

DESCRIPTION DETAILLEE

INTITULÉ DU SUJET DE THESE

Amélioration des systèmes de suivi des cultures à l'échelle du paysage à partir de la télédétection multi-sources et des techniques d'apprentissage automatique

CONTEXTE ET PROBLÉMATIQUE (positionnement du sujet par rapport à la littérature internationale)

Avec les perspectives d'une population à 9 milliards à l'horizon 2050 et les impacts attendus du changement climatique tant sur les écosystèmes naturels que cultivés, garantir la sécurité alimentaire de tous, tout en favorisant la mise en place d'une agriculture durable, préservant les écosystèmes terrestres et la biodiversité (Objectifs 2 et 15 du Développement Durable des Nations Unies), font parties des enjeux majeurs pour le développement futur de notre société.

Dans ce contexte, les systèmes modernes d'observation terrestre offrent de nouvelles opportunités pour suivre et caractériser les systèmes agricoles à l'échelle territoriale ou continentale. Parmi l'ensemble de mission satellitaires, le programme ESA SENTINEL [Sentinel] fournit tous les 5 jours des images à haute résolution spatiale (HSR) (10 mètres), radar (SENTINEL-1) et optique (SENTINEL-2). Ce programme a été lancé en décembre 2015. L'ensemble de ces informations, organisées sous forme de séries temporelles d'images satellitaires (STIS), permet de suivre dans le temps une même zone et d'en comprendre les processus dynamiques sous-jacents. D'autre part, des images de résolution spatiale très élevée (VHSR), avec une résolution spatiale de 0,5 à 5 mètres, sont aujourd'hui disponibles par l'intermédiaire d'infrastructure d'archivage d'images afin d'étudier et d'estimer les surfaces cultivées ainsi que leur rendement. L'EQUIPEX GEOSUD [Geosud], fondé par l'agence ANR, constitue en exemple d'infrastructure d'archivage de données satellitaires. L'EQUIPEX GEOSUD met à disposition des images SPOT 6/7 et Pleiades à une résolution, respectivement, de 6 et 1,5 mètres en multispectrale. En raison des coûts d'acquisition, les images VHSR ne sont disponibles qu'une ou deux fois par an sur une zone particulière. Ceci peut limiter leur utilisation pour l'analyse de l'occupation du sol et/ou l'estimation de rendements: deux classes de végétation qui partagent des réponses radiométriques similaires à une estampille donnée peuvent avoir différentes tendances temporelles. C'est pour cela qu'exploiter explicitement les dynamiques temporelles présentes dans les données STIS optiques [Lebourgeois17] et radar pourra certainement améliorer le suivi des évolutions des espaces agricoles [Hu16, Ienco17]. En raison de la complémentarité de ces sources d'information, fusionner des données STIS optiques et radar ensemble et, en même temps, intégrer de l'information fournie par des images VHSR, pourra aider à mieux suivre et caractériser les pratiques agricoles.

Comment fusionner toutes ces différentes sources d'information (STIS optique/radar et imagerie VHSR) produites à travers les récentes technologies d'observation terrestre, reste un problème ouvert [Schmitt16] auquel la communauté de télédétection, nationale et

international, commence à être confrontée et pour lequel des solutions efficaces et stables ne sont pas encore proposées.

L'objectif de cette thèse est donc de proposer de nouvelles techniques d'apprentissage automatique [Zhang16, Guttler16] pour l'intégration des différentes sources de données (séries temporelles optiques, séries temporelles radar, imagerie à Très Haute Résolution Spatiale - VHRS) dans un but d'améliorer i) la caractérisation des surfaces cultivées et ii) l'estimation des rendements agricoles [Leroux16] à partir de données collectées sur le terrain.

Positionnement par rapport aux projets structurants régionaux, nationaux et internationaux:

Positionnement Régional. Dans le cadre régional, la thèse s'intégrera dans les objectifs de l'ISITE Muse qui vise à apporter une réponse collective aux défis liés à la science du vivant, de l'environnement et de l'agriculture. Plus particulièrement, cette thèse répondra aux objectifs de l'Institut de Convergence DigitAg (entité auprès de laquelle nous avons soumis une demande de cofinancement pour cette proposition de thèse) qui vise à proposer des nouveaux outils liés au monde du numérique afin de contribuer à un développement durable et intelligent de l'agriculture d'aujourd'hui.

Positionnement National: Depuis 2015, la création du GdR CNRS MaDICS (Masses de Données, Information et Connaissances en Sciences) met en exergue la problématique de l'infobésité et du traitement de gros volumes de données hétérogènes qui constitue une des problématique majeure de la thèse proposée. Au niveau nationale, les avancées proposées dans le contexte de cette thèse seront aussi liés aux activités du Pôle Surface Continentale THEIA et liés aux différents CES (Centre d'Expertise Scientifique) associées à THEIA (i.e. CES OSO Occupation du sols).

Positionnement International: Dans un contexte international, la problématique de l'intégration et la fusion des données satellitaires multi-sources est mise en avant dans de nombreuses conférences et journaux internationaux comme IEEE IGARSS (International Geoscience and Remote Sensing Symposium), CVPR (Computer Vision and Pattern Recognition Conference), ECML-PKDD (European Conference on Machine Learning and Data Mining), IEEE TGRS (Transaction on Geoscience and Remote Sensing), Pattern Recognition (Elsevier Pattern Recognition). De plus en plus de programmes d'observation de la terre (ESA et NASA) sont en train d'être lancés pour assurer un suivi régulier des changements et des évolutions de la planète.

Face a ces volumes d'information disponibles, le défi majeur sera de comprendre comment coupler et exploiter les différentes sources des données satellitaire afin de créer de la nouvelle connaissance.

References:

- [Battude16] : Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., ... Demarez, V. (2016). Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment*, 184, 668–681.
- [Guttler16] F. Guttler, **D. Ienco**, P. Poncelet and M. Teisseire. “Combining Transductive and Active Learning to Improve Object-based Classification of Remote Sensing Images”. *Remote Sensing Letters*, 7(4): 358-367 (2016).
- [Hu16] Q. Hu, W. Wu, Q. Song, Q. Yu, M. Lu, P. Yang, H. Tang, Y. Long: “Extending the Pairwise Separability Index for Multicrop Identification Using Time-Series MODIS Images”. *IEEE Trans. Geoscience and Remote Sensing* 54(11): 6349-6361 (2016).
- [Ienco17] **D. Ienco**, R. Gaetano, C. Dupaquier, P. Maurel: “Land Cover Classification via Multi-temporal Spatial Data by Recurrent Neural Networks”. *IEEE GRSL* 14(10): 1685 - 1689 (2017)
- [Lebourgeois17] Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., & Bégué, A. (2017). A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sensing*, 9(3), 259.
- [Leroux16] **Leroux, L.**, Baron, C., Zoungrana, B., Traoré, S. B., Lo Seen, D., & Bégué, A. (2016). Crop Monitoring Using Vegetation And Thermal Indices For Yield Estimates: Case Study Of A Rainfed Cereal In Semi-Arid West Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(1), 347–362.
- [Sentinel] <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2>.
- [Schmitt16] M. Schmitt and X. X. Zhu: “Data Fusion and Remote Sensing: An ever-growing relationship”. *IEEE Geoscience and Remote Sensing Magazine*: 4(4): 6-23 (2016).
- [Valero16]: Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., ... Koetz, B. (2016). Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions. *Remote Sensing*, 8(55), 1–21.
- [Zhang16] L. Zhang, L. Zhang and B. Du: “Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art”. *IEEE Geoscience and Remote Sensing Magazine*: 4(2): 22-40 (2016).

RESUME DU TRAVAIL PROPOSE (partenariat, questions et hypothèse de recherche, méthodologie, partenariat)

Partenariat

Le travail de thèse sera mené en partenariat avec l'UR AIDA - CIRAD (Dr. Louise Leroux) situé à Montpellier avec des antennes à Dakar (Sénégal) et avec l'entreprise SMAG (Système d'Information et Logicielle pour l'Agriculture) situé à Montpellier et membre industrielle de l'Institut de Convergence DigitAg. Les terrains d'études qui seront utilisés pour développer et tester les méthodes de fusion des données satellitaires seront localisés au Sénégal (Bambey et Nioro du Rip) et en Région Occitanie.

Pour le site d'étude au Sénégal, le partenariat avec l'UR AIDA nous permettra d'accéder aux données de terrain collectées dans le cadre des programmes de recherche déjà menés par l'UR AIDA et nous permettra d'interagir avec des experts agronomes. Ces derniers pourront nous accompagner sur les phases d'organisation des données et dans l'interprétation/validation des résultats extraits par les techniques d'apprentissage automatique.

Pour le site d'étude en France (situé dans la région Occitanie), nous collaborerons avec le secteur R&D de l'entreprise SMAG. Cette collaboration nous permettra d'accéder à des données terrain collectés directement par les agriculteurs à travers des systèmes de collecte de données déjà mis en place de manière opérationnelle. Par ailleurs, la possibilité de pouvoir démarrer une collaboration avec l'entreprise SMAG, nous permettra tout d'abord de réaliser des premiers travaux ensemble et, par la suite, d'envisager de monter des projets de partenariat entreprise-recherche pour répondre à des appels d'offre Régionaux/National de façon conjointe sur des thématiques liées à la gestion du territoire et des espaces agricole.

Sites d'étude

Le premier site d'étude se situe dans le quart Nord-Est de la France et correspond à une agriculture dominée par les grandes cultures, notamment des cultures céréalières (blé tendre d'hiver, orge d'hiver et de printemps), des cultures oléagineux comme le colza ou encore des cultures industrielles comme la betterave sucrière. Le type d'agriculture pratiqué est majoritairement conventionnel, à destination de l'alimentation humaine mais également de l'élevage. Cette zone agricole est largement couverte par les solutions SMAG et permet d'avoir un nombre important de données historiques sur les sols et sur les pratiques agricoles notamment.

Le second terrain d'étude est situé dans le bassin arachidier sénégalais, dans la région de Bambey, dominé par de la petite agriculture familiale (petit parcellaire, hétérogénéité des pratiques et des couverts, présence d'arbres dans les parcelles). La production agricole y est dominée par les rotations céréales-légumineuses, notamment mil-arachide. Pour le site sénégalais, la thèse bénéficiera de l'appuis du projet SERENA (De la biodiversité des paysages agricoles à la sécurité alimentaire des ménages ruraux) porté par le CIRAD (financement GloFoodS) pour les enquêtes de terrain.

Questions de recherche (Problématique et application associée):

Cette proposition de thèse présente l'originalité d'aborder la question au travers de deux sites d'étude: l'un situé en France et caractérisé par une agriculture conventionnelle et l'autre situé au Sénégal et caractérisé par une petite agriculture familiale traditionnelle.

Les questions de recherche liées à cette thèse sont les suivantes:

- 1) Quelles potentialités offrent, dans un contexte de suivi agricole et agriculture digital, les systèmes modernes d'observation de la terre? Comment des techniques d'apprentissage automatique pourront nous aider pour intégrer différentes sources de données?
- 2) Est-ce que le couplage de plusieurs sources des données (séries temporelles optique/radar, imagerie à Très Haute Résolution) pourra améliorer le suivi et la gestion des systèmes de culture au niveau d'un territoire? Quelle source de données sera la plus/moins adaptée pour ce but?
- 3) A quel point est-il possible de réutiliser des techniques d'estimation de surfaces cultivées et de rendement développées sur un site spécifique pour, ensuite, analyser une autre zone avec un comportement différent (France / Sénégal)? Est-ce que les sources de données exploitées (optique/radar/VHRS) ont le même poids selon le site étudié ?

Méthodologie et Démarche:

Dans le cadre de cette thèse nous voulons attaquer deux problèmes liés: i) la caractérisation des surfaces cultivées [Ienco17,Valero16] et ii) l'estimation des rendements agricoles [Leroux16] à partir de données collectées sur le terrain.

Pour ce qui concerne l'estimation des surfaces cultivées, l'objectif sera de pouvoir caractériser et distinguer les différents types d'occupation du sol (avec un focus particulier sur les cultures agricoles) pour ensuite estimer leur surfaces. Pour ceci, des techniques d'apprentissage automatique supervisée seront mises en place. Ce point permettra de produire, de façon automatique, des masques de culture [Valero16] qui pour la suite seront utilisés pour isoler les parties agricoles pour estimer leur rendements.

Pour ce qui concerne l'estimation des rendements (spatialisée), deux approches (complémentaires entre-elles) sont généralement mises en place: la première repose sur le couplage entre des données de télédétection et des modèles de croissance de plantes [Leroux16] et la deuxième repose sur des relations statistiques empiriques entre les indices spectraux et des données mesurées sur le terrain [Battude16]. Dans le cadre de cette thèse nous allons nous pencher sur la deuxième famille d'approches en s'appuyant sur des techniques d'apprentissage automatique. Ces approches sont traditionnellement utilisées dans des analyses par convergence de preuve permettant d'exploiter la nature multi-sources des données. Les nouveaux indicateurs produits par les méthodes

d'apprentissage automatique pourront ainsi être assimilés par des approches de la première famille (modèles de culture et modèle de croissance des plantes) pour améliorer leur performances (i.e. indicateur de stade phénologique, indicateur de niveau d'intensification des pratiques agricoles).

Les sous-objectifs de cette proposition de thèse sont les suivants :

(1) Récupérer et organiser les séries temporelles multi-sources à haute résolution spatiale et les données à Très Haute Résolution Spatiale (accessible à travers l'Equipex GEOSUD) pour construire deux bases des données, une par site selon les spécificité de la zone d'étude (date de début et de fin du cycle culturelle). Pour ce point, l'étudiant(e) pourra s'appuyer sur les spécialistes de chaque site d'études et sur les expertises des partenaires de la proposition de thèse.

(2) Etudier comment fusionner, à travers des techniques d'apprentissage automatique, des séries temporelles multi-sources (optiques et radar) à haute résolution spatiale avec des données à Très Haute Résolution Spatiale (accessible à travers l'Equipex GEOSUD).

Dans cette phase, l'étudiant devra s'appuyer sur une étude bibliographique approfondi des techniques actuellement utilisés en télédétection pour pouvoir proposer des méthodes et des approches innovantes de fusion de données. Tout d'abord, l'étudiant(e) sera confronté(e) à la problématique de l'intégration des séries temporelles multi-sources (optique et radar) pour comprendre comment utiliser au mieux les deux sources des données. Dès que ce point sera résolu, la problématique de l'intégration de l'information de VHRS sera adressée. La méthodologie proposée sera implémentée pour les deux sites d'études, le plus possible en parallèle.

(3) Etudier la complémentarité des différentes sources d'information (temporelles, multi-sources, multi-résolution), avec un regard croisé entre les deux sites d'étude. Dans ce sous-objectif, une analyse plus poussée pour comprendre l'interaction entre les différentes sources des données et analyser l'apport de chaque capteurs (Sentinel-2, Sentinel-1, Pleiades et/ou SPOT6/7) sera proposée. Cette étude mettra aussi au centre une analyse comparative ainsi qu'un regard croisé entre les deux sites d'étude (site Nord et site Sud) pour une meilleur compréhension globale des processus d'analyse en milieu agricole.