

Classification non supervisées d’acquisitions hyperspectrales codées : quelles vérités terrain ?

Trung-Tin DINH^{1,2} Hervé CARFANTAN¹ Antoine MONMAYRANT² Simon LACROIX²

¹ Institut de Recherche en Astrophysique et Planétologie (IRAP) – Université de Toulouse/CNRS/CNES,
14 Avenue Edouard Belin, 31400 Toulouse, France

²LAAS-CNRS, Université de Toulouse, CNRS
7 Avenue du Colonel Roche, 31077 Toulouse, France

Résumé – Nous proposons une méthode de classification non supervisée exploitant un nombre réduit d’acquisitions codées issues d’un imageur hyperspectral de type DD-CASSI. Partant d’une modélisation simple de la variabilité spectrale intra-classe, cette approche permet d’identifier des classes et d’estimer des spectres de référence malgré une compression par un facteur dix des données. Nous mettons en évidence ici les limites des vérités terrain couramment utilisées pour évaluer ce type de méthode : absence de définition claire de la notion de classe, forte variabilité intra-classe, voire erreurs de classification. À partir de la scène *Pavia University*, nous montrons qu’avec des hypothèses simples, il est possible de détecter des régions spectralement plus cohérentes, ce qui souligne la nécessité de repenser l’évaluation des méthodes de classification, notamment dans les scénarios non supervisés.

Abstract – We propose an unsupervised classification method using a limited number of coded acquisitions from a DD-CASSI hyperspectral imager. Based on a simple model of intra-class spectral variability, this approach allow to identify classes and estimate reference spectra, despite data compression by a factor of ten. Here, we highlight the limitations of the ground truths commonly used to evaluate this type of method: lack of a clear definition of the notion of class, high intra-class variability, and even classification errors. Using the *Pavia University* scene, we show that with simple assumptions, it is possible to detect regions that are spectrally more coherent, highlighting the need to rethink the evaluation of classification methods, particularly in unsupervised scenarios.

1 Contexte

Contrairement à l’imagerie classique en couleur (RVB) qui utilise seulement trois bandes spectrales, l’imagerie hyperspectrale enregistre des centaines de bandes étroites couvrant une large gamme de longueurs d’onde. Les systèmes d’acquisition traditionnels, tels que les imageurs *pushbroom*, balayent progressivement la scène pour acquérir le cube hyperspectral, ce qui peut nécessiter un temps d’acquisition important et introduire des artefacts dus aux mouvements de la plateforme. À l’opposé, des imageurs dits *snapshot* permettent une acquisition instantanée de la scène, réduisant ainsi le temps d’acquisition, en sacrifiant cependant la résolution spatiale. En alternative, des imageurs à masques codés permettent d’effectuer des acquisitions en codant des mélanges spectro-spatiaux, sans sacrifier la résolution spatiale. En particulier le DD-CASSI (*Double Disperser - Coded Aperture Spectral Snapshot Imager*) [7], préserve la localisation des informations spectrales. Il est alors possible d’analyser la scène hyperspectrale à partir de quelques acquisitions, pour des masques différents. En effet, de nombreux travaux ont proposé des méthodes de reconstruction du cube hyperspectral à partir de telles acquisitions.

Dans ce cadre, nous avons proposé un algorithme de classification spectrale non supervisée [3], c’est-à-dire identifiant automatiquement les régions correspondant à des spectres similaires inconnus *a priori*, à partir de quelques acquisitions codées, en s’affranchissant d’une étape de reconstruction.

Cependant, l’évaluation des résultats d’une telle approche par comparaison à une « la vérité terrain » soulève des difficultés, en particulier à cause d’une grande variabilité spectrale intra-classe. Cette variabilité peut avoir des origines diverses.

Le cas le plus simple, que nous avons considéré ici, prend en compte des variations d’intensité globale du spectre d’un même matériau, variations dues à un éclairage variable. Mais cette variabilité peut bien sûr également provenir de différences spatiales au sein d’un matériau, voire de mélange entre matériaux, ce qui peut poser des questions sur la définition d’une classe. Enfin, une partie de cette variabilité peut être purement induite par des erreurs de labellisation.

Afin d’illustrer ces problématiques, nous nous appuyons dans ce travail sur une scène hyperspectrale largement utilisée dans la communauté appelée *Pavia University* (*Pavia-U*) [6], mais des observations similaires ont été obtenues sur d’autres jeux de données comme *Indian Pines* [6], *CAMCATT* [9].

Dans la section 2 nous présenterons les modèles, hypothèses et outils exploités pour la méthode de classification non supervisée d’images hyperspectrales à partir de données codées que nous avons proposé [3]. Cela nous amènera à préciser ce que nous définissons comme classe dans une telle méthode. Dans la section 3 nous analyserons les données de la classifications de référence fournie avec les données hyperspectrales de la scène *Pavia-U*, en particulier en terme de variabilité spectrale. Enfin, dans la section 4, nous analyserons sur cette base les résultats obtenus par notre méthode sur cette scène, avec 10 fois moins de données codées que le cube hyperspectral.

2 Modélisation et outils exploités

Dans un dispositif d’acquisition codé de type DD-CASSI la scène observée est dispersée spectralement puis projetée à travers un masque codé avant de subir une dispersion inverse

et d'être captée par un détecteur. Pour un objet hyperspectral \mathbf{o} de dimension $R \times C \times W$ (où W est le nombre de longueurs d'onde) on peut effectuer A acquisitions de dimension $R \times C$ pour A configurations différentes du masque. L'objectif est d'analyser la scène avec $A \ll W$, soit beaucoup moins d'acquisitions que de longueurs d'onde du cube (typiquement $A < W/10$).

Afin de combler le manque d'information, il est nécessaire de prendre en compte la corrélation spectro-spatiale de l'objet. Dans un cadre de classification, on peut considérer que les pixels d'une même région k partagent un spectre de référence \mathbf{s}_k avec une éventuelle variabilité. Dans nos travaux, nous considérons un modèle simple de variabilité spectrale à savoir une variabilité locale d'intensité [4, 8, 3], soit pour $\mathbf{o}_n \in \mathbb{R}^W$ le spectre pour le n ème pixel de la région \mathcal{R}_k : $\mathbf{o}_n = \psi_n \mathbf{s}_k$, où $\psi_n \in \mathbb{R}$ le coefficient de variation de l'intensité, est supposé proche de 1. Ce coefficient ψ_n peut aisément s'estimer à partir de l'image panchromatique, elle-même directement accessible par le dispositif (acquisition sans masque).

Partant de ce modèle, nous avons proposé une méthode de classification non supervisée [3], à partir d'un faible nombre d'acquisitions codées, qui exploite plusieurs ingrédients :

- estimation du spectre de référence $\hat{\mathbf{s}}_k$ par reconstruction sur une région homogène par la méthode SA [8] ;
- estimation du coefficient de variabilité $\hat{\psi}_n$ pour les pixels de la région \mathcal{R}_k ;
- suppression des pixels de trop forte variabilité, pour un seuil T : $|1 - \hat{\psi}_n| < T$;
- test de Gaussiannité sur les résidus $\hat{\mathbf{d}}_n - \mathbf{d}_n$, où $\mathbf{d}_n \in \mathbb{R}^A$, pour l'ensemble des pixels $n \in \mathcal{R}_k$ d'une région \mathcal{R}_k correspond aux données acquises pour le pixel n et $\hat{\mathbf{d}}_n$ la prédiction des données pour ce pixel pour le spectre $\hat{\mathbf{s}}_k$ et l'intensité $\hat{\psi}_n$.

Avec ces ingrédients, l'algorithme proposé fonctionne en plusieurs étapes : (1) Détection de régions homogènes \mathcal{R}_k de $P \times P$ pixels (avec $P^2 > W/A$) et du spectre de référence $\hat{\mathbf{s}}_k$ associé validant le test de Gaussiannité et le seuil de faible variabilité ; (2) Croissance de la région \mathcal{R}_k satisfaisant les mêmes conditions ; (3) Fusion de régions satisfaisant les mêmes conditions. Hormis le nombre de pixels P^2 des régions homogènes utilisées pour reconstruire le spectre de référence, seul le seuil de variabilité T doit être réglé par l'utilisateur.

L'objectif ici n'est pas de présenter en détail cet algorithme mais de mettre en avant l'analyse des résultats obtenus par un algorithme de ce type sur des données pour lesquelles on dispose de « vérités terrain. »

3 Pavia-U et variabilité spectrale

La scène *Pavia-U* pour laquelle quelques régions ont été labellisées et sont utilisées comme « vérité terrain, » a été largement utilisée pour évaluer les performances des algorithmes de classification hyperspectrale, qu'il s'agisse de méthodes supervisées ou non supervisées [5], ou de techniques d'apprentissage automatique [2].

Cependant, la variabilité spectrale intra-classe présente dans ces régions labellisées complique la définition même de classes et l'évaluation des résultats de classification. Pour illustrer cette variabilité, nous extrayons deux petites régions appartenant respectivement aux classes de référence *Meadows* et

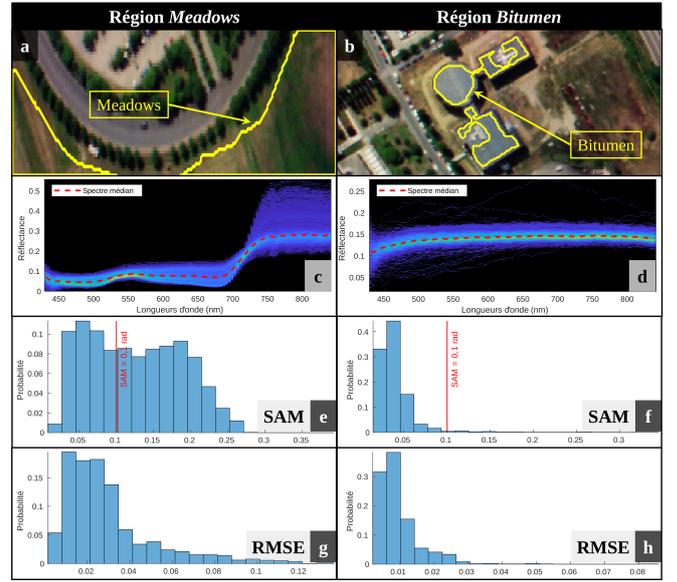


FIGURE 1 : Spectres extraits de la scène *Pavia-U*. De haut en bas : scène observée en couleurs RVB avec région d'intérêt – correspondant à une classe de référence – entourée en jaune (a, b), l'ensemble des spectres avec les valeurs médianes (c, d), histogrammes de SAM (e, f) et RMSE (g, h) entre les spectres de chaque région et leurs spectres médians.

Bitumen fournies avec le jeu de données (voir Fig. 1). En observant l'image en couleurs RVB de la région *Meadows* (Fig. 1a), nous constatons une grande diversité de couleurs parmi les pixels, suggérant une variabilité spectrale significative. Pour la région *Bitumen* (Fig. 1b), les pixels apparaissent plus homogènes en couleurs, mais des variations sont également visibles, notamment dans la partie circulaire du bâtiment.

Afin d'analyser l'impact des limites de la vérité terrain sur l'évaluation des résultats, nous utilisons deux métriques couramment employées en imagerie hyperspectrale, appliquées aux spectres de chaque classe en comparaison avec un spectre de référence (choisi ici comme le spectre médian). La similarité spectrale est quantifiée à l'aide de l'angle spectral (SAM - *Spectral Angle Mapper*), qui est indépendant de l'intensité des spectres. Cet angle est compris entre 0 et π rad, où 0 indique une similarité parfaite. Par ailleurs, l'erreur quadratique moyenne (RMSE - *Root Mean Square Error*) mesure l'écart global entre deux spectres. Pour des spectres de réflectance, les intensités spectrales sont comprises entre 0 et 1, la valeur du RMSE est également bornée entre 0 et 1, où 0 traduit une parfaite correspondance entre les spectres.

Avec l'analyse des spectres (Fig. 1c et 1d), on constate une importante variation de ces spectres au sein de chaque région, aussi bien en termes de forme que d'intensité, même dans la région *Bitumen* qui paraissait plus homogène dans l'image RVB. Dans la région *Meadows*, on distingue visuellement au moins trois groupes de spectres de formes différentes. Pour *Bitumen*, bien que la majorité des spectres soient proches, certains apparaissent clairement aberrants. Cette variabilité se retrouve dans les valeurs de SAM (Fig. 1e, 1f) et de RMSE (Fig. 1g, 1h). Par exemple, 59,8% des pixels de *Meadows* ont un SAM $> 0,1$ rad, ce qui dépasse le seuil usuel admis pour une même classe dans l'outil de classification d'ENVI [1]. Pour *Bitumen*, ce pourcentage est de 2,3%, confirmant

son homogénéité relative. Les RMSE suivent une tendance similaire, traduisant une variabilité plus marquée à *Meadows* qu'à *Bitumen*.

Nous sommes ainsi face à deux situations contrastées : la région *Meadows* présente une forte variabilité spectrale, suggérant que la vérité terrain gagnerait à être subdivisée en sous-régions plus homogènes ; à l'inverse, la région *Bitumen* apparaît globalement bien labellisée, hormis quelques spectres aberrants qui pourraient être exclus pour améliorer la cohérence spectrale.

4 Analyse des résultats

Afin de générer des données codées réalistes à partir de cubes hyperspectraux de référence, nous utilisons le simulateur SIMCA [10]. Ce simulateur intègre les paramètres instrumentaux du système d'acquisition, en prenant en compte les distorsions optiques ainsi que les caractéristiques du masque codé. Les données codées ont été simulées sur $A = 10$ acquisitions, soit une quantité de données dix fois inférieure à celle du cube hyperspectral complet ($W = 103$).

Notre algorithme de classification non supervisée est ensuite appliqué aux données codées générées, en fixant une valeur de T . Pour chaque valeur de T , nous obtenons un ensemble de régions homogènes et des régions non classées, ces dernières correspondant aux zones de mélange ou à des zones où l'information est insuffisante. Dans la Fig. 2, les régions homogènes sont représentées en couleurs, les pixels blancs indiquent des pixels non classés, et les pixels noirs, en dehors de la région d'intérêt, sont exclus de la comparaison.

Outre les régions labellisées, notre algorithme fournit un spectre reconstruit régularisé pour chaque région homogène, basé sur la méthode SA [8]. Toutefois, dans le contexte de classification, nous choisissons d'utiliser, pour chaque région détectée, le spectre médian calculé à partir des spectres complets des pixels de la région. Ce choix permet une comparaison équitable avec la classe de référence, elle-même représentée par son spectre médian. En effet, le spectre régularisé n'est disponible que pour les régions issues de notre algorithme, et ne peut être estimé pour les classes de la vérité terrain, ce qui introduirait un biais dans l'évaluation. Les spectres médians ainsi obtenus sont tracés dans la Fig. 2 avec les mêmes couleurs que les régions correspondantes, et le spectre de la vérité terrain est représenté en trait pointillé.

Pour la région *Meadows*, notre algorithme a identifié 7 régions distinctes avec $T = 0,2$ et 20 régions pour $T = 0,05$ (Fig. 2a et 2c). Leurs spectres médians diffèrent notablement de la référence, mais sont cohérents avec notre observation de la variabilité spectrale et de l'existence de plusieurs spectres différents dans la Fig. 1c. Pour *Bitumen*, notre algorithme a identifié 1 région homogène avec $T = 0,2$ et 4 régions avec $T = 0,05$ (Fig. 2b et 2d). Cela est peu surprenant, car cette région est plus homogène que celle de *Meadows*. Pour $T = 0,2$, quelques pixels aberrants ont été supprimés (Fig 3d), mais la majorité des pixels sont classés dans une même région homogène correspondant dans les grandes lignes à la classification de référence. Notons que plus le seuil T est petit, plus le nombre de régions homogènes détectées est important, et plus les spectres médians de ces régions sont différents du spectre de référence de la classe de référence.

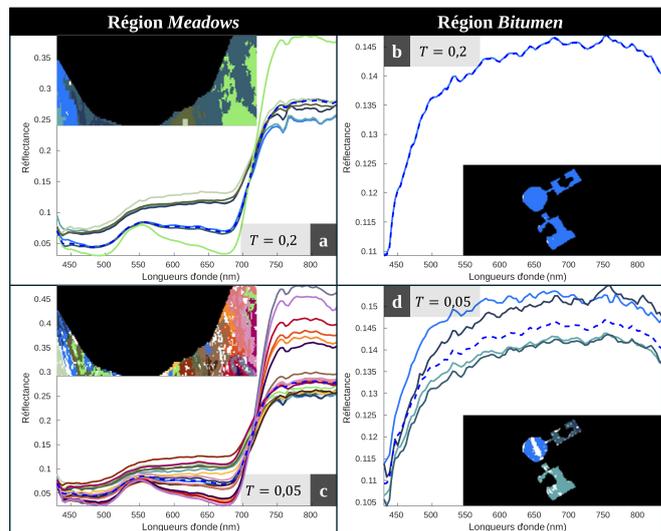


FIGURE 2 : Résultats de classification (régions classées et spectres médians correspondants) pour différentes valeurs de T . Images des régions labellisées : *couleurs* - régions homogènes, *blanc* - régions non classées, *noir* - non concernées ; spectres médians : *trait pointillé* - spectre de référence, *trait plein* - spectre médian de chaque région homogène.

Pour quantifier la variabilité intra-classe, dans chaque région labellisée, nous calculons le SAM entre les spectres des pixels et le spectre médian. Les cartes de SAM correspondantes sont présentées en Fig. 3. Pour la région *Meadows*, la

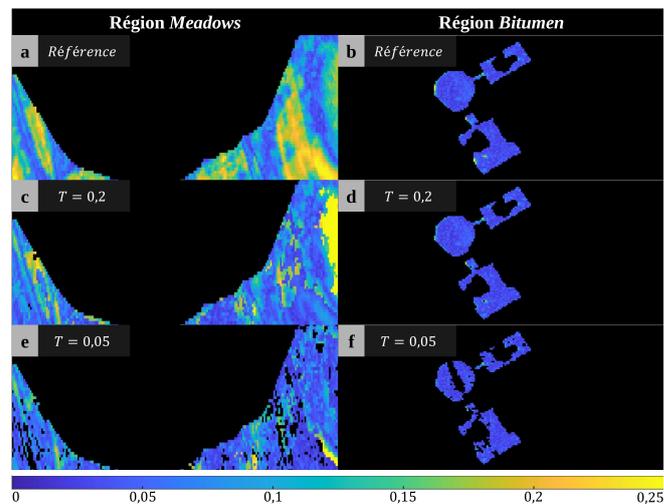


FIGURE 3 : Cartes de SAM de la vérité terrain pour différentes valeurs de T .

carte SAM associée à la classification de référence (Fig. 3a) montre une forte variabilité spectrale, avec de larges zones présentant des valeurs élevées, certaines étant même nettement discernables sur l'image RVB par des couleurs distinctes. En appliquant notre méthode avec $T = 0,2$ (Fig. 3c), on observe une nette diminution des valeurs de SAM dans la majorité des régions, traduisant une plus faible variabilité intra-classe. Pour $T = 0,05$ (Fig. 3e), la réduction est encore plus marquée, sauf dans une petite zone localisée où il existe probablement un mélange complexe de matériaux. Pour la région *Bitumen*, la carte SAM de la classe de référence (Fig. 3b) montre une variabilité globalement faible, mais quelques petites zones

isolées présentent des valeurs plus élevées. Avec $T = 0,2$ (Fig. 3e), notre méthode élimine ces pixels aberrants, réduisant la variabilité. Enfin, pour $T = 0,05$ (Fig. 3f), la carte de SAM devient encore plus homogène. Ainsi, l'analyse des cartes de SAM montre que notre méthode permet de mieux détecter les régions véritablement homogènes, présentant une faible variabilité intra-classe, contrairement à la classification de référence qui regroupe parfois des régions présentant des spectres dissemblables, voire des spectres aberrants.

Les histogrammes de probabilité (Fig. 4) permettent de visualiser la distribution des valeurs de SAM et de RMSE suivant la valeur du seuil T , ainsi que pour la classification de référence. Les histogrammes sont présentés sous forme d'images où chaque ligne correspond à une valeur de T , la dernière ligne représentant les valeurs issues de la classification de référence ; les valeurs des probabilités étant indiquées par des couleurs. Nous constatons que les distributions de SAM et de RMSE des

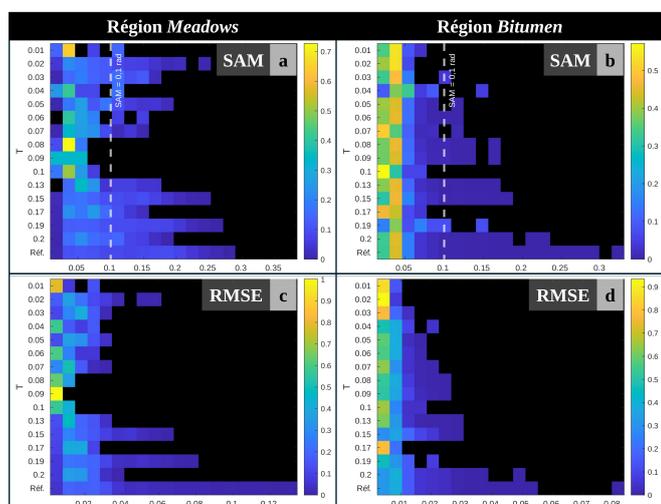


FIGURE 4 : Histogrammes de probabilité des SAM et RMSE pour différentes valeurs de T et de la vérité terrain.

régions classées par notre algorithme sont plus concentrées autour de 0 et présentent des valeurs moyennes plus faibles que celles des régions de la classification de référence, traduisant une meilleure cohérence spectrale intra-classe.

Pour la région *Meadows*, on observe que les histogrammes issus de notre classification deviennent progressivement plus étroits et concentrés autour de 0 lorsque T diminue jusqu'à un seuil $T \sim 0,09$, ce qui reflète une diminution de la variabilité spectrale intra-classe. En deçà de ce seuil, les distributions s'élargissent à nouveau. Cela est dû à des difficultés rencontrées par notre algorithme pour l'estimation de spectres de référence sur certaines régions de petite taille, ce qui déborde du cadre de cette étude. Néanmoins, l'histogramme associé à la classification de référence reste globalement plus large, confirmant sa moindre homogénéité spectrale.

Pour la région *Bitumen*, bien que la vérité terrain montre déjà une certaine homogénéité (distributions initialement resserrées), nos résultats en réduisant la valeur de T permettent de réduire encore davantage la dispersion des valeurs de SAM et RMSE (Fig. 4b et 4d) en excluant progressivement des spectres aberrants.

Ces résultats confirment que notre méthode identifie des régions plus cohérentes du point de vue de variabilité spectrale, tandis que la vérité terrain, utilisée comme référence, ne reflète

pas toujours fidèlement l'homogénéité spectrale réelle des régions, regroupant parfois des spectres hétérogènes, ce qui peut fausser l'évaluation.

5 Conclusions

Ce travail met en évidence les limites des vérités terrain dans les jeux de données hyperspectrales, souvent considérées comme des références absolues, notamment pour l'évaluation des méthodes de classification non supervisée. Tout en cherchant à analyser les résultats d'un algorithme spécifique de classification à partir d'acquisitions codées, nous avons montré que ces vérités terrain, bien qu'issues d'annotations expertes, peuvent introduire un biais significatif dans l'évaluation, en regroupant sous une même classe des pixels aux signatures spectrales pourtant dissemblables. Ce problème n'est pas spécifique à une scène donnée : il se retrouve aussi bien dans les bases de données classiques [6] (*Indian Pines, Pavia Center, Salinas, ...*) que dans des jeux récents et rigoureusement construits comme *CAMCATT* [9], où des spectres de référence sont pourtant mesurés *in situ* à l'aide de spectromètres.

Plus généralement, la modélisation et la prise en compte de variabilité spectrale intra-classe est un enjeu difficile, qui remet en question la définition de la notion de classe et donc celle de vérité terrain. Comme nous l'avons illustré, cela rend délicat l'évaluation des méthodes de segmentation, mais on peut également se questionner sur l'influence de telles définitions de classes sur des algorithmes d'apprentissage s'appuyant sur ces « vérités terrain. »

Références

- [1] NV5 Geospatial Software, ENVI Classic Tutorial: Spectral Angle Mapper (SAM) and Spectral Information Divergence (SID) Classification.
- [2] M. AHMAD, S. SHABBIR, S. K. ROY, D. HONG, X. WU et J. YAO : Hyperspectral Image Classification—Traditional to Deep Models : A Survey for Future Prospects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:968–999, 2021.
- [3] T.-T. DINH, H. CARFANTAN, A. MONMAYRANT et S. LACROIX : Statistical Tests for Hyperspectral Coded Data Unsupervised Classification. *In IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Helsinki, Finland, 2024.
- [4] L. DRUMETZ, M. VEGANZONES, S. HENROT, R. PHLYPO, J. CHANUSSOT et C. JUTTEN : Blind Hyperspectral Unmixing Using an Extended Linear Mixing Model to Address Spectral Variability. *IEEE Transactions on Image Processing*, 25(8):3890–3905, août 2016.
- [5] M. FAUVEL, Y. TARABALKA, J. A. BENEDIKTSSON, J. CHANUSSOT et J. C. TILTON : Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proc. IEEE*, 101(3):652–675, mars 2013.
- [6] P. GAMBA, M. GRAÑA, M. A. VEGANZONS et B. AYERDI : Pavia University, 2001.
- [7] M. E. GEHM, R. JOHN et D. J. BRADY : Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):1013–14027, nov. 2007.
- [8] E. HEMSLEY, I. ARDI, T. ROUVIER, S. LACROIX, H. CARFANTAN et A. MONMAYRANT : Fast reconstruction of hyperspectral images from coded acquisitions using a separability assumption. *Journal of the Optical Society of America A*, 30(5):8174–8185, fév. 2022.
- [9] L. ROUPIOZ, X. BRIOTTET, K. ADELIN, A. AL BITAR, D. BARBONDUBOSC, R. BARDA-CHATAIN, P. BARILOTT, S. BRIDIER, E. CARROLL et C. CASSA : Multi-source datasets acquired over Toulouse (France) in 2021 for urban microclimate studies during the CAMCATT/AI4GEO field campaign. *Data in Brief*, 48, 2023.
- [10] A. ROUXEL, A. MONMAYRANT, S. LACROIX, H. CAMON et S. LOPEZ : Accurate ray-tracing optical model for coded aperture spectral snapshot imagers. *Applied Optics*, 63(7):1828–1838, 2024.