



Laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes
MICS

Présente

L'AVIS DE SOUTENANCE

De **Monsieur Walid HAMMACHE**

Laboratoire MICS, CentraleSupélec, Université Paris Saclay, soutiendra publiquement
ses travaux de thèse de doctorat intitulés :

**"Machine learning methods for automatic crop classification and prediction of yields
on a large scale"**

Le 09 Décembre 2022 à 10h

À l'école CentraleSupélec, en Amphi **e.068**- Bâtiment Bouygues en présentiel et en
distanciel (à suivre ici https://teams.microsoft.com/l/meetup-join/19%3ameeting_ZjJmM2E3Y2YtZDE1OC00NjY2LThkNjUtNTIIMGM5NjU5YTk1%40thread.v2/0?context=%7b%22Tid%22%3a%2261f3e3b8-9b52-433a-a4eb-c67334ce54d5%22%2c%22Oid%22%3a%22bc3f8bb3-45da-471b-b701-dd32a87b9e03%22%7d).

Membres du jury :

Céline Hudelot, Professeur, CentraleSupélec, Université Paris-Saclay, (Présidente du jury)

Adel Hafiane, Professeur, INSA Centre Val de Loire, (Rapporteur)

Xiaopeng Zhang, Professeur, Institute of Automation, Chinese Academy of Sciences, Chine
(Rapporteur)

Maria Vakalopoulou, Maîtresse de conférence, CentraleSupélec, Université Paris-Saclay (Invitée)

Direction de thèse :

Paul-Henry Cournède, Professeur, CentraleSupélec, Université Paris-Saclay (Directeur de thèse)
Théophile Lohier, Docteur, Entreprise CybeleTech (Co-encadrante)

Abstract :

The main thematic studied in this work concerns the identification of crop types using radar satellite imagery by using statistical learning models. Crop type identification allows the automatic monitoring of land cover and the estimation of planted areas as well as the improvement of the prediction of agricultural yields, this is the other topic that will be addressed in this work. For these works, we could count on a major ally, the satellite imagery which provides multi-temporal images with high resolution offering a real opportunity to track the evolution of the reflective properties of plants during their growth, according to the variations related to the phenology and the pedoclimate. We use radar remote sensing imagery ensuring independence from climatic conditions, more precisely, images from the Sentinel-1 satellite. Our approach for crop recognition is based on the use of plant growth dynamics that can be captured by satellite imagery to discriminate crops. It is a pixel-based approach for which we have chosen to use two supervised statistical learning models, LSTM recurrent and CNN convolutional neural networks. Our approach is designed so that the model can produce a vegetation coverage assessment without having to label parcels at the beginning of the season, which is a very complex operation to perform. We, therefore, sought to calibrate the models using data from previous seasons. We propose a method in which the phenological development of the plant is considered to identify its species using the concept of thermal time. This allows compensating the inter-annual variability, whether it is pedo-climatic, related to the characteristics of the crop such as the resistance to frost episodes, or due to the cultivation practices. Crop type identification can be useful for several applications, such as improving yield forecasting through a better knowledge of the planted areas. This is another research theme of our thesis. Crop recognition can be an essential element in the process of obtaining the most accurate crop yield forecasts possible. By identifying the type of crop in the field, we know the phenotypic characteristics of that crop to be able to estimate the biophysical parameters, which, coupled with soil-climate data and cultivation practices, allows us to move towards accurate crop yield forecasting using plant growth models, calibration and data assimilation methods. This methodology for yield forecasting at the scale of the parcel or the region is also studied in this thesis. The anticipated knowledge of agricultural yields is a major challenge in agriculture: in a production basin, for the silo manager to predict the logistics of cultivation and storage of grains or on a larger scale, to anticipate agricultural crises. In this context, we started by evaluating the potential of radar imagery for the development of statistical models allowing the estimation of plant biophysical variables such as biomass. We also proposed a tool allowing to recommend the positioning of measurement points in agricultural parcels according to intra-parcel heterogeneities and using an unsupervised K-means clustering of the radar response dynamics of each pixel. Next, we present an implementation of data assimilation methods to show how they can be used to improve the predictive powers of plant growth models. We use a data assimilation algorithm based on the particle convolution filter applied to a mechanistic wheat growth model derived from the STICS model.

Résumé :

La principale thématique étudiée dans ce travail concerne l'identification des types de cultures par imagerie satellite radar avec des modèles d'apprentissage statistique. L'identification des cultures

permet le suivi automatique de la couverture du sol et l'estimation des surfaces plantées ainsi que l'amélioration de la prévision des rendements agricoles, c'est l'autre sujet traité dans ce travail. Pour ces travaux, nous avons pu compter sur un allié majeur, l'imagerie satellitaire qui fournit des images multi-temporelles à haute résolution offrant une réelle opportunité de suivre l'évolution des propriétés réfléchissantes des plantes durant leur croissance, en fonction des variations liées à la phénologie et au pédoclimat. Nous utilisons l'imagerie radar Sentinel-1 assurant une indépendance vis-à-vis des conditions climatiques. Notre approche pour la reconnaissance des cultures est basée sur l'utilisation de la dynamique de croissance des plantes pouvant être capturée par l'imagerie satellite pour discriminer les cultures. Il s'agit d'une approche par pixels pour laquelle nous avons choisi d'utiliser deux modèles supervisés d'apprentissage statistique, les réseaux de neurones récurrents LSTM et les réseaux de convolution CNN. Notre approche est conçue de telle sorte que le modèle puisse produire une évaluation du couvert végétal sans avoir à labelliser des parcelles en début de campagne, opération complexe à réaliser. Nous avons donc cherché à calibrer les modèles en utilisant les données des campagnes précédentes. Nous proposons une méthode dans laquelle le développement phénologique de la plante est considéré pour l'identifier en utilisant le concept du temps thermique. Cela permet de compenser la variabilité inter-annuelle, qu'elle soit pédo-climatique, liée aux caractéristiques de la culture comme la résistance aux épisodes de gel, ou due aux pratiques culturales. L'identification des cultures peut servir l'amélioration des prévisions de rendement résultant d'une meilleure connaissance des surfaces plantées. C'est une autre thématique de recherche de notre thèse. La reconnaissance des cultures peut constituer un élément essentiel dans le processus visant à obtenir des prévisions de rendements les plus précis possible. En identifiant le type de culture dans le champ, nous connaissons les caractéristiques phénotypiques de cette culture pour pouvoir estimer les paramètres biophysiques, ce qui, couplé aux données pédoclimatiques et aux pratiques culturales, nous permet d'aller vers une prévision précise des rendements agricoles en utilisant les modèles de croissance des plantes, les méthodes de calibration et d'assimilation de données. Cette méthodologie pour la prévision des rendements à l'échelle de la parcelle ou de la région est également étudiée dans cette thèse. En effet, la connaissance anticipée des rendements agricoles est un enjeu majeur en agriculture: dans un bassin de production, pour le chef de silo afin de prévoir la logistique de culture et de stockage des grains ou à plus grande échelle, pour anticiper les crises agricoles. Dans ce contexte, nous avons commencé par évaluer le potentiel de l'imagerie radar pour le développement de modèles statistiques permettant l'estimation des variables biophysiques des plantes telles que la biomasse. Nous avons également proposé un outil permettant de conseiller le positionnement des points de mesure dans les parcelles agricole en fonction des hétérogénéités intra-parcellaires et en utilisant un clustering non supervisé de type K-means de la dynamique de la réponse radar de chaque pixel. Ensuite, nous présentons une mise en œuvre des méthodes d'assimilation de données pour montrer comment elles peuvent être utilisées pour améliorer les pouvoirs prédictifs des modèles de croissance des plantes. Nous utilisons un algorithme d'assimilation de données basé sur le filtre particulaire par convolution appliqué à un modèle mécaniste de croissance du blé dérivé du modèle STICS.