



**DOSSIER DE CANDIDATURE  
POUR THESE EN COTUTELLE  
POUR LA RENTREE 2022  
FINANCEMENT : BOURSE**

Dossier complété et revêtu des signatures à transmettre impérativement pour le :

**09 mai 2022 12h00 au plus tard,**

A la Direction de la Recherche et Valorisation

[secretariat.recherche@univ-littoral.fr](mailto:secretariat.recherche@univ-littoral.fr)

**Titre de la thèse : Apprentissage automatique profond par factorisation matricielle multi-objective évolutionnaire. Applications en imagerie satellitaire urbaine**

**Laboratoire d'accueil ULCO : LISIC**

**Directeur de thèse ULCO : Gilles ROUSSEL et Matthieu Puigt**

**Si nouveau partenariat (absent ci-dessous) merci de nous indiquer l'université d'origine, le directeur de thèse, le laboratoire, ainsi que les conditions de financement du doctorant dans le pays d'origine**

**☐ LIBAN - Université Libanaise (2 financements)**

Pour ce dispositif, merci d'indiquer en plus :

- le nom du codirecteur étranger et le laboratoire partenaire

- Thématique :

- (1) La qualité de l'air
- (2) Le milieu aquatique
- (3) L'obésité, la nutrition et les activités sportives,
- (4) Les énergies propres et renouvelables
- (5) La gestion et le traitement des déchets
- (6) L'urbanisme

**LIBAN - CNRS Libanais (4 financements)**

Pour ce dispositif, merci d'indiquer en plus :

- le nom du codirecteur étranger et le laboratoire partenaire : **discussions en cours avec Ghaleb Faour, Centre National de Télédétection, CNRS Libanais**

- Thématique :

- (1) La qualité de l'air
- (2) Le milieu aquatique



- (3) L'obésité, la nutrition et les activités sportives
- (4) Les énergies propres et renouvelables
- (5) La gestion et le traitement des déchets
- (6) L'urbanisme

#### \* LABORATOIRE D'ACCUEIL

Nom du laboratoire d'accueil : LISIC

Nombre de HDR dans le laboratoire : 18

Nombre de thèses encadrées dans le laboratoire (rentrée 2020) : 26

Cotutelles en cours au sein du laboratoire : 11

Durée moyenne des thèses soutenues dans le laboratoire, sur la période 2015-2020 : 3,57

#### ENCADREMENT

Nom, Prénom du directeur de laboratoire : VEREL Sébastien

Nom, Prénom du directeur de thèse (si différent du directeur de laboratoire) : ROUSSEL Gilles

Nombre de doctorats en préparation sous la direction du directeur de thèse : 2 doctorants

Avis détaillé du directeur de thèse :

Ce projet de thèse est une opportunité très intéressante pour engager un partenariat avec les équipes du Centre national de télédétection du CNR Libanais, vu la complémentarité évidente avec nos travaux engagés depuis deux ans sur l'analyse et la fusion d'image satellitaires et l'expérience acquise dans le domaine des images multi et hyperspectrales. Les applications en imagerie urbaine est une intersection évidente avec les travaux menés par le Centre National de Télédétection du CNRS Libanais. Les méthodes envisagées sont à la croisée de l'apprentissage profond par décomposition faible rang (DeepNMF) en Machine Learning (ML), l'optimisation des fonctionnelles multi-objectif (MOE) et l'imagerie hyperspectrale. L'expérience scientifique des co-directeurs doit permettre une collaboration très prolifique. Je suis bien sûr enthousiaste à l'idée de co-diriger avec Matthieu Puigt cette thèse pour le côté ULCO.

Signature du directeur de thèse



Avis détaillé du directeur de laboratoire :

Le développement des activités du laboratoire dans le domaine de l'apprentissage automatique (machine learning) correspond à une priorité de notre unité de recherche. Cette thèse s'inscrit dans les travaux de l'équipe Specifi qui concerne l'apprentissage non-supervisé à l'aide des techniques de factorisation matricielle (NMF). De plus, le sujet fait appel à d'autres expertises du laboratoire concernant l'optimisation multi-objectif. Ce projet de thèse viendrait parfaitement renforcer plusieurs aspects prioritaire du laboratoire.

L'application cible et support des travaux de thèse fait suite à des travaux principalement développer à l'antenne de Saint Omer du LISIC sur l'imagerie hyperspectrale et les dynamiques urbaines. Pour toutes ces raisons, j'émets un avis favorable à son financement.

Signature du directeur de laboratoire

A handwritten signature in black ink, appearing to read 'S. Amis', is written below the text 'Signature du directeur de laboratoire'.



## **PROJET DE THESE**

Intitulé du projet de thèse : **Apprentissage automatique profond par factorisation matricielle multi-objective évolutionnaire. Applications en imagerie satellitaire urbaine**

Domaine scientifique : Traitement du signal et des images

Résumé (1/2 page maxi.) :

L'imagerie hyperspectrale satellitaire a connu de nombreux développements durant ces dernières décennies, fournissant notamment pour chaque pixel une signature spectrale propre au(x) matériau(x) observé(s). Le démélange de ces images fournit à la fois l'ensemble des signatures spectrales des matériaux observés et des cartes d'abondance de chacun de ces matériaux, permettant par la suite de nombreux post-traitements (classification, segmentation, modélisation, etc). La factorisation en matrices non-négatives (NMF) est une des méthodes phare de démélange. Cet outil est très populaire en apprentissage, traitement du signal et traitement des images, de part la meilleure interprétabilité qu'il permet par rapport à des approches sans contrainte de signe. Cependant, elle nécessite souvent dans son optimisation le rajout de fonctions de pénalisation pour générer de meilleurs résultats. Choisir les bonnes fonctions de pénalisation et leur pondération a priori nécessite un degré d'expertise qui peut être relâché par l'utilisation d'approches d'optimisation multi-objectif évolutionnaires. Quelques récents travaux sur l'utilisation de ces méthodes pour l'optimisation bi-objective ont montré l'intérêt de ces techniques mais leur extension à de nombreuses fonctions objectif reste un problème ouvert. De plus, des extensions profondes de NMF (ou Deep NMF) ont été proposées récemment et ont montré leur supériorité à la NMF conventionnelle. Ces méthodes, qui peuvent aussi nécessiter des termes de pénalisation dans les différentes couches de la Deep NMF, n'ont pas encore été combinées à l'optimisation multi-objectif. L'objectif de cette thèse consistera à développer de nouvelles approches de Deep NMF combinées à l'optimisation multi-objectif évolutionnaire. Ces méthodes seront appliquées aux données hyperspectrales satellitaires, pour des applications urbaines (caractérisation des couvertures des sols, analyse des points chauds urbains, détection automatique de plans d'eau, etc).

Projet de thèse (5 pages maxi.) :



## **Apprentissage automatique profond par factorisation matricielle multi-objective évolutionnaire. Applications en imagerie satellitaire urbaine.**

*Mots clés : factorisation en matrices non-négatives ; données de grande dimension ; optimisation multi-objective évolutionnaire ; factorisation multi-couches ; apprentissage profond ; imagerie satellitaire*

### **I. Contexte scientifique :**

L'imagerie satellitaire a connu de nombreux développements depuis plusieurs décennies, en particulier via l'ajout d'information couleur puis spectrale associée à chaque pixel de l'image observée. En parallèle de ces développements instrumentaux, de nombreux développements méthodologiques en traitement du signal et des images ont été réalisés. Parmi ceux-ci, le démixage hyperspectral [1] a connu un vif intérêt de la part de la communauté scientifique. Il permet notamment d'extraire à partir d'un cube de données hyperspectrales un ensemble de signatures spectrales (*endmembers*) des matériaux observés depuis l'espace, mais aussi une carte d'abondances – ou proportions – de chacun de ces matériaux. Ces *endmembers* et cartes d'abondances permettent de nombreux post-traitements comme la segmentation ou la classification. L'un des intérêts majeurs du démixage réside dans sa capacité à extraire à la fois les *endmembers* et les abondances dans les pixels où plusieurs matériaux sont présents.

Parmi les méthodes phares de démixage, la factorisation en matrices non-négatives (NMF) a aussi connu un vif intérêt de la part de la communauté dès la fin des années 1990 et les travaux de Lee et Seung [2]. Elles ont notamment été appliquées avec succès en traitement d'images [2], traitement du son [3], observation spatiale [4], chimie environnementale [5], analyse de réseaux sociaux [6], ou encore en étalonnage de capteurs mobiles [7]. Cependant, lorsqu'elles sont appliquées à l'imagerie satellitaire, elles doivent souvent être informées pour fournir de bonnes performances. En particulier, de nombreuses extensions de la NMF pour le démixage hyperspectral proposent de pénaliser sa fonction de coût en rajoutant des contraintes de parcimonie [8], de non-linéarité parcimonieuse [9], de volume [11], etc. Le choix de ces contraintes et le réglage a priori des pondérations associées nécessitent un degré d'expertise avancé. De manière alternative, l'optimisation multi-objectif évolutionnaire (MOE) déplace le champs de compétences « métier » dans l'analyse a posteriori des différentes fonctions de pénalisation [12]. Un petit nombre de travaux existe en NMF-MOE [13-16] : Les travaux dans [13] et [14] proposent de combiner la NMF avec des stratégies de recuit simulé et d'optimisation évolutionnaire pour contraindre un niveau de parcimonie sur un des facteurs. Les travaux dans [15] proposent de trouver un compromis entre deux NMF différentes pour le démixage hyperspectral, l'une étant le pendant non-linéaire de l'autre. Pour cela, une extension de la NMF, consistant en une co-factorisation (CoNMF) est proposée et le front de Pareto est étudié de manière extensive (sans aucun apport de l'optimisation évolutionnaire). Enfin, les travaux dans [16] proposent d'étudier plusieurs fonction de coût qui présentent divers degrés de robustesse à différents types de bruit. Nous proposons de nous appuyer sur ces travaux pour les généraliser.

Par ailleurs, l'apprentissage profond (ou *deep learning* en anglais) a révolutionné l'intelligence artificielle depuis une dizaine d'années [17]. Il est démontré depuis longtemps qu'une



couche de réseau de neurones peut être ré-écrite comme un produit matriciel [2]. Suite aux travaux pionniers en *multi-layered NMF* [18], les approches de deep NMF (qui rajoutent une étape de back-propagation dans l'optimisation de la *multi-layered NMF*) ont été proposées [19]. En démixage hyperspectral, des premiers travaux ont été ainsi proposés [20]. Cependant, des termes de pénalisation peuvent aussi être rajoutés dans les différentes couches de la factorisation profonde. Une fois de plus, nous pensons que la MOE peut aider le développement des approches de Deep NMF.

Comme indiqué précédemment, les méthodes de NMF permettent un grand nombre d'applications. Dans le cadre de ce projet, d'un point de vue applicatif, nous nous focaliserons sur l'apport du démixage hyperspectral pour l'urbanisme. En effet, les signatures spectrales et les cartes d'abondance apportent de nombreuses connaissances qui peuvent être utiles. Par exemple, les signatures spectrales permettent de mieux caractériser les matériaux et de régler les coefficients d'émissivité IR pour mieux détecter et estimer les îlots de chaleurs urbains (soit via les données satellitaires qui auraient dans bandes spectrales dans le domaine infra-rouge, soit en mesure proche in-situ aéroporté avec des caméras thermiques). Les signatures spectrales permettent aussi de détecter les plans d'eau et donc potentiellement détecter les fraudes dans des régions où l'eau peut manquer ou encore les foyers probables de propagation de moustiques.

## II. Sujet de recherche :

Ce sujet de thèse sera découpé en 2 parties :

1. Dans un premier temps, nous chercherons à proposer une méthode "universelle" de NMF utilisant l'optimisation multi-objective (MOE). Nos méthodes de NMF-MOE tiendront compte d'un ensemble fonctions de contraintes propres aux problèmes considérés et proposeront un ensemble de solutions "optimales au sens de Pareto" pour l'aide à la décision. L'objectif est de généraliser les travaux [13-16] en proposant de gérer bien plus de contraintes que ce qui est actuellement réalisé dans la littérature. Dans un deuxième temps, nous chercherons à automatiser l'apprentissage de ces fonctions de contraintes en cherchant à éliminer automatiquement celles qui seraient inutiles pour répondre au problème d'apprentissage. Pour cela, un formalisme original consisterait à apprendre les hyper-paramètres de la NMF-MOE comme un problème de MOE en soit. Ce premier travail est déjà ambitieux car il permettrait de sérieuses avancées à la fois en apprentissage pour le traitement du signal et des images, mais aussi en recherche opérationnelle qui est le domaine scientifique dans lequel l'optimisation MOE est la plus active.
2. Dans un deuxième temps, il s'agira d'étendre ces approches "universelles" au cas de la factorisation profonde. Chaque couche cachée de NMF sera potentiellement contrainte et il s'agira de trouver, pour chaque couche, les contraintes les mieux adaptées au problème. Ces techniques automatisées pourraient alors concurrencer les approches basées sur des réseaux de neurones pour l'analyse de données.

## III. Travail et échéancier proposés :



Tâche	Durée	Période
Bibliographie	6 mois	De Mois-1 à Mois-34
Développement d'approches de NMF-MOE	12 mois	De Mois-3 à Mois-18
Développement d'approches de Deep NMF-MOE	9 mois	De Mois-18 à Mois-33
Validation méthodes et rédaction d'articles	6 mois	De Mois-18 à Mois-34
Rédaction thèse	3 mois	De Mois-30 à Mois-36
<b>Total</b>	<b>36 mois</b>	<b>De Mois-1 à Mois-36</b>

#### **IV. Etat du sujet dans le laboratoire et l'équipe d'accueil :**

La NMF fait l'objet de nombreux travaux dans l'équipe depuis de nombreuses années. Nous avons notamment obtenu des contrats industriels (ArcelorMittal) et institutionnels (DREAL, Région, CNRS, FEDER) autour de cette thématique. Nous avons pu développer de nouvelles approches robustes et informées que nous appliquées à la séparation de sources de pollution atmosphérique, en collaboration avec l'UCEIV de l'ULCO. Nous avons aussi proposé des approches spécifiques pour conjointement étalonner des réseaux de capteurs mobiles de la qualité de l'air (internet des objets) et réaliser des cartographies de la qualité de l'air. Dans cette application, la matrice de données à factoriser est de grande taille mais partiellement connue (présence de nombreuses données manquantes). Nous avons proposé plusieurs approches structurées et informées de NMF pondérée à cet effet et en avons proposé des extensions rapides dans le cadre d'un formalisme d'apprentissage comprimé. Nos travaux en cours revisitent des problèmes d'imagerie hyperspectrale computationnelle via la NMF. Dans ces problèmes, nous proposons des approches de NMF localement de rang-1 qui permettent de restaurer un cube de données (dématriçage) tout en réalisant le démixage hyperspectral.

Le travail de thèse que nous proposons ici serait donc parfaitement complémentaire des travaux de l'équipe puisque nous tirerions partie de notre expertise. Par ailleurs, la MOE est un objet d'étude de l'équipe OSMOSE du LISIC, qui pourra collaborer avec nous sur ces travaux.

#### **V. Retombées scientifiques et économiques attendues :**

Ces travaux nous permettront de nous positionner de manière originale par rapport aux communautés en traitement du signal et des images d'une part, en apprentissage automatique (*machine learning*) d'autre part, et enfin en recherche opérationnelle. Ces travaux nous permettront de renforcer les activités de l'antenne de Longuenesse du LISIC et les collaborations internes au LISIC, nous permettant ainsi de pouvoir répondre de manière originale à des appels à projets nationaux (ANR) et européens (H2020).



## VI. Collaborations prévues :

Comme indiqué précédemment, ces travaux seront réalisés avec le concours d'une autre équipe du LISIC, spécialiste de l'optimisation multi-objective évolutionnaire. Par ailleurs, des discussions sont en cours avec le centre national de télédétection du CNRS libanais, qui possède une solide expérience dans l'analyse et les usages de l'imagerie satellitaire.

## VII. Publications significatives de l'équipe en lien avec le sujet (membres actuels de l'équipe en gras, stagiaires et doctorants de l'équipe soulignés) :

1. K. Abbas, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Rank-One Completion-Based Joint Unmixing and Demosaicing Method for Snapshot Spectral Images*, submitted to EUSIPCO'22.
2. **F. Yahaya**, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Random Projection Streams for (Weighted) Nonnegative Matrix Factorization*, Proc. ICASSP 2021, pp. 3280-3284, Virtual, June 2021.
3. O. Vu thanh, **M. Puigt**, F. Yahaya, **G. Delmaire**, **G. Roussel**, *In situ calibration of cross-sensitive sensors in mobile sensor arrays using fast informed non-negative matrix factorization*, Proc. of ICASSP 2021, Virtual, pp. 3515-3519, June 2021.
4. **A. Alboody**, **M. Puigt**, **G. Roussel**, V. Vantrepotte, C. Jamet, T. K. Tran, *Experimental comparison of multi-sharpening methods applied to Sentinel-2 MSI and Sentinel-3 OLCI images*, Proc of IEEE WHISPERS 2021, Virtual, March 24-26, 2021.
5. **G. Delmaire**, **M. Omdivar**, **M. Puigt**, F. Ledoux, A. Limem, **G. Roussel**, D. Courcot, *Informed Weighted Non-negative Matrix Factorization Using  $\alpha\beta$ -Divergence Applied to Source Apportionment*, Entropy, Volume 21, Issue 3, Article number 253, special issue on "Information Theory Applications in Signal Processing", March 2019.
6. C. Dorffer, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration*, IEEE Transactions on Signal and Information Processing over Networks, Volume 4, Issue 4, pp. 667-682, December 2018.
7. A. Limem, **G. Delmaire**, **M. Puigt**, **G. Roussel**, D. Courcot, *Non-negative matrix factorization under equality constraints—a study of industrial source identification*, Applied Numerical Mathematics (APNUM), Volume 85, pp. 1-15, November 2014.

## VIII. Références :

- [1] Ma, W. K., Bioucas-Dias, J. M., Chan, T. H., Gillis, N., Gader, P., Plaza, A. J., ... & Chi, C. Y. (2014). A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1), 67-81.
- [2] Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788.
- [3] Virtanen, T. (2007). Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE TASLP*, 15(3), 1066-1074.
- [4] Berne, O., Joblin, C., Deville, Y., Smith, J. D., Rapacioli, M., Bernard, J. P., ... & Abergel, A. (2007). Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using blind signal separation methods. *Astronomy & Astrophysics*, 469(2), 575-586.



- [5] Limem, A., Delmaire, G., Puigt, M., Roussel, G., & Courcot, D. (2014). Non-negative matrix factorization under equality constraints—a study of industrial source identification. *Applied Numerical Mathematics*, 85, 1-15.
- [6] Yang, J., & Leskovec, J. (2013, February). Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proc. of the sixth ACM intl. conf. on Web search and data mining* (pp. 587-596). ACM.
- [7] Dorffer, C., Puigt, M., Delmaire, G., & Roussel, G. (2018, December). Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration. *IEEE TSIPN*, 4(4):667-682.
- [8] Yang, Z., Zhou, G., Xie, S., Ding, S., Yang, J. M., & Zhang, J. (2010). Blind spectral unmixing based on sparse nonnegative matrix factorization. *IEEE Transactions on Image Processing*, 20(4), 1112-1125.
- [9] Févotte, C., & Dobigeon, N. (2015). Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. *IEEE Transactions on Image Processing*, 24(12), 4810-4819.
- [10] Li, J., Bioucas-Dias, J. M., Plaza, A., & Liu, L. (2016). Robust collaborative nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6076-6090.
- [11] Ang, A. M. S., & Gillis, N. (2019). Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12), 4843-4853.
- [12] Emmerich, M., & Deutz, A. H. (2018). A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, 17(3), 585-609.
- [13] Foley, K., Greene, D., & Cunningham, P. (2010, August). Optimizing conflicting objectives in NMF using Pareto simulated annealing. In *Paper presented at the 21st National Conference on Artificial Intelligence and Cognitive Science (AICS 2010), Galway, Ireland, 2010*.
- [14] Gong, M., Jiang, X., Li, H., & Tan, K. C. (2018). Multiobjective Sparse Non-Negative Matrix Factorization. *IEEE Transactions on Cybernetics*.
- [15] Zhu, F., & Honeine, P. (2016). Biobjective nonnegative matrix factorization: Linear versus kernel-based models. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7), 4012-4022.
- [16] Gillis, N., Leplat, V., & Tan, V. (2021). Distributionally robust and multi-objective nonnegative matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [17] Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, (5), 173-220.
- [18] Cichocki, A., Zdunek, R., & Amari, S. I. (2007, September). Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In *Proc. ICA* (pp. 169-176). Springer, Berlin, Heidelberg.
- [19] Chen, W. S., Zeng, Q., & Pan, B. (2022). A survey of deep nonnegative matrix factorization. *Neurocomputing*, 491, 305-320.
- [20] Feng, X. R., Li, H. C., Li, J., Du, Q., Plaza, A., & Emery, W. J. (2018). Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 6245-6257.