitons pouvoir appliquer les méthodes existantes, et même éventuellement des classifications déjà prêtes, détachées des méthodes, supervisées ou non, sans avoir à retoucher les algorithmes pour les emboîter entre eux (*hybridation*). 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. Nous la présentons ici avec quelques résultats qui nous semblent intéressants.

Dans la partie suivante, nous commençons par décrire de manière abstraite la notion de *mesure consensuelle*. Ensuite nous présentons la méthode de visualisation basée sur cette mesure, permettant de produire 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, [TID 05]).

## 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  $(p_0, \dots, p_i, \dots, p_n)$  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  $(M_1, \dots, M_n)$  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.

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  $P$ , est définie par l'équation :

$$I(P) = - \sum_{i=0}^n p_i \log(p_i) \quad [1]$$

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 classificateurs, alors la mesure de l'entropie peut être utilisée comme une mesure de l'homogénéité de la décision. Trivialement, d'après l'équation :

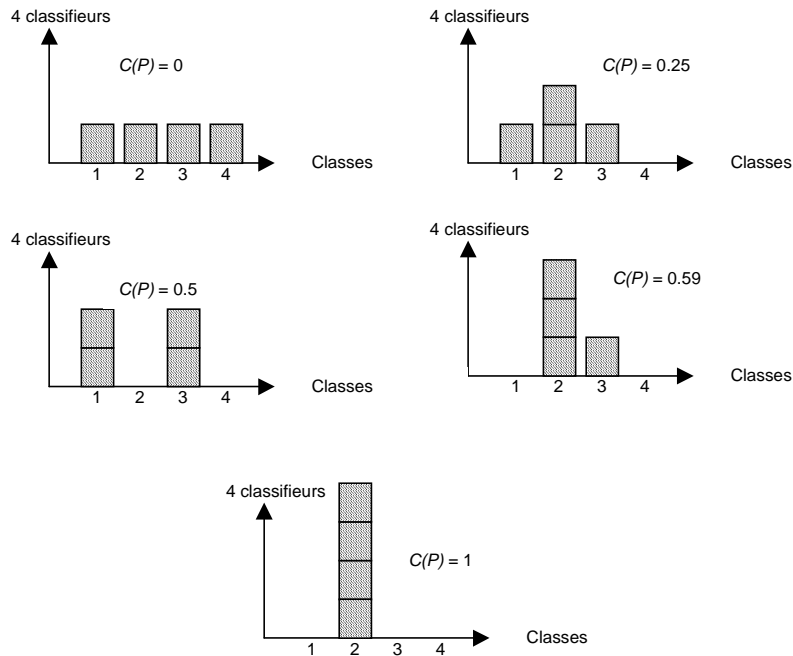
$$\lim_{p \rightarrow 0} p \log(p) = 0 \quad [2]$$

$I(1, 0, \dots, 0) = 0$  et l'entropie est comprise entre 0 et  $\log(n)$ , où  $n$  est le nombre de classes. 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 :

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



**Figure 1.** Comparaison de différentes valeurs de consensualité.

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.

### 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. Les valeurs sont affichées en niveaux de gris. Par exemple, si tous les classifieurs sont d'accord pour classifier 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 correspondant à la classe en fonction d'une légende proposée par l'utilisateur. Si la consensualité d'un échantillon est inférieure à un certain seuil  $C_{min}$ , alors le pixel est imprimé en noir, signifiant « non classifié ». Un seuil de 0.5 a été choisi arbitrairement pour les figures qui vont suivre.

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'échantillons : ceux qui sont consensuels, et ceux qui ne le sont systématiquement pas à 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 de l'histogramme. L'analyse des résultats à la lumière d'un seul classifieur, dans une démarche non collaborative, ne permettrait pas de telles identifications.

## 4. Etude de cas

### 4.1. Présentation

Notre étude de case se base sur une image de la région de San Felice (Italie), d'une taille de 754x293 pixels. 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  $\mu m$ ). L'expertise utilisée est une expertise de terrain (*ground truthing*), renvoyée par des experts après étude d'échantillons au sol (spectrométrie). Nous disposons à la fois d'expertise *dure* (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.

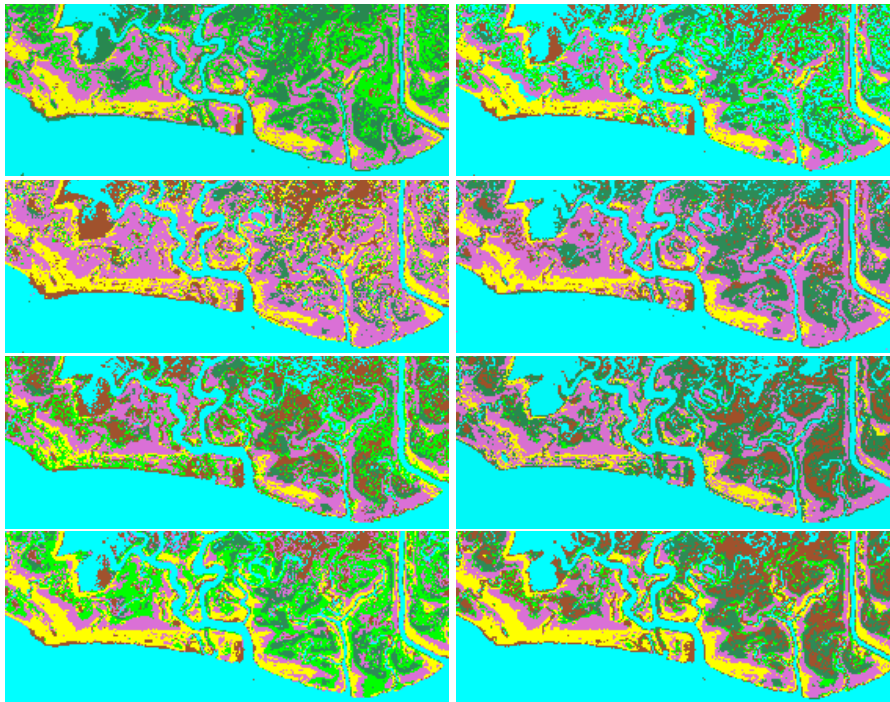
Nom	Supervisée ?	Floue ?	Paramètres	Forme des résultats	Durée CPU (3 GHz)
C4.5	Oui	Non	Confiance (0.25), feuilles (3)	Arbre de décision	Courte (5 min)
ICU	Oui	Non	Population (50 à 200), générations (100 à 500)	Règles de classification	Moyenne (30 min)
ICUX	Oui	Oui	Population (100 à 300), générations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

0. (a) *Multi-Layer Perceptron*

0. (b) *Spectral Angle Mapper* [YUH 92]

0. (c) *Self-Organizing Map* [KOH 82]

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



**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).

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 développé par Quinlan [QUI 86].

**ICU** (*I See You*) est une méthode évolutive permettant de générer des classifieurs sous la forme de règle lisible et intelligible par l'expert, en vue de pouvoir comprendre la connaissance extraite. Ces classifieurs sont ensuite utilisés pour classifier l'image [KOR 03].

**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'affectation d'un échantillon d'image aux différentes classes d'étude [QUI 05].

**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 [WIL 00].

**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 [YUH 92].

**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.

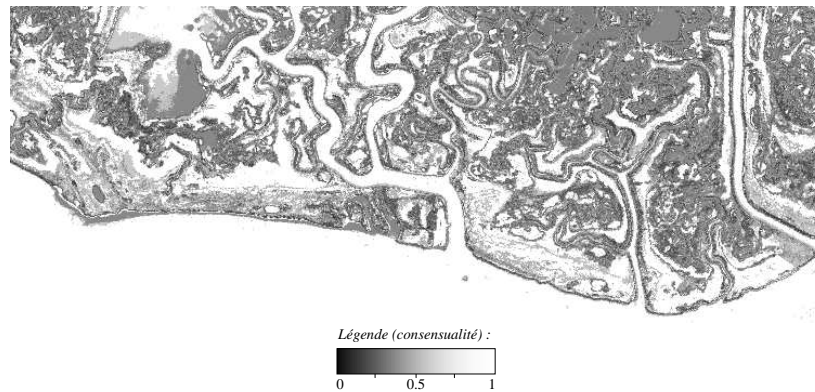
**SOM** (Self Organizing Maps) est une méthode de clustering non supervisée basée sur une grille de neurones artificiels dont les poids sont adaptés pour correspondre aux échantillons d'entrée [KOH 82]. Pour cette méthode, nous avons aussi réaffecté et fusionné les classes des échantillons de la même manière que pour K-Means.

#### 4.2. Cartes de vote et de consensualité

Nous pouvons y voir que de nombreuses zones ne sont pas toutes classifiées à l'identique, notamment pour les zones de végétation situées au nord-est. La figure suivante (figure 3) montre la carte du vote : 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, en effet, la végétation y est extrêmement mélangée. La figure 4 montre la carte de consensualité. Les pixels grisés sont les plus difficiles à classifier. On remarque que la région du nord-est, surtout constituée de plantes fines, couvrant un sol



**Figure 3.** Carte de vote pour CASI.



**Figure 4.** Carte de consensualité pour CASI.

très humide voire inondé, est la moins consensuelle. Heureusement, le sol et l'eau ne représentent pas des classes intéressantes pour le projet TIDE. 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 ( $C(P) > 0.5$ ). Il faut cependant noter que la majorité des pixels ayant une consensualité maximale correspondent à l'eau, 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 (voir Fig.7). Les intitulés des lignes et des colonnes représentent les différentes classes de l'image (essentiellement de la végétation rurale).



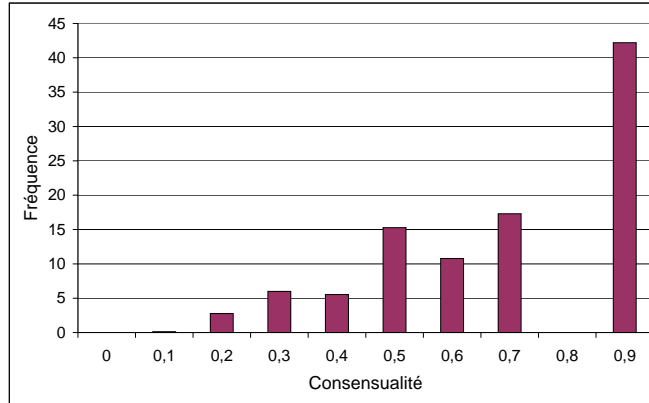


Figure 5. Histogramme de la consensualité pour CASI.

C4.5							XCS-R						
	lim	soil	sar	jun	spa	water		lim	soil	sar	jun	spa	water
lim	0,97	0,00	0,05	0,02	0,00	0,00	lim	0,99	0,04	0,04	1,00	0,06	0,09
soil	0,00	1,00	0,00	0,00	0,00	0,00	soil	0,00	0,98	0,00	0,00	0,00	0,00
sar	0,02	0,00	0,99	0,00	0,00	0,00	sar	0,01	0,00	0,99	0,00	0,00	0,00
jun	0,00	0,00	0,00	0,92	0,00	0,00	jun	0,00	0,00	0,00	0,00	0,00	0,00
spa	0,00	0,00	0,00	0,02	0,99	0,00	spa	0,00	0,00	0,00	0,00	0,94	0,00
water	0,00	0,00	0,00	0,00	0,01	1,00	water	0,00	0,00	0,00	0,00	0,00	0,91

ICU							MLP						
	lim	soil	sar	jun	spa	water		lim	soil	sar	jun	spa	water
lim	0,98	0,00	0,00	0,00	0,00	0,00	lim	1,00	0,00	0,06	0,00	0,00	0,00
soil	0,00	1,00	0,00	0,00	0,00	0,00	soil	0,00	1,00	0,00	0,33	0,02	0,00
sar	0,02	0,00	1,00	0,00	0,00	0,00	sar	0,00	0,00	0,94	0,00	0,00	0,00
jun	0,01	0,00	0,00	0,00	0,00	0,00	jun	0,00	0,00	0,00	0,87	0,00	0,00
spa	0,00	0,00	0,00	0,00	1,00	0,00	spa	0,00	0,00	0,00	0,00	0,98	0,00
water	0,00	0,00	0,00	0,00	0,00	1,00	water	0,00	0,00	0,00	0,00	0,00	1,00

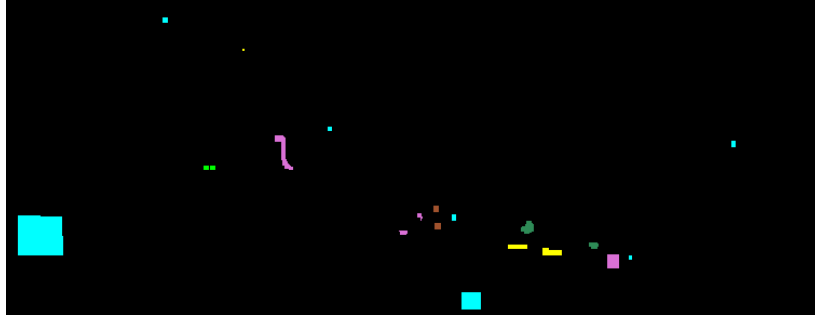
  

ICUX							SAM						
	lim	soil	sar	jun	spa	wat		spa	jun	soil	sar	lim	
lim	0,99	0,00	0,00	0,00	0,00	0,00	spa	0,98	0,25	0,00	0,00	0,00	
soil	0,00	1,00	0,00	0,00	0,00	0,00	jun	0,02	0,75	0,00	0,00	0,00	
sar	0,01	0,00	1,00	0,00	0,00	0,00	soil	0,00	0,00	1,00	0,00	0,00	
jun	0,00	0,00	0,00	1,00	0,00	0,00	sar	0,00	0,00	0,00	0,95	0,02	
spa	0,00	0,00	0,00	0,00	1,00	0,00	lim	0,00	0,00	0,00	0,04	0,99	
wat	0,00	0,00	0,00	0,00	0,00	1,00							

Figure 6. Matrices de confusion obtenues pour chacune des méthodes supervisées.

Globalement la qualité des résultats individuels supervisés est plutôt bonne (0,99 pour C4.5, 0,99 pour ICU, 0,99 pour ICUX, 0,94 pour XCS-R, 0,98 pour MLP, 0,96 pour SAM), mais elle est moins bonne pour les méthodes non supervisées (0,36 pour K-Means, 0,84 pour SOM).

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 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  $\kappa$ -index est à 0,99). 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 est 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.



**Figure 7.** ROI (Regions Of Interest) experts utilisés pour valider les résultats supervisés et pour affiner les résultats non supervisés.

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	0	6	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.

## 5. Conclusion

Ces résultats sont très 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.

## Remerciements

Les auteurs tiennent à remercier Marco Marani, Enrica Belluco, Monica Camuffo et Sergio Ferrari de l'Université UNPADU (Padoue) pour les données multispectrales

utilisées (images de San Felice et expertises dures et floues acquises durant le projet TIDE [TID 05]).

## 6. Bibliographie

- [GAN 05] 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.
- [KOH 82] KOHONEN T., « Self-organized formation of topologically correct feature maps », *Biological Cybernetics*, vol. 43, 1982, p. 59-69.
- [KOR 03] 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.
- [QUI 86] QUINLAN J. R., « Induction of Decision Trees », *Machine Learning*, vol. 1, 1986, p. 81-106.
- [QUI 05] QUIRIN A., KORCZAK J. J., « Representation of Genetic Individuals for Unmixing Multispectral Data », *2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, Edinburg, 2005.
- [TID 05] TIDE, 2005, Tidal Inlets Dynamics and Environment, Research Project Supported by the European Commission under the Fifth Framework Programme, contract n° EVK3-CT-2001-00064, document en ligne disponible sur <http://www.istitutoveneto.it/tide>.
- [WIL 00] WILSON S. W., « *Learning Classifier Systems : From Foundations to Applications*, *LNAI 1813* », chapitre Get real ! XCS with continuous-valued inputs, p. 209-220, 2000.
- [YUH 92] YUHAS R. H., GOETZ A. F. H., BOARDMAN J. W., « Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm », *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, Pasadena, Jet Propulsion Laboratory, 1992, p. 147-149.