

# Étude de données multisources par simulation de capteurs et clustering collaboratif

Germain Forestier\*, Cédric Wemmert\*, Pierre Gançarski\*

\*Université de Strasbourg - LSIT - CNRS - UMR 7005  
Pôle API, Bd Sébastien Brant - 67412 Illkirch, France  
{forestier,wemmert,gancarski}@unistra.fr

**Résumé.** Depuis quelques années les données issues de capteur satellitaire deviennent de plus en plus accessibles. Différents systèmes satellitaires sont maintenant disponibles et produisent une masse de données importante utilisée pour l'observation de la Terre. Pour mieux comprendre la complexité de la surface terrestre, il devient courant d'utiliser plusieurs données provenant de capteurs différents. Cependant, il est souvent difficile de prédire le gain potentiel d'un système multisource avant de réellement acquérir les données. Dans cet article nous présentons une approche par simulation qui permet de créer différentes vues de données satellitaire à partir des caractéristiques des capteurs. Ces différentes vues sont ensuite étudiées à travers des algorithmes de classification supervisée, d'outils statistiques et de clustering collaboratif pour évaluer l'intérêt d'utiliser ces données conjointement. Les premières expériences permettent de mettre en avant des couples de capteurs pouvant tirer partie de cette utilisation multisource.

## 1 Introduction

Un nombre important de capteurs satellitaires sont disponibles pour capturer des images de télédétection de la surface terrestre. Ces données de télédétection sont utilisées intensivement pour étudier la Terre et les systèmes satellitaires sont utilisés pour collecter des données dans les domaines de l'agriculture, la production de nourriture, la géologie, la prospection pétrolière, l'exploration minière, la géographie ou l'étude des milieux urbains.

Chaque satellite a ses propres caractéristiques, l'une des plus importantes étant sa résolution spectrale. La résolution spectrale d'un capteur peut être caractérisée par le nombre de bandes spectrales, leurs largeurs, et leurs positions sur le spectre (Herold et al., 2003).

Plusieurs études ont montrées (Herold et al., 2003; Meyer et Chander, 2007; Heidena et al., 2007) que la résolution spectrale des capteurs est un point critique, particulièrement pour discriminer différents types d'occupation du sol dans des environnements complexes tels que le milieu urbain. Malgré l'augmentation de la disponibilité des données hyperspectrales, les capteurs multispectraux embarqués à bord de plusieurs satellites acquièrent chaque jour une très grande masse de données avec une résolution spectrale relativement pauvre. La plupart des systèmes satellitaires possède 4 à 7 bandes spectrales allant du visible à l'infrarouge sur le

spectre électromagnétique (Govender et al., 2007). Il existe cependant quelques capteurs qui utilisent également des bandes thermiques. L'un des avantages de ces capteurs multispectraux, comparés aux capteurs hyperspectraux, est la couverture spatiale plus importante, qui permet de cartographier plus rapidement de grandes zones.

Comme le nombre de systèmes satellitaires existant augmente avec la complexité des données, un challenge important est d'évaluer la complémentarité de ces capteurs pour une application donnée. En effet, les informations fournies par les différents capteurs peuvent être complémentaires, et un problème clé est de proposer des systèmes capables d'utiliser ces sources d'informations hétérogènes dans un processus unique. Cependant, acquérir des images satellites est toujours très coûteux, c'est pourquoi il serait intéressant d'évaluer a priori la complémentarité des capteurs. Pour résoudre ce problème, nous utilisons dans cet article une approche par simulation. La simulation de capteurs, également appelée simulation de bande, consiste à générer des données multispectrales à partir de données acquises par un autre capteur existant ayant une meilleure résolution spatiale. Cette simulation consiste à combiner des bandes fines en bandes plus larges. Ce type d'approche a déjà été utilisé, particulièrement pour la calibration de capteur. L'étape de simulation utilise les Relative Spectral Response (RSR) des capteurs multispectraux, qui décrivent la réponse spectrale de chaque bande simulée.

Les spectres utilisés pour la simulation proviennent de bibliothèques spectrales qui sont des dépôts de spectres de différents matériaux (par exemple des minéraux, de la végétation etc...) généralement capturés sur le terrain en utilisant des spectromètres. Nous avons utilisé ces bibliothèques ainsi que les caractéristiques techniques de plusieurs capteurs pour simuler différentes vues des spectres de ces bibliothèques. À l'issue de cette étape de simulation nous disposons donc d'un ensemble d'objets tels qu'ils auraient été perçus par les différents capteurs. Chaque objet est donc décrit par un certain nombre d'attributs correspondant au nombre de bandes spectrales du capteur. Ce nombre ainsi que la nature des bandes sont différents pour chaque capteur. On dispose bien alors de vues différents de la même donnée originelle (la librairie spectrale). La seule différence entre les données étant les caractéristiques du capteur utilisé lors de la simulation.

Pour évaluer l'intérêt d'utiliser plusieurs vues conjointement, nous nous sommes intéressés à évaluer l'utilisation de couples de capteurs. L'objectif est d'étudier si l'utilisation de couples de vues provenant de satellites différents améliore la qualité des résultats. Dans la Section 2, nous présentons la problématique de la simulation de capteur, et comment celle-ci est utilisée pour créer des données multisources. Dans la Section 3, nous étudions les approches mises en oeuvre pour traiter ces données et évaluer leur intérêt en utilisation conjointe. Finalement, la Section 4 conclut cet article.

## **2 Simulation de capteurs**

### **2.1 Définition et applications**

Comme indiqué précédemment, le principe de la simulation de capteurs est de générer un spectre multispectral simulé à partir de données acquises à partir de capteurs ayant une meilleure résolution spectrale (hyperspectrale). La simulation consiste à combiner plusieurs bandes hyperspectrales voisines pour former la bande plus large correspondante dans la simulation multispectrale. Elle est réalisée par l'utilisation des fonctions de réponse spectrale rel-

ative ou *Relative Spectral Response* (RSR) du capteur multispectral à simuler. Ces fonctions décrivent la réponse spectrale de chacune des bandes spectrales du capteur. La FIG. 1 présente la fonction de RSR de trois systèmes satellites bien connus : Quickbird, Spot 5 et Landsat.

La simulation de capteurs a été utilisée dans plusieurs types d'applications. Par exemple, Salvatore et al. (1999) ont simulé la réponse d'un nouveau capteur à partir de données AVIRIS hyperspectrales. Cela a permis aux scientifiques d'évaluer le potentiel de leur nouveau capteur multispectral et de paramétrer au mieux les RSR afin d'obtenir de meilleurs résultats en fonction de leurs objectifs. Herold et al. (2003) ont étudié différentes résolutions spectrales pour l'analyse de tissu urbain. Pour cela, ils ont utilisé des données hyperspectrales AVIRIS et une librairie de spectres mesurés sur des matériaux (bitume, tuiles, végétation, etc.) afin de trouver quelles bandes spectrales permettaient de séparer au mieux les classes urbaines d'occupation du sol (bâtiments, routes, végétation, etc.). Les données AVIRIS ont aussi été utilisées pour simuler des données Landsat et Ikonos. Les résultats ont montré que Ikonos et Landsat manquaient de finesse spectrale pour séparer certaines classes urbaines.

## 2.2 Méthodes de simulation

Afin de simuler des données multispectrales à partir de données hyperspectrales, il faut fusionner les bandes hyperspectrales voisines afin de simuler les bandes multispectrales. Cependant, les réflectances des bandes proches à fusionner ne peuvent pas être simplement additionnées. En fait, elles doivent être pondérées en fonction du RSR des bandes multispectrales. Cette sensibilité est décrite pour chaque capteur par sa fonction de RSR.

Comme évoqué par Clark et al. (2002), plusieurs stratégies différentes ont été proposées pour calculer les pondérations à appliquer à chaque bande hyperspectrale. Pour la simulation effectuée dans cet article, chaque longueur d'onde hyperspectrale a été liée avec la moyenne de la RSR (dans l'intervalle de la largeur à mi-hauteur de chaque bande hyperspectral simulée). Cette approche est similaire à celle proposée par Franke et al. (2006) et est décrite en détail dans Forestier et al. (2009).

Il est important de préciser que certains paramètres externes ne sont pas utilisés dans cette étude. Par exemple, d'autres approches (Kavzoglu, 2004) prennent en compte d'autres variables comme les paramètres atmosphériques ou les différences géométriques entre les différents capteurs. Dans nos travaux, nous nous intéressons à la capacité des différents capteurs à séparer les différentes classes en fonction de leur RSR, c'est pourquoi nous nous concentrons sur les différences spectrales entre les capteurs uniquement. Cependant, d'autres aspects comme la résolution spatiale devraient aussi être étudiés afin de mieux appréhender les différences entre les capteurs. Les six capteurs étudiés ici sont : Spot 5, Quickbird, Pleiades, Landsat TM, Ikonos and Formosat (voir TAB. 1).

## 2.3 Librairies spectrales

Une librairie spectrale est une base de données de spectres de plusieurs types de matériaux (minéraux, objets artificiels, végétation, etc.) mesurés *in situ* à partir de spectromètres.

Plusieurs librairies spectrales libres de droit existent. Dans ces travaux, nous avons utilisé la librairie ASTER (Baldrige et al., 2008) qui se compose de spectres provenant du *Jet Propulsion Laboratory* (JPL), de la *John Hopkins University* (JHU) et du *United States Geological Survey* (USGS). Cette librairie comporte des spectres de différentes roches, minéraux,

Étude de données multisources par simulation de capteurs

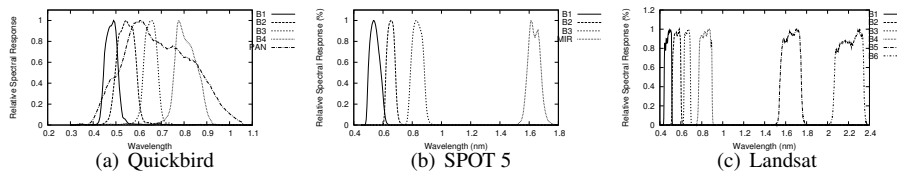


FIG. 1 – Exemple de trois réponses spectrales relatives (RSR).

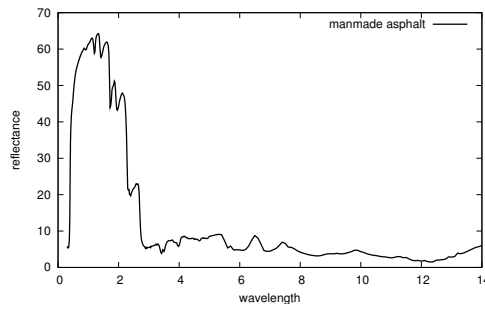


FIG. 2 – Exemple du spectre complet de l'asphalte extrait de la librairie ASTER.

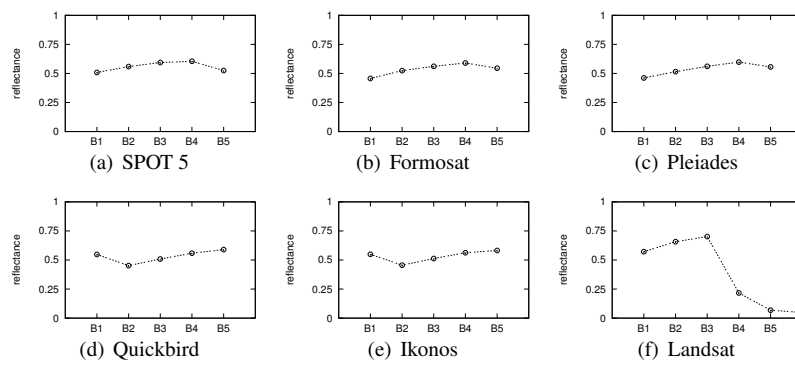


FIG. 3 – Exemple de spectre simulé extrait de la librairie ASTER.

Nom	# Bandes spectrales	Propriétaire
Spot 5	5	CNES
Quickbird	5	Digital Globe
Pleiades	5	CNES
Landsat TM	6	NASA
Ikonos	5	Satellite Imaging Corporation
Formosat	5	Taiwan

TAB. 1 – Liste des capteurs étudiés dans cet article.

sols lunaires, sols terrestres, matériaux artificiels, météorites, végétation, neige et glace, qui couvrent les longueurs d’onde du visible à l’infrarouge thermique (0.4-15.4  $\mu\text{m}$ ). La première version date de juillet 1998 et la seconde est disponible depuis 2007 sur simple demande via le site web d’ASTER<sup>1</sup>. La librairie ASTER est, à notre avis, la plus simple et complète disponible librement.

Les spectres de cette librairie ont été convolués avec le RSR de chaque capteur afin de créer plusieurs points de vue de la librairie. En fait, le processus de simulation permet de construire la *vue* qu’aurait chaque capteur des données disponibles dans la librairie. Ceci est illustré sur la FIG. 2 qui représente le spectre complet de l’asphalte extrait de la librairie ASTER, et la FIG. 3 qui présente le spectre simulé obtenu pour chacun des capteurs étudiés. Le but est d’utiliser ces différentes vues de la même donnée afin d’évaluer le gain apporté par l’utilisation conjointe de plusieurs satellites pour identifier les différentes occupations du sol.

### 3 Évaluation des données multisources

L’objectif de cette section est d’étudier l’intérêt d’utiliser conjointement plusieurs données simulées par la méthode proposée dans la section précédente. Trois approches simples ont été mises en oeuvre pour évaluer le gain potentiel d’utiliser ces données multisources conjointement. Nous nous sommes intéressé ici à étudier la collaboration entre couple de capteurs. L’objectif est d’étudier si le fait d’utiliser des données provenant de deux capteurs permet d’améliorer les performances par rapport à l’utilisation monosource de chacune des données individuellement.

#### 3.1 Clustering collaboratif

Un nombre important de nouveaux algorithmes de clustering ont été développés ces dernières années, et des méthodes existantes ont également été modifiées et améliorées. Cette abondance de méthodes peut être expliquée par la difficulté de proposer des méthodes génériques s’adaptant à toutes les types de données disponibles. En effet, chaque méthode comporte un biais induit par l’objectif choisi pour créer les clusters. Par conséquent, deux méthodes différentes peuvent proposer des résultats de clustering très différents à partir des mêmes données. De plus, le même algorithme peut fournir des résultats différents en fonction de son initialisation ou de ses paramètres. Pour résoudre ce problème, certaines méthodes proposent

<sup>1</sup><http://speclib.jpl.nasa.gov>

d'utiliser plusieurs résultats de clustering différents pour mieux refléter la diversité potentielle des résultats. Ces approches tirent parti des informations fournies par les différents résultats de manière sensiblement différente. Certaines cherchent à créer un résultat unique à partir de plusieurs résultats (Strehl et Ghosh, 2002) en s'intéressant uniquement à la fusion de plusieurs partitions produites par les différentes méthodes. D'autres utilisent des parties de chaque résultat pour construire le résultat final (Law et al., 2004). Une approche faisant également appel à plusieurs résultats est appelée le clustering collaboratif (Wemmert et al., 2000). Celle-ci consiste à utiliser plusieurs méthodes de clustering qui vont collaborer pour proposer un clustering commun d'un même jeu de données. La collaboration peut être définie comme un processus, où deux acteurs ou plus travaillent ensemble pour arriver à un but commun en partageant des connaissances. La première étape du clustering collaboratif consiste à effectuer plusieurs clusterings différents des données. Puis, ces différents résultats sont modifiés pendant une étape de raffinement. Lors de cette étape, chaque résultat est remis en cause à partir des informations proposées par les autres résultats. Dans nos travaux précédents, nous nous sommes majoritairement intéressés à l'utilisation monosource du clustering collaboratif, c'est à dire que chaque méthode de clustering travaille sur la même donnée (voir FIG. 4 (a)). Dans cet article nous avons adopté une approche multisource où chaque méthode va travailler sur une vue différente des données, ici la vue de la librairie spectrale par un des capteurs (voir FIG. 4 (b)). Pour vérifier l'intérêt d'utiliser des données multisources provenant de plusieurs capteurs différents, plusieurs configurations ont été évaluées pour chaque couple de capteurs. Soit  $D_1$  les données issues du premier capteur et  $D_2$  les données issues du second capteur. Les configurations évaluées sont les suivantes :

- $D_1$  : seulement la première vue
- $D_2$  : seulement la seconde vue
- $D_1 + D_2$  : fusion de  $D_1$  et de  $D_2$
- $D_1 \circ D_1$  : collaboration utilisant deux fois  $D_1$
- $D_2 \circ D_2$  : collaboration utilisant deux fois  $D_2$
- $D_1 \circ D_2$  : collaboration utilisant les deux vues  $D_1$  et  $D_2$

La FIG. 5 illustre ces différentes configurations. Dans chacune des expériences, nous avons utilisé l'algorithme KMeans comme méthode de base. Pour chaque configuration ne contenant qu'une seule vue (les trois premières), l'algorithme des KMeans a été initialisé pour trouver un nombre de clusters égal au nombre de classes. Pour les configurations en mode collaboratif (les trois autres), chaque méthode a été initialisée aléatoirement avec un nombre de clusters choisi dans  $\{5 \dots 10\}$ . Ce choix a été fait pour assurer une certaine diversité des deux résultats impliqués dans la collaboration, ce qui est nécessaire pour obtenir une collaboration intéressante entre les deux résultats. Les résultats de clustering obtenus avec et sans collaboration ont été évalués à l'aide d'indices d'évaluation de partition (ex : Rand, Jaccard). Le tableau TAB. 2 illustre les résultats sous forme d'une matrice pour chaque couple de capteurs. Le symbole  $\bullet$  indique que la configuration  $D_1 \circ D_2$  a fourni de meilleurs résultats que toutes les autres configurations,  $\circ$  indique que au moins une autre configuration a donné un meilleur résultat que  $D_1 \circ D_2$ .

Dans ce tableau, il apparaît que les capteurs ayant des fonctions RSR proches ne tirent pas parti de la collaboration. Par exemple, la collaboration des capteurs Quickbird et Pleiades dont les RSR sont très proches n'est pas bénéfique. A contrario, la collaboration de Spot 5 avec Quickbird ou Pleiades semble bénéfique, Spot 5 possédant une bande dans le moyen infra-

$D_1 \backslash D_2$	formosat	pleiades	quickbird	spot5	landsat	ikonos
formosat	-	○	○	●	●	○
pleiades		-	○	●	●	●
quickbird			-	●	●	○
spot5				-	○	●
landsat					-	●
ikonos						-

TAB. 2 – Evaluation de la collaboration de couples de capteurs

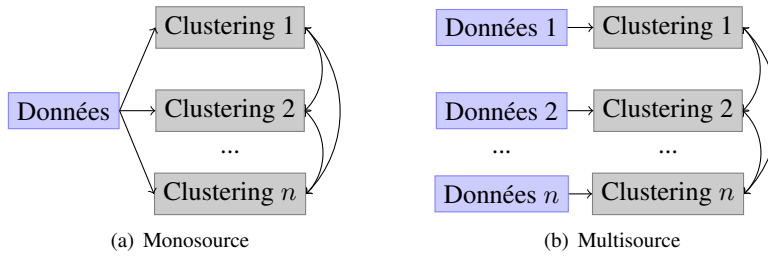


FIG. 4 – Cas monosource (a) et multisource (b)

rouge que ces deux autres capteurs ne possèdent pas. La conclusion que l'on peut tirer de ces premiers résultats est qu'il semble que la collaboration est bénéfique quand les capteurs ne portent pas exactement la même information.

### 3.2 Évaluation par classifieurs supervisés

Les résultats obtenus dans les expériences menées dans la section précédente, montrent qu'il semble être intéressant de faire collaborer plusieurs vues si les données ne sont pas similaires. En effet, d'après les résultats obtenus, plus les capteurs sont différents, et plus ceux-ci apportent des informations différentes et potentiellement complémentaires. Pour valider cette hypothèse, nous avons appris un modèle prédictif à l'aide d'une méthode de classification supervisée (Bayésien Naïf) pour chacune des vues. Nous avons ensuite comparé les prédictions des différents modèles appris. Soit  $p_1$  et  $p_2$  deux modèles prédictifs et  $p_1(o_i)$  et  $p_2(o_i)$  les prédictions pour ces deux modèles pour l'objet  $o_i$  (même objet mais représenté de manière différente par les deux vues). Nous avons calculé un coefficient d'accord qui correspond au pourcentage d'accord sur l'ensemble des prédictions des  $N$  objets :

$$pr = \sum_{i=0}^N \frac{(p_1(o_i) = p_2(o_i))}{N} \quad (1)$$

Le tableau TAB. 3 présente les résultats pour chaque couple de capteurs. Si on met ces résultats en rapport avec ceux obtenus en clustering collaboratif on peut y observer un lien fort (voir valeurs en gras). On observe que pour que l'utilisation de différentes vues soit bénéfique

## Étude de données multisources par simulation de capteurs

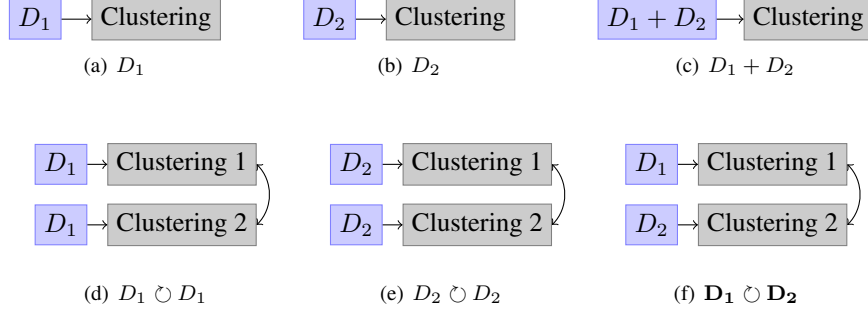


FIG. 5 – Les différentes configurations évaluées.

$D_1 \backslash D_2$	formosat	pleiades	quickbird	spot5	landsat	ikonos
formosat	100	91, 85	97, 78	<b>85, 93</b>	<b>80, 74</b>	99, 26
pleiades		100	94, 07	<b>92, 59</b>	<b>83, 7</b>	<b>92, 59</b>
quickbird			100	<b>88, 15</b>	<b>82, 96</b>	98, 52
spot5				100	91, 11	<b>86, 67</b>
landsat5					100	<b>81, 48</b>
ikonos						100

TAB. 3 – Pourcentage du nombre de fois où les deux classifieurs sont en accord.

il est nécessaire que celles-ci portent des informations différentes. En effet, on peut imaginer qu'utiliser des données fortement similaires risque de ne pas améliorer le processus. Il cependant nécessaire que ces données partagent une certaine cohérence commune, et ne diffèrent que sur certains objets spécifiques. Cela pose le problème de pouvoir évaluer ces différences entre les vues et savoir à quel point ces données doivent diverger, et également de connaître l'impact de données trop dissimilaires.

### 3.3 Évaluation par critère statistique

Enfin, dans cette section, nous avons évalué les similarités entre les différentes vues en utilisant un coefficient de corrélation dans le but de vérifier les résultats obtenus dans les deux sections précédentes. Ce coefficient de corrélation a été calculé comme la moyenne entre les corrélations des différents attributs des différentes vues. Comme chaque satellite possède un nombre de bandes différent et que celles-ci ne sont pas similaires ni ordonnées, il est nécessaire de calculer ce coefficient pour tous les couples de bandes pour un couple de capteurs donné. Soit  $D_1 = (a_1, \dots, a_{n_1})$ ,  $D_2 = (a_1, \dots, a_{n_2})$  les attributs (ou bandes) pour deux capteurs. La corrélation moyenne est évaluée telle que :

$$\mu_{corr} = \sum_i^{n_1} \sum_j^{n_2} \frac{r(a_i, a_j)}{(n_1 * n_2)} \quad (2)$$



$D_1 \backslash D_2$	formosat	pleiades	quickbird	spot5	landsat	ikonos
formosat	-	0,630	0,642	<b>0,592</b>	<b>0,632</b>	0,643
pleiades		-	0,640	<b>0,590</b>	<b>0,631</b>	<b>0,641</b>
quickbird			-	<b>0,602</b>	<b>0,628</b>	0,632
spot5				-	0,676	<b>0,654</b>
landsat					-	<b>0,653</b>
ikonos						-

TAB. 4 – *Corrélation moyenne entre les différentes vues.*

avec  $r(a_i, a_j)$  la coefficient de corrélation linéaire de Bravais-Pearson entre les valeurs des bandes considérées comme deux variables aléatoires :

$$r(a_i, a_j) = \frac{\rho_{a_i a_j}}{\rho_{a_i} \rho_{a_j}} \quad (3)$$

avec  $\rho_{a_i a_j}$  la covariance de  $(a_i, a_j)$  et  $\rho_{a_i}$  et  $\rho_{a_j}$ , respectivement l'écart type de  $a_i$  et  $a_j$ .

## 4 Conclusion

Dans cet article nous avons présenté des travaux sur l'utilisation de données multisources provenant de la simulation de capteurs. Une étape de simulation utilisant une librairie spectrale permet de générer des vues de cette librairie à la résolution de différents capteurs. Ces données ont ensuite été utilisées dans un processus de clustering collaboratif pour étudier l'intérêt de faire collaborer des données issues de couples de capteurs. Une étude utilisant des modèles prédictifs supervisés ainsi que le calcul de coefficients de corrélation entre différentes vues ont permis d'identifier qu'il est intéressant de faire collaborer les vues si celle-ci sont légèrement dissimilaire. La question maintenant posée est de pouvoir mieux évaluer ces dissimilarités, les quantifier, et évaluer quel niveau de dissimilarité est nécessaire pour obtenir une amélioration significative des résultats.

## Références

- Baldrige, A. M., S. J. Hook, C. I. Grove, et R. g. (2008). The aster spectral library version 2.0. *Remote Sensing of Environment*.
- Clark, R., G. A. Swayze, K. Livo, R. F. Kokaly, T. V. V. King, J. B. Dalton, J. S. Vance, B. W. Rockwell, T. Hoefen, et R. R. McDougal (2002). Synthesis of multispectral bands from hyperspectral data : Validation based on images acquired by aviris, hyperion, ali, and etm+.
- Forestier, G., J. Inglada, C. Wemmert, et P. Gancarski (2009). Mining spectral libraries to study sensors' discrimination ability. In *SPIE Europe Remote Sensing*.
- Franke, J., V. Heinzl, et G. Menz (2006). Assessment of ndvi- differences caused by sensor specific relative spectral response functions. *IEEE International Geoscience and Remote Sensing Symposium*, 1138–1141.

## Étude de données multisources par simulation de capteurs

- Govender, M., K. Chetty, et H. Bulcock (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA* 33(2), 145–152.
- Heidena, U., K. Segl, S. Roessner, et H. Kaufmann (2007). Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. *Remote Sensing of Environment* 111, 537–552.
- Herold, M., M. Gardner, et D. Roberts (2003). Spectral resolution requirements for mapping urban areas. *Geoscience and Remote Sensing, IEEE Transactions on* 41(9), 1907–1919.
- Kavzoglu, T. (2004). Simulating landsat etm+ imagery using dais 7915 hyperspectral scanner data. *International journal of remote sensing* 25(22), 5049–5067.
- Law, M., A. Topchy, et A. Jain (2004). Multiobjective data clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 424–430.
- Meyer, D. et G. Chander (2007). The effect of variations in relative spectral response on the retrieval of land surface parameters from multiple sources of remotely sensed imagery. *IEEE International Geoscience and Remote Sensing Symposium*, 5150–5153.
- Salvatore, E., C. Esposito, T. Krug, et R. Green (1999). Simulation of the spectral bands of the ccd and wfi cameras of the cbers satellite using aviris data.
- Strehl, A. et J. Ghosh (2002). Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research* 3, 583–617.
- Wemmert, C., P. Gançarski, et J. Korczak (2000). A collaborative approach to combine multiple learning methods. *International Journal on Artificial Intelligence Tools (World Scientific)* 9(1), 59–78.

## Summary

In recent years, satellite sensor data have become easier to acquire. Several different satellite systems are now available and produce a large amount of data used for Earth observation. To better grasp the complexity of the Earth surface, it became usual to use different images from different satellites. However, it is generally difficult to predict the potential gain of using multisource satellite sensor data before actually acquiring the data. In this paper, we present a simulation approach to create different views of remote sensing sensor data according to different satellite characteristics. These different views are then used in a collaborative clustering approach to assess the interest of using these multisource data together. Experiments provide some insights on couple of satellite systems able to leverage the complementary of the sources.