Mapping Irrigated Areas using Radar and Optical Remote Sensing Data

Hassan Bazzi

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT - AGROPARISTECH

N° : 2021AGPT0013

En Géomatique

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau – n°584
Portée par l'Université de Montpellier

Unité de recherche TETIS-Territoires, Environnement, Télédétection et Information Spatiale

Mapping Irrigated Areas using Radar and Optical Remote Sensing Data

Présentée par Hassan BAZZI
Le 15 Octobre 2021

Sous la direction de Nicolas BAGHDADI, Hatem BELHOUCHETTE et Mehrez ZRIBI

Devant le jury composé de

Nicolas BAGHDADI, Directeur de recherche, INRAE Montpellier	Directeur de thèse
Hatem BELHOUCHETTE, Enseignant-chercheur, CIHEAM-IAMM	Co-encadrant
Dominique COURAULT, Directrice de recherche, INRAE Avignon	Rapportrice
Zohra LILI CHABAANE, Professeure, Université de Carthage, Tunisie	Rapportrice
Clément ALBERGEL, Chargé de recherche CNRS, ESA	Examinateur
Gilles BELAUD, Professeur, Institut Agro - Montpellier SupAgro	Président
Mehrez ZRIBI, Directeur de recherche, CNRS	Co-directeur & Invité

AgroParisTech

UNIVERSITÉ DE MONTPELLIER

ACKNOWLEDGEMENTS

First of all, I would like to deeply thank M. Nicolas Baghdadi, my thesis director, who has given me the full trust since the very first days. Thank you M. Baghdadi for having accompanied me since my Engineering study there in Lebanon, and for giving me, the opportunity to come to Montpellier for my internship followed by my master's study here in Montpellier, which subsequently allowed me to carry out this thesis. Thank you for the continuous support and faith you invested in me. Your unstoppable availability, whenever needed, was indeed the foundation stone of my work. I would like to thank my thesis supervisors, M. Hatem Belhouchette and M. Mehrez Zribi for their continuous support, help and encouragement that were dedicated to me throughout the past three years. Your wise advices have indeed assisted me in conducting this work. I would like to thank the Occitanie region, the Mediterranean Agronomic Institute of Montpellier (IAMM), and the National Research Institute for Agriculture Food and Environment for financially supporting this work. To my friend, my office partner and my colleague M. Ibrahim Fayad, a sincere thanks for all the support, help, encouragement and assistance. You are a confirmed copy of my personality, motivation, intelligence and hard work. Thanks to every person, I have met at the UMR-TETIS; Ghaith, Reem, Nuria, and lastly a true new friend Mohamad Hamze. My special thanks go to M. Mohammad El Hajj who was always there for assisting me in the very beginning stages.

To my father and mother. I owe you everything, and if I am where I am today, it is thanks to you, to your presence and your sacrifices, to your love. Thank you for all the encouragement, for all the support and trust. My three sisters, Alaa, Farah, and Zeina I give my deepest expression of appreciation for the love, care and support you gave me in this long journey. Distance was never a barrier to feel your support.

My second family in Montpellier. I will never forget good times and adventures we had together. For Jacinthe *Boulos Saeed Yaacob Mekary*, Jona Karam *Adira*, Amani Wehbe *Jarte*, Ibrahim El-Moussawi *Akhowe*, Ali Nasrallah *Hajj* Baraa Nachar *Jelo*, Fatima Haidar and Adham Safeiddine *Jare* special thanks for the unforgettable moments and memories we shared together. Finally, the list of friends ends with precious people to be thanked. Ayman El Jundi, Rita Tanos, Mohammad Hamoud, Sana Kassem, Ziad Kantar, Diala Kantar and Imad Safar a big thanks for all the support you gave me, for the jokes, nights, sleepovers and endless laugh we have together. You were my way for a joyful life to escape from my stressful PhD moments.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER II

ARTICLE 1

CHAPTER II

ARTICLE 2

CHAPTER III

ARTICLE 3

LIST OF TABLES

CHAPTER IV

ARTICLE 6

CHAPTER IV

ARTICLE 7

ABSTRACT

The climate change and the increase in the global population are putting more pressure on natural resources in order to satisfy the global food requirement. To fulfill the increasing food demand, the global agricultural intensification is mainly based on increasing the resource use especially water. Irrigation is the primary source of freshwater consumption. Knowing the extent of irrigated areas and the irrigation frequency could help decision makers adapt to future water policies in the agricultural sector. The aim of this research work is to develop approaches capable of mapping the spatial extent of irrigated areas and detecting the irrigation episodes at plot scale using optical and radar remote sensing.

The first part of this thesis concentrates on the evaluation of the soil moisture estimation at plot scale S^2MP (Sentinel-1/2 Soil Moisture Product) recently developed at INRAE-TETIS. The S^2MP was evaluated by comparing it to precipitation records, in situ soil moisture measurements and the newly derived Copernicus surface soil moisture product on a large region of south France (Occitanie region). The results showed good accuracy of the S^2MP at plot scale (RMSE = 4.0 vol.%). Furthermore, the S^2MP showed better accuracy for soil moisture estimation than the Copernicus product. In addition, the results showed high consistency between the S^2MP soil moisture estimates at grid scale of 10 km x 10 km and the precipitation records. The S^2MP product at grid scale (10 km x 10 km) was used next in mapping irrigated areas in order to distinguish between irrigation events and precipitation events.

The second part of the thesis presents supervised classification approaches for irrigation mapping at plot scale. The classification approaches are based on the use of radar Sentinel-1 (S1) and optical Sentinel-2 (S2) temporal series. Over a semi-arid region in Catalonia (Spain), the results showed that irrigated areas are accurately mapped using S1 and S2 data with a classification accuracy between 89% and 94% depending on the used machine-learning model. Then, I proposed a transfer-learning framework capable of transferring the model built on Catalonia to map irrigated areas in Adour-Amont watershed of southwest France (humid zone).

The final part of the thesis shows the potential of Sentinel-1, Sentinel-2 and S^2MP for detecting irrigation events at plot scale. It presents also an operational methodology for mapping irrigated areas. In this context, I developed first a near-real time irrigation event detection model (IEDM) at plot scale, capable of identifying the existing irrigation events at the agricultural plots. The validation of the IEDM with in situ data showed that irrigation events are detected

with an accuracy reaching 75%. Finally, I proposed an operational methodology for mapping irrigated areas at plot scale without using terrain data since the approach collects automatically the training data for classification without the help of in situ data. The operational method was tested over a study site located in Orléans (northcentral France) for four different years. The main results showed that the proposed operational framework provides very good accuracy for irrigation mapping (between 72% and 94%).

Keywords: *Irrigation extent, Irrigation events, Sentinel-1, Sentinel-2, Soil Moisture, S^2MP, Supervised Classification, Unsupervised Classification*

RÉSUMÉ

Cartographie des zones irriguées à l’aide de données de télédétection radar et optique

Le changement climatique et l'augmentation de la population mondiale exercent une pression sur les ressources naturelles afin de satisfaire les besoins alimentaires mondiaux. Pour répondre à la demande alimentaire croissante, l'intensification de l'agriculture est principalement basée sur l'augmentation de l'utilisation des ressources, en particulier de l'eau. L'irrigation est la principale source de consommation d'eau douce. Connaitre l'étendue des zones irriguées et la fréquence d'irrigation permettra d’aider les principaux acteurs à adapter les futures politiques concernant la gestion des ressources en eau. L'objectif de ce travail de recherche est de développer des approches capables de cartographier l'étendue spatiale des zones irriguées et de détecter les épisodes d'irrigation à l'échelle de la parcelle en utilisant la télédétection optique et radar.

La première partie de cette thèse se concentre sur l'évaluation du produit d'estimation de l'humidité du sol S^2MP (Sentinel-1/2 Soil Moisture Product) récemment développé à l'INRAE-TETIS. Le produit S^2MP a été évalué en comparant l’humidité estimée aux enregistrements de précipitations, aux mesures in situ de l'humidité du sol et au nouveau produit de l’estimation d’humidité du sol fournit par Copernicus sur une large région du sud de la France (région Occitanie). Les résultats ont montré une bonne précision du produit S^2MP à l'échelle de la parcelle (RMSE = 4.0 vol.%). De plus, le S^2MP montrait une meilleure précision que le produit Copernicus. En outre, les résultats ont montré une grande cohérence entre les estimations de l'humidité du sol du S^2MP à l'échelle d’une grille de 10 km x 10 km et les précipitations enregistrées. Le produit S^2MP sur des grilles de 10 km x 10 km a été utilisé par la suite dans la cartographie de l'irrigation pour distinguer les événements d'irrigation des événements de précipitation.

La deuxième partie de la thèse présente des approches de classification supervisée pour la cartographie de l'irrigation à l'échelle de la parcelle. Les approches de classification sont basées sur l'utilisation des séries temporelles d’images radar Sentinel-1 (S1) et optique Sentinel-2 (S2). Sur une région semi-aride en Catalogne (Espagne), les résultats montrent que les zones irriguées sont cartographiées en utilisant les données S1 et S2 avec une précision de classification entre

89% et 94% selon le modèle d'apprentissage automatique utilisé. Ensuite, j'ai proposé un cadre d'apprentissage profond capable d'adapter le modèle d'irrigation construit sur le site de la Catalogne pour cartographier l'irrigation sur le bassin Adour-Amont de l'Occitanie (zone humide).

La dernière partie de la thèse montre le potentiel de Sentinel-1, Sentinel-2 et S^2MP pour la détection des événements d'irrigation à l'échelle de la parcelle. Elle présente également une méthodologie opérationnelle pour cartographier des zones irriguées. Dans ce contexte, j'ai développé en premier temps un modèle de détection d'événements d'irrigation (IEDM) en temps quasi-réel à l'échelle de la parcelle, capable d'identifier les événements d'irrigation existants sur les parcelles agricoles. La validation de l'IEDM avec des données in situ a montré que les événements d'irrigation sont détectés avec une précision de 75%. Enfin, j'ai proposé une méthodologie opérationnelle pour cartographier les zones irriguées à l'échelle de la parcelle sans l'utilisation de données de terrain car l'approche est basée sur une collecte automatique des données d'entrainement sans l'aide de données in situ. Elle était testée sur un site d'étude situé à Orléans (centre-nord de la France) pendant quatre années différentes. Les principaux résultats montrent que le cadre opérationnel proposé offre une très bonne précision pour la cartographie de l'irrigation (entre 72% et 94%).

Mots-clés : *Étendue de l'irrigation, Événements d'irrigation, Sentinel-1, Sentinel-2, humidité du sol, S^2MP, classification supervisée, classification non supervisée.*

LIST OF PUBLICATIONS

Published Journal Articles

1. **Bazzi, H.;** Baghdadi, N.; Amin, G.; Fayad, I.; Zribi, M.; Demarez, V.; Belhouchette, H. An Operational Framework for Mapping Irrigated Areas at Plot Scale Using Sentinel-1 and Sentinel-2 Data. *Remote Sens.* 2021, *13*, 2584. https://doi.org/10.3390/rs13132584

2. **Bazzi, H**.; Baghdadi, N.; Fayad, I.; Charron, F.; Zribi, M.; Belhouchette, H. Irrigation Events Detection over Intensively Irrigated Grassland Plots Using Sentinel-1 Data. Remote Sensing 2020, 12, 4058, doi:10.3390/rs12244058.

3. **Bazzi, H.**; Baghdadi, N.; Fayad, I.; Zribi, M.; Belhouchette, H.; Demarez, V. Near Real-Time Irrigation Detection at Plot Scale Using Sentinel-1 Data. Remote Sensing 2020, 12, 1456, doi:10.3390/rs12091456.

4. **Bazzi, H.**; Ienco, D.; Baghdadi, N.; Zribi, M.; Demarez, V. Distilling Before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas Using Sentinel-1 Time Series. IEEE Geosci. Remote Sensing Lett. 2020, 1–5, doi:10.1109/LGRS.2019.2960625.

5. **Bazzi, H.;** Baghdadi, N.; Ienco, D.; El Hajj, M.; Zribi, M.; Belhouchette, H.; Escorihuela, M.J.; Demarez, V. Mapping Irrigated Areas Using Sentinel-1 Time Series in Catalonia, Spain. *Remote Sensing* **2019**, *11*, 1836, doi:10.3390/rs11151836.

6. **Bazzi, H.;** Baghdadi, N.; El Hajj, M.; Zribi, M.; Belhouchette, H. A Comparison of Two Soil Moisture Products S ^{2} MP and Copernicus-SSM Over Southern France. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* **2019**, *12*, 3366–3375, doi:10.1109/JSTARS.2019.2927430.

7. **Bazzi, H.;** Baghdadi, N.; El Hajj, M.; Zribi, M. Potential of Sentinel-1 Surface Soil Moisture Product for Detecting Heavy Rainfall in the South of France. *Sensors* **2019**, *19*, 802, doi:10.3390/s19040802.

8. Hamze, M.; Baghdadi, N.; El Hajj, M.M.; Zribi, M.; **Bazzi, H.;** Cheviron, B.; Faour, G. Integration of L-Band Derived Soil Roughness into a Bare Soil Moisture Retrieval Approach from C-Band SAR Data. Remote Sens. 2021, 13, 2102. https://doi.org/10.3390/rs13112102

9. Fayad, I.; Baghdadi, N.; **Bazzi, H.;** Zribi, M. Near Real-Time Freeze Detection over Agricultural Plots Using Sentinel-1 Data. Remote Sensing 2020, 12, 1976, doi:10.3390/rs12121976.

10. El Hajj, M.; Baghdadi, N.; **Bazzi, H.;** Zribi, M. Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. *Remote Sensing* **2018**, *11*, 31, doi:10.3390/rs11010031.

11. Baghdadi, N.; Bazzi, H.; El Hajj, M.; Zribi, M. Detection of Frozen Soil Using Sentinel-1 SAR Data. Remote Sensing 2018, 10, 1182, doi:10.3390/rs10081182.

12. Baghdadi, N.; El Hajj, M.; Choker, M.; Zribi, M.; Bazzi, H.; Vaudour, E.; Gilliot, J.-M.; Ebengo, D. Potential of Sentinel-1 Images for Estimating the Soil Roughness over Bare Agricultural Soils. Water 2018, 10, 131, doi:10.3390/w10020131.

13. El Hajj, M.; Baghdadi, N.; Zribi, M.; **Bazzi, H.** Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. Remote Sensing 2017, 9, 1292, doi:10.3390/rs9121292

Oral Presentations

14. **Bazzi, H**., Belhouchette, H. & Baghdadi, N. (2021) The utopia of increasing crop diversity and intensification by increasing access to irrigation water in dry conditions. In International Symposium for Farming Systems Design – March 2021.

15. **Bazzi, H**., Baghdadi, N., Ienco, D., Zribi, M. & Belhouchette, H. (2020) Irrigation mapping using sentinel-1 time series. In IGARSS 2020 IEEE International Geoscience and Remote Sensing Symposium

16. **Bazzi, H**., Baghdadi, N., Ienco, D., Zribi, M. & Belhouchette, H. (2020). Potential of Sentinel-1 and Sentinel-2 data for mapping irrigated areas at plot scale. In EGU General Assembly Conference Abstracts (p. 10367)

17. Baghdadi, N., **Bazzi, H**., Hajj, M. & Zribi, M. (2020). Comparison between Theïa and Copernicus surface soil moisture products over southern France. In 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS) (p. 168-171). IEEE

18. Lozac'h, L., **Bazzi, H**., Baghdadi, N., El Hajj, M., Zribi, M. & Cresson, R. (2020). Sentinel-1/Sentinel-2-Derived Soil Moisture Product At Plot Scale (S 2 MP). In 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS) (p. 168-171). IEEE

19. Baghdadi, N., **Bazzi, H**., El Hajj, M. & Zribi, M. (2019). Analysis of Sentinel-1 derived soil moisture maps over Occitanie, South France. In IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium (p. 6275-6278). IEEE.

CHAPTER I: INTRODUCTION

1. General Context

In the 2030 Agenda of Sustainable Development of the United Nations (UN), several challenges have been recognized concerning the growing human population (Hambrey, 2017). With an estimated population of 9 billion in 2050, higher food demands are expected along with changing consumption patterns and agricultural cultivation regimes (Tilman and Clark, 2015; Tilman et al., 2011). Moreover, the human and natural systems are already stressed by the climate change, which is threatening both natural resources and food production (Schaldach et al., 2012; White et al., 2011).

As more risk is present on food security, several actions around food security and agriculture are to be considered to avoid unsatisfactory destinations in the future. In 2011 and under the French presidency of the Group of 20 (G20), the G20's Agricultural Ministers requested a full proposal for agricultural monitoring as a part of their actions to reach sustainable agriculture (Whitcraft et al., 2019). Moreover, reaching the sustainability in the agricultural sector and protecting food security through agricultural monitoring are part of the Sustainable Development Goals (SDGs) defined by the UN for the year 2030 (UN Water, 2018). In 2015, 17 SDGs have been fixed by the UN including hundreds of targets spanning on multiple science domains including the agricultural sector (Hambrey, 2017). At national, regional and global scales, the success of the policy decisions related to the SDGs, especially in the agricultural sector, requires moving to a quantitative domain spatially and temporally (Griggs et al., 2013). In this context, several international communities such as GEOGLAM (Group on Earth Observation Global Agricultural Monitoring) are working on enhancing the utilization of Earth Observations to produce and distribute accurate and reliable information on food production (Whitcraft et al., 2015, 2019). With the advanced space technologies and remote sensing techniques, monitoring of the earth surface including the agricultural areas has been made available. In the mission of the GEOGLAM for agricultural monitoring, they first defined a set of Essential Agricultural Variables (EAVs) and fundamental products considered mandatory for agricultural monitoring. Among the fundamental products for agricultural monitoring the GEOGLAM mentions the vital need of cropland mask, crop type maps and irrigated vs non-irrigated cropland maps. While the cropland mask and the crop type map can help decision makers develop land use policies, the irrigated vs non-irrigated cropland information is essential

for the water resource management and monitoring. Indeed, water resource management is currently the main concern of several international and national groups aiming to save and protect natural water resources (UN Water, 2018). For example, the G20 held in Hamburg-Germany in 2017 clearly declared that in order to achieve food security, the agricultural production should increase in a sustainable and resilient manner with the aim to protect, manage and efficiently use water resources (FAO, 2017).

Currently, global agricultural intensification, required to fulfill the increasing food demand, is mainly based on ever-increasing resource use especially water (García-Tejero et al., 2011; Jägermeyr et al., 2017; Tilman et al., 2011). However, the over exploitation of the freshwater resources for irrigation is causing a rapid degrading of water resources in several regions especially arid and semi-arid areas (Vörösmarty et al., 2010). At both local and global scales, irrigation is the human activity introducing the largest imbalances in the natural water cycle. Irrigation accounts to more than 70% of the water withdrawals, which is considered thus the main consumer of water resources (Ozdogan et al., 2010; Siebert and Döll, 2010). This percentage reaches 95% in some developing countries. Today's human water withdrawals reaches 2,409 km^3 for irrigation and harm many river stretches around the world (Jägermeyr et al., 2017). Due to the crop intensification based on the extensive use of water resources, cropland only increased by 12% whereas, the agricultural food production has increased by more than 100% in the last 30 years (Hunter et al., 2017). Even though only 17% of the cropland are irrigated, the irrigated agriculture accounts to more than 40% of the global food production. According to the Food and Agricultural Organization of the United Nations (FAO), 60% more food are needed by 2050 to meet the food requirements of a the global population. Indeed, the irrigated food production is estimated to increase by more than 50% by 2050 as estimated by the FAO (FAO, 2017, 2018). To boost the agricultural production with the expected global population increase, irrigation rates are expected to increase in the coming decades. In fact, current regional and national policies are pushing toward the increase of the water available for irrigation for farmers in order to achieve higher food production. Meanwhile, FAO estimates that freshwater resources could be sufficient to meet the future requirements by 2050 if and only if appropriate technologies and investments are considered (FAO, 2017). However, significant inconsistency in the water availability is expected between countries and extensive water scarcity will persist in arid and semi-arid areas (Jägermeyr et al., 2017). Besides the challenging scenario described for water use in the agricultural sector, climate change is found also to add a significant impact over the water cycle. With the changing climatic conditions,

altering rainfall patterns are witnessed affecting thus the availability and quality of water resources, agricultural production and associated ecosystems (Ferguson et al., 2018; Vörösmarty et al., 2000).

The continuing expansion of irrigated areas in an unsustainable manner has necessitated the implementation of different management strategies to mobilize water resources with the aim to reach the sustainable water management and water saving (García-Tejero et al., 2011; Griggs et al., 2013). These strategies have been primarily based on actions to reduce water losses through the establishment of facilities and conservation structures. Referring to the EAVs defined by the GEOGLAM (Whitcraft et al., 2019), the most important step towards better water management and water saving is the spatial quantification of irrigated areas at large scale using earth observation techniques. Delineating the extent of irrigation is crucial for several aspects including, and not limited to, the analysis of the impact of climate change (Vörösmarty et al., 2000), the water management affecting the global food security (Tilman et al., 2011) and the land-atmosphere water exchange modeling (Boucher et al., 2004). Concerning the aspects of food security, water resources and agricultural intensification, accurate information about irrigated areas could help assess the food production, estimate crop yields, evaluate water consumption and accurately measure the water discharge from rivers and ground aquifers.

Unfortunately, the extent of irrigated areas at global and national scales is still uncertain and mostly rely on national statistical data derived from farmer's declaration, which could be not reliable. In the last decade, the FAO developed a map for areas equipped for irrigation called the Global Map of Irrigated Area (GMIA). GMIA is primarily based on combining sub-national irrigation statistics with geospatial information to compute the fraction of five arc minute cells (5' x 5') that are equipped for irrigation hence called the irrigation density (Siebert et al., 2005, 2015). However, the irrigation infrastructure (irrigation equipment) provided by GMIA does not necessarily reflect actual irrigation activities. The same limitation is observed with the MIRCA2000 product also available at five arc-minute spatial resolutions (Portmann et al., 2010). These global datasets could be suitable for large-scale studies (Pokhrel et al., 2016; Wada et al., 2014) but are generally inadequate for local-scale studies. At local scales, detailed information with relatively fine spatial resolution are required. In addition, given that irrigation may change between years, such global datasets may not accurately represent the real irrigation.

In these presented contexts, my thesis will therefore be structured around one main axes, which is analyzing the potential and capability of remote sensing data to map and monitor the

irrigation activity at high spatial resolution. The arrival of Sentinel-1 radar satellite and Sentinel-2 optical satellite, offering both high spatial and temporal resolutions, opened the way toward precise agricultural monitoring including irrigation activities at high spatial resolution. The output including the accurate detection of irrigated areas and the irrigation events at plot scale will reply to the local, national and international end-users who are urgently demanding accurate quantification of irrigation spatial extent and the irrigation frequency for better resource water management.

2. State of Art

2.1 Remote sensing data for monitoring soil water conditions

Soil moisture is a key variable in the water cycle parameter since it varies as a function of precipitation, irrigation and soil evaporation rates. When the aim is to detect the irrigation signal, surface soil moisture (SSM) is a significant indicator of the existence or absence of irrigation. Since irrigation principally increases the surface soil moisture, the latter can play an important role in irrigation detection.

The monitoring of the water status of the soil using in situ sensors helps optimizing the management of water resources to obtain an optimal yield. However, such techniques are very expensive and limited to local scales. Thus, it is important to move towards global scale to monitor soil water status (soil moisture) which can later allow wide monitoring irrigated areas. Now, satellite imageries are capable of providing spatial information on soil water status and vegetation parameters in near-real time and mostly in a free of charge cost. Among spatial imageries, the synthetic aperture radar (SAR) data have been extensively used to estimate and map the soil physical parameters particularly the SSM and the surface roughness (Aubert et al., 2013; Baghdadi et al., 2011a, 2018a; El Hajj et al., 2016a)

Through literature, soil moisture estimation based on SAR data has been realized using either physical models such as the integral equation model (IEM) (Fung et al., 1992) or empirical models such as the Dubois and Oh model (Dubois et al., 1995). Contrary to physical models, empirical models require in situ terrain data for calibration. In the case of bare soil, the SAR backscattering coefficient σ^0 depends on the dielectric constant of the soil, the surface roughness and the SAR configuration such as the frequency (f), the incidence angle (θ) and the polarization. Several studies have been carried out to estimate surface soil moisture on bare soil mainly using the X band for the top 3 cm of soil (f ~ 9.6 GHz) and the C band for the top 6 cm

of the soil (f ~5.6 GHz) and showed an estimation precision on soil moisture between 3 and 6 Vol.% (Aubert et al., 2011; Baghdadi et al., 2011a, 2011b, 2016b). However, in the presence of vegetation the soil moisture estimation from SAR data is more complicated. In this case, the radar backscattering signal is also attenuated by the existing vegetation cover. For the soil moisture estimation with vegetation cover using SAR data, the water cloud model (WCM) (Attema and Ulaby, 1978) is mostly used. In the WCM the σ^0 is modelled as a sum of the vegetation contribution $\sigma^0_{vegetation}$ and the soil contribution σ^0_{soil} multiplied by an attenuation factor. The vegetation contribution is principally calculated using a vegetation biophysical parameter such as the biomass, the vegetation height or the vegetation water content. However, recent studies have started exploiting the effectiveness of using vegetation indices such as the Normalized Difference Vegetation Index (NDVI) easily derived from optical remote sensing data to describe the vegetation contribution in the WCM (Bousbih et al., 2018; El Hajj et al., 2017). In the presence of vegetation, several studies have reported a good accuracy for estimating surface soil moisture using the C-band SAR data and the WCM (Baghdadi et al., 2016b; Gherboudj et al., 2011; Mattia et al., 2003).

With the advancement in the satellite technologies and the availability of the several SAR and optical sensors in free and open access mode, several soil moisture products have been recently available for users. In 2009, the European Space Agency (ESA) launched the SMOS mission (Soil Moisture Ocean Salinity) with an objective to measure surface soil moisture and the ocean salinity (Kerr et al., 2001, 2010). SMOS is the first mission dedicated to provide global observation of soil moisture using the L-band interferometric radiometer operating at a frequency of 1.4 GHz. After processing of the level-0 product, the level-2 product of the SMOS provides soil moisture estimation, vegetation optical depth and other ancillary data such as the surface temperature and roughness parameter with their corresponding uncertainties. With a revisit time of 3 days at the equator, the SMOS provides soil moisture estimation at 25 km spatial resolution. Similarly, the National Aeronautics and Space Administration (NASA) launched the Soil Moisture Active Passive mission (SMAP) in January 2015 with a revisit period of 2 to 3 days (Entekhabi et al., 2010). SMAP provides soil moisture estimation at 9 km and 36 km using L-band radiometer at 1.4 GHz. Recently, SMAP product has been merged with Sentinel-1 and Sentinel-2 data to deliver a new soil moisture estimation product at 3 km spatial resolution. In addition, Advanced Scatterometer (ASCAT) provides SSM estimates at coarse spatial resolution and very high revisit time (up to 1 day) (Wagner et al., 2013). ASCAT SSM products are developed by the Department of Geodesy and Geo-information of the Vienna

University of Technology and the data service is provided in partnership with the EUMETSAT's Satellite Application Facility (SAF). ASCAT operates using the C-band radar data at a frequency of 5.3 GHz in vertical-vertical polarization (VV). The surface soil moisture of ASCAT is retrieved from the radar backscattering coefficients by using the change detection method developed in the Institute of Photogrammetry and Remote Sensing (IPF) at Vienna University of Technology (TU Wien) (Naeimi et al., 2009; Wagner et al., 1999).

Regardless of the great contribution in the field of soil water monitoring insured by the numerous soil moisture satellite missions, the spatial resolution offered by the mentioned products remained under investigation. In coarse resolution SSM products, the SSM estimations are retrieved from radar signals averaged at grid scales which itself could be affected by the heterogeneous land cover (such as forests). This heterogeneity in the land cover at grid scales can lead to inaccurate SSM estimation when used for agricultural applications. Moreover, considering irrigation application, soil moisture needs to be accurately estimated on one hand, and spatially refine on the other hand. Indeed, the distribution of the irrigated plots in an agricultural basin is heterogeneous. This heterogeneity can affect the accurate detection of irrigated extent when using coarse resolution soil moisture product at several km^2. For this reason, accurate and detailed (higher resolution) surface soil moisture product is required to reach satisfactory results in irrigation mapping at plot scale.

Recently, a new algorithm for soil moisture retrieval at the agricultural plot scale has been developed at the National Research Institute for Agriculture, Food and Environment (INRAE) particularly with the "Territories, environment, remote sensing and spatial information research unit" (TETIS) in Montpellier France. The developed algorithm for soil moisture mapping is based on the inversion of the WCM using artificial neural networks (NN). As described El Hajj et al. (2017a), the soil contribution σ^0_{soil} has been simulated using the IEM model modified by Baghdadi et al. (2011b) for C-band SAR data. Then, using a parametrized WCM for cereals and grassland (Baghdadi et al., 2017), the total backscattering coefficient σ^0 has been simulated by considering the NDVI as a vegetation descriptor. Therefore, a synthetic database of σ^0 has been created covering a wide range of radar incidence angle θ (20° to 45°), SSM (2 to 40 vol. %), surface roughness (0.5 to 3.8 cm), and NDVI (0 to 0.75) values. To estimate SSM, neural networks (NN) have been trained using a simulated dataset including σ^0, the radar incidence angle (θ) and NDVI. In their study, they validated the proposed NN using SAR data from Sentinel-1 sensor, NDVI data from Sentinel-2 sensor and in situ SSM measurements collected over a study site near Montpellier France. The comparison between the estimated SSM and the

in situ real SSM values showed a significant accuracy for soil moisture estimation with a root mean square difference (RMSD) reaching approximately 5 vol.% at plot scale. Following this study, the TETIS research unit started delivering soil moisture product called the "Sentinel-1/Sentinel-2 Soil Moisture Product" S^2MP at plot scale for different regions across the world. To extract the σ^0 in VV polarization and the SAR incidence angle (θ), the S^2MP mainly uses Sentinel-1 data. To calculate the NDVI values also required by the NN for SSM estimation, the Sentinel-2 images are used. With the availability of the Sentinel-1 and Sentinel-2 data, the operational mapping of SSM has been achieved and the S^2MP products are available in free open access mode via the Theia French Land Data Center (http://www.theia-land.fr/en/thematic-products).

In addition to the validation of the S^2MP with in situ data, the S^2MP has been also compared with the previously mentioned SSM products including SMOS, SMAP and ASCAT. In a study conducted by El Hajj et al. (2018), the S^2MP, SMOS, SMAP and ASCAT were compared to with in situ data derived from terrain measurements and weather station networks (SMOSMANIA). They concluded that the S^2MP showed the best accuracy compared to other products when evaluated against in situ data (RMSD about 5 vol.%).

Since the Sentinel-1 satellite has opened the way towards the operational mapping of soil moisture, the Copernicus Global Land Service (CGLS) has also started delivering soil moisture estimation at 1 km x 1 km grid scale (C-SSM) (Bernhard Bauer Marschallinger and Christoph Paulik, 2018). Similar to the ASCAT data, the SSM retrieval algorithm of the C-SSM is based on the TU Wien Change Detection Model. The C-SSM product is derived from the VV polarization of the S1 SAR signal mainly over the European continent. In a recent validation developed by the CGLS, the C-SSM provided an average accuracy of 10 vol.% in terms of RMSD when compared to several in situ SSM networks over Europe (Bauer-Marschallinger and Pfeil, 2021).

2.2 Remote sensing data for irrigation mapping

The importance of irrigation for crop production along with water scarcity caused by the unsustainable use of water in the agricultural sector encouraged several studies to quantify the irrigated area extent. Before the huge advancement of remote sensing technologies and the availability of massive remotely sensed datasets, large scale irrigation quantification have been initially based on the use of national statistical data rather than remote sensing data. For

example, the FAO GMIA map (Siebert et al., 2005, 2015) and the MICRA map (Portmann et al., 2010) provided only census irrigation data derived from the national declaration resampled at 5 arc-minute spatial resolution. However, the low spatial resolution and the lack of credibility of these products remain an obstacle for irrigation management especially in small and medium agricultural areas.

Currently, the advanced technologies in remote sensing allowed the monitoring of irrigated areas at large scales. Large scale monitoring of irrigated areas using satellite remote sensing can offer a global view with less time and money resources compared to terrain surveys. Several studies have concentrated on the potential of remote sensing data for mapping irrigated areas at different spatial resolutions at and at both local and global scales. Through literature, three main satellite data sources have been exploited for irrigation mapping including optical data, radar (SAR) data and satellite soil moisture data. The following subsections provides a sufficient overview about several studies dealing with irrigation mapping using the three mentioned data source (optical, radar and satellite SSM).

2.2.1 Optical remote sensing for mapping irrigated areas

Using remote sensing, optical data has been firstly manipulated to perform irrigation mapping at different scales. Fundamentals of optical remote sensing in irrigation mapping are based on the difference of the optical spectral signature between irrigated and non-irrigated cropland. Most studies that use optical data demonstrates that irrigated/non-irrigated areas can be distinguished based on the difference in temporal signals of the spectral signature or the difference in the optical derived vegetation indices, such as the NDVI and Greenness index (GI), between irrigated and non-irrigated areas. The NDVI, which represents a proxy measure for absorbed photosynthetic active radiation, is a commonly used vegetation index to map irrigated and non-irrigated crops. In fact, prior studies assessing the effect of water abundancy on the NDVI values showed that increasing available soil moisture for vegetation increases the NDVI value over different crops (Ji and Peters, 2003; Kawabata et al., 2001; Potter and Brooks, 1998; Wulder et al., 2004). When the crop benefits from additional amounts of water through irrigation, highest levels of photosynthesis could be achieved along with highest biomass and densest vegetation cover. These three mentioned biophysical properties induces high NDVI values for irrigated crops. Indeed, several studies have demonstrated that irrigated crops, especially maize and wheat, show higher NDVI than non-irrigated crops (Wardlow and Egbert, 2008). To perform irrigation mapping using satellite data, digital image classification

approaches have been deployed. Previously, traditional image classification techniques such as image segmentation, spectral matching and maximum likelihood classifiers have been employed for mapping irrigated and non-irrigated cropland using optical spectral data (Dheeravath et al., 2010; Thenkabail et al., 2009b). With the advancement of machine learning techniques, studies started to use several machine-learning approaches for image classification. Machine learning models such as the Random Forest (RF), support vector machine (SVM) and neural networks (NN) have been recently used for irrigation mapping using satellite data (Bousbih et al., 2018; Demarez et al., 2019; Gao et al., 2018; Pageot et al., 2020).

The Global irrigated area map (GIAM) (Thenkabail et al., 2009b) was produced for the last millennium using several optical remote sensing products. To map irrigated areas globally at 10 km scale, Thenkabail et al. (2009) used the Advanced Very High Resolution Radiometer (AVHRR) data and SPOT (Satellite Pour l'Observation de la Terre) data in addition to secondary data including rainfall records at 50 km spatial resolution and elevation data from the GTOPO30 (digital elevation model) product at 30 arc-second resolution. The proposed methodology relies on image segmentation applied on both the satellite data bands and the auxiliary data followed by unsupervised decision tree clustering to group the irrigated/non-irrigated classes using the spectral matching technique (SMT). The GIAM had an accuracy ranging between 79% and 91% when compared to ground truth data. Dheeravath et al. (2010) used the same protocol for mapping irrigated areas using MODIS data at 500 m spatial resolution in India between 2001 and 2003. In their study, they used 7 optical reflectance bands of MODIS at 8 days temporal resolution combined with precipitation data, elevation data and national boundaries declared for irrigated areas. The correlation coefficient (R^2) between their annual irrigation map over India and the national data reached 0.84. Finally, they report that the uncertainty of the produced map is mainly linked to the limitation of the spatial resolution (500m), the irrigated area fraction in remote sensing and the lack of robust irrigation declaration in the national surveys.

Merging MODIS data with statistical data has been also used by Pervez and Brown, (2010) to map irrigated agricultural areas over the US continent in 2002. In their study, they used MODIS satellite data with the country statistics derived from the agricultural departments. The classification model was based on the difference in the NDVI time series between irrigated and non-irrigated crops. They rely on the assumption that irrigated crops have higher annual NDVI peaks than non-irrigated crops especially in dry areas with low precipitation records. Thus, irrigated and non-irrigated crops have been classified using the maximum NDVI value attained

during the growing season in addition to statistical and land cover data. The spatial resolution of the produced map was 250 m and accurately identified irrigated lands with reasonable accuracy between 75% and 92% across several sites in the US continent. Brown and Pervez, (2014) followed the same methodology of Pervez and Brown, (2010) to map irrigated areas in the US continent in 2002 and 2007 in order to extract the temporal change in irrigated areas. They achieved similar accuracy in irrigation identification as Pervez and Brown, (2010). Always using MODIS data, the Global Rain-fed Irrigated and Paddy Croplands (GRIPC) provided irrigated area extent at 500 m spatial resolution (Salmon et al., 2015). To perform irrigation mapping, Salmon et al. (2015) used the enhanced vegetation index (EVI), the normalized differential water index (NDWI), the land surface temperature (LST) and the NDVI derived from MODIS data conjointly with auxiliary data related to climate (climate moisture index, annual moisture index) and agro-ecological zoning. Using terrain-training data, a univariate decision tree classification algorithm was applied on both MODIS satellite data and auxiliary data to obtain the irrigation classification map. In the assessment of the GRIPC map against other products, the results showed that the GRIPC irrigated cropland agrees closely with national inventory statistics from FAO (FAOSTAT) but had 24% less surface area of irrigated cropland than GIAM map (Salmon et al., 2015). Recent studies continued exploring MODIS data for irrigation mapping. Xiang et al. (2019) mapped irrigated areas of northeast China by comparing the MODIS derived Land Surface Water Index (LSWI) of the agricultural areas to the surrounding natural vegetation such as forests. They assumed that for irrigated crops, the canopy moisture indicated by the LSWI should be higher than that of adjacent forest. Following this logic, they fixed a threshold value for the difference between the LSWI of the irrigated crops and that of the forest. However, their approach produced an overall accuracy not exceeding 77.2%.

For finer spatial resolution than MODIS, studies dealing with irrigation mapping have also exploited the potential of Landsat imageries at 30 m spatial resolution to map irrigated areas (Chen et al., 2018; Deines et al., 2017, 2019; Demarez et al., 2019; Ren et al., 2021). Chen et al. (2018) used Landsat 8 image combined with MODIS (250 m) and ancillary data (precipitation) to detect the irrigation extent. In their study, they used the GI to detect irrigation events during the first half of the growing season. The overall accuracy for detecting irrigation water supplements reached 87%. Combining Landsat 8 and MODIS data for irrigation mapping has been the also employed in a recent study by Ren et al. (2021) to map irrigated maize (corn) using multi-temporal image classification. In their study, they tested three image classification

algorithms including the neural network, random forest and support vector machine to map irrigated and rain-fed corn at 30 m spatial resolution in Nebraska. First, they performed a MODIS based classification of irrigated and non-irrigated areas, which was then used to support the Landsat irrigation classification. The overall accuracy for irrigation mapping in 2012 over Nebraska reached 89.6% using the NN classifier. (Demarez et al., 2019) tested the potential of Landsat 8 images combined with elevation data to map irrigated summer crops in the south of France. They showed that the combined use of optical imagery with elevation data produces an irrigation classification accuracy of 89%.

With the arrival of the S2 data of the Sentinel satellite constellation, optical data are available in free and open access at 10 m spatial resolution and 5 days revisit time (in Europe). Giving the high spatial resolution offered by S2, several studies started recently exploiting the potential of S2 data for mapping irrigated areas and reported good accuracies for irrigation mapping (Maselli et al., 2020a; Pageot et al., 2020). Pageot et al., 2020 tested the use of S2 derived vegetation indices (NDWI and NDVI) time series to map irrigated and rain-fed crops in a temperate area in south-west France. However, they only achieved an accuracy of 49%.

In conclusion, mapping irrigated areas using optical data have been extensively exploited through literature. Several image classification approaches for irrigation mapping have been proposed using different satellite sensors. Coarse resolution satellite data such as MODIS and AVHRR have been firstly used. With the arrival of Landsat 8 and S2 satellites, maps at finer spatial resolution have been proposed. Using either coarse or fine spatial resolution, optical imagery for irrigation mapping is primary based on the difference in the spectral signature between irrigated and non-irrigated croplands usually expressed using vegetation indices such as the NDVI, NDWI and GI. The main restriction for using optical data for irrigation mapping is the cloud cover that limits the number of exploitable images. When the revisit time of the optical sensor is moderate (such as the Landsat at 16 days revisit time), the availability of high number of cloud free images decreases.

2.2.2 Radar remote sensing for mapping irrigated areas

To overcome the limitation of cloud cover present in optical images, microwave data (SAR) presents new perspectives in irrigation mapping. Microwave data are not limited to weather conditions and can be available during day and night. SAR data has been recently exploited for irrigation mapping in few studies. The key point that links SAR data to irrigation detection is

the soil moisture. As discussed previously in section 2.1, microwave remote sensing is sensitive to the water content of soil due to the increase in the dielectric constant with the increase of the soil water content (Aubert et al., 2013; Baghdadi et al., 2011a, 2011b, 2016a; El Hajj et al., 2017; Hajj et al., 2014). Following this logic, Hajj et al. (2014) analyzed the sensitivity of the TerraSAR X-band SAR data over irrigated grassland plots in the Crau plain of southeast France. They reported that SAR signal could be used to identify three-day-old irrigation event. In the study, they showed that after an irrigation event, the X-band SAR signal increased by approximately 1.4 dB due to an irrigation occurring one day before the satellite acquisition.

Unlike optical data, very few studies have reported the use of SAR data for irrigation mapping. The lack of studies using SAR data for irrigation mapping is primarily due to the lack of operational satellites providing SAR data in free and open access. Sharma et al., (2019) evaluated the use of Radarasat-2 quad pol SAR (C-band) time series to monitor groundwater irrigation in the Berambadi watershed in India. From the quad polarization of Radarsat-2 (HH, HV, VH and VV) 15 polarimetry variables and 11 polarimetry indices were extracted and used with ground-collected data to perform a SVM classification for irrigated and non-irrigated crops. They report a classification accuracy of 91.4% and 90.4% for winter and summer irrigated crops respectively. The arrival of the free and open access Sentinel-1 SAR data (C-band) at 6 days temporal resolution enabled researchers to explore deeply the potential of SAR images for irrigation mapping. Gao et al. (2018) first reported the use of statistical derived metrics from S1 temporal series to map irrigated areas in a semi-arid region in Catalonia, Spain. They assume that due to high persisting soil moisture values in irrigated plots during the year (compared to non-irrigated plots) the SAR temporal signal over irrigated plot must have high mean value, low variance value and high signal correlation length. Thus, they used these three statistical metrics (mean, variance and correlation length) in a RF classifier with collected in situ data to map irrigated areas at plot scale. They achieved an overall accuracy of 81%. Bousbih et al. (2018) followed the same methodology to map irrigated wheat crops in an arid area of Tunisia. The validation of the irrigation mapping showed an overall accuracy up to 77%. Demarez et al. (2019) also reported the importance of joining S1 data and Landsat data for irrigated summer crop mapping in south France. The study showed that the joint use of SAR and Optical data provided the best accuracy for irrigation classification.

Moreover, some studies have focused on the use of microwave satellite soil moisture products for irrigation mapping and estimation. Lawston et al. (2017) reported the potential use of SMAP data at 9 km spatial resolution to map irrigated areas over three semi-arid regions of

the US. In a study performed over Catalonia-Spain, Escorihuela and Quintana-Seguí (2016) reported an inconsistency between satellite soil moisture estimations (ASCAT, SMOS) and simulated surface soil moisture from land surface model over irrigated areas. On the other hand, higher consistency was found over non-irrigated areas. This low correlation between satellite soil moisture and modelled land surface soil moisture over irrigated areas was mainly attributed to the existence of irrigation activities. Particularly, irrigation caused several variation in the satellite estimated soil moisture not accounted in the land surface model leading thus to low correlation values between both SSM estimations. They concluded thus the efficiency of using satellite soil moisture and land surface models for irrigation mapping. Dari et al. (2021) followed the same assumption as Escorihuela and Quintana-Seguí (2016) to map irrigated areas in a watershed in Catalonia. In their study, they assessed the potential of five different satellite soil moisture (SMOS 1km, SMAP 1 and 9 km, Sentinel-1 1 km, and ASCAT 12.5 km) for irrigation mapping by comparing the satellite soil moisture estimation to land surface modelled soil moisture. The method was able distinguish irrigated areas from rain-fed areas with an accuracy up to 78% using SMAP 1km product.

In conclusion, studies using microwave data for irrigation mapping are mainly based on the soil moisture variation as a key variable. Using either raw SAR data (SAR images) or microwave remote sensing soil moisture, recent studies have started reporting good accuracies for irrigation mapping. However, the low spatial resolution of soil moisture products remains an obstacle for a precise quantification of the irrigation extent. Moreover, recent studies dealing with irrigation mapping using S1 data are restricted to the studied zones and lack for the generalization in order to use operationally over several study sites and across several years.

3. Problematic

Based on the previous studies dealing with irrigation mapping and monitoring using remote sensing data, several controversial questions could be addressed in this context. First, the spatial resolution of the maps offered in previous irrigation mapping studies are considered moderately low for irrigation monitoring and agricultural management. The low spatial resolution is mainly due to the use of coarse resolution satellite data (such as MODIS) or soil moisture satellite data such as SMAP and SMOS. Although the low spatial resolution could offer global coverage, it remains an obstacle for irrigation management at local scales in small to medium irrigated agricultural areas. Next, the accurate identification of the irrigation frequency (irrigation events) has not yet received great attention despite its great importance in the field of irrigation

management and water consumption estimation. Given the vital need of the detection of irrigation events and the rare studies focusing on this aspect, the detection of irrigation events at plot scale using remote sensing data is a challenge to address. Finally, we can observe through almost all the studies addressing irrigation mapping that the developed approaches for irrigation mapping heavily rely on supervised image classification models. Although supervised classification approaches such as the RF, the SVM or the NN can provide very good accuracy for irrigation mapping, there exists two important issues to discuss. On one hand, supervised classification models highly rely on the studied geographical contexts and the studied year. Thus, the application of a supervised model (built on one area at a given year) on another geographic area or on another year in the same area is extremely difficult and mostly provide unsatisfactory results. For this reason, previous studies are providing irrigation mapping in a limited geographical area and lack for the generalization over other areas. On the other hand, supervised classification models require the continuous availability of terrain. Thus, the application of any proposed supervised approach requires extensive terrain data measurements especially when using neural networks. However, it is well known that irrigation is a time dynamic activity and may vary from one year to another. Therefore, terrain data (irrigated vs. non-irrigated or irrigation timing) is required yearly or even seasonally in order to perform yearly basis irrigation mapping. Nonetheless, obtaining yearly terrain data at large scales is costly and time consuming. Moreover, the collection of irrigation terrain data could be difficult especially when farmers tend to hide information about the irrigation existence or the irrigation frequency.

In conclusion, this thesis will reply to several controversial points exposed in irrigation monitoring using remote sensing data. The first question concerns the potential of the Sentinel-1 (SAR) and Sentinel-2 (Optical) data for irrigation mapping at high spatial resolution. In particular, what is the potential of the S1 and/or S2 data for mapping irrigated areas at agricultural plot scale? Second, how can we apply a supervised irrigation classification model developed on one area on other regions? Third, the thesis will analyze the capability of detecting the irrigation events at each agricultural plot using remote sensing data. In other words, can we detect the irrigation events occurring at plot scale using Sentinel-1 and Sentinel-2 data in a near real time scenario? Finally, the thesis will analyze the capability of obtaining a semi-supervised classification model for irrigation mapping that overcomes the limitation of the terrain data availability. Expressly, can we build an operational model for irrigation mapping, which requires no in situ terrain data and is easily transferable between geographic regions?

4. Scientific Approach

4.1 General overview

This thesis is fundamentally constructed on the potential of Sentinel-1 and Sentinel-2 images as well as S^2MP (soil moisture at plot scale) data for the monitoring of irrigation activities. Sentinel-1 and Sentinel-2 offers a high spatial resolution at 10 m each, which is appropriate for agricultural applications at plot scale. The use of Sentinel-1 and Sentinel-2 data allows obtaining irrigation mapping at high spatial resolution (up to plot scale). Moreover, Sentinel-1 and Sentinel-2 offers exceptional revisit time among other satellites (6 days and 5 days respectively over Europe) which helps tracking temporal changes in the agricultural field during short periods. The high revisit period of the used satellites is essential for the detection of irrigation events since irrigation is a time dynamic activity. In addition, the plot scale surface soil moisture estimation delivered from the S^2MP product could be an additional data entry for irrigation monitoring tasks thanks to the direct relation between soil moisture and irrigation. Therefore, the complementary between SAR data (Sentinel-1), optical data (Sentinel-2) and accurate surface soil moisture estimation (S^2MP) opens the way towards accurate quantification of irrigated areas and precise detection of irrigation events.

Three important tasks are to be addressed in this thesis. First, it is important to assess the accuracy of the S^2MP product that offers soil moisture estimation at plot scale. As discussed previously, the addition of soil moisture estimation as auxiliary data helps improve the irrigation detection. Thus, an assessment of the S^2MP against in situ data, other SSM products and water cycle parameters (precisely rainfall) has been performed. Second, the thesis focus on discovering the potential of Sentinel-1 and/or Sentinel-2 data for irrigation mapping at plot scale. In this context, supervised classification approaches are investigated with the additional opportunity to transfer the irrigation-mapping model from one region to another. Finally, operational approaches for irrigation event detection and irrigation mapping are to be investigated. In this perspective, a model capable of detecting irrigation events using S1, S2 and S^2MP is proposed. Then, an operational irrigation-mapping framework is constructed.

Our research work replies for these three aspects through different chapters. The second chapter investigates the accuracy of the S^2MP product by comparing it to precipitation data from the Global Precipitation Mission (GPM) and the new Copernicus Global Land Service Soil Moisture product (C-SSM). In the third chapter, we present an irrigation mapping

methodology at plot scale using supervised classification approaches. Moreover, we highlight a transfer-learning approach that permits the use of the developed irrigation-mapping model over different regions. Then, in the fourth chapter, we present a near-real time irrigation event detection approach followed by an operational methodology for mapping irrigated areas, which does not require in situ terrain data. Finally, we finish by a conclusion and a perspective of this work.

4.2 Assessment of the S^{2}MP product

This chapter presents an evaluation of the S^{2}MP product against in situ data, Copernicus SSM product and rainfall data. Before integrating the surface soil moisture estimation provided by the S^{2}MP into the irrigation classification approaches, it is important first to study the correlation of this product with other water cycle parameters (precisely rainfall) in order to understand the interaction between rainfall and soil moisture estimations derived from SAR data. Moreover, it is essential to compare this product with terrain data in addition to the newest surface soil moisture products. This chapter summarizes the principal results detailed in a first paper published in IEEE JSTARS Journal ***"A Comparison of Two Soil Moisture Products S^{2}MP and Copernicus-SSM over Southern France (Bazzi et al., 2019a)*** and a second paper published in published in Sensors Journal ***"Potential of Sentinel-1 Surface Soil Moisture Product for Detecting Heavy Rainfall in the South of France (Bazzi et al., 2019b)"***.

In the first paper, (Bazzi et al., 2019a), we present a detailed comparison between the S^{2}MP and the C-SSM over the Occitanie region of south France. Both products are derived from the S1 data with different soil moisture retrieval methodology. In this comparison, both products were first evaluated versus in situ terrain measurements. Then, an inter-comparison was made between both products in order to determine the correlation between both products. In the second paper (Bazzi et al., 2019b) we present a comparison between the surface soil moisture of the S^{2}MP and the rainfall data of the GPM over the Occitanie region of south France. This comparison allows us to show the consistency between rainfall data and S^{2}MP surface soil moisture at grid scale (10 km x 10 km). This proved consistency was used later in irrigation mapping to discriminate between irrigation and rainfall events. In the second paper

4.3 Supervised classification approaches for mapping irrigated areas using Sentinel-1 and Sentinel-2 data

This chapter focuses on the potential of S1 and S2 data for mapping irrigated areas at plot scale. This chapter presents the results detailed in a first paper published in Remote Sensing Journal: ***"Mapping Irrigated Areas Using Sentinel-1 Time Series in Catalonia, Spain (Bazzi et al., 2019c)*** and a second paper published in IEEE GRSL journal ***"Distilling before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas using Sentinel-1 Time Series (Bazzi et al., 2020a)".***

In the first paper ***(Bazzi et al., 2019c)***, we present a supervised classification approach for irrigation mapping at plot scale based on the use of S1 and/or S2 temporal series using random forest classifier and the convolutional neural network. The approach is based on integrating gird scale SAR backscattering coefficient (10 km x 10 km) as a rainfall indicator (findings of chapter 2) in order to accurately differentiate between irrigation and rainfall using SAR data. The proposed approach was tested over Catalonia region in northeast Spain (semi-arid conditions). Then in the second paper ***(Bazzi et al., 2020a),*** we propose a framework capable of transferring the model built on Catalonia to map irrigated maize plots in Adour-Amont basin of southwest France (humid conditions).

4.4 Towards operational mapping and monitoring of irrigated areas

In this chapter, we discuss first the capability to detect irrigation events at plot scale in a near real-time scenario through a decision tree algorithm using S1, S2 and the S^2MP data. Irrigated and non-irrigated plots were investigated over three different study sites in Catalonia (Spain), Montpellier (southeast France) and Tarbes (southwest France) were examined. The findings are summarized in the paper published in Remote Sensing Journal ***"Near Real-Time Irrigation Detection at Plot Scale Using Sentinel-1 Data (Bazzi et al., 2020b)"***.

The proposed decision tree algorithm for irrigation event detection was then validated against in situ recorded irrigation events over intensively irrigated grassland plots in the Crau plain of southeast France. The findings of this study are summarized in the paper published in Remote Sensing Journal ***"Irrigation Events Detection over Intensively Irrigated Grassland Plots Using Sentinel-1 Data (Bazzi et al., 2020c)".***

In order to obtain an operational methodology for irrigation mapping at plot scale we proposed to merge both the supervised classifiers and the decision tree algorithm in a new semi-supervised framework capable of mapping irrigated areas with no need for in situ terrain data. In the proposed approach, the decision tree is first used to select a training dataset of irrigated and non-irrigated plots. Then a random forest classifier is constructed using S1 and S2 temporal series and selected training dataset. The framework was tested over a new study site located in Orleans (northcentral France) for four different years. The main results are reported in the paper published in Remote Sensing Journal ***"An Operational Framework for Mapping Irrigated Areas at Plot Scale using Sentinel-1 and Sentinel-2 Data".***

4.5 Conclusions and perspectives

This chapter details the main conclusions and presents the limitations of the remote sensing data (particularly Sentinel-1 and Sentinel-2) for irrigation mapping and irrigation event detection. Prospects for further research are thus proposed.

CHAPTER II: ASSESSMENT OF THE S^{2}MP SURFACE SOIL MOISTURE PRODUCT

1. Analytical Summary

1.1 Overview

Soil moisture is a key variable in irrigation mapping. The artificial application of water by irrigation causes an increase of the soil moisture values. Currently, several satellites deliver soil moisture products at coarse spatial resolution (several km²). Recently, the TETIS research unit at INRAE developed an algorithm for surface soil moisture (SSM) retrieval at plot scale called the S^{2}MP (Sentinel-1/2 soil moisture product). This algorithm uses radar Sentinel-1 and optical Sentinel-2 images to estimate surface soil moisture. The product is available in free and open access via the THEIA French land data center for several regions across the world.

The objective of this part is to assess the accuracy of the S^{2}MP product in order to use it next in the irrigation mapping approaches. The assessment of the S^{2}MP was performed in three different approaches. First, the S^{2}MP was evaluated against in situ surface soil moisture measurements to find the accuracy of the S^{2}MP estimations. Next, the S^{2}MP was compared to the recent Copernicus surface soil moisture (C-SSM) product at 1 km x 1 km grid scale developed by the European Space Agency. In order to understand the relation between estimated surface soil moisture and rainfall data, the S^{2}MP was compared finally to precipitation records derived from the Global Precipitation Mission (GPM) at 10 km x 10 km grid scale. The S^{2}MP assessment was performed over the Occitanie region in southern France.

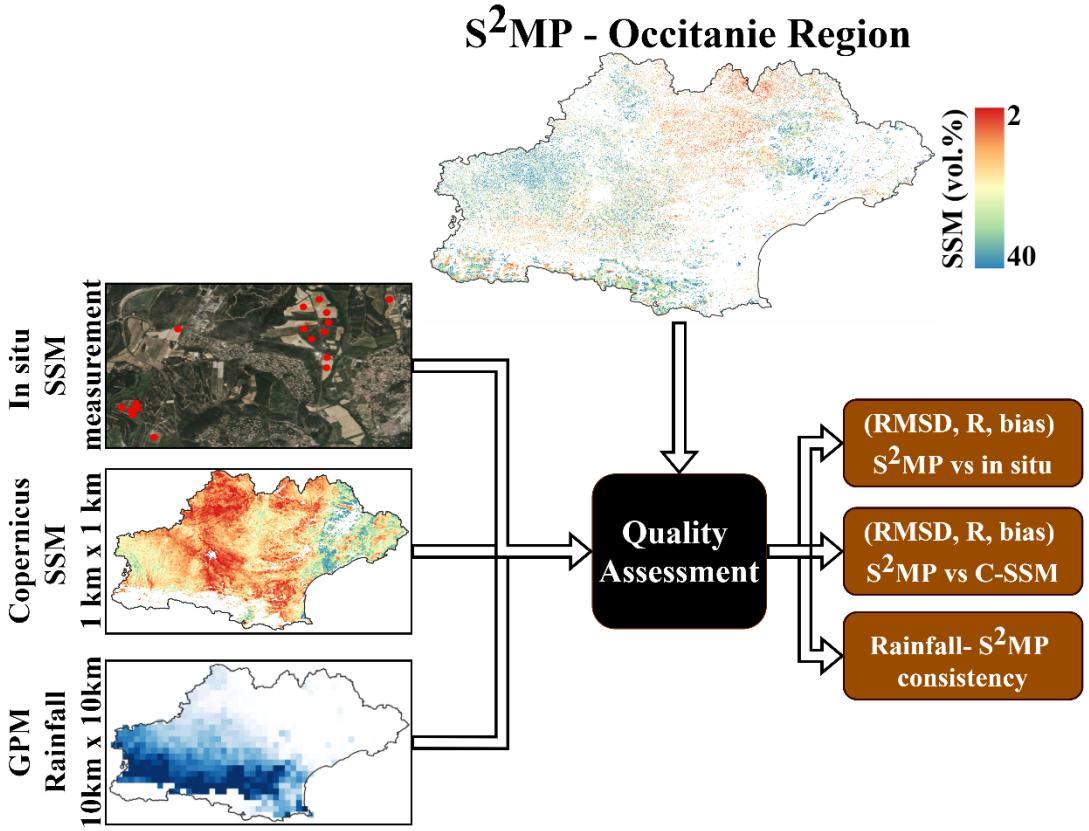


Figure A. Graphical abstract for the S^{2}MP assessment using in situ data, Copernicus SSM product and GPM rainfall data.

1.2 Materials and Methods

1.2.1 Study site

The study site examined for the S^{2}MP assessment is the Occitanie region of south France. It covers an area of 72,724 km^2. The region has a variety of landscape and is mainly covered by agricultural areas in the middle and western parts. In the eastern part of the region, the climate is Mediterranean (average annual precipitation 700 mm), whereas in the western part the climate is humid and oceanic (average annual precipitation 1200 mm).

1.2.2 In situ soil moisture measurements

The SSM values were measured for 23 reference plots (grassland and wheat) at each Sentinel-1 acquisition date (same as S^{2}MP and C-SSM dates) over an agricultural area near Montpellier city (southeast France) for the period between January 2017 and June 2018. For each reference plot, twenty-five to thirty measurements of volumetric soil moisture were

acquired in the top 5 cm of soil by means of a well-calibrated TDR (Time Domain Reflectometry). All soil moisture measurements within each plot were averaged to provide a mean value for each plot at each S1 acquisition date (S^2MP date).

1.2.3 S^2MP product

The S^2MP provides SSM estimations at plot scale for agricultural areas. The SSM estimation is based on coupling S1 and S2 data as proposed by El Hajj et al. (2017). The approach is based on the inversion of the S1 backscattering coefficient using the Neural Network (NN) technique. To estimate SSM values by NN inversion, the S^2MP uses the C-band SAR signal in VV polarization derived from the S1 satellite, the S1 incidence angle (θ) and the NDVI value derived from the S2 optical images. The S^2MP provides soil moisture estimations in absolute volumetric units (vol.%). S^2MP products acquired with the same dates of in situ measurements were used to compare S^2MP to terrain SSM measurements. To compare S^2MP to C-SSM, all the S^2MP maps available, between October 2016 and October 2017, with the same acquisition time of C-SSM were used.

1.2.4 Copernicus Surface Soil Moisture Product

Copernicus Global Land Service delivers the first soil moisture estimations over the European continent at a 1-km spatial resolution using C-band S1 data (C-SSM). The SSM retrieval in the C-SSM is based on the Wein Change Detection Model developed by the TU (University of Technology) of Vienna, Netherlands. The product delivers relative SSM values in percentages ranging between 0% (extremely dry soil conditions) and 100% (very wet soil conditions). C-SSM provides soil moisture estimations in all the S1 overpasses (ascending and descending). Since C-SSM provides soil moisture estimations in relative unit (%), the C-SSM product values were converted to a volumetric unit (vol.%) in order to compare the C-SSM estimations with measured soil moisture in absolute volumetric unit (vol.%).

1.2.5 GPM rainfall data

The GPM (Global Precipitation Mission) provides rainfall measurements globally between 60°N and 60°S at $0.1^{\circ} \times 0.1^{\circ}$ (~10 km x 10 km) spatial resolution. From GPM data, the Late IMERG (Integrated Multi-satellite Retrievals for GPM) rainfall maps at 30 minutes temporal

resolution were downloaded. The comparison between the S^{2}MP and the rainfall data of the GPM was performed on one year between 1 September 2016 and 31 August 2017.

1.2.6 Comparison with in situ measurements

S^{2}MP and C-SSM were first evaluated by comparing SSM values from each product with the in situ soil moisture. The comparison of each product with in situ measurements was performed at 1 km grid scale (same grid as C-SSM). For this reason, at each in situ measurement date, the in situ SSM measurements and the S^{2}MP soil moisture estimations at plot scale were averaged at to 1 km grid scale. The accuracy of the products against in situ data was determined using the statistical metrics including the Pearson correlation coefficient (R), the root mean square difference (RMSD), the bias (estimated SSM – measured SSM) and the unbiased root mean square difference.

1.2.6 S^{2}MP vs C-SSM

Over the entire Occitanie region, an inter-comparison was performed between the S^{2}MP and the C-SSM products for the period between 1 October 2016 and 1 October 2017. The comparison was performed using the statistical metrics R, RMSD, bias, and ubRMSD. The comparison was achieved at each grid cell of the C-SSM (1 km x 1 km) using SSM estimates of only common dates between the S^{2}MP and C-SSM. At each common date, the S^{2}MP estimations (of several plots) existing within each C-SSM grid cell (1 km x 1 km) were averaged to obtain a 1km x 1km S^{2}MP value.

1.2.7 S^{2}MP vs GPM rainfall

For each S^{2}MP map date, one cumulative rainfall map for the 6 days prior to the S^{2}MP map date (revisit of S^{2}MP is 6 days) was computed by summing the 30 minutes time interval of the Late IMERG GPM product. Then, each S^{2}MP map was overlaid with the corresponding 6 days of cumulative GPM rainfall. A temporal analysis was performed between the variation of the SSM values estimated by the S^{2}MP and the GPM rainfall registrations.

1.3 Main results

The comparison between the soil moisture estimation of S^{2}MP and in situ measurements collected over agricultural areas and grasslands showed that the S^{2}MP provides very good

accuracy for soil moisture estimation (RMSD = 4.0 vol.%, ubRMSD = 3.9 vol.% and R=0.77). However, the C-SSM is less accurate than the S^2MP but still provides satisfactory soil moisture estimations (RMSD = 6.0 vol.%, ubRMSD = 6.0 vol.% and R=0.48). The comparison between the two SSM products over one year shows very high correlation between the S^2MP and the C-SSM over agricultural areas mainly used for cereals (R value between 0.5 and 0.9 and RMSE between 4 vol.% and 6 vol.%). However, over agricultural areas mixed with forests and vineyards, the C-SSM values tend to overestimate the S^2MP values (bias > 5 vol.%).

The comparison between the S^2MP SSM and the GPM rainfall data showed that a global consistency exists between the SSM values and precipitation records at 10 km spatial resolution. Very high soil moisture values are estimated after rainfall events, whereas with the absence of precipitation over a period, the soil dries due to evaporation, and the SSM values of the S^2MP generally decrease. For example, between two consecutive S^2MP maps acquired in 15 November 2016 and 21 November 2016 (6 days difference), the SSM values increases by more than 9 vol.% due to 40-mm cumulative precipitation recorded between the two S^2MP dates. Moreover, the absence of precipitation along with evaporation (temperature between 22 °C and 25 °C) for a period of 25 days between 01 April 2017 and 25 April 2017 caused soil moisture values to drop down from 25 vol.% to less than 15 vol.%. The temporal analysis of the S^2MP derived SSM values over a period of time shows the direct effect of raining episodes or dry conditions on the SSM values at grid scale (10 km x 10 km).

1.4 Conclusions

The S^2MP product was compared to both in situ data and C-SSM product. The S^2MP shows very good accuracy for soil moisture estimations and provides SSM estimations better than the C-SSM product. However, for both products, the soil moisture estimations are less accurate and sometimes impossible in the presence of very dense vegetation cover due to the limited penetration of the SAR signal in C-band. In the case of well-developed vegetation cover, the S^2MP does not provide SSM estimations (NDVI > 0.7) while the C-SSM provides unreliable estimations. Over dense vegetation (the case of the wheat with NDVI > 0.7), the soil contribution in the C-band SAR signal becomes negligible, and the estimation of soil moisture could not be obtained using C-band SAR data (El Hajj et al., 2018b).

The comparison between the S^2MP SSM estimations and the GPM rainfall data showed high correlation between presence/absence of rainfall and SSM values at grid scale (10 km x

10 km). The increase/decrease of the SSM values at grid scale could be an indicator of the presence/absence of rainfall events. Thus, to separate the increase in soil moisture (at plot scale) due to rainfall from the increase due to irrigation, the S^2MP SSM estimations at grid scale were subsequently used in the mapping of irrigated areas at plot scale in our irrigation event detection algorithm.

2. Article 1: A Comparison of Two Soil Moisture Products S^{2}MP and Copernicus-SSM over Southern France

Author version of the article published in IEEE JSTARS: Volume: 12, Issue: 9, September 2019

Hassan Bazzi[1], Nicolas Baghdadi[1], Mohammad EL Hajj[1], Mehrez Zribi[2], and Hatem Belhouchette[3]

[1]INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2]CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

[3]CIHEAM-IAMM, UMR-System, Montpellier 34090, France; belhouchette@iamm.fr

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 12 March 2019; Accepted: 28 June 2019; Published: 25 September 2019

Abstract

This paper presents a comparison between the Sentinel-1/Sentinel-2-derived Soil Moisture Product at plot scale (S^{2}MP) and the new Copernicus Surface Soil Moisture "C-SSM" product at 1-km scale over a wide region in southern France. In this study, both products were first evaluated using in situ measurements obtained by the calibrated TDR (Time Delay Reflectometer) in field campaigns. The accuracy against the in situ measurements was defined by the correlation coefficient R, the RMSD (Root Mean Square Difference), the bias and the ubRMSD (unbiased Root Mean Square Difference). Then, the soil moisture estimations from both SSM products were intercompared over one year (October 2016 - October 2017). Both products show generally good agreement with in situ measurements. The results show that using in situ measurements collected over agricultural areas and grasslands the accuracy of the C-SSM is good (RMSD = 6.0 vol.%, ubRMSD = 6.0 vol.% and R=0.48) but less accurate than the S^{2}MP (RMSD = 4.0 vol.%, ubRMSD = 3.9 vol.% and R=0.77). The intercomparison between the two SSM products over one year shows that both products are highly correlated over agricultural areas that are mainly used for cereals (R value between 0.5 and 0.9 and RMSE between 4 vol.% and 6 vol.%). Over areas containing forests and vineyards, the C-SSM values tend to overestimate the S^{2}MP values (bias > 5 vol.%). In the case of well-developed vegetation

cover, the S^2MP does not provide SSM estimations while C-SSM sometimes provides underestimated SSM values.

Index Terms: Soil Moisture, S^2MP, Copernicus SSM, Sentinel-1, South France

1. Introduction

Understanding the water cycle and climate change requires the spatial and temporal monitoring of the Surface Soil Moisture (SSM) (Tebbs et al., 2016). The spatiotemporal monitoring of soil moisture is important in the fields of hydrology, climatology, and agronomy (Herrmann et al., 2016; Sellami et al., 2014). For example, monitoring natural phenomena, such as drought and floods, requires continuous soil moisture data because soil moisture is a key variable in the global water cycle. Moreover, soil moisture data at high spatial resolution is essential for agricultural management, such as monitoring of irrigation practices and water requirements of agricultural areas (Baghdadi et al., 2016b; Bousbih et al., 2018; El Hajj et al., 2016a; Gao et al., 2018). Therefore, continuous soil moisture data at both high spatial and temporal resolutions are important for many hydrological and agricultural models.

To estimate soil moisture, Synthetic Aperture Radar (SAR) data are widely used in both C and X-bands. Soil moisture values are mainly obtained through the inversion of either physical (Chen et al., 2003; Fung, 1994) or statistical models (Baghdadi et al., 2016a; Dubois et al., 1995; Paloscia et al., 2013). While statistical models require site calibration, physical models can always be used to simulate the radar backscattering from radar configuration. Among the physical models, the Integral Equation Model (IEM) (Fung, 1994) is the most commonly used to estimate soil moisture over bare soil or soil with little vegetation cover. Baghdadi et al. (Baghdadi et al., 2006a, 2011b) proposed a semi-empirical calibration of the IEM to compensate between the large difference observed between IEM simulation and real SAR data. To estimate soil moisture over vegetated areas, the Water Cloud Model (WCM) developed by Attema and Ulaby, (1978a) is commonly used. In the WCM the radar backscattering signal is modeled as a sum of the direct vegetation contribution and the soil contribution multiplied by the attenuation factor. Several studies have parametrized the WCM for different SAR band configurations and for several crop types (El Hajj et al., 2016a; He et al., 2014). In addition to the inversion of statistical or physical models for soil moisture retrieval, recent studies have started reporting the estimation of soil moisture using the change detection method. In this method, the change in backscattering coefficients between two successive dates is linked to the

change in soil moisture values where vegetation and roughness are considered stable (Gao et al., 2017; Mattia et al., 2017; Van doninck et al., 2012). Currently, several satellite missions provide surface soil moisture estimations at different spatial resolutions: the Soil Moisture Active Passive "SMAP" (level 3: 36 km x 36 km, level 3 enhanced: 9 km x 9 km, and Level 2 SMAP/Sentinel-1: 1 km x 1 km) (Entekhabi et al., 2010), the Advanced Scatterometer "ASCAT" (level 2 with three spatial resolution 25 km x 25 km, 12.5 km x 12.5 km, and 1 km x 1 km) (Wagner et al., 2013), and the Soil Moisture and Ocean Salinity "SMOS" (SMOS INRA-CESBIO level 3: 25 km x 25 km) (Kerr et al., 2001). However, the low spatial resolution could not be adequate for water cycle monitoring and for agricultural management at plot scale.

Recently, the arrival of the Sentinel-1 (S1) SAR satellite provided users with free open access SAR data at a high spatial resolution (10 m x 10 m) and high revisit time (six days over Europe). The S1 mission from the European Space Agency (ESA) is a constellation of two polar orbiting SAR satellites (Sentinel-1A and Sentinel-1B) operating in the C-band (~5.4 GHz). The S1A and S1B SAR sensors operate in four acquisition modes: Strip Map (SM), Interferometric Wide Swath (IW), Extra-Wide Swath (EW), and Wave (WV). Among the four modes, the IW mode provides images with a spatial resolution of 10 m x 10 m. The SAR data of the S1 mission at high spatial and temporal resolutions have encouraged mapping soil moisture in an operational mode. Paloscia et al. (2013) used the neural network technique to invert the Sentinel-1 SAR C-band and estimate the SSM values. Mattia et al. (Mattia et al., 2017) estimated soil moisture at 1 km x 1 km spatial resolution using the change detection technique applied to Sentinel-1 SAR time series. Recently, El Hajj et al. (2017) developed an operational method to map soil moisture at the plot scale over agricultural areas based on coupling S1-SAR data and Sentinel-2 (S2) optical data using the neural network technique. The synergic use of S1 and S2 data allowed the provision of soil moisture product at the plot scale (S^2MP) for several sites over the world (south France, Lebanon, etc.) Most recently, the Copernicus Global Land Service has started providing SSM of the soil's topmost 5 cm over the European continent from Sentinel-1 sensors at 1 km x 1 km spatial resolution using the change detection method (Bauer-Marschallinger et al., 2019).

The objective of this study is to evaluate the performance of the Copernicus SSM product and S^2MP product over a wide region in south France. In this study, both products are first evaluated against in situ measurements of soil moisture. The new Copernicus SSM product was then intercompared with the SSM estimations from the S^2MP. The correlation between the products was analyzed as a function of the land cover. Section 2 presents the SSM products

used, the in situ SSM measurements, and the applied methods. The results and discussions are presented in Section 3. Finally, Section 4 presents the main conclusions.

2. Materials and Methods

2.1 Study Site

The study site examined is the Occitanie region of southern France (centered on 2˚30′ E and 43˚30′ N, Figure 1), with an area of 72,724 km2 (Figure 1a). Figure 1b presents the land cover of the studied region derived from the land cover map produced by Inglada et al. (Inglada et al., 2017) and available via the Theia French Land Data Center (http://www.theia-land.fr/en/thematic-products). The region has a variety of landscape and is mainly covered by agricultural areas in the middle and western parts. The northern regions are generally covered with a mix of forests, grassland and agricultural crop land. The eastern region is generally covered with a mix of agricultural crop land, grassland and vineyards. Figure 1c shows the elevations derived from the digital elevation model (DEM) of the Shuttle Radar Topography Mission (SRTM) at a 1 arc-second spatial resolution (~30 m). The climate in the eastern part of the region is considered Mediterranean (approximately 700 mm average annual precipitation), whereas that of the western part is more humid and oceanic (approximately 1200 mm average annual precipitation) (http://www.meteofrance.com/climat/france/).

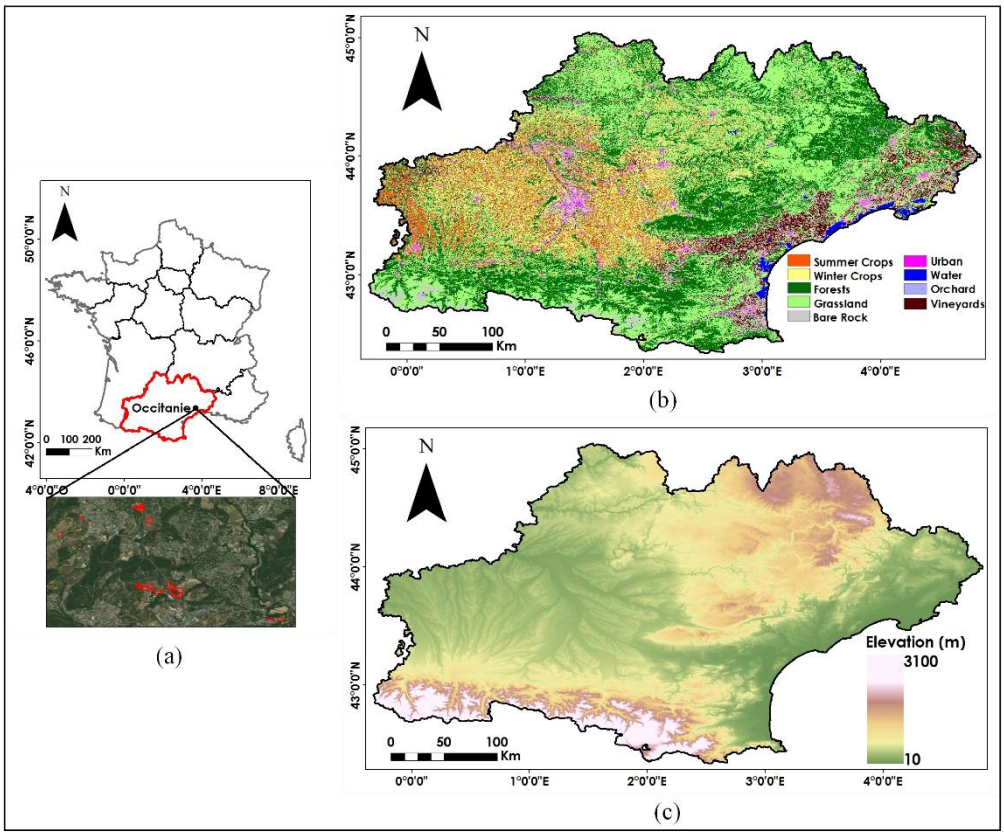


Figure 1 : (a) Location of the Occitanie region, south France, with the in situ reference plots. (b) Land cover map of Occitanie produced by Theia (http://www.theia-land.fr/en/thematic-products) (c) Elevation derived from the SRTM DEM of 30 m spatial resolution

2.2 In situ Soil Moisture Measurements

Over an area near Montpellier, France (Figure 1a), the SSM values were measured on 23 reference plots (10 grassland and 13 wheat) during the period between 1 January 2017 and 31 May 2017 (El Hajj et al., 2017). The SSM values were measured within 2 hours of the S1 acquisition time in 22 different campaigns at different dates of S1 acquisition. For each plot, 25 to 30 volumetric soil moisture measurements were taken from the top 5 cm layer using a calibrated TDR (Time Delay Reflectometer). Then, within each plot, the SSM measurements were averaged to obtain a mean value for each plot. The measured soil moisture values varied between 7.0 vol.% and 36.3 vol.%.

2.3 S^{2}MP Product

The S^{2}MP product is obtained by coupling Sentinel-1 SAR data and Sentinel-2 optical data. The S^{2}MP product provides SSM estimates over the agricultural areas at plot scale with 6 days revisit time (El Hajj et al., 2017). To estimate the SSM values, El hajj et al. (El Hajj et al., 2017)

inverted the WCM parameterized by Baghdadi et al. (Baghdadi et al., 2017) for C-band combined with the IEM, as modified by Baghdadi et al. (Baghdadi et al., 2006a) . The inversion approach uses the neural network (NN) technique to invert the radar signal into SSM value. To operationally map the soil moisture, the developed NN uses the C-band SAR signal in VV polarization, SAR incidence angle, and the Normalized Differential Vegetation Index "NDVI" as the inputs. While the SAR signal and incidence angle are derived from the S1 data, the NDVI value is derived from Sentinel-2 images. To overcome the cloud limit usually present in optical images, one NDVI image each month is usually used to obtain the NDVI value required for the inversion. The S^2MP maps are produced for agricultural areas (it is not applied to vineyards and orchards). Forest and urban areas are masked using the land cover map of Inglada et al. (Inglada et al., 2017). Additionally, areas with slope greater than 20% are masked in the soil moisture product (calculated from SRTM DEM at 30 m spatial resolution). The S^2MP are available in free open access mode via the Theia French Land Data Center (http://www.theia-land.fr/en/thematic-products). In this study only the S^2MP maps for images in ascending mode were produced and compared to in situ and C-SSM data for two main reasons. First S1 images in descending mode are acquired at 05:40 UT with a high probability of freeze in winter causing an under-estimation of SSM (Baghdadi et al., 2018b). Second, in situ measurements were acquired within two hours of the ascending S1 acquisitions. Thus, our in situ measurements could be not applicable to SSM estimations obtained from S1 images in descending mode.

2.4 Copernicus SSM product

Recently, Copernicus Global Land Service began to distribute the first soil moisture estimations over the European continent at a 1-km spatial resolution using S1 data in C-band (Bauer-Marschallinger et al., 2019). The SSM retrieval algorithm is based on the TU (University of Technology) Wien Change Detection Model already applied for the ASCAT data and adapted for Sentinel-1 data. In the applied model, the changes observed in the SAR backscattered coefficient (σ^0) are interpreted as changes in the soil moisture values, whereas other surface properties such as the geometry, the surface roughness and the vegetation cover are interpreted as static parameters. To estimate relative surface soil moisture at time t (SSM_t) in percent (%), the backscattered coefficient $\sigma^0(\theta, t)$ observed at time t with a given SAR incidence angle θ is normalized to a reference angle Θ and linearly scaled between dry and wet reference values:

$$SSM_t = \frac{\Delta\sigma^0(\Theta, t)}{S(\Theta)} \ [\%] \quad (1)$$

where $\Delta\sigma^0(\Theta, t)$ is the change in the normalized backscatter (relative to dry conditions), and $S(\Theta)$ is the sensitivity to the SSM changes at reference angle $\Theta = 40°$ expressed as follows:

$$S(\Theta) = \sigma^0{}_{wet}(\Theta) - \sigma^0{}_{dry}(\Theta) \ [dB] \quad (2)$$

where $\sigma^0{}_{wet}(\Theta)$ and $\sigma^0{}_{dry}(\Theta)$ are the highest and lowest backscattered values at the individual location respectively.

The C-SSM products at 1-km spatial resolution are retrieved from the Sentinel-1 radar backscattering images acquired in IW mode and VV polarization. The product delivers relative SSM values in percentages ranging between 0% and 100%. In the case of extremely dry conditions, frozen soil, snow-covered soil and flooding, the C-SSM retrieval is ill posed. In these cases, the C-SSM does not provide estimation, and it encodes values of 241 and 242 in the product. Moreover, the product masks water surfaces (sea, lakes, rivers…), urban areas and dense forests with values equal to 252. High undulating terrains with slopes greater than 30% are also masked in the product and encoded as 253.

The C-SSM product values were converted to a volumetric unit (vol.%) in order to compare the C-SSM estimations produced in relative units (%), with measured soil moisture being given in absolute volumetric unit (vol.%). Thus, the C-SSM values were converted to vol.% unit (SSM_α) by using the 90% confidence interval of a Gaussian distribution (Amri et al., 2012) equal to $\mu + 1.65\sigma$, where μ and σ are the mean and the standard deviation of the TDR ground data:

$$SSM_\alpha(t) = SSM(t) \times (SSM_{max} - SSM_{min}) + SSM_{min} \quad (3)$$

where $SSM_\alpha(t)$ is the volumetric surface soil moisture at a time t (in vol.%), $SSM(t)$ is the relative C-SSM soil moisture content (in %), SSM_{max} is the maximum measured wetness value (in vol.%) equal to $\mu + 1.65\sigma$, and $SSM_{min}$is the minimum measured wetness value (in vol.%) equal to $\mu - 1.65\sigma$. Using the in-situ measured soil moisture values, $SSM_{max} = 39.7$ vol. % and $SSM_{min} = 7.5$ vol. %.

Figure 2 shows an example of the C-SSM product covering a part of southern France on 25/03/2017. The physical soil moisture values of the C-SSM product are converted to volumetric units (vol.%). Strong discontinuities are observed in the C-SMM product between

the different sub-swaths of Sentinel-1 SAR data (black arrows). As an example, the statistical mean calculated for the two magenta polygons in Figure 2 (at both sides of the two sub-swaths) shows that the difference in SSM estimates is 11 vol.%

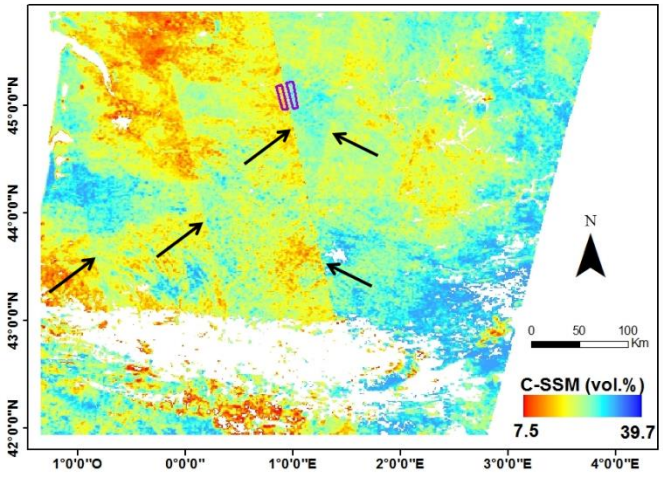


Figure 2 : C-SSM product over south France. White areas represent the mask applied in the C-SSM product. Arrows are used to highlight the discontinuity observed between the sub-swaths with the C-SSM product. Magenta polygons are used to calculate the C-SSM mean value at the sub-swath limits.

2.5 Methods

Since the C-SSM product is produced at a 1 km spatial resolution, and the S^{2}MP product is produced at the plot scale, the evaluation was carried out at a 1 km $\times$ 1 km spatial resolution (same grid of the C-SSM product). For each field campaign, the in situ soil moisture measurements carried out on all plots within the same 1 km $\times$ 1 km of the C-SSM grid cell were averaged. Moreover, the S^{2}MP product was also aggregated by averaging the high-resolution soil moisture pixels of S^{2}MP within each C-SSM cell of 1 km $\times$ 1 km. This means that, at a given date, for each C-SSM cell (1 km $\times$ 1 km) the values of the S^{2}MP estimations within the cell were averaged to obtain a 1 km $\times$ 1 km S^{2}MP value. Therefore, for each available S^{2}MP map, the mean soil moisture values were calculated for the corresponding C-SSM grid.

First, S^{2}MP and C-SSM were evaluated by comparing SSM values from each product (at 1 km spatial resolution) with the in situ soil moisture, which were also averaged at 1 km C-SSM grid cell. Since both SSM products have been produced at different temporal resolution (in this study S²MP was applied only on S1 acquisitions in ascending mode whereas that C-SSM was generated for all S1 overpasses), only those SSM maps with common dates between S^{2}MP, C-SSM and in situ campaigns were considered in the comparison. Thus, the sample size used for

the comparison is the same for both SSM products. The accuracy of the products was determined using the statistical metrics including the Pearson correlation coefficient (R), the root mean square difference (RMSD), the bias (estimated SSM – measured SSM) and the unbiased root mean square difference. These metrics are calculated as follows:

$$R = \frac{\sum_{i=1}^{n}[(O_i - \overline{O})(P_i - \overline{P})]}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2}} \quad (4)$$

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \quad (5)$$

$$Bias = \frac{1}{n}\sum(O_i - P_i) \quad (6)$$

$$ubRMSD = \sqrt{RMSD^2 - Bias^2} \quad (7)$$

where P_i is the in situ measurement at grid cell i, O_i is the SSM product estimation at grid cell i, $\bar{P}$ is the average value of the in situ measurements at all compared grid cells, and $\bar{O}$ is the average value of the SSM product estimation at all grid cells. The four statistical metrics were calculated for each product. The obtained results are presented in Section 3.1.

Then, an intercomparison was performed between the S^{2}MP and the C-SSM products over the entire Occitanie region for the period between 1 October 2016 and 1 October 2017 using the statistical metrics *R*, RMSD, bias, and ubRMSD. The intercomparison was performed at each grid cell using SSM estimates over one year. Since the two products have a different temporal resolution, only those SSM values with common dates between the S^{2}MP and C-SSM were considered in the intercomparison. The same grid of C-SSM was used for the re-gridding of S^{2}MP. For each C-SSM cell (1km x 1km), the S^{2}MP SSM estimations existing within the C-SSM cell were averaged to obtain a 1km x 1km S^{2}MP value. The statistical metrics were not calculated for those cells where the S^{2}MP product or the C-SSM product does not provide soil moisture estimation over the entire year due to the applied masks. Four maps, each representing one of the statistical metrics, are presented and discussed in Section 3.2.

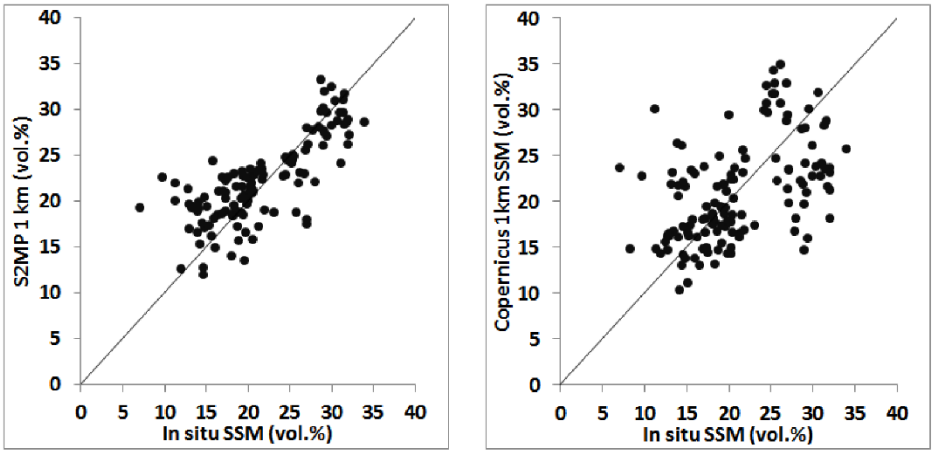
Finally, the temporal evolution of the soil moisture estimations derived from both SSM products was investigated. This temporal evolution was performed for one wheat and one maize

grid cell (the cell is predominantly covered by wheat and maize, respectively). Temporal profiles are further presented in Section 3.3.

3. Results and Discussion

3.1 Comparison between SSM products and in situ SSM

The SSM values of each product were evaluated using the in situ SSM measurements during the period between January 2017 and May 2017. Figure 3 shows the comparison between the SSM products and the in situ SSM measurements, whereas Table 1 summarizes the statistical metrics obtained from the comparison of each product with the in situ measurements. In general, the accuracy of the S^2MP product is higher than that of the Copernicus SSM. The S^2MP shows a higher R value than that of the Copernicus SSM. Additionally, the RMSD is lower for the S^2MP than for the Copernicus SSM. Both products show approximate unbiased estimations. However, the bias between in situ and satellite soil moisture could be driven from the soil texture maps used for converting dielectric constant in volumetric soil moisture as well as from the accuracy of inversion approaches. Nevertheless, the Copernicus SSM product performs well with an RMSD of approximately 6 vol.%. Notably, the p-value of the comparison for both products is lower than 0.01, which indicates that the correlation is significant.



(a) (b)

Figure 3 : SSM products against in situ measurements. (a) S2MP-1 km, (b) Copernicus SSM-1 km.

Table 1 : Statistics of the comparison between SSM products and in situ SSM measurements

Product	R	RMSD (vol.%)	Bias (vol.%)	ubRMSD (vol.%)	Sample size
S²MP	0.78	4.0	0.52	3.9	122
C-SSM	0.48	6.0	-0.32	6.0	122

3.2 Comparison between S²MP and C-SSM Products

Figure 4 represents the maps of statistical metrics calculated from the comparison between the SSM from the S²MP and C-SSM products during the period between October 2016 and October 2017. Figure 4a represents the correlation coefficient R between the two products, Figure 4b represents the RMSD values, Figure 4c shows the bias values (C-SSM - S²MP), and Figure 4d shows the ubRMSD values over the Occitanie region. Moderate to high correlation values between the compared products are observed, mainly in the western, middle and eastern part of Occitanie (R between 0.5 and 0.9). Figure 4b shows that the RMSD values for these parts are approximately homogeneous and ranging between 4 and 6 vol.%. The ubRMSD (Figure 4d) for these parts show also low values. Between these parts, only the eastern part shows high positive bias values (between 2 and 4 vol.%). The correlation map (Figure 4a) shows low values mainly in the northern part of Occitanie, with moderate RMSD values (between 6 and 8 vol.%). High RMSD values (more than 8 vol.%) are observed for the northeastern part. The bias map for the northern area shows high positive values (Figure 4c). The ubRMSD is generally homogeneous over the Occitanie, with values ranging between 3 and 7 vol.%. Notably, the fringes observed in the maps are due to the heterogeneity of the SSM estimates in the C-SSM product at the borders of the sub-swath, as illustrated in Section II.4. The great difference in SSM-values observed at the limit of the sub-swath induces a difference of bias values between two adjacent areas at the limit of the sub-swath (case of far western part in R, bias and ubRMSD maps).

For some Copernicus cells (1km x 1km) only a small number of S²MP pixels are sometimes used for calculating the average soil moisture estimation from S²MP. For example, only 8.5% of 1 km x 1 km cells contain between 20 and 200 S²MP pixels (the minimum was fixed to 20 pixels) whereas 75% of the 1 km x 1 km cells contain more than 1000 S²MP pixels. The correlation between the two products for the cells with low number of S²MP pixels was analyzed mainly for the cells containing between 20 and 200 S²MP pixels. Results showed that

there are as many low values as high values of the correlation coefficient (R) between both products at these cells (same number of cells with R less than 0.3, between 0.3 and 0.5 and with R higher than 0.5).

The statistical metrics obtained from the comparison of both SSM products are analyzed as a function of the land cover (Figure 1b). The S^{2}MP product provides soil moisture estimation over agricultural crop land only grown in summer and winter (urban, forest, grassland and vineyards are masked), which should be considered before analyzing the effect of the land cover. Thus, we suppose that the soil moisture calculated from the S^{2}MP product is the same for the whole cell of 1 km $\times$ 1 km (not composed only of crops). The results reveal that the western and middle parts show the highest R values with smallest bias and RMSD values. In the land cover map, this area shows a majority of summer and winter crops (more than 65% of the area). The eastern part that shows high correlation values but moderate to high bias values is composed of a mix of vineyards (30%), grassland (30%) and summer-winter crops (14%). The northern part, with the lowest *R* and moderate RMSD and bias values, is generally occupied with a mix of grassland (60% of the area) and forests (23% of the area). This part comprises no more than 9% agricultural areas. Finally, the northeastern part (high RMSD and bias values) is mostly covered by a mix of forests (40%) and grassland (40%).

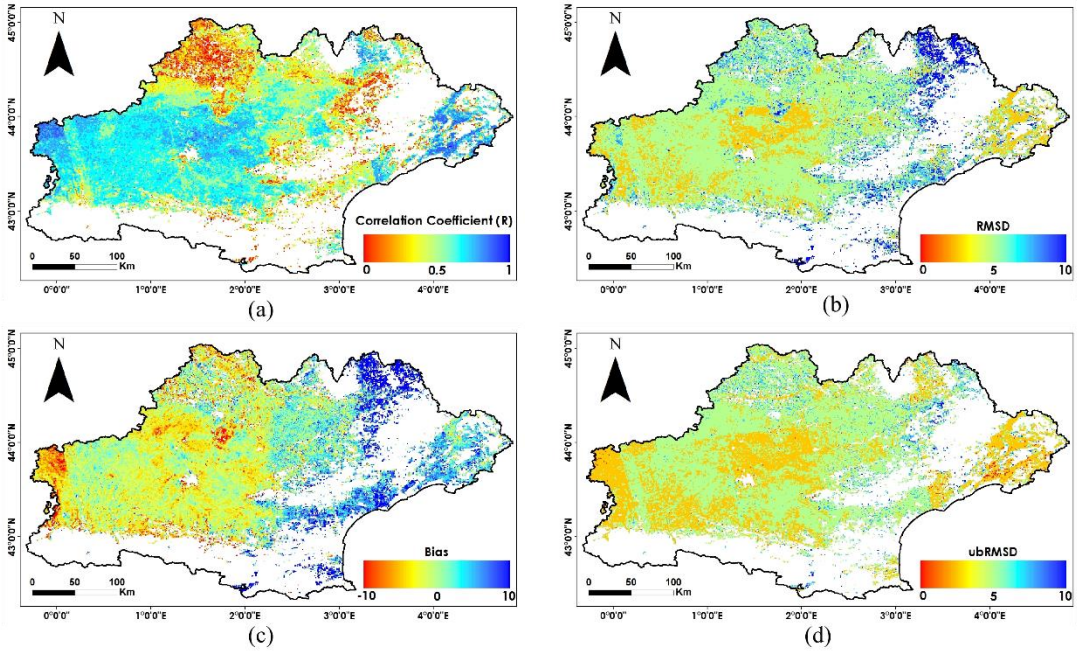


Figure 4 : Statistical metrics maps at 1-km grid generated from the comparison between the C-SSM and S2MP over the Occitanie region. (a) Correlation coefficient R map, (b) RMSD map, (c) Bias map, and (d) ubRMSD map

Figure 5 shows the scatter plots for the comparison between the C-SSM and S^2MP over four different cells (each of 1 km × 1 km) with four different land cover types. Figure 5a shows the scatter plot comparison over a cell with mainly grasslands (89% grassland and 6% summer crops). The comparison indicates low correlation (R=0.2) but moderate RMSD value (7.6 vol.%) and unbiased estimation. Figure 5b shows the scatter plot for a cell covered with mainly agricultural areas (80% summer and winter crops and 5% grassland). In this case, the C-SSM and S^2MP are highly correlated (R=0.8, RMSD=4.6 vol.%, bias = 1 vol.%). This result confirms the evaluation results obtained in Section 3.1, where both the C-SSM and S^2MP show good agreement with in situ measurements over agricultural areas. Figure 5c shows the SSM estimations of both products over a cell with mainly forests (86% forests and 7% summer crops). In this cell, the overestimation of C-SSM with respect to the S^2MP is visible (bias = 8 vol.%). In fact, over forests, the backscattered signal in C-band is highly affected by the forest canopy because the penetration of the SAR signal to the ground surface is very low. High backscattering coefficients that are observed in the forest could induce high soil moisture estimations in C-SSM and thus an overestimation of SSM by the C-SSM product. Similarly, Figure 5d shows the intercomparison at a cell with a mix of agricultural and vineyard areas (75% vineyards and 18% summer crops) with high bias values, where the C-SSM highly overestimates the S^2MP. High biased values over vineyards (5.6 vol.%) are due to the high backscattering signal reflected from metals and wooden stakes that are usually present in vineyards. Baghdadi et al. (Baghdadi et al., 2006b) demonstrated that vineyards have strong radar backscattering signal because of these stakes. This strong backscattering signal, which is not related to the soil contribution, could induce higher soil moisture estimations.

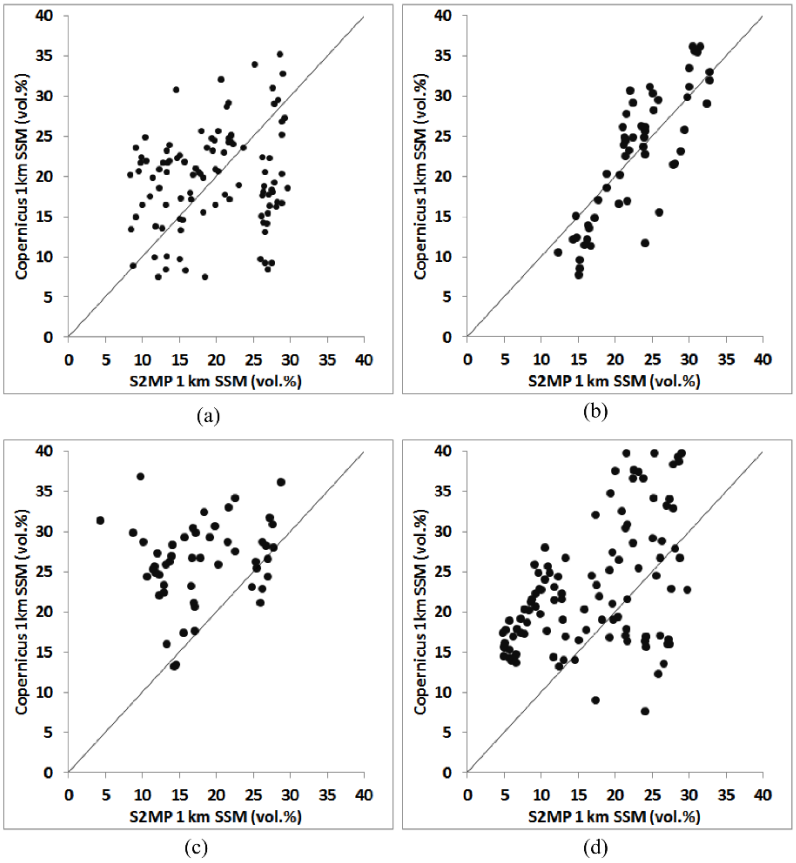


Figure 5 : C-SSM against S²MP over cell of 1 km x 1 km for (a) mainly Grassland (89%); (b) mainly agricultural areas (80% summer and winter crops), (c) mainly Forests (86%), and (d) mainly Vineyards (75%).

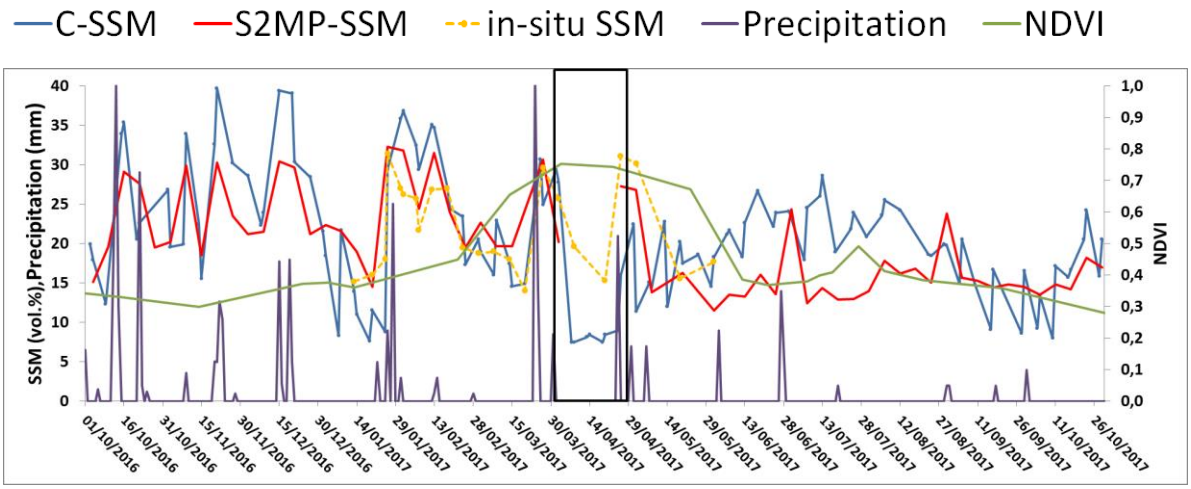
The analysis of the SSM products' quality according to land cover type reveals that over areas mainly used for cereals and market gardening (the case of the middle and western parts), the correlation between the SSM products is high, indicating high consistency in soil moisture estimation between the products. When the area becomes more occupied by grassland (the case in the northern part), the correlation decreases but keeps moderate RMSD and bias values. However, when the forested area dominates and the land cover becomes more heterogeneous (forests, grassland, agricultural…) (the case in the northeastern part) an overestimation of C-SSM with respect to S^2MP is observed. Moreover, when the vineyards are dominant in the land cover (the case of the eastern part) overestimation of C-SSM with respect to S^2MP is observed. The limitation observed over heterogeneous land cover for the C-SSM product could be related to the spatial resolution because the product is produced at a 1-km pixel size and not at a fine scale.

3.3 Temporal behavior of SSM Series

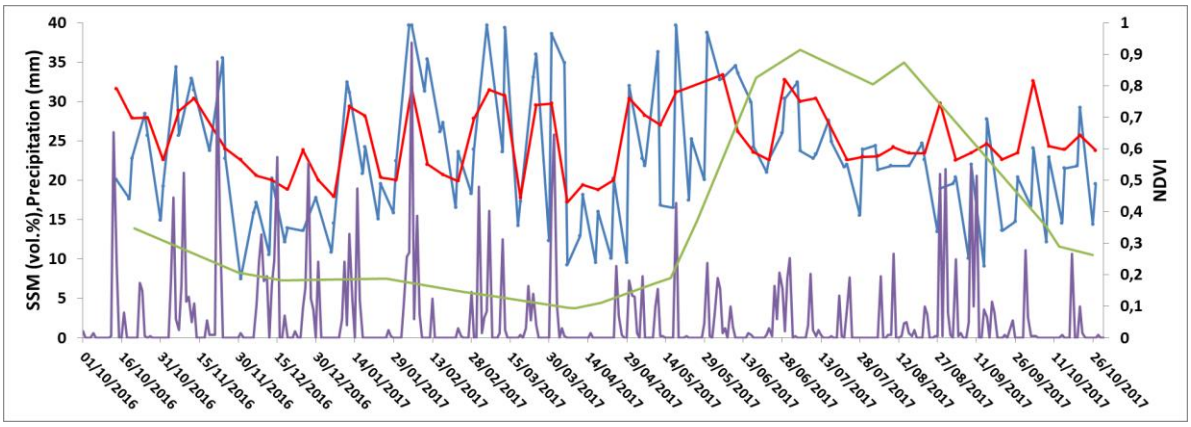
The temporal evolution of the soil moisture values was investigated over two cells (1 km × 1 km): one totally covered with winter wheat cultivation (Figure 6a) near Montpellier, France, and the second with summer maize cultivation in the western part near Tarbes, France (Figure 6b). For the wheat cell, in situ soil moisture values are also available. Unfortunately, in situ measurements are not available over the maize cell of the western part. In addition to S^2MP and C-SSM estimates, the NDVI –values derived from Sentinel-2 temporal series images and averaged over the 1-km grid cell were also plotted. The precipitation amounts derived from a local metrological station near each cell (local station in Montpellier for Figure 6a and local station in Tarbes for Figure 6b) were used to qualitatively analyze the behavior of the SSM products. The temporal resolution of the C-SSM is also higher than that of the S^2MP; thus, more points are obtained for the C-SSM product. As mentioned before, the correlation coefficient could only be calculated for soil moisture values of the common estimation dates. For both land cover types, the correlation coefficient between the SSM products is significant (0.63 for maize and 0.77 for wheat). For the maize cell, the RMSD value is of 5.7 vol.%, whereas that of the wheat is 5.2 vol.%. Moreover, both products show a consistency with the precipitation events. After a precipitation event, high soil moisture values are estimated in both products (in both maize and wheat cases). However, the C-SSM values frequently show very high SSM values following rainfall events, and the estimation reaches 100% in relative soil moisture (39.7 vol.% in volumetric soil moisture). Following a dry period, both products show low soil moisture values, thus expressing dry soil conditions.

An important point in the temporal series concerns the winter wheat cultivation. During the period between 30 March 2017 and 29 April 2017 (black box in Figure 6a), the C-SSM fails to estimate accurate soil moisture values. The relative soil moisture obtained from the C-SSM product is approximately 0% during this period (which corresponds to volumetric soil moisture of approximately 7.5 vol.%). When comparing these values to the in situ soil moisture measurement, we find that C-SSM abruptly drops down and underestimates the soil moisture. For the same period, the S2MP does not provide any surface soil moisture estimation (gap in the red curve). This is due to the fact that, over dense vegetation (the case of the wheat with NDVI > 0.7), the soil contribution in the SAR signal in C-band becomes negligible, and the estimation of soil moisture could not be obtained. El Hajj et al. (2018b) studied the penetration of C-band SAR data over wheat, maize, and grassland. They showed that the C-band in VV

polarization is not able to penetrate the wheat canopy when the vegetation is well developed (NDVI > 0.7). As a result, the estimation of the SSM values is not reliable. For this reason, in the S2MP algorithm, a mask was applied to eliminate those likely unreliable SSM values for NDVI greater than 0.7 and SSM estimations that are less than 5 vol.%.



(a)



(b)

Figure 6: Temporal series of SSM values derived from C-SSM and S²MP products over (a) a winter wheat cultivation cell and (b) a summer maize cultivation cell

For the maize canopy (Figure 6b), both products are able to estimate the soil moisture values, even with well-developed vegetation. For example, between the end of July and the end of August, the NDVI remains high (approximately > 0.8), and both products are able to show reliable SSM values. The SSM values decrease during a time span without rainfall events and increase following a rainfall event, even when the canopy is well developed. However, both El Hajj et al. (2018b) and (Joseph et al., 2010) reported that the soil contribution to the C-band backscattering is notable when the maize canopy is well-developed due to the high order scattering along the soil-vegetation pathway that contains the soil contribution.

4. Conclusion

The aim of this study was to assess the accuracy of Surface Soil Moisture (SSM) in southern France estimated by the Sentinel-1/Sentinel-2-derived Soil Moisture Product at plot scale (S^2MP) and the new Copernicus Surface Soil Moisture "C-SSM" product at 1-km scale. Using in situ measurements, the accuracy of each product was first evaluated and determined by the Pearson correlation coefficient (R), the root mean square difference (RMSD), the bias, and the unbiased root mean square difference. Then, an intercomparison between the products was performed for one year (between October 2016 and October 2017) over the entire Occitanie region, south France. The results of the intercomparison were further discussed as a function of the land cover. The results reveal that both products show good reliability with in situ measurements. However, S^2MP shows higher accuracy (RMSD = 4.0 vol.%, R = 0.77) than the C-SSM (RMSD = 6.0 vol.%, R=0.48) when compared to in situ measurements over agricultural areas and grasslands. The analysis of the intercomparison between the products as a function of the land cover shows that the S^2MP and C-SSM are highly correlated over dense agricultural areas (mainly cereals of winter and summer). When the land cover becomes more heterogeneous (mix of forests, grassland and vineyards with agricultural areas) the correlation between the products decreases. The C-SSM tends to overestimate SSM values over cells containing vineyards and forests. In addition, the C-SSM sometimes shows underestimated SSM values when the vegetation is well developed. In wheat areas, the underestimation is approximately 7.5 vol.% when NDVI is greater than 0.7. On the other hand, the algorithm of S^2MP eliminates likely unreliable SSM values for NDVI greater than 0.7 because the SSM estimations in this case are less than 5 vol.%.

5. Acknowledgment

The authors wish to thank the French Space Study Center (2019 TOSCA) and the National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA) for supporting this work.

3. Article 2: Potential of Sentinel-1 Surface Soil Moisture Product for Detecting Heavy Rainfall in the South of France

Author version of the article published in IEEE JSTARS: Volume: 12, Issue: 9, September 2019

Hassan Bazzi[1], Nicolas Baghdadi[1], Mohammad EL Hajj[1] and Mehrez Zribi[2]

[1]INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2]CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 7 January 2019; Accepted: 14 February 2019; Published: 16 February 2019

Abstract

The objective of this paper is to present an analysis of Sentinel-1 derived surface soil moisture maps (S1-SSM) produced with high spatial resolution (at plot scale) and a revisit time of six days for the Occitanie region located in the South of France as a function of precipitation data, in order to investigate the potential of S1-SSM maps for detecting heavy rainfalls. First, the correlation between S1-SSM maps and rainfall maps provided by the Global Precipitation Mission (GPM) was investigated. Then, we analyzed the effect of the S1-SSM temporal resolution on detecting heavy rainfall events and the impact of these events on S1-SSM values as a function of the number of days that separated the heavy rainfall and the S1 acquisition date (cumulative rainfall more than 60 mm in 24 hours or 80 mm in 48 hours). The results showed that the six-day temporal resolution of the S1-SSM map doesn't always permit the detection of an extreme rainfall event, because confusion will appear between high S1-SSM values due to extreme rainfall events occurring six days before the acquisition of S1-SSM, and high S1-SSM values due to light rain a few hours before the acquisition of Sentinel-1 images. Moreover, the monitoring of extreme rain events using only soil moisture maps remains difficult, since many environmental parameters could affect the value of SSM, and synthetic aperture radar (SAR) doesn't allow the estimation of very high soil moistures (higher than 35 vol.%).

Keywords: Sentinel-1; global precipitation measurement; soil moisture; heavy rainfall

1. Introduction

Understanding the Earth's global water cycle requires the spatiotemporal monitoring of precipitation (Kidd and Huffman, 2011; Liu, 2016). Moreover, precipitation monitoring is essential, since rainfall directly affects human life. Indeed, very high rainfall episodes may cause runoff, which damages life and property, whereas the absence of precipitation can produce droughts (Liu, 2016). Thus, knowing the extent of rain helps us better understand the effect of climate and weather on agriculture and water availability.

The registration of in situ rain gauges allows only the measurement of local precipitation amounts. Thus, these rain gauge instruments are not sufficient to provide robust continuous mapping of rainfall, because rainfall events can be very local. Moreover, in situ rain gauge registrations are not usually freely available. On the other hand, remote sensing metrological sensors enable the continuous mapping of precipitation in open access mode. Currently, several metrological sensors provide rainfall mapping with a high revisit time and low spatial resolution: the Tropical Rainfall Measuring Mission (TRMM) (resolution $\sim 0.25^{\circ} \times 0.25^{\circ}$) (Kummerow et al., 1998), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN) (resolution $\sim 0.25^{\circ} \times 0.25^{\circ}$) (Hsu et al., 1997), the Climate Prediction Center Morphing (CMORPH) (resolution ~8 km × 8 km at the equator) (Joyce et al., 2004). The most recent sensor, Global Precipitation Measurement (GPM), provides precipitation mapping at $0.1^{\circ} \times 0.1^{\circ}$ spatial resolution and a high revisit time (up to 30 minutes) (Skofronick-Jackson et al., 2018). The low spatial resolution makes the available metrological sensors not adequate for hydrological application, flood forecasting, and water resource management. For example, the spatial detection of extreme rainfall events remains difficult due to the low spatial resolution of available precipitation data. Extreme rainfall events can be defined as rainy episodes with high cumulative precipitation in short periods (24 to 48 hours), which is equal to that usually received in one month.

An important variable affected by precipitation events is the surface soil moisture. In fact, surface soil moisture can be a key tool for detecting rain events, because soil moisture is a function of the precipitation rate, evaporation, runoff, irrigation, and snow melt. Eventually, an increase in the soil moisture values could be primarily associated with a precipitation event, assuming that no irrigation activity is performed. Currently, several satellite missions provide surface soil moisture estimations at different spatial resolutions such as the Soil Moisture Active Passive "SMAP" (level 3: 36 km × 36 km, level 3 enhanced: 9 km × 9 km, and Level 2

SMAP/Sentinel-1: 1 km × 1 km) (Entekhabi et al., 2010), the Advanced Scatterometer "ASCAT" (level 2 with three spatial resolutions: 25 km ×25 km, 12.5 km × 12.5 km, and 1 km × 1 km) (Wagner et al., 2013), and the Soil Moisture and Ocean Salinity "SMOS" (SMOS INRA-CESBIO level 3: 25 km × 25 km) (Kerr et al., 2001), the Advanced Microwave Scanning Radiometer 2 "AMSR2" (Imaoka et al., 2012) (with two spatial resolutions $0.1^{\circ} \times 0.1^{\circ}$ and $0.25^{\circ} \times 0.25^{\circ}$), and the Chinese Fengyun-3B "FY3B" (spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$) (Shi et al., 2006).

Using soil moisture information, Pellarin et al. (2013) proposed a simple method to correct the TRMM and PERSIANN precipitation estimations at low spatial resolution ($\sim 0.25^{\circ} \times 0.25^{\circ}$). However, in the last three decades, numerous approaches have been developed to retrieve soil surface parameters using synthetic aperture radar (SAR) remote sensing (Aubert et al., 2013; Baghdadi et al., 2012a; Gao et al., 2017). From soil surface parameters, Sentinel-1 gives soil moisture estimations at very high spatial resolution (plot scale). In fact, with the arrival of Sentinel-1 (S1 SAR data at 10-m spatial resolution and a six-day revisit period) and Sentinel-2 (S2 optical data at 10-m spatial resolution and a five-day revisit period) satellites, an operational algorithm has been developed for soil moisture mapping over agricultural areas with six days' revisit time and high spatial resolutions (up to plot scale) (El Hajj et al., 2017).

The objective of this paper is to analyze the variation in surface soil moisture values obtained from the very high spatial resolution Sentinel-1 surface soil moisture maps (S1-SSM) during one year over the Occitanie region in the south of France. In this study, the variation of the SSM is analyzed only according to the variation of rainfall derived from GPM data. Although the water cycle components, such as evaporation rate and snow melt, play a critical role in SSM variation, these components were not considered in this study. This study will allow us to investigate the opportunity to detect heavy rainfall using Sentinel-1 surface soil moisture products obtained with a revisit time of six days at high spatial resolution in the south of France. After a detailed description of the study site and the data used in Section 2, Section 3 includes a comparison between S1-SSM and precipitation, as well as an analysis of the possibility of detecting heavy rainfall using S1-SSM with a revisit time of six days in the south of France (Occitanie region). A discussion is presented in Section 4, including the effect of S1 temporal resolution on detecting heavy rainfall and investigating the potential of SMAP soil moisture product with a revisit time of one day to detect heavy rainfall. Finally, Section 5 presents the main conclusions.

2. Study site and dataset description

2.1 Study site

The study site is the Occitanie region in the south of France (centered on 2˚30′ E and 43˚30′ N, Figure 1) covering an area of 72724 km^2. The region is formed out of 13 departments (Figure 1). Having a rich variety of landscapes, Occitanie is mainly covered by agricultural areas (especially wheat and corn) in the western part with a mix of crops, mountains, and forests in the eastern and southern parts. The region comprises different climatic zones. The eastern part is considered Mediterranean (about 700 mm of average annual precipitation in Hérault, Figure 1), while the western part is more humid (about 1200 mm of average annual precipitation in Gers, Figure 1).

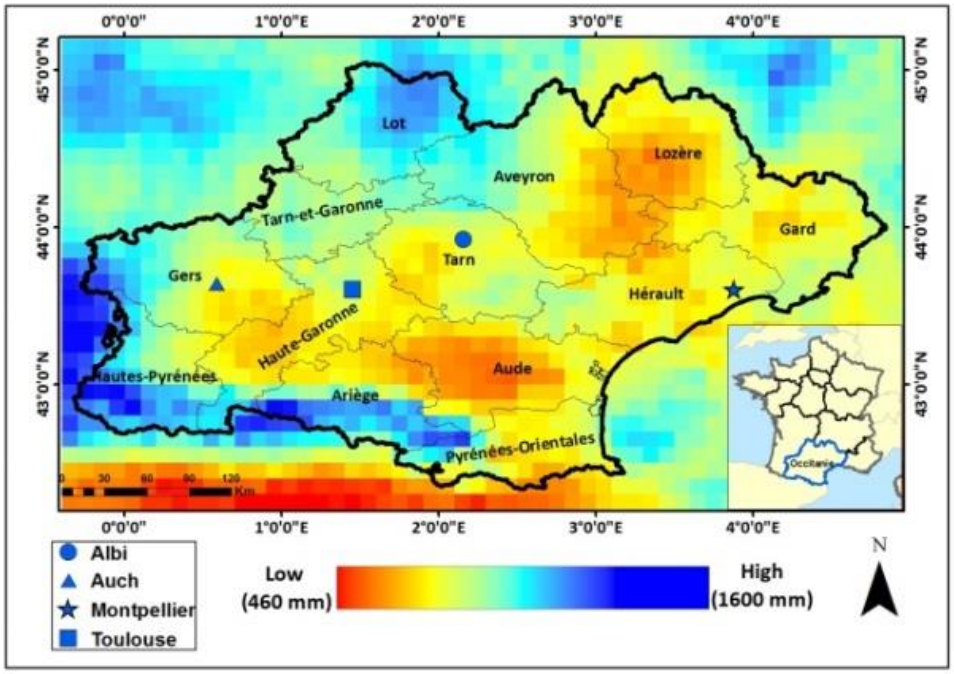


Figure 1: Annual cumulative precipitation calculated using Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG-GPM) data over the study site (Occitanie region, South France) between 1 September 2016 and 31 August 2017. Black lines represent the department limits according to French administrative nomenclature.

2.2 Dataset Description

2.2.1 S1-Derived Soil Moisture Maps

Sentinel-1 soil moisture maps (S1-SSM) at very high spatial resolution (up to a plot scale with a minimum area of 0.2 hectares) over our study area were analyzed. These soil moisture

maps were operationally produced by coupling SAR data from both the S1A and S1B satellite constellations operating at the C-band (wavelength ~6 cm) and optical data from both the S2A and S2B satellite constellations (El Hajj et al., 2017). To estimate the soil moisture over agricultural areas, El Hajj et al. (2017a) inverted the Water Cloud Model (WCM) (Attema and Ulaby, 1978) combined with the Integral Equation Model (IEM) (Baghdadi et al., 2006a). In their study, they used the WCM parameterized by Baghdadi et al. (Baghdadi et al., 2017) for the C-band. The total backscattering coefficient in the WCM is considered the sum of the direct vegetation contribution and the soil contribution multiplied by the attenuation factor. Using a vegetation descriptor derived from optical images (Sentinel-2), the direct contribution of vegetation can be calculated. El Hajj et al. (2017a) validated the proposed approach for the operational mapping of soil moisture over a study site in Occitanie in South France. They showed that the soil moisture in agricultural areas could be estimated with an accuracy of approximately 5 vol.%. Moreover, El Hajj et al. (2018a) evaluated SMOS, SMAP, ASCAT, and S1-SSM products (El Hajj et al., 2018a), and revealed that the S1-derived soil moisture maps provide the most accurate estimation of SSM. The higher accuracy of estimated SSM moisture is probably due to the well-calibrated IEM combined with the well-parameterized WCM and the use of high spatial resolution (10 m × 10 m) land cover maps derived from S2 images to eliminate SAR scattering from forest and urban areas. S1-SSM maps are produced for agricultural areas masking forest, urban, and high-slope areas (slope >20% calculated based on elevation data provided by the Shuttle Radar Topography Mission “SRTM” at 30 m x 30 m spatial resolution). S1 soil moisture maps for the Occitanie region are available as open access data via the Theia French Land Data Center (http://www.theia-land.fr/en/thematic-products). The SSM map at each date was generated using the ascending acquisition mode of the Sentinel-1 images (acquisition time ~17:40 TU). Four maps are required to cover the region on a single date. Thus, a total of 250 soil moisture maps were downloaded to cover the whole region on the studied dates.

2.2.2 IMERG GPM Products

Rainfall measurements with 30-minute revisit time that were obtained from the GPM (Global Precipitation Measurements) sensor were considered in this study. The GPM mission is an international satellite mission initiated by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA) to unify and advance global precipitation measurements from space (Skofronick-Jackson et al., 2018).

GPM provides several precipitation records at different scales by using different sensor combinations. From the GPM mission, the IMERG (Integrated Multi-satellite Retrievals for GPM) data product was selected to perform our analysis. IMERG data offer global precipitation estimations in both real-time and late mode. While real-time IMERG data are proposed for disaster monitoring and flood risk assessment, the higher quality late IMERG data are intended for meteorological and climatological applications. The spatial resolution of IMERG maps is $0.1^{\circ} \times 0.1^{\circ}$, which fully covers the globe between 60°N and 60°S. These data provide precipitation maps with several time intervals (30 minutes, 24 hours, three days…). The NASA GPM program established several campaigns to collect ground-based observations to validate GPM products (Houze et al., 2017). Omranian and Sharif (2018) validated the accuracy of GPM-IMERG data over the Colorado River basin in Texas, and found that the GPM-IMREG slightly underestimates the precipitation over the basin. Several studies have also reported good correlation between GPM-IMREG products and rain gauges with small under or overestimation depending on the study site, and better estimation when compared to other precipitation products (Dezfuli et al., 2017). IMERG data at different spatial and temporal resolutions are available via the NASA Precipitation Measurements Mission (PMM) website (https://pmm.nasa.gov/data-access). In this study, the fifth version (V05) of GPM-IMERG products was used.

3. Data Analysis

3.1 Comparison between S1-SSM and Precipitation

A first analysis of soil moisture variation according to precipitation events was performed. For a grid cell of the GPM product ($0.1^{\circ} \times 0.1^{\circ}$) over Montpellier (Figure 1), GPM daily cumulative precipitation records were obtained over one year between September 2016 and August 2017. Then, at each available S1-SSM map between 1 September 2016 and 31 August 2017, mean soil moisture values were calculated for the corresponding GPM grid cell. Figure 2 shows the temporal variation in the mean soil moisture values, the daily precipitation records, and the daily air temperature records (minimum and maximum daily air temperature). A global consistency is observed between the S1-SSM values and precipitation records. Figure 2 shows that after a precipitation event, high soil moisture values are estimated, whereas with the absence of precipitation over a period of time, the soil dries due to evaporation, and S1-SSM values generally decrease. For example, an increase in soil moisture values from 16 vol.% on

15 November 2016 to 25 vol.% on 21 November 2016 was observed due to 40-mm precipitation recorded on 20 November 2016. Also, soil moisture values increased from 19 vol.% on 9 December 2016 to 25 vol.% on 15 December 2016due to 17 mm of precipitation recorded on 15 December 2016. Moreover, the absence of precipitation along with evaporation (maximum temperature between 22 °C and 25 °C) between 1 April 2017 and 25 April 2017 caused soil moisture values to drop down below 15 vol.%, followed then by a sharp increase in soil moisture values on 26 April 2017 (24 vol.%) due to rainfall event occurring on 25 April 2017 (13 mm). On the other hand, 18 mm of precipitation on 4 March 2017 did not highly affect the soil moisture values on 9 March 2017 (18 vol.%) due to the five-day difference between the rainfall date and the following S1-SSM date. For the period between May and August, the increase of temperature (and consequently the evaporation) didn't cause a sharp decrease in SSM values. This is probably due to the existence of irrigation activities for summer crops. According to the six-day revisit time of the Sentinel-1 sensor, it is difficult to quantitatively analyze the correlation between the precipitation and the soil moisture, because this correlation also depends on other parameters such as the evaporation rate, the soil texture, and the previous state of soil before rainfall events. For this reason, the correlation coefficient R was only calculated between the average annual soil moisture for each department and the corresponding annual precipitation (Figure 1). A correlation coefficient R of 0.89 was obtained between the mean annual soil moisture at each department and the average annual cumulative precipitation of each department.

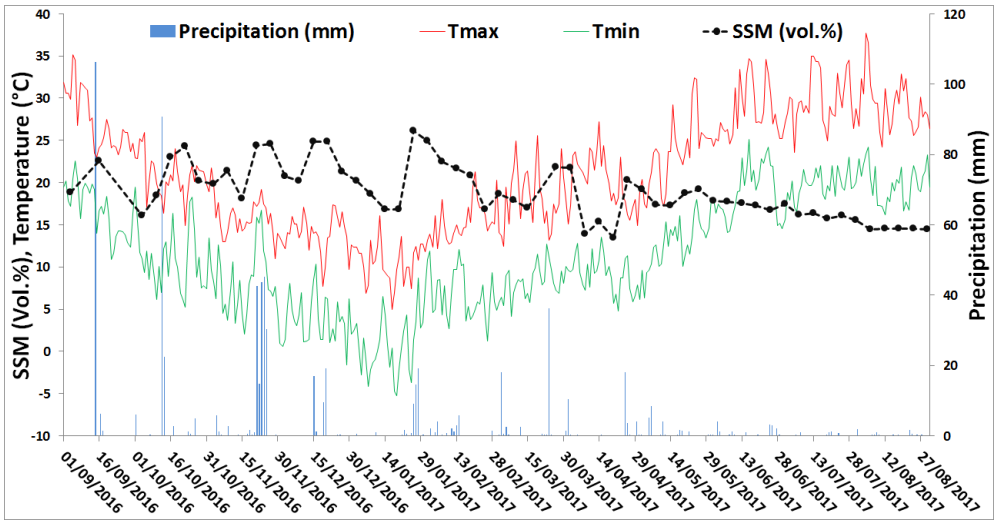


Figure 2: Temporal evolution of soil moisture values of Sentinel-1 derived surface soil moisture maps (S1-SSM) with daily precipitation records for a GPM grid cell over Montpellier ($0.1^{\circ} \times 0.1^{\circ}$), and daily temperature records (minimum and maximum temperature) at a local station in Montpellier, Occitanie region, France.

The temporal analysis of the S1-SSM values along with GPM data was then studied during the period between 21 December 2016 and 1 February 2017. Different climatological events were observed during this period: the absence of precipitation for 24 consecutive days in some parts of the region between 27 December 2016 and 25 January 2017 and rain episodes at several dates (Figure 2). The objective here is to analyze the temporal evolution of S1-derived SSM as a function of cumulative rainfall events (existence or absence of precipitation).

As mentioned in Section 2.2.1, an S1 soil moisture map can be obtained every six days (during the revisit of the S1 satellite). Thus, a total of eight soil moisture maps acquired between 21 December 2016 and 1 February 2017 were used. For these maps, soil moisture values were classified into four different intervals that were found to best describe the spatiotemporal variation in soil moisture values.

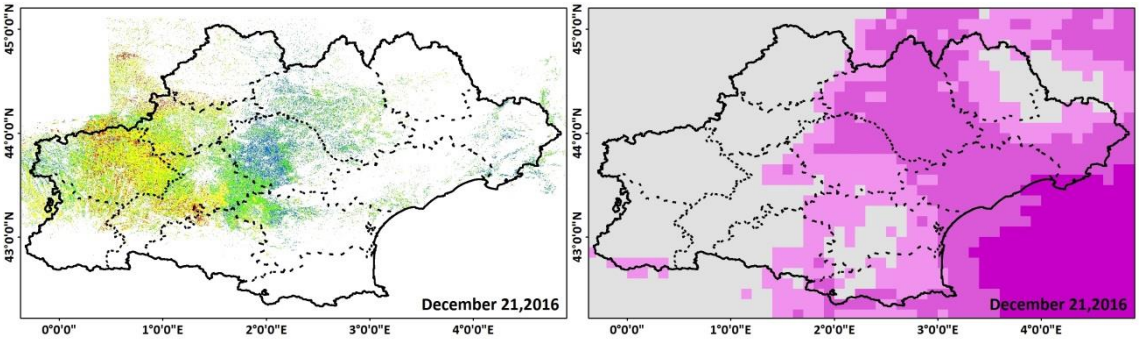
To perform the temporal evolution analysis, first, for each S1-SSM map date, one cumulative rainfall map for the six days prior to the S1-SSM map date was computed by summing the 30-minute time interval of the late IMERG-GPM product. Then, each S1-derived SSM map was overlaid with the corresponding six days of cumulative GPM rainfall. From the qualitative analysis of the overlaid maps, several arguments can be extracted and discussed. Starting with the initial map on 21 December 2016 (Figure 3a), departments as Tarn, Hérault, Lozère, and Aveyron of the eastern and middle parts of the region encountered high soil moisture values (more than 20 vol.%) corresponding to 10 to 20 mm of cumulative precipitation during the past six days, whereas Gers and Haute-Garonne in the western part showed low soil moisture values (20 vol.% at most) corresponding to the absence of precipitation in the area for the same prior period. A map derived six days later on 27 December 2016 (Figure 3b) shows a decrease in soil moisture values in the Hérault and Gard regions due to the absence of precipitation between 21 December 2016 and 27 December 2016. Moreover, high soil moisture values were estimated for Tarn due to continuous raining episodes between 21 December 2017 and 27 December 2017, and a slight increase in soil moisture values was observed in western part due to three mm of precipitation recorded one day prior to the S1-SSM map, according to the GPM data (Figure 3b). A general decrease in the soil moisture values throughout the region was observed between 27 December 2016 (Figure 3b) and 2 January 2017 (Figure 3c), as no precipitation was recorded. A continuous lack of precipitation for 12 days from 27 December 2016 until 8 January 2017 caused soil moisture values to drop to less than 15 vol.% on 8 January 2017 in the whole region, as shown in Figure 3d. Six days later on 14 January 2017 (Figure 3e), soil moisture values remained low at the east, as no precipitation was recorded for an additional

six days, whereas a moderate increase in soil moisture values was observed in the middle and western parts (Tarn, Haute-Garonne, and Gers), corresponding to precipitation between 8 January 2017 and 14 January 2017 (a cumulative rainfall increase of five to 20 mm). Twenty-four days with no precipitation recorded in the Hérault and Gard region between 27 December 2016 and 20 January 2017 caused the soil moisture values to drop more, with values less than 15 vol.% in this area on 20 January 2017 (Figure 3f).

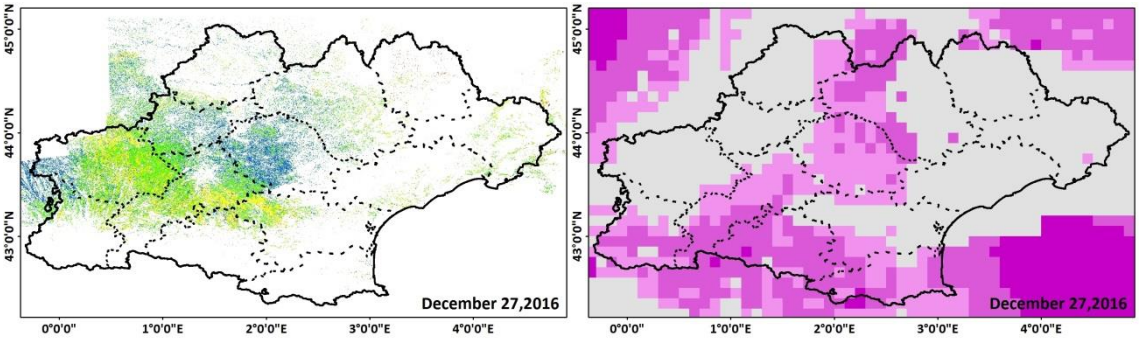
In addition, for the Gers department, low soil moisture values (less than 15 vol.%) were estimated on 20 January 2017 and high cumulative precipitation was recorded between 14 and 20 January 2017 (15 mm) (Figure 3f). This low soil moisture estimates could be linked to the presence of frozen soil in the area. In a recent study using Sentinel-1 SAR data, Baghdadi et al. (2018b) showed that a decrease of at least three dB in the radar backscattered signal can be observed over frozen soil conditions. Such a decrease in the SAR signal yields an underestimation of soil moisture in the S1-derived SSM maps. In our case, it seems that the low soil moisture values for this area on 20 January 2017, although with the presence of rainfall, were due to frozen soil conditions. To support our assumption, the temperature curve and precipitation records for a local meteorological station in Auch city located in Gers-Occitanie were analyzed (Figure 4). On 19 January 2017 (S1 acquisition over Gers on 19 January 2017 17:56 UTC), the temperature varied between –8 °C and 2 °C with a mean value of –3 °C throughout the day. Three days before (15 and 16 January), 15 mm of cumulative precipitation was recorded. Thus, a precipitation event followed by a decrease in the air temperature to less than 0 °C justifies the presence of frozen soil conditions as the cause of a decrease in the estimated soil moisture values in the S1-derived SSM maps.

After 29 days of dry conditions in the eastern part (Hérault and Gard between 27 December 2016 and 25 January 2017), a sudden increase in the soil moisture was observed on 26 January 2017 with the beginning of a rainfall event (Figure 3g). The mean S1-SSM value in this area increased from 13 vol.% on 20 January 2017 to 32 vol.% on 26 January 2017. On the other hand, the absence of precipitation in the remaining parts of the region between 20 January 2017 and 26 January 2017 caused the soil moisture to attain lower values (16 vol.%) on 26 January 2017.The slight increase in soil moisture values in Gers between 20 and 26 January is probably due to the disappearance of the frozen conditions observed in the previous acquisition (20 January 2017). The map on 1 February 2017 (Figure 3h) shows high soil moisture values for the eastern part, which was affected by continuous raining events, and a general decrease in the western part due to the absence of rainfall between 20 January 2017 and 1 February 2017. The

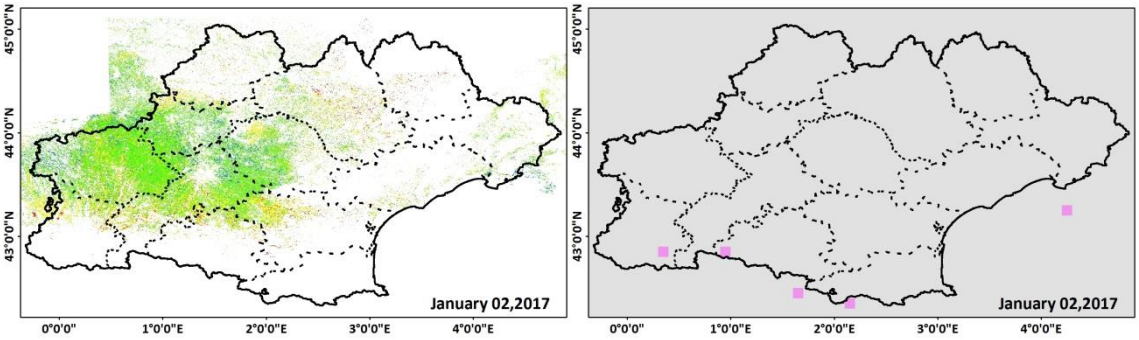
temporal analysis of S1-derived SSM values over a period of time shows the direct effect of raining episodes or dry conditions on S1-SSM values.



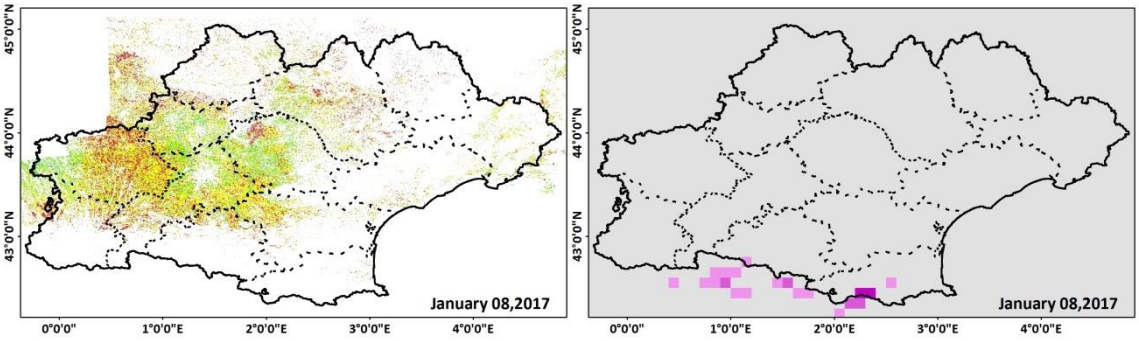
(**a**)



(**b**)



(**c**)



(**d**)

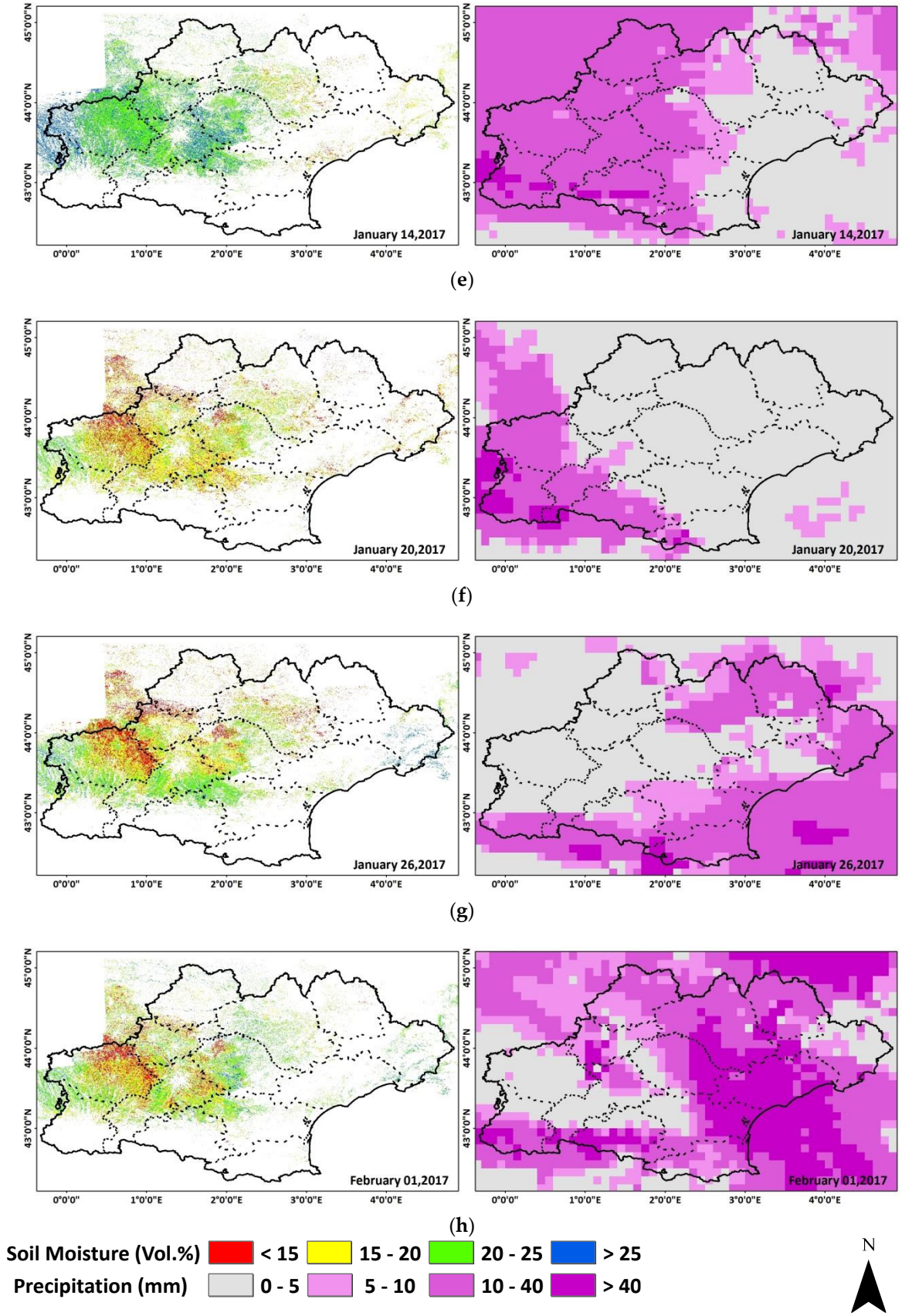


Figure 3: On the left, an S1-derived SSM map. On the right, the corresponding six days of GPM cumulative precipitation data. (a) 21 December 2016, (b) 27 December 2016, (c) 2 January 2017, (d) 8

January 2017, (e) 14 January 2017, (f) 20 January 2017, (g) 26 January 2017, and (h) 1 February 2017. Departments are represented by dashed lines.

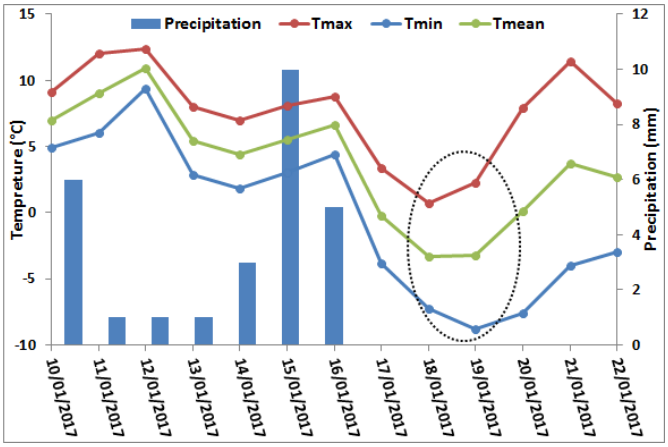


Figure 4: Temporal evolution of temperature and precipitation data recorded at a local meteorological station in Auch, Occitanie region.

3.2 Effect of S1-SSM Temporal Resolution

The behavior of soil moisture values following rainfall events was then studied as a function of the time lag between the rainfall date and the S1-SSM date. In fact, soil moisture values are mainly affected by precipitation events, evaporation, runoff, and snow melt. In this analysis, only the effect of precipitation was studied. To study the effect of the time difference between the date of S1-SSM maps and the date of rainfall events, S1-SSM values were plotted as a function of the recorded six days of cumulative precipitation values (Figure 5). For the $0.1^{\circ} \times 0.1^{\circ}$ GPM grid, the mean S1-SSM values were calculated from each available S1-SSM map. The cumulative precipitation was computed for six days, because the revisit time of derived S1-SSM maps is for six days over our study site. The results showed that for low cumulative precipitation for six days preceding the S1-SSM map date (for example, less than 20 mm), the soil moisture values are distributed on a wide range between 10–25 vol.% (Figure 5). This wide range of S1-SSM values was expected, since low precipitation would induce high S1-SSM values if the precipitation occurs in the 24 hours preceding S1 acquisition, whereas low precipitation would not induce an increase of S1-SSM if it occurs a couple of days before S1 acquisition. As the cumulative precipitation for six days is high (for example 100 mm, Figure 5), the S1-SSM values become more homogenous and high. Indeed, high cumulative precipitation for six days (for example, higher than 100 mm) induces high S1-SSM values, even if the precipitation occurs six days before S1 acquisition, in particular during winter, and the evaporation rate is low.

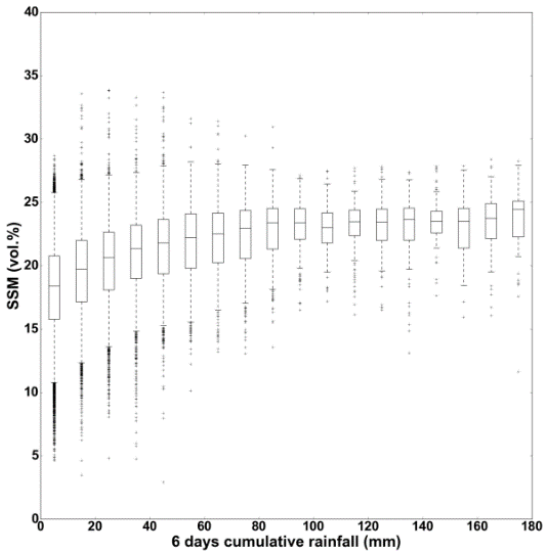


Figure 5: S1-SSM distribution as a function of six days of cumulative precipitation (revisit time of S1-SSM = six days).

To illustrate the results presented in Figure 5, three cases of rainfall events with different time intervals between the S1-SSM map's date and the rainfall date were studied for Gard Department of Occitanie. First, for the S1-SSM map of 16 October 2016, the latest rainfall event before 16 October 2016 took place on 12 October 2016 with an average cumulative precipitation of 100 mm in 24 hours (Figure 6a.) Then, between 12 October 2016 and 16 October 2016, no precipitation was recorded. The soil moisture map of 16 October 2017 shows wet soil conditions with S1-SSM values between 20–25 vol.%. On the other S1-SSM map acquired on 26 January 2017, a rainfall event occurred on the same date a few hours before the S1 acquisition date (26 January 2017) with 20 mm of cumulative precipitation (Figure 6b). Consequently, the soil moisture attained very high values: higher than 25 vol.%. The department didn't encounter any rainfall event for 29 days prior to 26 January 2017 (the soil moisture map that was acquired six days before 26 January 2017 showed soil moisture values of 13 vol.%). Similar soil moisture values were recorded on the S1-SSM map of 27 January 2018 (Figure 6c), where the department was affected by a light rain event 24 hours before the S1-SSM date (nine mm). Wet soil moisture values higher than 25 vol.% were estimated over the department. Prior to this rainfall event on 26 January 2018, no events were detected for the previous six days, and the soil moisture values six days before presented dry values (15 vol.%). This result assures that the detection of rainfall events on S1-SSM is highly dependent on the S1-SSM revisit time. Indeed, a few mm of precipitation just 24 hours before the S1-SSM map shows higher soil moisture values (for example on 27 January 2018) than an extreme event rainfall event (for example on 12 October 2016) four days before the S1-SSM map (16 October 2016). Thus, the six days of temporal resolution of the S1-SSM map doesn't always permit the

detection of an extreme rainfall event. For example, confusion will appear between high values due to extreme rainfall events occurring six days before the S1-SSM acquisition and high values due to light or moderate rain a few hours before the S1 acquisition date.

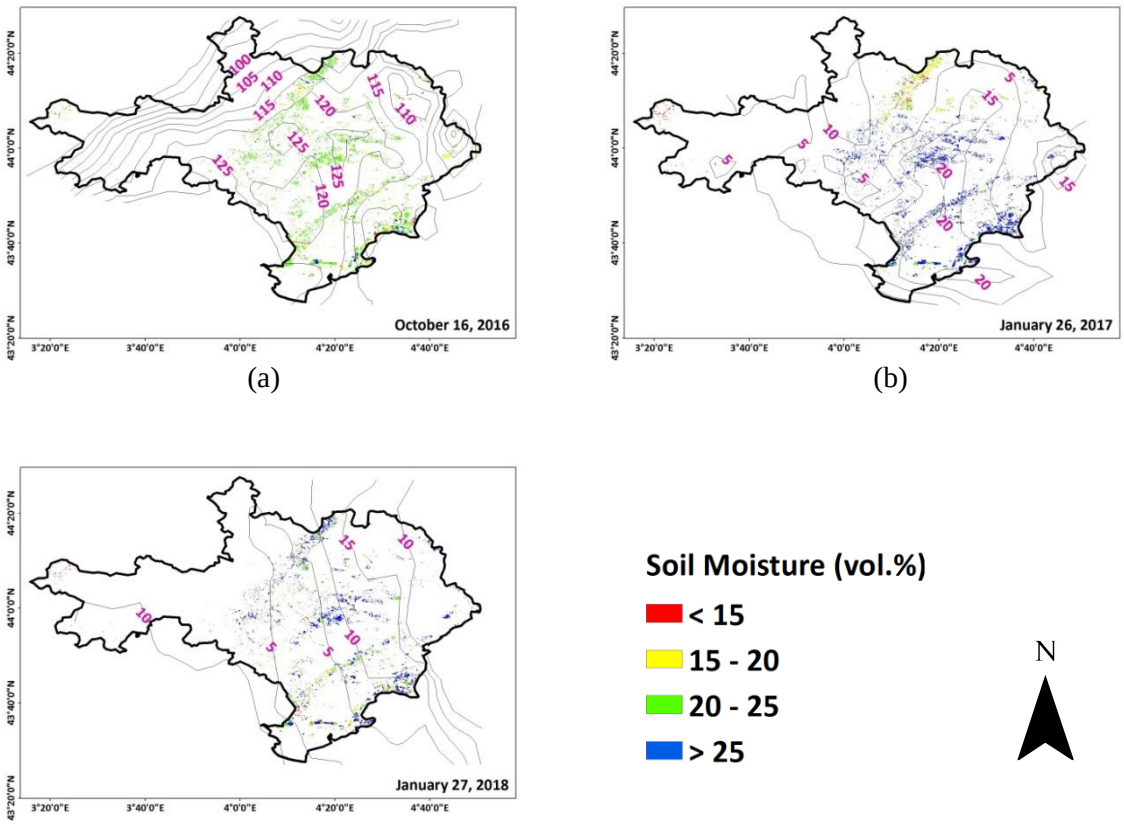


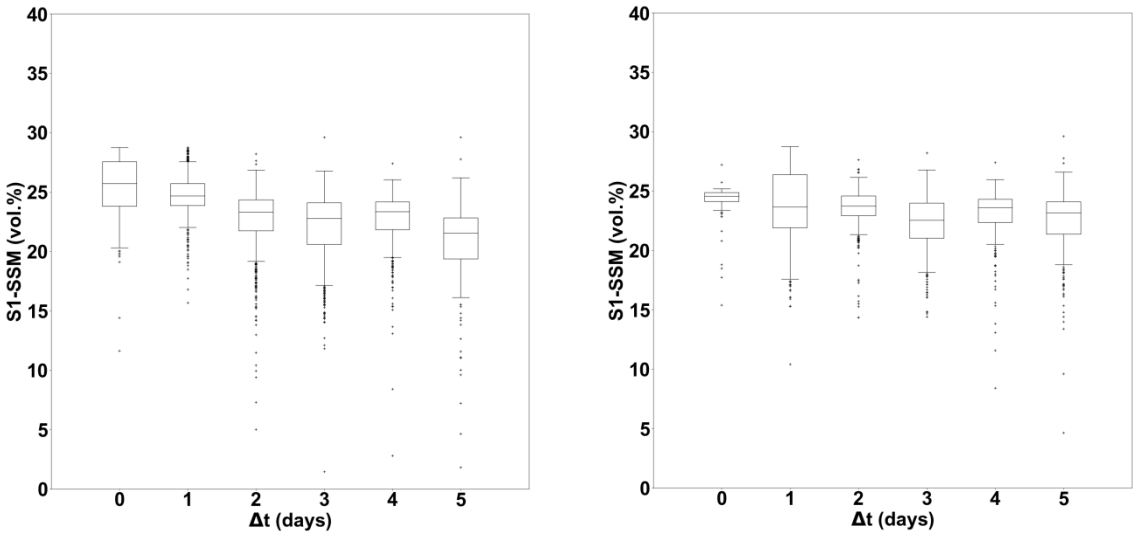
Figure 6: S1-derived SSM maps overlaid with GPM precipitation data. (a) 16 October 2016, (b) 26 January 2017, and (c) 27 January 2018. Contour lines represent precipitation amounts in mm.

4. Discussion

The results obtained in Section 3 show that the detection of heavy rainfall using S1-SSM values is highly dependent on the time lag between the heavy rainfall event and the S1-SSM acquisition. Figure 5 showed that, on average, SSM values varied between 15 vol.% for five mm of accumulated precipitation in six days to about 22 vol.% for accumulated precipitation greater than 100 mm in six days. In addition, when the accumulated precipitation was less than 30 mm, 75% of the pixels (0.1° × 0.1°) showed SSM values lower than 22 vol.%, and the remaining 25% had SSM values between 20–28 vol.%. On the other hand, when the accumulated precipitation is higher than 100 mm in six days, 75% of the pixels have SSM values higher than 22 vol.%, while the remaining 25% of the pixels have SSM values between 20–22 vol.%. Moreover, for six days of cumulative precipitation between 30–100 mm, the SSM

values varied between 15–28 vol.%. Thus, the results show that low rainfall of less than 30 mm in six days and very heavy rainfall greater than 100 mm in six days generally result in different soil moisture conditions in approximately 75% of the cases. However, it remains difficult to distinguish between high SSM values (around 28 vol.%) corresponding to (1) light rain the day before the S1 acquisition, (2) heavy rainfall a few days before the radar acquisition, or (3) various light rainy episodes during several days preceding the radar acquisition. Therefore, the SSM maps with a revisit period of six days do not always allow us to discriminate wet soil following a light rain at a date very close to the date of acquisition of the S1 image, and wet soil following a heavy rain that took place several days before the S1 acquisition date.

Moreover, we have analyzed the impact of heavy rainfall (cumulative rainfall more than 60 mm in 24 hours or more than 80 mm in 48 hours) on S1-SSM values as a function of the number of days that separated the heavy rainfall event and the S1 acquisition date ($\Delta t = \mathrm{t}_{S1} - \mathrm{t}_{Heavy\ Rainfall}$). These threshold values of 60 mm and 80 mm were selected according to Meteo-France recommendations (http://pluiesextremes.meteo.fr/france-metropole/Statistiques-et-records.html). Figure 7 shows the SSM values as a function of $\boldsymbol{\Delta t}$ for cumulative rainfall greater than 60 mm in 24 hours (Figure 7a) and 80 mm in 48 hours (Figure 7b). For example, $\boldsymbol{\Delta t = 0}$ corresponds to a heavy rainfall event that occurred during the past 24 hours (or 48 hours) of the S1 acquisition. In the case of a cumulative rainfall greater than 60 mm in 24 hours (Figure 7a), we find that the SSM average values decreases from 25.4 vol.% for $\boldsymbol{\Delta t = 0}$ to 20.8 vol.% for $\boldsymbol{\Delta t = 5}$ days (S1 image acquired six days after the heavy rainfall event). In the case where the cumulative rainfall is greater than 80 mm in 48 hours (Figure 7b), we find that SSM average values are similar (23 vol.%) after a heavy rainfall that took place during the past 48 hours of the S1 acquisition or after a heavy rainfall during the 48 hours that took place six days before the S1 acquisition.



(a) (b)

Figure 7: S1-SSM distribution as a function of the number of days that separate the heavy rainfall event and the acquisition date of S1. (a) Heavy rainfall with 60 mm of cumulative precipitation in 24 hours, (b) heavy rainfall with 80 mm of cumulative precipitation in 48 hours

The effect of the temporal resolution of satellite images on detecting heavy rainfall events was further studied with soil moisture data from SMAP at higher temporal resolution (Entekhabi et al., 2010). For this reason, we have re-established Figure 5 with SMAP data at 9 km × 9 km spatial resolution (Figure 8). SMAP data were processed for the period between September 2016 and August 2017, over the Occitanie region, with one day of temporal resolution. Figure 8 shows the distribution of SMAP soil moisture values as a function of one day of cumulative precipitation obtained from the IMERG-GPM data. In average, SMAP shows an estimation of 15 vol.% following light rain one day before the SMAP acquisition, and about 30 vol.% following 100 mm of cumulative rainfall during one day before the SMAP acquisition. Although the difference in soil moisture between the two events (light rain and heavy rain) is greater for SMAP (about 15 vol.%) than that of S1 (7 vol.%), the results obtained with SMAP at 9 km × 9 km and one-day revisit do not also allow the easier detection of heavy rainfall events. However, the potential of the SMAP in detecting heavy rainfall appears to be better than the potential of S1 for rainfall events >150 mm, where SMAP-SSM shows values greater than 30 vol.%.

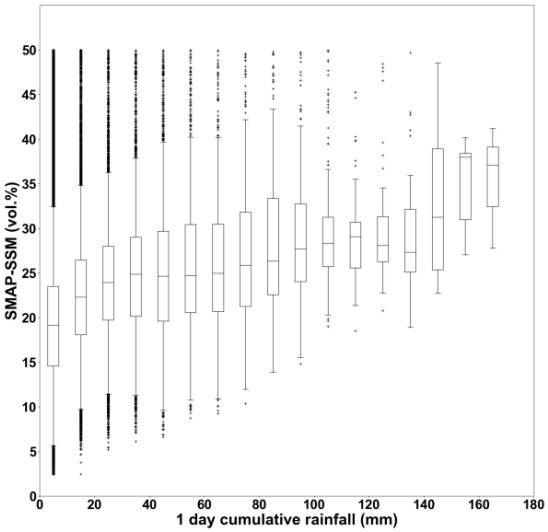


Figure 8: Soil Moisture Active Passive (SMAP)-SSM distribution as a function of one day of cumulative precipitation (revisit time of SMAP-SSM = one day)

The previous discussion leads to the conclusion that soil moisture, which is obviously correlated with the amount of rainfall, depends also on other water cycle parameters such as evaporation, soil type, and the previous state of soil moisture before the heavy rainfall. Although it is obvious that heavy rainfall will lead to very wet soil, a slight rain over already wet soil will also result in extremely wet soil. A revisit time of several days (case of S1) makes the detection of heavy rainfall more complicated, because an extreme rainfall event that takes place four or five days before the S1 acquisition and is then followed by a significant evaporation will not lead to very high soil moisture values. In addition, radar sensors don't correctly estimate soil moisture for SSM values above 30 to 35 vol.%. Baghdadi et al. (2008) demonstrated that the radar signal is stable when the SSM values ranged between 30–35 vol.%, and it decreases beyond this threshold. This eventually leads to an underestimation of SSM beyond this threshold value (Baghdadi et al., 2012b) Thus, the configurations for having very wet soil are numerous, and do not always depend on the amount of rain that falls. That many environmental parameters affect the value of SSM and the limitation of SAR to estimate the very high SSM values makes the monitoring of extreme rain events very delicate when it is based only on soil moisture maps. Therefore, the analysis carried out in this paper confirms that the use of the S1-SSM maps with a revisit period of six days (at plot scale) or the SMAP-SSM maps with very high temporal resolution of one day (at 9 km × 9 km) makes the detection of heavy rainfall not always obvious.

5. Conclusion

This letter presents an analysis of Sentinel-1 derived surface soil moisture maps (S1-SSM) produced with high spatial resolution (at plot scale) and a revisit time of six days for the Occitanie region located in the south of France as a function of precipitation data derived from the GPM mission, in order to investigate the potential of S1-SSM maps in detecting heavy rainfalls. First, we investigated the temporal evolution of the soil moisture according to precipitation records derived from the GPM data. A general coherence was observed between the soil moisture estimations and precipitation records over the study site. The S1-SSM values increase following rainfall events and decrease after a period without rainfall due to evaporation. Then, we studied the behavior of soil moisture values following rainfall events in order to explore the possibility of detecting heavy rainfall using the revisit time of six days of the S1-SSM product. The influence of heavy rainfall over 24 and 48 hours (precipitation accumulation of 60 mm and 80 mm, respectively) on the S1-SSM values was analyzed as a function of the time lag between the rainfall event and the S1 acquisition. The results show that similar soil moisture conditions could be obtained on the S1-SSM maps that are acquired after an intensive rainfall event a couple of days prior to the S1-SSM maps, and the S1-SSM maps acquired after light rain (few mm) 24 hours before the S1-SSM maps. Finally, we studied the effect of the temporal resolution of satellite images on detecting heavy rainfall events using the SMAP soil moisture product at one-day temporal resolution. The results show that despite the difference in soil moisture values in SMAP data after heavy and light rainfall events, the one-day revisit time does not also allow an easier detection of heavy rainfall events. Therefore, the detection of heavy rainfall is not always obvious, whether using the S1-SSM maps with a revisit period of six days (at plot scale) or the SMAP-SSM maps with a very high temporal resolution of one day (at 9 km × 9 km) since many environmental parameters (evaporation, soil texture…) could also affect the value of SSM.

Author Contributions: H.B., N.B., and M.E.H. conceived and designed the experiments; H.B. and N.B. performed the experiments; H.B., M.E.H., N.B., and M.Z. analyzed the results; H.B. wrote the letter; and H.B., N.B., M.E.H., and M.Z. revised the paper.

Funding: This research received funding from the French Space Study Center (CNES), the National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA), and the European Space Agency (ESA).

Acknowledgment: The authors thank the French Space Agency (CNES, TOSCA 2018 project, POMME-V) and the National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA) for supporting this work.

Conflicts of Interest: The authors declare no conflict of interest.

CHAPTER III: SUPERVISED CLASSIFICATION APPROACHES FOR MAPPING IRRIGATED AREAS USING SENTINEL-1 AND SENTINEL-2 DATA

1. Analytical Summary

1.1 Overview

The new satellite constellations SAR Sentinel-1 (S1) and optical Sentinel-2 (S2) has encouraged researchers to evaluate their potential for mapping irrigated areas at high spatial resolution. Until now, only few studies have reported the use of these two high-resolution satellites for irrigation mapping (Demarez et al., 2019; Gao et al., 2018). For irrigation mapping, optical data from S2 can provide the difference in vegetation indices between irrigated and non-irrigated plots while SAR (Synthetic Aperture Radar) data can offer a 6 days temporal monitoring of the soil water status; particularly soil moisture.

In this chapter, two complementary objectives are presented. The first objective is to study the potential of S1 and/or S2 data for mapping irrigated areas using supervised classification models such as the random forest (RF) and the convolutional neural networks (CNN). This study has been performed over Catalonia region of northeast Spain characterized by a semi-arid climatic condition. Since supervised classification models are highly dependent on the studied region, it is of great importance to resolve this region dependency of the developed classification model. Thus, the second main objective is to propose a framework capable of transferring the supervised classification model from one region to other region with different climatic properties. This transfer has been applied on the model developed on Catalonia to map irrigated areas in a humid climatic watershed in southwest France (Adour Amont watershed). Figure B presents a general graphical abstract summarizing the two principle objectives.

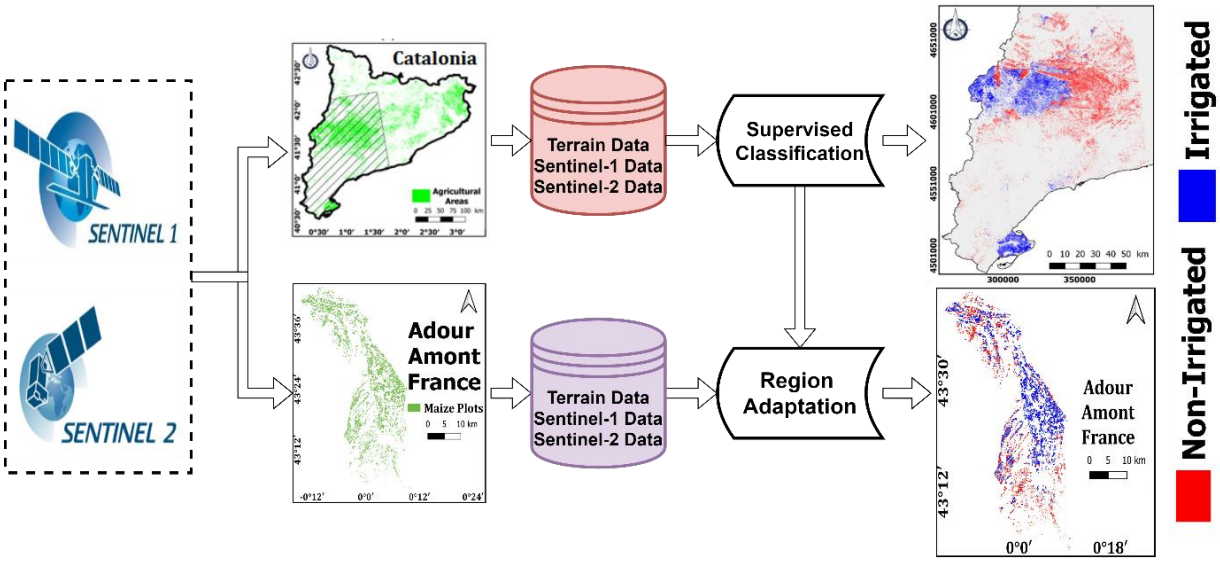


Figure B. Graphical abstract representing a general overview of supervised classification approaches for mapping irrigated areas on two different regions (Catalonia and Adour Amont watershed).

1.2 Materials

1.2.1 Study sites

Two study sites are examined for mapping irrigated areas using S1/2 data and supervised classification approaches. The first site is the Catalonia region located along the northeastern Spanish Mediterranean coast. The climate in the region is typically Mediterranean (semi-arid); mild in winter and dry in summer. A huge database of labelled irrigated and non-irrigated plots (~193,000 plots) is available over the region thanks to the SIGPAC (Geographic Information System for Agricultural Parcels) data. SIGPAC provides the plots' boundaries and textual data for each plot including the crop type and existence or absence of irrigation. On the other hand, the Adour Amont basin is located in the Occitaine region of southwest France near the city of Tarbes. The basin is characterized by temperate climate (humid to oceanic) with humid summer season usually encountering several rainfall events. In Adour Amont, less number of terrain data is available including 300 non-irrigated plots and 151 irrigated plots. The terrain data over Adour Amont was collected by terrain campaigns for the year 2017.

1.2.2 Remote sensing data

Over the Catalonia region, 82 C-band (5.405 GHz) SAR images acquired by S1A and S1B satellites were used for the period between September 2017 and December 2018 corresponding to the irrigation information of the used SIGPAC data. Over Adour Amont basin, the same

number of images (82 images) was collected for the period between September 2016 and December 2017 (terrain data collected in summer 2017). The 164 images (82 for Catalonia and 82 for Adour Amont) are acquired in both VV and VH polarizations. The Sentinel-1 toolbox (S1TBX) developed by the European Spatial Agency (ESA) was used to calibrate the S1 images radio-metrically and geometrically. In addition, 17 cloud free S2 images (level 2A) were used for each study site covering the same period as the S1 images at each site.

1.3 Methodology

1.3.1 Irrigated area mapping in Catalonia

The spatial resolution of irrigation mapping is the agricultural plot scale. First, the average S1 SAR backscattering coefficients (σ°) were calculated at both plot scale (SIGPAC plots) and grid scale (10 km × 10 km). The mean signal at grid scale is mainly used to discriminate between rainfall and irrigation events. Based on the results of chapter 2, we suppose that if the mean SAR signal increases at the base of 10 km × 10 km (global increase of soil moisture at grid cell) then a rainfall event most probably occurred. Next, the principal component analysis (PCA) and the wavelet transformation (WT) were applied on the obtained S1 time series at plot and grid scales. In addition, the PCA was applied on the NDVI temporal series at plot scale calculated using the S2 optical images. Finally, two classification approaches including the RF classifier and the CNN were examined. In each approach, three scenarios have been tested. The first scenario is based on the use of SAR data only, the second scenario considers the use of optical data only, and the third scenario includes the combined use of both optical and SAR data.

1.3.2 Transfer learning for mapping irrigated areas in Adour Amont

Among the several scenarios tested over Catalonia, the CNN built using S1 time series provided the best classification accuracy (94%). However, applying exactly the same model (built on Catalonia) on Adour Amont basin produces unsatisfactory results. The challenge thus is to transfer the supervised classification model from a source geographic area (Catalonia) to a target geographic area (Adour Amont). The transfer learning is based on “distilling before refine” strategy. The CNN model trained in Catalonia (source geographical area) is characterized by large volume of labelled samples (from SIGPAC data). This model is denoted as the teacher model. A lighter model is then distilled from the teacher model following the

distillation procedure developed by Hinton et al. (2015). This lighter model (denoted as the student model) contains less number of parameters. The student model is then refined by data terrain data from the Adour Amont basin. Then irrigation mapping was performed using the refined student model. The proposed strategy is compared to different approaches including RF and CNN directly trained using the collected terrain data.

1.4 Main results

Using the PC dimensions and WT coefficients of the S1 and S2 temporal series, irrigated and non-irrigated plots were separated. Over Catalonia, irrigated areas were first classified in two different random forest classifiers one using the PC dimensions and the other using the WT coefficients. The results show a good overall accuracy (OA) for the RF classifier using the PC values (90.7%) and the WT coefficients (89.1%). Moreover, the results showed that the RF classifier using optical data (NDVI) performs well with OA = 89.5%. The combined use of optical and SAR data (in PC values) slightly improved the classification accuracy (OA = 92.3%). The results of the validation of CNN approach, applied directly on the S1 time series, showed very high accuracy (OA = 94.1%) compared to the RF classifiers. When using the CNN, the accuracy of the irrigated class increased thus allowing better detection of irrigated plots. The use of NDVI data only in a CNN classifier produced lower overall accuracy (91.5%) than that obtained using only S1 data. However, the combined use of both S1 and S2 data in the CNN did not significantly improve the accuracy. The results thus validate the capability of mapping irrigated areas with high accuracy and at high spatial resolution using either S1, S2 or coupled S1 and S2 data.

Over Adour Amont Basin, we observed that the proposed framework (distilling before refine) obtains best results regarding both Accuracy and F-Measure when compared to other alternative approaches. The overall accuracy and the F-Measure for irrigation mapping in Adour Amont reaches 83% using the proposed transfer-learning framework. This good accuracy supports our distill before refine framework in which a smaller network distilled from a bigger one will be easier to adapt to a target area characterized by a very limited number of labeled samples. This result constitutes a preliminary experiment to perform transfer learning from a source dataset (a geographical area) to a target dataset (another geographical area).

1.5 Conclusion

The results obtained in this chapter demonstrates the capability of obtaining high-resolution irrigation mapping using S1 and S2 data. However, a supervised machine-learning classifier built on a zone and applied on a second zone does not lead to good accuracy in the second zone. Moreover, applying the same classifier on the same zone but at different year could also lead to lower accuracies. To build supervised classification models for irrigation mapping, terrain in situ data are always required. Therefore, the availability of terrain data remains an obstacle for performing irrigation mapping using supervised classifiers. As discussed in the thesis problematic, obtaining terrain data of irrigated and non-irrigated plots at yearly basis is difficult, time consuming and expensive. To handle the terrain data availability problem, our the next chapter will focus on developing operational semi-supervised approach to map irrigated/non irrigated plots that does not require in situ reference data. The semi-supervised approach will be capable of automatically generating its own reference data before building the supervised classification model. The development of the operational semi-supervised approach, that does not require in situ reference data, will allow mapping irrigated areas regardless of the studied year, geographic region, and climatic properties.

Although huge effort has been performed to map the extent of irrigated areas, the timing and frequency of irrigation has not yet received important attention, despite their high importance in managing water resources. Therefore, the next chapter will provide a detailed methodology for detecting irrigation events at plot scale using S1, S2 and S^2MP data.

2. Article 3: Mapping Irrigated Areas Using Sentinel-1 Time Series in Catalonia, Spain

Author version of the article published in Remote Sensing Journal: Volume: 11, Issue: 15, August 2019

Hassan Bazzi[1], Nicolas Baghdadi[1], Dino Ienco[1], Mohammad EL Hajj[1], Mehrez Zribi[2], Hatem Belhouchette[3], Maria Jose Escorihuela[4] and Valerie Demarez[2]

[1] INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

[3] CIHEAM-IAMM, UMR-System, 34090 Montpellier, France

[4] isardSAT, Parc Tecnològic Barcelona Activa, Carrer de Marie Curie, 8, 08042 Barcelona, Catalunya, Spain

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 1 July 2019; Accepted: 4 August 2019; Published: 6 August 2019

Abstract

Mapping irrigated plots is essential for better water resource management. Today, the free and open access Sentinel-1 (S1) and Sentinel-2 (S2) data with high revisit time offers a powerful tool for irrigation mapping at plot scale. Up to date, few studies have used S1 and S2 data to provide approaches for mapping irrigated plots. This study proposes a method to map irrigated plots using S1 SAR (synthetic aperture radar) time series. First, a dense temporal series of S1 backscattering coefficients were obtained at plot scale in VV (Vertical-Vertical) and VH (Vertical-Horizontal) polarizations over a study site located in Catalonia, Spain. In order to remove the ambiguity between rainfall and irrigation events, the S1 signal obtained at plot scale was used conjointly to S1 signal obtained at a grid scale (10 km × 10 km). Later, two mathematical transformations, including the principal component analysis (PCA) and the wavelet transformation (WT), were applied to the several SAR temporal series obtained in both VV and VH polarization. Irrigated areas were then classified using the principal component (PC) dimensions and the WT coefficients in two different random forest (RF) classifiers. Another classification approach using one dimensional convolutional neural network (CNN)

was also performed on the obtained S1 temporal series. The results derived from the RF classifiers with S1 data show high overall accuracy using the PC values (90.7%) and the WT coefficients (89.1%). By applying the CNN approach on SAR data, a significant overall accuracy of 94.1% was obtained. The potential of optical images to map irrigated areas by the mean of a normalized differential vegetation index (NDVI) temporal series was also tested in this study in both the RF and the CNN approaches. The overall accuracy obtained using the NDVI in RF classifier reached 89.5% while that in the CNN reached 91.6%. The combined use of optical and radar data slightly enhanced the classification in the RF classifier but did not significantly change the accuracy obtained in the CNN approach using S1 data.

Keywords: irrigation; Sentinel-1; principal components; wavelet transformation; random forest; deep learning; convolutional neural network; Spain

1. Introduction

Under changing climatic conditions, irrigation plays a significant role in agricultural production in order to meet the global food requirement. In fact, better planning of irrigation is required to fulfil the high demand of food with the increase in the global population (Godfray et al., 2010; Tilman and Clark, 2015). Moreover, irrigated agriculture accounts now for more than 80% of water withdrawn from rivers, lakes, and groundwater aquifers being the principal consumer of fresh water resources (Cai and Rosegrant, 2002; Shiklomanov, 2000; Thenkabail et al., 2009b). Therefore, more focus is being set on the assessment of irrigation performance for improving water management in order to achieve higher water productivity and increase the agricultural water sustainability.

Accurate information on the irrigated area extent helps in managing water resources that affect global food security and in analyzing the impact of climate change on the irrigation water requirement. However, the extent and distribution of irrigated areas remain indefinite in spite of the significant impact of irrigation on food security and water resources. At a global scale, several irrigation products are available at different spatial resolutions. Existing irrigation extent maps such as the FAO (Food and Agriculture Organization of the United Nations) database (Ozdogan et al., 2010), the Global Map of Irrigated Areas version 5 (GMIA 5.0) (Siebert et al., 2005, 2015) and the MIRCA2000 product (Portmann et al., 2010) have been primary derived from national statistical data. The Global Rain-Fed Irrigated and Paddy Croplands (GRIPC) with spatial resolution of 500 m was produced using MODIS (Moderate

resolution imaging spectroradiometer) satellite data (Salmon et al., 2015). The low spatial resolution of these products is an obstacle for using them in irrigation management especially in regards to small to medium agricultural areas.

At a regional scale, several efforts have been made to map irrigated areas using remote sensing data including both optical and radar data (Boken et al., 2004; Thenkabail et al., 2005; Kamthonkiat et al., 2005; Biggs et al., 2006; Gumma et al., 2011; Dheeravath et al., 2010). Most studies that use optical data demonstrate that irrigated/non-irrigated areas can be classified based on the difference in temporal signals of vegetation indices such as NDVI (Normalized Differential Vegetation Index), NDWI (Normalized Differential Water Index) and GI (Greenness index) between irrigated and non-irrigated areas. Among optical sensors, MODIS and Landsat have been extensively used to map irrigated areas (Biggs et al., 2006; Dheeravath et al., 2010; Kamthonkiat et al., 2005; Thenkabail et al., 2005). For example, Pervez and Brown, 2010 merged the NDVI derived from MODIS at 250 m spatial resolution and used statistical data to map irrigated areas of United States of America in 2002. In their study, they assumed that irrigated crops have a higher annual NDVI peak than non-irrigated crops especially under non-optimal precipitation conditions such as drought. Chen et al. (2018) proposed a new approach to map irrigation extent, time and frequency in an arid region located in Hexi Corridor of northwest China by merging the 30 m spatial resolution Landsat images with 250 m MODIS data and ancillary data. In their study, they used the GI to detect irrigation events during the first half of the growing season. The overall accuracy for detecting irrigation water supplements reached 87%. Recently, Xiang et al. (2019) mapped irrigated areas of northeast China by comparing the MODIS derived LSWI (Land Surface Water Index) of the agricultural areas to the surrounding natural vegetation such as forests. They assumed that the canopy moisture indicated by the LSWI is higher for irrigated crops than the adjacent forest. Thus, they mapped irrigated areas by fixing a threshold for the difference between LSWI of agricultural plots and that of the forest. However, the overall accuracy of the produced map did not exceed 77.2% with a 0.49 kappa coefficient.

Apart from optical data, recent studies have started exploiting the potential of SAR (synthetic aperture radar) data to map different agricultural irrigated areas (Bazzi et al., 2019d; Bousbih et al., 2018; Demarez et al., 2019; Fieuzal et al., 2011; Gao et al., 2018). The assumption established for using SAR data relies on the fact that radar remote sensing is sensitive to the water content of soil due to the increase in the dielectric constant with the increase of the soil water content. In fact, several studies have reported that the SAR signal is

highly affected by soil moisture content (Aubert et al., 2013; Baghdadi et al., 2011a, 2011b, 2016a; El Hajj et al., 2017; Hajj et al., 2014). For example, Hajj et al. (2014) analyzed the sensitivity of radar signals in the X-band over irrigated grassland conditions using time series of TerraSAR-X and Cosmo-SkyMed X-band SAR data. Their results showed that the radar signal in X-band increased by approximately 1.4 dB due to irrigation events occurring one day before the radar acquisition. They concluded that the radar signal could be used to identify three-day-old irrigated plots.

The recently launched Sentinel-1 (S1) SAR satellite offers a powerful tool for agricultural area classification and monitoring under various weather conditions. The S1 satellites (S1A and S1B) provide an exceptional combination of high spatial and high temporal resolutions for dual polarization SAR data (six days of temporal resolution and a 10 m × 10 m pixel spacing) available in free open access mode. S1 time series has been newly exploited by several studies for land cover classification (Bazzi et al., 2019d; Carrasco et al., 2019; Feng et al., 2019; Ndikumana et al., 2018). High quality classification mapping has been produced by applying either classical machine learning, such as random forest and the support vector machine, or complex deep learning methods on S1 time series. Recently, deep learning (DP) techniques (Ahishali et al., 2019; Feng et al., 2019; Zhu et al., 2017) have shown that neural network models are well adapted tools to automatically produce land cover mapping from information coming from both optical (Interdonato et al., 2019) and radar (Ho Tong Minh et al., 2018) sensors. The main characteristic of these models is the ability to simultaneously extract features optimized for image classification. However, only few studies have started exploiting the potential of S1 time series for land cover classification using the DP method. In Ho Tong Minh et al. (2018) the authors proposed to leverage DP to perform winter quality mapping over a zone in West France. In Ndikumana et al. (2018) another DP approach was proposed to perform land cover mapping considering several agricultural classes. Concerning the joint exploitation of radar/optical satellite images based on deep learning techniques, some approaches were recently proposed considering really precise tasks such as image simulation (He and Yokoya, 2018) or change detection (Liu et al., 2018). Unfortunately, the same opportunity has not been yet fully exploited in the context of land cover mapping.

This study proposes an innovative approach to map irrigated areas at plot scale using the Sentinel-1 and Sentinel-2 (S2) time series. The potential of S1 and S2 acquired at high spatial resolution and high revisit time for irrigation mapping did not yet receive enough attention by the scientific committee. The approach is based on the use of S1 and/or S2 time series combined

with statistical and mathematical functions such as the principal component analysis (PCA) and the wavelet transformation (WT) to map irrigated areas. The novelty of this study resides in eliminating the ambiguity between irrigation and rainfall through the comparison between the SAR backscattering signal of each plot and that of the grid (10 km × 10 km) including this plot. To map irrigated areas, the PCA and WT transformation were applied to the S1 time series prepared over a study site located in west of Catalonia, Spain. The PCA was also applied to the NDVI time series derived from the S2 images. Then, random forest (RF) and the convolutional neural network (CNN) approaches were used to build up classification models in three different scenarios: The first using only S1 data, the second using only S2 data, and the third using both S1 and S2 data. Finally, the two classification approaches (RF and CNN) were inter-compared.

2. Materials

2.1 Study Site

The study site examined is the western part of Catalonia region of north east Spain (Figure 1a). Figure 1b shows the location of the S1 footprints used in this study. The area covered by the used footprints corresponds to 55% of the total area of Catalonia and compromises 64% of the total agricultural areas in the region. Figure 1c is a digital elevation model (DEM) derived from the shuttle radar topography mission (SRTM) data and Figure 1d shows the agricultural areas derived from the SIGPAC (geographical information system for agricultural parcels) data of Catalonia. The hatched zone in Figure 1d represents the area finally used in the classification.

Catalonia is a region of 32,000 km^2 located in the northeast corner of the Iberian Peninsula along the northeastern Spanish Mediterranean coast. The agricultural areas of Catalonia are concentrated in the northeastern part and the interior plane in the western part as shown in the agricultural map of Figure 1d. The climate in the region is typically Mediterranean; mild in winter and warm in summer with very dry season. Figure 2 presents the precipitation and temperature data recorded at a local meteorological station located in Tornabous village of the interior plane (Figure 1b) (https://ruralcat.gencat.cat/web/guest/agrometeo.estacions). The average annual precipitation in this station is around 400 mm.

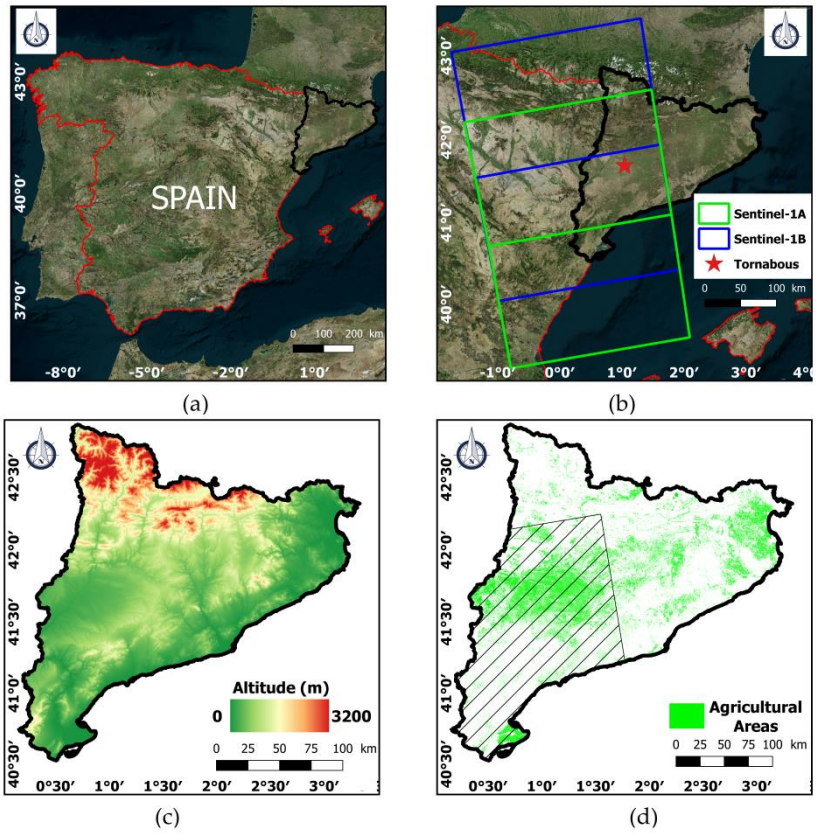


Figure 1: (a) Location of the study site (in black), Catalonia Spain, (b) Sentinel-1 footprints over Catalonia used in the study, (c) digital elevation model (DEM) from shuttle radar topography mission (SRTM) data, (d) agricultural areas of Catalonia derived from geographical information system for agricultural parcels (SIGPAC) data. The hatched area represents the zone finally used for classification.

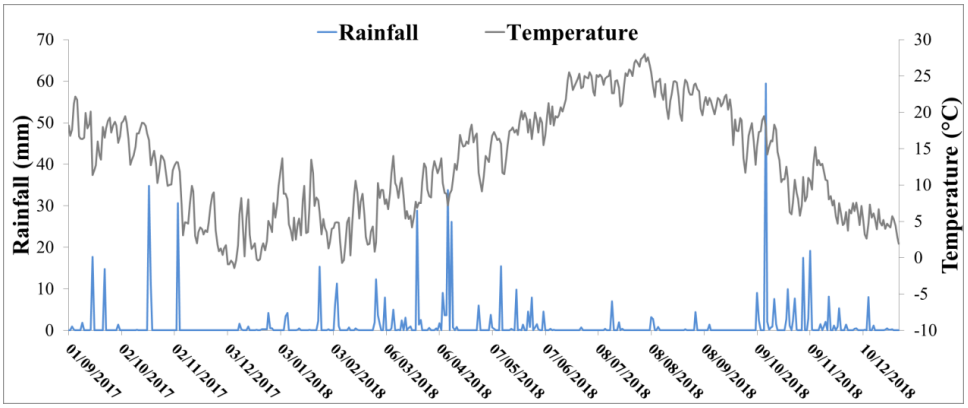


Figure 2: Precipitation and temperature records for a local meteorological station in Tornabous of the interior plain of Catalonia, Spain.

During the dry summer season between May and September, irrigation mainly occurs in the study site. Different irrigation techniques exist. In the old irrigation systems, water is distributed using open channels and irrigation is performed by inundation every two weeks. In the new irrigation systems, water is distributed by pressure where sprinklers or drip irrigation are used depending on the culture type. Irrigation here can be applied daily.

2.2 SIGPAC Data

SIGPAC data provides referenced geographic boundaries of plot land in Catalonia attributed with alphanumeric information. The data are produced and maintained by the General Direction of Rural Development of the Department of Agriculture, Livestock, Fishing and Food of the Generalitat of Catalunya. The plots are digitized using orthophoto images of 25-cm spatial resolution at scale 1:2000. Group of fields are associated in the database of the provided geographical data describing each plot by a unique identification code, surface area, land use, irrigation coefficient (0 for non-irrigated and 100 for irrigated), average slope and others. The SIGPAC data can be freely downloaded for the whole region in vector format. For each plot, the crop type could be obtained from the agricultural crop map of Catalonia (DUN). This map is produced based on the SIGPAC geographic boundaries and the annual declaration of the agricultural exploitation. The agricultural map of Catalonia is also available in free and open access mode.

In this study, SIGPAC data was used as a reference and validation of our proposed classification approaches. In our study, only agricultural crops (summer and winter crops) were considered for the irrigation classification. Forests, urban and orchards plots were removed from the SIGPAC database using the DUN crop map. Finally, a total of 193,327 plots of different crop types were used in the study forming a total area of 3795 km^2. The pie chart of Figure 3 summarizes the distribution of dominant crop types present in the studied zone. It shows the number of plots per crop type present in the study area. In general, winter cereals such as wheat, oat and barley are rarely irrigated. However, it may happen that these crops are exceptionally irrigated very few times. On the other hand, irrigated plots mainly include alfalfa, maize, grassland, beans, rapeseed and rice.

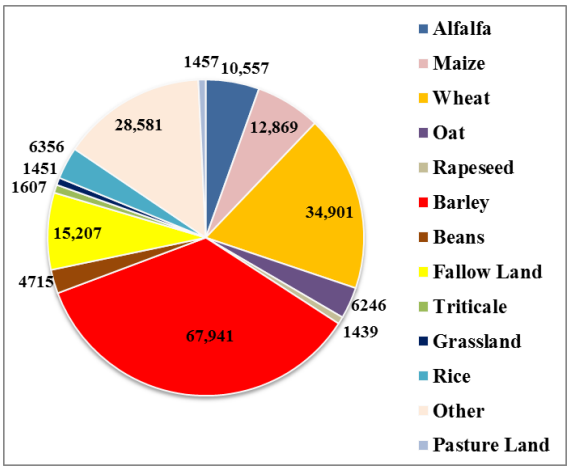


Figure 3: Distribution of the number of agricultural plots per class of crop type.

2.3 Remote Sensing Data

In this study, a database of radar (SAR) and optical data was prepared over the study site described in Section 2.1. While the S1 satellite constellation provides SAR images, the S2 satellite provides optical images. Table 1 summarizes the characteristics of remote sensing data used in the study. More details on each data type are further provided in the following subsection.

Table 1: Summary of the used remote sensing data

Data	Type	Satellite	Spatial Resolution	Images Number	Frequency Used	Calibration Performed	Bands Polarization
Satellite Images	SAR	Sentinel-1	10 m	82	6 days	Geometric/Radiometric	VV VH
	Optical	Sentinel-2	10 m	17	One month	Ortho-rectified Reflectance	Near Infrared, Red

2.3.1 Sentinel-1 SAR Data

A time series of S1 images collected over our study site was used to generate SAR temporal profiles at both VV and VH polarizations. Eighty-two S1 images obtained from the S1A and S1B satellite constellation operating at the C-band (frequency = 5.406 GHz, wavelength ~6 cm) were collected for the period between 28 August 2017 and 27 December 2018. Four S1 (two S1A and two S1B) footprints cover the western part of Catalonia with a six days temporal resolution as shown in Figure 1b. The 82 S1 images were generated from the high-resolution Level-1 ground range detected (GRD) product with 10 m × 10 m pixel spacing and acquired in the interferometric wide swath (IW) imaging mode with the VV (Vertical-Vertical) and VH (Vertical-Horizontal) polarizations. These S1 images are available via the Copernicus website (https://scihub.copernicus.eu/dhus/#/home). The 82 S1 images were calibrated (radiometric and

geometric correction) using the Sentinel-1 Toolbox (S1TBX) developed by the ESA (European Spatial Agency). The radiometric calibration aims to convert the digital number values of the S1 images into backscattering coefficients (σ°) in a linear unit while the geometric correction ortho-rectifies the SAR image using the DEM of the SRTM at 30 m spatial resolution. The radiometric accuracy of the S1 SAR backscattering coefficient is approximately 0.70 dB (3σ) for the VV polarization and 1.0 dB (3σ) for the VH polarization (El Hajj et al., 2016b; Schwerdt et al., 2017). Since the terrain in the study site is very complex, the radar backscattering coefficients corresponding to areas with high slope values (slope > 20%) were masked out. The slope mask was generated from the SRTM DEM (Figure1c) at 30 m spatial resolution available via the earth explorer website (https://earthexplorer.usgs.gov/) and applied to the 82 radar image.

2.3.2 Sentinel-2 Optical Data

Currently, S2 offers fine spectral and spatial resolution optical images with 13 bands ranging from 10 m to 60 m spatial resolution with 5 days revisit time. Seventeen free optical images were obtained over our study site during the period between August 2017 and December 2018 at a frequency of one image per month. The S2 images were downloaded from the Theia (French land data center) website (https://www.theia-land.fr/). The Theia website provides ortho-rectified S2 images at level-2A corrected for atmospheric effect. However, starting from April 2018, Theia website started providing a monthly synthesis of cloud free images (level-3A). These images were used for the period between April 2018 and December 2018.

3. Methodology

3.1 Overview

An overview of the methodology is presented in Figure 4. First, the average S1 SAR backscattering coefficients (σ°) were calculated at both plot scale (SIGPAC plots) and grid scale (10 km × 10 km). The mean signal at grid scale will be used to discriminate between rainfall and irrigation events. In general, we suppose that if the mean signal increases at the base of 10 km × 10 km then a rainfall event had occurred. Next, the PCA and the WT were performed on the obtained S1 time series. Moreover, the PCA was applied on the temporal series of the NDVI established at plot scale using the S2 optical data. Finally, two classification approaches including the RF classifier and the CNN were investigated. In each approach, three scenarios

have been tested. The first scenario is based on the use of SAR data only, the second scenario considers the use of optical data only, and the third scenario includes the combined use of both optical and SAR data. A detailed description of the methodology is further elaborated in the following sub-sections.

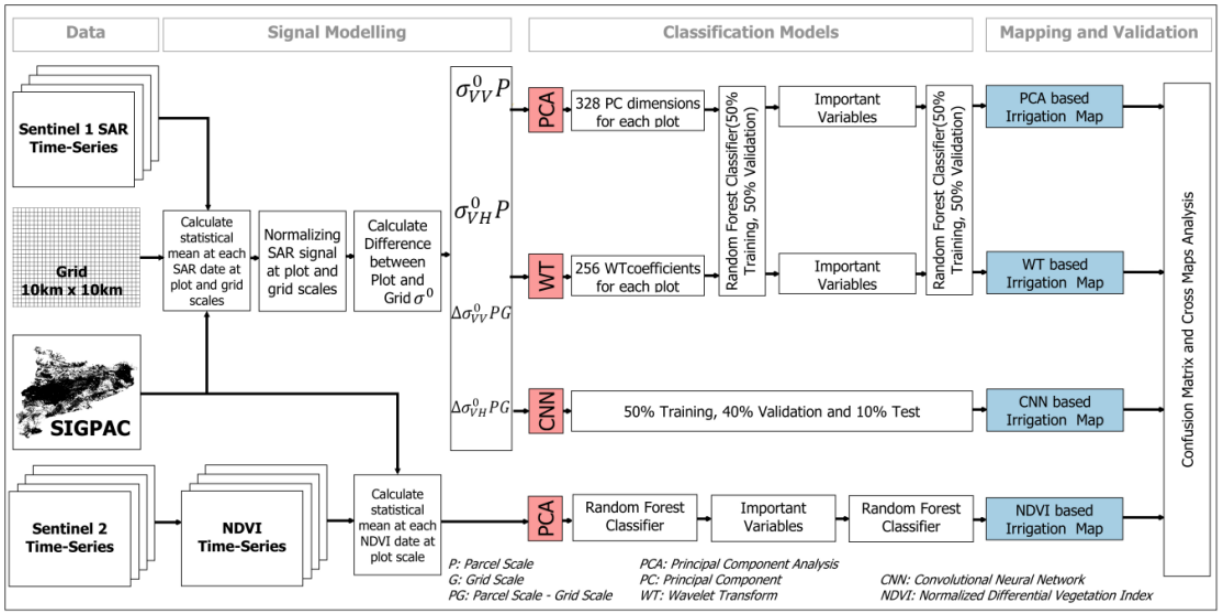


Figure 4: Workflow overview using random forest (RF) and the convolutional neural network (CNN).

3.2 σ° SAR Backscattering at Plot and Grid Scale

The detection of irrigation activities using SAR data at plot scale requires a good separation of irrigation events from rainfall events. In fact, both events are considered as water supplements and may have the same effect on the radar backscattering signal. In this case, additional information about the rainfall, such as the daily precipitation records, is required to remove the ambiguity between rainfall events and irrigation events. However, rainfall data is normally hard to be obtained freely and accurately. Thus, another indicator is required to separate irrigation events from rainfall events.

Following a rainfall event, the soil becomes wet thus causing an increase of the SAR backscattering signal due to the change in the dielectric constant of the soil (Amri et al., 2012; Baghdadi et al., 2011b, 2016a; Hajj et al., 2014). If all the plots within a limited spatial extent (5 km, 10 km, 20 km, etc.) show an increase in the SAR backscattering signal between two consecutive radar acquisitions, then a rainfall event could have probably occurred. Thus, an indicator of an existing rainfall event could be the increase of the SAR backscattering signal, between two consecutive SAR dates, obtained within a grid scale (5 km, 10 km, 20 km, etc.).

This increase of the SAR backscattering signal at a grid cell is mainly due to increase of soil moisture due to a rainfall event. In fact, the relation between rainfall events and soil moisture has been investigated in several studies (Brocca et al., 2009, 2014). Most recently, Bazzi et al. (2019a) compared the soil moisture values estimated at plot scale with six days revisit time from S1 data (El Hajj et al., 2017) with GPM (global precipitation mission) rainfall data at 0.1° × 0.1° grid scale (~10 km × 10 km) (Skofronick-Jackson et al., 2018). The study site examined was the Occitanie region of France (~72,000 km^2). They showed that following a rainfall event occurring couple of days before the radar acquisition, the average soil moisture of plots with low vegetation cover (NDVI < 0.4) in a grid of ~10 km × 10 km increases. This increase in soil moisture estimation is directly related to the increase of the radar backscattering signal at grid scale.

The increase of the SAR backscattering signal at plot scale accompanied with the stability or decrease of the radar signal at grid scale could be attributed to an irrigation event that took place on the plot. On the other hand, the increase of the SAR backscattering signal at the grid scale could be linked to a rainfall event. Therefore, to remove the ambiguity between rainfall and possible irrigation events a comparison is considered between the SAR backscattering coefficient at each plot and the SAR signal obtained from bare soil plots with low vegetation cover at 10 km × 10 km grid cell containing this plot. However, this comparison may not be useful for detecting irrigated plots if the irrigation occur simultaneously for almost all plots existing within the same 10 km grid cell. In this case, an increase of the SAR backscattering signal at grid scale would be due to irrigation that took place on all the plots within the 10 km cell and not rainfall. It is also good to mention that cropland plots could encounter a change in the radar backscattering coefficient due to change in the surface roughness mainly at sowing stage (smooth) or after harvesting (moderate to rough) (Baghdadi et al., 2002). Nevertheless, very rough conditions in an agricultural plot are not permanent over the year and may exist only for couple of days.

The temporal series of the S1 SAR backscattering coefficients over each agricultural plot was obtained by averaging the σ° values of all pixels within the plot at each available date and at both VV and VH polarizations. The average of σ° pixels' values reduces the speckle noise. For grid scale, a 10 km × 10 km grid was first generated to cover the whole region. Then, the SAR backscattering coefficients were averaged at each date for each grid cell (10 km × 10 km) using pixels within all the agricultural plots having NDVI values less than 0.5 (NDVI values calculated using S2 images explained later in Section 3.3). The threshold value of the NDVI

was selected in order to consider $\sigma°$ values for bare soil plots and plots with a low vegetation cover. For each plot, the SAR σ° temporal series in both VV and VH polarizations and σ° temporal series for its corresponding 10 km × 10 km grid cell in both VV and VH polarizations were obtained.

In order to adjust σ° values at a common scale (between 0 and 1), the σ° time series at plot and grid scale were normalized using the minimum-maximum normalization:

$$Y' = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (1)$$

where Y' is the normalized value, Y is the initial value of the time series (σ° at plot scale " σ^0_{plot} " and at grid scale "$\sigma^0_{10\text{km}}$ " at VV and VH polarizations), Y_{min} is the minimum value of the time series, and Y_{max} is the maximum value of the time series.

The difference between the normalized σ° at plot scale and the normalized σ° at grid scale was then computed for each plot at each date in both VV and VH polarizations:

$$\Delta\sigma^0_{\text{PG}} = \sigma^0_{\text{plot}} - \sigma^0_{10\text{km}} \quad (2)$$

Finally, for each plot we obtain four temporal series including:

– Normalized σ° time series at plot scale in VV polarization ($\sigma^0_{\text{P,VV}}$);

– Normalized σ° time series at plot scale in VH polarization ($\sigma^0_{\text{P,VH}}$);

– Difference between the normalized σ° at plot and grid scales in VV polarization ($\Delta\sigma^0_{\text{PG,VV}}$);

– Difference between the normalized σ° at plot and grid scales in VH polarization ($\Delta\sigma^0_{\text{PG,VH}}$) .

3.3 NDVI Temporal Series at Plot Scale

Using the S2 images, 17 NDVI maps were derived from the red and infrared bands at each month between September 2017 and December 2018:

$$\text{NDVI} = \frac{\rho\text{NIR} - \rho\text{RED}}{\rho\text{NIR} + \rho\text{RED}} \quad (3)$$

To obtain an NDVI value for each plot at each month, the NDVI pixels within each plot were averaged. Finally, a temporal series of 17 NDVI values were obtained at plot scale.

3.4 Principal Component Analysis

To reduce the volume of the data prior to the application of any classification method, a data compression is necessary to be applied especially when a huge amount of data is involved in the study. This data compression reduces the dimensionality of the classifier workspace and eliminates redundancies and nonessential information in raw data that may lead to inaccurate classification. Among data compression techniques, the principal component analysis is a statistical compression technique which converts high dimensional data to low dimensional data by selecting the most important features capturing maximum information about the dataset (Jain, 1989; Jolliffe and Cadima, 2016). Data compression by means of PCA is accomplished by projecting each sample block of data along the directions of the individual orthonormal eigenvectors of the covariance matrix of the data. The conventional approach for the PC extraction involves the computation of the input data covariance matrix and then the application of a diagonalization procedure to extract the eigenvalues and the corresponding eigenvectors.

For each plot, the PCA was applied on each one of the four SAR temporal series presented in Section 3.2 ($\sigma^0_{P,VV}, \sigma^0_{P,VH}, \Delta\sigma^0_{PG,VV}, \Delta\sigma^0_{PG,VH}$). Since each temporal series consists of 82 different values, we obtain 82 different PC values for each time series. Finally, each plot is defined by 328 different PC values (82 × 4).

For optical images, the PCA was also applied to the NDVI temporal series of each plot. Since the temporal series consists of 17 NDVI values, the PCA produces 17 different PC values for each plot. The obtained 17 PC values are later used in a random forest classifier in order to classify irrigated plots.

3.5 Haar Wavelet Transformation

Wavelet transformation (WT) is used to decode a real periodic or non-periodic signal into a linear combination of elementary functions. The principal of the wavelet transformation is based on the decay of the signal on a set of sinusoidal functions. Several wavelet transformation functions exist from which the 'Haar' wavelet transformation was chosen. The idea behind this wavelet analysis is to decompose the signal of $N = 2^k$ points on a basis of discontinuous and orthogonal functions resulting from the translation and dilations of functions called the father scaling function $\varphi(t)$ and the mother wavelet function $\psi(t)$.

The real SAR time series obtained from the available S1 images consists of 82 values corresponding to 82 different dates. Since 82 is not a perfect square of 2 ($N = 2^k$), the number of values to be considered in the wavelet decomposition should be limited to the 1st perfect square less than 82. Thus only 64 values were considered out of the 82 values in the time series. For 64 σ° values, the temporal series could be decomposed on six levels where $2^6 = 64$. However, the 64 values of the time series were able to cover a period of one year (from 15 September 2017 until 28 September 2018) which is considered sufficient to detect irrigation activities since most of the irrigation in the region mainly exist between May and September of each year.

The 'Haar' wavelet transformation was computed at each plot for $\sigma^0_{P,VV}, \sigma^0_{P,VH}, \Delta\sigma^0_{PG,VV}, \Delta\sigma^0_{PG,VH}$. Thus, each SAR temporal series of 64 values is transformed into 64 independent wavelet coefficients corresponding to six levels of decomposition. Each plot attains finally 256 different wavelet coefficients (64 coefficients × four time series).

3.6 Random Forest Classifier

Random forest (RF) is an ensemble of machine learning algorithms consisting of large number of decision tree classifiers where each tree is constructed using a different subset of the training set. In its architecture, the RF classifier fits a number of decision tree classifiers on various subsamples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting (Breiman, 2001). When compared to other classical machine learning algorithms, the random forest classifier has demonstrated its ability to yield high quality classification mapping with fast computation time (Inglada et al., 2017; Rodriguez-Galiano et al., 2012). The tree depth and minimum sample size are the two most important parameters in the RF classifier.

Three different scenarios were tested using the RF classifier. First, the RF classifier was modeled using SAR data only (one for PC values and another for the WT coefficients). Second, NDVI optical data was tested in another RF classifier. Finally, we combined SAR and optical data in a third RF classifier.

Using SAR data only, the RF classifier was applied for the PC variables and the WT coefficients separately in two different classifiers denoted later by PC-RF and WT-RF respectively. First, each RF classifier (PC-RF and WT-RF) was trained using all variables of each method (328 values for the PC-RF and 256 coefficients for the WT-RF). This step gives

the importance of each variable which is a number between 0 and 1 evaluating the contribution of the variable in the classification. The importance of the variables was then analyzed and a threshold value of the importance was determined to select the significant variables only. This step allows reducing the number of variables for each classifier while retaining good accuracy. Then, both RF classifiers were regenerated using only important variables for each method. In all tested RF classifiers, the number of trees was set to 500. This number is fixed by increasing the number of trees until the accuracy of the classification converges. The tree depth was kept at its default value which is the square root of the number of variables available. In order to train and validate the model on two independent datasets, we randomly split the database into 50% for training and 50% for validation (Demarez et al., 2019; Wang et al., 2018).

Using optical data only, the 17 PC values obtained from the NDVI data are implemented in a random forest classifier later denoted by NDVI-RF. As done for SAR data, important variables were then retained based on a threshold value of a variable's importance, and the RF classifier was reestablished using these variables only.

Finally, a RF classifier was modeled using both SAR and optical data. From the two SAR classifiers (PC-RF and WT-RF) the PC-RF was chosen since its results show slightly better performance. Therefore, the important variables obtained in the PC-RF classifier were combined with the important variables obtained in the NDVI-RF in a new RF classifier.

3.7 Convolutional Neural Network

Convolutional Neural Networks are deep learning models that achieve state-of-the-art performance on image related classification tasks (Bengio et al., 2013) and, nowadays, they are getting increasing attention in the field of remote sensing (Zhu et al., 2017). Here, one dimensional (1D) CNN model (CNN1D) was proposed to leverage the convolutional characteristics of such networks to manage the time dimension of Sentinel-1 time series data. In this scenario, a simple and effective CNN architecture that involves standard operation was conceived: Convolution, nonlinear activation function, batch normalization and dropout. Each convolutional layer is applied on the valid portion of the time series (without any kind of padding) and it is associated with a convolution combined with a rectifier linear unit (ReLU) activation function (Nair and Hinton, 2010) to induce non-linearity. Successively, a batch normalization step (Ioffe and Szegedy, 2015) is employed to accelerate the network convergence and reduce the internal covariate shift. Finally, Dropout (Dahl et al., 2013) was

adopted with a drop rate equal to 0.4, i.e., 40% of the neurons are randomly deactivated at each propagation step.

Due to the fact that S1 time series and NDVI time series have different lengths, a two branch architecture (one CNN1D for the S1 time series and one CNN1D for the NDVI time series) that performs late fusion (Benedetti et al., 2018) of the input sources was proposed. The fusion is achieved with a simple concatenation of the features extracted from each branch. The concatenated features are successively used by two fully connected layers to produce the final classification. The fully connected layers are still associated with a nonlinear activation function (ReLU) and both of them contain 512 neurons. The two-branch network architecture is trained end-to-end. Figure 5 represents the architecture of the developed CNN. Each layer of the network involves the following operations: Convolution with ReLU function (Conv), batch normalization (BN), and dropout (DropOut): DropOut(BN(Conv(.))). For each convolution, the number of filters used are 64, 128, or 256, the Kernel size is 7×1, 5×1, 3×1 or 1×1 and the stride value is 1or 2. Once the features are combined (concatenation), two free fully connected layers are employed to provide the final decision. The output layer (the last layer of the network) has as many neurons as the number of classes to predict (2).

For all deep learning approaches, the Adam method (Kingma and Ba, 2014) to learn the model parameters with a learning rate equals to 1×10^{-4} was used. The training process is conducted over 1000 epochs with a batch size of 256. Considering the train/test splits (50%, 50%) for the neural network strategies, the data was split as 50%, 40% and 10% for training, validation and test set respectively (Interdonato et al., 2019). The model that achieves the best performances on the validation set will be successively blindly evaluated on the test set.

Similar to what was proposed for the RF classifiers, two scenarios were first considered for the deep learning approach one using only the S1 time series data and the other using only the NDVI time series. To do this, one of the two branches of the model is completely dropped and the one corresponding to the specific data source is kept. Then, for the third scenario using SAR and optical data, the full model with both branches was used.

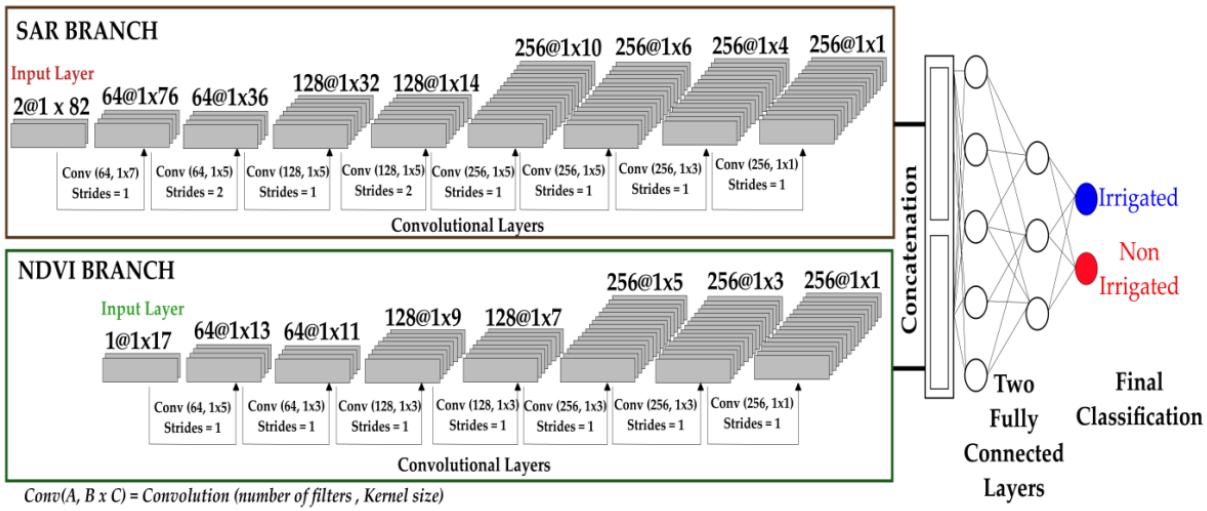


Figure 5: Architecture of the one dimensional (1D) CNN model (CNN1D) used for classification of irrigated/non-irrigated plots using SAR and optical data.

3.8 Accuracy Assessment

For each tested method, the assessment of the classifier results was expressed as a function of the overall accuracy (OA), Kappa and the F-score by applying the generated model on the validation dataset (50%). The precision, recall, and F-score were also calculated for each class (irrigated/non-irrigated). In fact, several studies report the use of these indices to assess the accuracy of the classification (Bazzi et al., 2019d; Gaetano et al., 2018; Ho Tong Minh et al., 2018; Ndikumana et al., 2018). While OA can be easily interpreted as the percentage of correctly classified plots to the total number of plots, Kappa and F-score can be used to assess the statistical differences between classifications. The obtained accuracy indices were calculated for the obtained results of each method as proposed by (Fawcett, 2006; Landis and Koch, 1977; Powers, 2011). Table 2 summarizes the equations used to calculate these accuracy indices.

Table 2: Accuracy indices calculated for each classifier. TP is the number of the irrigated plots truly classified as irrigated, TN is the number non-irrigated plots truly classified as non-irrigated plots, FP is the number of non-irrigated plots falsely classified as irrigated, FN is the number of irrigated plots falsely classified as non-irrigated plots, N is the total number of plots, and EA is the expected accuracy.

Index	Equation
OA	$\frac{TP + TN}{TP + TN + FN + FP}$
EA	$\frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{(TP + TN + FN + FP)^2}$
Kappa	$\frac{OA - EA}{1 - EA}$
F1 score	$\frac{2TP}{2TP + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$

4. Results

4.1 Comparison of σ° SAR Backscattering at Plot and Grid Scale

To remove the ambiguity between rainfall and irrigation events, this study proposes to compare between the SAR signal obtained at plot scale and the SAR signal obtained at 10 km grid scale. Figure 6 represents an example of the temporal behavior of σ° values in VV polarization obtained at plot and grid scale for a non-irrigated plot (Figure 6a) and an irrigated plot (Figure 6b). In both figures, the green curve represents the σ° temporal evolution at the plot scale and the red curve represents the σ° temporal evolution calculated at the 10 km grid cell. Precipitation data obtained from a local meteorological station is added to the figures to better understand the relation between grid scale σ° variations and the existence or absence of rainfall events. For the non-irrigated plot in Figure 6a, the σ° SAR values at plot scale and grid scale attain the same behavior for almost all the dates. Following a rainfall event, both curves increase and then decrease following a dry period. The consistency between both curves at both scales for the non-irrigated plot during the whole period indicates that the plot did not receive any water supplement other than rainfall events. For the irrigated crop in Figure 6b, the temporal behavior of the σ° SAR temporal series at plot and grid scales are coherent between September 2017 and April 2018. Both σ° values for both scales increase following rainfall event, and decrease with the absence of rainfall. However, between May 2018 and September 2018 (black dashed circle in Figure 6b) the irrigated plot shows a frequent change in σ° values accompanied

with the absence of rainfall events due to possible irrigation events occurring. On the other hand, the 10 km grid cell shows low and constant values of σ° during the same period indicating very dry soil conditions.

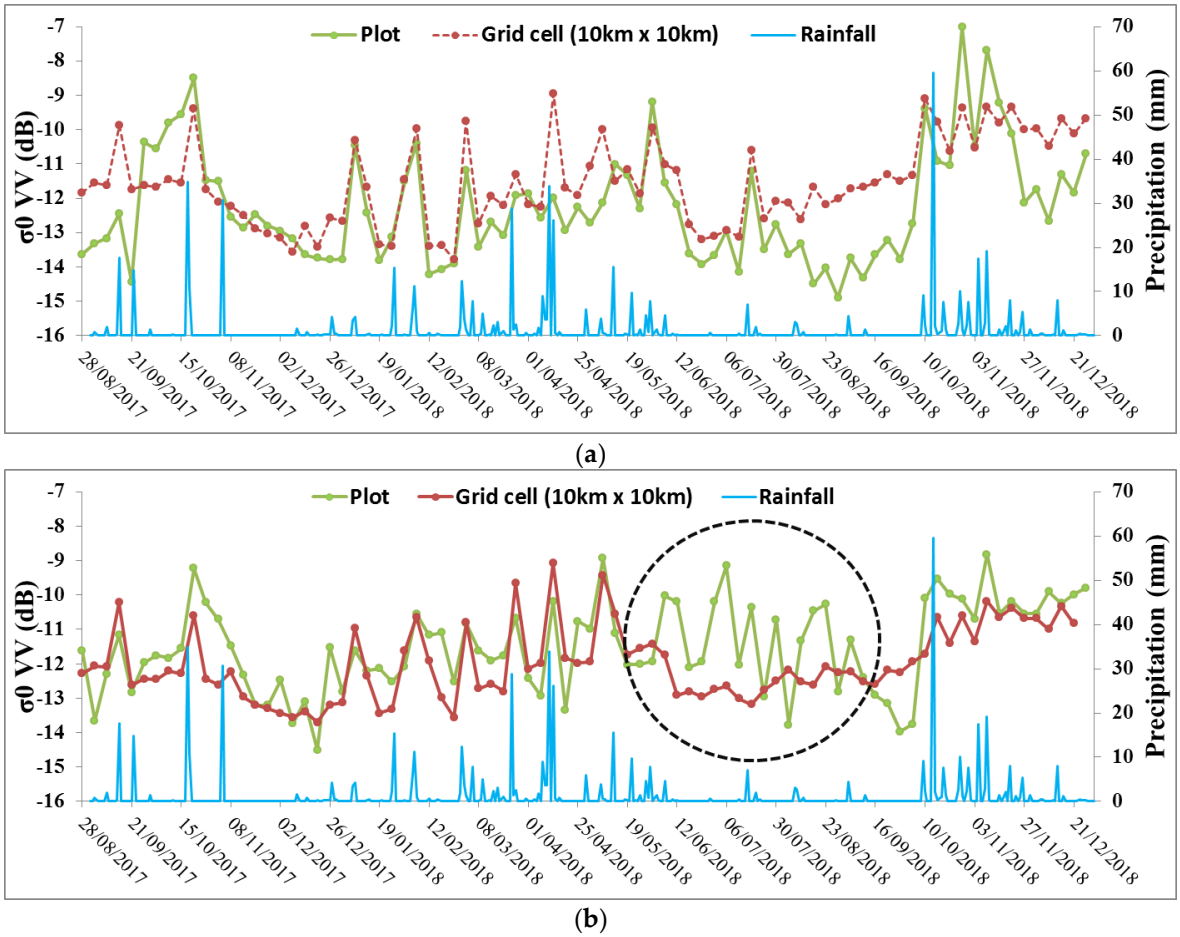


Figure 6: Temporal evolution of SAR backscattering coefficient σ° in VV polarization at plot scale (green curve) and 10 km grid scale (red curve) with precipitation data recorded at a local meteorological station for (a) the non-irrigated plot, (b) the irrigated plot.

4.2 Classification using Random Forest Classifier

4.2.1 PC-RF

As discussed in Section 3.6, the RF classification was first performed using the total number of PC values obtained from the application of the PCA on the four different temporal series (328 variables). Then, the significant PC variables that most contributed in the classification of the irrigated/non-irrigated plots were determined using the increase in the mean square error of the predictions (%IncMSE > 1%). The use of this threshold value allows reducing the number of predictive variables from 328 to 15 while keeping accurate classification. These variables contain the first, second, fourth and sixteenth PC dimension of $\sigma^0_{P,VV}$, the first third and

sixteenth PC dimension of $\sigma^0_{P,VH}$, the second and the fifth PC dimension of $\Delta\sigma^0_{PG,VV}$, and the first and second dimension of $\Delta\sigma^0_{PG,VH}$. Figure 7 shows the scatter plot of 2000 irrigated and 2000 non-irrigated plots (randomly selected) with different combinations of the obtained important PC values. It clearly shows that irrigated and non-irrigated plots can be separated using several PC combinations as they distinctly appear in two different point clouds. The RF classifier is thus capable of obtaining good results when introducing these variables for classifying irrigated and non-irrigated plots.

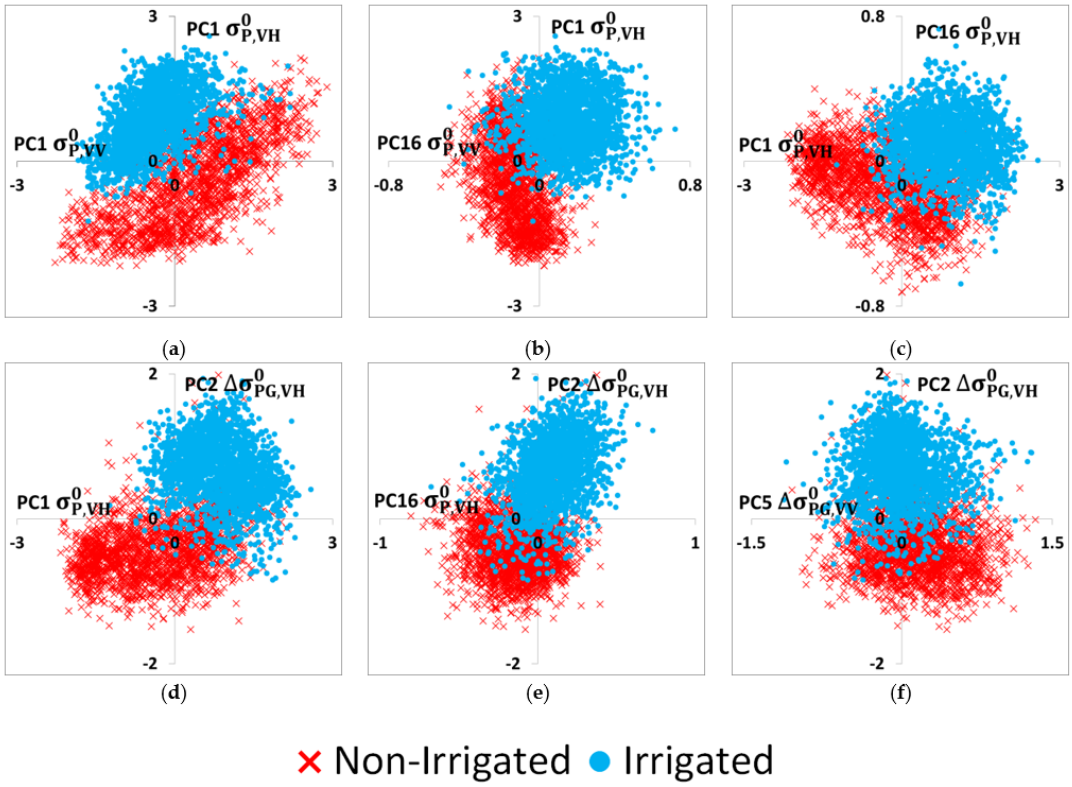


Figure 7: Scatter plot of a random sample of 2000 irrigated and 2000 non-irrigated plots using different combinations of important principal component (PC) variables. Irrigated plots are presented in blue and non-irrigated plots represented in red. (a) PC1 of $\sigma^0_{P,VV}$ with PC1 of $\sigma^0_{P,VH}$, (b) PC16 of $\sigma^0_{P,VV}$ with PC1 of $\sigma^0_{P,VH}$, (c) PC1 of $\sigma^0_{P,VH}$ with PC16 of $\sigma^0_{P,VH}$, (d) PC1 of $\sigma^0_{P,VH}$ with PC2 of $\Delta\sigma^0_{PG,VV}$, (e) PC16 of $\sigma^0_{P,VH}$ with PC2 of $\Delta\sigma^0_{PG,VH}$ and (f) PC5 of $\Delta\sigma^0_{PG,VV}$ with PC2 of $\Delta\sigma^0_{PG,VV}$. $\sigma^0_{PG,VV} = \sigma^0_{P,VV} - \sigma^0_{G,VV}$ and $\sigma^0_{PG,VH} = \sigma^0_{P,VH} - \sigma^0_{P,VH}$. "P" means plot scale and "G" means grid scale.

Table 3 summarizes the accuracy obtained when applying the trained RF classifier on the 50% validation data using all the variables and the selected important variables. The validation of the RF classifier using the 328 PC values generally shows very high accuracy. In fact, the overall accuracy reaches 91.2% while the kappa coefficient reaches 0.79 and the F-score reaches 0.91. The F-score of the irrigated class is 0.86 while that of the non-irrigated class is 0.94. The reestablishment of the RF classifier using the 15 important PC variables produced

similar accuracy compared to the RF classifier using the total number of variables. In general, both classifiers present nearly the same results. However, the overall accuracy slightly decreased from 91.2% to 90.7% but yet keeping an accurate classification. The Kappa coefficient and F-score remained unchanged for both approaches.

Table 3: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the RF classifier using the 328 SAR PC variables and 15 important SAR PC variables.

Method	Class	Precision	Recall	F-Score
PC-RF 328 Variables	**Irrigated**	0.95	0.79	0.86
	Non-Irrigated	0.90	0.98	0.94
	OA		91.2%	
	Kappa		0.79	
	F-score		0.91	
PC-RF 15 Important Variables	**Irrigated**	0.92	0.81	0.86
	Non-Irrigated	0.90	0.96	0.93
	OA		90.7%	
	Kappa		0.79	
	F-score		0.91	

4.2.2 WT-RF

The linear combination of all the levels of the WT allows the reconstruction of the real SAR signal using the obtained 64 wavelet coefficients. Figure 8 presents an example of the reconstruction of the σ° SAR temporal series of an alfalfa irrigated plot using the linear combination of consecutive decomposing levels (functions) of the wavelet transformation. The first father scale function consists of two coefficients (level 1). The mother scale functions of levels two, three, four, five and six consist of two, four, eight, 16, and 32 coefficients respectively. The linear combination of the father level with additive mother levels allows the reconstruction of the real SAR signal (Figure 8). The obtained 64 wavelet coefficients are used in a random forest classifier in order to classify irrigated plots.

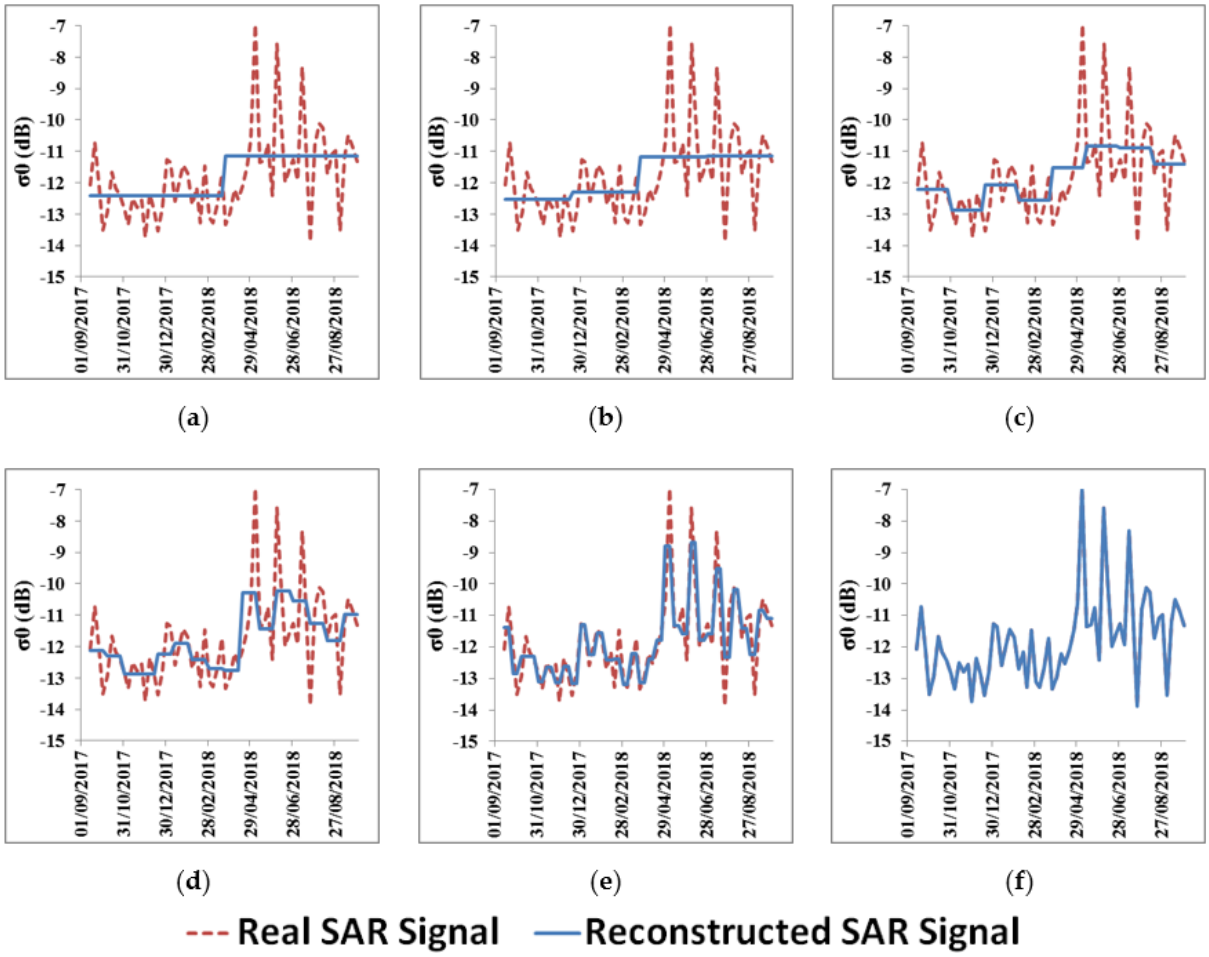


Figure 8: Reconstruction of SAR signal in VV polarization at plot scale through the linear combinations of the 'Haar' wavelet coefficients using (a) 2 coefficients, (b) 4 coefficients, (c) 8 coefficients, (d) 16 coefficients, (e) 32 coefficients, and (f) 64 coefficients.

The same process applied for the RF classification using the PC values is also performed using the Haar wavelet coefficients. The RF classifier was first trained using the 256 values (discussed in Section 3.5) of the wavelet coefficients. When validating the obtained classifier, the overall accuracy of the classification reached 91.4% with a 0.81 kappa coefficient and a 0.91 F-score. The F-score of the irrigated class (0.87) is slightly lower than that of the non-irrigated class (0.94) (Table 4). The significant wavelet coefficients that most contributed in the classification were also determined using the threshold value on the increase in the mean square error of the predictions (%IncRMSE >1%). This also reduced the number of variables from 256 coefficients to 18. Figure 9 similarly shows the scatter plot of 2000 irrigated and 2000 non-irrigated plots (randomly selected) with different combinations of the obtained important WT coefficients. As shown with the PC values, Figure 9 also clearly shows that irrigated and non-irrigated plots can be separated using the WT coefficients as they distinctly appear in two different point clouds.

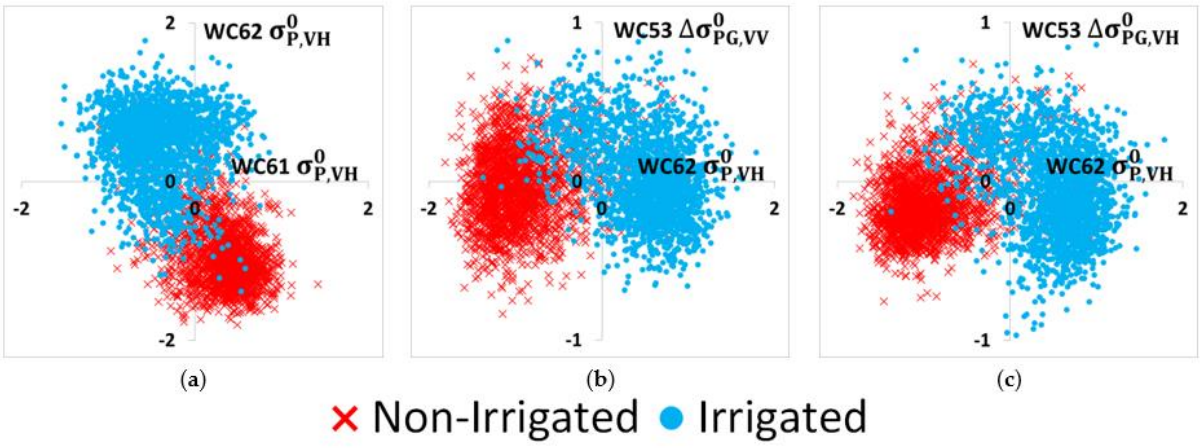


Figure 9: Scatter plot of a random sample of 2000 irrigated and 2000 non-irrigated plots using different combinations of important wavelet transformation (WT) coefficients. Irrigated plots are presented in blue and non-irrigated plots represented in red (a) WC61 of $\sigma^0_{P,VH}$ with WC62 of $\sigma^0_{P,VH}$, (b) WC62 of $\sigma^0_{P,VH}$ with WC53 of $\Delta\sigma^0_{PG,VH}$, (c) WC62 of $\sigma^0_{P,VH}$ with WC53 of $\Delta\sigma^0_{PG,VH}$. $\sigma^0_{PG,VV} = \sigma^0_{P,VV} - \sigma^0_{G,VV}$ and $\sigma^0_{PG,VH} = \sigma^0_{P,VH} - \sigma^0_{P,VH}$. "P" means plot scale and "G" means grid scale and WC means wavelet coefficient.

The validation of the RF classifier trained using important data only (18 variables) produces slightly lower accuracy than that using the full dataset (Table 4). The overall accuracy decreased by 1.3% only. The F-score of the irrigated class decreased from 0.87 to 0.83 while that of the non-irrigated class decreased from 0.94 to 0.92. The kappa value also decreased from 0.81 to 0.75. Although the accuracies reported using the 18 important WT coefficients show lower values compared to those obtained using the whole dataset, the proposed RF classifier keeps a very good performance.

Table 4: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the RF classifier using the 256 wavelet coefficients and the 16 important wavelet coefficients.

Method	Class	Precision	Recall	F-Score
WT-RF 256 Variables	**Irrigated**	0.94	0.81	0.87
	Non-Irrigated	0.90	0.97	0.94
	OA		91.4%	
	Kappa		0.81	
	F-score		0.91	
WT-RF 18 Important Variables	**Irrigated**	0.89	0.78	0.83
	Non-Irrigated	0.89	0.95	0.92
	OA		89.1%	
	Kappa		0.75	
	F-score		0.89	

4.2.3 NDVI-RF

The RF using NDVI data was first developed with the 17 PC values obtained from the application of the PCA on the NDVI temporal series. The accuracy obtained when validating

the model is reported in Table 5. Then, the analysis of the variables' importance allows us to fix a threshold value of 4% for the increase in mean square error. This threshold reduced the number of variables from 17 PC values to seven. The important PC variables obtained includes the first five, the eighth and 11th PC dimensions. The regeneration of the RF classifier using the seven important PC dimensions of NDVI produced approximately similar accuracy to that using all the PC variables. The overall accuracy decreased by only 1% while the kappa and F-score decreased by 2% and 3% respectively.

Table 5: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the RF classifier using the 17 normalized differential vegetation index (NDVI)-PC and the seven important NDVI PC values.

Method	Class	Precision	Recall	F-Score
NDVI-RF 17 Variables	**Irrigated**	0.94	0.78	0.85
	Non-Irrigated	0.89	0.97	0.93
	OA		90.5%	
	Kappa		0.78	
	F-score		0.91	
NDVI-RF 7 Important Variables	**Irrigated**	0.92	0.76	0.84
	Non-Irrigated	0.88	0.96	0.92
	OA		89.5%	
	Kappa		0.76	
	F-score		0.88	

4.2.4 RF using Combined Optical and SAR Data

The RF classifier using the 15 important PC variables shows slightly higher accuracy than that using the WT coefficients. For this reason, the important variables of the PC-RF were combined with the important variables of the NDVI-RF in a new RF classifier. The accuracies obtained are presented in Table 6. In general, a slight increase in the overall accuracy is obtained when adding the NDVI data to the SAR data using the PC values. The overall accuracy increased by 1.6% while kappa increases by 3%. The F-score of the classification remains unchanged.

Table 6: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the RF classifier using 15 important SAR PC variables and the seven important NDVI PC variables.

Method	Class	Precision	Recall	F-Score
15 variable SAR PC-RF + 7 variable NDVI-RF	**Irrigated**	0.95	0.82	0.88
	Non-Irrigated	0.91	0.98	0.94
	OA		92.3%	
	Kappa		0.82	
	F-score		0.91	

4.3 Classification using Convolutional Neural Network

4.3.1 CNN on SAR Temporal Series

As discussed in Section 3.7, the CNN was applied on the SAR temporal series of $\sigma^0_{P,VV}, \sigma^0_{P,VH}, \Delta\sigma^0_{PG,VV}, \Delta\sigma^0_{PG,VH}$ using the SAR branch of the developed CNN1D model. The data was divided into 50% for training, 40% for validation and 10% for testing. Table 7 summarizes the accuracy obtained when applying the CNN method on the SAR temporal series with the corresponding standard deviation of 10 different iterations. The CNN approach reports very high accuracy where the overall accuracy reaches 94.1%. The kappa value of the classification reaches 0.87 and the F-score of is 0.94. The F-score of the irrigated class reaches 0.91 which is less than that of the non-irrigated class (0.96).

Table 7: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the CNN method on SAR data.

Method	Class	Precision	Recall	F-Score
CNN on SAR Data	**Irrigated**	0.93	0.89	0.91
	Non-Irrigated	0.95	0.96	0.96
	OA		94.1% $\pm$ 0.06	
	Kappa		0.87 $\pm$ 0.0014	
	F-score		0.94 $\pm$ 0.0006	

4.3.2. CNN Using NDVI Temporal Series

Using the optical branch of the developed CNN, the approach was also applied on the NDVI temporal series composed of 17 different NDVI values. The data was also divided into 50% for training, 40% for validation and 10% for testing. The accuracy obtained from the 10 iterations is reported in Table 8. The overall accuracy reaches 91.6% while the kappa and F-score values reach 0.81 and 0.91 respectively. The irrigated class also shows a lower F-score value (0.87) than the non-irrigated class (0.94).

Table 8: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the CNN method on NDVI data.

Method	Class	Precision	Recall	F-Score
CNN on Optical Data	**Irrigated**	0.93	0.81	0.87
	Non-Irrigated	0.91	0.97	0.94
	OA		91.6% $\pm$ 0.06	
	Kappa		0.81 $\pm$ 0.0016	
	F-score		0.91 $\pm$ 0.0006	

4.3.3. CNN Using Combined SAR and Optical Data

The application of the two branches of the CNN model (SAR and optical branches) discussed in Section 3.7 shows that the classification model using both datasets did not considerably increase the accuracy of the classification. The overall accuracy of the combined model reaches $94.5\% \pm 0.05$. The kappa coefficient and the F-score increased by 1% only. The accuracy of the irrigated class has increased with 1% while that of the non-irrigated class remained stable (Table 9).

Table 9: The values of the overall accuracy, Kappa, F-score, precision, and recall obtained from the CNN method on combined SAR and optical data.

Method	Class	Precision	Recall	F-Score
CNN on Combined SAR and Optical Data	**Irrigated**	0.94	0.90	0.92
	Non-Irrigated	0.95	0.97	0.96
	OA		$94.5\% \pm 0.05$	
	Kappa		0.88 ± 0.0016	
	F-score		0.95 ± 0.0005	

4.4 Irrigation Mapping

Using the SAR based classification, three different maps were produced which are the WT-RF based map (Figure 10a), the PC-RF based map (Figure 10b) and the CNN based map (Figure 10c). In the three maps, the blue and the red colors represent the irrigated and non-irrigated plots, respectively. The irrigated areas in the classified maps are centered mainly in the western part of the study site. The irrigated part present in the south is the Ebro Delta typically used for rice cultivation which is inundated from April to November.

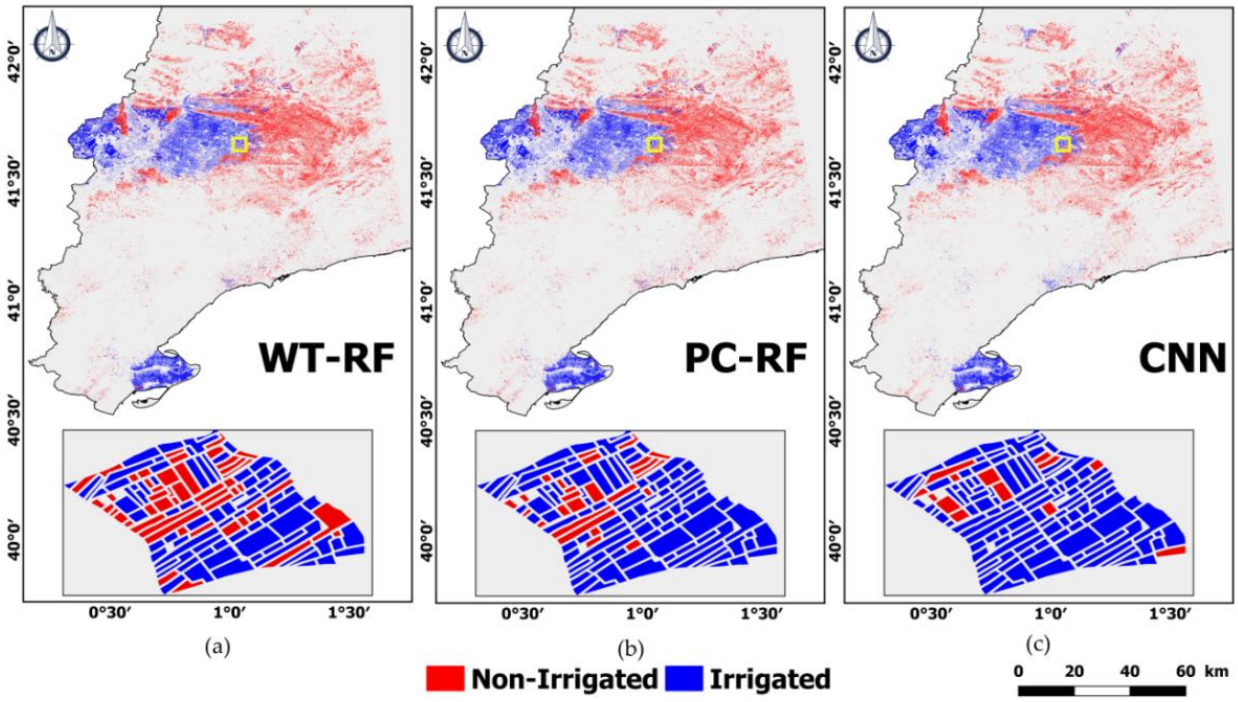


Figure 10: Irrigation mapping using the (**a**) WT-RF model, (**b**) PC-RF model and (**c**) the CNN model. Irrigated areas are presented in blue while non-irrigated areas are shown in red. A zoom version of the yellow box in each map is provided to better visualize different classification results.

5. Discussion

In this study, irrigated areas were mapped over a study site in Catalonia, Spain using Sentinel-1 (S1) SAR time series and optical NDVI time series by applying two classification approaches: Random forest (RF) and a convolutional neural network (CNN). The potential of each data type was investigated, and the combined use of both data types was also presented for both classification approaches.

5.1 Comparison of σ° SAR Backscattering at Plot and Grid Scale

Using S1 data only, the analysis of the σ° SAR temporal series showed that an irrigated plot encounters a frequent change of the σ° SAR backscattering coefficient due to the artificial application of water. This change in the σ° should be differentiated from the change in σ° due to the rainfall events. Thus, the ambiguity between rainfall events and irrigation event was the most challenging point using S1 data. To overcome this uncertainty, the σ° SAR backscattering coefficient at plot scale was compared to that obtained at grid scale (10 km × 10 km). The assumption says that if the mean S1 signal within 10 km × 10 km grid cell, obtained from the bare soil plots with low vegetation cover, increases between two consecutive dates then a rainfall event took place. On the other hand, if σ° value increases at plot scale with stability or

decrease of σ° SAR backscattering coefficient obtained at grid scale then an irrigation event most probably took place. This dependency was clearly shown in Figure 6 for the irrigated and non-irrigated plots. Therefore, the difference between σ° SAR at plot and grid scales helps remove the ambiguity between rainfall and irrigation events.

5.2 Random Forest with PC and WT Transformation

A transformation of the S1 data by the means of principal component analysis and wavelet transformation was applied to the four temporal series of each plot ($\sigma^0_{P,VV}, \sigma^0_{P,VH}, \Delta\sigma^0_{PG,VV}, \Delta\sigma^0_{PG,VH}$). This data transformation of the S1 data showed that using either the PC dimensions or the WT coefficients, the irrigated/non-irrigated plots were classified with high accuracy using a classical random forest classifier. As a classical machine learning method, the RF classifier remains a powerful tool. It is good to mention that adding $\Delta\sigma^0_{PG,VV}$ and $\Delta\sigma^0_{PG,VH}$ remarkably improved our classification accuracy where the overall accuracy increased by more than 15%. This enhancement confirms the relevance of our assumption of using conjointly σ^0_P to σ^0_{PG} to remove the rainfall-irrigation ambiguity for better detection of irrigation. However, in the performed RF classification, the majority of the irrigated plots misclassified as non-irrigated plots belong to winter cereals agricultural including wheat, barley and oat. To fully understand whether those plots were actually irrigated or not, the NDVI temporal behavior of these plots was analyzed and compared to similar non-irrigated agricultures. The cycle and maximum values of NDVI did not significantly differ from similar non-irrigated agricultures. Thus, some plots being in an irrigated district could be registered as irrigated but actually the irrigation did not take place. Moreover, if some winter cereal plots (which are usually not irrigated) were exceptionally irrigated and thus misclassified, then this may be due to the fact that these plots are irrigated only very few times during the growing cycle, and these few irrigation activities have occurred at a date far from the date of S1 acquisition. In this case, the soil moisture of a given plot irrigated five or six days before the S1 acquisition date will certainly decrease to reach a level equivalent to that attained before the irrigation. Hajj et al. (2014) who have demonstrated that the radar signal could be used to identify three-day-old irrigated plots, but it could be difficult to detect irrigation event if the irrigation occurs far away from the SAR acquisition time (more than four days).

Moreover, the assessment of the variable importance of the PC dimensions and the WT coefficients allowed us to significantly reduce the number of variables in the RF approach. From the results obtained, only 15 out of 328 variables were conserved in the PC-RF and 18

out of 256 variables were kept in the WT-RF. Using the conserved important variables in both methods, the accuracy slightly decreased compared to that obtained using all variables (in PC-RF the OA decreased by 0.5% while in WT the OA decreased by 2.3%). This reduction of variables' numbers allows reducing considerably the training time of the model.

5.3 CNN Approach

Another approach tested was the convolutional neural network (CNN). The CNN applied directly to the four S1 derived temporal series showed a remarkable capability to classify irrigated/non-irrigated areas. It is clearly observed that the CNN-based classification approach has a better performance than the classical RF machine learning using either the PC dimensions or WT coefficients. The CNN approach shows higher overall accuracy (94.1%) than the PC-RF (90.7%) and the WT-RF (89.1%). In terms of the F-score and the Kappa coefficient, the CNN is also superior to the RF classifier. The gain in the performance offered by the CNN is clearly visible on the irrigated class where the precision, recall and F-score are higher for CNN than that for the RF classifier. For the non-irrigated class, the obtained accuracies by the two approaches are quite similar. Thus, the increase of the overall accuracy when using the CNN approach is mainly caused by better detection of irrigated plots.

5.4 Inter-Comparison and Quality Assessment

An inter-comparison was performed between the maps obtained from the PC and that from the CNN in the three different scenarios: SAR only, optical only and combined SAR and optical. Figure 11 summarizes the accuracy indices (OA, Kappa, and F-score) obtained for the RF and CNN classifiers in the three different scenarios. The comparison between different classifications was first evaluated through the obtained accuracy indices. Moreover, to properly assess the performance of the tested machine learning algorithms (CNN and RF), the significance of the overall accuracy between classifications was calculated using the McNemar statistical test (McNemar, 1947; Dietterich, 1998). Several studies have reported the use of the McNemar test to compare between two classification approaches (Georganos et al., 2018; Hidayat et al., 2018; Wang et al., 2018). In this test, a null hypothesis that there is no significant difference between OA values of the two compared classifications is proposed. The McNemar test will reject this null hypothesis if the calculated probability (*p*-value) is less than 0.05 (i.e., considering 95% confidence level). In this case, the two classification schemes show a statistically significant difference.

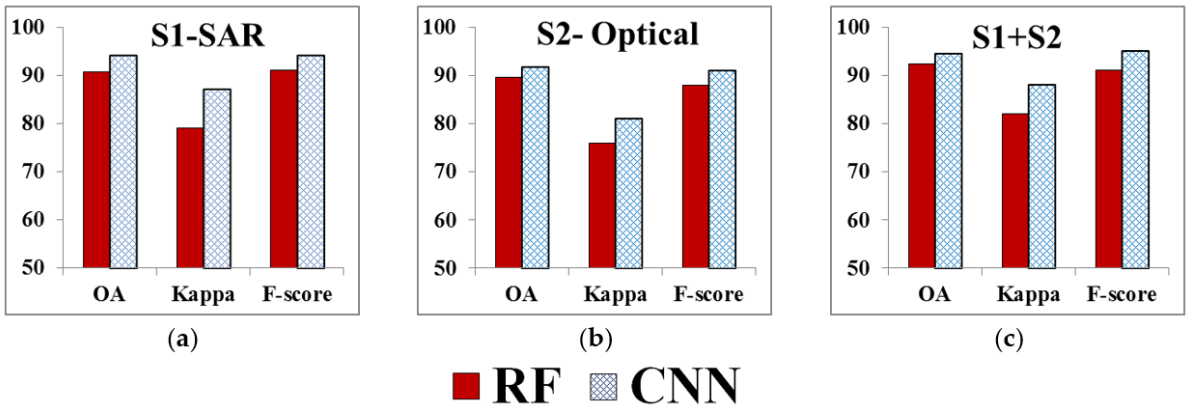


Figure 11: Comparison of accuracy indices between RF and CNN classifications in three different scenarios: (a) Using the S1 SAR data, (b) using S2 optical data and (c) using S1 SAR and S2 optical data.

Using SAR data only, the cross-comparison showed a 92% of agreement between both maps. This means that 92.2% of the plots were commonly correctly classified using both models. Among this percentage, 95.5% of the plots being non-irrigated were classified as non-irrigated plots while 88.0% of the plots registered as irrigated are correctly classified as irrigated in both models. The CNN model was able to correctly classify 4317 plots (forming 6.4% of the total number of irrigated plots) as irrigated which were not classified using the RF approach. Among these plots, 43% are irrigated cereals plots including wheat, barley and oat. The CNN thus was able to improve the accuracy for irrigated cereals compared to the RF classifier. This improvement was clearly visible in terms of the precision, recall and F-score of the irrigated class in the CNN compared to that in the RF-classifier. Moreover, the McNemar test reveals that the two classification approaches using SAR data have a significant statistical difference with *p*-value less than 0.05.

The use of the NDVI temporal series only shows approximately similar performance between the RF classifier and the CNN in terms of overall accuracy (89.5% and 91.6% for RF and CNN respectively). However, the inter-comparison between the classifications' results using NDVI time series shows that 90.2% of the plots were commonly correctly classified using both models. Among this percentage, 96.0% of the non-irrigated plots were correctly classified in both maps while 80.0% of the irrigated plots were commonly classified as irrigated plots. In terms of *p*-value obtained from the McNemar test, the results also show that the two approaches have significant statistical difference with a p-value < 0.05. Finally, it is good to mention that, compared to the use of S1 data, the use of the NDVI data performed well in classifying irrigated areas in both approaches. As irrigation in the region usually takes place between May and

September, the probability of a plot being irrigated increases when a growing cycle exists in summer season.

The combined use of optical and radar data in the RF classifier has slightly increased the classification results compared to that obtained using SAR data only while the results did not significantly change in the CNN approach. The inter-comparison between the results here shows that 94.0% of the plots were commonly correctly classified including 97.1% of the non-irrigated plots and 87.8% of the irrigated plots. The *p*-value obtained from the McNemar test also indicates that the classification approaches have significant statistical difference.

Moreover, to properly assess the quality of the obtained results, the proposed method was compared with the method recently adopted by Gao et al. (2018). In their study, they mapped irrigated areas in a region (20 km × 20 km) located in Urgell Catalonia, Spain (contained within our study site) using the SIGPAC data and the S1 multi-temporal SAR data. To map irrigated/non-irrigated plots, Gao et al. (2018) used the statistical metrics including the mean, variance and correlation length of S1 SAR time series obtained at each plot in VV and VH polarization. They obtained an overall accuracy of 81.1 % using the SVM (support vector machine) and similar accuracy using the RF classifier. For the study site examined in this study, the approach proposed by Gao et al. (2018) was tested using the RF classifier. For each plot, the same metrics were calculated: The mean of the SAR time series, the variance and the correlation length of the SAR signal in both VV and VH polarizations. Results showed low accuracy compared to our proposed RF classifiers. The overall accuracy obtained was 80.1% with accuracy of irrigated class reaching 60.8% only while that of the non-irrigated class reaching 90.3%. Moreover, the addition of these metrics (mean, variance and correlation length) to our RF classifier of PC variables or WT coefficients did not enhance the accuracy of the classification obtained.

5.5 Strength, Limitations and Future Directions

The S1 satellite with the 6 days revisit time allows more precise temporal follow-up of agricultural crops. S1 is now the only operational satellite system providing dense time series in free and open access mode with global coverage. Thus, any method based on the use of S1 time series could interest a wide range of end users. Moreover, the advantage of using S1 temporal series rather than the optical time series is that radar sensors are not limited to good weather conditions. The cloud cover, that may be present in the optical images of certain

geographical regions, could be a drawback when working with optical satellites. Even though the combined use of optical and radar data did not highly enhance the classification results, the synergic use of microwave and optical data could increase the possibility of transferring the model across different regions.

The six days temporal resolution of the S1 satellite allows the detection of irrigation events especially when the irrigation frequency is high. As shown by Hajj et al. (2014), irrigation could be still detected even if the irrigation took place three days before the radar acquisition. On the other hand, the detection of irrigation becomes difficult if the irrigation event takes place five or six days before the SAR acquisition. However, the effect of irrigation on the radar backscattering can change depending on the irrigation technique used (inundation or sprinkler) and the frequency of irrigation (weekly or daily).

Although the RF classifier was less accurate than the deep learning model, the RF classifier remains quite precise and competitive. With the dense time series now available, the comparison between classical and advanced machine learning techniques allows us to study the importance of both techniques to intelligently exploit the temporal behavior of SAR signal in the multi-temporal remote sensing data.

The proposed model was trained on Catalonia region of Spain which is a Mediterranean climatic region. Other regions may have different climatic and soil conditions which could lead to unsatisfactory results when applying the trained model. Hence, for future studies, including other climatic regions in the training process, we would make the proposed model more generic. However, with an accuracy exceeding 90%, the trained model could be applied across regions with close climatic conditions while maintaining good classification results. Currently, studies are concentrating on region adaptation and transfer learning which allow the use of one model on several regions. Therefore, future work should concentrate on adapting the proposed model to use it over other study sites. Moreover, future work should also concentrate on building effective unsupervised approaches that do not require any training procedure. In this case, the transferability of the approach would be easier.

6. Conclusions

In this paper, a new methodology for mapping irrigated areas using Sentinel-1 time series was introduced. First, a temporal series of SAR backscattering coefficients from the S1 data was obtained at plot scale in VV and VH polarization. To remove the ambiguity between

rainfall and irrigation events, the S1 signal at plot scale was compared to that obtained at grid scale (10 km × 10 km). Then, two different transformations including the PCA and the WT were applied to both the S1 temporal series at plot scale (σ^0_P) and the temporal series of the difference between plot and grid ($\Delta\sigma^0_{PG}$). Using the PC dimensions and the WT coefficients, irrigated areas were classified in two different random forest classifiers. The results show a good overall accuracy (OA) for the RF classifier using the PC values (90.7%) and the WT coefficients (89.1%). Moreover, the PC transformation was applied on the NDVI time series obtained from Sentinel-2 optical images. Results showed that the RF classifier using optical data (NDVI) performs well with OA = 89.5%. The combined use of optical and SAR data (in PC values) slightly improved the classification accuracy (OA = 92.3%).

Another approach tested in this study was the one-dimensional convolutional neural network (CNN). The convolutional neural network was applied to the S1 time series obtained at plot scale σ^0_P and difference between plot and grid scales ($\Delta\sigma^0_{PG}$). The results of the validation of CNN approach showed very high accuracy (OAm = 94.1%) compared to that obtained using the RF classifiers. The accuracy of the irrigated class increased when using the CNN approach thus allowing better detection of irrigated plots. The use of NDVI data only in a CNN classifier produced lower overall accuracy (91.5%) than that obtained using S1 data only. However, the combined use of both data types in the CNN did not significantly improve the accuracy.

The proposed approach should be applied to other agricultural areas to better assess its relevance and possible usage in operational mode. Given the very good accuracy obtained and the fact that S1 data is free and open access, the use of Sentinel-1 data is relevant. Even though the optical data presents good results, its use could be problematic in certain geographical areas due to the presence of cloud cover. However, machine learning algorithms including classical or advanced approaches still require training data to obtain good classification results. Thus, future work should also focus on building effective approaches to detect irrigation using less training data.

Author Contributions: H.B. (Hassan Bazzi) and N.B. conceived and designed the experiments; H.B. (Hassan Bazzi) and D.I. performed the experiments; H.B. (Hassan Bazzi) and N.B. analyzed the results; H.B. (Hassan Bazzi) wrote the article; N.B., D.I., M.E.H., M.Z., H.B. (Hatem Belhouchette), M.J.E., and V.D. revised the paper.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2019 project), and the National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA).

Acknowledgments: The authors wish to thank the French Space Study Center (TOSCA 2019) and the National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA). The authors wish also to thank the European Space Agency (ESA) for the Sentinel-1 and Sentinel-2 data and Theia pole for the calibration of the Sentinel-2 images. Finally, authors would like to thank the General Direction of Rural Development of the Department of Agriculture, Livestock, Fishing and Food of the Generalitat of Catalunya for providing the SIGPAC data.

Conflicts of Interest: The authors declare no conflicts of interest.

3. Article 4: Distilling before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas using Sentinel-1 Time Series

Author version of the article published in IEEE GRSL: Volume: 17, Issue: 11, November 2020

Hassan Bazzi[1], Dino Ienco[1], Nicolas Baghdadi[1], Mehrez Zribi[2] and Valerie Demarez[2]

[1] INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 29 October 2019; Accepted: 15 December 2019; Published: 07 January 2020

Abstract

This letter proposes a deep learning model to deal with the spatial transfer challenge for the mapping of irrigated areas through the analysis of Sentinel-1 data. First, a convolutional neural network (CNN) model called "Teacher Model" is trained on a source geographical area characterized by a huge volume of samples. Then, this model is transferred from the source area to the target area characterized by a limited number of samples. The transfer learning framework is based on a *distill and refine* strategy, in which the teacher model is first distilled into a student model and, successively, refined by data samples coming from the target geographical area. The proposed strategy is compared with different approaches including a random forest (RF) classifier trained on the target data set and a CNN trained on the source data set and directly applied on the target area as well as several CNN classifiers trained on the target data set. The evaluation of the performed transfer strategy shows that the "distill and refine" framework obtains the best performance compared with other competing approaches. The obtained findings represent a first step toward the understanding of the spatial transferability of deep learning models in the Earth observation domain.

***Index Terms*—**Deep learning, knowledge distillation, satellite image time series, Sentinel-1 (S1), transfer learning.

1. Introduction

Irrigation, nowadays, plays a significant role in agricultural production in order to meet the global food requirement (Cai and Rosegrant, 2002). Due to this fact, a better management of irrigation policies is required to deal with the high demand of food with the increase in the global population (Godfray et al., 2010). To support such policies, accurate information on the irrigated area extent are essential to manage water resources or evaluate irrigation water requirements. Unfortunately, the extent and distribution of irrigated areas remain indefinite and the large scale mapping of such property remains a challenge for modern remote sensing analysis. Recent works have pointed out that SAR (Synthetic Aperture Radar) signal seems to be more suitable to map different agricultural irrigated areas (Gao et al., 2018). The main assumption is related to the fact that radar signal is sensitive to the water content of soil due to a positive correlation between the dielectric constant and the soil water content. Following this direction, time series of SAR information acquired via the recently Sentinel-1 (S1) SAR constellation provides an effective tool for large-scale area mapping and monitoring due to their high revisit period (six days revisit time). To perform such mapping, machine (and deep) learning based techniques are becoming the standard tools since they allow large-scale analysis and they provide acceptable results (Pelletier et al., 2019; Zhu et al., 2017). One of the key question about the adoption of machine learning based solution is related to their ability, given a particular task, to be spatially transferred from a geographical area to another spatially uncorrelated one.

Considering the task of mapping irrigated areas (Bazzi et al., 2019c), the challenge is to build a predictive model from a rich set of labelled samples available on a particular area and, successively, adapt this model on another geographical area, with a limited set of labelled samples. This adaption could be crucial to conceive large-scale monitoring systems in operational scenario.

With the aim to tackle this issue, we propose a deep learning (Zhu et al., 2017) framework capable to adapt a predictive model trained from a source dataset (characterized by a large volume of labeled samples) to deal with the irrigation mapping on a target dataset (characterized by a limited number of labeled samples). To this end, we propose a distill before refine framework in which first a teacher model is learned on a source dataset, then a smaller student model is distilled from the teacher model and finally the student model is fine-tuned to deal with the classification task on the target dataset.

Here, the challenge behind our framework is that models with a huge number of parameters are difficult to fine-tune (refine) on classification tasks where a very limited number of labeled samples are available.

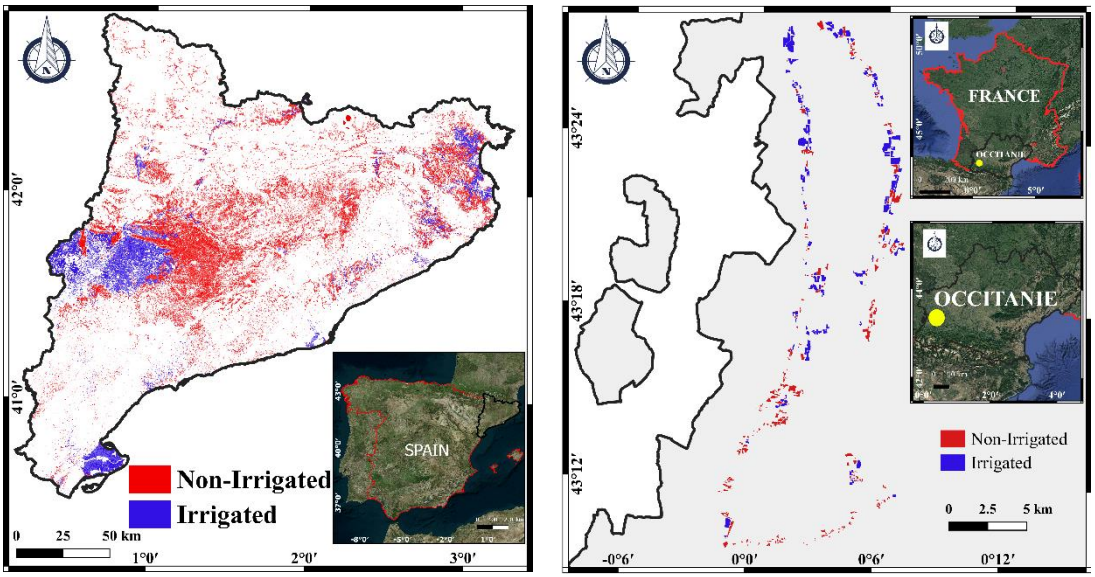
The innovative component of our framework is associated to the use of Knowledge Distillation (Hinton et al., 2015) to distill a new smaller model that can be easily fit with new limited labelled samples on the target study area. In the general context of fine-tuning approaches (Yin et al., 2019; Zhou et al., 2017), the model learnt on the source data is directly fine-tuned on the target data without considering the issues related to scarce or limited information. To the best of our knowledge, we are not aware about any other research work that adopts a similar pipeline to deal with spatial transfer learning in the field of remote sensing.

The proposed methodology is implemented on mapping irrigated areas in which the classification problem involves Irrigated vs. Non Irrigated area detection. The experiments are conducted considering a source study area from Catalonia region of north east Spain where a time-consuming and cost-intensive field campaign were conducted collecting around 193,000 labelled samples and a target study area, from West Occitanie region of South France, where the reference data is constituted of less than 500 examples.

2. Data

In this section, we introduce the geographical areas involved in this study: the source dataset over the Catalonia region and the target dataset over the West Occitanie (South of France) region. We also introduce the Sentinel-1 SAR satellite images involved in this study and detail the different pre-processing steps performed to obtain the Sentinel-1 time series backscattering coefficients at plot scale.

We point out that, in terms of climate, both studied zones (Catalonia and West Occitanie) are extremely different. In fact, the climate of the Catalonia region is typically Mediterranean where the average annual precipitation is around 376 mm. On the other hand, the climate in West Occitanie is humid to oceanic with average annual precipitation of 1200 mm. In both regions, irrigation mainly occurs in the summer season between May and October of each year. However, the summer season in Catalonia is very dry with rare rainfall events whereas the summer season in West Occitanie is more humid with average precipitation of 300 mm between May and October.



(a) (b)

Figure 1: Location of the source and the target study sites. (a) Catalonia region of North East Spain. (b) Western Occitanie region of South France

Catalonia dataset

Over the Catalonia region of northeast Spain (Figure 1a), the SIGPAC (Geographic Information System for Agricultural Parcels) data are provided by the General Direction of Rural Development of the Generalitat of Catalonia. The SIGPAC data represents the agricultural plots digitized using ortho-photo images of 25-cm spatial resolution at scale 1:2000. Each plot in the provided dataset is identified by a unique identification code, surface area, land use and an irrigation coefficient (0 for non-irrigated and 100 for irrigated). Each year, an update of the database is provided based on an annual large field campaign in order to maintain the credibility of the dataset (mainly irrigation information and crop type). In our study, a total of 193 000 plots covering an area of 3795 km^2 of different crop types and irrigation management have been used to develop the later called teacher model. Among different land cover types, only agricultural crops (summer and winter crops) were considered for the irrigation classification. Forests, urban, and orchards plots were eliminated. In general, winter cereals such as wheat, oat and barely are rarely irrigated with some exceptions. On the other hand, irrigated plots mainly include alfalfa, maize, grassland, beans, rapeseed and Rice. Among the total number of plots, 126 000 are non-irrigated whereas 67 000 plots are irrigated.

West Occitanie dataset

Over the western part of the Occitanie region of south France (Figure 1b), a terrain campaign was conducted over different summer crop plots in 2017. In this field campaign, the existence

or absence of irrigation activity over the summer crops was registered for each visited plot. The land cover of the plot was also recorded. Finally, a total of 451 plots including 300 non-irrigated and 151 irrigated plots were registered for the agricultural season of summer 2017. The dominant crop types of the collected plots are maize and soybeans.

Sentinel-1 Time Series

Over the Catalonia region, a total of 82 C-band (5.405 Ghz) SAR images acquired by S-1A and S-1B satellites were used for the period between September 2017 and December 2018. Over the West Occitanie region, the collected irrigation information correspond to the year 2017. Thus, the same number of images (82 images) were collected over this region for the period between September 2016 and December 2017. The 164 images (82 for Catalonia and 82 for Occitanie) are acquired in the Interferometric Wide (IW) mode in both VV and VH polarizations. These S1 images are of Level-1 Ground Range Detected (GRD) product with a pixel spacing of 10 m x 10 m and six days revisit time. The Sentinel-1 toolbox (https://step.esa.int/main/toolboxes/snap) (S1TBX) developed by the European Spatial Agency (ESA) was used to calibrate the S1 images. This calibration converts the digital number of downloaded SAR images into backscattering coefficients in linear units and ortho-rectifies the images using a 30-m digital elevation model of the Shuttle Radar Topography Mission (SRTM). The temporal series of the S1 SAR backscattering coefficients over each agricultural plot in each region was then obtained by averaging the σ^0 values of all pixels within each plot at each available date and at both VV and VH polarizations.

3. Contributions

In this section, we describe our Distill before refine framework devoted to deal with spatio-temporal transfer learning over different geographical areas. Conversely to other scenarios in which the transfer is made between different classification problems or considering different type of data (Huang et al., 2017; Zhao et al., 2017) the process here focuses on the transfer between two distinct geographical areas fixing the classification problem (irrigation mapping) as well as the type of considered data (Sentinel-1 time series data). Figure 2 visually depicts the different steps of our proposed pipeline.

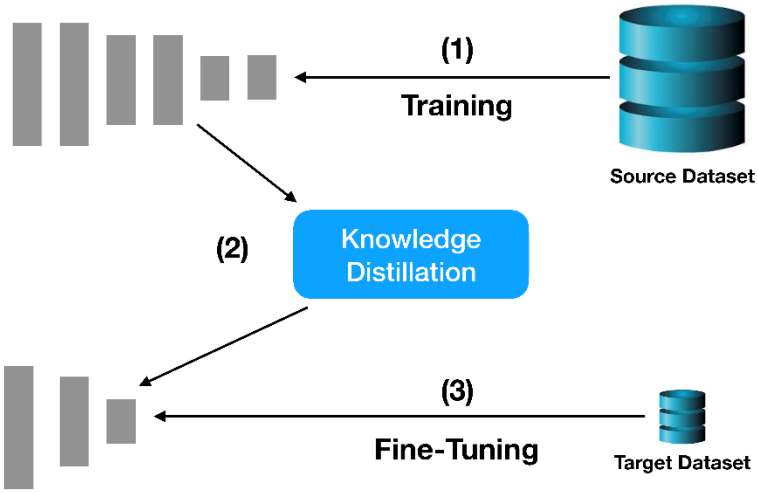


Figure 2: The overview of our distill and refine pipeline. (1) A complex deep learning approach (teacher) is learned on a certain study area involving a large volume of labeled samples. (2) A smaller (student) model is distilled from the teacher one. (3) The student model is finally fine-tuned with the limited labeled samples available on the target study area.

Firstly, a model is trained from scratch on the source dataset that is characterized by a large volume of labelled samples (Step (1) in Figure 2). Such model, considering knowledge distillation literature (Hinton et al., 2015; Shi et al., 2019), is commonly denoted as teacher model. Secondly, a lighter model (commonly named student) is distilled from the teacher model (Step (2) in Figure 2). The distillation step allows to transfer the knowledge from the teacher to the student model. More in details, during this step, we are trying to synthesize the teacher behavior in a usually smaller network that should behave similarly to the bigger one. Once the knowledge is distilled in the student network, such model is finally fine-tuned considering the limited labeled samples coming from the target study area (Step (3) in Figure 2).

The main ratio behind our framework is the follows: models involving a huge number of parameters such as the teacher model can hardly be fine-tuned on a target dataset characterized by a scarce number of labelled samples. Conversely, smaller and lighter models such as the student one, that behave similar to bigger ones, can be adapted more easily in the presence of scarce training data since they involve a smaller number of parameters to modify.

Considering the classical framework of knowledge distillation (Hinton et al., 2015), given a teacher T and a student S networks, the main objective is to distill the knowledge of T inside the network S. Usually, the common assumption is that the network T is much bigger than the network S in terms of parameters.

To deal with the knowledge distillation task, (Hinton et al., 2015) proposes to learn the student network considering the following loss function L:

$$L = \frac{1}{|X|} \sum_{x_i \in X} \alpha * L_1 + (1 - \alpha) * L_2$$
$$L_1 = L_{CE}\left(Softmax\left(f_s(x_i)\right), y_i\right) \quad (1)$$
$$L_2 = KL\left(Softmax\left(\frac{f_s(x_i)}{\tau}\right), Softmax\left(\frac{f_T(x_i)}{\tau}\right)\right)$$

where the main loss (L) is a linear combination of two other task-specific losses (L_1 and L_2) and the linear combination is weighted by the parameter α. The first task-specific loss (L_1) is the classical cross-entropy loss usually employed in multi-class classification task. The second loss L_2 is the Kullback-Leibler divergence between the predicted output distribution of the teacher and the student $x_i \in X$ (resp y_i) is an example of the dataset (resp. the associated label). In Hinton et al. (2015) the authors employ a temperature scaling factor τ to smooth the last output layer of the neural models before perform the softmax normalization. We remind that $f_T(.)$ and $f_S(.)$are the pre-softmax outputs for the teacher and student model, respectively (the Softmax (https://en.wikipedia.org/wiki/Softmax_function) operator is a standard component of modern neural network classification model). The objective of the loss L_2 is to force the student model to simulate the output of the teacher model with the aim to distill the teacher behavior into the student network.

3.1 Teacher and Student Model Implementation

In our case, both the teacher and the student models are implemented as Convolutional Neural Networks (CNNs). Table 1 resumes the architecture of the teacher and student CNNs. Each convolutional operation (*Conv*) is associated with a successive Batch Normalization and Dropout layer. The convolution takes as input the number of filters (*nf*), the kernel size (*k*), the stride (*s*) and the activation (*act*).

Our CNN follows the idea of temporal CNN introduced in (Pelletier et al., 2019)where the convolution is performed on the time dimension. Considering the Fully Connected layer (*FC*), we apply Batch Normalization and Dropout except for the output layer. The *FC* layer takes as inputs the number of neurons *nn* and the activation function *act*. For all the layers we adopt the Rectifier Linear Unit (Nair and Hinton, 2010) except for the last output layer that is associated to a linear function followed by the Softmax operator.

Comparing the teacher and the student models, we can observe that the former is deeper than the latter (11 vs. 6 layers) and the teacher model involves more than the double of

parameters of the student model (about 1.69M vs. 0.69M). Considering the student network, we eliminate all the fully connected layers. Moreover, the Global Average Pooling (GAP) layer, at the end of the convolutional layers, is directly connected with the output layer. The GAP layer aggregates each feature maps via the average operator producing a layer with as many neurons as the number of the feature maps at the precedent step.

Table 1: Architectures of the Teacher and Student Convolutional Neural Networks

Teacher Model	**Student Model**
Conv(nf=64, k=7x1, s=1, act=ReLU)	Conv(nf=96, k=7x1, s=1, act=ReLU)
Conv(nf=64, k=5x1, s=2, act=ReLU)	Conv(nf=96, k=5x1, s=2, act=ReLU)
Conv(nf=128, k=5x1, s=1, act=ReLU)	Conv(nf=192, k=5x1, s=1, act=ReLU)
Conv(nf=128, k=5x1, s=2, act=ReLU)	Conv(nf=192, k=5x1, s=1, act=ReLU)
Conv(nf=256, k=5x1, s=1, act=ReLU)	Conv(nf=384, k=5x1, s=1, act=ReLU)
Conv(nf=256, k=5x1, s=1, act=ReLU)	GlobalAveragePooling
Conv(nf=256, k=3x1, s=1, act=ReLU)	FC(nn=2, act=None)
Conv(nf=256, k=1x1, s=1, act=ReLU)	
Flatten	
FC(nn=512, act=ReLU)	
FC(nn=512, act=ReLU)	
FC(nn=2, act=Linear)	
Number Parameters	**Number Parameters**
1 693 570	697 538

4. Experiments

In this section we evaluate our *distill before refine* framework in the context of irrigation area mapping using Sentinel-1 time series data and the datasets introduced in Section 2. We refer to the Catalonia dataset as source dataset and the West Occitanie dataset as a target dataset. We compare the performance of our proposal with respect to several competing methods to pinpoint the benefits of compressing the deep learning model before performing fine-tuning on a study area characterized by a limited number of samples.

4.1 Experimental Settings

To assess the performances of the different methods we consider two standard metrics: Accuracy and F-Measure (Tan et al., 2006). While the former is a standard metric in remote sensing, the latter corresponds to the harmonic mean between precision and recall and it is well suited to evaluate classification performances in unbalanced scenario. Considering the Teacher CNN model (trained on the Catalonia region characterized by a large volume of labeled samples), we consider 80% of the dataset as training data while the rest as validation set to choose the best model. More in detail, the model that achieves the best accuracy on the validation set is the one that is retained for the subsequent steps.

Considering the fine-tuned step (trained on the French area characterized by a limited number of labeled samples), we adopt the following settings: we employ 33% (150 samples) of the data as test set, 20% (60 samples) of the data as validation set while we use the rest of the samples as training data (240 samples). Regarding the training set, given the validation and test set, we evaluate different models considering different amounts of training samples: {60,120,180,240}. In this way, we can study the behavior of the different approaches considering a varying amount of training data to feed the learning process. For each method, we repeat the procedure 10 times and we report the averaged value for each metric. For the different competing approaches, we involve: a Random Forest classifier learned directly on the target study area (Occitanie), we name such method *RF*; a CNN model with the same configuration of the teacher network (resp. the same configuration of the student network) learned from scratch directly on the target study area named CNN_T (resp. CNN_S); and the teacher model directly fined tuned on the target study area, we name such method CNN_{trans}. The proposed distill before refine framework is named CNN_{KD}. For the *RF* models, we optimize the model via the maximum depth of each tree (in the range {20, 40, 60, 80, 100}) and the number of trees in the forest (in the set {100, 200, 300, 400, 500}). For the CNN models we use Adam to optimize the parameters weights with a learning rate of 1×10^{-4}. CNNs learned on the source dataset (Catalonia) are trained for 1000 epochs with a batch size equals to 256 while the CNN models trained (or fine-tuned) on the target dataset (West Occitanie), are trained for 500 epochs with a batch size equals to 2. For the CNN_{KD} method, we empirically set the α and τ values to 0.5 and 1, respectively.

4.2 Results

Figures 3a and 3b resume the Accuracy and F-Measure results, respectively. Considering both metrics, the *RF* methods is quite stable over the considered range. No bigger difference can be noted between the smallest amount of training samples (60) and the biggest one (240). Regarding the CNN_T and CNN_S methods trained from scratch on the target dataset, as we can expect, they fail to be competitive considering small amount of training data. We can observe that when 60 labeled samples are considered, the models performances are really poor compared to the competitors. On the other hand, they achieve similar performances to the *RF* method when bigger amounts of training samples are considered. Inspecting the performance of the teacher model fine-tuned on the target study area (CNN_{trans}) it exhibits good performances starting from a training size of labeled examples equals to 120 but it does not show any successive improvement for larger training size. This is probably due to the fact that this deep learning model would need more labeled information to effectively modify the huge number of parameters it involves. Such behavior is highlighted regarding both evaluation metrics.

Generally, we can observe that the CNN_{KD} obtains best (or comparable) results for all the considered amount of training samples regarding both Accuracy and F-Measure. Such difference is clearly visible when the number of training samples is greater or equals to 180. This behavior supports our *distill before refine* framework in which a smaller network distilled from a bigger one will be easier to adapt to a target domain characterized by a very limited number of labeled samples. This result constitutes a preliminary experiment in which the benefit of knowledge distillation is assessed to perform transfer learning from a source dataset (a geographical area) to a target dataset (another geographical area).

Both Accuracy and F-Measure depicts almost the same behaviors among the different competing methods. The only point that changes between the evaluation of the two metrics is related to the comparison between the two *CNN* models trained from scratch (CNN_T and CNN_S) and the *RF* method on the range 120-240. Here we can observe opposite behaviour. Regarding the accuracy, it seems that the *RF* model slightly outperform the two *CNN* models while the opposite happens when F-Measure is considered. This is due to the fact that the test dataset, as the whole target dataset, is unbalanced with respect to the involved classes. The *RF* method is biased towards over-represented classes and this is why it achieves best (resp. worst) performances in terms of Accuracy (resp. F-Measure). On the other hand, the two *CNN* models

trained from scratch deals better with the unbalanced scenario achieving slightly better performance in terms of F-Measure since they are less biased towards over-represented classes. As additional test, we also evaluate the performance of the teacher model without any additional fine-tuning directly on the test set of the target dataset. In this experiment, we obtain a value of Accuracy and F-measure equal to 27.5% and 36.86%, respectively. Considering that we are dealing with a binary task, such performances are lower than the average random performances we can obtain on such task. This poor behavior indicates that the data distribution associated to the source (Catalonia area) and target (Occitanie area) datasets are really different and a heavy distribution shift exists. Such result confirms once more that the necessity to study and develop new machine learning approaches to deal with shifts in data distribution (due to spatial or temporal autocorrelation) to cope with spatio-temporal model transfer for remote sensing data.

With the increasing availability of remote sensing data, coming from large scale monitoring systems (i.e. Copernicus data), issues related to transfer models learned from a particular area (resp. in a particular time period) to deal with data coming from another geographical area (resp. another time period) will get more and more attention in a near future and, probably, greater effort will be made in this direction.

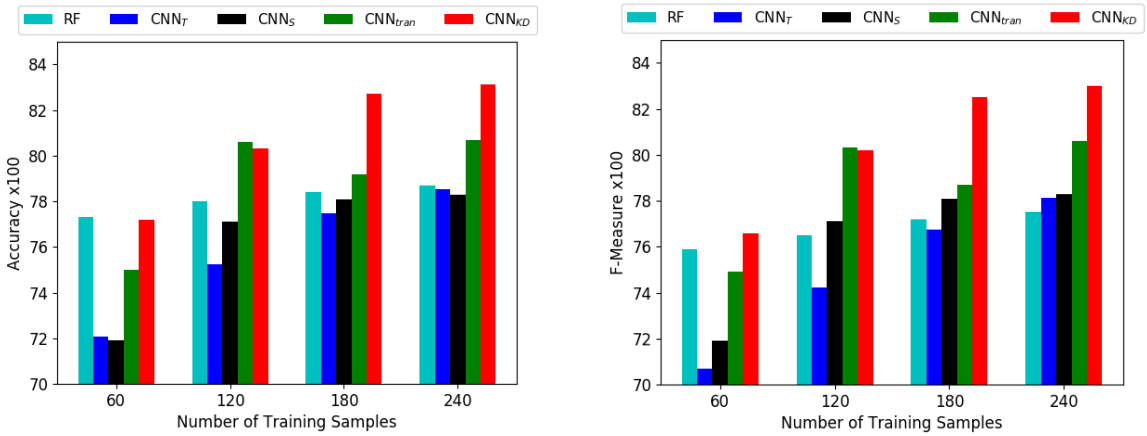


Figure 3: (a) Accuracy and (b) F-Measure results considering the different competing methods varying the amount of (target) train samples to learn the model on the West Occitanie dataset.

5. Conclusions

We have introduced a new *distill before refine* framework to deal with spatial transfer between two geographical areas coming from different countries. We have deployed the proposed framework in the context of irrigation mapping leveraging Sentinel-1 time series data. Experimental evaluations have underline the effectiveness of the proposed framework in the

context of irrigated area mapping. As future work, we plan to investigate more precisely the influence of the α and τ hyper-parameters in the knowledge distillation phase of our framework.

Acknowledgments

This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004 (DigitAg). The authors which to thank the CNES for financing this project in the frame of TOSCA program. The authors would like to thank Maria Jose Escorihuel from the isardSAT, Catalunya for providing the SIGPAC data. Authors would also like to thank Fabien Dauriac of the Chamber of Agriculture of Hautes-Pyrénées, France.

CHAPTER IV: TOWARDS OPERATIONAL MAPPING OF IRRIGATED AREAS

1. Analytical Summary

1.1 Overview

The use of S1 and S2 data in supervised classification approaches is capable of providing good accuracy for mapping irrigated areas. However, the dependency of supervised classifiers on terrain data remains an obstacle for operational mapping of irrigated areas. In addition, the detection of irrigation events at plot scale using remote sensing data did not yet receive important attention despite its vital need in irrigation water management. Before the arrival of the S1 and S2 satellites, previous satellite sensors did not have the sufficient temporal resolution required for irrigation detection. Indeed, before S1 and S2, most of the sensors had a revisit time of 2 weeks at best. Now, S1 and S2 satellites provides SAR and optical images with a revisit time better than one week, which could be adequate for irrigation detection at plot scale.

In this chapter, a near real-time operational model for detecting irrigation events at plot scale (IEDM) is first proposed. The IEDM is based on a decision tree-based method built using S1, S2 and S^2MP data. It principally relies on the change detection in the S1 backscattering coefficients following irrigation events at plot scale. To build the IEDM, irrigated and non-irrigated plots in three study sites with different climatic properties were examined: Montpellier, Catalonia and Adour Amont watershed of southwest Occitanie. The behavior of the S1 backscattering coefficients following irrigation events was analyzed in the three study sites. In the IEDM, the soil moisture estimation derived from the S^2MP at grid scale (10 km x 10 km) and the S1 backscattering coefficients at grid scale were used to eliminate the uncertainty between rainfall and irrigation at plot scale. Additional filters were also integrated to reduce ambiguities due to vegetation development (phenology stages) and surface roughness using S2 NDVI values. To assess the accuracy of the IEDM for irrigation event detection, the IEDM was tested over a fourth study site located in the Crau plain of southeast France. The IEDM was applied over 46 irrigated grassland plots and the detected irrigation events were compared with in situ registered irrigation timing.

Then, the IEDM was integrated into an operational framework for mapping irrigated areas at plot scale. To insure the operational mapping of irrigated areas, a method capable of generating its own terrain training data is required. To generate automatically the training data (irrigated and non-irrigated plots), SAR and optical derived metrics were proposed. First, the IEDM was applied over all agricultural plots. Plots with no detected irrigation events are considered potentially non-irrigated whereas plots with very high number of detected irrigation events are considered as irrigated plots. From the selected plots using the IEDM, additional filter using the maximum NDVI value from S2 images was applied to select the final training dataset of irrigated/non-irrigated plots. Using the selected training dataset as well as S1 and S2 temporal series, random forest classifier is constructed. The proposed framework was tested over a study site near Orléans city (northcentral France) for four distinct years. Figure xx presents a graphical abstract for the development of the IEDM and its integration in the operational mapping of irrigated areas.

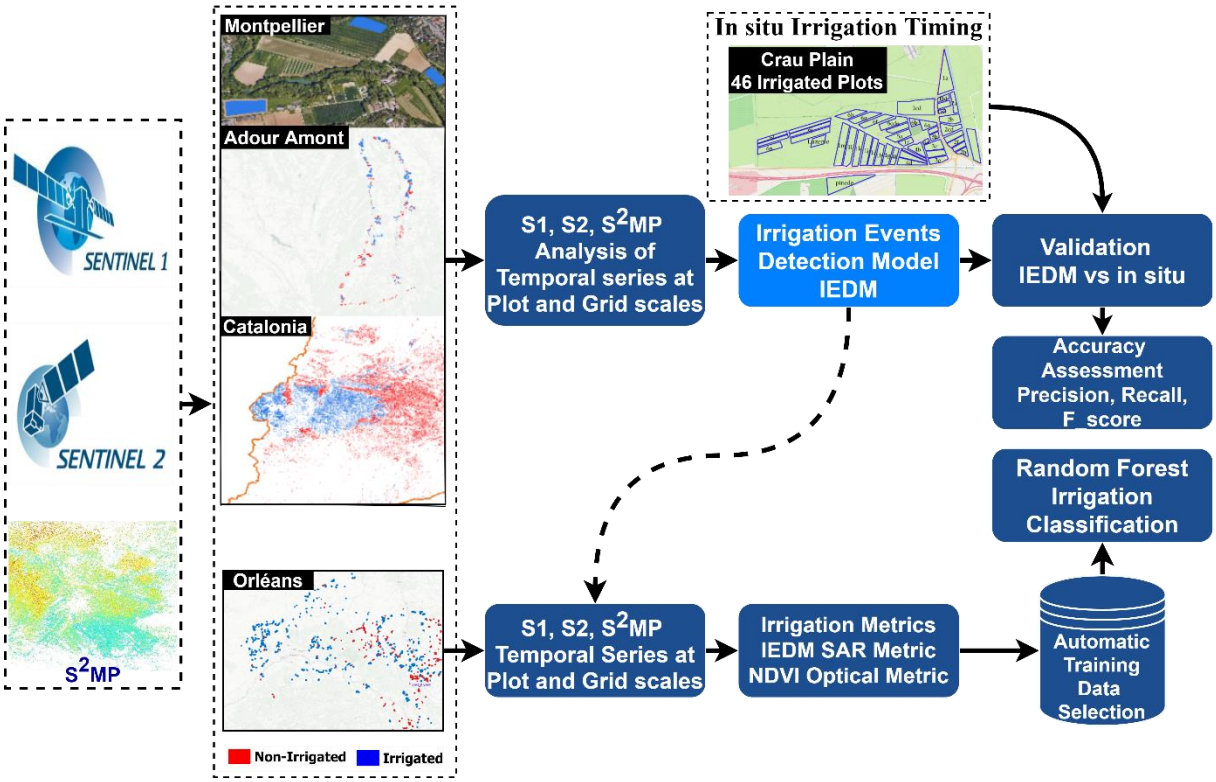


Figure C: Graphical abstract representing a general overview of the IEDM development and its use in the operational mapping of irrigated areas

1.2 Materials and Methods

1.2.1 Study sites and reference data

In this chapter, five different study sites were examined. First, three irrigated plots in Montpellier (south France) were examined in the IEDM development. For these three plots, exact irrigation dates are registered for the period between May and October 2017 (48 episodes). These three plots were first used to analyze the effect of irrigation events on the backscattered SAR signal. The SIGPAC data from Catalonia region and the terrain data from Adour Amont (presented previously in Chapter 3) were also used to analyze the performance of the IEDM in irrigation detection. For a robust validation of the IEDM, a fourth study site was examined. In the Crau plain of southeast France (PACA region), 46 irrigated grassland plots were used to validate the accuracy of the IEDM in irrigation event detection. The exact irrigation date at each plot is registered for the period between March and September 2019 (814 irrigation episodes). Irrigation over the 46 plots is performed using the gravity irrigation system. The plots usually encounter three different crop cycles from February to September.

For the operational mapping of irrigated areas, a fifth study site located near Orléans city is examined. In this study sites, several terrain campaigns were conducted to obtain irrigation information (absence or existence of irrigation) for four different years (2017, 2018, 2019 and 2020). Each year, the collected terrain data was used to validate the irrigation mapping performed using the proposed operational irrigation mapping methodology.

1.2.2 Remote sensing data

Over each study site, all possible Sentinel-1 images acquired with different S1 overpasses (ascending and descending) were downloaded and processed. Over Montpellier and Adour Amont sites, 92 images (46 ascending and 46 descending) images were obtained for each site (for the irrigation period between March 2017 and November 2017). For the Catalonia site, 162 images (82 ascending and 82 descending) were used covering a period between September 2017 and December 2018 (period of SIGPAC irrigation information). 180 S1 images were used over the Crau plain between February and September 2019. Finally, 578 S1 images were used for Orléans study site during the four studied years. All possible cloud free S2 images were also downloaded for each study site corresponding to the same period of the S1 acquisition

(frequency of approximately one to two S2 images per month). Moreover, all available S^2MP maps over each study site were used.

1.3 Methodology

1.3.1 Development and assessment of the IEDM

The IEDM is based on detecting the change in the SAR backscattering values at plot scale (σ_P^0) between the SAR acquisition at time t_i and the previous SAR acquisition at time t_{i-1} ($\Delta\sigma_p^0 = \sigma_p^0(t_i) - \sigma_p^0(t_{i-1})$). The IEDM assumes that the increase of the σ_P^0 between two consecutive SAR dates (t_{i-1} and t_i) is mainly caused by the increase of the surface soil moisture due to either rainfall or irrigation event. On the other hand, the increase of the SAR backscattering values at grid scale (σ_G^0) (10 km x 10 km) and high estimated SSM values from S^2MP at grid scale could be an evidence of rainfall event that occurred between the two SAR acquisitions. The stability or decrease of σ_G^0 values could indicate the absence of rainfall between the two dates. Thus, the increase of the σ_P^0 values between consecutive acquisitions accompanied with the stability or decrease of σ_G^0 values is considered as an irrigation event. The IEDM provides at each SAR image, an irrigation possibility weight that represents the chance of having irrigation event. The possibility weight could be either 0 (no irrigation), 25 (low possibility), 50 (medium possibility) or 100 (high possibility):

- The possibility weight 0 represents either a decrease in the σ_P^0 between t_{i-1} and t_i ($\Delta\sigma_P^0 \leq -0.5$ dB) indicating a decrease in soil moisture values (no irrigation possibility) or an important increase of the σ_G^0 ($\Delta\sigma_G^0 \geq 1$ dB) between t_{i-1} and t_i indicating the presence of a rainfall event.
- Low irrigation chance (25) corresponds to a very slight change in the σ_P^0 between t_{i-1} and t_i ($-0.5 \leq \Delta\sigma_P^0 < 0.5$ dB) accompanied with high estimated soil moisture values at plot scale at t_i along with an important decrease of the σ_G^0 at grid scale between t_{i-1} and t_i ($\Delta\sigma_G^0 \leq -1$ dB), indicating the absence of a rainfall event. In this case, the SSM estimations from S^2MP at plot scale are used to guarantee the existence of an irrigation event. The stability or slight increase of the SAR signal at plot scale between t_{i-1} and t_i could be interpreted as an irrigation event (low chance) if the σ_P^0 at time t_{i-1} already attains high values (due to irrigation or rainfall). To ensure this situation, we say that at

time t_{i-1}, SSM estimation should be greater than or equals to 20 vol.% in order to guarantee that a humid soil conditions at time t_{i-1} have continued to time t_i.

- Medium irrigation chance (50) corresponds to a moderate increase in the σ_P^0 between t_{i-1} and t_i ($0.5 \leq \Delta\sigma_P^0 < 1$ dB) accompanied with a decrease of the σ_G^0 ($\Delta\sigma_G^0 \leq 0.5$ dB).
- High irrigation chance (100) corresponds to important increase in the σ_P^0 between t_{i-1} and t_i ($\Delta\sigma_P^0 \geq 1$ dB) accompanied with a decrease of the σ_G^0 (increase of soil moisture at plot scale at t_i with no rainfall event between t_{i-1} and t_i).

Since the SAR backscattering signal could be affected by factors other than the soil moisture, such as the vegetation cover (growth cycle for example) and soil surface roughness, additional filter considering the NDVI values was added to the IEDM. During the sowing or harvesting periods, the surface roughness increases due to soil work which causes an increase in the σ_P^0. For each detected irrigation episode, the difference between the NDVI value at the detected irrigation event and the next NDVI value of the next optical image after 20 to 30 days ($\Delta\text{NDVI} = \text{NDVI}_{ti} - \text{NDVI}_{ti+(20 \text{ to } 30 \text{ days})}$) is obtained. If the $NDVI_{ti}$ value is less than 0.4, (bare soil conditions with small vegetation cover) and $\Delta NDVI \leq 0.1$ then the detected irrigation is most likely to be a soil work effect and is consequently discarded.

The IEDM was applied over the three study sites (Montpellier, Catalonia and Adour Amont). To validate the performance of the IEDM, the latter was applied on 46 irrigated grassland plots (Crau plain) and the detected irrigation events were compared with the irrigation schedule recorded in situ.

1.3.2 Operational mapping of irrigated areas

The proposed irrigation mapping methodology consists of two major steps for mapping irrigated areas. In the first step, the irrigated/non-irrigated training plots are selected based on multi criteria derived from both SAR and optical data. The selection of the training dataset is based on an irrigation possibility weight obtained at each plot using the irrigation event detection model (IEDM) and on the maximum NDVI for the plot.

Using the IEDM, plots with very high number of detected irrigated events are potentially considered as irrigated plots for the training database whereas plots with no detected events are considered as potentially non-irrigated plot. By applying the IEDM on the different S1 time series of the S1 orbits (in VV and VH polarization), an irrigation possibility metric is obtained.

This metric represents an irrigation possibility weight directly related to the number of detected irrigation events on the plot. The irrigation possibility weight is calculated by cumulating the irrigation possibilities (0, 25, 50 and 100) obtained at each SAR image of all S1 temporal series in VV and VH polarization (for each plot). When more irrigation events are detected on the plot, the irrigation possibility weight increases. Consequently, when no irrigation events are detected on the plot, the irrigation possibility weight attains very low value. Thus, irrigated and non-irrigated plots were considered for training data based on very high and very low threshold values of the irrigation possibility weight respectively. From the plots selected based on the irrigation possibility weight, another criterion based on the maximum NDVI value during the crop cycle is applied. It is assumed here that for non-irrigated plots the maximum NDVI should not exceed 0.7 whereas for irrigated plots the maximum NDVI should be greater than 0.8.

After selecting the training dataset that corresponds to the plots deemed as irrigated and non-irrigated with a high confidence degree, the second step consists of implementing S1 data (at plot and grid scales), S2 data (NDVI) and the selected training plots into a random forest classifier to build a classifier for mapping irrigated areas (Irrigation Classifier). Finally, using the in situ dataset, the performance of the classifier was assessed by means of several accuracy metrics. The methodology was performed and validated for four years separately (2017, 2018, 2019 and 2020) over Orléans study site.

1.4 Main results

In the development of the IEDM, the results show that 84.8% of the irrigation events occurring over agricultural plots in Montpellier have been correctly detected using the IEDM. By applying the IEDM over the Catalonian site, the results show that 90.2% of the non-irrigated plots encountered no detected irrigation events whereas 72.4% of the irrigated plots had one and more detected irrigation events. In Adour Amont watershed, the analysis shows that irrigation events could still be detected even in the presence of abundant rainfall events during the summer season where two and more irrigation events have been detected for 90% of the irrigated plots.

The validation of the IEDM over the 46 irrigated grassland plots of the Crau plain show that using only the VV polarization, 82.4% of the in situ registered irrigation events are correctly detected with an F_score value reaching 73.8%. Less accuracy is obtained using the VH polarization where 79.9% of the in situ irrigation events are correctly detected with an F_score

of 72.2%. The combined use of the VV and VH polarization showed that 74.1% of the irrigation events are detected with a higher F_score value of 76.4%.

The proposed operational irrigation classification was validated using real in situ data collected for four years in Orléans. The results show that, using the proposed classification procedure, the overall accuracy for the irrigation classification reaches 84.3%, 93.0%, 81.8% and 72.8% for the years 2020, 2019, 2018 and 2017 respectively. The comparison between the proposed classification approach and the RF classifier built directly from in situ data (fully supervised method) showed that our approach reaches an accuracy nearly similar to that obtained using in situ RF classifiers with a difference in overall accuracy not exceeding 6.2%.

1.5 Conclusion

Currently, the Sentinel-1 satellite is the only operational satellite that provides SAR data free and open access with high revisit time (6 days). However, the 6 days revisit time of the S1 satellite can restrict the detection of irrigation events. The first limitation of the IEDM is the effect of the time lag between the irrigation episode and the satellite passage. In fact, the detection of irrigation can become difficult if the irrigation event takes place more than three days before the SAR acquisition. Over the Crau plain, a sensitivity analysis for irrigation events detection as a function of the time lapse between the acquisition date of the S1 image and the irrigation date was performed. For NDVI ≤ 0.7, the results show that if the S1 image is acquired at the same day of the irrigation event or one day after the irrigation event, 75% of the irrigation events could be detected. Two and three days after the irrigation event, the percentage of the detected and the undetected events becomes approximately the same (50%). Beyond 4 days of the irrigation event, the percentage of the detected irrigation events decreases. However, for NDVI values greater than 0.7 the irrigation could only be detected if it exists within the same day of the S1 acquisition. In the presence of very well developed vegetation cover, higher uncertainty in irrigation detection is observed in the Crau plain, where 80% of the undetected events correspond to an NDVI value greater than 0.8. Moreover, the analysis of false detection of the IEDM showed that small-sized plots encounter more false irrigation detections than large-sized plots because the pixel spacing of S1 data (10 m × 10 m) is not adapted to small size plots with very narrow width or very short length.

Mapping irrigated areas using the proposed operational classification method showed that irrigated plots could be classified with very good accuracy with no need for terrain data.

However, the analysis of the obtained classification accuracies of the proposed method with precipitation data revealed that years with higher rainfall amounts during the crop-growing season (irrigation period) had low overall accuracy (72.8% for 2017) whereas years encountering drier summer had very good accuracy (93.0% for 2019).

2. Article 5: Near Real-Time Irrigation Detection at Plot Scale Using Sentinel-1 Data

Author version of the article published in Remote Sensing Journal: Volume: 11, Issue: 15, August 2019

Hassan Bazzi[1], Nicolas Baghdadi[1], Ibrahim Fayad[1], Mehrez Zribi[2], Hatem Belhouchette[3] and Valerie Demarez[2]

[1] INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

[3] CIHEAM-IAMM, UMR-System, 34090 Montpellier, France

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 15 April 2020; Accepted: 2 May 2020; Published: 4 May 2020

Abstract

In the context of monitoring and assessment of water consumption in the agricultural sector, the objective of this study is to build an operational approach capable of detecting irrigation events at plot scale in a near real-time scenario using Sentinel-1 (S1) data. The proposed approach is a decision tree-based method relying on the change detection in the S1 backscattering coefficients at plot scale. First, the behavior of the S1 backscattering coefficients following irrigation events has been analyzed at plot scale over three study sites located in Montpellier (southeast France), Tarbes (southwest France), and Catalonia (northeast Spain). To eliminate the uncertainty between rainfall and irrigation, the S1 synthetic aperture radar (SAR) signal and the soil moisture estimations at grid scale (10 km × 10 km) have been used. Then, a tree-like approach has been constructed to detect irrigation events at each S1 date considering additional filters to reduce ambiguities due to vegetation development linked to the growth cycle of different crops types as well as the soil surface roughness. To enhance the detection of irrigation events, a filter using the normalized differential vegetation index (NDVI) obtained from Sentinel-2 optical images has been proposed. Over the three study sites, the proposed method was applied on all possible S1 acquisitions in ascending and descending modes. The results show that 84.8% of the irrigation events occurring over agricultural plots in Montpellier

have been correctly detected using the proposed method. Over the Catalonian site, the use of the ascending and descending SAR acquisition modes shows that 90.2% of the non-irrigated plots encountered no detected irrigation events whereas 72.4% of the irrigated plots had one and more detected irrigation events. Results over Catalonia also show that the proposed method allows the discrimination between irrigated and non-irrigated plots with an overall accuracy of 85.9%. In Tarbes, the analysis shows that irrigation events could still be detected even in the presence of abundant rainfall events during the summer season where two and more irrigation events have been detected for 90% of the irrigated plots. The novelty of the proposed method resides in building an effective unsupervised tool for near real-time detection of irrigation events at plot scale independent of the studied geographical context.

Keywords: irrigation; plot scale; near real-time; Sentinel-1

1. Introduction

Efficient management of water resources is required to achieve environmentally sustainable development especially under changing climatic conditions and limited water resources. Fresh water is mainly consumed in the agricultural sector, which is considered the world's largest water user. In fact, with the increase of the global population, irrigating agricultural crops is essential in order to achieve satisfactory agricultural production and income. However, with the decreasing supplies of fresh water due to climate change, better management of irrigation policies is required to deal with the high demand of food and limited water resources.

To support the management of irrigated agricultural policies, a spatially detailed quantification of the irrigation extent and timing is required. This quantification is crucial to monitoring fresh water consumption in the agricultural sector especially for regions suffering scarce water resources. Unfortunately, the extent and distribution of irrigated areas as well as the irrigation timing remain indefinite especially at large scale. Moreover, existing irrigation maps such as the Global Rain-fed, Irrigated and Paddy Croplands (GRIPC) (Salmon et al., 2015) and the Global Map of Irrigated Areas (GMIA) (Siebert et al., 2005) products remain inadequate for irrigation management at plot scale due to their low spatial resolutions (500 m and 5 arc minutes, respectively).

With modern remote sensing, mapping irrigated areas has been the main concern for several studies (Boken et al., 2004; Gumma et al., 2011; Ozdogan and Gutman, 2008). Both optical and radar data have been exploited to perform an irrigated/non-irrigated classification maps. The

use of multi-band optical data is mainly related to the assumption that irrigated/non-irrigated crops could be classified using the temporal series of vegetation indices such as the normalized differential vegetation index (NDVI) (Boken et al., 2004; Gumma et al., 2011; Ozdogan and Gutman, 2008), normalized differential water index (NDWI) (Gumma et al., 2011) or greenness index (GI) (Chen et al., 2018). However, optical data is not only limited to weather conditions but also to the specific studied crop type. For this reason, several studies tend to map irrigated/non-irrigated areas focusing on one specific crop type such as rice (Kamthonkiat et al., 2005) wheat (Bousbih et al., 2018; Fieuzal et al., 2011) or maize (Demarez et al., 2019).

Recent works have shown that a synthetic aperture radar (SAR) signal seems to be more adequate to map irrigated areas over different agricultural crops (Bazzi et al., 2019c; Bousbih et al., 2018; Gao et al., 2018). The use of a SAR signal for mapping irrigated/non-irrigated areas over any vegetation cover is related to the fact that the radar signal is sensitive to soil and vegetation water content (Baghdadi et al., 2011a, 2016a). Since irrigation eventually increases the soil and the vegetation water content, the sensitivity of the radar signal to soil and vegetation water could help detect these irrigation events. Through literature, it has been widely demonstrated that the SAR backscattering coefficient (σ^0) is directly related to the soil and vegetation water content (Aubert et al., 2013; Baghdadi et al., 2011b, 2016a, 2016b; El Hajj et al., 2019; Gao et al., 2018). Mainly for the irrigation task, Hajj et al. (2014) have reported that a three-day-old irrigation point could still be detected using X-band SAR data. In their study, they showed that the X-band radar signal increases by more than 1.4 dB due to irrigation events occurring one day before the acquisition with 90 cm vegetation height. They also showed that for low vegetation cover (vegetation height = 25 cm) the X-band SAR signal increases by 2.6 dB due to irrigation event one day before the SAR acquisition. Similarly, Benabdelouahab et al. (2018) have shown that C-band SAR data could be used to detect irrigation activities over irrigated wheat plots with an interval of three days between the irrigation date and the SAR acquisition date. Since irrigation is a time dynamic activity, an extensive multi-temporal dataset is required to detect consecutive irrigation activities on the studied fields. In addition, high spatial resolution SAR data is required to obtain irrigation information at plot scale. In fact, irrigation information at plot scale is favorable especially in small agricultural areas. Among several SAR satellite constellations, the time series acquired via Sentinel-1 (S1) SAR constellation (S1A and S1B) provides an effective tool for large-scale irrigated area mapping and monitoring due to the unique combination of high revisit time (6 days revisit period) and high spatial resolution (10 m × 10m pixel spacing).

To preform irrigated area mapping, as well as other large-scale area mapping, both deep and machine learning approaches have been extensively exploited since they provide acceptable results and allow large-scale analysis (Pelletier et al., 2019; Zhu et al., 2017). Using the S1 temporal series and the machine-learning approaches, several studies have achieved high-quality classification mapping with a good accuracy (Bazzi et al., 2019d; Feng et al., 2019; Ndikumana et al., 2018; Zhu et al., 2017). However, one of the most important questions about using machine learning approaches relates to the dependency of these models on the terrain calibration data and the studied geographical context (Bazzi et al., 2020a). Such supervised classifications always depend on a training-validation procedure that requires a rich set of labelled samples in order to build the predictive model. Moreover, for better performance, almost all classifications performed via machine learning techniques require complete temporal series data over the growing season, thus making the near real-time mapping more complicated.

Most of the remote-sensing applications for irrigation monitoring mainly focus on the mapping of irrigation extent without taking frequency and timing into account. On the other hand, obtaining information about the period and frequency of irrigation over each agricultural cropland is more significant in the context of irrigation management (Mila et al., 2017; Ozdogan, 2011). In fact, to understand the sustainability of the water resources, irrigation timing and frequency are important especially in arid and semiarid regions. Furthermore, in the context of irrigation water management, the early detection of existing irrigation episodes over each cropland is of great importance for crop modelling in order to estimate the water status and hence better schedule the irrigation episodes over croplands. Better scheduling of irrigation activities can save water that may be used to irrigate more land particularly where water is a limiting factor of agricultural production. In addition, the improvement of the water-use efficiency (WUE) in irrigation requires a real-time control and optimization of the irrigation activities. In irrigated agriculture, improvement of WUE is achieved by optimizing the timing and quantity of irrigation applications (Koech and Langat, 2018). This optimization of the irrigation schedule requires an early detection of irrigation episodes over each irrigated agricultural plot. Moreover, the near real-time detection of irrigation episodes can help monitor and assess the water consumption over agricultural areas. The arrival of the Sentinel satellites (Sentinel-1/2), with high spatial and temporal resolutions, opens the way toward building operational models capable of detecting irrigation events at plot scale. Therefore, the challenge is to build an effective tool capable of detecting irrigation events at plot scale using simple

models that may not require extensive labelled samples (training/validation) and independent of the studied geographical context.

In the context of irrigation water management, the objective of this study is to build a near real-time irrigation detection approach at plot scale using Sentinel-1 time series. First, we analyzed the sensitivity of the radar signal following irrigation events over irrigated plots for a study site located in Montpellier, South-East France. Then, we build a tree-like approach for detecting irrigation events based on the change detection of the SAR signal at plot scale co-jointly with the change detection of the SAR signal obtained at grid scale (10 km × 10 km) which was used to eliminate the ambiguity between rainfall and irrigation. Since the SAR signal obtained at plot scale could be also affected by the vegetation contribution and the surface roughness, several filters considering these effects were introduced in the proposed tree-based approach. The method was tested over irrigated plots in Catalonia, Spain and in Tarbes, South-west France.

2. Materials

2.1 Study Sites

In this study, three different irrigation sites are examined. The first site is located in Montpellier, southeast France (Occitanie region), the second is located in the Catalonia region of northeast Spain and the third in Tarbes of southwest France (Occitanie region) (Figure 1). It is important to mention that both the Montpellier and Catalonia sites are nearly similar in terms of climatic conditions given that both zones are typically Mediterranean. The average annual precipitation in Montpellier is 629 mm where that of Catalonia is 500 mm. However, the summer season in Catalonia is drier than that of Montpellier. On the other hand, the climate in Tarbes is humid to oceanic with an average annual precipitation of 1200 mm. The summer season in Tarbes is more humid with an average precipitation of 300 mm in this season. However, in the three regions, irrigation mainly occurs in the summer season between May and October of each year.

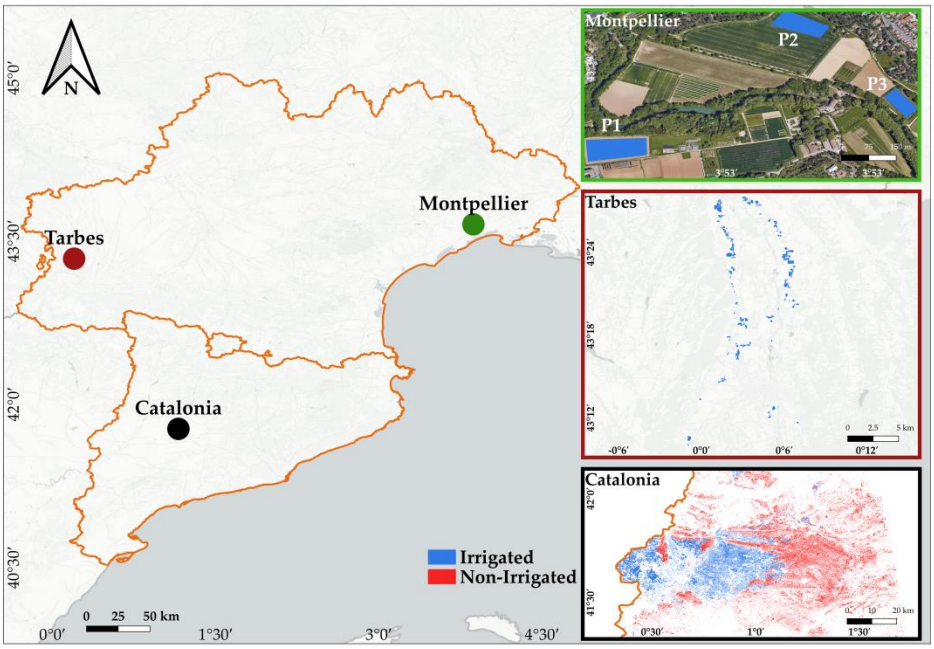


Figure 1: Location of the three studied sites. In green, Montpellier of southeast France; in black, Catalonia region in northeast Spain; and in red, Tarbes of southwest France. Irrigated plots are presented in blue while non-irrigated plots are in red.

2.2 Montpellier Dataset

Irrigation dates over three plots in Montpellier are registered for the year 2017. The three plots includes one maize plot denoted by (P1), one soya plot denoted by (P2) and one sorghum plot denoted by (P3) (Figure 1). Table 1 summarizes the frequency and period of irrigation for each studied plot. In general, irrigation over Montpellier takes place between May and October of each year corresponding to the dry summer season. The sprinkler irrigation technique is generally used. It is good to mention that not only the frequency of irrigation is available but also the exact date of irrigation for these plots. Therefore, these three plots were first used to analyze the effect of irrigation events on the backscattered SAR signal in order to build the suitable method capable of describing irrigation events at the plot.

Table 1: Irrigation information over three plots in Montpellier, south-east France.

Plot	Crop Type	Surface (ha)	Number of Irrigations	Period of Irrigation	Irrigation Type
P1	Maize	1.2	30	01 June–12 Oct. 2017	Sprinkler
P2	Soya	0.8	13	29 May–13 Sept. 2017	Sprinkler
P3	Sorghum	0.44	5	01 June–08 August 2017	Sprinkler

2.3 Catalonia SIGPAC (Geographic Information System for Agricultural Parcels) Dataset

Over the Catalonia region of northeast Spain, the General Direction of Rural Development of the Generalitat of Catalonia provides the Geographic Information System for Agricultural Parcels (SIGPAC) data. The SIGPAC data are based on cadastral plots digitized using aerial images at scale of 1:5000 and 25 cm spatial resolution. The graphical data of the SIGPAC provide the field boundaries. On the other hand, alphanumerical data define each plot by several elements of information including an identification code, surface area, land cover and irrigation indicator. The irrigation indicator shows the presence (100) or absence of irrigation (0). Annual field campaigns are performed each year, in order to update the database mainly for irrigation and land-cover information. In our study, 159 850 plots (123,428 non-irrigated and 36,423 irrigated plots) of different crop types and irrigation coefficients have been used for the year 2018. Considering only agricultural crops (summer and winter crops) in the study, forests, urban, and orchards plots were eliminated. The surface area of the plots varies between 0.1 ha and 65 ha. In general, winter cereals such as wheat, oat, and barely are rarely irrigated with some exceptions. On the other hand, irrigated plots mainly include alfalfa, maize, grassland, beans, rapeseed, and rice. The study area is mainly irrigated using inundation in the old irrigation district and sprinkler or dripper irrigation in the new irrigation district. Fields that have access to water are always irrigated, and fields that do not have access to water are not. The irrigation period occurs mainly in summer, from May to September, and the frequency depends on the irrigation district (old and new). Since irrigation frequency and dates are not available via the SIGPAC data, the irrigation information over the plots was used to analyze the performance of the proposed approach.

2.4 Tarbes Dataset

A field campaign was conducted in Tarbes, southwest France (Figure 1) over irrigated summer crops in 2017 where information about the existence of irrigation was registered for each plot. During this field campaign, 150 irrigated plots including 135 irrigated maize plots and 15 irrigated soya plots were localized. The surface area of the plots varies between 0.15 ha and 28 ha. Irrigation over Tarbes usually takes place between May and October of each year. The most common irrigation technique used in this site is the sprinkler irrigation technique.

Unfortunately, irrigation frequency and dates were not available over these plots. For this reason, these plots were used for analyzing the performance of the proposed model.

2.5 Sentinel-1 Synthetic Aperture Radar (SAR) Time Series

In this study, a total of 348 C-band (5.405 GHz) S1 SAR images acquired by S1A and S1B in both ascending (afternoon at 18:00 UT) and descending modes (morning at 06:00 UT) were used. Over the Montpellier and Tarbes sites, 92 images (46 ascending and 46 descending) images were obtained for each of the two sites for the period between March 2017 and November 2017. This period corresponds to the irrigation period over these two sites. However, for the Catalonia site, 162 images (82 ascending and 82 descending) were used covering a period between September 2017 and December 2018 that correspond to the irrigation information obtained by SIGPAC for 2018. All the images were acquired in the interferometric-wide (IW) swath with VV (Vertical-Vertical) and VH (Vertical-Horizontal) polarizations. In this study, only VV polarization was considered since it is more sensitive to the soil water content than the VH polarization (El Hajj et al., 2017). The 348 images are derived from the Ground Range-Detected (GRD) product with pixel spacing of 10 m x 10 m. The images were downloaded via the European Space Agency (ESA) website (https://scihub.copernicus.eu/dhus/#/home). The S1 toolbox developed by ESA was used to calibrate the 348 S1 images. The calibration (radiometric and geometric calibrations) converts the digital number into backscattering coefficients in linear units (radiometric calibration) and ortho-rectifies the images (geometric calibration) using a 30-m digital elevation model of the Shuttle Radar Topography Mission (SRTM).

Figure 2 shows the repetitiveness of the S1 data in ascending “A” and descending “D” acquisition modes over the three study sites for August 2017 (Montpellier and Tarbes) and August 2018 (Catalonia). For each month, 10 SAR images (5 ascending and 5 descending images) are acquired over each study site. For the Montpellier site (Figure 2a), the descending SAR image (morning) is acquired 36 h before the ascending evening image with an incidence angle of 38.1° and 39.3°, respectively. For Tarbes (Figure 2b), the morning acquisition is also 36 h prior to the evening acquisition with an incidence angle of 36.2° and 40.3°, respectively. Over Catalonia (Figure 2c), the morning acquisition is only 12 h prior to the evening acquisition (both images acquired on the same date) with an incidence angle of 39.5° and 43.2°, respectively.

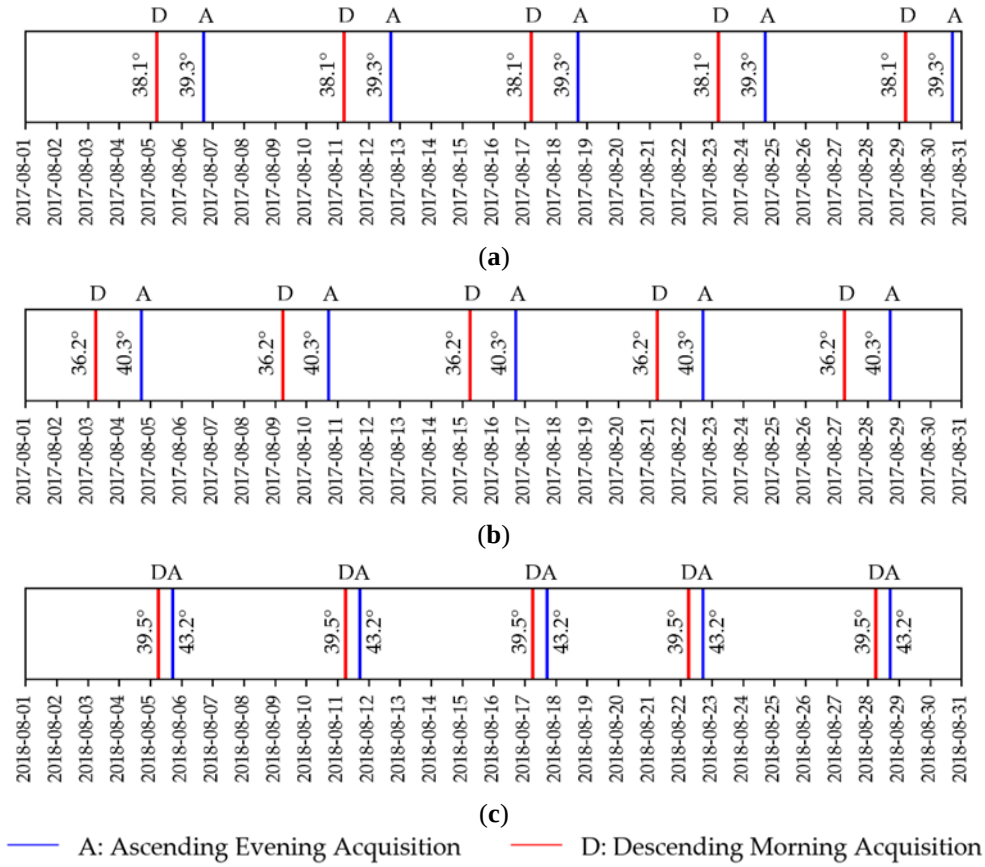


Figure 2: Frequency of the Sentinel-1 images in ascending "A" (afternoon) and descending "D" (morning) modes for one month over (a) Montpellier, (b) Tarbes and (c) Catalonia. The number next to each line represents the incidence angle for each acquisition.

2.6. Sentinel-2 Optical Time Series

We obtained 15, 22 and 17 S2 optical images over Montpellier, Tarbes and Catalonia, respectively, covering the same period of the S1 acquisitions (Section 2.3). The optical cloud-free images were downloaded over each study site with a frequency of approximately one image per month. These optical data were obtained via the Theia website (https://www.theia-land.fr/) which provides S2 images corrected for atmospheric effects (Level-2A). Optical images were used to calculate the NDVI values. The NDVI values were integrated as an additional post-processing filter in the proposed change detection method.

3. Methodology

3.1 Overview

For the three study sites, the average Sentinel-1 SAR backscattering coefficients (σ^0) in VV polarization and the incidence angle values (θ) were calculated at both plot scale (σ_P^0) and grid

scale (σ_G^0) (10 km × 10 km). Since rainfall and irrigation events have the same effect on SAR σ^0 values (σ^0 increases after rainfall or irrigation), this study proposes to remove or minimize the false detection of irrigation events due to rainfall events by using the mean SAR backscattering signal at grid scale (10km × 10km). The SAR signal at grid scale obtained from bare soil areas with low vegetation cover is assumed be a descriptor of rainfall events. For the plot scale, σ_P^0 was obtained by averaging the backscattering coefficient (σ^0) in the linear unit of all the pixels within each plot. For the grid scale, σ_G^0 was obtained by averaging σ^0 values in a linear unit of all agricultural bare soil pixels existing within each grid cell. The average backscattering coefficient, at plot and grid scales, helps reduce the speckle noise in the SAR data. Moreover, at each available S2 image, the average NDVI values were calculated for each plot and grid cell (10 km × 10 km). The proposed methodology consists of two major phases. Phase 1 consists of the tree-based conditions applied over σ°-values at both plot and grid scales in which, at a given SAR date, each plot is evaluated for encountering or not an irrigation event. Phase 2 is a post-processing phase where the irrigation events obtained in Phase 1 at each plot were filtered using additional criteria based on optical data. In order to ensure near real-time detection in phase 1, the absence or presence of an irrigation event at each plot for a given SAR time t_i is evaluated using only the SAR temporal series collected in previous dates (only before t_i) at each plot. However, the optical filter in phase 2 requires subsequent optical data one month after the SAR date. The significance of each proposed filter (criteria) is discussed separately in the coming sections and finally a practical overflow of the whole chain is presented.

3.2 σ° SAR Backscattering at Plot Scale

The proposed method is based on detecting the change in the σ_P^0 backscattering coefficient at plot scale. When the surface soil moisture (SSM) increases between two consecutive SAR acquisitions, σ_P^0-value between these two dates increases. Since irrigation is an artificial application of water, it causes obviously an increase in the SSM values. Thus, the SAR σ_P^0-value could increase between two SAR dates if an irrigation event occurred between these two SAR acquisitions (Benabdelouahab et al., 2018; Hajj et al., 2014). It is good to mention that only the VV polarization was considered in this study since VV is more sensitive to the soil water content than VH which is more sensitive to the vegetation cover (El Hajj et al., 2017).

Not only could the increase in the σ_P^0-values be an indicator of an irrigation event, but also the stability or a slight decrease of σ_P^0-values between two dates could be linked to an irrigation event if the σ_P^0-values attain high values. In fact, with no actions, such as irrigation, rainfall, soil work, or vegetation development, the SAR signal between two near dates (at maximum 6 days) tends to decrease, especially during spring and summer season, due to several water cycle parameters such as the infiltration, evaporation and evapotranspiration that cause a decrease in the SSM values. Thus, stabilization or slight decrease of σ_P^0 between time t_1 and time t_2 could be evidence of an additional water supply if and only if the level of the SAR signal at t_1 is already high due to a previous irrigation or rainfall event occurring at time t_1.

Since the method is based on detecting the change is σ_P^0 values, we propose to calculate, at each SAR acquisition, the difference between the σ_P^0 value at this acquisition and σ_P^0 at the previous SAR acquisition. We calculate therefore ΔVV_P at plot scale as:

$$\Delta VV_P = VV_{P_{ti}} - VV_{P_{ti-1}} \quad (1)$$

where $VV_{P_{ti}}$ is the σ^0 in VV polarization at the present SAR date and $VV_{P_{ti-1}}$ is the σ^0 in VV polarization at the first previous date.

Finally, for the σ_P^0 value, three main thresholds were determined for probable irrigation events. First, if ΔVV_P is less than or equals –0.5 dB (which correspond to a decrease in σ_P^0-value more than 0.5 dB), then we assume that no chance of irrigation exists. If the ΔVV_P is greater than or equal to 1 dB, then a possible irrigation event could have existed. In addition, if the ΔVV_P is between −0.5 dB and 1 dB (stabilization or slight decrease) then an irrigation event could have also existed. However, it is necessary to add other criteria to confirm the possible irrigation event between date t_{i-1} and t_i. These additional criteria will be discussed in the coming sections.

3.3 σ° SAR Backscattering at Grid Scale

The increase in the SSM values could not only be attributed to an irrigation event but also to rainfall events that are the main contributor in the SSM variation. Then, a good separation between rainfall events and irrigation events should be also performed. Actually, rainfall events and irrigation events are both considered as water supplements and thus may have the same effect on the value of ΔVV_P. For example, the threshold value of $\Delta VV_P \geq 1dB$ (presented in Section 3.2) could be a result of a rainfall event and not only an irrigation event. Therefore, the

ambiguity between rainfall and irrigation is the principal factor to be resolved for a good detection of irrigation events. Thus, better detection of irrigation events requires information about rainfall.

In this study, information about rainfall has been determined through the σ° value obtained at grid scale (10 km × 10 km). We assume that if the bare soil plots with low vegetation cover within the spatial extent of 10 km x 10 km show an increase of the SAR backscattering signal (increase in SSM values) between two consecutive radar acquisitions, a rainfall event probably occurred. This correlation between rainfall and σ_G^0 has been presented by Bazzi et al. (2019a) where they compared SSM estimations at bare soil plots in 10 km grid scale to rainfall events at the same scale and concluded that strong consistency exists between rainfall events and SSM values at 10 km grid scale. Moreover, Bazzi et al. (2019c) used the SAR signal at grid scale conjointly with the SAR signal at plot scale to map irrigated areas at plot scale. They reported that the use of σ_G^0 has remarkably improved the classification accuracy of irrigated/non-irrigated plots by 15%.

The SAR backscattering coefficients at grid scale are obtained by averaging $\sigma°$ values of all bare soil pixels within each grid cell (10 km × 10 km). A mask for bare agricultural soil pixels has been determined using first a land-cover map to delineate only agricultural areas (excluding urban, forests…) and then a threshold applied over the NDVI values obtained from S2 images. For this reason, the land-cover map proposed by (Inglada et al., 2017)is used for the two French sites while the agricultural plots of the SIGPAC data have been used for the Catalonia site. A maximum NDVI value of 0.4 is fixed to extract bare soil pixels with low vegetation cover. Thus, at each SAR date, and for each grid cell a σ_G^0 value is obtained describing the σ° backscattered from bare soil pixels of agricultural areas only within each grid cell (10 km × 10 km).

At a given date, the change in SAR σ° at each 10 km cell could be obtained through the difference of σ° value at the given date t_i and the σ° at the first previous date t_{i-1}

$$\Delta VV_G = VV_{G\,ti} - VV_{G\,ti-1} \quad (2)$$

where $VV_{G\,ti}$ is the σ_G^0 in VV polarization at the present SAR date and $VV_{G\,ti-1}$ is the σ_G^0 in VV polarization at the first previous date.

Using a threshold value of ΔVV_G, rainfall events could be determined and irrigation events detected in Section 3.2 could be filtered. First, the grid based filters are only applied in the case

where ΔVV_P has a value more than or equal to −0.5 (Section 3.2) (probable irrigation event). We consider that if the ΔVV_G is greater than 1 dB then a rainfall event occurred and, therefore, there is no chance of irrigation detection regardless of the value of ΔVV_P. This filter helps reduce an important part of the ambiguity between rainfall and irrigation. Next, we consider that if the value of ΔVV_G is between 0.5 and 1 dB (low rainfall possibility), then a probability of irrigation can exist based on the value of ΔVV_P. In this case, if ΔVV_P is less than 0.5 dB then irrigation did not take place (σ_G^0 at 10 km bare soil plots have slightly increased more than that at the plot scale). On the other hand, if the value of ΔVV_P is greater than 0.5 dB, then we calculate the difference (Δ) between ΔVV_P and ΔVV_G ($\Delta = \Delta VV_P - \Delta VV_G$). If the value of Δ is greater than 1 dB, then irrigation has more chance than rainfall and the point is assumed corresponding to an irrigation event. Inversely, if Δ is less than 1 dB then no irrigation occurs. The indicator Δ was also used to help confirm whether a point is an irrigation event or not in the case of slight change in ΔVV_P ($-0.5\ dB \leq \Delta VV_P < 0.5\ dB$). When the value of ΔVV_P is between 0 and 0.5, the value of Δ must be greater than or equal to 1.5 dB in order to consider the point as an irrigation event. Similarly, Δ should be greater than or equal to 2 dB if the value of ΔVV_P is between −0.5 and 0.

Finally, we consider that if the value of ΔVV_G is less than 0.5 dB, then irrigation could be detected in case the value of ΔVV_P respects certain criteria which will be detailed in Section 3.7.

3.4 Reducing Vegetation Contribution

Over vegetated areas, σ_P^0 is not only affected by the soil water content but also by the characteristics of the vegetation cover. Indeed, for certain agricultural crops, the SAR backscattering signal is attenuated by the existing vegetation cover (Attema and Ulaby, 1978; El Hajj et al., 2017). For example, for crop types such as soya, sorghum and sunflower the SAR backscattering signal between two dates could increase due to the development of the vegetation cover. The direct effect of vegetation cover on the SAR signal should be also considered to ensure accurate detection of irrigation events.

Recently, Nasrallah et al. (2019) used a smoothed Gaussian filter on SAR temporal series in order to describe the phenology stages of the vegetation growth over wheat crops. They demonstrated that the smoothed Gaussian could be used to describe the vegetation contribution in the SAR temporal series over wheat plots. Therefore, to minimize the effect of vegetation

growth on the radar signal, we propose to smooth the temporal series of the SAR backscattering signal. This smoothing gives the general behavior of the vegetation contribution in the SAR signal. However, in order to ensure the near real-time detection of irrigation, we propose to smooth the signal only using SAR dates existing before the examined SAR date. This means that at each SAR date, a smoothing value σ^0_{smooth} is obtained by applying a Gaussian smooth on all points before this date (from t_0 to t_i). In this study, the multidimensional Gaussian smooth has been used and the standard deviation for the used Gaussian kernel was set equal to 4. Finally, we obtain the index S considered as a vegetation descriptor where:

$$S = \sigma^0_{P_{ti}} - \sigma^0_{smooth} \quad (3)$$

Therefore, if $\sigma^0_{P_{ti}}$ is less than σ^0_{smooth} (i.e., $S < 0$), then the point should not be further considered for irrigation detection. This smoothing allows us to determine a vegetation indicator capable of reducing the vegetation effects in the SAR signal (at each date) without losing possible irrigation events.

Another vegetation contribution filter has been suggested particularly for winter cereals usually grown in the period between September and July of each year (wheat, barley and oats). In their study over wheat crops, Nasrallah et al. (2019) showed that the C-band SAR backscattering signal in VV polarization decreases gradually between the germination phase occurring by the beginning of January and the heading phase occurring between mid-March and mid-April. In the heading phase, the C-band SAR backscattering signal attains extremely low values due to extreme vegetation attenuation (less than –15 dB for incidence angle between 32° and 34°). The σ^0_P gradually increases then, between mid-April and the end of May, when cereals move from the heading phase to the soft dough phase. This increase in the C-band SAR signal is mainly due to the change in the phenology phase of the cereals and is not linked to irrigation episodes. To reduce this ambiguity, we propose to eliminate detected irrigation points between mid-April and the end of May if and only if σ^0_P attains extremely low value (less than −15 dB) between mid-March and mid-April. In this way, we can ensure that these detected points are most probably a phenology change of cereals where σ^0_P increases from extremely low values between March and April ($< -15\ dB$) to higher values in May. Practically, for any plot, if a point is detected between mid-April and the end of May, we find the minimum of all σ^0_P values (denoted by MIN) acquired between mid-March and mid-April. If the MIN value is less than –15 dB (most probably heading phase of cereals), then we eliminate the detected point.

3.5 Surface Soil Moisture Filter

As discussed in Section 3.2, irrigation activities obviously cause an increase in the SSM values at plot scale. Thus, an important factor that can describe the presence or absence of irrigation events is the SSM value. We consider here that low soil moisture values are not correlated with an occurring irrigation event since irrigation must increase the SSM values especially over plots with small vegetation cover. However, the integration of any proposed SSM filter requires proper estimation of SSM values. Recently, an algorithm using the neural network (NN) technique was developed by El Hajj et al. (2017) to estimate SSM values at plot scale over agricultural areas with vegetation cover. In their study, SSM values were estimated with an accuracy of 5 vol.% (volumetric water content in percent = 0.05 cm^3cm^{-3}) over agricultural areas which is considered a pleasant accuracy for any hydrological application at plot scale. In addition, the NN developed by El Hajj et al. (2017) has showed the most accurate SSM estimations when evaluated against several SSM products such as the Soil Moisture and Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP) (El Hajj et al., 2018a) and Copernicus Surface Soil Moisture (C-SSM) (Bazzi et al., 2019a)

In this study, the NN developed by El Hajj et al. (2017) has been used to estimate SSM values at plot scale. The NN requires as input data: SAR signal in VV polarization (σ^0), the SAR incidence angle (θ), and an NDVI value. As a result, SSM values were estimated at each plot and for each SAR date denoted by (SSM_P). However, in another study, El Hajj et al. (2018b) recommend that the SSM estimation could be limited in the presence of very dense vegetation cover due to the high vegetation attenuation on SAR signal. Thus, SSM_P has only been considered for NDVI values less than 0.5. Finally, an additional filter is applied in which irrigation points have been restricted to SSM estimation ≥15 vol.% when the NDVI value is less than 0.5. For NDVI values greater than 0.5, the SSM filter was not applied.

SSM values at plot scale were also used to help confirm whether a point is an irrigation point or not in the case of slight change in ΔVV_P ($-0.5\ dB \leq \Delta VV_P \leq 1\ dB$). As presented in Section 3.2, the stability or slight decrease of the SAR signal at plot scale between t_i and t_{i-1} could be interpreted as an irrigation event if and only if the σ^0 at time t_{i-1} already attains high values (due to irrigation or rainfall). To ensure this situation, we say that at time t_{i-1}, SSM estimation should be greater than or equals to 20 vol.% in order to guarantee that humid soil conditions at time t_{i-1} have continued to time t_i.

Over the grid scale, the SSM estimation for the σ_G^0 obtained for bare soil plots with low vegetation cover (NDVI <0.4) at grid scale (Section 3.3) was also performed at each grid cell and for each SAR date (SSM_G). This estimation presents the surface water content over bare soil plots on the basin scale. In fact, we assume that high soil moisture values at grid scale (10 km × 10 km) are more likely to be linked to possible rainfall events rather than irrigation events (humid soil conditions at grid scale). For this reason, we propose to eliminate all the points where the SSM estimation at grid scale is greater than 20 vol.%. For the grid scale, the effect of vegetation attenuation on SSM estimations (high NDVI values) does not exist since the σ_G^0-values were only calculated for bare soil plots with low vegetation cover.

3.6 Optical Normalized Differential Vegetation Index (NDVI) Filter

During the sowing or harvesting periods, the cropland plots usually encounter an increase in the surface roughness due to soil work. In fact, the backscattered radar signal strongly depends on the geometric characteristics such as the surface roughness that is usually expressed by the height root mean square (*Hrms*). The *Hrms* is the standard deviation of surface height (root mean square) which specifies the vertical scale of surface roughness. Several studies have discussed the sensitivity of the radar backscattering signal to the surface roughness (Aubert et al., 2011; Baghdadi et al., 2008b, 2018a). Baghdadi et al. (2008a) reported that a difference of 4 dB could be observed between backscattering signal from smooth surface (*Hrms* = 0.5 cm) and rough surfaces (*Hrms* = 3 cm). Therefore, between two near-date SAR acquisitions, an increase in the surface roughness could cause an increase in the backscattering coefficient. This increase is related to the change of the geometrical characteristics of the soil and not to the change of the water content (irrigation or rainfall). To overcome this limitation, an additional optical filter is suggested in order to better distinguish irrigation peaks from soil works such as sowing or harvesting.

Generally, irrigation activities must occur during a crop-growing cycle. This means that irrigation must be followed by a development of the vegetation cycle. When croplands receive water, high soil moisture causes better photosynthesis resulting in an increase in the leaf area index (LAI) values and therefore an increase in the NDVI values (Chen et al., 2018). Therefore, we consider that an irrigation event should be followed by the development of NDVI values. For this reason, a post processing filter is proposed to eliminate some false detected irrigation events due to soil work. At each plot and for each detected irrigation event (at a given SAR date), we first obtain the difference between the NDVI value at the detected irrigation event and

the next NDVI value of the next optical image (after 20 to 30 days according to cloud limitation):

$$\Delta NDVI = NDVI_{ti} - NDVI_{ti+(20\ to\ 30\ days)} \quad (4)$$

where $NDVI_{ti}$ is the NDVI value at the current SAR date and $NDVI_{ti+(20\ to\ 30\ days)}$ is the NDVI value one month later.

The $\Delta NDVI$ value is considered as a vegetation indicator which helps in detecting whether an increase of the NDVI values is observed or not (growing cycle) after the probable detected irrigation event. The filter suggests that if the $NDVI_{ti}$ value is less than 0.4 (bare soil conditions with small vegetation cover) and $\Delta NDVI \leq 0.1$, then the point is a falsely detected irrigation point and is eliminated. For $NDVI_{ti}$ greater than 0.4, the filter is discarded because in the presence of vegetation the existence of irrigation event is more probable. This filter ensures that if the bare soil condition is permanent and a vegetation growth cycle does not exist (or was in a decreasing stage), then the detected event is most likely to be a soil work point and not an irrigation point. However, this filter is a post-processing filter that can be applied after obtaining another NDVI image after 20 to 30 days.

3.7 Global Overflow for Irrigation Event Detection

Figure 3 presents a detailed overflow of the proposed tree-based change detection methodology. For a given plot and at a given SAR image acquired at time t_i, seven main indicators could be extracted for the plot:

- ΔVV_P: Change in SAR signal at plot scale
- ΔVV_G: Change in SAR signal at grid scale
- S: Smoothed vegetation descriptor S
- SSM_P: SSM value at plot scale
- SSM_G: SSM value at plot at grid scale containing this plot
- $NDVI_{ti}$: NDVI value at time t_i
- $\Delta NDVI$: Vegetation growth indicator

The chain starts with the ΔVV_P value where a value less than −0.5 dB is considered as the non-irrigation point. If the value of ΔVV_P is greater than −0.5 dB, the smoothed vegetation descriptor S is then checked for being positive and the point is considered as non-irrigation if $S < 0$. If S is positive, the SSM_P is then checked for the threshold value of 15 vol.% and the

point is considered as non-irrigation if the $SSM_P < 15\ vol.\%$ with $NDVI_{ti} \leq 0.5$ (both conditions should occur simultaneously). When the studied point arrives to pass by all the previous filters, the chance of having a water supplement (irrigation or rainfall) increases, and thus the change of SAR signal at grid scale is required at this stage in order to eliminate the irrigation-rainfall ambiguity. Filters applied at grid scale could be divided into four main cases:

- **Case i**: If $\Delta VV_G \geq 1\ dB$ then a rainfall event have occurred and the point is not an irrigation point.
- **Case ii**: If $SSM_G > 20\ vol.\%$ then a rainfall event probably occurred before and there is low chance to have an irrigation event (humid soil conditions at basin scale).
- **Case iii**: If $0.5 \leq \Delta VV_G \leq 1\ dB$ we check the value of ΔVV_P for two cases:
  - **Case iii.1**: If $\Delta VV_P \leq 0.5$ then no irrigation took place.
  - **Case iii.2**: If $\Delta VV_P > 0.5$ and $\Delta\ (\Delta VV_P - \Delta VV_G) \geq 1$ then it is considered as irrigation point with high certainty.
- **Case iv**: If $\Delta VV_G \leq 0.5\ dB$ then we check the ΔVV_P for four different cases:
  - **Case iv.1**: $\Delta VV_P \geq 1\ dB$ then the point is an irrigation point with high certainty.
  - **Case iv.2**: $0.5 \leq \Delta VV_P < 1\ dB$ then the point is an irrigation point with medium certainty if and only if $SSM_P \geq 20\ vol.\%$ OR $\Delta \geq 1.5\ dB$.
  - **Case iv.3**: $0 \leq \Delta VV_P < 0.5\ dB$ then the point is an irrigation point with low certainty if and only if $SSM_P \geq 20\ vol.\%$ OR $\Delta \geq 2\ dB$.
  - **Case iv.4**: $-0.5 \leq \Delta VV_P < 0\ dB$ then the point is an irrigation point with low certainty if and only if $SSM_P \geq 20\ vol.\%$ AND the previous point at t_{i-1} is a high certainty irrigation point or a rainfall point ($\Delta VV_G \geq 1\ dB$).

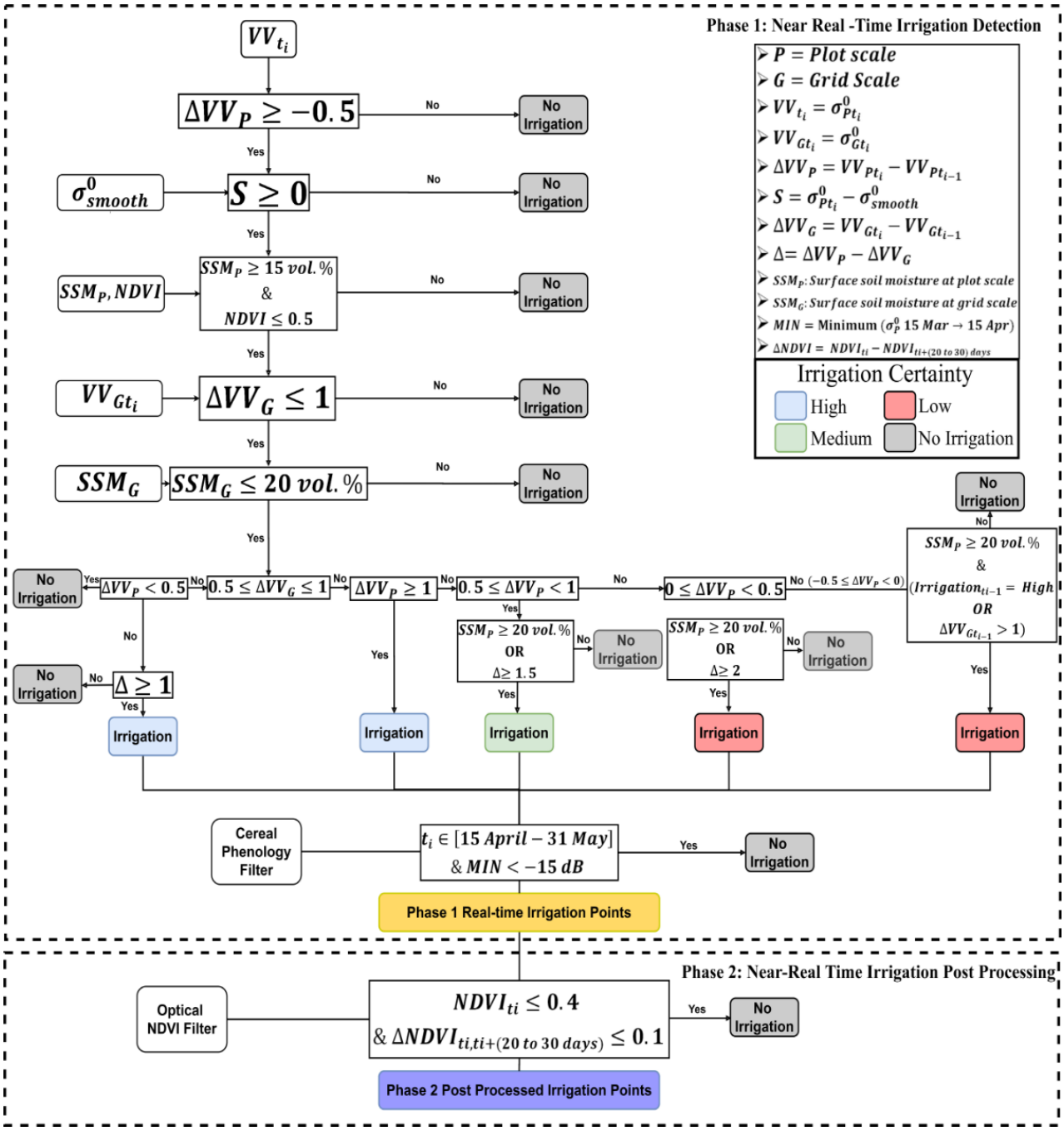


Figure 3: Workflow overview using the tree-based classification for irrigation detection at plot scale

A certainty indicator of irrigation (high, medium and low) is also associated for each detected irrigation point. The high certainty irrigation points are associated for the points considered as irrigation points from cases iii.2 and iv.1 (significant increase in ΔVV_P). For the points of case iv.2 a medium certainty is associated, while low certainty is considered for irrigation points detected from cases iv.3 and iv.4.

The implementation of the proposed method in a near real-time scenario depends principally on the delivery time of S1 images. S1 images are usually delivered by ESA in a “fast 24 h”

delivery mode. This mode insures that the S1 image is available for download 24 h after the satellite acquisition. Considering that the pre-processing of S1 images and applying the proposed method could be automatically performed with minimum human involvement, the irrigation event could then be detected about one hour after receiving the S1 images.

Finally, the post NDVI filter (Section 3.6) is applied for each detected irrigation point in order to eliminate the ambiguity with soil work. When another NDVI image is acquired at $t_{i+(20\ to 30\ days)}$, the $\Delta NDVI$ is calculated. Then, for each detected irrgation event, if $NDVI_{ti} < 0.4$ and $\Delta NDVI \leq 0.1$, then this irrigation event is eliminated. This NDVI filter is a long-term real-time scenario because it requires obtaining a new NDVI value one month later. Therefore, this optical filter remains as a post processing of the obtained irrigation points.

It is important to mention that the methodology is applied separately for the ascending (evening overpass) and descending (morning overpass) SAR acquisitions. In fact, morning and evening acquisitions could not be joined in one temporal series due to the presence of the diurnal variations between the two acquisitions. The diurnal variation is a result of the difference in the vegetation water content (VWC) between the morning and the evening. This difference in VWC causes high difference in the radar backscattering signal over vegetated plots between the morning and the evening acquisitions. Several studies have reported that σ^0 in the morning overpass registers higher values than σ^0 in the evening overpass (Brisco et al., 1990; van Emmerik et al., 2015; Friesen et al., 2012). Therefore, it was suggested to investigate separately each SAR temporal series acquired in the morning and evening.

4. Results

4.1 Grid Scale σ° Temporal Profile

Since irrigation and rainfall events are both considered a water supplement and have the same effect on the SAR backscattering coefficients, it was proposed to minimize the irrigation-rainfall ambiguity using the σ° values obtained at basin scale (10 km × 10km). Figure 4 shows an example of the temporal behavior of σ^0_G SAR values in VV polarization obtained at a 10 km grid cell (red curve) for Montpellier (Figure 4a), Tarbes (Figure 4b) and Catalonia (Figure 4c). Daily precipitation records obtained from the Global Precipitation Mission (GPM) data are added to the figures (blue curve) to help understand the consistency between the grid scale σ^0_G values and the rainfall events. The green line shows the SSM estimation at grid scale

(for bare soil plots with low vegetation cover) at each SAR date while the dotted black line shows the threshold value fixed for SSM_G at 20 vol.%. Following a rainfall event, the σ^0_G value increases (more than 1 dB) due to important precipitation before the SAR acquisition (black dashed circle) accompanied with high SSM estimations (more than 20 vol.%) indicating humid soil conditions at 10 km scale. On the other hand, the absence of precipitation causes a decrease or stability of σ° value at grid scale with low soil moisture values (less than 15 vol.%) indicating dry soil conditions (yellow dashed circle). This consistency between rainfall, σ^0_G and SSM estimations at grid scale ensures that both σ^0_G and SSM estimations at 10 km scale are a good representative for rainfall events and, therefore, could be used to eliminate the uncertainty between rainfall and irrigation.

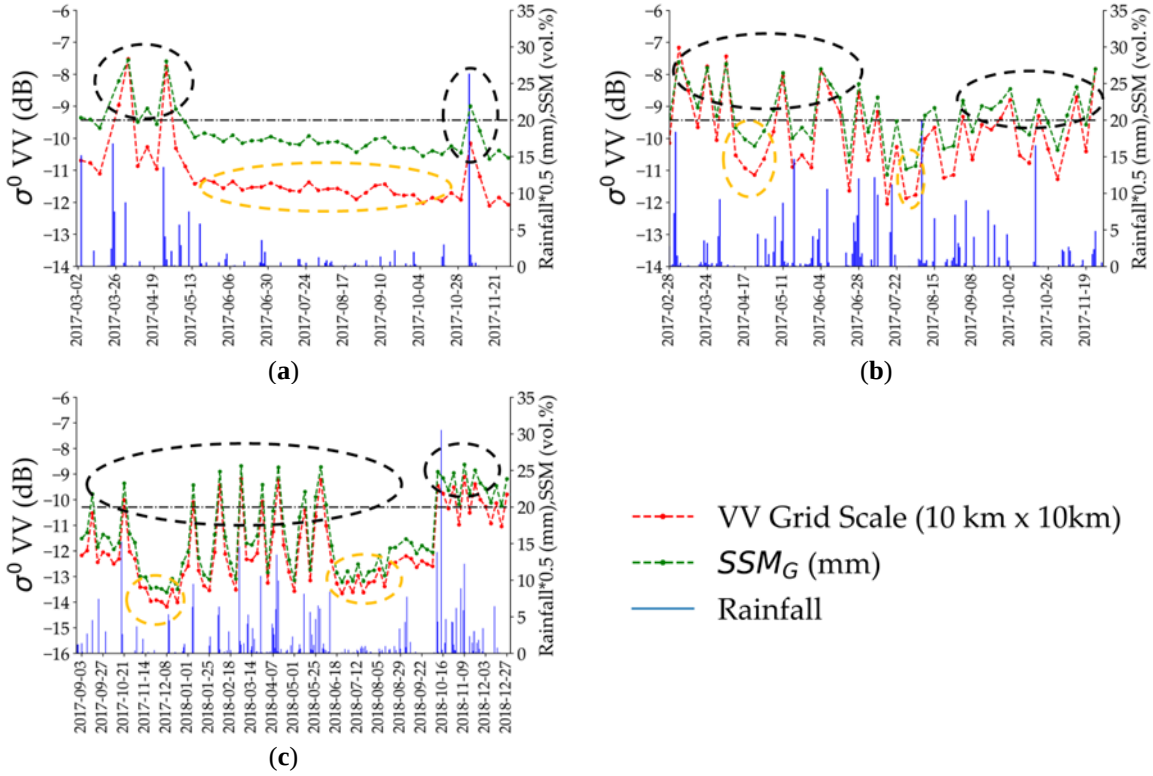


Figure 4: Temporal evolution of Synthetic Aperture Radar (SAR) backscattering coefficient σ° in VV polarization at 10 km grid scale (red curve) with surface soil moisture (SSM) estimation at 10 km grid scale (green curve) and daily precipitation data from the Global Precipitation Mission (GPM) mission (blue curve) for (a) Montpellier, (b) Tarbes and (c) Catalonia. Black dashed circle corresponds to existing rainfall events while yellow circle shows the absence of rainfall.

4.2 Results over Montpellier

Figure 5 presents the results of the application of the proposed methodology (Phase 1 and 2) over the three plots P1, P2 and P3 located in Montpellier, France using the ascending (morning) and descending (evening) SAR images separately. Over Montpellier site, the morning acquisition is 36 h prior to the evening acquisition.

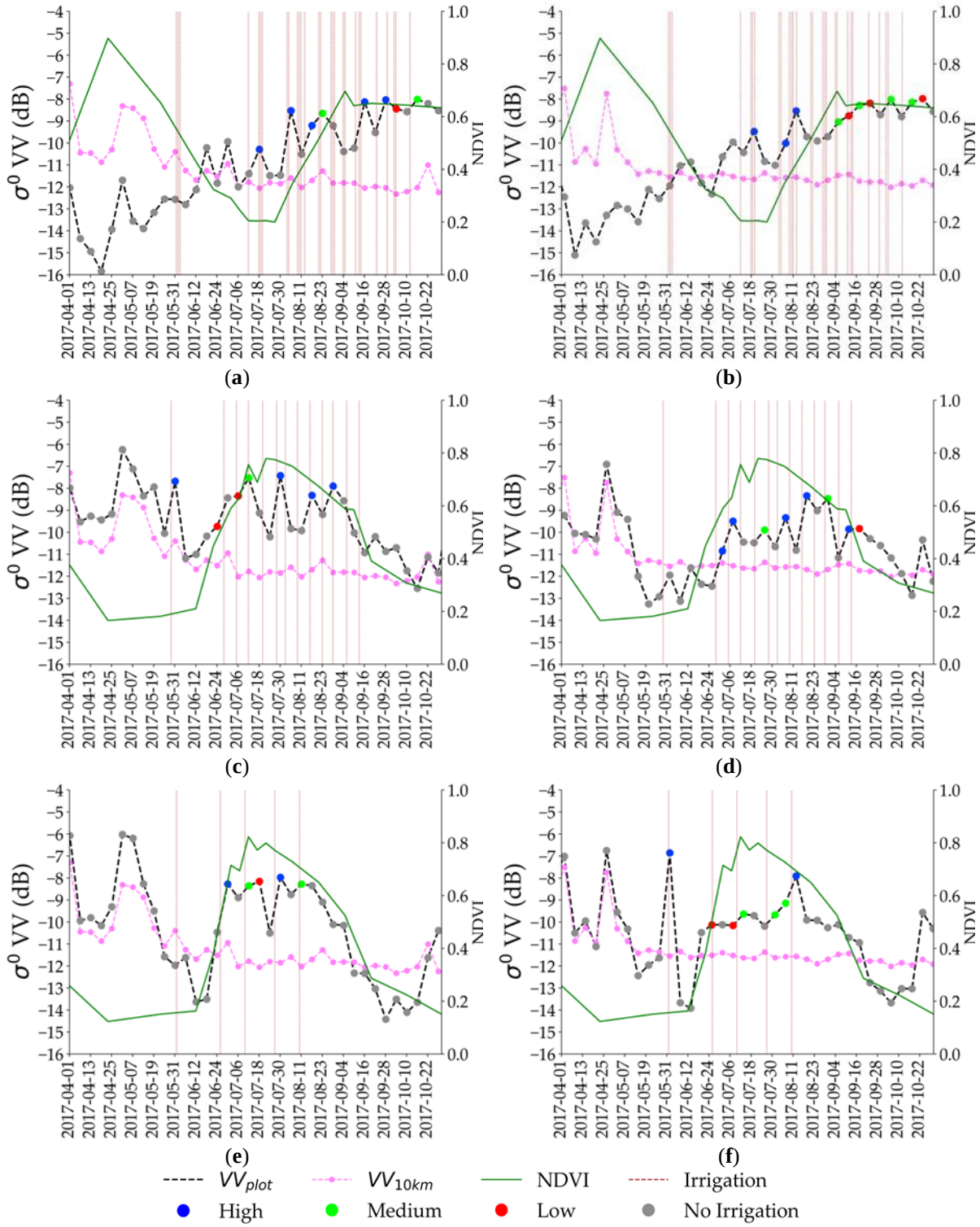


Figure 5: Irrigation detection over P1 (a, b), P2 (c, d), and P3 (e, f) using the morning SAR acquisition (a, c, e) and the evening SAR acquisition (b, d, f), in Montpellier, France. SAR VV signal at plot scale in dashed black line and VV at grid scale in dashed pink line. Points with blue, green and red are irrigation points detected on SAR signal at high, medium and low certainty respectively. The green line represents the normalized differential vegetation index (NDVI). High, medium and low correspond respectively to detected irrigation events with high, medium and low certainty.

Figures 5a and 5b show the morning and evening SAR acquisitions for P1 plot (maize), respectively. The first detected irrigation point was on 17/07 (morning acquisition) and 18/07

(evening acquisition) due to two irrigation episodes that took place on 16/07 and 18/07. Between 19/07and 31/07 no irrigation occurred on the plot and no peaks where detected for two consecutive SAR acquisitions. Later, on both 05/08 (morning) and 06/08 (evening), an irrigation peak was detected due to two irrigation episode occurring on 03/08 and 04/08. In addition, two irrigation episodes that occurred on 04/09 and 05/09 appeared as an irrigation peak on the evening image of 06/09. However, the morning image acquired on 04/09/2017 did not show any detected irrigation since the episodes occurred after the SAR acquisition (SAR acquired at 06h00 while irrigation generally takes place after 09h00). Moreover, an irrigation episode occurring on 11/09 was detected on the evening image acquired on 12/09/2017. Two irrigation episodes occurring on 13/09 and 14/09 appeared on both the morning and evening acquisitions on 14/09 and 15/09, respectively. By combining the irrigation events detected from the morning and evening acquisition, 12 out of 15 possible irrigation episodes over P1 were detected including 5 points with high certainty, 5 points considered as medium certainty and 2 points with low certainty. However, it is important to mention that several irrigation episodes occurring between two consecutive SAR dates are considered as only one irrigation episode. For plot P1 only one point with low certainty was falsely detected as irrigation event.

For the soya plot (P2), the results of the combined use of the morning (Figure 5c) and evening (Figure 5d) acquisitions show that 11 irrigation episodes were detected out of 13 possible irrigation episodes. Among the 11 detected irrigation points, eight points were with high certainty, two medium and one low certainty point. Moreover, among the 92 SAR acquisitions (morning and evening), only one false irrigation detection was obtained with low certainty. The first irrigation episode on 29/05 was detected in the morning SAR image of 31/05 (Figure 5c). For the morning acquisition, the radar signal then decreased between 31/05 and 06/06 due to the decrease in soil moisture but started to increase between 06/06 and 24/06 without any rainfall or irrigation events. This increase in the σ_P^0 values for three consecutive acquisitions was strongly correlated with the development of vegetation cover (NDVI values increases sharply during this period). However, among these three SAR points, two points were correctly not detected as irrigation points since they were filtered by the smoothed Gaussian filter (S) that helped eliminate the effect of the vegetation development on σ_P^0. Figure 6 explains the effect of the smoothed-Gaussian filter over P2 plot of Montpellier where the σ_{smooth}^0 (red curve) is added to the results obtained by Figure 5d to demonstrate the importance of the smoothed-Gaussian filter. In Figure 6, the green dashed circle shows the three consecutive increase of σ_P^0 between 06/06/2017 and 24/06/2017. However, for two out of these three points,

the σ^0_{smooth} curve is above σ^0_P curve thus the calculated S value is negative and these points were not detected as irrigation points.

The sorghum plot (P3) encountered five irrigation episodes between 31/05/2017 and 10/08/2017 (Figure 5e and Figure 5f). Using the proposed change detection method, five out of the five episodes were correctly detected as irrigation events. The first irrigation episode was detected on the evening SAR image (Figure 5f) of 01/06 due to an irrigation event occurring on the same date. Then, the 4 irrigation episodes that occurred on 26/06, 07/10, 27/10 and 10/08 were detected with both acquisition modes (ascending and descending). However, on this plot two false irrigation points were detected with medium and low power.

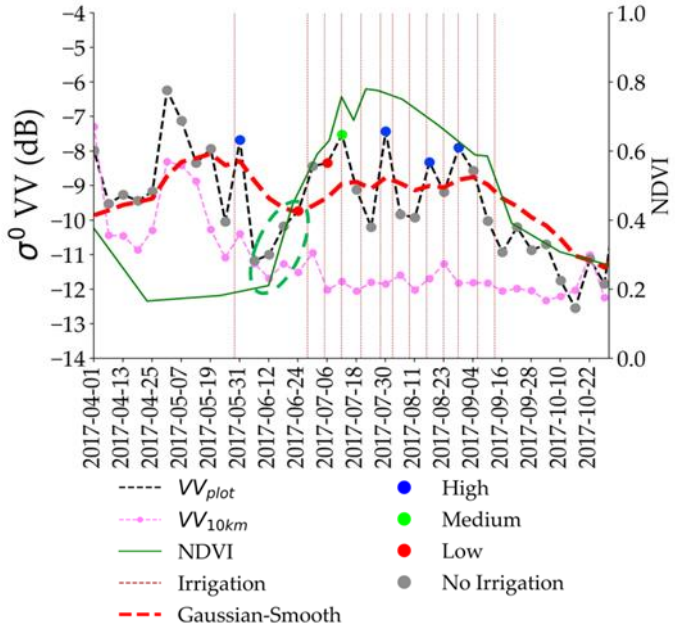


Figure 6: Effect of smoothed-Gaussian filter on the detection of irrigation points for plot P2, Montpellier. The red dashed curve represents the σ^0_{smooth} calculated as each SAR date.

Table 2 summarizes the results obtained on the three plots of Montpellier site. The total number of irrigation events over the three plots was 48 events. However, at each plot, several irrigation events occurring between two consecutive SAR dates are considered as only one irrigation event. We call these events the possibly detectable irrigation events. For example, three irrigation events occurring on 10/08, 11/08 and 13/08 on plot (P1) are considered as one irrigation event since the three events were followed by only one available SAR acquisition on 13/08/2017. Therefore, among the 48 irrigation events occurring on the three plots, 33 events could be possibly distinguished by the acquired SAR temporal series. This means that, 69% of the total number of irrigation events could be detected by the available SAR temporal series. Out of 33 possibly detectable irrigation events over the three plots, 28 irrigation episodes have

been correctly detected. This means that among the possible detectable irrigation events 84.4% of these events are correctly detected. However, a total error of five falsely detected irrigation events has been registered over the three plots. The falsely detected irrigation points corresponds to the saturation of the radar signal caused by very well developed vegetation cover in plots P1 (maize) and P3 (Sorghum).

Table 2: Irrigation events detected over three plots in Montpellier, South-east France

Plot	Number of Irrigation Events	Possibly Detectable Irrigation Events	Detected Irrigation Events	False Detection
P1	30	15	12	2
P2	13	13	11	1
P3	5	5	5	2
Total	48	33	28	5

4.3. Results over Catalonia

Over the Catalonia region, the huge database available allows us to investigate in depth the performance of the proposed method. The method was applied not only over irrigated plots but also over non-irrigated plots in order to analyze its capability to distinguish between irrigated and non-irrigated fields. First, it is important to remember that Catalonia region is a semi-arid region with a dry summer where most of the irrigation activities occur during the summer season (between May and September). Figure 7 presents the temporal profile of σ_P^0 and σ_G^0 over the period between September 2017 and December 2018 (evening acquisitions) along with the irrigation points detected during this period for an irrigated maize (Figure 7a) plot, irrigated alfalfa plot (Figure 7b), and non-irrigated wheat plot (Figure 7c). The daily GPM precipitation data are also presented in the three figures (blue curve). During the period between September 2017 and December 2018, 11 irrigation points were detected on the maize plot whereas 12 irrigation points were detected on the alfalfa plot. In both irrigated plots (Figure 7a and 7b), the irrigation points correspond to the dry summer period (between 25 April and 30 September). For example, 6 irrigation points were detected on the maize plot (Figure 7a) between 12/06/2018 and 15/08/2018. This is correlated with the existence of the maize vegetation cycle in the summer season (Increasing NDVI values). The alfalfa plot (Figure 7b) shows also frequent irrigation points detected between 12/06/2018 and 30/08/2018. For both maize and alfalfa plots, the σ_G^0 between 01/06/2018 and 30/09/2018 shows stable low values indicating dry conditions and the absence of rainfall events. On the other hand, the frequent change of the σ_P^0 in both plots indicates that possible irrigation events have occurred. This increase of the σ_P^0 was detected as irrigation events based on our proposed method.

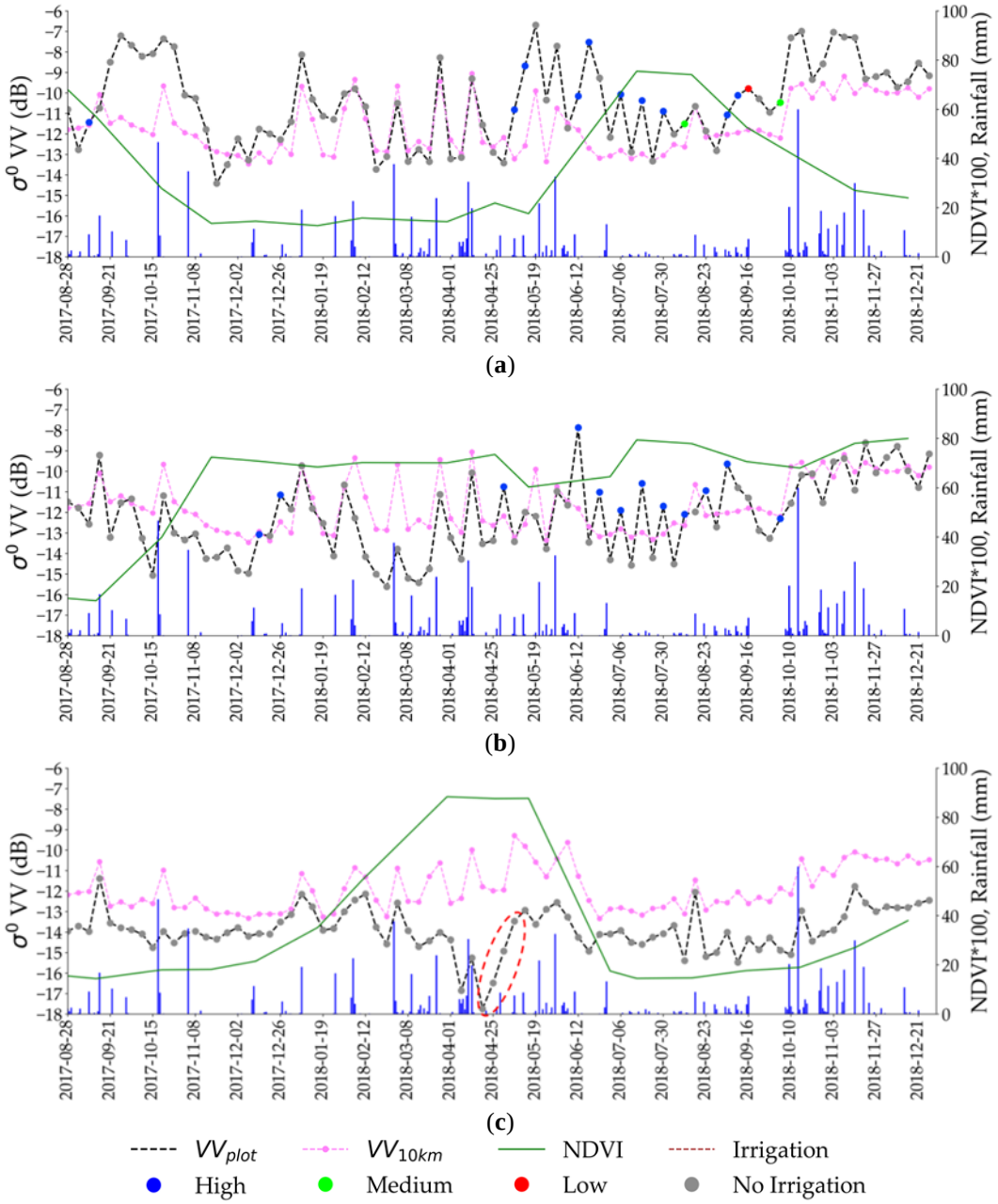


Figure 7: Irrigation detection over agricultural plots in Catalonia, Spain for (a) irrigated maize plot, (b) irrigated alfalfa plot, and (c) non-irrigated wheat plot. The red dashed ellipse represents the period of the transition between the heading and soft dough phase of the wheat plot.

Over the non-irrigated wheat plot presented in Figure 7c, no irrigation points have been detected for 82 SAR images (18 months). Such results support our criteria for detecting irrigation events and eliminating ambiguities with rainfall or vegetation effects. Following a rainfall event both σ^0_P and σ^0_G increase and then decrease following a dry period. The consistency between both σ^0_P and σ^0_G for the non-irrigated plot during the whole period indicates that the plot did not receive any water supplement other than rainfall events. Despite several frequent changes of the σ^0_P over the complete temporal series due to several rainfall events and

vegetation change, the method was able to eliminate all possible irrigation ambiguities and judge that the plot did not receive any irrigation event.

An important point regarding the wheat plot in Figure 7c is the period between 19/04/2018 and 13/05/2018. During this period, the SAR backscattering signal increases gradually between each two SAR dates (red dashed circle). This increase in the SAR signal is related to the transition between the heading phase (minimum point on 19/04) and the soft dough phase (maximum point at 19/05). However, these points were eliminated using the proposed cereal phenology filter discussed in Section 3.4.

The proposed method was applied over a wide area of Catalonia region including 159,850 plots (123,428 non-irrigated and 36,423 irrigated plots) of different crop types. Despite the absence of information about the exact irrigation dates over Catalonia, a quantitative analysis was performed by comparing the results obtained by applying the method over irrigated and non-irrigated plots. Over Catalonia, the morning acquisition was 12 h prior to the evening acquisition. Figure 8 presents the histogram of the distribution of the number of events detected over irrigated and non-irrigated plots. Figure 8a corresponds to the morning SAR acquisition and Figure 8b corresponds to the evening SAR acquisition. The intersection between the evening and morning acquisitions is presented in Figure 8c. The intersection means that at each SAR date the point is considered as irrigation point if it exists within both the ascending and the descending acquisition modes (only 12 h difference between the two acquisitions is approximately equal to same day). Finally, Figure 8d shows the result of the combined use of both acquisitions. The combination between the results of the both acquisition modes means that at a given SAR date, the point is considered irrigation point if it exists in either the ascending or the descending acquisition modes. For morning acquisition mode (Figure 8a), the distribution shows that 68.3 % of the non-irrigated plots had no detected irrigation events for the period between September 2017 and December 2018 (82 SAR images). Moreover, 20.3 % of the non-irrigated plots encountered only one detected event. Therefore, 88.6% of the non-irrigated plots have maximum one detected irrigation point in the morning acquisition mode. The non-irrigated plots with 2 and 3 detected irrigation points represent only 6% and 2% respectively. On the other hand, only 12.1% of the irrigated plots failed to register any irrigation event. Thus, 87.9% of the irrigated plots had one and more detected irrigation points. The percentage then increases gradually where 54.9% of the irrigated plots had between two and five detected irrigation points. In addition, 20.4% of the irrigated plots encountered between five and 10 detected irrigation points. Similar results are obtained for the evening acquisition

(Figure 8b). For example, 58.2% of the non-irrigated plots had no detected irrigation events while 25.1% had only one detected event. Thus, 83.3% of the non-irrigated plots encountered a maximum of one detected point for evening acquisition mode. In contrast, only 9.1 % of the irrigated plots failed to gain any detected irrigation point and therefore 91.9. % of the irrigated plots had one and more detected irrigation point in the evening acquisition mode. Moreover, 51.1% of the irrigated plots had between two and five events and 27.3% of the irrigated plots had between six and ten detected points.

The intersection between the morning and evening acquisitions modes (Figure 8c) shows that 90.2% of the non-irrigated plots have no detected peaks whereas 72.4% of the irrigated plots had one or more detected irrigation peak. The combination of both acquisition modes (Figure 8d) increases the number of detected irrigation points but leads to an accuracy of the same order of magnitude as in the case of a separate use of the two SAR acquisition modes or in the case of the intersection of the two acquisition modes. For non-irrigated plots, 45.9 % of the plots had no detected points and 27.0 % had one detected points. The percentage decreases gradually to then reach 7% for three detected irrigation points. Inversely, the irrigated plots with no detected irrigation points consist of only 6.8% while the percentage of plots varies between 5% and 10% for irrigation points between 1 and 14 points.

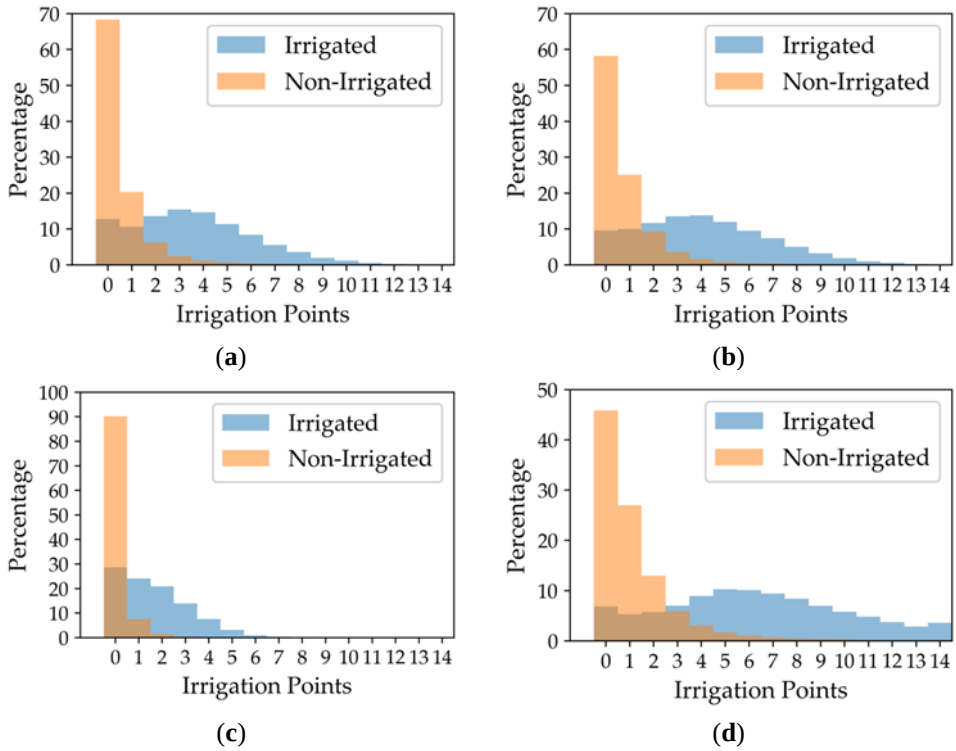


Figure 8: Distribution of the irrigation points detected over irrigated (blue) and non-irrigated (orange) plots in Catalonia Spain for (a) morning SAR acquisition, (b) evening SAR acquisition, (c) intersection

between morning and evening acquisitions and (d) combination between morning and evening acquisitions

Figure 9 represents the distribution of the number of irrigation points detected over irrigated plots in Catalonia, Spain as a function of months. For each month (including approximately five SAR acquisitions at each acquisition mode), the number of irrigation peaks detected over all the irrigated plots is calculated. While Figure 9a represents the distribution of irrigation points for the morning acquisitions, Figure 9b shows this distribution for the evening acquisition. For the morning acquisition mode (Figure 9a), the number of irrigation points reaches a low value in February 2018 (1040 points ~0.7%) and then starts to increase gradually between March 2018 and May 2018. A sharp increase of the irrigation points exists between May 2018 (3827 ~3%) and June 2018 (21,787 ~16%) and then continues to increase to reach a maximum value in July 2018 (25,380 ~19%). In August 2018, the number of irrigation points slightly decreases to attain approximately 15%. Then the number of detected irrigation points starts to decrease until reaching a minimum value in November 2018 (103 ~0.01%). Similar behavior is registered in the evening acquisition (Figure 9b) where the number of irrigation points reaches low value in February 2018 (2735 ~1%) and then increases until reaching its maximum value in July 2018 (29,286 ~18%). A slight decrease is recorded in August 2018 (27,252 points ~16%). The number of detected irrigation points then starts to decrease to reach a minimum value of 257 points ~0.01% in November 2018.

As a result, Figure 9 shows that most of the detected irrigation points exist within the period between March 2018 and September 2018. Indeed, 80.0% of the total detected points using the evening acquisition mode (Figure 9b) exist within the period between 01 March 2018 and 30 September 2018. Similarly, 74.5% of the total detected points exist within the same period for the morning acquisition mode (Figure 9a). Only 12% of the irrigation points (in both morning and evening) correspond to the period between November 2017 and February 2018. This low percentage mostly corresponds to the false detection of irrigation events since irrigation rarely occurs over this period and rainfall events are abundant. Therefore, the results show that most of the detected points using our proposed irrigation detection method correspond to the dry season where irrigation mostly occurs in the region.

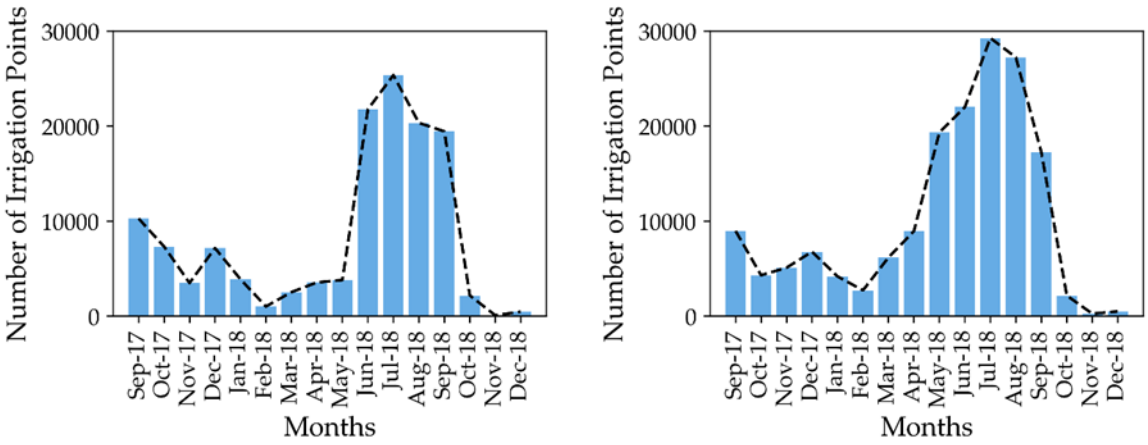


Figure 9: Distribution of the irrigation points detected over irrigated plots in Catalonia Spain as a function of months for (a) morning SAR acquisition, (b) evening SAR acquisition

4.4. Classifying Irrigated and Non-Irrigated Plots over Catalonia

Based on the results obtained in Section 4.3, we propose to classify irrigated and non-irrigated plots over the Catalonia site using the obtained distributions of detected irrigated points present in Figure 8. Four different classifications were performed. The first classification was performed using the results obtained from the morning SAR acquisitions (Figure 8a), and the second classification was executed using the results of evening SAR acquisitions (Figure 8b). A third classification was undertaken using the intersection between the morning and evening results (Figure 8c), and finally a fourth classification was executed using the combination of the morning and evening results (Figure 8d). As stated in 4.4, the intersection means that at each SAR date, a point is considered an irrigation point if it exists in both SAR acquisition modes while the combination means that the point is considered as irrigation point if it exists in either one of the two modes. For the morning and the evening classification scenarios, we suppose that a plot is considered irrigated if our method identifies two and more irrigation points within the complete SAR temporal series (morning or evening). For the intersection classification scenario, we consider that a plot is an irrigated plot if the intersection between the morning and the evening results gives one and more irrigation points. Finally, for the combined classification we consider that a plot is irrigated if the combination of the morning and the evening results gives three and more irrigation points. For each of the four classifications, a confusion matrix is built and Table 3 reports the accuracy metrics by means of the overall accuracy, weighted F-measure, and the F-Measure at each class (irrigated, non-irrigated). The overall accuracy is a standard metric used for remote-sensing applications. The weighted F-measure corresponds to the harmonic mean affected by the number of samples.

Since the number of non-irrigated plots is bigger than the irrigated plots, the weighted F-measure is well suited to evaluate the performance of the classification (Tan et al., 2006)

The highest overall accuracy (85.9%) was recorded for the intersection scenario with a weighted F-measure of 86.0%. The F-score of the irrigated class in the intersection scenario reaches 0.70 and that for the non-irrigated class reaches 0.90. On the other hand, the lowest overall accuracy was recorded for the evening SAR acquisition (82.5%) accompanied with the lowest weighted F-measure (83.4%). The combined scenario shows the highest accuracy for the irrigated class (F-measure = 0.72) with an overall classification accuracy of 85.4%. Generally, the F-measure of each class was nearly the same between the four scenarios. In fact, the F-measure varies between 0.68 and 0.72 for the irrigated class, and varies between 0.88 and 0.90 for the non-irrigated class.

Table 3: The values of the overall accuracy and F-measure obtained for classification of irrigated and non-irrigated plots using the detected irrigation points for four different scenarios.

Scenario	Condition to be Irrigated	Class	F-Measure	Weighted F-Measure	Overall Accuracy
SAR Morning	Two points and more	Non-irrigated Irrigated	0.90 0.71	86.1%	85.7%
SAR Evening	Two points and more	Non-irrigated Irrigated	0.88 0.68	83.4%	82.5%
Intersection Morning and Evening	One point and more	Non-irrigated Irrigated	0.90 0.70	86.0%	85.9%
Combined Morning and Evening	Three points and more	Non-irrigated Irrigated	0.89 0.72	84.7%	85.4%

The four proposed classification scenarios were re-established using only the detected irrigation events between April and September 2018 which corresponded to the irrigation period in Catalonia. Figure 10 shows a comparison between the accuracy metrics previously obtained when using all the detected irrigation events (between September 2017 and December 2018) and that obtained when using only irrigation events between April 2018 and September 2018. For the morning and evening classification scenarios, the results show that the overall accuracy increased by 2.8% and 4.8%, respectively, when considering only detected irrigation events between April and September (Figure 10a). The F-measure of the irrigated class increased by 2.2% for the morning scenario and significantly increased by 5.0% for the evening classification (Figure 10b). For the intersection scenario, the obtained results remained nearly the same for the four accuracy metrics. Moreover, in the combined classification scenario, the overall accuracy increased by 4.0% (Figure 10a) where the F-measure of the irrigated class increased by 4.5% (Figure 10b). The F-measure of the non-irrigated class (Figure 10c) also

increased for the four different scenarios. Finally, the weighted F-measure also increased by 2.0%, 4.0% and 3.4% for the morning, evening and the combined classification scenarios, respectively (Figure 10d). Therefore, a priori information about the irrigation period can help reduce the uncertainty in the proposed model for better irrigation detection. This priori information helps limiting the size of the studied temporal series and thus reduces the chance of detecting additional false irrigation events.

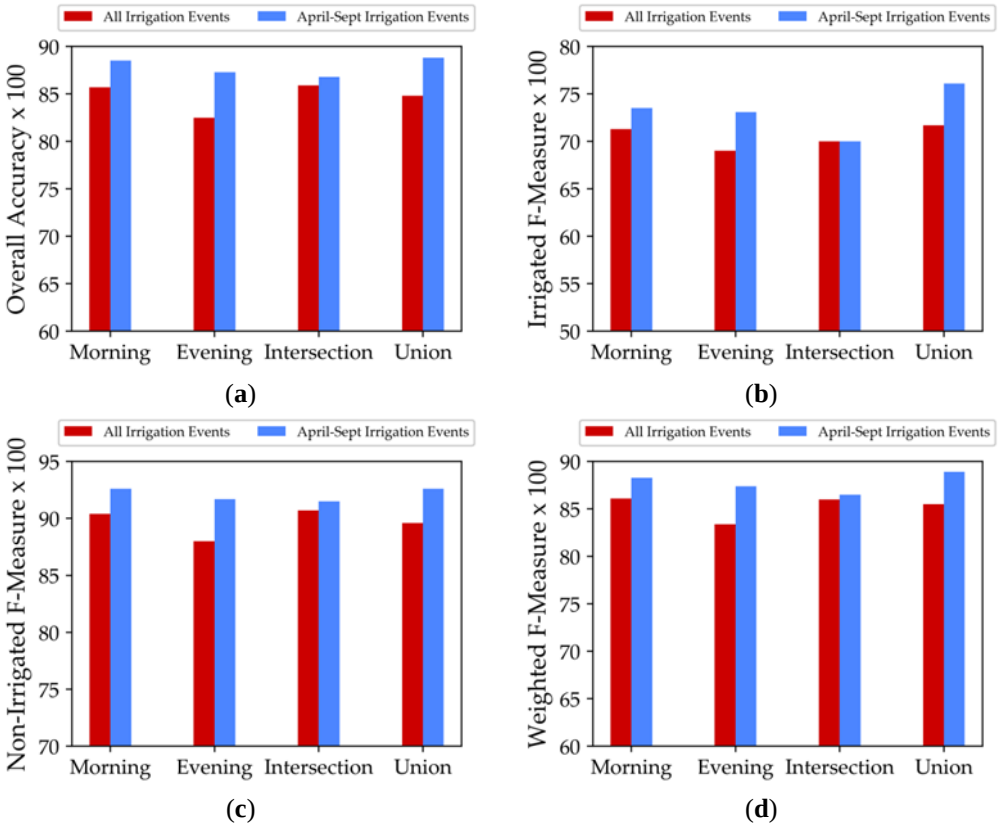


Figure 10: Comparison between accuracy metrics obtained for classification of irrigated and non-irrigated plots using all the detected irrigation events (red bar) and the detected irrigation events between April and September 2018 (blue bar). (a) Overall accuracy, (b) F-measure of irrigated class, (c) F-measure of non-irrigated class and (d) weighted F-measure.

4.5. Results over Tarbes

The proposed method was applied over irrigated plots in Tarbes, south-west France. Figure 11a represents the σ^0_P and σ^0_G in the morning acquisition mode for an irrigated maize plot with the irrigation points detected using the proposed change detection method. While Figure 11a shows the complete series between 01 March and 30 November 2017, Figure 11b represents the period where the irrigation points were detected. However, irrigation dates were not available over this site so a qualitative analysis was performed to show the performance of the method. During the period between 01 March 2017 and the end of June 2017 (Figure 11a), the

σ_P^0 and σ_G^0 followed the same behavior for almost all the dates. Following a rainfall event both curves increased and then decreased following a period without rainfall. This consistency could be related to the absence of any additional water supplement on the plot scale. However, between March and mid-May, the NDVI values were low (less than 0.2) indicating bare soil conditions. This also supports the possibility that no irrigation had occurred for this period. The NDVI then started to increase showing a vegetation cycle between May and October.

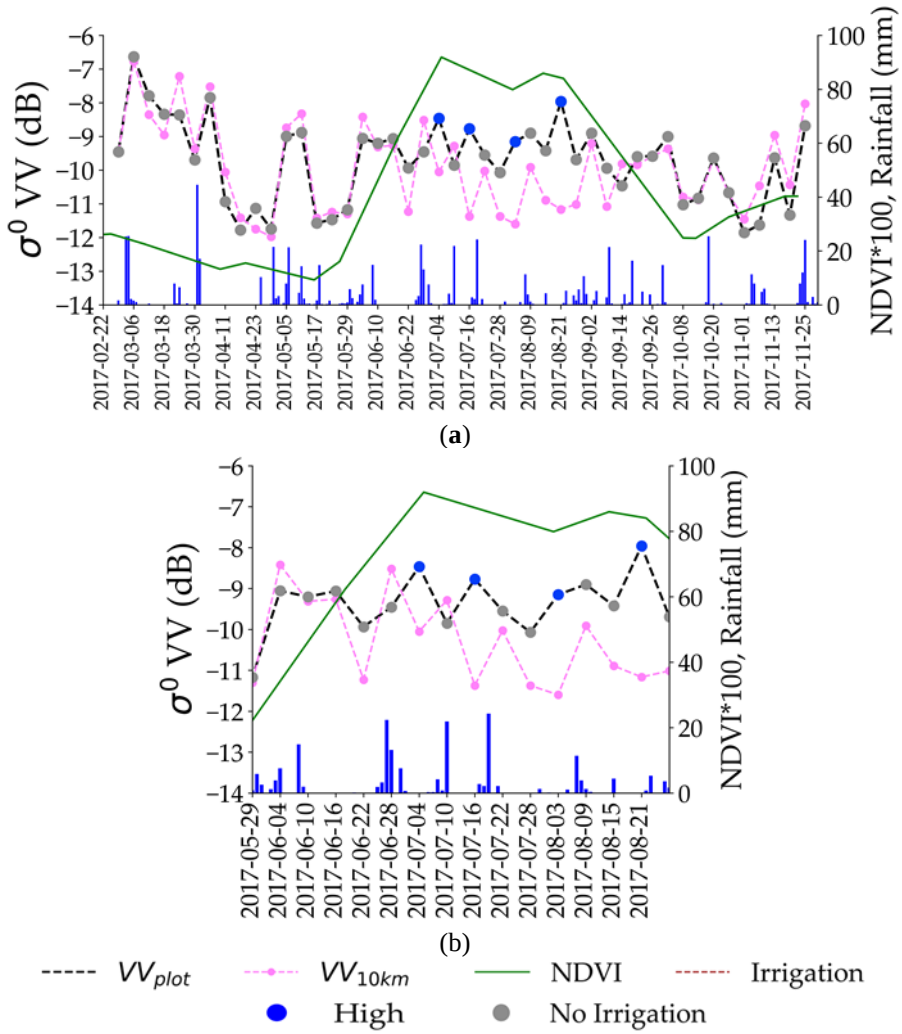


Figure 11: Irrigation events detection over a reference irrigated maize plot in Tarbes, France. (a) Complete temporal series between March and November 2017; (b) irrigation events detected between 29 May and 21 August 2017.

Using the proposed method for detecting irrigation events, four irrigation points were detected over this plot (Figure 11b). In general, all these irrigation points corresponded to an important increase of σ_P^0 accompanied with absence of rainfall events as shown by the GPM data (blue curve) and the σ_G^0-values at grid scale (pink curve). For example, an important cumulative rainfall of approximately 40 mm three days before the SAR image acquired on 28/06/2017 caused σ_G^0 to increase by 2.5 dB. Six days later, no rainfall events were registered

and the σ^0_G decreased by 1.5 dB. For the same period, σ^0_P increased by 1 dB. This point was the first irrigation point detected on the plot. Similarly, between 10/07 and 16/07, σ^0_G decreased by 2.1 dB due to the absence of precipitation while σ^0_P increased by 1.1 dB indicating that a water supplement could have occurred. The absence of precipitation for 12 days between 22/07 and 04/08 caused σ^0_G to decrease gradually. However, the SAR image acquired on 04/08 showed an increase of the SAR signal at plot scale which was evidence of an irrigation event occurring. Likewise, between 15/08 and 21/08 the σ^0_P increased by 1.5 dB indicating the presence of irrigation event along with a stability of low σ^0_G (~ -11.5 dB) indicating the absence of any rainfall events. After the fourth detected point on 21/08, the NDVI started to decrease. Usually during this phase irrigation events rarely occur. After 02/09, the σ° values at both plot and grid scale regained their consistent behavior indicating the absence of irrigation. Finally, it is important to mention that the summer season in Tarbes is humid and encountered several rainfall events.

Figure 12 presents the histogram of the distribution of the number of events detected over irrigated plots in Tarbes for morning acquisition (Figure 12a), evening acquisition (Figure 12b), intersection between morning and evening acquisitions (Figure 12c) and the combination of both acquisitions (Figure 12d). It is important to remember that the intersection means that the detected points are considered irrigation events if they exist within the morning and the evening modes while the combination means that the points are considered irrigation events if they exist in either the morning or the evening acquisition modes. Figure 12a shows that using the morning acquisition, 91% of the irrigated plots has one or more detected irrigation events and 75% of the plots have two and more detected irrigation events. Figure 12b shows that fewer points are detected in the evening acquisition than the morning acquisition where only 50% of the plots have one and more irrigation events. In Figure 12c, the intersection shows that with an interval of 36 h between both acquisitions, only 38% of the irrigation events were commonly detected within both acquisition modes. Finally, the combined use of the morning and the evening acquisitions in Figure 12d shows that 97% of the irrigated plots had and more detected irrigation events. Moreover, 90% of the irrigated plots had two and more detected irrigation events and 65% of the plots had three and more irrigation events. This indicates that the combined use of the morning and the evening acquisitions helps detect more irrigation events than using only one S1 overpass each 6 days. Figure 12d also shows that the maximum percentage of plots was registered with three irrigation events (29.7%) and the percentage then decreased as the number of detected irrigation events increase (four events and more). The low number of detected

irrigation events in Tarbes, compared to those obtained over Catalonia, was expected since Tarbes is a humid region and the frequency of irrigation is less than that in the arid or semi-arid regions. Moreover, the abundance of rainfall events during the summer season could restrict the possibility of detecting all the possible irrigation events.

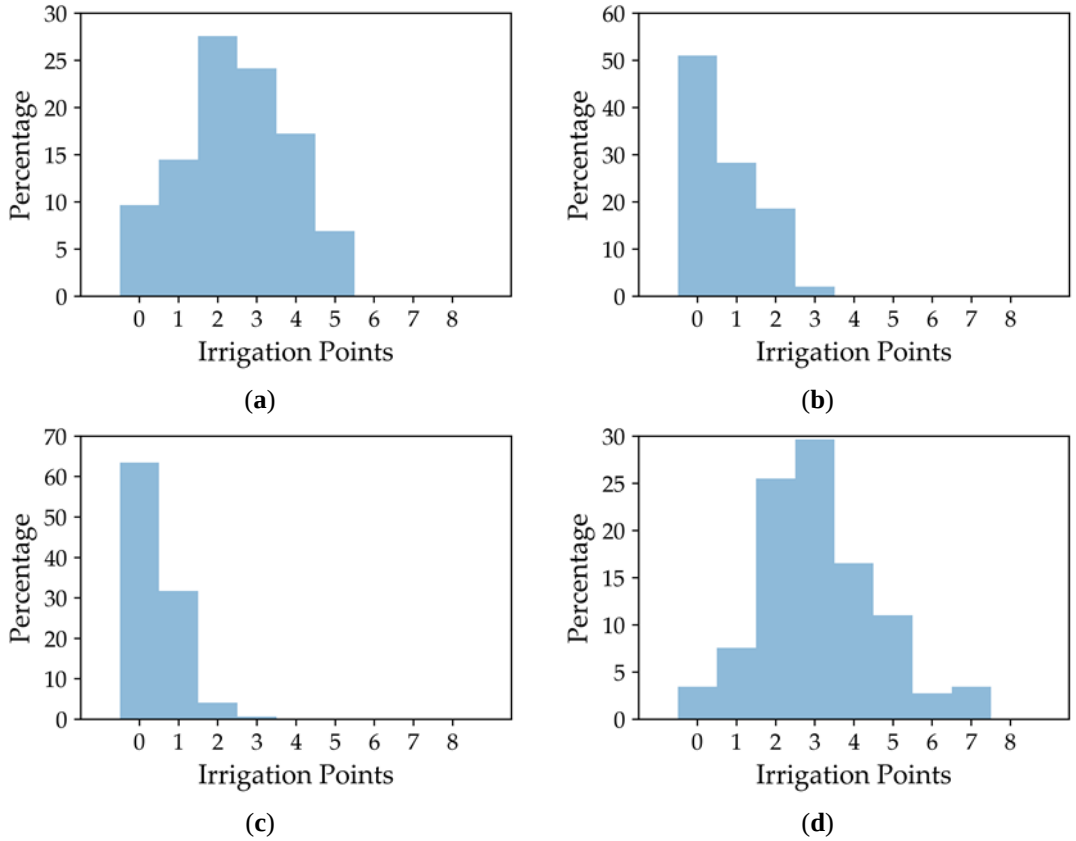


Figure 12: Distribution of the irrigation points detected over irrigated plots in Tarbes, France for (a) morning SAR acquisition, (b) evening SAR acquisition, (c) intersection between morning and evening acquisitions and (d) combination between morning and evening acquisitions.

5. Discussion

5.1. Change Detection in σ° SAR Backscattering

In this study, a new methodology is presented for detecting irrigation events at plot scale using the change detection in the SAR backscattering signal. The first important fact is that the change of the surface soil moisture due to irrigation events (artificial application of water) between two SAR dates may cause an increase in the SAR backscattering coefficient between these two dates. However, this assumption remains limited with the changes in SAR backscattering signal due to a rainfall event or vegetation development. Therefore, the separation between the changes in SAR backscattering signal due to irrigation events from similar changes, which could be caused by rainfall or vegetation development, was challenging.

To remove the uncertainty with rainfall events, the method proposes using the SAR backscattering signal at basin scale (10km × 10km) obtained from bare soil plots with low vegetation cover as a descriptor of rainfall events. The assumption stipulates that if the mean SAR signal within 10km x 10km grid cell increased between two consecutive dates then a rainfall event took place and thus irrigation, if it had occurred, could not be detected. This dependency between rainfall and σ_G^0 was presented in Figure 4. For this reason, we judge that any increase in the SAR signal at 10km grid scale by more than 1 dB is linked to rainfall events. In addition, the surface soil moisture estimation performed at grid scale (SSM_G) was used to enhance the detection of rainfall events. Thus, a threshold of 20 vol.% for SSM_G was proposed to describe humid soil conditions at basin scale which is probably linked to a rainfall event that occurred couple of days before the SAR acquisition. The combination of these two conditions at grid scale allowed a good separation between rainfall and irrigation episodes occurring at each SAR acquisition.

The second main contributor in the SAR backscattering signal at plot scale was the development of the vegetation cover. This development could cause an increase in the SAR backscattering signal at plot scale without any change in the surface soil moisture content and thus without any irrigation or rainfall events. In this study, a descriptor of the vegetation growth is suggested. This descriptor (S) is a Gaussian-smoothing filter applied to the SAR temporal series. The S values allow describing the vegetation growth pattern at the plot scale and thus permit the separation of irrigation events from vegetation growth events. This contribution of the smoothed-Gaussian is shown in Figure 6. For the soya plot (P2) of Figure 6, the vegetation descriptor (S) helped eliminate two points considered as vegetation development points. Another vegetation development filter was proposed for cereal plots in order to remove the increase in the SAR backscattering signal due to the transition between the heading and the soft dough phenology phases. The importance of this filter was demonstrated in Figure 7c for a non-irrigated winter wheat plot. However, the date limits present in this filter (mid-March until May) could be adjusted for other geographical contexts in order to follow the cereal growth cycle where cereals could be cultivated in other months.

5.2. Effect of NDVI Optical Filter

Throughout the study, an optical post-processing filter has been suggested to ameliorate the detection of irrigation events and remove the ambiguity between the increase in the SAR backscattering signal due to irrigation events and soil work. The role of the NDVI filter is to

insure that a vegetation growth cycle truly exists within or after the detected irrigation point. As presented in Section 3.6, the NDVI optical filter suggests that if the $NDVI_{ti}$ (NDVI value at a detected irrigation event) is less than 0.4 and $\Delta NDVI \leq 0.1$ then the point is a falsely detected irrigation point, and is further eliminated. For $NDVI_{ti}$ greater than 0.4, the filter was no longer applied because the vegetation cover is already developed (NDVI >0.4) and the existence of irrigation event is more probable.

To assess the importance of this NDVI optical filter, we obtain first the results over Catalonia without applying the NDVI optical filter. Then we apply the NDVI optical filter using two threshold values on $NDVI_{ti}$ accompanied with the criterion $\Delta NDVI \leq 0.1$. The filter was first applied using $NDVI_{ti} \leq 0.2$ and then applied using $NDVI_{ti} \leq 0.3$. The results obtained previously with $NDVI_{ti} \leq 0.4$ were also used in this comparative analysis. For each case, the threshold of $\Delta NDVI$ was kept equal to 0.1. To present the effect of this NDVI optical filter, we intended to monitor the change in the percentage of plots that did not encounter any irrigation event for both irrigated and non-irrigated plots (0 detected irrigation events). The class with no detected irrigation events was chosen for this analysis since it describes the capability of discriminating between irrigated and non-irrigated plots. Figure 13 presents the evolution of the percentage of the plots that did not register any irrigation event as a function of the value of $NDVI_{ti}$ used in the optical filter for both morning (Figure 13a) and evening (Figure 13b) acquisition modes.

For the morning acquisition mode, Figure 13a shows that when the NDVI filter is not applied only 29% of the non-irrigated plots have no detected irrigation events. When adding the NDVI filter using $NDVI_{ti} \leq 0.2$ ($\Delta NDVI \leq 0.1$), this percentage increases to reach 44%. The optical filter applied at $NDVI_{ti} \leq 0.3$ increases the percentage of non-irrigated plots with no detected irrigation events to reach 59%. Finally, using the NDVI optical filter at $NDVI_{ti} \leq 0.4$, 68% of the non-irrigated plots had no detected irrigation events. Beyond 0.4, the filter was no longer applied since vegetation already exists and the occurrence of an irrigation event was more probable than soil work. For irrigated plots in morning acquisition (Figure 13a), the percentage of irrigated plots with no detected irrigation events was 4% when no NDVI filter was added and slightly increased to 12% as the NDVI optical filter with $NDVI_{ti} \leq 0.4$ was used. Therefore, when applying the NDVI filter using $NDVI_{ti} \leq 0.4$ in the morning acquisition mode, the percentage of the non-irrigated plots with no detected irrigation events significantly increased by 39% (68%–29%) compared to the percentage obtained when no NDVI filter was

used. The percentage of the irrigated plots with no detected irrigation events increased only by 8% (12% minus 4%) compared to that obtained when no NDVI filter was used.

For the evening acquisition mode (Figure 13b) the results show that the percentage of non-irrigated plots with no detected events increased from 22% when no filter was applied to reach 35%, 49% and finally 58% when the $NDVI_{ti} \leq 0.2$, $NDVI_{ti} \leq 0.3$ and $NDVI_{ti} \leq 0.4$ were used respectively. For the irrigated plots in evening acquisition, the percentage increased from 3% when no filter was applied to reach 9% when NDVI optical filter was applied at $NDVI_{ti} \leq 0.4$. Therefore, when applying the NDVI filter using $NDVI_{ti} \leq 0.4$ in the evening acquisition mode, the percentage of the non-irrigated plots with no detected irrigation events increased by 36% (58%–22%) compared to the percentage obtained when no NDVI filter was applied. On the other hand, the percentage of the irrigated plots with no detected irrigation events increased only by 6% (9%–3%) compared to that obtained when the filter was not applied.

The significant increase in the percentage of non-irrigated plots with no detected peaks, between the case where NDVI filter was not applied and NDVI filter was applied at $NDVI_{ti} \leq 0.4$, encouraged the use of the post-processing optical filter despite of the loss of 8% and 6% in the irrigated plots for morning and evening acquisitions, respectively.

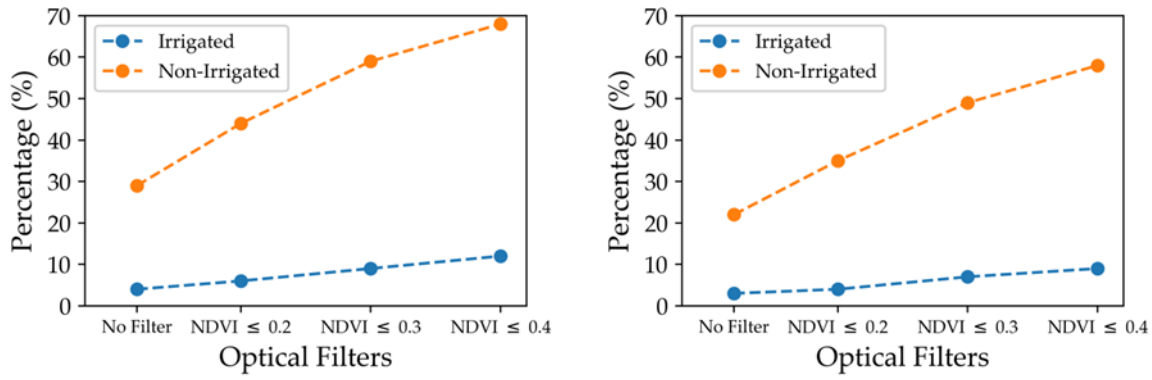


Figure 13: Percentage of plots with no detected irrigation points a function of the value of $NDVI_{ti}$ used for the optical filter in Catalonia, Spain for irrigated (blue) and non-irrigated (orange) plots using (a) morning acquisition mode and (b) evening acquisition mode.

5.3. Strengths, Limitations and Future Directions

The results of the study demonstrate that irrigation events could be detected using the Sentinel-1 satellite data with 6 days temporal resolution. Using the proposed method, the spatial distribution of the plots that encountered an irrigation event could be obtained at each available

SAR acquisition. Currently, the Sentinel-1 satellite is the only operational satellite that provides SAR data with high revisit time (6 days). However, this revisit time could still affect the detection of irrigation events. The first limitation of the proposed method is the effect of the time lag between the irrigation episode and the satellite passage. In fact, the detection of irrigation can become difficult if the irrigation event takes place more than three days before the SAR acquisition. Hajj et al. (2014) showed that irrigation could be detected if the irrigation took place three days and less before the radar acquisition. In their study using X-band SAR data, they showed that more than three days after the irrigation event, the surface soil moisture could attain the same value as that before the irrigation due to evaporation. Thus, a SAR acquisition more than three days after the irrigation event could not show any change in the surface soil moisture and the irrigation could not be detected. However, an analysis of the detected and non-detected irrigation events over Montpellier site using the C-band SAR data reveals that detection of irrigation events is not necessarily limited to the three-day time lag. For example, some detected irrigation points were found to be occurring four days before the SAR acquisition whereas few points were not detected even two days after the SAR acquisition. Thus, additional information such as the type of irrigation, the quantity of water, the evaporation rate and the plant water needs could be used to confirm the presence or absence of such irrigation events.

In general, to overcome the limitation of the time lag between the SAR date and irrigation date, a dense SAR temporal series is required. Indeed, within the overlapping area of the different orbits of the Sentinel-1, four images (two morning and two evening images) could be obtained each 6 days. Thus for some sites, we could obtain up to 20 images each month (10 morning and 10 evening images) which could help easily detect all the irrigation events. This configuration mainly depends on the surface area of the studied site in order to be covered by the overlapping area of different Sentinel-1 overpasses. The dense temporal series of SAR acquisitions (20 images per month) could also help increase the number of possibly detectable irrigation events. The number of possibly detected irrigation events mainly depends on the availability of the SAR images. As shown in the results over Montpellier (Section 4.2), 69% of the irrigation events could be possibly detected by SAR data using the available SAR configuration over the study area (10 images each month). Thus, as the number of SAR images increases, the chance of detecting all the existing irrigation events increases.

Another limitation for detecting irrigation events could be the existence of a very well developed vegetation cover. In fact, the value of ΔVV_P over high vegetation cover could be

lower than that for bare soil (Gao et al., 2017). However, this difference in ΔVV_P between bare soil and vegetated soil is a difficult aspect to quantify but can cause some uncertainty.

The strength of the method resides in the near real-time detection of irrigation episodes. Despite the post-processing step required in the NDVI optical filter, the method could still be operationally applied in near real-time. Unlike other approaches that may require the complete temporal series to perform irrigation mapping or detection, the proposed method could be a near real time application for irrigation detection.

The classification results obtained in Section 4.4 show that the proposed method could be a competing method for classifying irrigated plots versus machine-learning approaches that require an extensive training database in order to obtain good classification results. For example, Gao et al. (2018) and Bazzi et al. (2019c) used machine learning models such as the random forest and the convolutional neural network with S1 temporal series to classify irrigated/non-irrigated plots in Catalonia, Spain. Both studies achieved an overall accuracy of 82% and 90%, respectively. Using our approach, we were able to separate irrigated and non-irrigated plots with an overall accuracy of 85.9% over the same study site (4.4).

Generally, machine-learning models used for classification tasks usually depend on the studied sites and the studied period of the year. In our case, the proposed method was tested over three different study sites with different geographical and climatic properties. This indicates that the proposed method is not site-dependent. Since no specific site calibration is required in the proposed approach, the proposed method is thus capable of providing a non-supervised tool for monitoring irrigation activities at plot scale. This reflects the strength of the proposed method against the transfer of the machine-learning models from one site to the other. However, the threshold values used could still be modified to obtain optimum results over another study sites. The change of the thresholds, if needed, does not require extensive dataset from in situ observations for irrigated plots.

By analyzing the number of obtained irrigation points on each field (as for Section 4.3), our proposed method could be used to generate labelled samples of irrigated/non-irrigated plots. This means that the proposed method could be used to create a training/validation dataset, over any study site, which could be then used in machine-learning models such as random forest or deep-learning approaches. Our future work will concentrate on merging our proposed method with machine learning methods to obtain a semi-supervised classification model in which the training/validation data will first be selected using the proposed approach in this study, and then

a machine-learning model will be built to perform further irrigation classification. In this case, the proposed method could replace the terrain campaigns usually done for collecting rich set of labelled samples in order to perform machine-learning classification approaches.

6. Conclusions

To deal with the aspect of managing water resources in the agricultural sector, this paper proposes a new approach for detecting irrigation events at plot scale in a near real-time scenario. The proposed method is a decision tree-based approach for detecting irrigation events using the change detection in the S1 SAR backscattering coefficients at plot scale. Several filters were applied in order to remove the ambiguity between irrigation events and rainfall, vegetation development and soil surface roughness. A thresholds-based method was proposed to detect irrigation events at each available S1 SAR date and for each agricultural plot mainly using the backscattering S1 SAR signal at plot and at grid scale (10 km × 10 km). Finally, a post-processing filter based on the NDVI value was integrated to ameliorate the detection of irrigation events.

To ensure the transferability and the possible operational application of the proposed approach, three study sites were examined in this study (Montpellier, Catalonia and Tarbes) with two different climatic properties. In terms of climate, both Montpellier and Catalonia are similar (Mediterranean) with dry summer season whereas Tarbes is a humid region with frequent rainfalls in the summer season. Results showed, first, that the proposed method was capable of detecting 84.8% of the irrigation events occurring at three plots encountering 33 irrigation events in Montpellier. Then, an analysis performed over the semi-arid region in Catalonia revealed that the proposed method was capable of classifying irrigated and non-irrigated plots with an overall accuracy of 85.9%. Finally, the analysis performed over irrigated plots in Tarbes revealed that our proposed method is able to detect irrigation events even in the presence of frequent rainfall events in the summer season where 90% of irrigated plots were detected with two and more irrigation events.

Our irrigation events detection method opens the way toward building new semi-supervised approaches for irrigated area mapping at plot scale. Since several machine-learning models require an extensive dataset with costly terrain measurements, this method could be used to create a dataset of irrigated/non-irrigated labelled samples in order to be used in a machine-learning model that could be more efficient. Our future work will concentrate on

combining this new method with a supervised classification model to obtain a semi-supervised model for irrigation mapping. This combination can permit the transfer of machine-learning classification models over several regions to perform irrigation mapping at plot scale.

Author Contributions: Conceptualization, H.B. and N.B.; Data curation, H.B.; Formal analysis, H.B., N.B. and I.F.; Methodology, H.B., N.B. and I.F.; Supervision, N.B.; Validation, M.Z., H.B. and V.D.; Writing – original draft, H.B.; Writing – review and editing, N.B., I.F., M.Z., H.B. and V.D.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2020 project), the National Research Institute for Agriculture, Food and the Environment (INRAE), the Occitanie region, France and the Mediterranean Agronomic Institute of Montpellier (CIHEAM-IAMM).

Acknowledgments: Authors wish to thank the French Space Study Center (TOSCA 2020), the National Research Institute for Agriculture, Food and the Environment (INRAE) and the Mediterranean Agronomic Institute of Montpellier (CIHEAM-IAMM). The authors wish to thank the Occitanie region, France for supporting this work. The authors wish also to thank the European Space Agency (ESA) for the Sentinel-1 and Sentinel-2 data and Theia pole for the calibration of the Sentinel-2 images. The authors would like to thank Maria Jose Escorihuel from the isardSAT, Catalunya for providing the SIGPAC data. Finally, authors would like to thank Fabien Dauriac of the Chamber of Agriculture of Hautes- Pyrénées, France, for providing terrain data over Tarbes, France.

Conflicts of Interest: The authors declare no conflict of interest.

3. Article 6: Irrigation Events Detection over Intensively Irrigated Grassland Plots Using Sentinel-1 Data

Author version of the article published in Remote Sensing Journal: Volume: 12, Issue: 24, December 2019

Hassan Bazzi[1], Nicolas Baghdadi[1], Ibrahim Fayad[1], François Charron[2], Mehrez Zribi[3], Hatem Belhouchette[4]

[1] INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2] G-EAU Unit, University of Montpellier, AgroParisTech, CIRAD, INRAE, Institut Agro, IRD, Domaine du Merle, 13300 Salon de Provence, France

[3] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

[4] CIHEAM-IAMM, UMR-System, 34090 Montpellier, France

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Received: 10 November 2020; Accepted: 9 December 2020; Published: 11 December 2020

Abstract

Better management of water consumption and irrigation schedule in irrigated agriculture is essential in order to save water resources, especially at regional scales and under changing climatic conditions. In the context of water management, the aim of this study is to monitor irrigation activities by detecting the irrigation episodes at plot scale using the Sentinel-1 (S1) C-band SAR (synthetic-aperture radar) time series over intensively irrigated grassland plots located in the Crau plain of southeast France. The method consisted of assessing the newly developed irrigation detection model (IDM) at plot scale over the irrigated grassland plots. First, four S1-SAR time series acquired from four different S1-SAR acquisitions (different S1 orbits), each at six-day revisit time, were obtained over the study site. Next, the IDM was applied at each available SAR image from each S1-SAR series to obtain an irrigation indicator at each SAR image (no, low, medium, or high irrigation possibility). Then, the irrigation indicators obtained at each image from each S1-SAR time series (four series) were added and combined by threshold value criteria to determine the existence or absence of an irrigation event. Finally, the performance of the IDM for irrigation detection was assessed by comparing the in situ

recorded irrigation events at each plot and the detected irrigation events. The results show that using only the VV polarization, 82.4% of the in situ registered irrigation events are correctly detected with an F_score value reaching 73.8%. Less accuracy is obtained using only the VH polarization, where 79.9% of the in situ irrigation events are correctly detected with an F_score of 72.2%. The combined use of the VV and VH polarization showed that 74.1% of the irrigation events are detected with a higher F_score value of 76.4%. The analysis of the undetected irrigation events revealed that, in the presence of very well-developed vegetation cover (normalized difference of vegetation index (NDVI) $\geq$ 0.8); higher uncertainty in irrigation detection is observed, where 80% of the undetected events correspond to an NDVI value greater than 0.8. The results also showed that small-sized plots encounter more false irrigation detections than large-sized plots certainly, because the pixel spacing of S1 data (10 m × 10 m) is not adapted to small size plots. The obtained results prove the efficiency of the S1 C-band data and the IDM for detecting irrigation events at the plot scale, which would help in improving the irrigation water management at large scales especially with availability and global coverage of the S1 product

Keywords: irrigation; grassland; Sentinel-1; change detection; Crau plain; France

1. Introduction

In order to meet the global food demand, affected by the global increase in the world population and climate change, current policies are pushing toward the expansion of the irrigated areas especially in arid and semi-arid regions (Godfray et al., 2010; Tilman and Clark, 2015; Tilman et al., 2011). According to the FAO (Food and Agriculture Organization of the United Nations), irrigated agriculture currently shares more than 40% of the global food production and accounts for approximately 70% of the total freshwater consumption considered thus the main consumer of freshwater (Pokhrel et al., 2016). Under limited water resources, efficient management of irrigation schedule and timing is important to achieve the environmental sustainability in the agricultural sector (Ozdogan, 2011). In fact, obtaining information about irrigation timing is of great importance for several studies dealing with food security (Burney et al., 2010) and water consumption (Paredes et al., 2014).

Satellite remote sensing has proven its high capability and effectiveness for mapping and monitoring irrigated areas (Thenkabail et al., 2009a, 2005, 2009b). Recent studies have shown that irrigated areas could be spatially quantified over large scale using either passive optical

sensors (Dheeravath et al., 2010; Gumma et al., 2011; Xiang et al., 2019) or active radar sensors (Bazzi et al., 2019c; Bousbih et al., 2018; Gao et al., 2018). Optical images has been widely used to map irrigated areas using the difference between the spectral signature of irrigated crops and that of non-irrigated crops in the time series domain. The difference in the temporal signal between irrigated and non-irrigated crop is based on the fact that irrigation mixed with fertilizers makes the crop grow healthier and faster causing a difference in the captured spectral signature between irrigated and non-irrigated plot. This difference in the spectral signature has been investigated using several vegetation indices such as the NDVI (Normalized Differential Vegetation Index) (Pervez and Brown, 2010) or the GI (Greenness Index) (Chen et al., 2018). On the other hand, the Synthetic Aperture Radar (SAR) data has proven a superior efficiency over optical data for mapping irrigated areas (Bazzi et al., 2019c, 2019d; Fieuzal et al., 2011). The use of SAR data is mainly driven by the fact that the backscattered SAR signal is sensitive to the soil water content that is a key point in the irrigation activity (Aubert et al., 2013; Baghdadi et al., 2011a, 2016a; El Hajj et al., 2017). Since irrigation increases the soil water content, the detection of irrigated areas could be possible using SAR data that are highly affected by the soil moisture values.

Although huge effort has been performed to quantify the extent and spatial distribution of irrigated areas, the timing and frequency of irrigation has not yet received important attention despite their high importance in managing water resources. Using optical data, few studies have reported the detection of irrigation events. For example, Chen et al. (Chen et al., 2018) used MODIS (Moderate Resolution Imaging Spectroradiometer) and Landsat optical images with ancillary data (precipitation and soil moisture) to detect irrigation timing and frequency during the first half of the growing season. In their study, they performed a threshold-based model using the GI (Greenness index) to count the possible water supplement stages. To distinguish between rainfall and irrigation events, which are both counted as water supplements, they used local daily precipitation records. Their results show that the overall accuracy of detecting a water supplement stage using the proposed method reaches 87%. However, the cloud limit could restrict the use of optical data despite the high revisit time of the new optical satellites such as the Sentinel-2 (S2) satellite that reaches 5 days.

On the other hand, SAR data could be used to detect the wetness information caused by irrigation episodes due to the sensitivity of the SAR backscattered signal to the soil water content (Bazzi et al., 2019c; Bousbih et al., 2017, 2018; Gao et al., 2018). However, the detection of irrigation by SAR data depends, in addition to the vegetation cover (crop type,

density, growing phase …), on the characteristics of the SAR satellite used (mainly revisit time and radar wavelength). The revisit time of the SAR satellite could constraint the detection of all the irrigation events due to the rapid drying out of the soil within few days after the irrigation. Hajj et al. (2014) had shown that a radar image in X-band acquired 3 days after irrigation makes it difficult to certainly detect the irrigation event. This difficulty for detecting the irrigation event is due to the rapid dry out of the surface soil moisture which increases after an irrigation episode from 15-20 vol.% to reach 30-40 vol.%, and then the dries out three days after the irrigation event to reach the same value attained before irrigation (15-20 vol.%). Moreover, the used radar wavelength could constrain the detection of irrigation events. In fact, the available free and open access radar sensors provides C-band data (wavelength around 6 cm in the case of Sentinel-1). Recent studies (El Hajj et al., 2018b) have showed that the C-band SAR data could present certain limits for mapping soil moisture due to the low penetration of the SAR signal over certain very well vegetation cover (case of irrigated maize). Moreover, El Hajj et al. (2018b) and Nasrallah et al. (2019) showed that the sensitivity of the C-band SAR signal to soil moisture becomes negligible over wheat crops between the germination and the heading growing phases due to the low penetration of the C-band signal to the soil surface. Therefore, it is sometimes difficult to detect an increase in the radar signal due to an irrigation event because the soil contribution is very low. However, recent studies have started exploiting the potential of SAR data for detecting irrigation events at plot scale. Recently, Bazzi et al. (2020b) developed a decision tree based model to detect irrigation events at plot scale using S1 C-band SAR data. The proposed model relies on separating irrigation events from rainfall events using rainfall information derived from grid scale (10 km x 10 km) SAR backscattering signal (Bazzi et al., 2019b). By applying their model on three regions with different climatic contexts, they report that irrigation events could be detected by SAR data independent of the studied geographic context. In their study, they achieved a success rate of 84% in detecting irrigation events. However, they recommend that dense SAR temporal series help detect more irrigation events at plot scale. Since irrigation is a dynamic activity, the six-days revisit time of the S1 satellite remains a constraint for the detection of all irrigation events at plots scale. For this reason, an extensive temporal dataset is still required for the detection of all irrigation events occurring at plot scale. In addition, over irrigated maize plots of southwest France, Le Page et al. (2020) investigated the potential of the S^2MP (Sentinel-1/Sentinel-2-derived soil moisture product El Hajj et al. (2017) to detect irrigation events at plot scale. The S^2MP is a soil moisture estimation product mainly derived by coupling Sentinel-1 SAR data and Sentinel-2 optical data using the neural network (Bazzi et al., 2019a; El Hajj et al., 2017). They showed that irrigation

timing was detected with a good accuracy reporting an F-score between 80% and 83% for all studied plots. Lower accuracy of 69% was obtained when the vegetation cover is well developed (NDVI > 0.7). However, the study was only performed over five irrigated maize plots, which is still not considered robust enough for an operational application.

In France, the agricultural water consumption reaches 4.8-billion m^3 of the total withdrawn water. Particularly in of southeast France, the Crau plain is well known irrigated agricultural region exploited mainly for hay production. It covers an area of 12,500 ha and produces 100,000 tons of hay each year. The plain is intensively irrigated area that uses mainly the gravity irrigation technique. Despite the low irrigation efficiency of gravity irrigation, which is estimated at 60-70% (Tiercelin, 2006), this technique is still used over the Crau plain since four centuries. Several studies have been conducted over the Crau plain in order to estimate surface soil moisture and monitor the irrigated grassland plots using SAR data. Using the X-band SAR data from TerraSAR-X and COSMO-SkyMed satellites, (Hajj et al., 2014) showed that it is possible to track gravity irrigation and soil moisture variations from SAR X-band images acquired at high spatial resolution. In their study, they showed that over the irrigated grassland plots, the penetration depth of the radar wave in the X-band was high, even for dense and high vegetation (vegetation height more than 1 m); especially using the HH polarization. Moreover, (Baghdadi et al., 2016b) showed that using the C-band RADARSAT data in the HH polarization, the surface soil moisture over the irrigated grassland plots could be estimated with an accuracy of 6 vol.%. In addition, El Hajj et al. (2016a) showed that the use of X-band SAR measurements in HH polarization for soil moisture estimation over the irrigated grassland plots produces an estimation accuracy of 3.6 vol.% for NDVI values between 0.45 and 0.75, and 6.1 vol.% for NDVI between 0.75 and 0.90.

In the context of water resource management especially under changing climatic conditions, the objective of this paper is to monitor irrigation activities over irrigated grassland plots in the Crau plain of southeast France using Sentinel-1 SAR data. In this study, the irrigation detection model proposed by Bazzi et al. (2020b) has been adopted to detect irrigation events at grassland plots using all possible Sentinel-1 SAR acquisitions. Given the exact irrigation dates at 46 different grassland plots, the performance of the used model is assessed and the accuracy of irrigation event detection is reported. The capability of detecting irrigation events using S1-SAR data was evaluated as a function of the plot geometrical structure, the vegetation cover and the specific type of grass. While Section 2 presents the study site and materials used,

Section 3 describes the methodology. Section 4 presents the main obtained results followed by a discussion in Section 5. Finally, Section 6 shows the main conclusions.

2. Materials

2.1 Study Site

In this study, forty-six intensively irrigated grassland plots located in the Crau Plain of Southeast France (N 43°38′; E 5°1′) were examined (Figure 1). The Crau plain has a Mediterranean climate characterized by mild winter and dry summer season. For the past 15 years, rainfall amounts are relatively variable between years and ranges between 350 and 800 mm. Figure 2 represents the daily rainfall and temperature records in 2019 registered over the nearest local meteorological station located in "Salon-de-Provence", six km away from the study site. According to the registered precipitation data, the cumulative rainfall amounts in 2019 reaches 580 mm. The minimum-recorded temperature is 5.9 °C on 24/01/2019, while the maximum-recorded temperature is 43.5 °C, occurring on 28/06/2019. The average temperature in the summer season (between June and September) reaches 31.6 °C, with a cumulative rainfall of 37 mm only. Due to high temperatures in the summer season, the evaporation rate reaches 10 mm/day. In the Crau plain, the source of water for irrigation originates from the Alps Mountains.

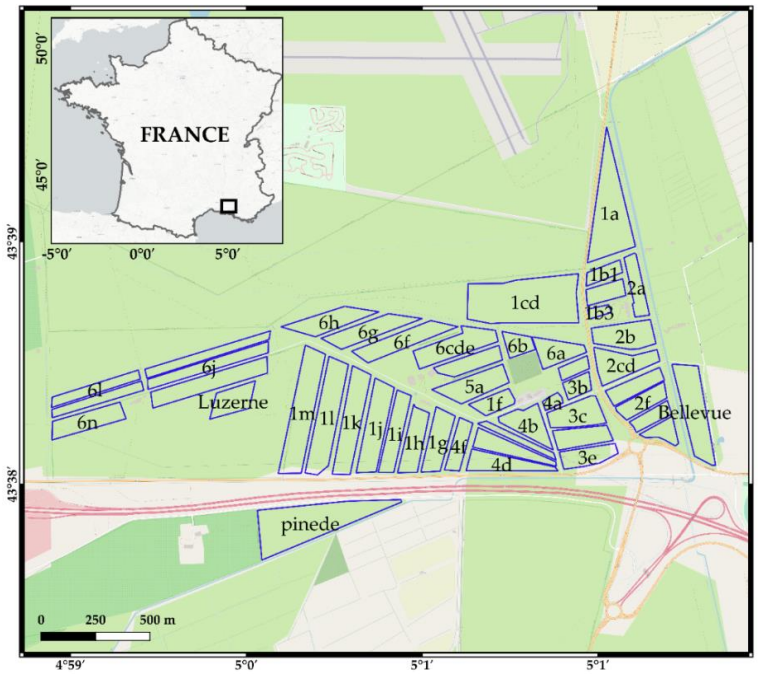


Figure 1: Location of the forty-six experimental plots in the Crau plain of Southeast France.

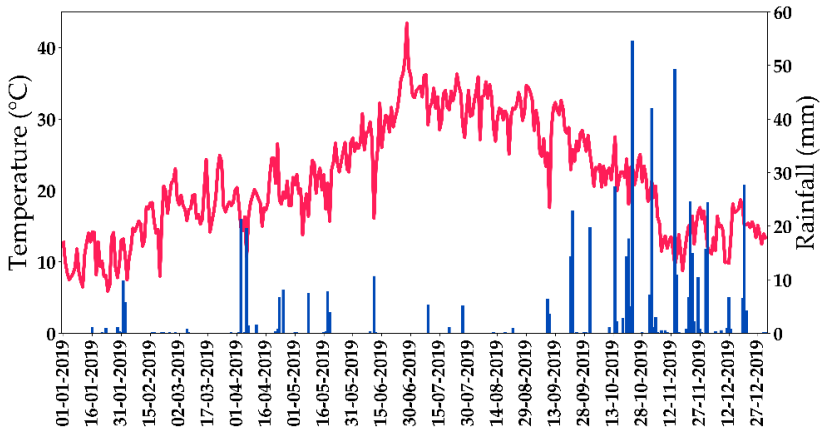


Figure 2: Temperature (in red) and rainfall (blue) profiles recorded at the local meteorological station in Salon de Provence near the study site.

2.2. Reference Data

The forty-six studied plots (Figure 1) are oriented toward hay production certified with the AOP French label (protected designation of origin), ensuring high quality of hay product (Mérot, 2007). Each plot undergoes three major cycles between February and September with three harvesting periods. The first cycle starts from the beginning of February until May with the first harvesting occurring in mid-May. The second cycle starts in May and is usually harvested in July which coincides with the beginning of the third cycle. The third cycle is finally harvested after mid-August. Different grass species are present in each plot with varying biomass proportions for each cycle. The 1st cycle is usually rich in grasses (60–65% of coarse hay). The grass percentage then decreases in the 2nd and 3rd cycles to reach 35% with the increase of the legumes proportion (Mérot, 2007). Therefore, the 1st cycle could be characterized as grass dominant cycle, whereas the second and the third cycle are legume dominant cycles (more leafy). Concerning the soil texture, loam is the dominant top soil texture among the plots, with depth varying between 30 and 80 cm. The area of the plots varies between 0.4 and 9.9 ha, with an average area of 2.8 ha.

Irrigation is performed using the gravity irrigation system where the gentle slope of the plots allows water to reach from the irrigation canals to cover the whole plot. The plots are usually irrigated between March and September of each year. The exact date of the end of each irrigation episode was registered at each plot for the period between March and September 2019. For each irrigation episode, the time between the start and the end of the irrigation usually varies between 1 and 24 h, depending on the surface of the plot (Hajj et al., 2014). Unfortunately, only the exact date of the end of irrigation was registered whereas the exact timing (in h) of the start and end of irrigation were not available. Some large surface plots are

split into two different subplots being irrigated sequentially with a time difference of less than 24 h. In general, the average frequency of the irrigation for each grassland plot reaches one irrigation each 10 days during the three vegetation cycles with approximately 14 irrigation episodes per plot and a total water quantity between 15,000 to 20,000 m3/ha. Finally, the mowing date for each cycle at each plot was also registered.

2.3. Sentinel-1 SAR Data

Over the study site, all available Sentinel-1 (S1) SAR images corresponding to the period between February 2019 and September 2019 were downloaded and processed. One hundred eighty C-band (f = 5.405 GHz) SAR images, acquired by the two polar-orbiting S1 satellites (S1A and S1B), were downloaded from the Copernicus website (https://scihub.copernicus.eu/dhus). The images are acquired in both ascending (afternoon ~18h00 UTC) and descending (morning ~06h00 UTC) modes. Each month, twenty images are available over the study site. Figure 3 shows the repeat cycle of the acquired S1 images in both ascending "A" and descending "D" modes for August 2019. The first image corresponds to the morning descending acquisition at an incidence angle of 32.2°. Twenty-four h later, another morning descending image is available at an incidence angle of 42.4°. A third image is then acquired twelve h later, corresponding to an afternoon ascending acquisition at 44.4° incidence angle. Finally, 24 h later, a new afternoon ascending image at 34.8° incidence angle is available. Thus, four S1 images are acquired within a period of 2.5 days. Since each one of the four acquisitions is repeated six days later (temporal resolution of S1 satellite), the time gap between the last afternoon acquisition and the first repeated morning acquisition is 3.5 days. Finally, four different temporal series were obtained where each temporal series represents the repeated images of the same acquisition mode each six days. The first series represents the 1st morning acquisition at incidence angle of 32.4° denoted by "MS1", while the second morning temporal series acquired at 42.4° is denoted by "MS2". The first afternoon SAR series at 44.4° is denoted by "AS1", while the second afternoon series at 34.8° is denoted by "AS2".

To perform the radiometric and geometric calibration of the downloaded GRD SAR images, the S1 toolbox developed by the European Space Agency (ESA) was used. The radiometric calibration converts the digital number (DN) into backscattering coefficient (σ^0) expressed in linear unit. The geometric calibration uses the digital elevation model at 30 m spatial resolution, offered by the Shuttle Radar Topography Mission (SRTM), to ortho-rectify the SAR images.

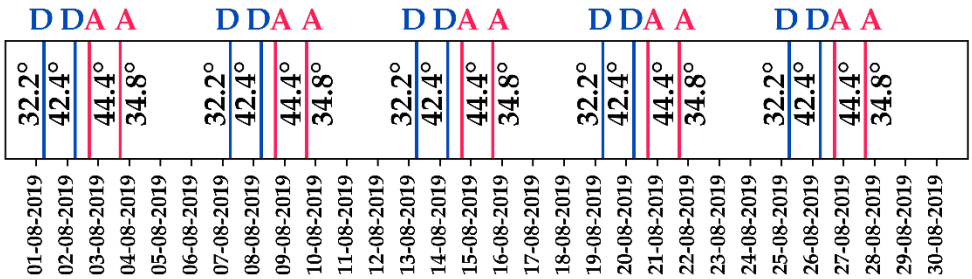


Figure 3: Repeat cycle of the Sentinel-1 images in ascending "A" (afternoon, red) and descending "D" (morning, blue) modes for the month of August 2019 over the examined plots.

2.4. Sentinel-2 Optical Data

The high temporal resolution of the Senintel-2 satellite (five days) allowed obtaining forty-six Sentinel-2 (S2) cloud free optical images acquired over the reference plots for the period between February and September 2019. The level-2 cloud free S2 images were downloaded from the Theia website (https://www.theia-land.fr/) which provides ortho-rectified level-2 products corrected for atmospheric effects. These optical images were used to calculate the NDVI values. The NDVI values are required as an input layer in the irrigation detection model used in this study (Bazzi et al., 2020b).

3. Methods

3.1. Irrigation Detection Model

The irrigation detection model (IDM) was recently developed by Bazzi et al. (2020b) to detect irrigation events at agricultural plots in a near real-time scenario. The IDM is a decision tree-based model that relies on the change detection in the SAR backscattering coefficient at plot scale (σ_P^0) accompanied by the change detection of the SAR backscattering coefficient at grid scale (σ_G^0) of 10 km × 10 km. The joint use of σ_P^0 and σ_G^0 is mainly used to remove the uncertainty between a rainfall event and an irrigation event because both events are water supplements that cause an increase in soil moisture value and therefore an increase in σ^0 values between two dates. The IDM is based on detecting the change is σ_P^0 values between the σ_P^0 value at the current SAR acquisition at time t_i and σ_P^0 value at the previous SAR acquisition at time t_{i-1}. Deep in details, the IDM assumes that the increase of the σ_P^0 between two consecutive SAR dates (t_{i-1} and t_i) is mainly caused by the increase of the surface soil moisture due to either rainfall or irrigation event. On the other hand, the increase of the σ_G^0 values (grid scale)

could be an evidence of rainfall event that occurred between the two SAR acquisitions, whereas the stability or decrease of σ_G^0 values could indicate the persistence of dry conditions between the two dates (absence of rainfall). Thus, the increase of the σ_P^0 values between consecutive acquisitions accompanied with the stability or decrease of σ_G^0 values is considered an irrigation event. Since the SAR backscattering signal could be affected by other factors, such as the vegetation cover (growth cycle for example) and surface roughness, additional filters considering the NDVI values and surface soil moisture estimations (SSM) are also added to the IDM to enhance the irrigation event detection. Therefore, the application of the IDM requires five principal data inputs:

- The SAR backscattering signal at plot (σ_P^0) and grid (σ_G^0) scales;
- Surface Soil Moisture at plot (SSM_P) and grid (SSM_G) scales;
- The NDVI value at plot scale.

3.1.1 σ° SAR Backscattering at Plot Scale

For each acquired S1 SAR image, the σ_P^0 at each of the forty-six plots was calculated by averaging the pixel values within each plot in both VV and VH polarizations. This average helps reducing the speckle filter present in the SAR images. As a result, four distinct temporal series are obtained for each plot using the four different SAR acquisitions discussed in Section 2.2 in both VV and VH polarizations. Although the study of Bazzi et al. (2020b) recommends the use of the VV polarization for better detection of irrigation events, the VH polarization was tested in order to assess the potential of VH polarization for detecting irrigation events. Moreover, the combined use of both VV and VH polarization for irrigation detection was also examined. The SAR incidence angle (θ_P) at each plot has been obtained for each SAR image.

3.1.2 σ° SAR Backscattering at Grid Scale

The σ_G^0 value at grid scale (10 km × 10km) was obtained at each SAR acquisition by averaging the SAR backscattering coefficient of all pixels corresponding to bare soil pixels or with small vegetation cover within each grid cell. Bare soil pixels with low vegetation cover has been extracted by using first a land cover map of France to delineate agricultural areas (Inglada et al., 2017) and then applying a threshold value of the NDVI obtained from the S2 images (NDVI < 0.4). The σ_G^0 has been also obtained in both VV and VH polarizations as well as the average SAR incidence angle at each grid cell (θ_G) for each SAR image.

3.1.3 NDVI Values

The NDVI was calculated for the forty-six available S2 images. Then, the average NDVI value at each plot was obtained by averaging the pixels' values within each plot. The plot NDVI value is used as an input data layer for the IDM and used for estimating the SSM value at plot scale. In addition, at each grid cell, the average NDVI value was obtained for bare soil pixels with low vegetation cover, as discussed in Section 3.1.2, and used to obtain the SSM estimation at grid cell.

3.1.4 Surface Soil Moisture Estimation

The IDM requires the SSM estimation, as one of the input parameters, at each SAR acquisition for both plot and grid scales. As recommended by Bazzi et al. (2020b), the SSM estimations are derived using the neural network (NN) developed by El Hajj et al. (2017). This NN provides satisfying SSM estimations (RMSE = 5 vol.%) and needs as inputs the σ^0 value in VV polarization, the SAR incidence angle (θ) and an NDVI value. Therefore, the SSM estimations are obtained for each plot and for each grid cell at each available SAR date in the four temporal series MS1, MS2, AS1 and AS2. However, the soil moisture estimations do not have a major role in the irrigation detection. They are mostly used to ensure the detection of irrigation events in some cases where the difference of the SAR backscattering signal at plot scale could not lead to a definitive decision on the existence of irrigation events.

3.2. Application and Assessment of the IDM

Figure 4 represents the overflow of the IDM application at a given plot using four SAR temporal series. First, the IDM was applied at each plot using inputs from each SAR temporal series (MS1, MS2, AS1 and AS2) independently. For example, at time t_i in the MS1 series, irrigation detection using the IDM has been performed using only SAR images from the MS1 time series between t_0 and t_{i-1}. While developing the IDM, Bazzi et al. (2020b) proposed a certainty indicator, which represents the chance of having an irrigation event at each SAR date. This certainty indicator depends mostly on the observed difference of σ_P^0 between two consecutive SAR dates from the same temporal series. Therefore, at each plot, each SAR image from each series is labelled with an irrigation indicator value (p) that represents irrigation chances as either no ($p = 0$), low (25), medium (50) or high (100). In the IDM, the indicator value 0 represents either a decrease in the σ_P^0 between t_{i-1} and t_i ($\Delta\sigma_P^0 \leq 0$) indicating a

decrease in soil moisture values (no irrigation possibility) or an important increase of the σ_G^0 ($\Delta\sigma_G^0 \geq 1\ dB$) between t_{i-1} and t_i indicating the presence of a rainfall event. Low irrigation chance indicator ($p = 25$) corresponds to a very slight increase in the σ_P^0 between t_{i-1} and t_i ($0 \leq \Delta\sigma_P^0 < 0.5\ db$) accompanied with high estimated soil moisture values at plot scale at t_i along with an important decrease of the σ_G^0 at grid scale between t_{i-1} and t_i ($\Delta\sigma_G^0 \leq 1\ dB$), indicating the absence of a rainfall event. In this case, the SSM estimations at plot scale presented in Section 3.1.4 are used to guarantee the existence of an irrigation event.

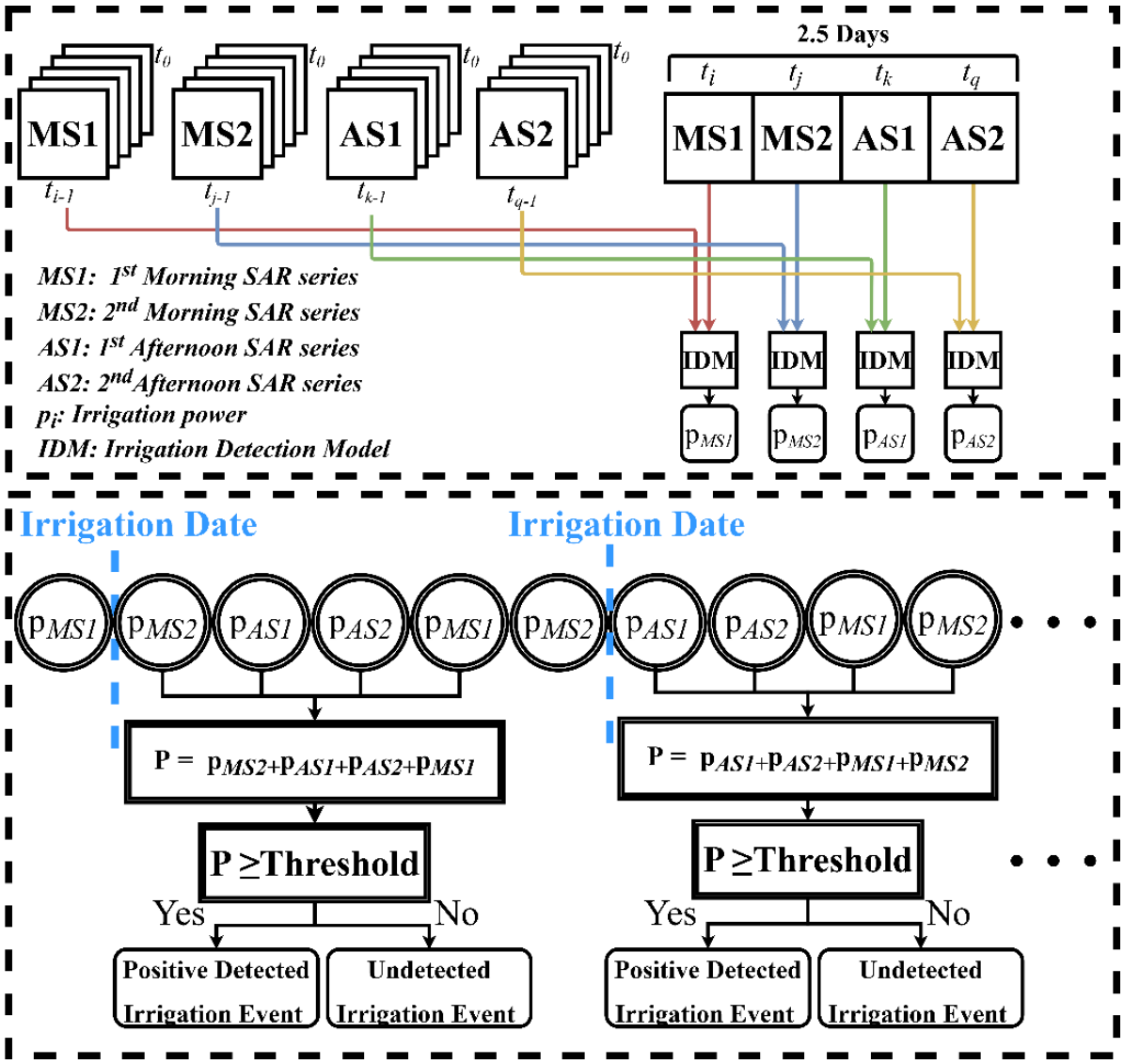


Figure 4: Overflow of the methodology for the detection of irrigation events using the IDM and four SAR temporal series. The time intervals (t_i - t_{i-1}), (t_j - t_{j-1}), (t_k - t_{k-1}), (t_q - t_{q-1}) are equal to 6 days.

The stability or slight increase of the SAR signal at plot scale between t_i and t_{i-1} could be interpreted as an irrigation event (low chance) if the σ_P^0 at time t_{i-1} already attains high values (due to irrigation or rainfall). To ensure this situation, we say that at time t_{i-1}, SSM estimation

should be greater than or equals to 20 vol.% in order to guarantee that a humid soil conditions at time t_{i-1} have continued to time t_i. The medium chance index ($p = 50$) corresponds to a moderate increase in the σ_P^0 between t_{i-1} and t_i ($0.5 \leq \Delta\sigma_P^0 < 1\ db$) accompanied with a decrease of the σ_G^0 ($\Delta\sigma_G^0 \leq 0.5\ dB$). Finally, high chance irrigation index ($p = 100$) corresponds to important increase in the σ_P^0 between t_{i-1} and t_i ($\Delta\sigma_P^0 \geq 1\ db$) accompanied with a decrease of the σ_G^0 (increase of soil moisture at plot scale at t_i with no rainfall event between t_{i-1} and t_i). Therefore, a value (p) equals to 100 represents the highest chance of existing irrigation. As the value of p decreases, the irrigation chance decreases to reach zero for no possible irrigation. As a result, the four SAR temporal series are transformed into labelled series of irrigation indicators showing the possibility of an existing irrigation event for each SAR date. It is good to mention that the irrigation indicator p is obtained at each SAR image using only the images from its corresponding SAR time series acquisition (orbit) at 6 days interval (Figure 4) where (t_i - t_{i-1}) = (t_j - t_{j-1}) = (t_k - t_{k-1}) = (t_q - t_{q-1}) = 6 days.

The second step is to compare the obtained labelled SAR dates of the four SAR series with the irrigation dates registered for each plot. To compare between in situ irrigation events and detected irrigation events from the four SAR time series, the p indicators coming from the four series were combined and compared to a threshold value, to determine the detection or not of in situ irrigation event. For each in situ irrigation date, four irrigation "p" values (one indicator from each image following this event from the four SAR series) are available. The four p values correspond to the irrigation chances detected by each SAR image that follows this irrigation event. The four obtained p values are then summed to obtain a value (P) for this irrigation event. Finally, the obtained P value is compared to a threshold in order to determine if the irrigation has been positively detected (considered as a true detection) (Figure 4). Inversely, if an irrigation episode attains a P value less than the fixed threshold then the event is considered as undetected irrigation event. In this study, the threshold values of 25, 100, 150 and 200 are tested. Finally, false detections are detected irrigation events but actually, no irrigation event is registered on the plot. This means that the P value (obtained from four SAR image encountering the false detection) is greater than the determined threshold but no real irrigation event is registered.

The IDM was applied on the four SAR series using the VV and VH polarizations separately. The assessment of the IDM for irrigation detection was then performed in four different configurations. The first and the second configurations correspond to the application of the IDM

using the VV and the VH polarizations, respectively. In the third configuration, “VV and VH”, a detected event is considered as an irrigation event of the IDM if it is detected within both VV and VH configurations. Finally, in the fourth configuration “VV or VH”, a detected event is considered as an irrigation event of the IDM if it exists within either the VV or VH the configurations. In the four configurations, the number of positively detected events, undetected events and the falsely detected events are registered.

Three accuracy metrics are calculated to assess the performance of the IDM for detecting irrigation events over grassland plots for each configuration. The first metric “S” is related to the sensitivity of the IDM for detecting irrigation events also known as the recall Equation (1). It reflects the ratio between the number of the positively detected irrigation events by the IDM and the total number of irrigation events on the plots. High S values represent the successful detection of irrigation events with low number of undetected events. The second metric “Pr” is the ratio between the positively detected irrigation events and all the detected events (true and false detections) by the IDM, also known as the precision Equation (2). High precision values mean that irrigation events are detected with low possibility of obtaining false detections (events detected as irrigation but no real irrigation events occur). To assess the performance of the model in detecting irrigation events, it is important to quantify simultaneously the number of irrigation events that could be positively detected represented by “S” and the additional false irrigations that could be detected represented by “Pr”. In order to obtain a balanced quantification between the sensitivity and the precision, the F_score value is calculated by Equation (3). The F_score is the harmonic mean between the recall (S) and the precision (Pr) that allows the comparison of the global accuracy for irrigation detection between the several thresholds and configurations used. In this study, the three accuracy metrics were calculated on all the tested threshold values of P (25, 100, 150 and 200) and for each of the four configurations.

$$S = \frac{\text{Number of Positive Detected Events}}{\text{Number of All Existing Irrigation Events}} \times 100\% \quad (1)$$

$$Pr = \frac{\text{Number of Positive Detected Events}}{\text{Number of All Detected Events}} \times 100\% \quad (2)$$

$$F\_score = \frac{2 \text{ x } S \text{ x } Pr}{S + Pr} \times 100\% \quad (3)$$

4. Results

4.1 Irrigation Detection using VV and VH separately

Table 1 summarizes the results obtained for the assessment of the irrigation detection using the IDM for the VV and the VH configurations separately. At each configuration, four threshold values of P were tested. The P value is obtained as a sum of the four p values from the four SAR images corresponding to each irrigation event. The p values are obtained in IDM for each SAR image using only its corresponding SAR time series (same orbit acquisition) at 6 days time interval. The first tested threshold value ($P \geq 25$) means that an irrigation event is detected if it exists, at least, on one SAR image from the four SAR time series with any obtained irrigation chance (25, 50 or 100). This threshold insures the positive detection of most of the existing irrigation events in both VV (82.4%) and VH (79.90 %) but with high chance of having false detections with Pr-value of 66.9% and 65.8% for VV and VH respectively. The second tested threshold corresponds to $P \geq 100$. This threshold occurs if at least one of the four SAR images shows high irrigation chance (p = 100) or two out of the four images insures a medium irrigation chance (2 x 50) or the four SAR images shows low irrigation probability simultaneously (4 x 25). Using the VV polarization, 66.5% of the events could be correctly detected with high detection precision of Pr-value reaching 81.6% (18.4% of false detection could be obtained). Using the VH polarization, 60.1% of the irrigation events could be detected with a precision of 76.8% where the chance of having false detections reaches 23.2%. For the threshold value 150, an irrigation event should exist at least with high chance on one of the four images accompanied with medium chance on another image or low chance on two other images (100+50 or 100+2 x 25). This condition ($P \geq 150$) could be also attained if at least three out of the four images shows medium irrigation chance (3 x 50) or two images with medium chance and two images with low chance (2x50+2x25). In the case of $P \geq 150$, the sensitivity value "S" decreases since less events could satisfy the threshold values but the precision increases indicating that less falsely detected irrigation events could be obtained. In fact, using the VV polarization, only 44.6% of the events could still be detected with $P \geq 150$ but with a high detection precision with Pr_value reaching 91.8% (only 8.2% of false detections could be obtained). For the VH polarization, only 29.7% of the irrigations could be detected with high precision value Pr reaching 86.6%. Finally, the threshold value $P \geq 200$ means that an irrigation could be detected if it exists at high chance on one image accompanied with either a high chance on another image (100+100) or two medium chances on two images (100+2 x 50).

It could be also satisfied if the four images shows medium irrigation chances (4x50). For very high tested P value ($P \geq 200$), the sensitivity drops for S=32.3% using the VV polarization and 18% using the VH polarization. This drop in the sensitivity value is expected since higher thresholds constraints the detection of irrigation and obliges an event to be present with high and medium chances on several images among the four SAR images.

Table 1 shows that low threshold values (i.e. $P \geq 25$) allows detecting most of the irrigation events (S equals to 79.9% and 82.4% for VH and VV respectively). However, this high detection is accompanied with moderate possibility of obtaining additional false detections where the Pr values reaches 66.9% and 65.8% for VV and VH respectively. On the other hand, high threshold values (i.e. $P \geq 200$) insures low chance of obtaining any false irrigation event with higher precision Pr reaching 93.7% in VV polarization. However, in this case only one-third of the irrigation events could be detected (32.3%).

Table 1: Accuracy metrics obtained for the assessment of the IDM for irrigation detection using VV and VH polarizations separately

Configuration	Metric	Thresholds			
		P ≥ 25	**P ≥ 100**	**P ≥ 150**	**P ≥ 200**
VV	S (%)	82.4	66.5	44.6	32.3
	Pr (%)	66.9	81.6	91.8	93.7
	F_score (%)	73.8	73.3	60.0	48.0
VH	S (%)	79.9	60.10	29.7	18.0
	Pr (%)	65.8	76.8	86.6	85.44
	F_score (%)	72.2	67.4	44.2	29.7

Figure 5 represents the average sensitivity (S) (Figure 5a), precision (Pr) (Figure 5b) and F_score (Figure 5c) at each month in VV and VH polarizations for a P threshold value equals to 25. This threshold value was used since it represents the highest F_score value compared to the other tested thresholds in both VV and VH polarizations. Figure 5a shows that in both VV and VH polarizations, the lowest percentage of detected irrigations occurs during the first cycle (Cycle I) in March, April and May. Using VV, the S-value reached 69.0%, 60.0% and 60.5% for March, April and May, respectively. For VH polarization, a sensitivity of 64.2%, 58.4% and 57.1% for March, April and May is obtained, respectively. For Cycle I, the precision value Pr (Figure 5b) and F_score (Fugure 5c) show low values for both VV and VH, indicating low precisions for irrigation detection with a high possibility of obtaining false detections. During the second cycle (Cycle II) and the third cycle (Cycle III), the percentage of truly detected irrigation events (S) increases for both VV and VH. In these two cycles the S value ranges

between 91.5% (in June) and 96.2% (in August) for the VV polarization. For the VH polarization, the S-value ranges between 83.9% (in June) and 96.2% (in August). The precision value Pr also increases during Cycle II and Cycle III to reach its maximum value of 74.5% for the VV polarization in July and 78.8% for the VH polarization in August. Similarly, Figure 5c shows that the F_score increases for Cycle II and Cycle III to reach its maximum value of 84.5% in July for VV and 86.6% in August for VH.

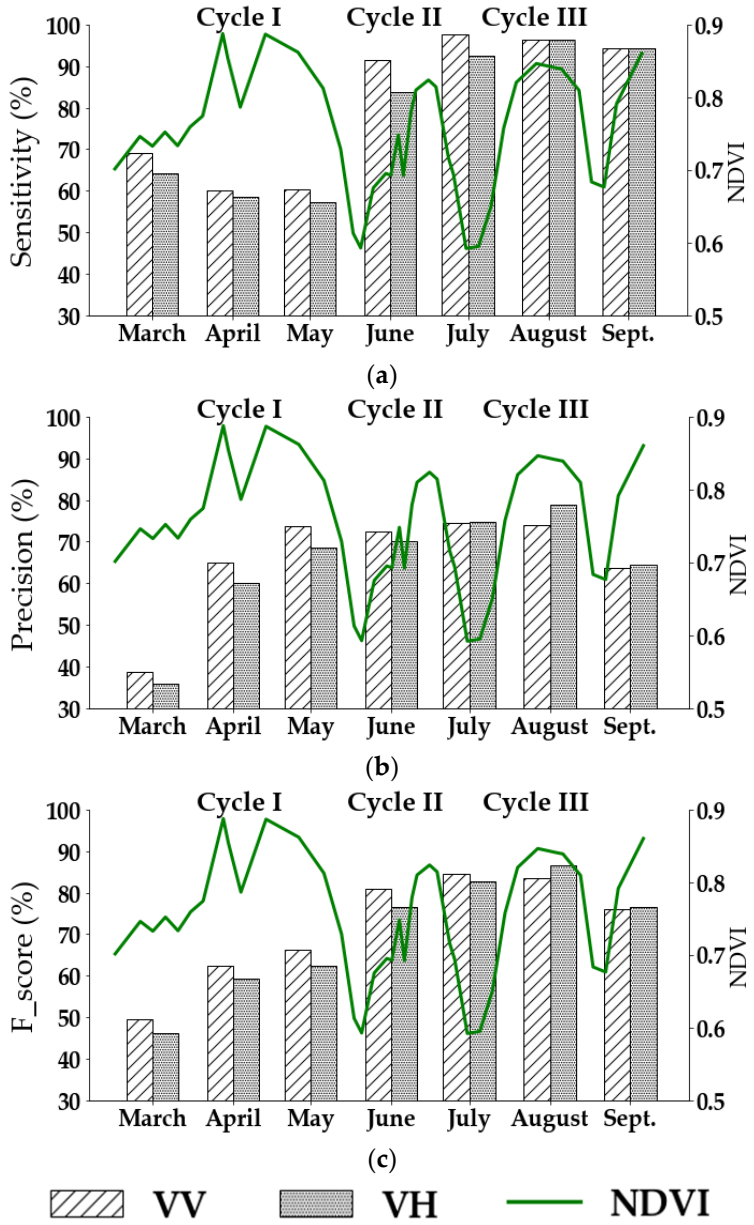


Figure 5: The accuracy of the detection of irrigation events in each month using the IDM for VV and VH over all the plots during the three vegetation cycles. (a) sensitivity "S", (b) precision "Pr" and (c) F_score. The green line represents the average normalized differential vegetation index (NDVI) values derived from S2 images for all the studied plots at the three distinct grass cycles

4.2. Irrigation Detection Using Combined VV and VH

In this section, we report the combined use of both VV and VH polarizations for irrigation detection. As mentioned in Section 3.2, the results of irrigation detection using VV and VH were joined in two different configurations. The first configuration "VV or VH" implies that a detected event is considered an irrigation event if it is detected within either one of the VV or the VH polarizations or both. The second configuration "VV and VH" means that a detected event is registered as a detected irrigation event if it is present within both polarizations simultaneously. The obtained accuracy metrics are reported in Table 2.

For low threshold value ($P \geq 25$), 88.2% of the total irrigation events (S) are detected in the "VV or VH" configuration with a moderate opportunity to obtain false detections with a precision (Pr) of 58.5% only. For the same value of *P*, the intersection configuration "VV and VH" shows less detection of irrigation events than "VV or VH" (S=74.1%) but with higher precision (Pr = 78.9%) and, thus, a lower possibility of false detections compared to that obtained with "VV or VH" configuration. For $P \geq 100$, half of the irrigation events could be detected using the "VV and VH" configuration with high precision of 91.8% (only 8.2% of possible false detections). However, using the "VV or VH" configuration, 76.8% of the irrigation events are detected with a precision value reaching 72.8% indicating a 27.2% chance of having false detections. For higher threshold value ($P \geq 150$), half of the irrigation events (51.6%) could be detected using the "VV or VH" scenario with a precision of 87.4%. However, for the intersection configuration "VV and VH" the detection of irrigation events becomes difficult for $P \geq 150$, as the S value does not exceed 22.7% but with very high precision of 95.2%. Finally, with the threshold of 200, both configurations detect not more than 38% for "VV or VH" and 12.1% for "VV and VH" despite the high precision of 89.5% and 94.1%, respectively.

Table 2: Accuracy metrics obtained for the assessment of the IDM for irrigation detection by the combined use of VV and VH in two different configurations.

Configuration	Metric	Thresholds			
		P ≥ 25	**P ≥ 100**	**P ≥ 150**	**P ≥ 200**
VV	S (%)	88.2	76.8	51.6	38.0
	Pr (%)	58.5	72.8	87.4	89.5
	F_score (%)	70.3	74.7	64.9	53.3
VH	S (%)	74.1	50.0	22.7	12.1
	Pr (%)	78.9	91.8	95.2	94.1
	F_score (%)	76.4	64.7	36.7	21.4

5. Discussion

In this section, the obtained results are analyzed as a function of group of factors that could limit the irrigation detection at plot scale. Among these factors, we consider the geometrical structure of the plot, the vegetation cover, the precipitation records and harvesting time of the plots. In order to discuss the obtained results, we chose the configuration attaining the highest F_score as the representative case for irrigation detection. The highest F_score has been chosen, since it represents the significant balance between the recall and the precision for the detection of irrigation events among the several tested thresholds and configurations. For this reason, the following discussion is based on the results obtained by the configuration "VV and VH" having a threshold value $P \geq 25$, which has the highest F_score of 76.4%, being a representative case for irrigation detection.

5.1 Effect of the Geometrical Structure of the Plots

The average σ_P^0 calculated for all pixels within the plot could be affected by the geometrical shape of the plot. For example, narrow plots with very small width could have border pixels attaining the backscattering values from the plot boundaries usually covered by trees, concrete or agricultural roads. Small plots as a function of width or length increases the chance of encountering speckle noise in the SAR backscattering coefficients, thus, causing undesirable fluctuation in the SAR signal (increase of σ_P^0 value). Therefore, one of the factors that could increase the possibility of encountering false detections is the geometrical structure of the plot such as the length, width and the surface area knowing that the Sentinel-1 images have pixel spacing of 10 m × 10 m. For this reason, we analyzed the false detections (events detected as irrigation where no registered irrigation exists) as a function of the length (l), width (w) and the surface area of the plots (S). For each plot the length, width and surface area have been calculated based on the RGF-93 projection system.

The grassland plots are generally elongated having a length significantly larger than the width (Figure 1). The length and width have been grouped into three classes to define small, average and large dimensional plots. The first class includes plots having a width less than 50 m or a length less than 150 m (10 plots). Independent of the first class, the second class contains averaged size plots having a width between 50 m and 100 m or a length between 150 m and 300 m (24 plots). Finally, the third class represents the large plots with a width more than 100 m and a length more than 300 m (12 plots). For each class, the average number of false detection

is obtained by averaging the number of the false detections registered over all the plots in the class. Similarly, the surface area of the plots has been classified into three classes. The first class represents the plots with surface area less than 2 ha (18 plots), the second class contains plots with a surface area between 2 and 3 ha (17 plots), while the third class having the plots with a surface area more than 3 ha (11 plots). The average number of false detection is also obtained by averaging the number of the false detections registered over all the plots in the class.

Figure 6a represents the average false detections in each class as a function of the length-width grouping for the three classes and Figure 6b represents the average false detections as a function of the three surface area classes for "VV and VH" configuration with $P \geq 25$. Figure 6a shows that when the width (w) of the plot is narrow (less than 50m) or the length (l) is short (less than 150 m) the average false detections reaches five false detections per plot. The average false detections decreases to three false detections when the length (l) or the width (w) attains moderate sizes ($50 < w(m) < 100$ and $150 < l(m) < 300$). Finally, big plots that have wide width (greater than 150 m) and long length (greater than 300 m) encounters the least false detections having an average of one false detection per plot.

Similar results are shown by Figure 6b using the surface area. Small surface plots having an area less than 2 ha have the highest average false detections (an average of four false detections per plot). As the surface area increases, the average false detections decrease to reach an average of two false detections per plot for medium surface area plots (between 2 and 3 ha) and an average of one false detection per plot for large surface area plots (greater than 3 ha).

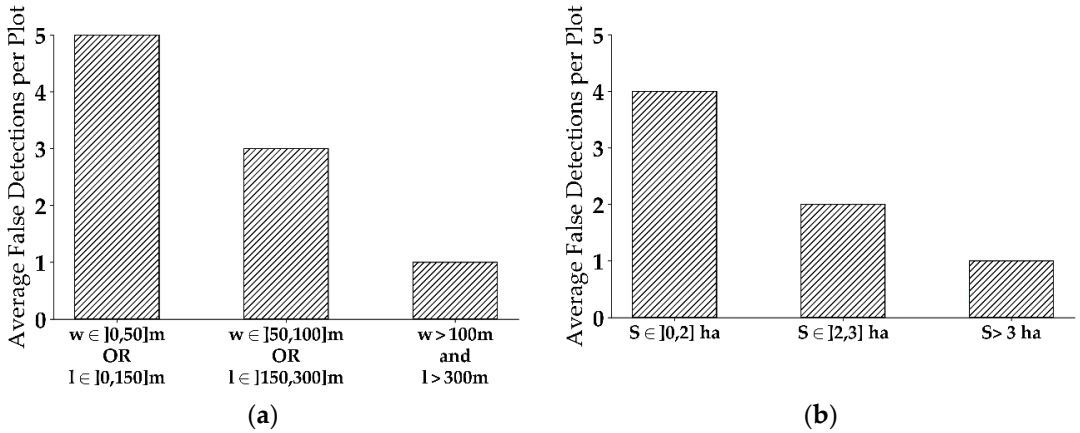


Figure 6: Average false detections per plot for each class as a function of (a) combined length-width and (b) surface area for "VV and VH" configuration with $P \geq 25$.

5.2 Effect of the Vegetation Cover

The detection of irrigation events could be limited by the existing vegetation cover, depending on the crop type or the vegetation density. In fact, this limitation is mainly related to the ability of the SAR signal to penetrate the vegetation cover in order to reach the surface soil. El Hajj et al. (2018b) analyzed the penetration of the C-band SAR signal over maize, wheat and grassland plots. In their study, they showed that over grassland plots, the penetration of the C-band SAR signal is limited to an NDVI value of 0.7. They also showed that over highly developed grassland cover, the C-band signal is insensitive to the SSM. For this reason, very well-developed vegetation cover could constrain the detection of the irrigation events, due to the lack of the SAR sensitivity to surface soil moisture. Figure 7 represents the distribution of the NDVI values for the non-detected irrigation events over the forty-six plots. The figure shows that more than 80% of the undetected irrigation events correspond to very high NDVI values exceeding 0.8, and 90% of the undetected events correspond to NDVI greater than 0.7. These very high NDVI values indicate the presence of highly developed vegetation cover. These results are in line with the findings of El Hajj et al. (2018b) for grassland plots and Le Page et al. (2020) for maize plots where both studies report the limitation of irrigation detection over dense vegetation cover (NDVI > 0.7) due to the limited penetration of C-band SAR signal.

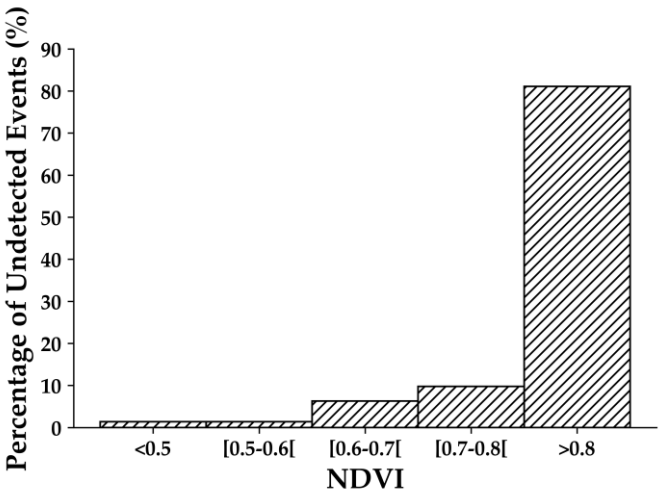


Figure 7: Distribution of the NDVI values for the non-detected irrigation events over all the plots among the three cycles for "VV and VH" configuration with $P \geq 25$.

As shown in Figure 5 in Section 4.1, the period with the least irrigation detection and highest false irrigation detection (in both VV and VH) corresponds to the first vegetation cycle from March to May, attaining the lowest sensitivity values compared to other months. In fact, independent of the NDVI values, the existing crop type could also affect the penetration of the

SAR signal. As presented in Section 2.2, different grass types exist within the three cycles with different biomass proportion of each type. Actually, the grass type for the first cycle is more oriented toward coarse hay, which represents more than 65% of biomass in the existing vegetation cover of the first cycle. The vegetation cover in the first cycle is rich in grasses with less leaves and more stems (Mérot, 2007). With the stem elongation in March and April, the C-band SAR backscattering decreases due to the attenuation of the direct ground scattering. In fact, the vertical stems and leaves highly absorbs the incident SAR wave causing weak direct scattering of the SAR signal and thus a decrease in the C-band backscattering (Cookmartin et al., 2000; Del Frate et al., 2004; El Hajj et al., 2018b; Mattia et al., 2003). This attenuation behavior is also similar to that present on the wheat crop between the germination and the heading phase, as shown by Nasrallah et al. (2019). Therefore, for the period with very well developed vegetation cover for the first cycle, (NDVI > 0.8) the backscattering from the canopy (with very low soil contribution) is the dominant mechanism, and the sensitivity of the SAR signal to the surface soil moisture is negligible. This high canopy attenuation of the soil contribution found in the first cycle is scarce in the second and the third cycles, since the grasses is less dominant, and the vegetation cover is leafier, containing more legumes than stems. In this case, the penetration of the C-band SAR signal is stronger and thus the soil contribution is more important. For this reason, higher accuracy in irrigation detection using the C-band SAR data is found for the second and third cycle than the first cycle (Figure 5).

To illustrate the vegetation attenuation mechanism that leads to the missing detection of irrigation events, Figure 8 presents a comparison between the backscattering SAR signal in VV polarization for the undetected irrigation events and that of the detected irrigation events for the four SAR temporal series MS1 (Figure 8a), MS2 (Figure 8b), AS1 (Figure 8c) and AS2 (Figure 8d).

The boxplots in Figure 8 shows that within all the SAR acquisitions, the average σ^0 value of the undetected events is at least 2 dB less than the average σ^0 value for the detected events. For example, the median of the σ^0 for undetected events in MS1 at incidence angle of 32° (Figure 8a) reaches -13.1 dB, while that of the detected events reaches -10.8 dB, with a difference of 2.2 dB. Figure 8a also shows that approximately 75% of undetected events attains a σ^0 value less than the minimum σ^0 value of any detected irrigation event. Similar behavior is found for other SAR acquisitions at different incidence angles (Figure 8b–d). The low σ^0 values for the undetected events validates the existence of the high canopy attenuation mechanism that weakens the backscattered SAR signal (low soil penetration) and, thus,

decreases the σ^0 values, indicating the absence of soil contribution in the SAR backscattered signal.

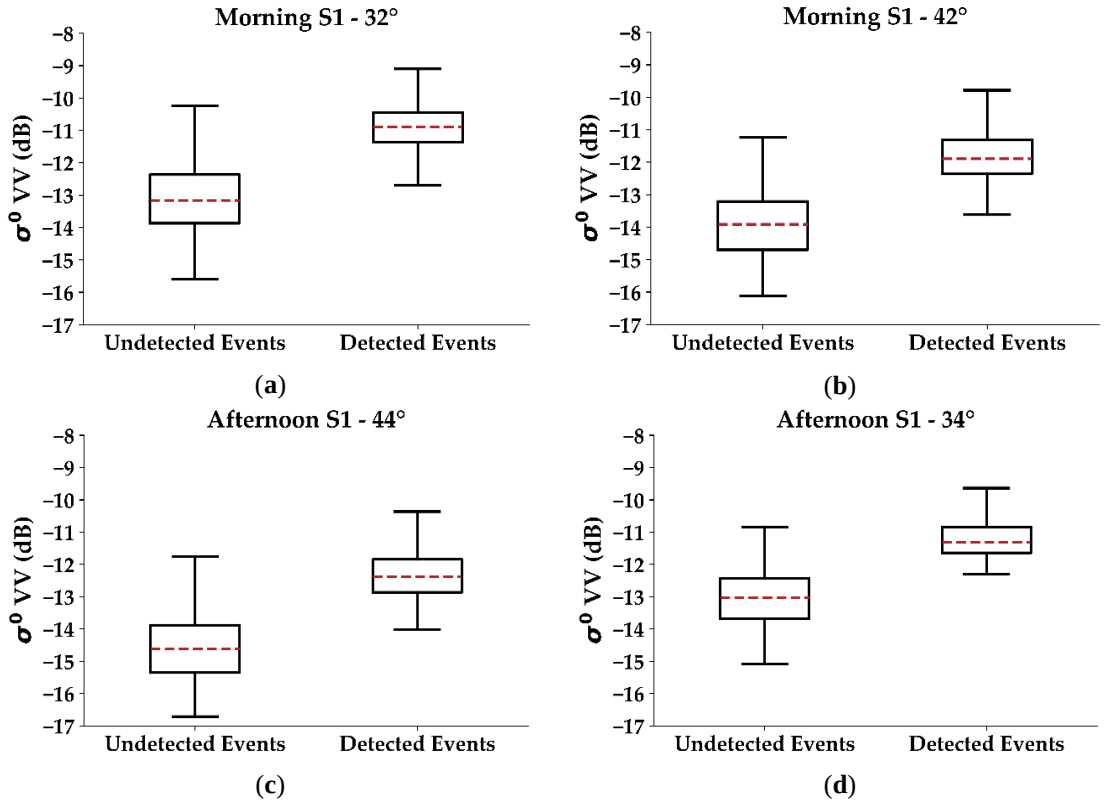


Figure 8: Boxplots of the distribution of σ^0-values in VV polarization for the detected and undetected events in the four temporal series. (a) MS1-32°, (b) MS2-42°, (c) AS1-44°, (d) AS2-34°.

5.3 Effect of precipitation and harvesting

In the Crau plain, the grassland plots are usually harvested three times per year corresponding to the three existing cycles between March and September. The harvesting mechanism usually increases the surface roughness of the plot (Figure 9). In fact, the backscattered radar signal strongly depends on the surface roughness that is usually expressed by the Height Root Mean Square (Hrms) (Baghdadi et al., 2018a). Baghdadi et al. (2008a) have reported that a difference of 4 dB could be observed between backscattering signal from rough surfaces (Hrms = 3 cm) and smooth surface (Hrms = 0.5 cm). When the surface roughness increases between two consecutive SAR images due to harvesting, the backscattering SAR signal will increase between these two SAR dates and thus a false irrigation event could be detected. In this study, the harvesting date is available for each plot at each cycle. The comparison between the false detected irrigation events and the harvesting dates shows that 28% of the false detections were preceded by a harvesting activity. Therefore, a priori knowledge about the exact harvesting dates could help reduce the false detections.

Figure 9: Harvesting of a grassland plot with an increase in surface roughness.

Rainfall events can also limit the capability of detecting irrigation events. Certainly, when both a rainfall event and an irrigation event occurs between the same two consecutive SAR image, the irrigation detection would not be possible since the increase in the SSM values could not be distinguished between the rainfall event and the irrigation. Over the study site, five important rainfall events occurred on 03, 09, and 27 April and on 09 and 21 May. These rainfall events constrained the detection of 50 irrigation events over the forty-six plots representing 20% of the total undetected events. The existence of rainfall events contributes also to the low sensitivity value of irrigation detection in March, April and May shown in Figure 5a.

5.4. Sensitivity Analysis of the IDM $\Delta\sigma_P^0$ Threshold Values

In order to assess the effect of the threshold values fixed in the IDM for the detection of irrigation events, a sensitivity analysis has been performed on the threshold values of the main contributor in the IDM ($\Delta\sigma_P^0$) which is the difference between the SAR backscattered signal at plot scale between time t_i and time t_{i-1}. The threshold of $\Delta\sigma_P^0$ has been modified in order to analyze their impact on the results of the three indicators: sensitivity (S), precision (Pr) and the F_score. In this sensitivity study, we used the four following configurations (1) VV with P = 25, (2) VH with P = 25, (3) "VV and VH" with P = 25", and (4) "VV or VH" with P = 100, which correspond to the best configurations identified in this study for the detection of irrigation events.

Two tests were carried out in the sensitivity analysis. The first test consists of adding 0.25 dB to the different threshold values considered for $\Delta\sigma_P^0$ in the irrigation events detection. This

means that for low irrigation chances (p = 25), a new threshold of $\Delta\sigma_P^0$ between 0.25 dB and 0.75 dB are used instead of the initial threshold values between 0 and 0.5 dB. For the medium irrigation chances (p = 50), the threshold values of $\Delta\sigma_P^0$ between 0.75 and 1.25 dB are newly used instead of the initial thresholds between 0.5 dB and 1 dB. Finally, for high irrigation chances (p = 100), a threshold value of $\Delta\sigma_P^0$ greater than 1.25 dB is used instead of the initial 1 dB value. The second test consists of adding 0.5 dB to the initial threshold values of the $\Delta\sigma_P^0$ making, thus, low irrigation chances (p = 25) with a threshold of $0.5 \leq \Delta\sigma_P^0 < 1$ dB, the medium irrigation chances (p = 50) with a threshold value $1 \leq \Delta\sigma_P^0 < 1.5$ dB, and the high irrigation chances (p = 100) with threshold values $\Delta\sigma_P^0 \geq 1.5$ dB.

Figure 10 shows the variation of the three accuracy metrics (S, Pr and F_score) as a function of the two-tested increase in the threshold values in the four configurations. The results show that, by increasing the threshold values, the sensitivity decreases (less event irrigation is detected) while the precision increases (less false detections could be obtained). Similarly, the F-score also decreases. For the case of VV (P = 25), the sensitivity decreases from 82.4% to 72.8% by adding 0.25 dB to the threshold values of $\Delta\sigma_P^0$ and to 69.9% by adding 0.5 dB to the thresholds. Thus, nearly 12% of the event irrigations initially detected with our starting thresholds will no longer be detected by increasing the values of the starting thresholds by 0.25 dB or 0.5 dB. For the same case, the precision increases from 66.9% with our starting thresholds to 74.3% by increasing the thresholds by 0.25 dB and to 76.4% by increasing the thresholds by 0.5 dB. Thus, the precision on the detection of irrigation events increases by approximately 10% by increasing the values of the starting thresholds by 0.25 dB or 0.5 dB. The effect of the increasing of the threshold values of $\Delta\sigma_P^0$ is clearly a decrease in the sensitivity when the precision increases and vice versa. What is also observed is that the sensitivity decreases more strongly than precision when the threshold value is increased. For example, for the configuration using VV and VH with P = 25, S decreases by 17.9% when the thresholds are increased by 0.5 dB, and Pr increases only by 8.3% when the thresholds are increased by 0.5 dB.

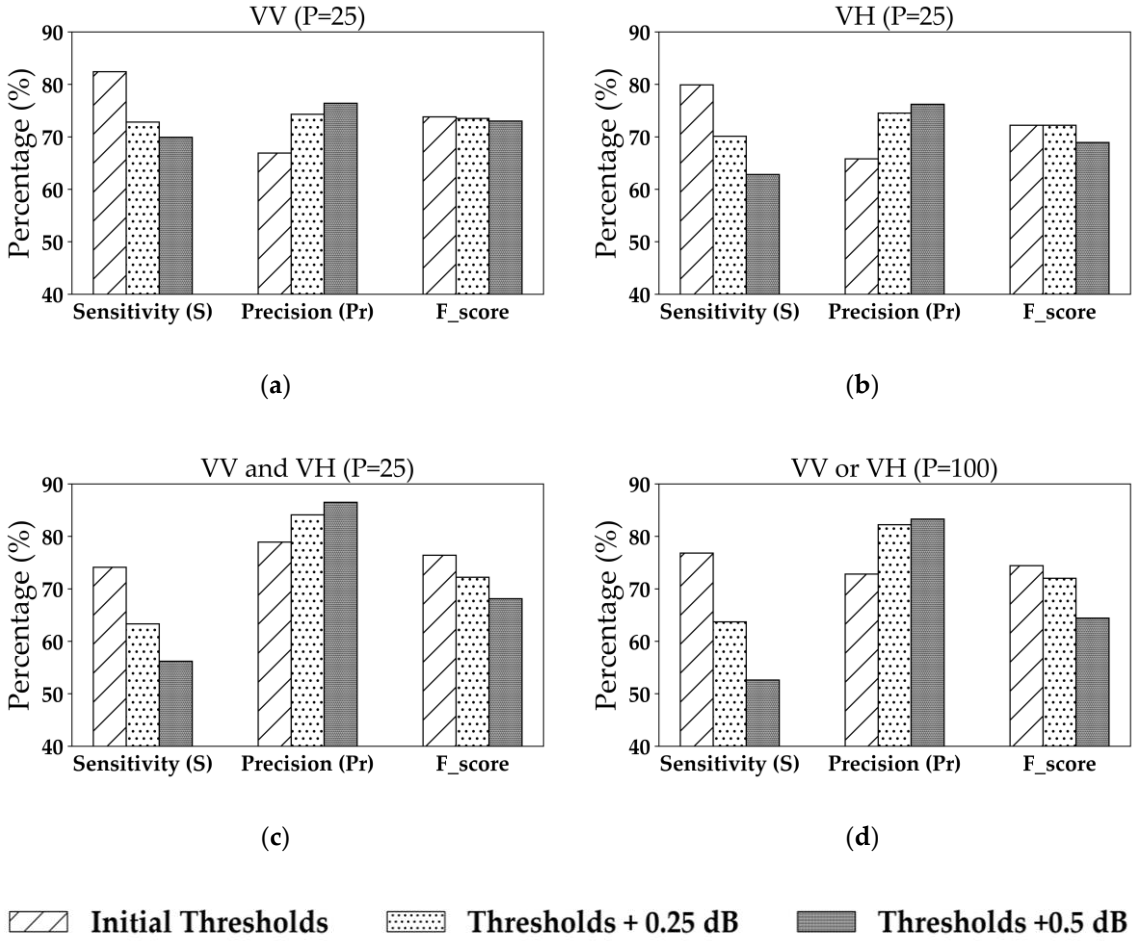


Figure 10: Variation of the accuracy metrics as a function of varying threshold values of $\Delta\sigma_P^0$ for (a) VV (P = 25), (b) VH (P = 25), (c) 'VV and VH' (P = 25) and (d) 'VV or VH' (P = 100).

5.5. Effect of the S1 Revisit Time on the Irrigation Detection

To study the effect of the S1 revisit time for the detection of irrigation events, a sensitivity analysis of the irrigation events detection as a function of the time lapse between the acquisition date of the S1 image and the irrigation date was performed. First, for each in situ irrigation event, we obtained the difference in days ($\Delta t = t_{S1} - t_{irrigation}$) between the acquisition date of S1 images used for the detection in each of the four temporal series (MS1, MS2, AS1 and AS2) and the date of the irrigation event. Since the revisit time of the S1 is 6 days, the values of Δt for an irrigation event varies between 0 days (irrigation occurred the same date as the S1 passage) and 5 days (the SAR image was acquired 5 days after the irrigation event). At each value of Δt (0, 1, 2, 3, 4, 5) we obtained the total number of irrigation events and then we calculate the percentage of the detected events and the percentage of the undetected events by the IDM. Since the vegetation cover can also play an important role in the detection of irrigation

events, we analyzed the percentage of detected and undetected irrigation events at each Δt value for two classes of NDVI values: (1) NDVI ≤ 0.7 and (2) NDVI > 0.7.

Figure 11 shows the percentage of detected and undetected irrigation events for each value of Δt between 0 and 5 days for NDVI ≤ 0.7 (Figure 11a) and NDVI > 0.7 (Figure 11b). For NDVI ≤ 0.7, the results show that if the S1 image is acquired at the same day of the irrigation event or one day after the irrigation event, we are able to detect approximately 75% of the irrigation events (Figure 11a). Two and three days after the irrigation event, the percentage of the detected and the undetected events are approximately the same (50%). Beyond 4 days of the irrigation event ($4 \leq \Delta t \leq 5$ days), the percentage of the detected irrigation events decreases. However, the percentage of undetected events increases and becomes greater than that of the detected events (60% and 40%, respectively).

For NDVI > 0.7 (Figure 11b), 70% of the irrigation events could still be detected when the S1 image is acquired at the same date of the irrigation event ($\Delta t = 0$), whereas 30% of the events are not detected. These results are similar to that obtained for NDVI ≤ 0.7 and $\Delta t = 0$. However, one day after the irrigation event, the percentage of undetected events (60%) becomes greater than that of the detected events (40%). In the case where NDVI > 0.7, the percentages of detected and undetected events remain approximately stable for the values of Δt greater than 1 day (Δt equals to 2, 3, 4 or 5 days), with a percentage of undetected events greater than that of the detected events (60–70% and 30–40%, respectively).

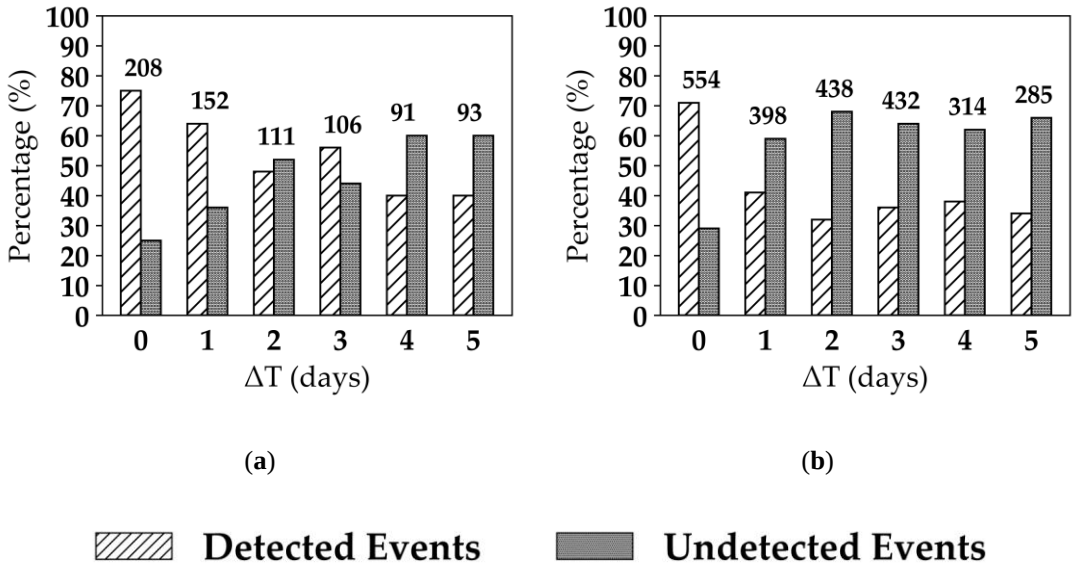


Figure 11: Percentage of detected and undetected irrigation events as a function of time laps between the S1 acquisition date and the irrigation date (Δt) for (a) NDVI ≤ 0.7 and (b) NDVI > 0.7. Numbers over the bars present the number of irrigation events at each value of Δt.

6. Conclusions

In this study, the potential of the Sentinel-1 C-band time series for detecting irrigation events has been evaluated over intensively irrigated grassland plots located in the Crau plain of southeast France. The irrigation event detection has been carried out using the newly developed near real-time irrigation detection model (IDM). Four SAR time series from four different S1 orbits were used in this study. The IDM was first applied on each SAR time series of the same orbit acquisition (at a 6 day interval) to obtain irrigation indicators (low, medium and high irrigation chance) at each SAR image. Then, the irrigation indicators obtained at each image of the four SAR time series were combined and compared to a threshold value to obtain detected irrigation events.

The application of the IDM over all possible S1 SAR images in ascending and descending modes proved the capability of detecting irrigation events with a good accuracy. The results of the comparison between the in situ registered irrigation events and the detected irrigation events by the IDM showed that using the VV polarization the accuracy for irrigation detection reaches 73.8% by the means of F_score. The use of the VH polarization seems to be slightly less accurate than the VV polarization, with an F_score reaching 72.2%. However, the combined used of the VV and VH polarization enhances the irrigation detection accuracy to obtain an F_score of 76.4%. The analysis of the threshold value used for combining the irrigation indicators from all possible SAR acquisitions showed that the use of low threshold value allows higher detection of existing irrigation events with a sensitivity (S) of 74.1%, but with slightly high possibility of obtaining false detections where the precision (Pr) reaches 78.9% (the case of “VV and VH”). On the other hand, using high threshold values for combining the irrigation indicators from the four SAR time series, less irrigation events could be detected (S = 50%), but with higher precision and low chance of obtaining false irrigation detection with a high precision value of 91.8% (the case of “VV and VH”).

The analysis of the undetected irrigation events as a function of vegetation cover and NDVI values showed that over leafy grass type (coarse hay) with very high NDVI values (NDVI > 0.8) the irrigation detection becomes difficult, due to the very low sensitivity of the C-band SAR signal to the surface soil moisture. The analysis of the falsely detected irrigation events, as a function of the plot size, showed that small sized plots in terms of narrow width or short length, (surface area less than 2 ha) encounter more false detections than large sized plots (surface area more than 3 ha), mainly due to the 10 m × 10 m pixel spacing of the S1 satellite.

The analysis of the irrigation detection as a function of the time laps between the S1 acquisition and the irrigation date showed that for NDVI values less than 0.7, the irrigation event could be detected until two to three days after the irrigation event. However, for NDVI values greater than 0.7 the irrigation could only be detected if it exists within the same day of the S1 acquisition.

The obtained results present a quantitative evaluation of the capability of S1 data to detect irrigation events, which opens the way towards the operational use of the S1 data with the IDM to detect irrigation events over regional scales. The operational detection of irrigation events is ensured with the free and open access availability of the S1 data and the direct use of the IDM that does not require priori calibration (training) phase.

Author Contributions: Conceptualization, H.B. (Hassan Bazzi) and N.B.; Data curation, H.B. (Hassan Bazzi) and F.C.; Formal analysis, H.B. (Hassan Bazzi), N.B. and I.F.; Investigation, H.B. (Hassan Bazzi), N.B. and F.C.; Methodology, H.B. (Hassan Bazzi), N.B. and I.F.; Software, H.B. (Hassan Bazzi); Validation, H.B. (Hassan Bazzi), N.B., I.F., F.C., M.Z. and H.B. (Hatem Belhouchette); Visualization, H.B. (Hassan Bazzi); Writing—original draft, H.B. (Hassan Bazzi); Writing—review & editing, N.B., I.F., F.C., M.Z. and H.B. (Hatem Belhouchette). All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2020 project), the National Research Institute for Agriculture, Food and the Environment (INRAE), the Occitanie region of France and the Mediterranean Agronomic Institute of Montpellier (CIHEAM-IAMM).

Acknowledgments: Authors wish to thank the French Space Study Center (CNES, TOSCA 2020), the National Research Institute for Agriculture, Food and Environment (INRAE) and the Mediterranean Agronomic Institute of Montpellier (CIHEAM, IAMM). The authors wish to thank the Occitanie region, France for supporting this work. The authors wish also to thank the European Space Agency (ESA) for the Sentinel 1 and Sentinel 2 data and Theia pole for the calibration of the Sentinel-2 images. The authors finally wish to thank Antoine Naudet from the Crau Hay Committee.

Conflicts of Interest: The authors declare no conflict of interest.

4. Article 7: An Operational Framework for Mapping Irrigated Areas at Plot Scale using Sentinel-1 and Sentinel-2 Data

Author version of the article published in Remote Sensing Journal: Volume: 13, Issue: 13

July 2021

Hassan Bazzi[1], Nicolas Baghdadi[1], Ghaith Amin[1], Ibrahim Fayad[1], Valérie Demarez[2], Mehrez Zribi[2], Hatem Belhouchette[3]

[1] INRAE, UMR TETIS, University of Montpellier, 500 rue François Breton, Montpellier CEDEX 5 34093, France; nicolas.baghdadi@teledetection.fr (N.B.), ibrahim.fayad@inrae.fr (I.F.)

[2] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, Toulouse CEDEX 9 31401, France; mehrez.zribi@ird.fr (M.Z.), valerie.demarez@univ-tlse3.fr (V.D.)

[3] CIHEAM-IAMM, UMR-System, 34090 Montpellier, France

*****Correspondence: hassan.bazzi@inrae.fr; Tel.: (33)-4-6704-6300

Submitted: 11 May 2021, Revised: 29 June 2021, Accepted: 30 June 2021

Abstract

In this study, we present an operational methodology for mapping irrigated areas at plot scale, which overcomes the limitation of terrain data availability, using Sentinel-1 (S1) C-band SAR (synthetic-aperture radar) and Sentinel-2 (S2) optical time series. The method was performed over a study site located near Orléans city of north-central France for four years (2017 until 2020). First, training data of irrigated and non-irrigated plots were selected using predefined selection criteria to obtain sufficient samples of irrigated and non-irrigated plots each year. The training data selection criteria is based on two irrigation metrics; the first one is a SAR-based metric derived from the S1 time series and the second is an optical-based metric derived from the NDVI (normalized difference vegetation index) time series of the S2 data. Using the newly developed irrigation event detection model (IEDM) applied for all S1 time series in VV (Vertical-Vertical) and VH (Vertical-Horizontal) polarizations, an irrigation weight metric was calculated for each plot. Using the NDVI time series, the maximum NDVI value achieved in the crop cycle was considered as a second selection metric. By fixing threshold values for both metrics, a dataset of irrigated and non-irrigated samples was constructed each year. Later, a random forest classifier (RF) was built for each year in order to map the summer agricultural plots into irrigated/non-irrigated. The irrigation classification

model uses the S1 and NDVI time series calculated over the selected training plots. Finally, the proposed irrigation classifier was validated using real in situ data collected each year. The results show that, using the proposed classification procedure, the overall accuracy for the irrigation classification reaches 84.3%, 93.0%, 81.8% and 72.8% for the years 2020, 2019, 2018 and 2017 respectively. The comparison between our proposed classification approach and the RF classifier built directly from in situ data showed that our approach reaches an accuracy nearly similar to that obtained using in situ RF classifiers with a difference in overall accuracy not exceeding 6.2%. The analysis of the obtained classification accuracies of the proposed method with precipitation data revealed that years with higher rainfall amounts during the summer crop-growing season (irrigation period) had lower overall accuracy (72.8% for 2017) whereas years encountering drier summer had very good accuracy (93.0% for 2019).

Keywords: Irrigation, synthetic aperture radar, normalized difference vegetation index, soil moisture, summer crops

Introduction

The current changing climate has altered the frequency and severity of extreme hydrological events (Grillakis, 2019) causing adverse impacts on crop production (Harkness et al., 2020) and endangering food security (Richardson et al., 2018). Insufficient precipitation and the significant increase in evaporative demand due to higher air temperatures have already affected agricultural regions particularly in arid and semi-arid regions (Jamshidi et al., 2019). As a natural response, water demand for crop cultivation has increased in the last decades (Tilman et al., 2011) despite the significant decrease in water resources in many regions worldwide (Scanlon et al., 2018). Over the last 50 years, irrigated areas is doubled (FAO, 2017) and it is projected to increase from 287 million hectares in 2005 to 318 million hectares in 2050 (Bruinsma, 2009). Given the regional water shortage, new agricultural policies should be adapted for a transition towards a more efficient and sustainable agriculture system to conserve water and enhance crop productivity.

Imposing sustainable water conservation policies at the core requires quantifying the spatial extent of the irrigated areas. Currently, the extent of irrigated areas at global scales is principally derived from country-level statistics and remains uncertain (Ozdogan et al., 2010; Portmann et al., 2010; Siebert et al., 2005; Thenkabail et al., 2009a). Although national statistical data gives the gist of irrigated areas and water use, these data may lack precision especially when irrigation

is not reported by the farmers. To address the need for a precise large-scale mapping of irrigated areas, remote sensing provides a powerful tool for mapping irrigated areas (Bazzi et al., 2019c; Gao et al., 2018; Hajj et al., 2014; Ozdogan et al., 2010; Pageot et al., 2020). With the availability of several operational cost-free and open access satellites (e.g., Landsat, Sentinels), remote sensing has been widely used for monitoring and managing agricultural crops from field level (Jamshidi et al., 2019; Segarra et al., 2020) to large domains (Bazzi et al., 2019d; Fayad et al., 2020; Ienco et al., 2019; Ndikumana et al., 2018).

Irrigation extent mapping using optical satellite data has been explored in several studies using various methodologies to distinguish between irrigated and rain-fed crops (Biggs et al., 2006; Kamthonkiat et al., 2005; Xiang et al., 2019). Nevertheless, these methods are developed based on a similar principle that the phenological differences between irrigated and non-irrigated crops (e.g., growth rate, greenness) are detectable by the vegetation spectral information derived from satellite optical sensors (Naser et al., 2020). Vegetation indices such as NDVI (Normalized Difference Vegetation Index) (Pervez and Brown, 2010), NDWI (Normalized Water Vegetation Index) (Xiang et al., 2019), NDRE (Normalized Difference Red-Edge) (Pageot et al., 2020) or GI (Greenness Index) (Chen et al., 2018) derived from Landsat, MODIS and Sentinel-2 data have been widely used to map irrigated areas. Most of the prior studies using optical data tend to classify irrigated/non-irrigated areas only for one specific crop type. The transferability of the methods based on the optical data is further limited in humid regions due to the cloud cover (Karakizi et al., 2018), and the marginal difference in the crop phenology between irrigated and rain-fed crops (Maselli et al., 2020b).

Synthetic Aperture Radar (SAR) data has been also exploited for mapping irrigated areas. The key element in the usage of the SAR data for irrigation mapping is the surface soil moisture (SSM) values that have been widely demonstrated to be correlated with the radar backscattering coefficients (Baghdadi et al., 2008a, 2011c, 2011b, 2016b; Bazzi et al., 2019a; El Hajj et al., 2016a, 2017). Several studies have shown that the C-band SAR temporal series derived from the Sentinel-1 (S1) satellite is efficient for mapping irrigated areas at plot scale (Bazzi et al., 2019c; Gao et al., 2018). The increase in the soil moisture due to an irrigation event causes an increase in the SAR backscattering coefficient between two consecutive SAR acquisitions if no rainfall is observed. However, since rainfall and irrigation have the same influence on the SSM values, it is important to distinguish between the increase of the SSM due to rainfall and irrigation.

In a recent study, Bazzi et al. (Bazzi et al., 2019c) used the differences between the S1 SAR signal calculated at grid-scale and plot scale to remove the rainfall-irrigation ambiguity. Using this technique and employing a convolutional neural network as a classification tool, Bazzi et al. (Bazzi et al., 2019c) mapped irrigated areas at plot scale using the S1 C-band SAR temporal series over a semi-arid region in Catalonia with an overall accuracy of 94%. Other studies including Gao et al. (Gao et al., 2018), Bousbih et al. (Bousbih et al., 2018) and Yann et al. (Pageot et al., 2020) have also used S1 SAR data for mapping irrigated areas with accuracy ranging from 78% to 82%. Bazzi et al. (Bazzi et al., 2020c) used a change detection model for detecting irrigation episodes at plot scale using S1 data. They achieved an overall accuracy of 76% in detecting irrigation episodes over grassland plots. The applicability of SAR data for irrigation mapping, however, could be limited in regions with frequent rainfall events (Bazzi et al., 2020c, 2020b). Additionally, the C-band SAR data has been reported to be more sensitive to canopy density than soil moisture in dense vegetative areas such as wheat and grasslands (Bazzi et al., 2020c; El Hajj et al., 2018b; Nasrallah et al., 2019). In the case of very well developed vegetation cover, the penetration of the C-band into the canopy is significantly reduced. El Hajj et al. (2018b) and Nasrallah et al. (2019) showed that the soil contribution in the C-band SAR backscattered signal (wavelength of 6 cm for S1) is negligible between the germination and heading growth stages in wheat due to the low penetration of the C-band signal to the soil surface. In this case, detecting an irrigation event using the increase in the backscattering coefficient could be challenging.

Another key factor in mapping irrigated areas using remotely sensed data is the classification method. Most of the previous studies on land cover classification and irrigation mapping are based on supervised classification techniques including Random Forest (RF) (Bazzi et al., 2019c, 2019d), Support Vector Machine (SVM) (Gao et al., 2018), or neural networks (NN) (Ienco et al., 2019; Ndikumana et al., 2018; Zhu et al., 2017). Supervised classification methods require obtaining in situ data (at yearly or half-yearly period) which is time and resource consuming, and may not be transferable spatially and temporally (Talukdar et al., 2020). To circumvent this issue, Bazzi et al. (2020a) proposed a spatiotemporal transfer-learning framework that transfers a CNN (Convolutional Neural Network) based irrigation classification model built on a source geographical region (Catalonia northeast Spain), to map irrigated areas on a target region (Tarbes of southwest France). Nevertheless, this method still requires in situ data over the target area in order to refine the source classification model.

Therefore, for continuous and yearly mapping of irrigated areas with fewer in situ measurements, an operational methodology capable of overcoming the limitation of terrain data collection is required. The proposed methodology should be reproducible and applicable for several years. In the context of operational mapping of irrigated areas, this study proposes an operational semi-supervised framework for mapping irrigated areas at a plot scale that overcomes the limitation of the availability of in situ data. The proposed methodology is based on a pre-step of selecting irrigated and non-irrigated plots to be served as training data for building a classification model. The training data selection is based on SAR and optical derived irrigation metrics (without using in situ data). The defined training data were then implemented in a RF classifier to map irrigated areas using SAR and optical temporal series. The obtained classifier was validated using real in situ data acquired over four years to assess the accuracy of the classification. The findings including irrigated area map could be later used by local authorities and stakeholders for estimating and managing water use at regional scales. These maps could help decision makers better follow the current irrigation situation and build future policies to manage water resources.

2. Materials

2.1 Study Site

The study site examined is located near Orléans city of north-central France (Figure 1). Located in the “Centre-Val de Loire” region, the study site is characterized by oceanic climate with an average rainfall of 730 mm per year with several rainfall events recorded in summer season. The cumulative precipitation, recorded from a local metrological station located in Orléans city, during the period between May and October for the years 2017, 2018, 2019 and 2020 reached 321 mm, 210 mm, 150 mm and 180 mm respectively. In the study site, irrigation mainly exists for summer crops (maize, sorghum, sunflower …) during the period between May and October of each year. The agricultural plots in the studied area are almost flat with very slight average slope of 3.2%.

2.2 Field Campaigns

In the study site, irrigation mainly occurs for spring and summer crops, which are generally sowed in April and May and harvested in September and October. For this reason, four field campaigns in the years 2017, 2018, 2019 and 2020 have been conducted over the study site in

June and July of each year to collect irrigation information. In the field campaigns, each plot has been registered as either irrigated, if irrigation was in progress during the campaign date or irrigation materials exist on the plot with summer crop cycle, or non-irrigated if neither irrigation nor irrigation materials has been observed.

Table 1 presents the number of terrain samples collected for each year. In 2020, a large database of 686 plots has been registered. However, a total of 92, 127 and 116 plots were registered for the years 2017, 2018 and 2019 respectively. Years 2017 and 2018 had the least percentage of non-irrigated plots from the total number of examined plots (approximately 28%). On the other hand, 49% and 42% of the collected plots are non-irrigated in the database of the years 2019 and 2020 respectively.

The average area of the visited plots in 2017, 2018, 2019 and 2020 is 10.96, 10.34, 8.35 and 7.18 hectares respectively (Table 1). For the four years together, the average area of all the visited plots is 8.0 ha. The area of the in situ plots varies between one ha and 48 ha. 70% of the plots have a surface area between 1 and 10 ha, 21% of the plots have a surface area between 10 and 20 ha and 9% of the plots have a surface area greater than 20 ha.

Table 1: Distribution of the number of in situ irrigated and non-irrigated plots for the four years.

Year	Irrigated	Non-Irrigated	Total	Average Area (ha)
2017	66	26	92	10.96
2018	91	36	127	10.34
2019	59	57	116	8.35
2020	395	291	686	7.18
Total	611	410	1021	8.00

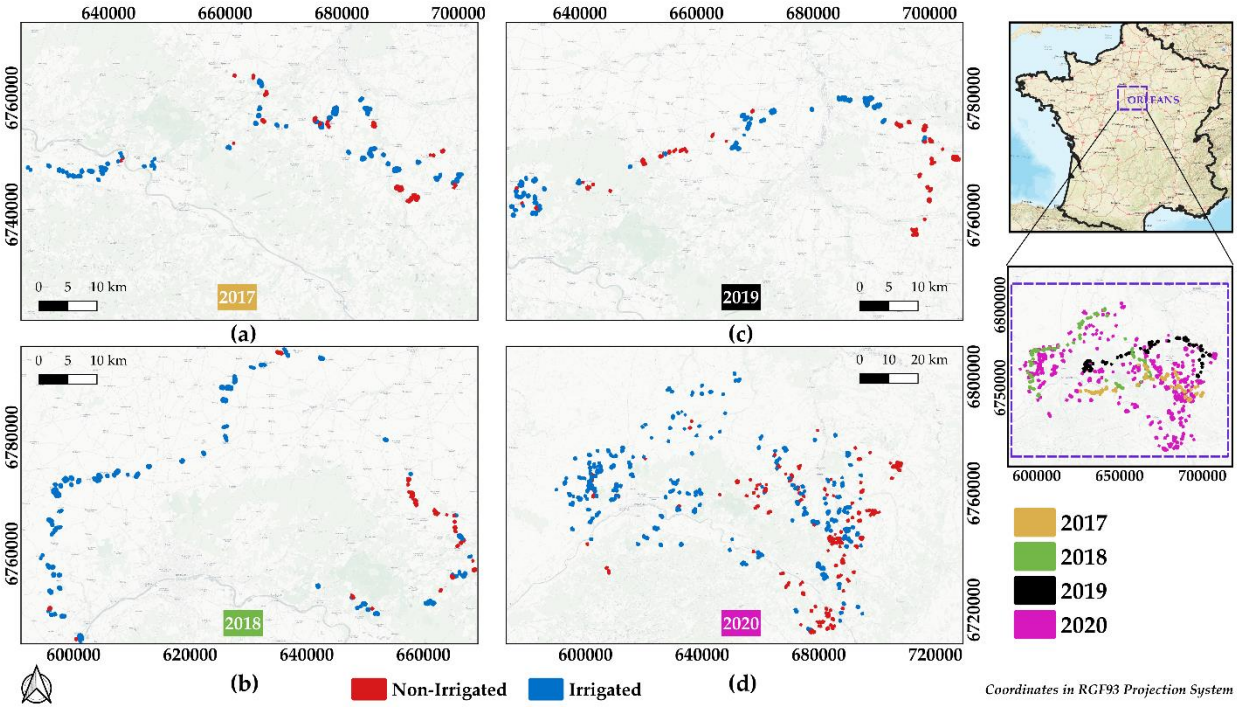


Figure 1: Location of the study site with collected in situ data near Orléans city in north-central France for (a) 2017, (b) 2018, (c) 2019 and (d) 2020. Non-irrigated plots are presented in red and irrigated plots are presented in blue.

2.3 RPG Data

The graphical parcel registry (RPG) is the French official graphical declaration system used by farmers, which provides an annual geo-localized representation of the agricultural parcels and their corresponding crop type. Covering around 26 million hectares, the RPG contains more than 6 million small parcels over the entire country. For each year, the corresponding RPG data was downloaded over the study site (Figure 1) (https://www.data.gouv.fr/en/datasets/) and used in this study. Among all the existing agricultural classes in the RPG, only classes corresponding to agricultural activities of summer and winter crops (including irrigated grasslands) were kept whereas vines and fruit trees were discarded. It is good to mention that the RPG data contains only the plot limits and their crop type and does not contain any information about irrigation. In this study, the RPG data was used to get the agricultural fields. Using the RPG agricultural fields, the training plot dataset (irrigated/non-irrigated) used for the irrigation classification model, are later selected with the training dataset selection criteria.

2.4 Sentinel-1 SAR Data

The high-resolution Level-1 ground range detected (GRD) product of the S1 satellite provides a 10 m x 10 m pixel spacing SAR image with a 6 days revisit time. However, with

several crossing orbits of the S1, four SAR images could be obtained in the period of 6 days within a study site. Thus, 578 available S1 images over the study site from different orbit acquisitions have been downloaded for the four years via the Copernicus website (https://scihub.copernicus.eu/dhus/#/home). Figure 2 shows an example of the frequency of the four SAR images acquired from the four S1 orbits for the year 2020 in both ascending 'A' and descending 'D' modes. Images is the descending mode are acquired in the morning (~05h00 UTC) whereas images in the ascending mode are acquired in the evening (~17h00 UTC). After the first acquired descending morning image (D1), another descending morning acquisition is available 24 hours later (D2). Thirty-six hours later, an ascending evening image is acquired (A1). Finally, the fourth image (A2) is acquired 24 hours after the first evening acquisition. The hatched area in Figure 2 shows the period where no images are acquired (2.5 days). After 6 days from the first descending image acquisition (D1), the first descending morning image is then repeated (revisit time of S1 satellite) and the same acquisition pattern is later repeated. Therefore, over the study site and for each year, four S1 temporal series (TS) are obtained later referred to as TS_I, TS_{II}, TS_{III} and TS_{IV} (for 2020, it corresponds to $TS_{D1}, TS_{D2}, TS_{A1}$ and TS_{A2} respectively). The four S1 temporal series correspond to S1 images from the four different S1 orbit acquisitions. Each TS in each year is composed of 42 to 46 S1 images (at 6 days revisit time) according to the studied year.

Radiometric and geometric calibration have been performed for all S1 images using the S1 toolbox developed by the ESA (European Space Agency). The radiometric calibration allows passing from the DN (digital number) of the pixel to backscattering coefficient (σ^0) in linear units while the geometric calibration insures the ortho-rectification of the S1 images using the digital elevation model (DEM) of the SRTM (Shuttle Radar Topography Mission) at 30 m spatial resolution.

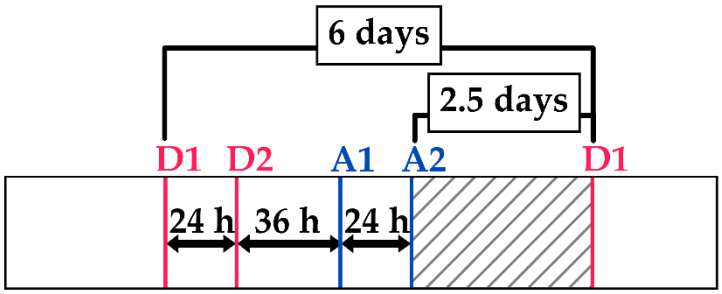


Figure 2: Example of the frequency of S1 images in ascending 'A' (evening, blue) and descending 'D' (morning, red) modes for the year 2020. Hatched area represents the period with no available SAR acquisitions.

2.5 Sentinel-2 Optical Data

For each year, temporal series of S2 optical images were downloaded from Theia website (https://www.theia-land.fr/) covering the period between February and October. This period largely fits the period of the irrigation of summer crops in the four years (usually between May and October). Optical S2 images were downloaded at a frequency of one to two images each month. The images downloaded from the French land data center (Theia) are calibrated for atmospheric correction and ortho-rectified for geometric correction (level 2A). Moreover, starting from 2018, Theia started providing monthly-synthesized cloud free S2 images (level 3A) which were also downloaded when available. The S2 images were mainly used to calculate the NDVI images used later in the training dataset selection criteria and the irrigation mapping classifier.

2.6 Global Precipitation Mission (GPM) data

The GPM mission is an international satellite mission initiated by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA) with the aim to provide global precipitation measurements from space (Huffman et al., 2014). The IMERG (Integrated Multi-satellite Retrievals for GPM) data product of the GPM offers global precipitation estimations at 0.1° spatial resolution (~10 km x 10 km) between 60°N and 60°S. In this study, the daily cumulative rainfall maps offered by the Final GIS (Geographic Information System) IMERG data (version 06) were downloaded for the period between February and October of each year (https://gpm.nasa.gov/data/directory). Nevertheless, the rainfall data from the GPM were not used in the proposed mapping method. They were only used to analyze and discuss the obtained results with rainfall registrations.

3. Methods

3.1 Overview

The proposed methodology, later referred to as S^2IM (Sentinel-1/Sentinel-2 Irrigation Mapping) consists of two major steps for mapping irrigated summer crops (Figure 3). In the first step, the irrigated/non-irrigated training plots are selected based on multi criteria derived from both SAR and optical data. The selection criteria of the training dataset are based on threshold values for the maximum attained NDVI for the plot during the studied period

(calculated from S2 data), and an irrigation possibility weight at each plot obtained using the newly derived irrigation event detection model (IEDM) at plot scale (Bazzi et al., 2020c, 2020b). After selecting the training dataset that corresponds to the plots deemed as irrigated and non-irrigated with a high confidence degree, the second step consists of implementing S1 data (radar backscattering coefficient at plot and grid scales), S2 data (NDVI) and the selected training plots into a random forest classifier to build a classifier for mapping irrigated areas (Irrigation Classifier). Finally, using the in situ dataset, the performance of the classifier was assessed using several accuracy metrics. The methodology was performed and validated for four years separately (2017, 2018, 2019 and 2020).

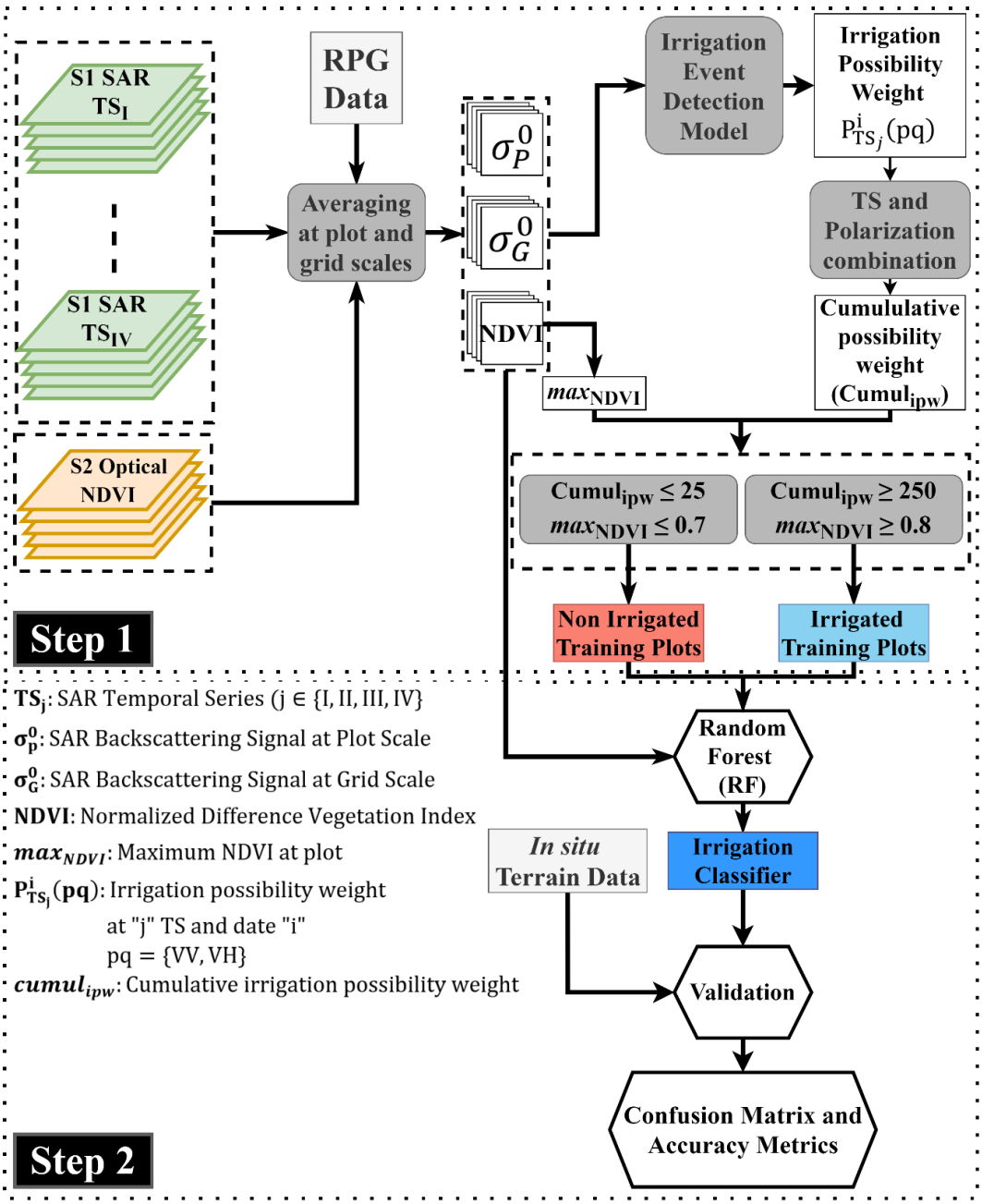


Figure 3: Overflow of the methodology for irrigation mapping at plot scale.

3.2 Sentinel-1 data

Using the plots' boundaries of the RPG data, the average S1 backscattering coefficient at plot scale σ_p^0 has been calculated for each acquired S1 image by averaging the pixel values within each plot. Before averaging the pixel values within each plot, an interior buffer of -10 m (10 m ~ one S1/S2 pixel) was applied to the RPG plots in order to avoid including boundary pixels from nearby plots, highways or surrounding vegetation. Moreover, RPG plots with a surface area less than 0.1 ha were not considered in order to avoid the speckle noise in the SAR data due to very small number of averaged pixels (0.1 ha ~ 10 S1 pixels). The σ_p^0 is calculated in both VV and VH polarizations. For each plot, four distinct S1 temporal series could be obtained ($TS_I, TS_{II}, TS_{III}, \text{and } TS_{IV}$) each at 6 days revisit time (maximum time difference between TS_I and TS_{IV} is 3.5 days). However, due to the limited overlapping zone of the S1 images, some plots of the RPG are only covered by two TS. In addition, the radar backscattering coefficients have been calculated at a grid of 10km x 10km (σ_G^0). Indeed, as shown by (Bazzi et al., 2019c, 2020b) the grid scale S1 backscattering helps reducing the uncertainty between rainfall and irrigation. In fact, the increase of the S1 backscattering signal at grid scale σ_G^0, between two consecutive S1 images, is mainly due to increase of soil moisture caused by a rainfall event. On the other hand, when σ_p^0 increases between two consecutive S1 images and σ_G^0 decreases or remain stable (no rainfall) then an irrigation event could have occurred. As such, the average S1 backscattering coefficient σ_G^0 at the grid scale was calculated by averaging the pixel values of bare soil pixels within each grid cell (10km x 10km) in both VV and VH polarizations. The extraction of bare soil pixels at each S1 date was performed by delineating the agricultural area from the French land cover map (Inglada et al., 2017) and applying a threshold value of the NDVI calculated from S2 images (NDVI < 0.4). Thus, for each agricultural plot, a temporal series of σ_p^0 at each TS is obtained at plot scale and its corresponding grid scale (σ_G^0). The plot and grid scale S1 backscattering coefficients for each TS were used as input layers for applying the irrigation event detection model (IEDM) and later in the RF classification of irrigated areas.

3.3 Sentinel-2 data

Using each available S2 image, the average NDVI value at plot scale was obtained by averaging the NDVI pixels' within each RPG parcel. The interior buffer of -10 m was also considered on the RPG plots to eliminate border pixels and plots with area less than 0.1 ha were

also not considered. Thus, an NDVI temporal profile is obtained for each plot in each year. The NDVI was first used as input in the IEDM, then in the selection criteria of the training dataset (maximum attained NDVI value) and finally as input data for the RF classifier.

3.4 Training dataset selection criteria

The selection of the irrigated/non-irrigated training plots from the RPG data is based on two threshold criteria fixed from S1 and optical derived metrics. Using the S1 data, an irrigation possibility weight was calculated for each plot by applying the IEDM on the several S1 temporal series in VV and VH polarizations. In addition, using the optical NDVI temporal profile, the maximum attained NDVI value at each plot was considered as additional selection metric.

3.4.1 Irrigation possibility metric

In a recent study, Bazzi et al. (2020b) developed a change detection model (called IEDM for irrigation event detection model) capable of detecting irrigation events at plot scale using S1 SAR temporal series. The IEDM principally uses the σ_p^0 and σ_G^0 for detecting the irrigation possibility at each S1 acquisition for each plot. In the IEDM, the increase in the SAR backscattering σ_p^0 between two consecutive SAR acquisitions is assumed be related mainly to the increase in the surface soil moisture (SSM) of the plot. However, since both rainfall and irrigation lead to an increase in the SSM values, the IEDM considers that rainfall/irrigation uncertainty could be removed by using the S1 backscattering at grid scale. Indeed, the increase in the σ_G^0 between two consecutive S1 acquisitions is most probably linked to a rainfall event occurring between the two S1 images.

For each plot and at each S1 image, the IEDM gives an irrigation possibility 0, 25, 50 and 100. The four irrigation possibility values are directly related to the change in the σ_p^0 between two dates t_i and t_{i-1} ($\Delta\sigma_P^0 = \sigma_p^0(t_i) - \sigma_P^0(t_{i-1})$). The value 0 corresponds to the absence of any irrigation chance between t_i and t_{i-1} caused either by the decrease of the σ_p^0 ($\Delta\sigma_p^0 \leq -0.5$ dB decrease of SSM) or increase of the σ_G^0 ($\Delta\sigma_G^0 \geq 1$ dB rainfall event occurred). The low irrigation possibility weight (value=25) correspond to the absence of rainfall events between t_i and t_{i-1} insured by the decrease of σ_G^0 ($\Delta\sigma_G^0 \leq 0.5$ dB) and a slight modification in the σ_p^0 between t_i and t_{i-1} ($-0.5 \leq \Delta\sigma_p^0 < 0.5$ dB). The medium irrigation possibility weight (value=50) is associated to a moderate increase in the σ_p^0 ($0.5 \leq \Delta\sigma_p^0 < 1$ dB) with the absence

of rainfall events ($\Delta\sigma_G^0 \leq 0.5$ dB). Finally, the high possibility of irrigation (value=100) is insured when the σ_p^0 strongly increases ($\Delta\sigma_p^0 \geq 1$ dB) with no rainfall event detected ($\Delta\sigma_G^0 \leq 0.5$ dB). However, additional filters are used in the IEDM to confirm the existence of an irrigation event or to remove falsely detected irrigation events. One of the additional filters considerers the SSM estimations at plot scale easily estimated using σ_p^0 and the neural network technique proposed by El Hajj et al. (El Hajj et al., 2017). For example, to confirm the existence of low irrigation possibility weight (value=25) with only slight modification of σ_p^0, the IEDM uses the SSM estimation at plot scale. In fact, the low possible irrigation event is detected if and only if the plot's SSM estimation at time t_{i-1} was high (SSM $\geq 20\ vol.\%$) and remained with high value to time t_i (humid soil conditions persisted from time t_{i-1} to t_i). Moreover, an NDVI filter is used to reduce some falsely detected irrigation events. The false detections could be due to the increase in the σ_p^0 values related to the change of surface roughness (Aubert et al., 2011; Baghdadi et al., 2018a). The NDVI filter proposes that if an event is detected with low NDVI value at date t_i (NDVI < 0.4) and the NDVI value one month later at t_{i+30} decreases or remains stable ($NDVI_{t_{i+30}} - NDVI_{t_i} \leq 0.1$) then the event is discarded (crop cycle in decreasing stage or persistent bare soil conditions).

The IEDM was validated on several geographical areas with different climatic contexts where Bazzi et al. (Bazzi et al., 2020b) proved the applicability of this algorithm on several study sites (semi-arid and temperate areas). Moreover, Bazzi et al. (Bazzi et al., 2020c) validated the IEDM for irrigation detection at intensively irrigated grassland plots. They reported that irrigation events could be detected with an F_score of 75% when using both the VV and VH polarizations and four S1 temporal series (all available acquisitions over a study site).

As proposed in Bazzi et al. (2020b) and Bazzi et al. (2020c), the IEDM was applied at each plot of the RPG, at each S1 temporal series separately. This means that, for a given year, the IEDM was applied at each plot using the S1 data (σ_p^0 and σ_G^0) of the same temporal series at 6 days revisit time. The separate use of the IEDM over the four TS is basically due to the diurnal effect between the morning and the evening acquisitions that may lead to uncertain irrigation detection (Bazzi et al., 2020c; Brisco et al., 1990; van Emmerik et al., 2015). Moreover, as the S1 images are acquired at different incidence angles, the use of all S1 temporal series together necessitates a normalization of the incidence angle, which generally does not allow eliminating completely the effect of incidence angle. Thus, an undesirable error will be added to the σ_p^0.

Therefore, at each plot and each date (i) of each TS an irrigation possibility weight $P^{i}_{TS}(pq)$ is obtained for each polarization (TS is the temporal series, i is the date of the image and pq = VV or VH). Then, the results of the irrigation possibilities of the four TS were combined in both VV and VH polarizations to obtain an irrigation possibility metric for each plot.

Figure 4 illustrates the combination procedure of the four temporal series and both VV and VH polarizations at each plot. For a given plot, four S1 TS are available where each TS is treated separately with the IEDM for both VV and VH. The application of the IEDM gives to each S1 date (i) in each TS an irrigation possibility indicator in both VV and VH polarizations $P^{i}_{TS}(pq)$ (pq=VV or VH). Then, the results of the IEDM from the four series are summed for both VV and VH separately. In other words, the irrigation possibilities appearing in the search window of the 4 consecutive S1 images (Figure 2) at 3.5 days interval are added. For example, in the VV and the VH polarizations, the four irrigation possibilities occurring at the first acquisition date P^{1}_{I}, P^{1}_{II}, P^{1}_{III}, and P^{1}_{IV} of the four S1 temporal series (3.5 interval between P^{1}_{I} and P^{1}_{IV}) are summed to obtain one value for VV (P^{1}_{VV}) and one value for VH (P^{1}_{VH}). As a result, two irrigation indicator series are obtained from the four TS; one for VV and the other for VH henceforth referred to P-VV and P-VH respectively.

As suggested by Bazzi et al. (2020c), the combined use of VV and VH provides better detection of irrigation events and reduces significantly the false detection. Thus, P-VV and P-VH were further combined. For each irrigation indicator value obtained at each date (i) of P-VV and P-VH series, if no irrigation possibility exists within one polarization (P^{i}_{VV} = 0 or P^{i}_{VH} =0) then the irrigation event is not considered. Thus, the irrigation indicator that combines VV and VH P^{i}_{VVVH} will be 0 (no irrigation event retained by the algorithm). If an irrigation possibility exists simultaneously within both VV and VH ($P^{i}_{VV} \neq 0$ and $P^{i}_{VH} \neq 0$) then the maximum of the possibility weight value between P^{i}_{VV} and P^{i}_{VH} is considered. Finally, the cumulative irrigation possibility weight for each plot "$cumul_{ipw}$" is the sum of all the irrigation possibilities of combined VV and VH ($\sum_{i=1}^{n} P^{i}_{VVVH}$) divided by the number of the used temporal series (maximum 4 in our case). In fact, some plots are not covered by four images due to the limited overlapping extent of the S1 images. Therefore, it was important to normalize the cumulative irrigation possibility weight by the number of used TS to get a discrete metric independent of the number of the used TS. Thus, each plot has an irrigation indicator deduced from the application of the IEDM on all possible TS using VV and VH polarizations. This indicator will be later used to select training samples of irrigated/non-irrigated plots. The

$cumul_{ipw}$ is a value that represents for each plot, the accumulation of the irrigation possibility weights $P^{i}_{TS}(pq)$ which are derived from the IEDM at each S1 image in both VV and VH and normalized for the number of TS used. Thus, as the detectable irrigation events on the plot increase, the $cumul_{ipw}$ value will increase. Consequently, very low $cumul_{ipw}$ values could be an evidence of absence of irrigation events at the plot and very high $cumul_{ipw}$ values could be an evidence of intensively irrigated plot. Therefore, a non-irrigated plot to be selected in the training dataset, the value of the $cumul_{ipw}$ should be very low to insure that approximately no possible irrigations are detected on the plot. On the other hand, high value of $cumul_{ipw}$ is required for a plot to be selected as irrigated training plot to be highly sure that the plot corresponds to an irrigated plot.

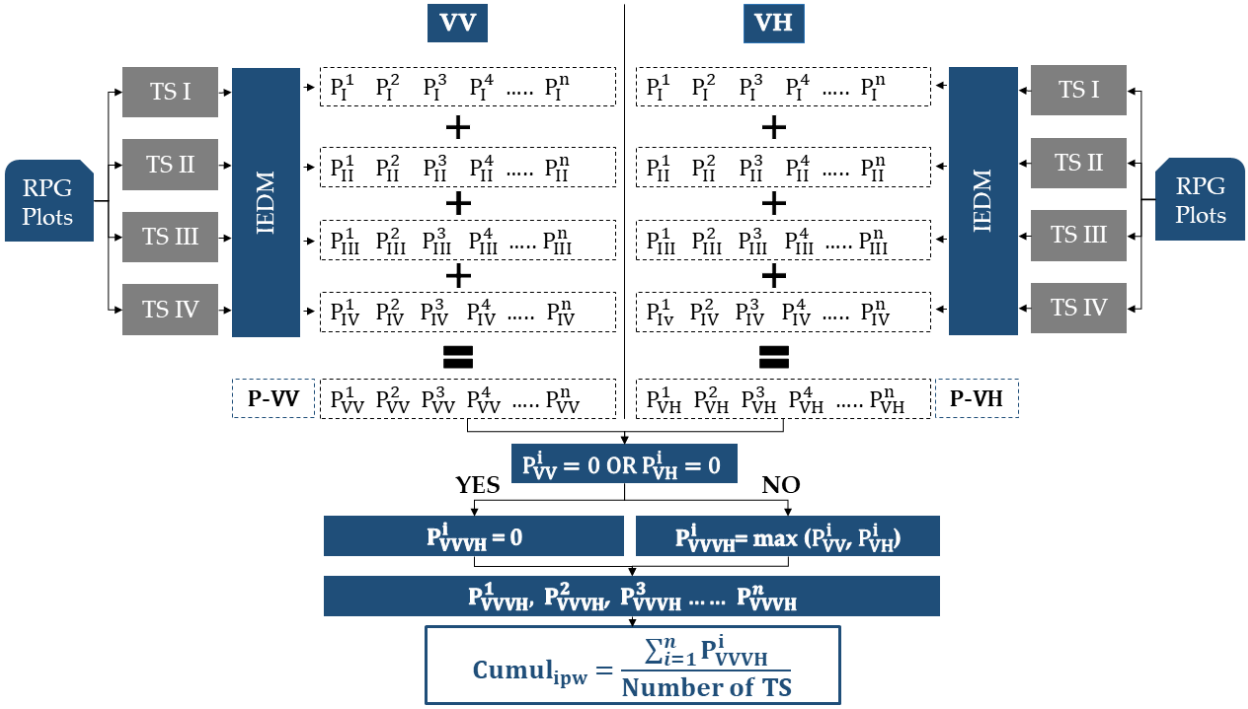


Figure 4: Detailed description of the combination of temporal series (TS) in VV and VH polarizations to calculate the irrigation metric for each plot using the IEDM. n is the number of images (dates) in each TS and i represents the date. Number of TS equals to 4 in our case

3.4.2 Maximum NDVI metric

Several studies have shown that vegetation indices derived from optical images could be used to separate irrigated and non-irrigated plots due to the different spectral response between irrigated and non-irrigated plots.

The NDVI, which represents a proxy measure for absorbed photosynthetic active radiation, is a commonly used vegetation index to map irrigated and non-irrigated crops (Bousbih et al.,

2018; Ozdogan et al., 2010; Pageot et al., 2020). In fact, several studies assessed the effect of water abundancy on the NDVI values and showed that the NDVI values of different crops increase when the available soil moisture for vegetation increases (Ji and Peters, 2003; Kawabata et al., 2001; Potter and Brooks, 1998; Wulder et al., 2004). When the crop benefits from additional amounts of water through irrigation, highest levels of photosynthesis could be achieved along with highest biomass and densest vegetation cover. These three mentioned biophysical properties induce high NDVI values for irrigated crops. Indeed, several studies have demonstrated that irrigated crops, especially maize and wheat, show higher NDVI than non-irrigated crops (Wardlow and Egbert, 2008). For example, Pervez and Brown (Pervez and Brown, 2010) used the maximum NDVI criterion to map irrigated areas over the entire US continent. In their study, they showed that the maximum NDVI (peak value) for non-irrigated crops (including corn, dry beans, pasture and millet) does not exceed 0.75. However, all irrigated crops showed a peak NDVI value higher than 0.8.The same methodology was later used by Pervez and Brown (Brown and Pervez, 2014) to map irrigated areas over USA and extract the temporal change of irrigated surface.

Unlike other low temporal resolution satellites, the high revisit time of the S2 satellite allows obtaining at least two cloud free images each month. The high revisit time of the S2 permits the detailed monitoring of the NDVI values at plot scale and thus extract the maximum NDVI or a value near to the maximum NDVI attained at the plot.

For these reasons, we propose to include additional criteria for the selection of the training irrigated/non-irrigated plots based on the maximum value of the NDVI. As mentioned in Section 2.2, the irrigation period in our study site mainly occurs for irrigated spring/summer crops, which are sowed in April and May and harvested between September and October. Therefore, the maximum NDVI value reached for each plot during the crop cycle between May and October was registered ($\max_{\mathrm{NDVI}}$). Based on the literature and the analysis of the in situ maximum NDVI for irrigated and non-irrigated plots (results shown later in the Results section), we were able to define two threshold values to separate irrigated and non-irrigated training plots. For the max_{NDVI} metric, we consider that for the plot to be selected as non-irrigated training plot the max_{NDVI} value must not exceed 0.7 whereas the max_{NDVI} value for a plot to be selected as irrigated training plot must exceed 0.8. The considered thresholds (< 0.7 for non-irrigated and > 0.8 for irrigated) allow obtaining irrigated and non-irrigated training plots with high confidence and less overlap between the two classes.

3.4.3 Selection criteria of Irrigated/Non-Irrigated plots

Based on the two calculated metrics, $cumul_{ipw}$ and $\max_{\mathrm{NDVI}}$, the irrigated and non-irrigated plots for the training phase of the RF each year are selected.

For the max_{NDVI} metric, we consider that for the plot to be selected as non-irrigated training plot the max_{NDVI} value must not exceed 0.7 whereas the max_{NDVI} value for a plot to be selected as irrigated training plot must exceed 0.8.

For the $cumul_{ipw}$ metric, a plot must have a low value of $cumul_{ipw}$ to be considered as non-irrigated training plot. In this study, we consider that for the plot to be selected in the non-irrigated class of the training dataset, the $cumul_{ipw}$ value should be less than or equal to 25. The cumulative possibility weight of 25 was considered as a very low value indicating very low number of detected irrigation events by the IEDM at the plot during the whole crop season. For example, using four temporal series as present in Figure 4, a plot achieving a $cumul_{ipw}$ value of 25 can have only one high possibility-weight event (100) present on only one image of the four images (from 4 TS: $P_{I}^{i}, P_{II}^{i}, P_{III}^{i}, P_{IV}^{i}$) during the whole season in both VV ($P_{VV}^{i} = 100$)and VH ($P_{VH}^{i} = 100$) ($P_{pq}^{i} = P_{I}^{i} + P_{II}^{i} + P_{III}^{i} + P_{IV}^{i} = 100 + 0 + 0 + 0$). In this case the $cumul_{ipw}$ is the maximum between P_{VV}^{i} and P_{VH}^{i} (100) divided by 4 TS (100/4 = 25). Another example to attain a $cumul_{ipw}$ value of 25, if a medium possibility-weight irrigation is detected on only one image of the four TS ($P_{I}^{i}, P_{II}^{i}, P_{III}^{i}, P_{IV}^{i}$) in VV ($P_{VV}^{i} = 50$) and VH ($P_{VH}^{i} = 50$) and another medium possibility-weight irrigation is detected over one image of different four TS (date "k") in VV and VH ($P_{VV}^{k} = 50$) and $P_{VH}^{k} = 50$). In this case, the $cumul_{ipw}$ value reaches 25 ((50+50)/4).

For the irrigated class, we consider that a plot must have a high value of $cumul_{ipw}$ to be taken for the irrigated training samples. The high $cumul_{ipw}$ value insures the detection of several irrigation events on the plot during the crop season. In this study, a value of $cumul_{ipw}$ greater than or equal to 250 is fixed to consider a plot in the irrigated training sample. For example, using the four TS, a value 250 could be achieved if five irrigation events are detected with high possibility weights (100) and each irrigation event is present on two of the four images from the 4 TS (e.g. $P_{pq}^{i} = P_{I}^{i} + P_{II}^{i} + P_{III}^{i} + P_{IV}^{i}$ = 100+100+0+0) in both VV ($P_{VV}^{i} = 200$) and VH ($P_{VH}^{i} = 200$). These 5 irrigation events will give a cumulative irrigation possibility weight

equals to 250 (5 irrigation events x 200 possibility weight / 4 TS). Thus, the value greater than or equal to 250 insures that a sufficient number of irrigation events are detected on the plot.

Finally, the selection of the training dataset for each class was fixed by combining both metrics (max_{NDVI} and $cumul_{ipw}$). This means that a plot is considered as "irrigated training plot" if the $max_{NDVI} \geq 0.8$ and the $cumul_{ipw} \geq 250$ simulatenously. In contrast, a plot is selected as "non-irrigated training plot" if the $max_{NDVI} \leq 0.7$ and the $cumul_{ipw} \leq 25$. For each of the four years (2017, 2018, 2019 and 2020), an independent training dataset of irrigated and non-irrigated plots has been selected from the RPG data using the explained selection criteria. This means that for each year, a corresponding classification model could be obtained.

3.5 Random Forest Classifier

3.5.1 Training Phase

Through literature, random forest (RF) has widely demonstrated its ability to perform a high quality classification. Particularly, irrigation mapping using random forests has been recently exploited by several studies (Bazzi et al., 2019c; Gao et al., 2018; Pageot et al., 2020). Moreover, studies dealing with mapping irrigated areas proved the reliable use of both SAR and optical temporal series for mapping irrigated areas. For example, Pageot et al. (Pageot et al., 2020) used both S1 time series and optical derived vegetation indices to map irrigated and rain-fed summer crops in a humid area. Moreover, Bazzi et al. (Bazzi et al., 2019c) showed that the use of the grid scale σ_G^0 S1 temporal series along with the plot scale σ_p^0 series and NDVI derived from S2 images in a RF classifier enhances the accuracy of irrigated area mapping. In addition, Gao et al. (Gao et al., 2018) demonstrated that using statistical metrics derived from S1 SAR series at plot scale in a random forest classifier leads to a good accuracy for mapping irrigated areas. Following these studies, the RF classifier was used for mapping irrigated areas using σ_p^0, σ_G^0 and NDVI temporal series (Figure 3). σ_p^0, σ_G^0 and NDVI temporal series of the training dataset derived from the previously explained selection criteria (Section 3.4) were used to train the RF classifier. It is good to mention that for each year a RF classifier was developed using the training dataset corresponding to each year (2017, 2018, 2019 and 2020).

3.5.2 Validation and assessment phase

To assess the accuracy of the obtained RF classifier in each year, we used the in situ terrain dataset of irrigated/non-irrigated plots (Section 2.2) as a validation dataset. This means that, for each year, the corresponding RF classifier, built using the selected training dataset, was applied on the in situ terrain plots to predict whether each plot is irrigated or not. From the obtained confusion matrix between predicted and in situ labels, the accuracy of the RF classifiers could be assessed each year. The classifier accuracy has been evaluated using several accuracy metrics including the overall accuracy (OA), the weighted F_score (F_score), the F_score of the irrigated class (F_score_Ir) and the F_score of the non-irrigated class (F_score_Nir). The OA shows the percentage of correctly classified plots to the total number of plots while the F_score presents the harmonic mean between the precision and the recall of each class. For each year, a classification accuracy report of the mentioned metrics could be obtained.

In addition to the validation of the RF classifier each year using its corresponding terrain data, we also compared the reported accuracy of each year using the proposed methodology (S^2IM) to the accuracy obtained when building a RF classifier using five folds cross validation of the in situ data. This comparison will allow us evaluate the robustness of the proposed method against the classical training/validation methods that directly use terrain data for classification.

4. Results

4.1 Irrigated vs non-irrigated plots

Figure 5 presents an example of the temporal evolution of the σ_p^0 (black dashed line) and σ_G^0 (brown dashed line) for one of the four TS along with the NDVI (green line) for an irrigated maize plot (Figure 5a) and non-irrigated maize plot (Figure 5b) between February and October 2019. The blue bars in the figure represent the daily rainfall amounts given by the IMERG GPM product described in Section 2.6. Between February 2019 and mid-May 2019, σ_p^0 and σ_G^0 of each plot show the same behavior for both irrigated and non-irrigated plots. Indeed, for this period both σ_p^0 and σ_G^0 increase with rainfall events (blue bars) and decrease with the absence of rainfall. Moreover, the low NDVI values in the same period for both plots (approximately 0.2) indicate the absence of a vegetation cycle (bare soil conditions). After mid-May, the maize cycle of both irrigated and non-irrigated plots started when the NDVI values increased from 0.2

to 0.6 for both irrigated and non-irrigated plots between 15 May and 15 June 2019. This increase in the NDVI took place with a cumulative rainfall of 110 mm. After a rainfall event on 23 June 2019, no rainfall events occurred for approximately 1.5 months (until 12 August 2019) (very dry conditions). During this period, the σ_G^0 values of both irrigated and non-irrigated plots decreased gradually between 26 June and 12 August indicating no rainfall events.

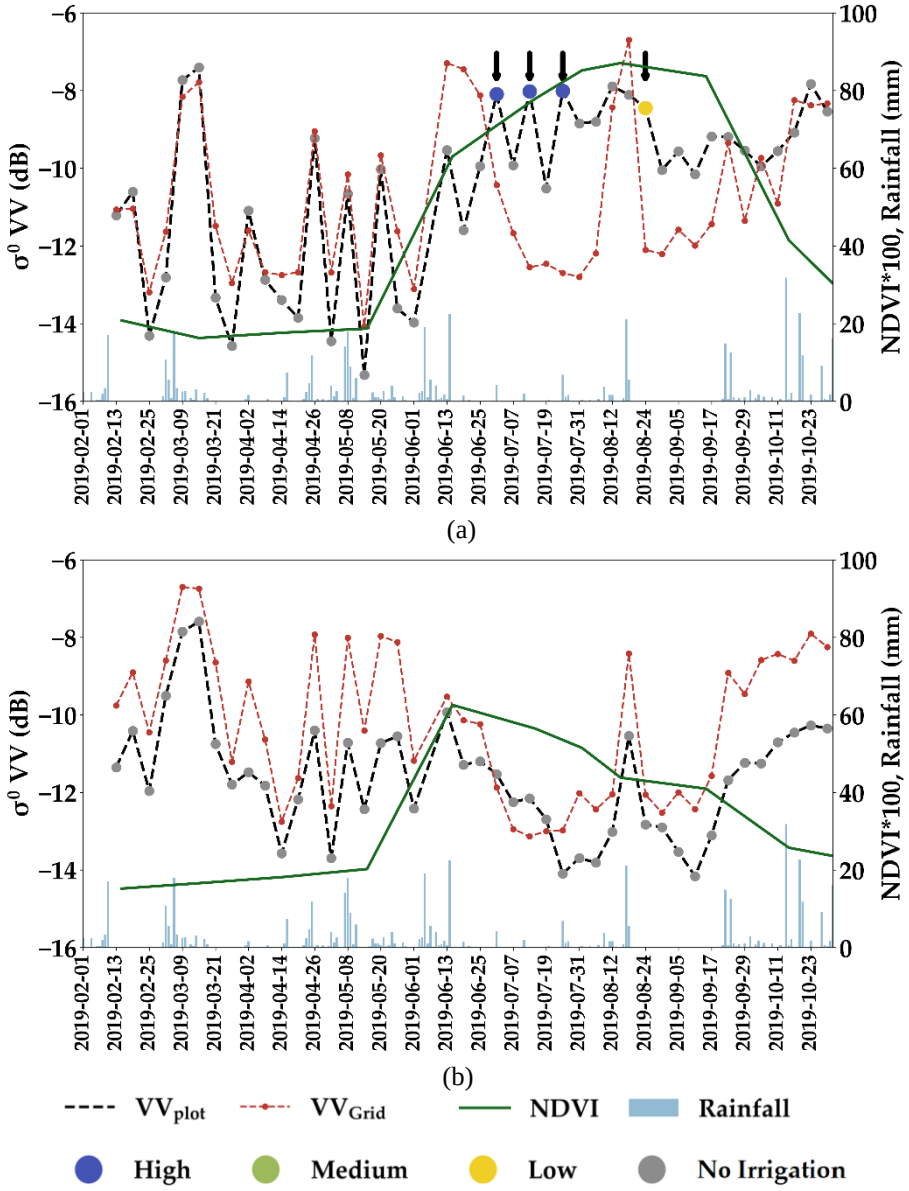


Figure 5: Example of σ_p^0 , σ_G^0 and NDVI temporal profiles for (a) irrigated maize plot and (b) non-irrigated maize plots for the year 2019. Black arrows highlights the dates with detected irrigation events by the IEDM. Blue bars show the daily rainfall from the GPM data.

For the irrigated plot, the σ_p^0 value increased between two consecutive S1 images on 26 June 2019 and 01 July ($\Delta\sigma_p^0$=1.9 dB) whereas for the same SAR images the σ_G^0 value decreased by 2.3 dB. Due to the important increase in σ_p^0 and important decrease in σ_G^0 between two

consecutive SAR images, the IEDM was capable of detecting the first irrigation event with high irrigation possibility weight on the irrigated plot on 01 July 2019. Similarly, two additional high probable irrigation events were detected on the S1 images of 13 July and 25 July (black arrows). The three high chance detected irrigations are mainly detected due to the important increase in σ_p^0 between two consecutive S1 image and the decrease of σ_G^0 indicating no rainfall events occurring. Due to these three detected irrigation events, the NDVI of the irrigated plot increased from 0.6 to 0.82 for the period between 15 June 2019 and 30 July 2019. Finally, a fourth irrigation event with low possibility weight has been detected on 24 August 2019 due to only slight decrease in σ_p^0 but at high level of radar signal at plot scale (-9 dB and SSM ≥ 20 vol.%) accompanied by a sharp decrease of σ_G^0 between 18 and 24 August 2019.

In contrast, the temporal profile of σ_p^0 of the non-irrigated plot (Figure 5b) decreases during the dry period (between 23 June and 18 August) showing no possible irrigation events on the plot. The similar behavior between σ_p^0 and σ_G^0 during the dry period of the season could be an evidence that the plot did not receive any additional water supplement. Therefore, the IEDM did not detect any irrigation events on the non-irrigated plot. Moreover, the NDVI value between 15 June and 30 July decreased from 0.6 to 0.5 and then continued its decreasing pattern until the end of the cycle.

4.2 Comparison of irrigation derived metrics using in situ data

In this section, we present a comparison between in situ irrigated and non-irrigated plots as a function of the irrigation derived metrics in the four years. We present first the max_{NDVI} metric derived from the NDVI temporal profile and then the $cumul_{ipw}$ metric derived using the IEDM. It is good to mention that the objective of this metrics comparison for in situ data is to only demonstrate the separability between the irrigated and non-irrigated classes using the proposed metrics. However, in situ data were not used to build the classification model of the S^{2}IM (step 1 of Figure 3) and were only used for the validation of the built RF classifier each year.

4.2.1 Maximum NDVI value (max_{NDVI})

Figure 6 presents the boxplot of the distribution of the maximum NDVI value (max_{NDVI}) acquired by the in situ plots in the summer cycle (between May and October) for the four

different years. The red line in the box plot represents the median value whereas the black point represents the mean value.

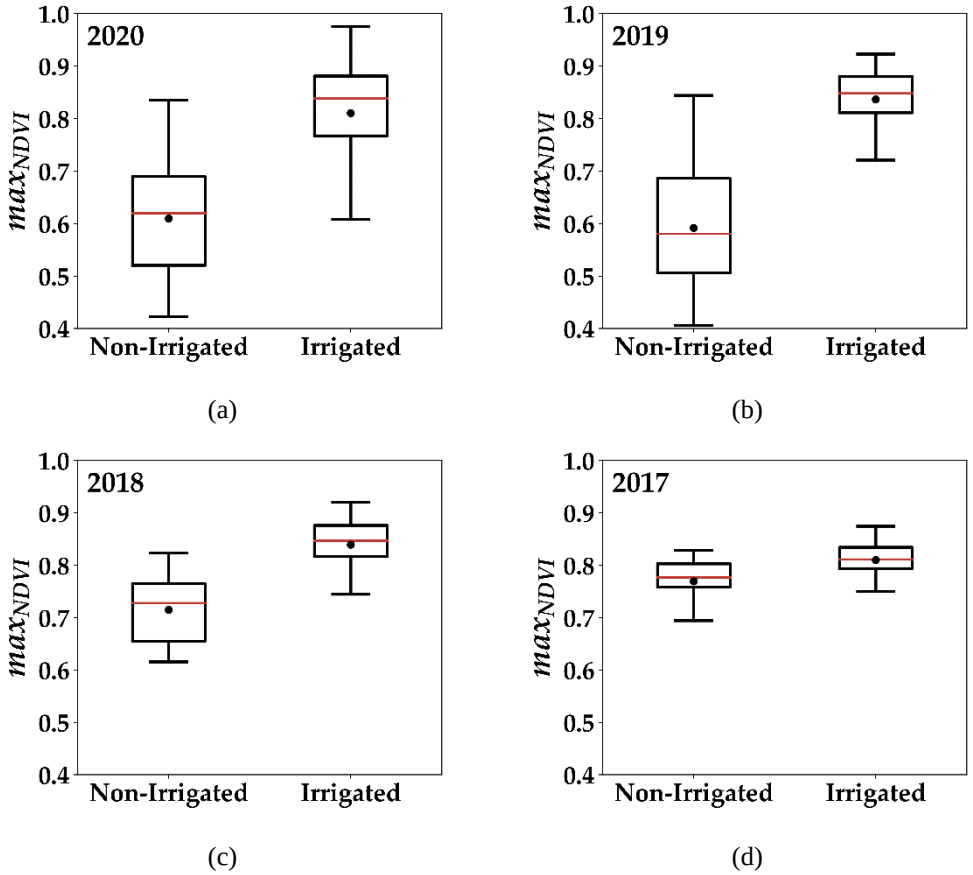


Figure 6: Boxplots of the distribution of maximum NDVI values (max_{NDVI}) for in situ irrigated and non-irrigated plots in (a) 2020, (b) 2019, (c) 2018 and (d) 2017. Red line in each box represents the median value while the black dot shows the mean value. The number of the irrigated and non-irrigated plots for each year corresponds to the number of in situ plots presented in Table 1 of Section 2.2.

In Figure 6a of 2020, the distribution of max_{NDVI} for non-irrigated plots shows that the average of max_{NDVI} value reaches 0.61 with approximately same median value (0.62). Moreover, the non-irrigated boxplot shows that 75% of the plots have a maximum NDVI value less than 0.69. In contrast, the irrigated plots have an average max_{NDVI} value of 0.82 and a median of 0.81 where 86.5% of the plots have a max_{NDVI} value greater than 0.7. In Figure 6b, the distribution of the max_{NDVI} value for the non-irrigated plots in 2019 shows approximately similar behavior as 2020 with a median value reaching 0.58 and a mean value of 0.59. Also in 2019, more than 75% of the non-irrigated plots attain a maximum NDVI value less than 0.7. However, irrigated plots in 2019 show that all the irrigated plots attain a maximum NDVI value greater than 0.7 with an average of 0.84 and a median value of 0.85. In 2018, the distribution of the maximum NDVI value for non-irrigated plots shows an average value of 0.71 and a median of 0.73 whereas that of the irrigated plots shows an average of 0.84 with a median value

of 0.85 (Figure 6c). In 2017 (Figure 6d), all the irrigated plots have a max_{NDVI} value greater than 0.7. However, Figure 6d of 2017 shows that, in situ irrigated and non-irrigated plots have close distributions with max_{NDVI} average value of 0.77 and 0.81 for non-irrigated and irrigated plots respectively. Given the poor separability between irrigated and non-irrigated classes in 2017, it will be difficult to map accurately irrigated plots for the year 2017.

4.2.2 IEDM cumulative irrigation ($cumul_{ipw}$)

The histograms in Figure 7 present the distribution of the $cumul_{ipw}$ metric obtained for in situ data for the four years. In Figure 7a, the $cumul_{ipw}$ distribution in 2020 shows good discrimination between irrigated and non-irrigated plots. In fact, 26% of the non-irrigated plots encounter a $cumul_{ipw}$ value less than 25, 41% have a $cumul_{ipw}$ less than 50 and 67% have a $cumul_{ipw}$ less than 100. In contrast, irrigated plots register higher $cumul_{ipw}$ values than the non-irrigated plots where 80% of the irrigated plots have a $cumul_{ipw}$ value more than 100, 54% have a $cumul_{ipw}$ value more than 150 and 33% have a $cumul_{ipw}$ value greater than 200. In 2019, Figure 7b shows that better discrimination between irrigated and non-irrigated plots is available using the $cumul_{ipw}$ metric. Indeed, 87% of the non-irrigated plots have a $cumul_{ipw}$ value less than 75 including 38% less than 25 and 22% between 25 and 50. Moreover, 61% of the irrigated plots in 2019 have a $cumul_{ipw}$ value greater than 150, 46% greater than 200 and 24% greater than 250. On the other hand, the distribution of $cumul_{ipw}$ values for irrigated and non-irrigated plots in both 2018 (Figure 7c) and 2017 (Figure 7d) have less separability than that present in 2020 and 2019. In 2017 and 2018, less number of irrigation events are detected over irrigated plots. Thus, the $cumul_{ipw}$ at irrigated plots did not reach high values as that reaches in 2019 and 2020 and therefore the histogram of the irrigated plots is closer to that of the non-irrigated plots. For example, in 2017, Figure 7d shows that 34% of the non-irrigated plots have a $cumul_{ipw}$ value less than 50 and 73% of the non-irrigated plots have a $cumul_{ipw}$ value less than 100. However, 48% of the irrigated plots have a $cumul_{ipw}$ value less than 100 and 52% of the irrigated plots have a $cumul_{ipw}$ value between 100 and 200. Moreover, the irrigated plots in 2017 have maximum $cumul_{ipw}$ of 200 whereas the maximum $cumul_{ipw}$value of irrigated plots in 2019 and 2020 is between 400 and 450.

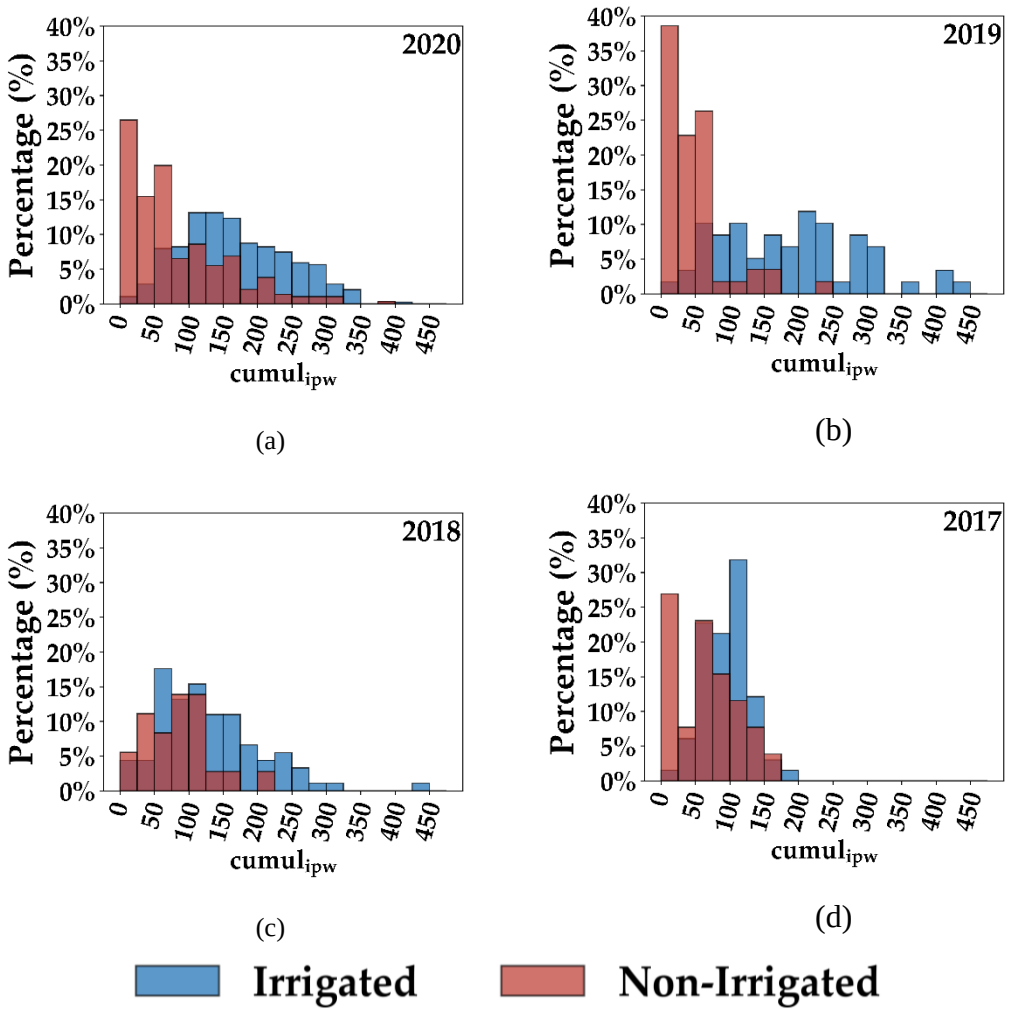


Figure 7: Distribution of the cumulative irrigation metric "$cumul_{ipw}$" for in situ irrigated (blue bars) and non-irrigated (red bars) plots for (a) 2020, (b) 2019, (c) 2018 and (d) 2017. The overlap between irrigated and non-irrigated classes appears with dark red color. The $cumul_{ipw}$ is unitless.

4.3 S^{2}IM selected training data

Table 2 presents the number of the training samples of irrigated and non-irrigated plots selected using the proposed metrics for each year. The total RPG represent the total number of investigated summer crop plots in the study area.

Table 2. Number of the selected irrigated and non-irrigated training samples for each year

Table 2: Number of the selected irrigated and non-irrigated training samples for each year

Year	Non-Irrigated Plots	Irrigated Plots	Total RPG Plots
2020	1486	2209	19938
2019	1033	614	15958
2018	1441	1176	14161
2017	852	289	23599
Total	4812	4288	73656

Figure 8 shows the NDVI temporal profiles of the selected irrigated and non-irrigated training data (Table 2) and the in situ data for each year. The blue dashed line represents the temporal profile of the average NDVI of all the selected irrigated plots surrounded by the standard deviation (shaded blue) while the black line represents the temporal NDVI profile of the in situ irrigated plots. The red dashed line shows the temporal profile of the average NDVI for the selected non-irrigated plots with the red shading of the standard deviation value compared to the in situ NDVI temporal profile in grey line.

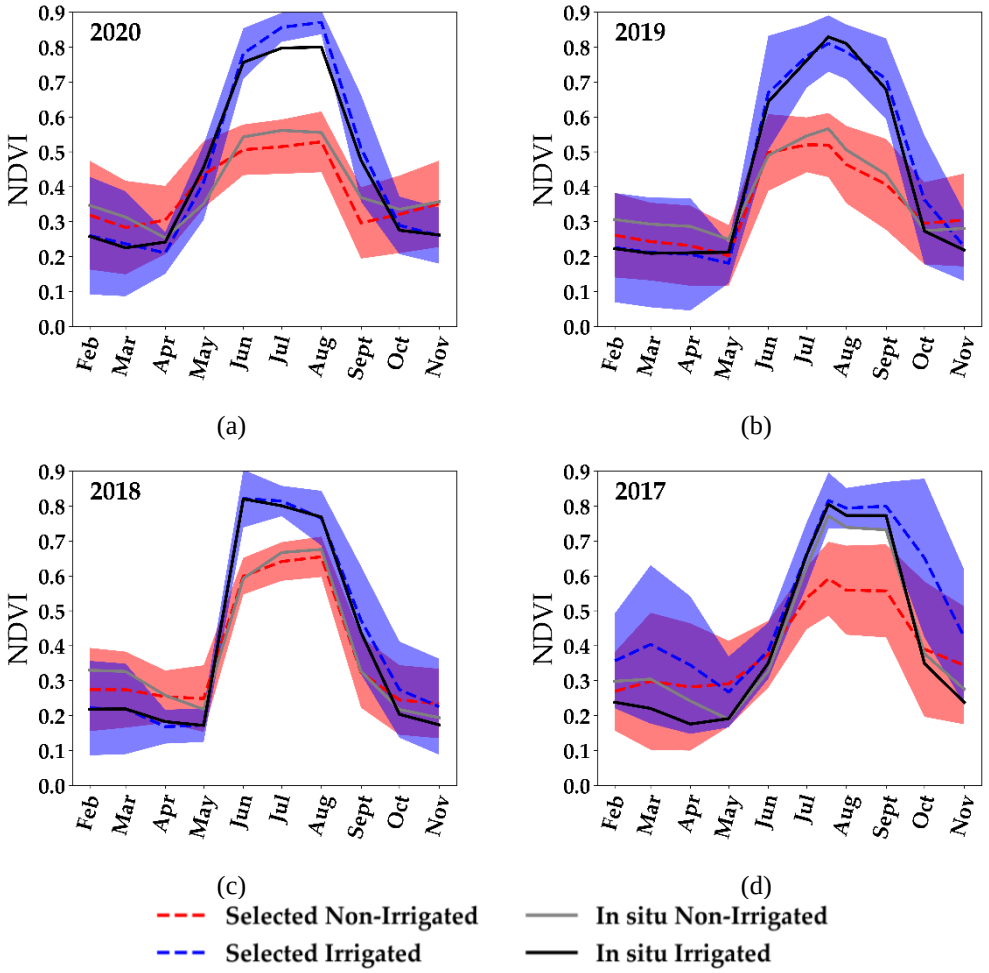


Figure 8: Comparison of average NDVI temporal profiles between selected training plots and in situ plots for (a) 2020, (b) 2019, (c) 2018 and (d) 2017. Red dashed line represents the selected non-irrigated plots while grey line represents the in situ non-irrigated plots. Blue dashed line represents the selected irrigated plots while the black line represents the in situ irrigated plots. Shaded regions represent the standard deviation of the average NDVI for selected samples for both irrigated (blue) and non-irrigated (red). Number of in situ plots is referred to Table 1 whereas the number of selected plots is referred to Table 2 for the irrigated and non-irrigated classes each year.

Figures 8a, 8b, 8c and 8d of the four years show that the selected irrigated plots have the same behavior of the NDVI as the in situ irrigated plots. The NDVI starts to increase between April and May with approximately the same increasing gradient for both in situ and selected

irrigated plots. The NDVI in both datasets (selected and in situ) then reach a maximum high NDVI value (approximately 0.85) between July and August. Similarly, the selected non-irrigated plots show the same NDVI pattern as the in situ non-irrigated plots for the years 2020, 2019 and 2018. The NDVI values increase between April and May to reach a maximum value of approximately 0.6 in July and August. Only in 2017 (Figure 8d), the NDVI temporal profile of the in situ non-irrigated plots looks far from the selected non-irrigated plots. In fact, in 2017, both irrigated and non-irrigated classes of the terrain in situ data show similar temporal profile with only small differences. In general, except for 2017, the selection criteria of by the S^2IM produced a training dataset of irrigated and non-irrigated plots that are nearly similar in terms of NDVI profile to the dataset of irrigated and non-irrigated plots collected through a terrain campaign.

4.4 Random Forests classification results

In this section, we present the results obtained by the S^2IM for each RF classifier built at each year using the selected training data of our proposed methodology. Moreover, for each year, we compare the obtained S^2IM results with RF 5-folds cross validation developed using the in situ terrain data (RF in situ). Table 3 summarizes the accuracy metrics obtained when applying the RF classifier, trained with selected training data of each year, on the in situ validation dataset of the same year.

The validation of the RF classifier built from selected training data in 2020 generally shows very good accuracy (Table 3). The overall accuracy (OA) obtained using the S^2IM reaches 84.3% with a similar F_score value (84.1%). The irrigated class seems to have higher accuracy (86.4%) than the non-irrigated class (81.3%). On the other hand, the RF 5-folds cross validation (RF in situ) shows slightly higher accuracy values than the RF S^2IM for the four accuracy metrics (Table 3). In terms of OA, F_score, and irrigated class F_score (F_score_irr) the 5-folds cross validation is approximately 5% higher than the proposed S^2IM. Higher difference of 7% between the 5-folds cross validation and the S^2IM is observed for the non-irrigated class.

In 2019, the results show that an optimum accuracy is obtained for the irrigation mapping (Table 3). Indeed, the validation of the S^2IM with in situ data shows that the four accuracy metrics attain high values (between 92.5% and 93%). In addition, the RF 5-folds cross validation built directly from in situ data shows approximately the same accuracy values as that obtained using the S^2IM for the four accuracy metrics.

Using the selected training dataset in the S^2IM, an overall accuracy of 81.8% is obtained for the year 2018 with a weighted F_score of 82.2% (Table 3). While the irrigated class show good accuracy (F_score_irr = 86.8%), the non-irrigated class shows lower accuracy than the irrigated class (70.0%). However, in the 5-folds cross validation using in situ data, similar results are obtained. The F_score_nirr of the non-irrigated class (73.6%) is lower than that of the irrigated class (92.0%) while the overall accuracy reaches 88% (6.2% more than that obtained with the S^2IM). In the RF 5-folds cross validation, the weighted F_score (86.9%) is also slightly higher than that obtained using the S^2IM (82.2%).

When validating the S^2IM using in situ data in 2017, an overall accuracy of 72.8% is obtained for irrigation mapping with an F_score value of 74.0% (Table 3). However, the accuracy between the two classes is different. While the irrigated class shows good accuracy (F_score 78.1%), the non-irrigated class show a moderate accuracy reaching 62% only. This trend is also present in the RF 5-folds cross validation where the irrigated class attains an accuracy of 85.7% greater than that of the non-irrigated class (53.7%). In general, the RF 5-folds cross validation has marginally higher accuracy than the proposed S^2IM. Although, the overall accuracy of the RF 5-folds cross validation (78.3%) is slightly higher than that of the S^2IM, the S^2IM gives higher accuracy for the non-irrigated class (53.7% vs 62%).

Table 3: Accuracy metrics of RF classifications obtained using the proposed S^2IM methodology and the in situ 5-folds cross validation for the four years

Year	Method	OA	F_score	F_score_irr	F_score_nirr
2020	RF S^2IM	84.3%	84.1%	86.4%	81.3%
	RF in situ	89.0%	87.5%	90.2%	88.1%
2019	RF S^2IM	93.0%	92.8%	93.0%	92.5%
	RF in situ	91.3%	91.3%	91.2%	91.3%
2018	RF S^2IM	81.8%	82.2%	86.8%	70.0%
	RF in situ	88.0%	86.9%	92.0%	73.6%
2017	RF S^2IM	72.8%	74.0%	78.1%	62.0%
	RF in situ	78.3%	76.5%	85.7%	53.7%

4.5 Method Generalization

The effectiveness of the S^2IM mainly resides in the ability of generating a training dataset each year. The training data generation helps obtaining an irrigation map each year even with the absence of terrain campaigns for in situ data collection. On the other hand, it is known that using machine-learning algorithms such as the RF, the transfer of the model from one year to another remains difficult due to the variable response obtained using SAR and optical data between different years. However, to explore the difficulty of transferring the RF model from

one year to another, we conducted an experiment to test the ability to transfer the RF classifier for mapping irrigated areas from one year to the other years. In this experiment, we built a RF model at each year using its own in situ data with S1 and optical data, and we applied it on the three other years to obtain the classification accuracies. For example, a model was built using in situ data of 2017 (considered for training) and applied over the in situ data of 2018, 2019, and 2020 (considered for validation). Table 4 summarizes the weighted F_score obtained when applying an in situ built RF model of one year on the other years. Among different scenarios of training and validation for the four years, the maximum accuracy of the transfer does not exceed 68.6%. All the models trained on one year and applied on the three other years present low accuracy for mapping irrigated areas. Indeed, the F_score value ranges between 51.5% and 68.6% only. The results thus confirm the difficulty of mapping irrigated areas using only one-year in situ data and applying over several years. Therefore, as the irrigation mapping using spatiotemporal machine learning transfer is not yet achieved with high accuracy, the need of a training dataset for each year remains important. However, since the terrain campaigns are still time and resource consuming, the automatic reference data generation of the S^2IM offers a powerful tool to achieve irrigation mapping at very good accuracy without the need of yearly terrain campaign.

Table 4: Accuracy metrics derived from training a RF classifier at a year and applying on the other years.

		Training			
		2017	2018	2019	2020
Validation	2017		61.3%	65.1%	65.8%
	2018	68.6%		54.2%	51.5%
	2019	67.1%	53.4%		67.4%
	2020	62.2%	61.7%	60.9%	

4.6 Thresholds Sensitivity Analysis

For the operational use of the S^2IM method, it is important to discuss the threshold values of the irrigation metrics fixed to select training datasets (irrigated and non-irrigated classes). In order to show the flexibility of the thresholds considered, we conducted a sensitivity analysis to test the effect of changing the threshold values of the $cumul_{ipw}$ metric on the accuracy of the classification. For this experiment, we chose two years, one representing a humid year (2017) and the other representing a dry year (2019). The selection of the training dataset (step 1 of Figure 3) was re-performed using different threshold values for $cumul_{ipw}$ for both irrigated and non-irrigated classes. In fact, instead of considering $cumul_{ipw} \leq 25$ for non-irrigated class

and $cumul_{ipw} \geq 250$ for irrigated class we considered new values in three different tests (Table 5). In the first test, a threshold less than or equal to 50 is considered for the $cumul_{ipw}$ value of non-irrigated class and greater than or equal to 225 for the irrigated class. In the second test, we considered a value of $cumul_{ipw} \leq 75$ for the non-irrigated class and $cumul_{ipw} \geq 200$ for the irrigated class. Finally, we considered the threshold values $cumul_{ipw} \leq 100$ and $cumul_{ipw} \geq 175$ for non-irrigated and irrigated classes respectively. Table 5 presents the weighted F_score obtained for both 2017 and 2019 when applying the S^2IM in the three tested thresholds compared with the initial thresholds (≤25 and ≥250). For the dry year 2019, the F_score value remains nearly constant with the change of the threshold values. The F_score decreased only 1% when the threshold values changes to $cumul_{ipw} \leq 100$ for non-irrigated and $cumul_{ipw} \geq 175$ for the irrigated class. In contrast, the F_score of 2017 decreases by 10% as the threshold values of the two classes becomes closer (≤100 and ≥175 for non-irrigated and irrigated classes respectively). However, for the threshold values ≤50 and ≥225, for non-irrigated and irrigated class respectively, both years showed the same accuracy as the initial thresholds considered in this study.

Table 5: Variation of obtained weighted F_score as a function of the threshold values of $cumul_{ipw}$ for 2017 and 2019.

Threshold Test	Non-Irrigated Threshold ≤	Irrigated Threshold ≥	F_score 2019	F_score 2017
Initial	25	250	0.93	0.74
Test 1	50	225	0.93	0.74
Test 2	75	200	0.92	0.66
Test 3	100	175	0.92	0.66

The results could be analyzed using the histograms of the distribution of $cumul_{ipw}$ for in situ data presented in Figure 7 of Section 4.2.2. For the dry year of 2019 (Figure 7b), the distribution of the $cumul_{ipw}$ for the in situ irrigated class is distinguished from that of the non-irrigated class where the separability between both classes is highly present using the $cumul_{ipw}$. This distribution could be the same for the RPG data when selecting the training dataset. For this reason, when narrowing the difference between the irrigated and non-irrigated threshold values from (≤25, ≥250) to (≤100, ≥175) both classes remains separable and distinct and thus the classification accuracy remained constant. However, for a humid year as 2017, the $cumul_{ipw}$ distributions of the in situ irrigated and non-irrigated classes are closer to each other than 2019 (Figure 7d). Thus narrowing the threshold window from (≤25, ≥250) to (≤100, ≥175) increases the ambiguity between the two classes. When using the RPG data, the

consideration of the thresholds (≤100, ≥175) most probably reduced the separability of the two classes in the selected training dataset (as shown for the in situ). As a result, the classification accuracy decreases significantly.

5. Discussion

5.1 Classification accuracies and rainfall data

In order to understand the variable performance of the irrigation mapping classifiers between the four studied years it is important to discuss the limitations that can affect the distinction between irrigated and non-irrigated plots. Among several limitation in both radar and optical data for irrigation mapping, the most important factor that can affect the irrigation classification is the amount of rainfall received during the growth cycle of the crop. In fact, several studies have reported that irrigation classification in humid areas is more difficult than that in semi-arid and arid regions due to abundant rainfall events (Bazzi et al., 2020b; Pageot et al., 2020). Therefore, we analyzed the performance of the proposed S^2IM method for each year as a function of the cumulative rainfall data received each year. Figure 9 shows the accumulation of the daily precipitation record from 01 May (considered as starting point for the cumulative calculation) until 01 October for the four years derived from the IMERG daily rainfall maps presented in Section 2.6

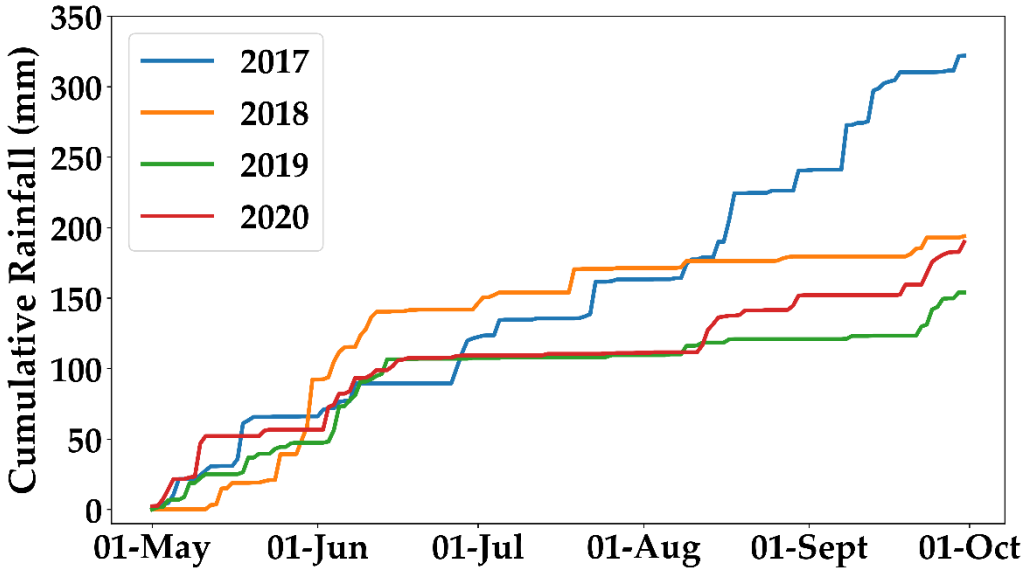


Figure 9: Daily cumulative rainfall for the period between May and October for 2020 (red), 2019 (green), 2018 (orange) and 2017 (blue) over the study site.

Between 01 May and 01 June, the study area received a cumulative rainfall of 66, 92, 50 and 56 mm in 2017, 2018, 2019, and 2020 respectively. In 2020, the cumulative rainfall from

May to mid-June reached 100 mm and then remains stable until the mid-August (approximately 2 months with no registered rainfall). In 2019, no rainfall was registered in the summer season for approximately 3 months between mid-June and late September where the cumulative rainfall remained stable at 100 mm during this period. On the other hand, in 2018, an important rainfall was registered at the beginning of June, causing the cumulative rainfall to reach 150 mm during the first two weeks of June 2018. From mid-June until the beginning of October, the study area received 60 mm of rainfall distributed over the summer season to reach a cumulative rainfall of 210 mm in the end of September. In 2017, rainfall events occurred during the whole summer season. After a stable cumulative rainfall in June (100 mm) for 2017, successive rainfall events occurred between July and August, which increased the cumulative rainfall to 150 mm by the beginning of August. In August, important rainfall events occurred causing an increase in the cumulative rainfall between the first week of August until the beginning of October. In August and September 2017, the area received a cumulative rainfall of 158 mm. Thus in 2017, a cumulative rainfall of 320 mm was registered for the period between May and October 2017 with continuous rainfall events during the whole summer season. Finally, we conclude that both 2019 and 2020 were the most dry years (2019 the driest) with a cumulative rainfall during the irrigation period (summer season) reaching 150 and 180 mm respectively. The year 2018 had moderate cumulative rainfall of 210 mm but with very important rainfall event occurring with the beginning of the crop cycle (end of May and beginning of June 2018). The most humid year was 2017 with continuous rainfall events during the whole irrigation period and the highest cumulative rainfall of 321 mm.

Concerning the irrigated/non-irrigated classification accuracy, we notice that the years 2019 and 2020 attain the highest overall accuracies of 93.0% and 84.3% respectively among the four years accompanied with the least amount of cumulative rainfall (150 and 180 mm respectively). In 2018, the overall accuracy reaches 81.8% with moderate amount of cumulative rainfall (210 mm). However, the discrimination between the irrigated and non-irrigated class is less accurate in 2018 than that of 2019 and 2020. In fact, the F_score_nirr (F_score of non-irrigated class) reaches 70% for the year 2018 compared to 92.5% and 81.3% for the years 2019 and 2020 respectively. This means that the discrimination of non-irrigated plots from irrigated plots is harder in 2018 than that in 2019 and 2020 due to higher rainfall amounts received during the summer season. In 2017, the overall accuracy is the lowest among the four years whereas the cumulative rainfall is the highest. With 320 mm of rainfall during the irrigation period of 2017, the overall accuracy reached 72.8% only, which is the lowest accuracy compared to the other

three years. Moreover, the F_score_nirr value attains in 2017 its lowest value of 62.0% among the four years. This indicates that in a humid year (2017) with abundant and continuous rainfall events during the summer crop-growing season, which corresponds to the irrigation season, the discrimination between irrigated and rain-fed plots is more complicated. Therefore, when less rainfall events and cumulative rainfall are registered during the irrigation period, the classification accuracy increases and the discrimination of irrigated and non-irrigated plots becomes easier.

The effect of rainfall on discriminating irrigated and non-irrigated plots appears on both radar and optical data. Using optical data, frequent rainfall events expose a similar vegetation index profile for both irrigated and non-irrigated plots. This is mainly due to the fact that with abundant rainfall, the non-irrigated plots are also benefiting from a sufficient amount of water capable of giving a well-developed canopy cover. This similarity in the NDVI between irrigated and non-irrigated plots due to high frequency of rainfall events is clearly visible in the in situ data for the year 2017. In Figure 6d of Section 4.2.1, the distribution of the maximum NDVI values (max_{NDVI}) of irrigated and rain fed plots shows that both classes from in situ have nearly the same distribution with a similar average value of 0.81 and 0.77 respectively. On the other hand, in both 2019 and 2020 with the least rainfall amounts, the distribution of the max_{NDVI} value showed that the separability between the irrigated and non-irrigated class of in situ data is high. In Figure 8d (Section 4.3), the NDVI temporal profile of the non-irrigated class (in situ) showed same behavior as that of the irrigated profile, which indicates that both classes had similar vegetation development. Therefore, more difficulty to separate both classes is present in 2017. In contrast, in the dry years (2019 and 2020) the temporal profile of the NDVI for non-irrigated plots was easily distinguished from the NDVI temporal profile of irrigated plots (in situ data). This indicates that the addition of water for irrigated plot in dry year induces a significant difference in the NDVI temporal profile when compared to non-irrigated plot.

Using radar data, frequent rainfall events can affect the detection of irrigated plots. In fact, if a rainfall event and an irrigation event occurred between two consecutive S1 images, the irrigation event will be difficult to detect. This is mainly due to the incapability to distinguish whether the increase of soil moisture is due to irrigation or rainfall since both have the same effect on the SSM. Therefore, with frequent rainfall events, the capability to detect irrigation events decreases. This fact has been demonstrated in Bazzi et al. (2020c) where they reported that frequent rainfall events in the spring season limited the detection of irrigation events over grassland plots. In Figure 7c and Figure 7d (Section 4.2.2), the histogram of the distribution of

the $cumul_{ipw}$ metric for 2017 and 2018 using the in situ data showed that both irrigated and non-irrigated plots have close distribution. This is due to the fact that over the irrigated plots in a humid season with frequent rainfall events, a limited number of irrigation events could be detected and thus the $cumul_{ipw}$ metric will not attain high values due to low numbers of detected irrigation events. In contrast, years 2019 and 2020 showed higher separability in the in situ data between irrigated and non-irrigated plots using the $cumul_{ipw}$ metric. In dry conditions with low number of rainfall events, the irrigation frequency at the plot increases and the capability to detect these irrigation events also increases with the absence of rainfall. For this reason, most of the irrigated plots in 2019 and 2020 encountered high $cumul_{ipw}$ values, which were separable from the non-irrigated plots.

5.2 Limitations of S^2IM

5.2.1 Threshold values and reference data selection

In this study, an innovative approach is proposed to map irrigated areas at plot scale (S^2IM). To overcome the limitation of terrain data availability, the power of the proposed method resides in its ability to generate automatically its own reference data. The generated reference data are then used in a RF classifier to map irrigated areas at plot scale. However, the selection of the training reference data is based on two metrics (SAR and optical) with threshold values to deem whether a plot is irrigated or not. The first metric is related to the number of detected irrigation events computed from the newly derived IEDM. However, the irrigation detection using IEDM presents some limitations. The IEDM is based on the detection of soil moisture change using the S1 C-band SAR data. Nevertheless, the detection of soil moisture change (therefore irrigation) using the S1 C-band SAR data could be limited to two main factors. First, the time lag between the irrigation time and the S1 acquisition time plays an important role in irrigation detection. When the S1 acquisition is acquired long time after the irrigation event (3 to 4 days), the detection of the irrigation event becomes difficult. This is mainly due to the evaporation (especially in summer), which causes soil moisture values to decrease 3 or 4 days after irrigation. This limitation has been discussed in both (Hajj et al., 2014) and Bazzi et al. (2020c). In the study of Hajj et al. (2014), they showed that using the X-band SAR data, a maximum of 3 days old irrigation could be detected. In the assessment of the IEDM by Bazzi et al. (2020c), they showed that for low vegetation cover (NDVI < 0.7), the irrigation event could be detected until two to three days after the irrigation event using S1 C-band data. For

NDVI > 0.7, they showed that irrigation event could be detected if it occurs on the same day of the S1 acquisition. Therefore, the time interval between the S1 acquisition time and the irrigation time can constraint the detection of irrigation events.

The second important factor that limits the detection of irrigation events by C-band SAR data is the penetration of the C-band SAR signal in developed vegetation cover. When the vegetation cover is well developed (NDVI > 0.7), the soil contribution to the backscattered SAR signal in C-band decreases. This limitation has been demonstrated by several studies (Bazzi et al., 2019a, 2020c; El Hajj et al., 2017, 2018b; Joseph et al., 2010; Nasrallah et al., 2019). In a study performed by El Hajj et al. (2018b), they compared between C and L bands penetration over wheat and maize. They showed that the C-band in VV polarization is able to penetrate the maize canopy even when the canopy is well developed (NDVI > 0.7) due to high-order scattering along the soil-vegetation pathway that contains a soil contribution. (Joseph et al., 2010) also showed that surface soil moisture in maize plot could still contribute to the C-band SAR backscattering signal even at maximum biomass stage. On the other hand, El Hajj et al. (2018b) showed that for wheat crops, the sensitivity of the C-band SAR signal to soil moisture estimations is negligible for NDVI > 0.7. For grassland, Bazzi et al. (Bazzi et al., 2020c) showed that for some grass types, the high vegetation canopy (NDVI > 0.7) attenuates the SAR backscattering signal (no soil contribution) and therefore makes the detection of the irrigation event difficult. Thus, the well-developed vegetation cover reduces the opportunity of detecting part of the irrigation events on the plot. This may lead to less number of detected irrigation events on the plot. Nevertheless, for less dense vegetation cover, the radar signal in C-band has the necessary penetration to detect soil moisture change. Thus, the irrigation events occurring from the sowing date until the stage before the vegetation is very well developed, could still be detected using the C-band SAR signal due to the existence of soil contribution in the backscattered signal.

Nowadays, the S1 satellite in C-band is the only operational radar satellite providing free and continuous data acquisitions at high spatial and temporal resolutions. It is well known that L-band SAR data can penetrate more the vegetation canopy (El Hajj et al., 2018b), but the current L-band satellites such as ALOS-2 does not provide continuous (high revisit time) and free acquisitions. The arrival of new L-band SAR satellites (some are planned in 2022) could help apply the IEDM using L-band data and acquire accurate detection of irrigation events with less uncertainty caused by the vegetation cover.

The second metric used for reference data generation is based on the maximum NDVI value during the crop cycle. For this metric, two thresholds were considered. The first considers that reference non-irrigated plots should not attain a maximum NDVI more than 0.7. The second states that reference irrigated plots must have a maximum NDVI greater than 0.8. The thresholds on the max_{NDVI} metric were derived through the analysis of this metric using in situ data and supported by previous studies using the max_{NDVI} for discriminating between irrigated and rain-fed crops. It is important to note that the criterion on NDVI thresholds is not exclusive and the thresholds allow only the selection of irrigated/non-irrigated plots. In fact, the RF can later detect irrigated plots with NDVI < 0.8 if the radar signal shows significant increases in σ_p^0 identified by the IDEM as potential irrigation events.

However, for other terrains having irrigated plots corresponding to NDVI lower than 0.6 (vegetables plots for example), a small number of irrigated plots may not be detected by the RF as irrigated while the IEDM may show many irrigation events. For this reason, once the classification is completed, a filter could be applied on the non-irrigated plots, which consists in transforming a non-irrigated plot into an irrigated plot if the IEDM detects many irrigation events with high certainty. On our database in Orlèans, the improvement of the mapping accuracy after applying this filter is not more than 1%. This improvement could be higher in another territory having irrigated crop types with low NDVI values (0.5-0.7)

For some other crop types, the max_{NDVI} of both irrigated and non-irrigated plots could be different from the proposed thresholds which may lead to uncertain accuracy. Generally, a priori information about the crop type in the studied area could help adjust the threshold values of the max_{NDVI} metric. The addition of crop type map could help better adjust the threshold values for each crop class and help distinguish irrigation/rain-fed for each crop class. This threshold value adjustment could be done to enhance the selection of the training data and thus ameliorate the classification accuracy. The crop type map could be either provided by local authorities or obtained by classifying crop types using S2 and/or S1 data. Nonetheless, crop type maps are not always available. In this case, the proposed general threshold that accounts to the most common summer crops could serve as general thresholds for reference data selection.

5.2.2 Irrigation mapping in humid and dry areas

Over our study site, four different years were examined. One of the years (2017) was characterized by very humid summer while another (2019) was characterized by very dry summer.

As shown in the results, using either our approach (the S^2IM) or a RF classifier directly performed using in situ data; the irrigation mapping in humid conditions is less accurate than that in dry conditions. Using either in situ RF or the S^2IM, the irrigation classification accuracy in 2017 was between 73% and 78%. The decrease in the classifier performance is mainly due to the minimized difference between irrigated and non-irrigated plots in both SAR and optical (NDVI) data for humid conditions. In fact, this is considered as one of the limitations for using remote sensing data (nowadays S1 and S2 data) in irrigation mapping in humid areas. Regardless of the proposed methodology, when abundant rainfall events occur during the irrigation period, the differences in NDVI between irrigated and non-irrigated cropland becomes negligible. Adequate moisture from precipitation available to non-irrigated crops can increase the NDVI value, potentially narrowing the difference in NDVI between irrigated and non-irrigated crops. This will make the separation of irrigated and non-irrigated crops difficult. Using SAR data, abundant rainfall events decreases the chance of detecting irrigation events. When irrigation and rainfall occurs between the same consecutive SAR acquisitions, it is difficult to distinguish between irrigation and rainfall. However, the results show acceptable accuracy in 2017 for irrigation mapping.

Several studies mapping irrigated areas have reported the same common limitation in humid conditions (Bousbih et al., 2018; Brown and Pervez, 2014; Demarez et al., 2019; Pageot et al., 2020). Recently, Pageot et al. (Pageot et al., 2020) tried to map irrigated summer crops in a humid area in southwestern France (Adour Amont watershed). Despite of using climatic data (precipitation) in addition to S1 and S2 data in the RF classifier, the overall accuracy for irrigation mapping did not exceed 78%. They showed that the difference between the vegetation indices NDVI and NDWI (Normalized Difference Water Index) was narrow due to frequent rainfall events occurring during the growing season. This is the same case as our study site in the year 2017 were the region received 320 mm of rainfall during the irrigation period.

Finally, mapping irrigated areas is more important in dry areas, which suffer from water scarcity and altering rainfall amounts. In humid climates, irrigation is a supplementary water applied usually to meet the additional crop water demand especially for crops that may require

water more than that offered with natural precipitation. In arid and semi-arid climates, continuous irrigation is usually required to assure agricultural production (Ozdogan et al., 2010).

6. Conclusion

In this study, an operational methodology for mapping irrigated areas at plot scale (S^2IM) has been proposed. To address the main issue related to the dependency of supervised classification models on in situ terrain campaigns for irrigation mapping, the methodology presented in this study is capable of automatically generating reference dataset of irrigated and non-irrigated plots to be used in a supervised classification model. The reference data selection was based on two metrics derived from S1 and S2 temporal series. The RF classifier was then used to map irrigated areas using the selected reference dataset, S1 and S2 data. The method was applied on a study site located in northcentral France for four years between 2017 and 2020.

The proposed methodology "S^2IM" delivered reasonable performance with overall accuracy between 93.0% and 72.8% depending on the climatic conditions. Dry years with slight rainfall events in the irrigation period had significant separability between both classes with a mapping accuracy reaching 93%. Humid years with frequent rainfall in summer irrigation period had low separability between both classes in SAR and NDVI data revealing moderate accuracy in irrigation mapping. The comparison of the S^2IM with traditional RF developed using in situ terrain data revealed that the proposed S^2IM performs well with accuracy nearly similar to that obtained using in situ RF. However, generating yearly basis reference data using the S^2IM showed widely better classification results than using one in situ based RF model built on a year and applied on others.

With the absence of terrain data to perform irrigated area maps, the strength of the S^2IM is the ability of generating yearly reference data. Thus, the proposed mapping approach is temporally transferable to other years, which can insure continuous monitoring of irrigated areas even in the absence of terrain data. The method could be also spatially transferred to other areas sharing similar climate, cropping landscapes and crop management. Areas with different climate and cropping land cover may only require some adaptation of the reference data selection thresholds before further application.

Author Contributions: Conceptualization, Hassan Bazzi and Nicolas Baghdadi; Data curation, Hassan Bazzi, Ghaith Amin and Ibrahim Fayad; Formal analysis, Hassan Bazzi, Nicolas Baghdadi and Ibrahim Fayad; Methodology, Hassan Bazzi, Nicolas Baghdadi and Ibrahim Fayad; Software, Hassan Bazzi; Supervision, Nicolas Baghdadi; Validation, Hassan Bazzi, Mehrez Zribi, Valérie Demarez and Hatem Belhouchette; Writing – original draft, Hassan Bazzi; Writing – review & editing, Nicolas Baghdadi, Ghaith Amin, Ibrahim Fayad, Mehrez Zribi, Valérie Demarez and Hatem Belhouchette.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2021 project), the National Research Institute for Agriculture, Food and the Environment (INRAE), the Occitanie region of France and the Mediterranean Agronomic Institute of Montpellier (CIHEAM-IAMM).

Acknowledgments: Authors wish to thank the French Space Study Center (CNES, TOSCA 2021), the National Research Institute for Agriculture, Food and Environment (INRAE) and the Mediterranean Agronomic Institute of Montpellier (CIHEAM, IAMM). The authors wish to thank the Occitanie region of France for supporting this work. The authors wish also to thank the European Space Agency (ESA) for the Sentinel 1 and Sentinel 2 data and Theia pole for the calibration of the Sentinel-2 images. Finally, authors would like to thank the Regional Directorate of Food, Agriculture and Forest of (D.R.A.A.F) of the “Centre Val de Loire” and particularly Gérard Guillaume (DRAAF Centre Val de Loire/SSI), Audrey Oddos (DRAAF Centre Val de Loire/SRISE) and Claudie Suzanne (DRAAF Centre Val de Loire/SSI) for the collection of field data and their expertise.

Conflicts of Interest: The authors declare no conflict of interest.

CHAPTER V: GENERAL CONCLUSIONS AND PERSPECTIVES

1. Research synthesis

Accurate information on the spatial extent of irrigated areas is fundamental in several aspects of water and agricultural management. In this thesis, two complementary objectives, including irrigation mapping and irrigation events detection using radar and optical remote sensing data, were investigated. Nowadays, Sentinel-1 satellite is the only operational SAR satellite that provides continuous acquisitions at six days revisit time and 10 m spatial resolution in C-band. Sentinel-2 satellite also provides exceptional revisit time over Europe (5 days) and 10 m spatial resolution (for some wavelengths). The temporal resolution of both S1 and S2 satellites allows obtaining temporal series information with suitable temporal interval (less than a week) permitting accurate spatio-temporal follow of the agricultural fields during the season.

Concerning the thesis problematic presented at the beginning of this manuscript, the performed work replied to all the posed questions. Primary, the thesis studied the potential of Sentinel-1 radar data and Sentinel-2 optical data combined with the S^2MP soil moisture estimation (at plot scale) for mapping irrigated areas. Before integrating the S^2MP product into irrigation mapping, it was first validated across in situ SSM estimations, and compared to other soil moisture product (C-SSM over 1 km x 1 km) and rainfall data (GPM). The results showed that the S^2MP provides good SSM estimations, highly correlated with rainfall events. To map irrigated areas at plot scale, several techniques were tested, including supervised classification models, transfer-learning approaches, and semi-supervised classification models. First, the results showed that the use of Sentinel-1, Sentinel-2, or coupling Sentinel-1 and Sentinel-2 data permits accurate mapping of irrigated plots mainly using supervised classification models. The study conducted over the semi-arid region of Catalonia showed significant accuracy for mapping irrigated area using either classical machine learning models (such as the random forest – 92.3%) or deep learning models (such as the CNN – 94.1%) with S1 and/or S2 data. However, it was proved through literature that irrigation mapping models at regional and global scales using remote sensing has not been entirely operational. Models that work in one place and time are not necessarily transferable to other locations and periods. For this reason, the thesis focused on proposing a transfer learning procedure capable of transferring the irrigation

model from one region to another. As a result, a transfer-learning framework was implemented to transfer the irrigation mapping model from a semi-arid area (Catalonia) to a humid area (Adour-Amont watershed). The model built on Catalonia was re-calibrated using some reference data in Adour Amont by applying the transfer technique "Distilling before refine". Using the new proposed transfer learning technique, the results demonstrated that a model built on a specific area could be transferred to another area in order to map irrigated plots. A reasonable accuracy of 83% has been obtained for irrigation mapping using the transfer learning technique. On the other hand, applying the Catalonia model in Adour Amont directly without any re-calibration produced very low accuracy of 27.5%. With the transfer learning approach, this accuracy was widely enhanced to reach 83%. Nevertheless, the transfer technique requires some in situ data on the target study site to re-calibrate the model.

The third objective in this thesis is the detection of irrigation events at the plot scale. In this context, the thesis evaluated the potential of using the C-band SAR data (Sentinel-1 data) to detect irrigation events at the plot scale. The Sentinel-1 (S1) data was used conjointly with soil moisture estimations from S^2MP and optical S2 data to detect the existence or absence of irrigation event at each available S1 image. The thesis thus proposed a decision tree algorithm (IEDM) capable of detecting the irrigation events at the plot scale. The algorithm mainly analyzes the change in SAR backscattering signal between two successive SAR acquisitions. The significant increase of the SAR signal at plot scale between two consecutive dates is mainly attributed to increase in soil moisture values. Then, using the grid scale SSM values from the S^2MP and the SAR signal at grid scale, irrigation events are distinguished from rainfall events.

The results demonstrated that irrigation events could be detected with an accuracy reaching 75%. Two main limitations were found in the detection of irrigation events using S1 C-band SAR data. These limitations include the time lag between the irrigation event and SAR image date in addition to the penetration of the C-band SAR data in very well developed vegetation cover. Both limitations will be deeply discussed in the following limitation section.

The most critical obstacle toward the operational mapping of irrigated areas across several regions is the reference data availability. Indeed, machine-learning algorithms, including classical or advanced approaches, still require training data to obtain good classification results. Therefore, any new approach for irrigation mapping should be independent of terrain data in order to be operational and applicable in several areas and over different years. In this context, the thesis tried to resolve the terrain data availability required to perform irrigation mapping.

To get over the availability in situ collected reference data, we proposed an innovative semi-supervised classification approach for irrigated area mapping capable of generating its reference data, which is used next in a machine-learning model. The model is based on the use of the IEDM and optical NDVI data to generate reference dataset of irrigated/non-irrigated plots. Then, irrigation mapping was performed using the automatically generated reference data in a supervised classification (RF) using S1 and S2 time series. The proposed approach has been validated over an irrigated basin near Orléans city. The results showed good accuracy in irrigation mapping for four different years where the accuracy varied between 72% and 93% depending on the climatic conditions of the year. Dry years showed higher classification accuracy than humid years.

The use of remote sensing techniques (mainly here S1 and S2) for irrigation mapping and irrigation events detection presents some limitations in producing high accuracy maps and event's detection. In this conclusion, a set of limitations which constraints the use of S1 and S2 data in irrigation monitoring (mapping and events detection) are highlighted. Then, future perspectives are presented.

2. Limitation

2.1 Climatic limitation

Throughout this thesis, a set of limits for irrigation mapping using remote sensing has emerged. The first limitation to be discussed is the studied climatic region. Many studies that used multi-spectral and multi-temporal remote sensing images for irrigation mapping were performed mainly over arid and semi-arid regions (Ambika et al., 2016; Cheema and Bastiaanssen, 2010; Dheeravath et al., 2010; Gumma et al., 2011). Over humid areas, mapping irrigated areas is more difficult and faces more challenges. The limitation of remote sensing in obtaining high precision irrigation mapping in humid area is present on both optical and radar data. Using optical data, the main assumption to map irrigated areas is the difference in the spectral response (thus vegetation indices) between irrigated and rain-fed areas. Several studies have demonstrated that vegetation indices such as the NDVI, GI (Green Index) and LAI (Leaf Area Index) are positively correlated with the available soil moisture of the vegetation (Aparicio et al., 2000; Wardlow and Egbert, 2008). However, climatology plays an important role in this assumption. When the studied area is humid, it encounters several rainfall events during the irrigation period. Thus, sufficient soil moisture caused by frequent rainfall makes NDVI values

of non-irrigated plots similar to NDVI values of irrigated plots. This change in the NDVI of non-irrigated crop can potentially narrow the difference in NDVI between irrigated and non-irrigated crops. When the difference between the satellite information for irrigated and non-irrigated plots becomes lesser, it becomes difficult to separate accurately both classes. In contrast, low soil moisture from time to time due to less rainfall events will affect the NDVI values and widen the difference between NDVI from irrigated and non-irrigated crops making them more separable. The narrow and large difference between irrigated and non-irrigated crops in humid and dry areas respectively is the same also over other vegetation indices such as the NDWI (Normalized Difference Water Index) and the GI.

In C-band SAR data, the difference in the SAR temporal series between irrigated and non-irrigated plots also decrease when abundant rainfall events are present in the area because soil moisture will be quite frequently similar on irrigated and non-irrigated plots. The high number of rainfall events can decrease the number of detectable irrigation events. When both an irrigation event and rainfall event occur between two S1 images at time t_i and t_{i-1}, it is difficult to distinguish between the rainfall and irrigation.

The limitation of optical and radar remote sensing for irrigation mapping in humid areas has been shown in some recent studies. Recently, Pageot et al. (2020) tried to map irrigated summer crops in a humid area (Adour Amont watershed). Despite of using climatic data (precipitation) in addition to S1 and S2 data, the overall accuracy for irrigation mapping did not exceed 78%. They showed that the difference between irrigated and rain-fed crops in the NDVI and NDWI was narrow due to frequent rainfall events occurring in 2018 (humid year) during the irrigation season. In my thesis, we also encountered this same limitation when applying the proposed semi-supervised approach in Orléans. For the humid year with the highest rainfall amounts in the summer (320 mm for summer 2017), the overall accuracy for irrigation mapping reached 72%. On the other hand, the driest year (2019) with less rainfall amounts in the summer season showed the best accuracy (93%). Nevertheless, the obtained accuracies even in humid conditions are not bad (between 70% and 80%) and could be still used for irrigation water management. Indeed, irrigation and water consumption in humid areas is less than that in arid and semi-arid areas due to higher temperatures and dries seasons in the latter. In humid climates, irrigation often takes place as a supplementary water supply to meet the excess demand of crops whose growth cycle may require water more than that offered with natural precipitation. In arid and semi-arid climates, continual irrigation is often necessary to assure agricultural production.

In this thesis, only spring-summer crops were studied. Thus, it would be interesting to validate the proposed models on winter crops (mainly wheat). For winter crops, distinguishing between irrigated and rain-fed crops could be also difficult. In some regions across the world, deficit irrigation is exceptionally applied for some winter crops (such as wheat in April and May). In such cases, the detection of irrigated and rain-fed crops could be difficult first due to the abundant rainfall events in winter, second due to the very limited number of irrigation events occurring on the plot and third due to the penetration limitation of the C-band signal in very well developed vegetation cover (discussed in the following section). Bousbih et al. (2018) tried to map irrigated winter wheat using S1 estimated soil moisture and S2 data. They reported a poor accuracy for winter wheat irrigation mapping using NDVI (58.1%) and moderate accuracy using S1 estimated soil moisture (71.8%). In humid area performing irrigation maps less relevant than dry and semi-arid areas since irrigation is limited. Mapping irrigated areas is more important in dry areas, which suffer from water scarcity and altering rainfall amounts. In humid climates, irrigation is a supplementary water applied usually to meet the additional crop water demand especially for crops that may require water more than that offered with natural precipitation. In arid and semi-arid climates, continuous irrigation is usually required to assure agricultural production. Therefore, the proposed methods could be to mobilize for the dry periods of the year which is consistent with the studies carried out in arid or semi-arid zones.

2.2 S1 C-band SAR limitation for irrigation event detection and irrigation mapping

One of the main concern regarding the use of the C-band SAR signal in irrigation event detection is the penetration of the C-band signal in very well developed vegetation cover. Over bare soil or soil with low vegetation cover, the emitted C-band SAR signal (wavelength ~ 6 cm) will interact without difficulty with the surface layer of the soil (top 5 cm) and thus, the radar backscattered signal will be highly correlated to soil moisture. In this case, the detection of irrigation events based on the increase in soil moisture values between two dates could be easily obtained. When the canopy cover develops, the behavior of SAR signals and its penetration in the canopy becomes also dependent on the crop type and canopy growth stage (Ulaby, 1982). In fact, in the case of very well developed canopy cover, the emitted SAR signal in C-band may not reach the soil layer and thus the soil contribution in the backscattering signal becomes negligible. We say here that the S1 backscattering signal is insensitive to the soil moisture content and thus cannot provide information about the soil moisture value. In this case, the

change in the S1 backscattering signal between two S1 dates could not always represent the change in the soil moisture values. As a result, the detection of irrigation events based on the change is soil moisture values becomes difficult.

The penetration constraint of the C-band SAR signal into very well developed vegetation cover could also differ according to crop types and phenology stage. For wheat crops as well as grassland, the crop growth until an NDVI value of 0.7 (corresponding to approximately an LAI of 1.5 m^2/m^2) does not strongly attenuate the C-band SAR backscattered signal. Thus, irrigation events could still be detected. Beyond NDVI of 0.7, the sensitivity of backscattered signal becomes negligible to soil moisture content due to very high attenuation by the wheat canopy (similarly for some grassland crop types). The extreme canopy attenuation of the S1 signal in wheat crops and some grass types (coarse hay) corresponds to the heading phenology phase. This phenology phase causes the SAR backscattering signal to reach very low values mainly due to the very low soil contribution in the backscattering signal (He et al., 2014; Nasrallah et al., 2019; Srivastava et al., 2011). In our study performed over the irrigated grassland in PACA region (southeast France), we showed that for very well developed canopy cover (NDVI > 0.7) the detection of irrigation events was extremely hard during the first crop cycle where the canopy was rich in grasses such as coarse hay which are similar in geometric structure and type to wheat crops. This was mainly due to the attenuation of the C-band SAR signal during the heading phenology phases of grasses (hay).

For maize crop which is the most dominant irrigated crop in France, the penetration of the C-band SAR data into the canopy cover is different from the wheat or grassland. In two studies by El Hajj et al. (2018) and (Joseph et al., 2010), they showed that surface soil moisture in maize plot could still contribute to the C-band SAR backscattering signal even at maximum biomass stage. This soil moisture contribution even in developed maize vegetation cover is mainly explained by the significant soil-vegetation scattering pathway that includes soil moisture information. Thus, the scattering along the soil-vegetation pathway present in dense maize cover allows obtaining SSM estimations (Macelloni et al., 2001).

In conclusion, the penetration of the C-band SAR signal in developed vegetation cover remains the most important limitation for the detection of some irrigation events using S1 satellite. The irrigation events occurring from the sowing date until the stage before the vegetation is very well developed for some crop types, could still be detected using the C-band SAR signal due to the existence of soil contribution in the backscattered signal. Thus, the IEDM

proposed in this thesis would still be capable of detecting most of the irrigation events for a large period of the cropping cycle. Indeed, the S1 satellite working with C-band SAR data is the only operational satellite that provides continuous and free access acquisitions at high revisit time and high spatial resolution. Hence, the interest today is working exclusively on this radar frequency (C-band) while waiting for other satellites that could provide P or L band data with high revisit time as well as a high spatial resolution.

The second concern for using S1 satellite in detecting irrigation events is the revisit of the S1 images. The S1 temporal resolution is six days. On the other hand, irrigation is a time dynamic activity. Thus, monitoring and detecting irrigation events requires a fine temporal follow of the plot using SAR data. When the S1 image is acquired long time after the irrigation event, it becomes difficult to detect the change in the SSM values before and after the irrigation. This is mainly due to the evaporation of the soil water content especially in summer. Thus, the detection of the irrigation events is also constrained by the time lapse between the irrigation and the following SAR image. In our work, we showed using the in situ data of the grassland plots that irrigation could be detected two to three days after the irrigation event if the vegetation cover is low (NDVI < 0.7). If the vegetation is well developed (NDVI > 0.7), the irrigation event could be mainly detected if it occurs at the same date as the SAR acquisition.

In order to overcome this limits, we tried to use all the possible S1 acquisitions from all the S1 orbits to increase the chance of detecting an irrigation event. In this case, we obtain 20 different S1 images in one month from four S1 orbits instead of only five images from only one orbit. However, the IEDM was applied to each orbit temporal series separately and thus the detection of irrigation events remained at 6 days basis but with several images acquired on consecutive dates (4 images in 6 days). Both the incidence angle difference and the diurnal effect prevented us from combining the four temporal series together. While the effect of different incidence angles could be minimized by normalizing the S1 signal at a reference incidence angle, the diurnal effect on the S1 backscattering signal between day and night could not be precisely quantified. The diurnal variation is a result of the difference in the vegetation water content (VWC) between the morning and the evening. This difference in VWC causes high difference in the radar backscattering signal (for some dates higher than 1dB) over vegetated plots between the morning and the evening acquisitions. Several studies have reported that σ^0 in the morning overpass registers higher values than σ^0 in the evening overpass (Brisco et al., 1990; Van doninck et al., 2012). Therefore, combining a morning acquisition and

an evening acquisition to make the time interval less than 6 days can cause uncertain detection of irrigation events and lead to various false detections.

3. Perspectives

Several research perspectives arise from my thesis work. Some perspectives concern, on one hand methodology related perspectives capable of enhancing the proposed approaches and thus increasing the output confidence. On the other hand, other perspectives include wider scale applications, which would open the way toward real implementation of the proposed approaches for agricultural and water management.

3.1 L-band SAR data for irrigation event detection and irrigation mapping

In this thesis, the S1 carrying a C-band SAR was used principally for irrigation mapping and for detecting irrigation events. The use of S1 was mainly due to the unique continuous acquisitions at high spatial and temporal resolution offered in free open access mode. However, as explained previously in the limitations, the C-band SAR signal for irrigation mapping could encounter some limitations over very well developed vegetation cover in several crop types. This is mainly due to the low penetration of the C-band signal in developed vegetation cover. For this reason, a SAR signal with higher penetration capabilities is required. In fact, it has been demonstrated that the L-band SAR signal (wavelength ~ 24 cm) penetrates the vegetation cover well more than the C-band (wavelength ~6 cm) for several crop types (El Hajj et al., 2018b; Joseph et al., 2010).The vegetation attenuation in the L-band SAR signal is less than that of the C-band (Ulaby, 1982). Therefore, the change in soil moisture values due to irrigation could be better detected using L-band in the presence of well-developed canopy cover since the L-band (~24 cm) is characterized by a large penetration depth into the crop canopy. As stated in the state of art of the introduction, the satellite based soil moisture products such as SMOS and SMAP provide SSM estimations using the L-band signal. However, their low spatial resolution makes the estimated SSM useless for irrigation mapping at plot scale.

Studies focusing on the potential of the L-band SAR data from ALOS-2 or PALSAR-2 satellites for soil moisture estimations are few (Narvekar et al., 2015). This is mainly because most of the L-band satellites does not offer free and open access data with high revisit time thus making the spatio-temporal monitoring of soil moisture and irrigation difficult. However, even though current satellites provide low number of images, the integration of the L-band data with

C-band data can partially ameliorate the detection of irrigation events. When the C-band fails to detect the irrigation event due to the high canopy attenuation (example for wheat or grassland in heading phenology phase), the integration of only some L-band images (one or two images in one month) could help detect some of the missing irrigation events by C-band. Since both signals are at different frequencies and could not be joined together, the study could be performed based on SSM values estimated from C-band compared to SSM values estimated at some available L-band images. For example, high L-band SSM estimations in the period of very well developed vegetation cover (where C-band fails to detect irrigation) could indicate the presence of irrigation event when no rainfall is registered during the same period.

Several L-band satellites are to be launched soon such as the NISAR NASA-ISRO SAR (planned in 2022), the Tandem-L (German Aerospace Center, planned in 2022) and ALOS-4 (JAXA planned in 2022). Some of these missions (NISAR and Tandem-L) aim providing open access images at weekly basis temporal resolution. Therefore, concerning our work, the application of the developed IEDM for irrigation detection could be later performed using the expected new L-band images. In this case, accurate detection of irrigation events with higher confidence could be achieved mainly for well-developed vegetation. Thus, the L-band would open the way toward better mapping of irrigated areas because (1) L-band will allow better mapping of irrigation events if the vegetation is well developed and (2) the next L-band data combined with C-band data will give several radar acquisitions per week, which will allow detecting almost all the irrigation events. Thus, instead of using the C-band temporal series only in the RF classification, the L-band time series could be used in addition to the C-band time series for more accurate mapping.

3.2 Mapping irrigated fruit trees

This thesis concentrated on mapping irrigated areas at plot scale mainly for major crops (maize, sunflower, sorghum, vegetables, legumes…) and grassland. On the other hand, irrigated trees were not investigated. Mapping irrigated trees using satellite imagery is still an unexplored subject. Among fruit trees, irrigating vineyards is the most important matter in several regions in Europe and especially in the south of Europe. In vineyards, the term irrigation has a negative connotation and is often associated with high yields and therefore low quality wines.

For some time now, irrigating vineyards has become the major political and trade union issue in the South of France. Ensuring the profitability and the durability of the vineyard will

not be possible without water. Today, the Languedoc-Roussillon region is the main vine irrigated region in France with 23,000 hectares of vines that benefit from water, i.e. 10% of the region's vineyards. This surface is increasing under the effect of the successive severe droughts. Public authorities in France decided to change the legislation that prohibited the irrigation of vineyards. They governed vineyard irrigation with many restrictions to guarantee the proper vine yields. For example, irrigation of vines is possible between 1 May and 15 August and remains prohibited for all vines between 15 August and the harvest.

Therefore, building approaches capable of detecting irrigation events at vine plots and mapping irrigated vineyards is of great importance for public authorities and wine manufacturer to insure the high quality of wine. Irrigation mapping for vineyards presents many challenges. Notably the use of micro-irrigation and low applied doses in vineyards makes it difficult to detect irrigation by the approaches developed for irrigated crops by sprinklers or gravity irrigation techniques. Methods developed for soil moisture estimation or irrigation mapping (detection) at major crops (cereals and grassland) cannot be directly applied on vineyards because of the complexity of vineyard plots (presence of metal and wooden stakes, metal wires…). Thus, one of the future perspectives is to develop a method capable of estimating soil moisture and another method capable of detecting irrigation events at vine plots. For these objectives, two major challenges are to be investigated including the age of vines (young or old), presence of metal stakes and wires, inter-rows between trees, irrigation technique (drip irrigation, etc.)

3.3 Irrigation mapping using remote sensing and land surface models

Recent studies have exploited the joint use of land surface models (LSMs) and satellite soil moisture estimations to map irrigated areas. Using LSMs such as the SURFEX (developed by Meteo-France), soil moisture values could be simulated using several environmental and atmospheric variables without taking into consideration the irrigation. Then, satellite soil moisture estimations, which are eventually affected by irrigation, are compared to simulated soil moisture values from land surface models. Studies have demonstrated that low correlation is observed between soil moisture estimations (by remote sensing) and soil moisture simulations (by land surface models) over irrigated areas (Escorihuela and Quintana-Seguí, 2016) whereas high correlation is observed over non-irrigated surfaces. The low correlation between simulated and estimated surface soil moisture over irrigated areas is mainly due to the presence of irrigation (not considered in the SSM simulations) which is altering the soil

moisture values. Recently, Dari et al. (2021) used this assumption to map irrigated areas at coarse spatial resolution using the SMAP, SMOS and S1 satellite soil moisture. However, such land surface models are usually applied at coarse spatial resolution (few km^2) which makes it difficult to use it at plot scale. Thus, as a future perspective, it is of great importance to investigate the potential of using crop models at plot scale (such as Aquacrop, Cropsyst, Optirrig …) instead of land surface models combined with estimated surface soil moisture from the S^2MP to map irrigated areas at plot scale. Soil moisture values could be first simulated from a crop model using several weather data and hydrological parameters such as the precipitation, temperature, evapotranspiration and soil type. Then, the simulated SSM values could be compared to S^2MP estimated SSM values to map irrigated/non-irrigated plots.

3.4 Estimating irrigation water amounts

One of the important future aspect of irrigation mapping is the estimation of the irrigation water amounts. It is difficult through only SAR and optical data to estimate the applied water amount at each irrigation. Nevertheless, integrating remote sensing derived data into a crop model could help estimate the water use. Using the detected irrigation events from the IEDM with satellite derived vegetation index such as the LAI, the actual amounts of applied water could be estimated. The optimized irrigation crop model developed by INRAE called the "Optirrig" (Cheviron et al., 2020) opens the way toward accurate estimation of water amounts if coupled with the satellite derived information such as the LAI and the irrigation events from the IEDM.

3.5 Participative GIS platform for irrigation management

Irrigation products developed in this thesis (irrigation maps and irrigation events) could be a beneficial tool to public authorities, decision makers and farmers. Thus, it would be useful to create a participatory GIS (Geographic Information System) platform of our irrigation area maps as well as of the detected irrigation events in order to share these supports with end users and decision makers concerned by the thematic of irrigation.

Through the GIS platform, the farmer can enrich our irrigation products (irrigation events and irrigated areas) by validating whether this or that plot is irrigated or not and confirming the detected irrigation events (complete some undetected events or correct some false detections). In such way, the database of irrigated/non-irrigated plots and the irrigation events will be

adjusted according to farmer's knowledge in order to ameliorate and refine the proposed irrigation detection and mapping models. Such participation from farmers through the GIS platform helps us collecting reliable information about the irrigation time and irrigated/rain-fed plots without the need to organize costly terrain campaigns by our research organizations.

The participative GIS platform can help authorities and decision makers understand better follow the current situation of irrigation (irrigated surfaces) and plan for the future policies concerning the water use for irrigation. For example, stakeholders can initiate a pedagogical work designated for farmers to explain whether they need to decrease the irrigation rates based on the availability of the water resources in the irrigated basin and on the irrigated surfaces derived from our irrigation maps. By visualizing the irrigation events detected on the plots, stakeholders can guide farmers to irrigate in a better way suitable to the planted crop since some farmers may sometimes irrigate much more than the needs of the plants. Through the continuous monitoring of irrigated plots using the GIS platform, authorities could propose a bonus/penalty system in order to financially gratify farmers who respect the regulations and penalize the farmers who does not follow the recommendations on water use for irrigation especially when water resources are deficit.

RÉSUMÉ EN FRANÇAIS

1. Introduction (Chapitre I)

L'augmentation de la population mondiale et le changement climatique exercent une pression croissante sur les ressources naturelles et leur capacité à répondre aux besoins alimentaires de la population mondiale. Ce risque patent sur la sécurité alimentaire et la disponibilité en ressources naturelles rend indispensable la mise en place d'un suivi des zones agricoles à grande échelle, notamment pour atteindre les objectifs de développement durable dans le secteur agricole. Actuellement, l'intensification de l'agriculture mondiale, nécessaire pour répondre à la demande alimentaire croissante, est principalement basée sur l'augmentation de l'utilisation des ressources, en particulier de l'eau. L'irrigation est déjà la principale source de consommation d'eau douce à l'échelle mondiale, entraînant une dégradation rapide des ressources en eau dans plusieurs régions. Avec le changement climatique, caractérisé par une diminution des précipitations et une augmentation des températures, les eaux de surface et les eaux souterraines sont soumises à une sur exploitation accrue. Davantage d'informations sur l'étendue des zones irriguées et sur la fréquence d'irrigation aidera les agriculteurs à mieux gérer l'irrigation pour augmenter son efficience d'utilisation et les décideurs politiques pour définir les leviers socio-économiques et techniques pour une mobilisation plus durable des ressources hydriques. L'objectif du présent travail de recherche est de i) développer des méthodes de cartographie de l'étendue spatiale des zones irriguées, ainsi que ii) de mobiliser la télédétection spatiale pour détecter les épisodes d'irrigation à l'échelle de la parcelle. Les données de télédétection optique et radar ont été exploitées dans le but de cartographier et de suivre les pratiques d'irrigation à l'échelle de la parcelle.

Pour répondre aux besoins de cartographie et de suivi de l'irrigation à l'aide de données de télédétection, cette thèse veut répondre à quatre principales questions. Premièrement, quel est le potentiel des données Sentinel-1 (S1) et/ou Sentinel-2 (S2) pour la cartographie des zones irriguées à l'échelle de la parcelle agricole ? Deuxièmement, comment pouvons-nous appliquer un modèle de classification supervisée développé sur une zone pour cartographier l'irrigation d'autres régions ? Troisièmement, pouvons-nous détecter les événements d'irrigation à l'échelle de la parcelle en utilisant les données Sentinel-1 et Sentinel-2 en temps quasi réel ? Enfin, pouvons-nous construire un modèle opérationnel de cartographie de l'irrigation qui ne nécessite

pas de données de terrain in situ et qui soit facilement transférable entre les régions géographiques ?

2. Évaluation de l'estimation d'humidité du sol de produit S^2MP (Chapitre II)

2.1 Objectif

L'humidité du sol (SSM pour Surface Soil Moisture) est un paramètre clé pour la détection de l'irrigation car cette dernière provoque de fait une augmentation de cette humidité. L'intégration d'informations sur l'humidité du sol obtenues à l'échelle de la parcelle ou d'une grille de quelques km^2 permet d'améliorer la précision de la cartographie des zones irriguées. Récemment, l'unité de recherche TETIS (équipe INRAE) a développé un algorithme pour estimer l'humidité du sol (SSM) à l'échelle de la parcelle appelé S^2MP (en anglais Sentinel-1/2 Soil Moisture Product). Afin d'utiliser le S^2MP dans la cartographie de l'irrigation, il est important d'abord d'évaluer la précision de l'algorithme. Dans cette thèse, l'évaluation du S^2MP a été réalisée selon trois approches différentes. Tout d'abord, le S^2MP a été évalué en utilisant des mesures in situ de l'humidité des sols afin de déterminer la précision de l'estimation d'humidité du sol par le S^2MP. Ensuite, le S^2MP a également été comparé au plus récent produit d'humidité du sol fournit par Copernicus « Global Land Service » à l'échelle de 1 km x 1 km (C-SSM). Enfin, le S^2MP a été évalué en comparant l'humidité estimée aux enregistrements de précipitations dérivées des données GPM « Global Precipitation Mission » à l'échelle de grille de 10 km x 10 km. L'évaluation du S^2MP dans les trois approches a été réalisée sur la région Occitanie dans le sud de la France.

2.2 Matériel et méthodes

Zone d'étude : Pour l'évaluation du S^2MP, le site d'étude examiné est la région Occitanie (sud de la France). La région présente une variété de paysages et est principalement couverte de zones agricoles au centre et à l'ouest. Dans la partie est, le climat est méditerranéen (précipitations annuelles moyennes de 700 mm), tandis que dans la partie ouest, le climat est humide et océanique (précipitations annuelles moyennes de 1200 mm).

Données terrain : Pour comparer les produits d'humidité S^2MP et C-SSM avec des données terrain, les valeurs d'humidité du sol ont été mesurées pour 23 parcelles de référence (prairie et

blé) à chaque date d'acquisition Sentinel-1 (identique aux dates S^2MP et C-SSM) dans une zone agricole proche de Montpellier (sud-est de la France) pour la période entre janvier 2017 et juin 2018.

Algorithme S^2MP : Le S^2MP fournit des estimations de l'humidité du sol à l'échelle de la parcelle. Basé sur le couplage de données S1 et S2 (El Hajj et al., 2017), l'algorithme inverse le coefficient de rétrodiffusion radar S1 en utilisant les réseaux de neurones (NN). Pour estimer les valeurs de l'humidité du sol, le S^2MP utilise le signal Sentinel-1 en polarisation VV (σ^0), l'angle d'incidence S1 (θ) et la valeur NDVI dérivée des images optiques S2. Le S^2MP fournit des estimations de l'humidité du sol en unité volumétrique (vol.%).

Données C-SSM : Copernicus « Global Land Service » fournit des estimations de l'humidité du sol à une résolution spatiale de 1 km en utilisant les données S1 (C-SSM). Dans le C-SSM, l'humidité du sol est estimée en pourcentage entre 0 % (conditions de sol extrêmement sèches) et 100 % (conditions de sol très humides). L'estimation de l'humidité du sol du produit C-SSM est convertie en unité volumétrique (vol.%) afin de comparer les estimations du C-SSM avec l'humidité du sol mesurée sur le terrain et celles fournies par S^2MP (fournies en unité volumétrique).

Données GPM : GPM (Global Precipitation Mission) fournit des mesures de précipitations à l'échelle mondiale entre 60°N et 60°S à une résolution spatiale de 0.1° x 0.1° (~10 km x 10 km) et une résolution temporelle de 30 minutes. La comparaison entre S^2MP et les données pluviométriques de la mission GPM a été réalisée sur la durée d'une année hydrologique, entre le 1er septembre 2016 et le 31 août 2017.

Comparaison entre S^2MP et C-SSM : La précision des produits S^2MP et C-SSM par rapport aux données in situ a été déterminée à l'aide du coefficient de corrélation de Pearson (R), la différence quadratique moyenne (RMSD), le biais (estimé - mesuré) et la différence quadratique moyenne non biaisée (ubRMSD). Sur la région Occitanie, une comparaison a été réalisée entre les produits S^2MP et C-SSM pour la période du 1er octobre 2016 au 1er octobre 2017. La comparaison a été réalisée par les métriques statistiques R, RMSD, biais et ubRMSD.

Comparaison entre S^2MP et GPM : Pour chaque carte S^2MP, une carte de précipitations cumulées pour les 6 jours précédant la date de la carte S^2MP (la revisite de S^2MP est de 6 jours) a été calculée en cumulant les cartes IMERG (Integrated Multi-satellite Retrievals for GPM) de la mission GPM toutes les 30 minutes. Ensuite, chaque carte S^2MP a été superposée avec la carte pluviométrie cumulative de 6 jours. Une analyse temporelle a été effectuée entre la

variation des valeurs d'humidité du sol estimées par le S^2MP et les enregistrements des précipitations GPM.

2.3 Résultats et conclusion

La comparaison entre l'estimation de l'humidité du sol du S^2MP et les mesures in situ a montré que le S^2MP fournit une bonne précision pour l'estimation de l'humidité du sol (RMSD = 4,0 vol.%, ubRMSD = 3,9 vol.% et R=0,77). Cependant, le C-SSM est moins précis que le S^2MP (RMSD = 6,0 vol.%, ubRMSD = 6,0 vol.% et R=0,48). La comparaison entre les deux produits sur une année montre une corrélation très élevée entre le S^2MP et le C-SSM principalement sur les zones agricoles qui cultivent des céréales (valeur R entre 0,5 et 0,9 et RMSE entre 4 vol.% et 6 vol.%). Sur les pixels de 1 km x 1 km du produit C-SSM qui correspondent aux parcelles agricoles avec des forêts et des vignes, les valeurs du C-SSM ont tendance à surestimer les valeurs du S^2MP (biais > 5 vol.%).

La comparaison entre le S^2MP et les données pluviométriques GPM a montré une cohérence globale entre les valeurs d'estimation d'humidité du sol et les enregistrements de précipitations à une résolution spatiale de 10 km. Par exemple, entre deux cartes S^2MP consécutives acquises le 15 novembre 2016 et le 21 novembre 2016, les valeurs SSM augmentent de plus de 9 vol.% à cause des 40 mm de précipitations enregistrées entre les deux dates S^2MP. De plus, l'absence de précipitations pendant une période de 25 jours entre le 01 avril 2017 et le 25 avril 2017 a fait chuter les valeurs d'humidité du sol de 25 vol.% à moins de 15 vol.%.

Les résultats ont montré que les estimations de l'humidité par S^2MP à l'échelle de 10 km x 10 km était très corrélées aux événements de précipitation. Pour la suite, notre algorithme de détection des événements d'irrigation a donc pu utiliser les cartes d'humidité S^2MP à l'échelle de grille pour distinguer l'augmentation de l'humidité due à une pluie de l'augmentation due à une irrigation. La cartographie des surfaces irriguées a été établie sur cette base.

3. Approche de classification supervisée pour la cartographie des zones irriguées à l'aide des données Sentinel-1 et Sentinel-2 (Chapitre III)

3.1 Objectif

Dans la deuxième partie de la thèse, nous nous sommes concentrés sur le potentiel des données radar Sentinel-1 et des données optique Sentinel-2 pour la cartographie des zones

irriguées à l'échelle de la parcelle. Deux objectifs complémentaires ont été poursuivis. Le premier objectif est l'étude du potentiel des données S1 et/ou S2 pour la cartographie des zones irriguées en utilisant des modèles de classification supervisée tels que la forêts d'arbres décisionnels (RF) et les réseaux de neurones convolutifs (CNN). Cette étude a été réalisée pour la région de Catalogne, au nord-est de l'Espagne, caractérisée par un climat semi-aride. Le deuxième objectif est de proposer une approche capable de transposer le modèle de classification supervisée d'une région à une autre région ayant des propriétés climatiques différentes. Le modèle développé en Catalogne pour cartographier les parcelles irriguées a ainsi été transféré au bassin versant « Adour Amont » au sud-ouest de la France.

3.2 Matériel et méthodes

Zones d'étude : Deux sites d'étude ont été retenus pour cartographier les zones irriguées en utilisant les données S1/2 et les approches de classification supervisée. Le premier site est la région de Catalogne située en nord-est de l'Espagne (climat semi-aride). Une grande base de données de parcelles irriguées et non irriguées (~193 000 parcelles) est disponible dans cette région grâce aux données SIGPAC (Système d'information géographique pour les parcelles agricoles). Le second site est le bassin Adour Amont, situé dans la région Occitanie du sud-ouest de la France (climat océanique) avec une saison estivale humide présentant généralement plusieurs événements pluvieux. Dans le bassin, 300 parcelles non irriguées et 151 parcelles irriguées ont été collectées par des campagnes de terrain pour l'année 2017.

Données de télédétection : Sur la région de Catalogne, 82 images SAR (Synthetic Aperture Radar = Radar à synthèse d'ouverture) en bande C (5,405 GHz) acquises par les satellites S1A et S1B ont été utilisées pour la période entre septembre 2017 et décembre 2018. Sur le bassin Adour Amont, le même nombre d'images (82 images) a été récupéré pour la période entre septembre 2016 et décembre 2017. En outre, 17 images S2 sans nuage (niveau 2A) ont été utilisé pour chaque site d'étude, couvrant la même période que les images S1.

Cartographie des zones irriguées en Catalogne : La résolution spatiale de la cartographie de l'irrigation est l'échelle de la parcelle. D'abord, les coefficients moyens de rétrodiffusion des images S1 (σ°) ont été calculés à l'échelle des parcelles (mis à disposition par SIGPAC « Système d'information géographique pour les parcelles agricoles de Catalogne » et à l'échelle de la grille (10 km × 10 km). Le signal moyen à l'échelle de la grille est principalement utilisé pour discriminer les événements pluvieux des événements d'irrigation. Basé sur les résultats du

chapitre 2, nous supposons que si le signal SAR moyen à l'échelle de 10 km × 10 km augmente, un événement pluvieux a très probablement eu lieu. Ensuite, l'analyse en composantes principales (ACP) et la transformation en ondelettes (WT) ont été appliqué sur les séries temporelles S1 à l'échelle de la parcelle et de la grille. En outre, l'ACP a été appliquée sur la série temporelle NDVI à l'échelle de la parcelle. Enfin, deux approches de classification, dont le classifieur RF et le CNN, ont été considérées. Dans chaque approche, trois scénarios ont été testés. Le premier scénario est basé sur l'utilisation des données SAR (S1) uniquement, le deuxième scénario considère l'utilisation des données optiques (NDVI) uniquement et le troisième scénario inclut l'utilisation combinée des données optiques et SAR.

Apprentissage par transfert pour la cartographie des zones irriguées dans l'Adour Amont : Parmi les différents scénarios testés sur la Catalogne, le CNN construit à partir des séries temporelles S1 a fourni la meilleure précision de classification (94 %). Cependant, l'application d'exactement le même modèle (construit sur la Catalogne) sur le bassin Adour Amont a donné des résultats insatisfaisants (précision globale de 27 %). Le défi est donc de trouver comment transposer le modèle de classification supervisée d'une zone géographique source (Catalogne) à une zone géographique cible (Adour Amont). Pour cette raison, nous avons proposé un cadre d'apprentissage de transfert basé sur un « distilling before refine ». Le modèle CNN formé en Catalogne est désigné comme le modèle « Teacher ». Un modèle plus léger est distillé à partir du modèle « Teacher » en suivant la procédure de distillation développée par (Hinton et al., 2015). Le modèle « Student » est ensuite raffiné par des données de terrain du bassin Adour Amont. Enfin, la cartographie de l'irrigation sur le bassin Adour Amont est réalisée en utilisant le modèle « Student ». La stratégie proposée a été comparée à différentes approches, y compris RF et CNN, directement entraînées à l'aide des données de terrain.

3.3 Résultats et conclusion

En Catalogne, les résultats montrent une bonne précision globale (OA) pour le classifieur RF en utilisant les composantes principales issues de l'ACP (90,7 %) et les coefficients de la transformation en ondelette (89,1 %). De plus, les résultats montrent que l'outil de classification RF construit en utilisant les données optiques (NDVI) est performant avec une OA = 89.5 %. L'utilisation combinée des données optiques et SAR (analyse en composantes principales) a légèrement amélioré la précision de la classification (OA = 92,3 %). Les résultats de la validation de l'approche CNN, appliquée directement sur les séries temporelles S1, ont montré

une très bonne précision (OA = 94,1 %) par rapport au RF. L'utilisation uniquement des données NDVI dans un classificateur CNN a produit une précision globale plus faible (91,5 %) que celle obtenue en utilisant uniquement les données S1. Cependant, l'utilisation combinée des données S1 et S2 dans le CNN n'a pas amélioré la précision de manière significative.

Dans le bassin d'Adour Amont, la technique d'apprentissage par transfert « distilling before refine » a permis de cartographier les zones irriguées avec une précision globale de 83%. Comparé à d'autres alternatives telles que RF et CNN construites directement à partir de données in situ d'Adour Amont, l'apprentissage par transfert a montré une meilleure précision.

Les résultats obtenus dans ce chapitre démontrent la possibilité d'obtenir une cartographie de l'irrigation à haute résolution en utilisant les données S1 et S2. Cependant, un classifieur supervisé construit sur une zone et appliqué sur une seconde zone ne donne pas une bonne précision. De plus, l'application du même classifieur sur la même zone mais à une année différente pourrait conduire à perte de précision. La construction du modèle supervisé nécessite toujours des données terrain. En outre, le transfert du modèle supervisé nécessite également des données de terrain sur le site cible. Par conséquent, la disponibilité des données terrain reste un obstacle pour réaliser la cartographie de l'irrigation. Notre étude dans le prochain chapitre se concentrera sur le développement d'approches opérationnelles semi-supervisées pour cartographier les zones irriguées sans avoir recours aux données de terrain.

4. Vers une cartographie et un suivi opérationnel des zones irriguées (Chapitre IV)

4.1 Objectif

L'utilisation de données S1 et S2 couplée à des classifieurs supervisés améliore la précision d'une cartographie des zones irriguées. Cependant, la dépendance des classifieurs supervisés aux données de terrain reste un obstacle pour la cartographie opérationnelle des zones irriguées. En outre, la détection des événements d'irrigation à l'échelle de la parcelle à l'aide de données de télédétection n'a pas encore été explorée en profondeur malgré son apport potentiel pour la gestion de l'eau d'irrigation. Ce chapitre veut donc d'abord proposer un modèle opérationnel en temps quasi réel pour la détection des événements d'irrigation à l'échelle de la parcelle (IEDM qui veut dire Irrigation Event Detection Model). L'IEDM est une méthode basée sur un arbre

de décision construit à partir des données S1, S2 et S^2MP. Cette méthode est ensuite validée sur des parcelles de prairies irriguées situées dans la plaine de la Crau, dans le sud-est de la France.

Une méthode opérationnelle de cartographie de l'irrigation demande d'être capable de générer ses propres données d'entraînement. Pour générer automatiquement les données d'entraînement (parcelles irriguées et non irriguées) nécessaires, le modèle IEDM a été utilisé à l'échelle de la parcelle en complément des valeurs NDVI. Ensuite, pour cartographier les zones irriguées, une classification RF a été construite en utilisant les données d'entraînement générées automatiquement, les données S1 et les données S2.

4.2 Matériel et méthode

Zones d'étude : Pour développer l'IEDM, trois sites d'étude ont été examinés. Trois parcelles irriguées situées à Montpellier avec des événements d'irrigation enregistrés (48 irrigations) ont été utilisées pour analyser l'effet de l'irrigation sur les coefficients de rétrodiffusion de S1 (σ^0). Les données SIGPAC en Catalogne et les données de terrain collectées à Adour Amont (détaillées dans le chapitre 2) ont été utilisées pour analyser la performance de l'IEDM dans la détection de l'irrigation. Pour valider l'IEDM, 46 parcelles de prairies (plaine de la Crau, sud-est de la France) avec des événements d'irrigation enregistrés in situ ont été utilisées.

Pour la cartographie opérationnelle des zones irriguées, un site d'étude situé à proximité de la ville d'Orléans (centre nord de la France) est examiné. Sur ce site d'étude, plusieurs campagnes de terrain ont été réalisées pour obtenir des informations sur l'irrigation (absence ou existence d'irrigation) pour quatre années (2017, 2018, 2019 et 2020). Chaque année, les données de terrain collectées ont été utilisées pour valider la cartographie de l'irrigation réalisée à l'aide de la méthodologie proposée.

Données de télédétection: Sur chaque site d'étude, toutes les images Sentinel-1 possibles acquises avec différents passages de S1 (ascendants et descendants) ont été téléchargées et traitées. Toutes les images S2 possibles (sans nuage) ont également été téléchargées pour chaque site d'étude correspondant à la même période d'acquisition que les S1. De plus, toutes les cartes S^2MP disponibles sur chaque site d'étude ont été utilisées.

Développement et évaluation de l'IEDM : L'IEDM est basé sur la détection du changement des coefficients de rétrodiffusion S1 à l'échelle de la parcelle (σ_p^0), entre l'acquisition SAR au temps t_i et l'acquisition SAR précédente au temps t_{i-1}. L'IEDM suppose que l'augmentation de

la σ_p^0 entre deux acquisitions S1 consécutives est principalement causée par l'augmentation de l'humidité du sol due soit à une pluie, soit à un événement d'irrigation. D'autre part, l'augmentation des valeurs de σ_G^0 (échelle de grille de 10 km x 10 km) et les valeurs élevées d'humidité du sol de S^2MP à l'échelle de la grille pourraient être la preuve d'un événement pluvieux. La stabilité à de faibles valeurs ou la diminution des valeurs de σ_G^0 pourraient, elles, indiquer la persistance de conditions sèches entre les deux dates (absence de précipitations). Ainsi, l'augmentation des valeurs σ_p^0 entre deux acquisitions consécutives accompagnée de la stabilité ou de la diminution des valeurs σ_G^0 est considérée comme un événement d'irrigation. L'IEDM fournit, pour chaque image S1, un poids de possibilité d'irrigation qui représente la probabilité d'avoir un événement d'irrigation. Le poids peut être soit 0 (pas d'irrigation), 25 (faible possibilité), 50 (possibilité moyenne) ou 100 (forte possibilité). Étant donné que le signal de rétrodiffusion SAR peut être affecté par d'autres facteurs tels que le couvert végétal (cycle de croissance de végétation) et la rugosité de la surface, des filtres supplémentaires ont été ajoutés en utilisant les valeurs NDVI et les valeurs d'humidité du sol du S^2MP. Pour valider l'IEDM, le modèle a été appliqué sur 46 parcelles de prairies irriguées (plaine de Crau), et les événements d'irrigation détectés ont été comparés avec le calendrier d'irrigation enregistré in situ.

Cartographie opérationnelle des zones irriguées : La méthodologie proposée pour la cartographie de l'irrigation consiste en deux étapes principales. Dans la première étape, les parcelles d'entraînement (irriguées/non irriguées) sont sélectionnées en fonction de deux critères dérivés des données S1 et optiques. Le premier critère est basé sur un poids de possibilité d'irrigation dérivé de l'IEDM. Le second critère est basé sur la valeur NDVI maximale de la parcelle.

En appliquant l'IEDM sur les différentes séries temporelles des orbites S1 (en polarisations VV et VH), on obtient une métrique de possibilité d'irrigation. Cette métrique est calculée en cumulant les possibilités d'irrigation (0, 25, 50 et 100) obtenues à chaque image S1 de toutes les séries temporelles S1 en polarisations VV et VH (pour chaque parcelle). Pour la sélection des données d'entraînement, les parcelles ayant une valeur de possibilité d'irrigation très élevée sont considérées comme irriguées et les parcelles ayant des valeurs très faibles sont considérées comme non irriguées. Sur ces parcelles sélectionnées, un autre filtre NDVI est ensuite appliqué. On suppose ici que pour les parcelles non irriguées, la valeur maximum de NDVI ne doit pas dépasser 0,7 alors que, pour les parcelles irriguées, la valeur maximum de NDVI doit être supérieur à 0,8. La deuxième étape consiste à utiliser les données S1 (à l'échelle de la parcelle

et de la grille), les données S2 (NDVI) et les parcelles d'entraînement sélectionnées dans un classifieur (RF) pour réaliser la cartographie de l'irrigation.

4.3 Résultats et conclusion

En termes de développement de l'IEDM, les résultats montrent que 84,8 % des événements d'irrigation sur des parcelles agricoles à Montpellier ont été correctement détectés en utilisant ce modèle. Sur le site de Catalogne, les résultats montrent aussi une bonne cohérence : 90,2 % des parcelles non irriguées n'ont rencontré aucun événement d'irrigation détecté alors que 72,4 % des parcelles irriguées ont eu un et plusieurs événements d'irrigation détectés. Dans le bassin versant de l'Adour Amont, l'analyse indique également que deux événements d'irrigation et plus ont été détectés pour 90 % des parcelles irriguées. La validation de l'IEDM sur les 46 parcelles de prairies irriguées montre qu'en utilisant uniquement la polarisation VV, 82,4 % des événements d'irrigation enregistrés in situ sont correctement détectés, avec une valeur F_score de 73,8 %. Une précision moindre est obtenue en utilisant la polarisation VH, où 79,9 % des événements d'irrigation in situ sont correctement détectés avec un F_score de 72,2 %. L'utilisation combinée des polarisations VV et VH permet de détecter 74,1 % des événements d'irrigation, avec une valeur F_score de 76,4 %.

La classification opérationnelle et semi-supervisée proposée a été validée en utilisant des données réelles in situ collectées pendant quatre ans à Orléans (centre-ouest de la France). Les résultats indiquent qu'en utilisant la procédure de classification proposée, la précision globale de la classification de l'irrigation atteint 84,3 %, 93,0 %, 81,8 % et 72,8 % pour les années 2020, 2019, 2018 et 2017, respectivement. La comparaison entre l'approche de classification proposée et le classifieur RF construit directement à partir de données in situ (méthode entièrement supervisée) montre que notre approche atteint une précision presque similaire à celle obtenue en utilisant les classifieurs RF avec des données in situ (différence de précision globale ne dépassant pas 6,2 %).

5. Conclusions générales et perspectives (Chapitre V)

Des informations précises sur l'étendue spatiale des zones irriguées sont fondamentales dans plusieurs aspects de la gestion de l'eau et de l'agriculture. Dans cette thèse, le produit d'humidité du sol S^2MP a d'abord été évalué afin de l'utiliser ensuite dans la cartographie de l'irrigation et la détection des événements d'irrigation. Ensuite, plusieurs approches de classification

supervisée ont été présentées pour cartographier les zones irriguées à l'échelle de la parcelle sur deux sites d'étude différents (Catalogne et Adour Amont). Puis, un algorithme a été proposé pour détecter les événements d'irrigation à l'échelle de la parcelle en utilisant les données S1, S2 et les cartes d'humidité du produit S^2MP. Enfin, une méthode de classification semi-supervisée a été développée pour la cartographie des parcelles irriguées.

Les principaux résultats sont :

- Le S^2MP fournit de bonnes estimations de l'humidité du sol (humidité de surface sur les 5 premiers centimètres), fortement corrélées avec les événements pluvieux.
- L'utilisation des données Sentinel-1 et Sentinel-2 permet une cartographie précise des parcelles irriguées en utilisant des modèles de classification supervisée (RF ou CNN).
- Les modèles supervisés de cartographie des surfaces irriguées qui fonctionnent dans une région à une année donnée ne sont pas nécessairement applicables à d'autres régions et périodes.
- En utilisant des techniques d'apprentissage par transfert, un modèle construit sur une zone spécifique peut être transféré à une autre zone afin de cartographier les parcelles irriguées.
- Le couplage de S1, S2 et S^2MP permet une bonne détection des événements d'irrigation à l'échelle de la parcelle en utilisant l'algorithme « IEDM » développé.
- En utilisant principalement l'IEDM et un classificateur RF, on obtient une classification semi-supervisée des zones irriguées qui ne nécessite pas de données in situ.

La cartographie des zones irriguées et la détection des événements d'irrigation en utilisant les données S1 et S2 présentent néanmoins quelques limites :

- Sur les zones humides, la cartographie des zones irriguées est plus difficile et présente plus de défis. Lorsque la zone étudiée est humide, elle rencontre plusieurs événements pluvieux pendant la période d'irrigation. Ainsi, les précipitations fréquentes rendent les valeurs NDVI des parcelles non irriguées similaires aux valeurs NDVI des parcelles irriguées. Ce changement dans le NDVI de la parcelle non irriguée peut potentiellement réduire la différence de NDVI entre les parcelles irriguées et non irriguées. Ainsi, il devient difficile de distinguer les deux classes avec une bonne précision.
- Dans le cas d'une couverture végétale très développée (NDVI > 0.7, principalement pour le blé et les prairies), la pénétration limitée du signal SAR en bande C (longueur d'onde ~ 6 cm) peut contraindre la détection des événements d'irrigation. Dans le cas d'une

couverture végétale très développée, le signal SAR en bande C peut ne pas atteindre le sol et, donc, la contribution du sol dans le signal de rétrodiffusion devient négligeable. Par conséquent, la détection des événements d'irrigation basée sur le changement des valeurs d'humidité du sol devient alors difficile.

- Lorsque l'image S1 est acquise longtemps après l'événement d'irrigation, il devient difficile de détecter le changement des valeurs d'humidité du sol après l'irrigation. Ceci est principalement dû à l'évaporation de la teneur en eau du sol, surtout en été. Nous avons montré que l'événement d'irrigation pouvait être détecté deux à trois jours après l'événement d'irrigation si la couverture végétale est faible (NDVI < 0,7). Si la végétation est bien développée (NDVI > 0.7), l'événement d'irrigation pourrait être principalement détecté s'il se produit à la même date que l'acquisition SAR. Cette limitation est donc due au temps de revisite des données S1 et la longueur d'onde utilisé des capteurs S1.

Quelques perspectives se dégagent de cette thèse :

1. Le signal SAR en bande L (longueur d'onde ~ 24 cm) a des capacités de pénétration plus élevées que le signal SAR en bande C, principalement dans une couverture végétale dense. L'intégration des données en bande L (ALOS-2/PALSAR-2) avec les données en bande C peut donc améliorer la détection des événements d'irrigation. Les futurs satellites en bande L (NISAR et Tandem-L qui seront lancés à partir de 2022) permettront une meilleure cartographie des événements d'irrigation si la végétation est bien développée. E effet, les prochaines données en bande L combinées aux données en bande C de Sentinel-1 augmenteront le nombre d'acquisitions radar par semaine, ce qui permettra de détecter presque tous les événements d'irrigation.
2. La cartographie des arbres fruitiers irrigués à l'aide d'images satellites est encore un sujet inexploré. Parmi les arbres fruitiers, l'irrigation des vignobles est le sujet le plus important dans plusieurs régions d'Europe, notamment dans le sud. L'élaboration d'approches capables de détecter les événements d'irrigation dans les parcelles de vigne et de cartographier les vignobles irrigués est d'une grande importance pour les autorités publiques et les fabricants de vin au moment où les dérèglements climatiques questionnent les pratiques de la filière. Ainsi, une des perspectives ouvertes par cette thèse est de développer, d'une part, une méthode capable d'estimer l'humidité du sol pour les parcelles de vigne et, d'autre part, une méthode capable de détecter les événements d'irrigation sur les parcelles de vigne

3. Il serait utile de créer une plateforme SIG (Système d'Information Géographique) participative de nos cartes d'irrigation ainsi que des événements d'irrigation détectés afin de partager ces supports avec les gestionnaires concernés par la thématique de l'irrigation. Les gestionnaires pourraient alors initier un travail pédagogique à destination des agriculteurs pour adapter la fréquence d'irrigation à la disponibilité de la ressource en eau. Grâce au suivi continu des parcelles irriguées à l'aide de la plateforme SIG, les autorités pourraient proposer un système de bonus/malus afin de gratifier les agriculteurs qui respectent les réglementations et pénaliser ceux qui ne suivent pas les recommandations sur l'utilisation de l'eau pour l'irrigation.

REFERENCES

Ahishali, M., Kiranyaz, S., Ince, T., and Gabbouj, M. (2019). Dual and Single Polarized SAR Image Classification Using Compact Convolutional Neural Networks. Remote Sensing *11*, 1340.

Ambika, A.K., Wardlow, B., and Mishra, V. (2016). Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. Scientific Data *3*, 1–14.

Amri, R., Zribi, M., Lili-Chabaane, Z., Wagner, W., and Hasenauer, S. (2012). Analysis of C-Band Scatterometer Moisture Estimations Derived Over a Semiarid Region. IEEE Transactions on Geoscience and Remote Sensing *50*, 2630–2638.

Aparicio, N., Villegas, D., Casadesus, J., Araus, J.L., and Royo, C. (2000). Spectral Vegetation Indices as Nondestructive Tools for Determining Durum Wheat Yield. Agron. J. *92*, 83–91.

Attema, E.P.W., and Ulaby, F.T. (1978). Vegetation modeled as a water cloud. Radio Science *13*, 357–364.

Aubert, M., Baghdadi, N., Zribi, M., Douaoui, A., Loumagne, C., Baup, F., El Hajj, M., and Garrigues, S. (2011). Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust. Remote Sensing of Environment *115*, 1801–1810.

Aubert, M., Baghdadi, N.N., Zribi, M., Ose, K., El Hajj, M., Vaudour, E., and Gonzalez-Sosa, E. (2013). Toward an Operational Bare Soil Moisture Mapping Using TerraSAR-X Data Acquired Over Agricultural Areas. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing *6*, 900–916.

Baghdadi, N., King, C., Bourguignon, A., and Remond, A. (2002). Potential of ERS and Radarsat data for surface roughness monitoring over bare agricultural fields: Application to catchments in Northern France. International Journal of Remote Sensing *23*, 3427–3442.

Baghdadi, N., Holah, N., and Zribi, M. (2006a). Calibration of the Integral Equation Model for SAR data in C-band and HH and VV polarizations. International Journal of Remote Sensing *27*, 805–816.

Baghdadi, N., Holah, N., Dubois-Fernandez, P., Dupuis, X., and Garestier, F. (2006b). Evaluation of polarimetric L- and P-bands RAMSES data for characterizing Mediterranean vineyards. Canadian Journal of Remote Sensing *32*, 380–389.

Baghdadi, N., Zribi, M., Loumagne, C., Ansart, P., and Anguela, T. (2008a). Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields. Remote Sensing of Environment *112*, 4370–4379.

Baghdadi, N., Cerdan, O., Zribi, M., Auzet, V., Darboux, F., El Hajj, M., and Kheir, R.B. (2008b). Operational performance of current synthetic aperture radar sensors in mapping soil surface characteristics in agricultural environments: application to hydrological and erosion modelling. Hydrological Processes: An International Journal *22*, 9–20.

Baghdadi, N., Camus, P., Beaugendre, N., Issa, O.M., Zribi, M., Desprats, J.F., Rajot, J.L., Abdallah, C., and Sannier, C. (2011a). Estimating Surface Soil Moisture from TerraSAR-X Data over Two Small Catchments in the Sahelian Part of Western Niger. Remote Sensing *3*, 1266–1283.

Baghdadi, N., Abou Chaaya, J., and Zribi, M. (2011b). Semiempirical Calibration of the Integral Equation Model for SAR Data in C-Band and Cross Polarization Using Radar Images and Field Measurements. IEEE Geosci. Remote Sensing Lett. *8*, 14–18.

Baghdadi, N., Saba, E., Aubert, M., Zribi, M., and Baup, F. (2011c). Evaluation of Radar Backscattering Models IEM, Oh, and Dubois for SAR Data in X-Band Over Bare Soils. IEEE Geosci. Remote Sensing Lett. *8*, 1160–1164.

Baghdadi, N., Cresson, R., Pottier, E., Aubert, M., Mehrez, M., Jacome, A., and Benabdallah, S. (2012a). A Potential Use for the C-Band Polarimetric SAR Parameters to Characterize the Soil Surface Over Bare Agriculture Fields. IEEE Transactions on Geoscience and Remote Sensing *50*, 3844–3858.

Baghdadi, N., Cresson, R., El Hajj, M., Ludwig, R., and La Jeunesse, I. (2012b). Estimation of soil parameters over bare agriculture areas from C-band polarimetric SAR data using neural networks. Hydrol. Earth Syst. Sci. *16*, 1607–1621.

Baghdadi, N., Choker, M., Zribi, M., Hajj, M., Paloscia, S., Verhoest, N., Lievens, H., Baup, F., and Mattia, F. (2016a). A New Empirical Model for Radar Scattering from Bare Soil Surfaces. Remote Sensing *8*, 920.

Baghdadi, N., El Hajj, M., Zribi, M., and Bousbih, S. (2017). Calibration of the water cloud model at C-band for winter crop fields and grasslands. Remote Sensing *9*, 969.

Baghdadi, N., El Hajj, M., Choker, M., Zribi, M., Bazzi, H., Vaudour, E., Gilliot, J.-M., and Ebengo, D. (2018a). Potential of Sentinel-1 Images for Estimating the Soil Roughness over Bare Agricultural Soils. Water *10*, 131.

Baghdadi, N., Bazzi, H., El Hajj, M., and Zribi, M. (2018b). Detection of Frozen Soil Using Sentinel-1 SAR Data. Remote Sensing *10*, 1182.

Baghdadi, N.N., El Hajj, M., Zribi, M., and Fayad, I. (2016b). Coupling SAR C-Band and Optical Data for Soil Moisture and Leaf Area Index Retrieval Over Irrigated Grasslands. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing *9*, 1229–1243.

Bauer-Marschallinger, B., and Pfeil, I. (2021). Copernicus Global Land Operations "Vegetation and Energy"SCIENTIFIC QUALITY EVALUATION 2020.

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., et al. (2019). Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. IEEE Transactions on Geoscience and Remote Sensing *57*, 520–539.

Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., and Belhouchette, H. (2019a). A Comparison of Two Soil Moisture Products S^2 MP and Copernicus-SSM Over Southern France. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing *12*, 3366–3375.

Bazzi, H., Baghdadi, N., El Hajj, M., and Zribi, M. (2019b). Potential of Sentinel-1 Surface Soil Moisture Product for Detecting Heavy Rainfall in the South of France. Sensors *19*, 802.

Bazzi, H., Baghdadi, N., Ienco, D., El Hajj, M., Zribi, M., Belhouchette, H., Escorihuela, M.J., and Demarez, V. (2019c). Mapping Irrigated Areas Using Sentinel-1 Time Series in Catalonia, Spain. Remote Sensing *11*, 1836.

Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., Minh, D.H.T., Ndikumana, E., Courault, D., and Belhouchette, H. (2019d). Mapping Paddy Rice Using Sentinel-1 SAR Time Series in Camargue, France. Remote Sensing *11*, 887.

Bazzi, H., Ienco, D., Baghdadi, N., Zribi, M., and Demarez, V. (2020a). Distilling Before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas Using Sentinel-1 Time Series. IEEE Geosci. Remote Sensing Lett. 1–5.

Bazzi, H., Baghdadi, N., Fayad, I., Zribi, M., Belhouchette, H., and Demarez, V. (2020b). Near Real-Time Irrigation Detection at Plot Scale Using Sentinel-1 Data. Remote Sensing *12*, 1456.

Bazzi, H., Baghdadi, N., Fayad, I., Charron, F., Zribi, M., and Belhouchette, H. (2020c). Irrigation Events Detection over Intensively Irrigated Grassland Plots Using Sentinel-1 Data. Remote Sensing *12*, 4058.

Benabdelouahab, T., Derauw, D., Lionboui, H., Hadria, R., Tychon, B., Boudhar, A., Balaghi, R., Lebrini, Y., Maaroufi, H., and Barbier, C. (2018). Using SAR Data to Detect Wheat Irrigation Supply in an Irrigated Semi-arid Area. JAS *11*, 21.

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R., and Dupuy, S. (2018). M3fusion: A deep learning architecture for multi-{Scale/Modal/Temporal} satellite data fusion. ArXiv Preprint ArXiv:1803.01945.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation Learning: A Review and New Perspectives. IEEE Trans. Pattern Anal. Mach. Intell. *35*, 1798–1828.

Bernhard Bauer Marschallinger and Christoph Paulik (2018). Copernicus Global Land Operations "Vegetation and Energy" ALGORITHM THEORETICAL BASIS DOCUMENT.

Biggs, T.W., Thenkabail, P.S., Gumma, M.K., Scott, C.A., Parthasaradhi, G.R., and Turral, H.N. (2006). Irrigated area mapping in heterogeneous landscapes with MODIS time series, ground truth and census data, Krishna Basin, India. International Journal of Remote Sensing *27*, 4245–4266.

Boken, V.K., Hoogenboom, G., Kogan, F.N., Hook, J.E., Thomas, D.L., and Harrison, K.A. (2004). Potential of using NOAA-AVHRR data for estimating irrigated area to help solve an inter-state water dispute. International Journal of Remote Sensing *25*, 2277–2286.

Boucher, O., Myhre, G., and Myhre, A. (2004). Direct human influence of irrigation on atmospheric water vapour and climate. Climate Dynamics *22*, 597–603.

Bousbih, S., Zribi, M., Lili-Chabaane, Z., Baghdadi, N., El Hajj, M., Gao, Q., and Mougenot, B. (2017). Potential of Sentinel-1 Radar Data for the Assessment of Soil and Cereal Cover Parameters. Sensors *17*, 2617.

Bousbih, S., Zribi, M., El Hajj, M., Baghdadi, N., Lili-Chabaane, Z., Gao, Q., and Fanise, P. (2018). Soil Moisture and Irrigation Mapping in A Semi-Arid Region, Based on the Synergetic Use of Sentinel-1 and Sentinel-2 Data. Remote Sensing *10*, 1953.

Breiman, L. (2001). Random Forests. Machine Learning *45*, 5–32.

Brisco, B., Brown, R., Koehler, J., Sofko, G., and Mckibben, M. (1990). The diurnal pattern of microwave backscattering by wheat. Remote Sensing of Environment *34*, 37–47.

Brocca, L., Melone, F., Moramarco, T., and Singh, V.P. (2009). Assimilation of observed soil moisture data in storm rainfall-runoff modeling. Journal of Hydrologic Engineering *14*, 153–165.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., and Levizzani, V. (2014). Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. Journal of Geophysical Research: Atmospheres *119*, 5128–5141.

Brown, J.F., and Pervez, M.S. (2014). Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture. Agricultural Systems *127*, 28–40.

Bruinsma, J. (2009). The resource outlook to 2050: by how much do land, water and crop yields need to increase by 2050. In Expert Meeting on How to Feed the World In, pp. 24–26.

Burney, J., Woltering, L., Burke, M., Naylor, R., and Pasternak, D. (2010). Solar-powered drip irrigation enhances food security in the Sudano–Sahel. Proc Natl Acad Sci USA *107*, 1848–1853.

Cai, X., and Rosegrant, M.W. (2002). Global Water Demand and Supply Projections: Part 1. A Modeling Approach. Water International *27*, 159–169.

Carrasco, L., O'Neil, A., Morton, R., and Rowland, C. (2019). Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. Remote Sensing *11*, 288.

Cheema, M.J.M., and Bastiaanssen, W.G. (2010). Land use and land cover classification in the irrigated Indus Basin using growth phenology information from satellite data to support water management analysis. Agricultural Water Management *97*, 1541–1552.

Chen, K.S., Tzong-Dar Wu, Leung Tsang, Qin Li, Jiancheng Shi, and Fung, A.K. (2003). Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations. IEEE Transactions on Geoscience and Remote Sensing *41*, 90–101.

Chen, Y., Lu, D., Luo, L., Pokhrel, Y., Deb, K., Huang, J., and Ran, Y. (2018). Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, Landsat imagery, and ancillary data. Remote Sensing of Environment *204*, 197–211.

CHEVIRON, B., SERRA-WITTLING, C., DOMINGUEZ BOHORQUEZ, J.D., MOLLE, B., LO, M., ELAMRI, Y., and DELMAS, M. (2020). Irrigation efficiency and optimization: the Optirrig model.

Cookmartin, G., Saich, P., Quegan, S., Cordey, R., Burgess-Allen, P., and Sowter, A. (2000). Modeling microwave interactions with crops and comparison with ERS-2 SAR observations. IEEE Trans. Geosci. Remote Sensing *38*, 658–670.

Dahl, G.E., Sainath, T.N., and Hinton, G.E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, (IEEE), pp. 8609–8613.

Dari, J., Quintana-Seguí, P., Escorihuela, M.J., Stefan, V., Brocca, L., and Morbidelli, R. (2021). Detecting and mapping irrigated areas in a Mediterranean environment by using remote sensing soil moisture and a land surface model. Journal of Hydrology *596*, 126129.

Deines, J.M., Kendall, A.D., and Hyndman, D.W. (2017). Annual Irrigation Dynamics in the U.S. Northern High Plains Derived from Landsat Satellite Data: Satellite-Derived Irrigation Dynamics. Geophys. Res. Lett. *44*, 9350–9360.

Deines, J.M., Kendall, A.D., Crowley, M.A., Rapp, J., Cardille, J.A., and Hyndman, D.W. (2019). Mapping three decades of annual irrigation across the US High Plains Aquifer using Landsat and Google Earth Engine. Remote Sensing of Environment *233*, 111400.

Del Frate, F., Ferrazzoli, P., Guerriero, L., Strozzi, T., Wegmuller, U., Cookmartin, G., and Quegan, S. (2004). Wheat Cycle Monitoring Using Radar Data and a Neural Network Trained by a Model. IEEE Trans. Geosci. Remote Sensing *42*, 35–44.

Demarez, V., Helen, F., Marais-Sicre, C., and Baup, F. (2019). In-Season Mapping of Irrigated Crops Using Landsat 8 and Sentinel-1 Time Series. Remote Sensing *11*, 118.

Dezfuli, A.K., Ichoku, C.M., Huffman, G.J., Mohr, K.I., Selker, J.S., van de Giesen, N., Hochreutener, R., and Annor, F.O. (2017). Validation of IMERG Precipitation in Africa. Journal of Hydrometeorology *18*, 2817–2825.

Dheeravath, V., Thenkabail, P.S., Chandrakantha, G., Noojipady, P., Reddy, G.P.O., Biradar, C.M., Gumma, M.K., and Velpuri, M. (2010). Irrigated areas of India derived using MODIS 500 m time series for the years 2001–2003. ISPRS Journal of Photogrammetry and Remote Sensing *65*, 42–59.

Dietterich, T.G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation *10*, 1895–1923.

Dubois, P.C., van Zyl, J., and Engman, T. (1995). Measuring soil moisture with imaging radars. IEEE Trans. Geosci. Remote Sensing *33*, 915–926.

El Hajj, M., Baghdadi, N., Zribi, M., Belaud, G., Cheviron, B., Courault, D., and Charron, F. (2016a). Soil moisture retrieval over irrigated grassland using X-band SAR data. Remote Sensing of Environment *176*, 202–218.

El Hajj, M., Baghdadi, N., Zribi, M., and Angelliaume, S. (2016b). Analysis of Sentinel-1 Radiometric Stability and Quality for Land Surface Applications. Remote Sensing *8*, 406.

El Hajj, M., Baghdadi, N., Zribi, M., and Bazzi, H. (2017). Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. Remote Sensing *9*, 1292.

El Hajj, M., Baghdadi, N., Zribi, M., Rodríguez-Fernández, N., Wigneron, J., Al-Yaari, A., Al Bitar, A., Albergel, C., and Calvet, J.-C. (2018a). Evaluation of SMOS, SMAP, ASCAT and Sentinel-1 Soil Moisture Products at Sites in Southwestern France. Remote Sensing *10*, 569.

El Hajj, M., Baghdadi, N., Bazzi, H., and Zribi, M. (2018b). Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. Remote Sensing *11*, 31.

El Hajj, M., Baghdadi, N., Bazzi, H., and Zribi, M. (2018c). Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. Remote Sensing *11*, 31.

El Hajj, M., Baghdadi, N., and Zribi, M. (2019). Comparative analysis of the accuracy of surface soil moisture estimation from the C- and L-bands. International Journal of Applied Earth Observation and Geoinformation *82*, 101888.

van Emmerik, T., Steele-Dunne, S.C., Judge, J., and van de Giesen, N. (2015). Impact of Diurnal Variation in Vegetation Water Content on Radar Backscatter From Maize During Water Stress. IEEE Trans. Geosci. Remote Sensing *53*, 3855–3869.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., et al. (2010). The Soil Moisture Active Passive (SMAP) Mission. Proc. IEEE *98*, 704–716.

Escorihuela, M.J., and Quintana-Seguí, P. (2016). Comparison of remote sensing and simulated soil moisture datasets in Mediterranean landscapes. Remote Sensing of Environment *180*, 99–114.

FAO (2017). Water for sustainable food and agriculture (Rome, Italy: FAO).

FAO (2018). The future of food and agriculture – Alternative pathways to 2050 (Rome: FOOD & AGRICULTURE ORG).

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters *27*, 861–874.

Fayad, I., Baghdadi, N., Bazzi, H., and Zribi, M. (2020). Near Real-Time Freeze Detection over Agricultural Plots Using Sentinel-1 Data. Remote Sensing *12*, 1976.

Feng, Q., Yang, J., Zhu, D., Liu, J., Guo, H., Bayartungalag, B., and Li, B. (2019). Integrating Multitemporal Sentinel-1/2 Data for Coastal Land Cover Classification Using a Multibranch Convolutional Neural Network: A Case of the Yellow River Delta. Remote Sensing *11*, 1006.

Ferguson, C.R., Pan, M., and Oki, T. (2018). The Effect of Global Warming on Future Water Availability: CMIP5 Synthesis. Water Resour. Res. *54*, 7791–7819.

Fieuzal, R., Duchemin, B., Jarlan, L., Zribi, M., Baup, F., Merlin, O., Hagolle, O., and Garatuza-Payan, J. (2011). Combined use of optical and radar satellite data for the monitoring of irrigation and soil moisture of wheat crops. Hydrology and Earth System Sciences *15*, 1117–1129.

Friesen, J., Steele-Dunne, S.C., and van de Giesen, N. (2012). Diurnal Differences in Global ERS Scatterometer Backscatter Observations of the Land Surface. IEEE Trans. Geosci. Remote Sensing *50*, 2595–2602.

Fung, A.K. (1994). Microwave scattering and emission models and their applications (Boston: Artech House).

Fung, A.K., Li, Z., and Chen, K.S. (1992). Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sensing *30*, 356–369.

Gaetano, R., Ienco, D., Ose, K., and Cresson, R. (2018). A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery. Remote Sensing *10*, 1746.

Gao, Q., Zribi, M., Escorihuela, M., and Baghdadi, N. (2017). Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. Sensors *17*, 1966.

Gao, Q., Zribi, M., Escorihuela, M., Baghdadi, N., and Segui, P. (2018). Irrigation Mapping Using Sentinel-1 Time Series at Field Scale. Remote Sensing *10*, 1495.

García-Tejero, I.F., Durán-Zuazo, V.H., Muriel-Fernández, J.L., and Rodríguez-Pleguezuelo, C.R. (2011). Water and sustainable agriculture. In Water and Sustainable Agriculture, (Springer), pp. 1–94.

Georganos, S., Grippa, T., Lennert, M., Vanhuysse, S., Johnson, B., and Wolff, E. (2018). Scale Matters: Spatially Partitioned Unsupervised Segmentation Parameter Optimization for Large and Heterogeneous Satellite Images. Remote Sensing *10*, 1440.

Gherboudj, I., Magagi, R., Berg, A.A., and Toth, B. (2011). Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. Remote Sensing of Environment *115*, 33–43.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. Science *327*, 812–818.

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M.C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., and Noble, I. (2013). Sustainable development goals for people and planet. Nature *495*, 305–307.

Grillakis, M.G. (2019). Increase in severe and extreme soil moisture droughts for Europe under climate change. Science of The Total Environment *660*, 1245–1255.

Gumma, M.K., Thenkabail, P.S., Hideto, F., Nelson, A., Dheeravath, V., Busia, D., and Rala, A. (2011). Mapping Irrigated Areas of Ghana Using Fusion of 30 m and 250 m Resolution Remote-Sensing Data. Remote Sensing *3*, 816–835.

Hajj, M., Baghdadi, N., Belaud, G., Zribi, M., Cheviron, B., Courault, D., Hagolle, O., and Charron, F. (2014). Irrigated Grassland Monitoring Using a Time Series of TerraSAR-X and COSMO-SkyMed X-Band SAR Data. Remote Sensing *6*, 10002–10032.

Hambrey, J. (2017). The 2030 Agenda and the Sustainable Development Goals: The challenge for aquaculture development and management. 73.

Harkness, C., Semenov, M.A., Areal, F., Senapati, N., Trnka, M., Balek, J., and Bishop, J. (2020). Adverse weather conditions for UK wheat production under climate change. Agricultural and Forest Meteorology *282–283*, 107862.

He, W., and Yokoya, N. (2018). Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. IJGI *7*, 389.

He, B., Xing, M., and Bai, X. (2014). A Synergistic Methodology for Soil Moisture Estimation in an Alpine Prairie Using Radar and Optical Satellite Data. Remote Sensing *6*, 10966–10985.

Herrmann, F., Baghdadi, N., Blaschek, M., Deidda, R., Duttmann, R., La Jeunesse, I., Sellami, H., Vereecken, H., and Wendland, F. (2016). Simulation of future groundwater recharge using a climate model ensemble and SAR-image based soil parameter distributions — A case study in an intensively-used Mediterranean catchment. Science of The Total Environment *543*, 889–905.

Hidayat, S., Matsuoka, M., Baja, S., and Rampisela, D.A. (2018). Object-Based Image Analysis for Sago Palm Classification: The Most Important Features from High-Resolution Satellite Imagery. Remote Sensing *10*, 1319.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. ArXiv Preprint ArXiv:1503.02531.

Ho Tong Minh, D., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., and Maurel, P. (2018). Deep Recurrent Neural Networks for Winter Vegetation Quality Mapping via Multitemporal SAR Sentinel-1. IEEE Geosci. Remote Sensing Lett. *15*, 464–468.

Houze, R.A., McMurdie, L.A., Petersen, W.A., Schwaller, M.R., Baccus, W., Lundquist, J.D., Mass, C.F., Nijssen, B., Rutledge, S.A., Hudak, D.R., et al. (2017). The Olympic Mountains Experiment (OLYMPEX). Bulletin of the American Meteorological Society *98*, 2167–2188.

Hsu, K., Gao, X., Sorooshian, S., and Gupta, H.V. (1997). Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. Journal of Applied Meteorology *36*, 1176–1190.

Huang, Z., Pan, Z., and Lei, B. (2017). Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. Remote Sensing *9*, 907.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., and Xie, P. (2014). Integrated multi-satellite retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center.

Hunter, M.C., Smith, R.G., Schipanski, M.E., Atwood, L.W., and Mortensen, D.A. (2017). Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. BioScience *67*, 386–391.

Ienco, D., Interdonato, R., Gaetano, R., and Ho Tong Minh, D. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. ISPRS Journal of Photogrammetry and Remote Sensing *158*, 11–22.

Imaoka, K., Maeda, T., Kachi, M., Kasahara, M., Ito, N., and Nakagawa, K. (2012). Status of AMSR2 instrument on GCOM-W1. H. Shimoda, X. Xiong, C. Cao, X. Gu, C. Kim, and A.S. Kiran Kumar, eds. (Kyoto, Japan), p. 852815.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. Remote Sensing *9*, 95.

Interdonato, R., Ienco, D., Gaetano, R., and Ose, K. (2019). DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. ISPRS Journal of Photogrammetry and Remote Sensing *149*, 91–104.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv Preprint ArXiv:1502.03167.

Jägermeyr, J., Pastor, A., Biemans, H., and Gerten, D. (2017). Reconciling irrigated food production with environmental flows for Sustainable Development Goals implementation. Nat Commun *8*, 15900.

Jain, A.K. (1989). Fundamentals of digital image processing (Englewood Cliffs, NJ: Prentice Hall).

Jamshidi, S., Zand-parsa, S., Pakparvar, M., and Niyogi, D. (2019). Evaluation of evapotranspiration over a semiarid region using multiresolution data sources. Journal of Hydrometeorology *20*, 947–964.

Ji, L., and Peters, A.J. (2003). Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. Remote Sensing of Environment *87*, 85–98.

Jolliffe, I.T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A *374*, 20150202.

Joseph, A.T., van der Velde, R., O'Neill, P.E., Lang, R., and Gish, T. (2010). Effects of corn on C- and L-band radar backscatter: A correction method for soil moisture retrieval. Remote Sensing of Environment *114*, 2417–2430.

Joyce, R.J., Janowiak, J.E., Arkin, P.A., and Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology *5*, 487–503.

Kamthonkiat, D., Honda, K., Turral, H., Tripathi, N.K., and Wuwongse, V. (2005). Discrimination of irrigated and rainfed rice in a tropical agricultural system using SPOT VEGETATION NDVI and rainfall data. International Journal of Remote Sensing *26*, 2527–2547.

Karakizi, C., Karantzalos, K., Vakalopoulou, M., and Antoniou, G. (2018). Detailed land cover mapping from multitemporal landsat-8 data of different cloud cover. Remote Sensing *10*, 1214.

Kawabata, A., Ichii, K., and Yamaguchi, Y. (2001). Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. International Journal of Remote Sensing *22*, 1377–1382.

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., and Berger, M. (2001). Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sensing *39*, 1729–1735.

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., et al. (2010). The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. Proc. IEEE *98*, 666–687.

Kidd, C., and Huffman, G. (2011). Global precipitation measurement: Global precipitation measurement. Meteorological Applications *18*, 334–353.

Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. ArXiv Preprint ArXiv:1412.6980.

Koech, R., and Langat, P. (2018). Improving Irrigation Water Use Efficiency: A Review of Advances, Challenges and Opportunities in the Australian Context. Water *10*, 1771.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J. (1998). The tropical rainfall measuring mission (TRMM) sensor package. Journal of Atmospheric and Oceanic Technology *15*, 809–817.

Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 159–174.

Lawston, P.M., Santanello, J.A., and Kumar, S.V. (2017). Irrigation Signals Detected From SMAP Soil Moisture Retrievals: Irrigation Signals Detected From SMAP. Geophys. Res. Lett. *44*, 11,860-11,867.

Le Page, M., Jarlan, L., El Hajj, M.M., Zribi, M., Baghdadi, N., and Boone, A. (2020). Potential for the Detection of Irrigation Events on Maize Plots Using Sentinel-1 Soil Moisture Products. Remote Sensing *12*, 1621.

Liu, Z. (2016). Comparison of Integrated Multisatellite Retrievals for GPM (IMERG) and TRMM Multisatellite Precipitation Analysis (TMPA) Monthly Precipitation Products: Initial Results. Journal of Hydrometeorology *17*, 777–790.

Liu, J., Gong, M., Qin, K., and Zhang, P. (2018). A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. IEEE Trans. Neural Netw. Learning Syst. *29*, 545–559.

Macelloni, G., Paloscia, S., Pampaloni, P., Marliani, F., and Gai, M. (2001). The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops. IEEE Transactions on Geoscience and Remote Sensing *39*, 873–884.

Maselli, F., Battista, P., Chiesi, M., Rapi, B., Angeli, L., Fibbi, L., Magno, R., and Gozzini, B. (2020a). Use of Sentinel-2 MSI data to monitor crop irrigation in Mediterranean areas. International Journal of Applied Earth Observation and Geoinformation *93*, 102216.

Maselli, F., Chiesi, M., Angeli, L., Fibbi, L., Rapi, B., Romani, M., Sabatini, F., and Battista, P. (2020b). An improved NDVI-based method to predict actual evapotranspiration of irrigated grasses and crops. Agricultural Water Management *233*, 106077.

Mattia, F., Thuy Le Toan, Picard, G., Posa, F.I., D'Alessio, A., Notarnicola, C., Gatti, A.M., Rinaldi, M., Satalino, G., and Pasquariello, G. (2003). Multitemporal c-band radar measurements on wheat fields. IEEE Trans. Geosci. Remote Sensing *41*, 1551–1560.

Mattia, F., Balenzano, A., Satalino, G., Lovergine, F., Loew, A., Peng, J., Wegmuller, U., Santoro, M., Cartus, O., Dabrowska-Zielinska, K., et al. (2017). Sentinel-1 high resolution soil moisture. In 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), (Fort Worth, TX: IEEE), pp. 5533–5536.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika *12*, 153–157.

Mérot, A. (2007). Analyse et modélisation du fonctionnement biophysique et décisionnel d'un système prairial irrigué-Application aux prairies plurispécifiques de Crau en vue de l'élaboration d'un Outil d'Aide à la Décision.

Mila, A.J., Ali, Md.H., Akanda, A.R., Rashid, Md.H.O., and Rahman, Md.A. (2017). Effects of deficit irrigation on yield, water productivity and economic return of sunflower. Cogent Food & Agriculture *3*, 1287619.

Naeimi, V., Bartalis, Z., and Wagner, W. (2009). ASCAT Soil Moisture: An Assessment of the Data Quality and Consistency with the ERS Scatterometer Heritage. Journal of Hydrometeorology *10*, 555–563.

Nair, V., and Hinton, G.E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814.

Narvekar, P.S., Entekhabi, D., Kim, S.-B., and Njoku, E.G. (2015). Soil moisture retrieval using L-band radar observations. IEEE Transactions on Geoscience and Remote Sensing *53*, 3492–3506.

Naser, M.A., Khosla, R., Longchamps, L., and Dahal, S. (2020). Using NDVI to differentiate wheat genotypes productivity under dryland and irrigated conditions. Remote Sensing *12*, 824.

Nasrallah, A., Baghdadi, N., El Hajj, M., Darwish, T., Belhouchette, H., Faour, G., Darwich, S., and Mhawej, M. (2019). Sentinel-1 Data for Winter Wheat Phenology Monitoring and Mapping. Remote Sensing *11*, 2228.

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., and Hossard, L. (2018). Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. Remote Sensing *10*, 1217.

Omranian, E., and Sharif, H.O. (2018). Evaluation of the Global Precipitation Measurement (GPM) Satellite Rainfall Products over the Lower Colorado River Basin, Texas. JAWRA Journal of the American Water Resources Association *54*, 882–898.

Ozdogan, M. (2011). Exploring the potential contribution of irrigation to global agricultural primary productivity: IRRIGATION AND PRIMARY PRODUCTIVITY. Global Biogeochem. Cycles *25*, n/a-n/a.

Ozdogan, M., and Gutman, G. (2008). A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. Remote Sensing of Environment *112*, 3520–3537.

Ozdogan, M., Yang, Y., Allez, G., and Cervantes, C. (2010). Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. Remote Sensing *2*, 2274–2304.

Pageot, Y., Baup, F., Inglada, J., Baghdadi, N., and Demarez, V. (2020). Detection of Irrigated and Rainfed Crops in Temperate Areas Using Sentinel-1 and Sentinel-2 Time Series. Remote Sensing *12*, 3044.

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., and Reppucci, A. (2013). Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. Remote Sensing of Environment *134*, 234–248.

Paredes, P., Rodrigues, G.C., Alves, I., and Pereira, L.S. (2014). Partitioning evapotranspiration, yield prediction and economic returns of maize under various irrigation management strategies. Agricultural Water Management *135*, 27–39.

Pellarin, T., Louvet, S., Gruhier, C., Quantin, G., and Legout, C. (2013). A simple and effective method for correcting soil moisture and precipitation estimates using AMSR-E measurements. Remote Sensing of Environment *136*, 28–36.

Pelletier, C., Webb, G., and Petitjean, F. (2019). Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. Remote Sensing *11*, 523.

Pervez, M.S., and Brown, J.F. (2010). Mapping irrigated lands at 250-m scale by merging MODIS data and national agricultural statistics. Remote Sensing *2*, 2388–2412.

Pokhrel, Y.N., Hanasaki, N., Wada, Y., and Kim, H. (2016). Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models: Recent progresses in incorporating human land-water management into global land surface models. WIREs Water *3*, 548–574.

Portmann, F.T., Siebert, S., and Döll, P. (2010). MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling: MONTHLY IRRIGATED AND RAINFED CROP AREAS. Global Biogeochem. Cycles *24*, n/a-n/a.

Potter, C.S., and Brooks, V. (1998). Global analysis of empirical relations between annual climate and seasonality of NDVI. International Journal of Remote Sensing *19*, 2921–2948.

Powers, D.M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Ren, J., Shao, Y., Wan, H., Xie, Y., and Campos, A. (2021). A two-step mapping of irrigated corn with multi-temporal MODIS and Landsat analysis ready data. ISPRS Journal of Photogrammetry and Remote Sensing *176*, 69–82.

Richardson, K.J., Lewis, K.H., Krishnamurthy, P.K., Kent, C., Wiltshire, A.J., and Hanlon, H.M. (2018). Food security outcomes under a changing climate: impacts of mitigation and adaptation on vulnerability to food insecurity. Climatic Change *147*, 327–341.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing *67*, 93–104.

Salmon, J.M., Friedl, M.A., Frolking, S., Wisser, D., and Douglas, E.M. (2015). Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data. International Journal of Applied Earth Observation and Geoinformation *38*, 321–334.

Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Schmied, H.M., Van Beek, L.P., Wiese, D.N., Wada, Y., Long, D., and Reedy, R.C. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proceedings of the National Academy of Sciences *115*, E1080–E1089.

Schaldach, R., Koch, J., Aus der Beek, T., Kynast, E., and Flörke, M. (2012). Current and future irrigation water requirements in pan-Europe: An integrated analysis of socio-economic and climate scenarios. Global and Planetary Change *94–95*, 33–45.

Schwerdt, M., Schmidt, K., Tous Ramon, N., Klenk, P., Yague-Martinez, N., Prats-Iraola, P., Zink, M., and Geudtner, D. (2017). Independent System Calibration of Sentinel-1B. Remote Sensing *9*, 511.

Segarra, J., Buchaillot, M.L., Araus, J.L., and Kefauver, S.C. (2020). Remote sensing for precision agriculture: Sentinel-2 improved features and applications. Agronomy *10*, 641.

Sellami, H., La Jeunesse, I., Benabdallah, S., Baghdadi, N., and Vanclooster, M. (2014). Uncertainty analysis in model parameters regionalization: a case study involving the SWAT model in Mediterranean catchments (Southern France). Hydrology and Earth System Sciences 2393–2413.

Sharma, A.K., Hubert-Moy, L., Sriramulu, B., Sekhar, M., Ruiz, L., Bandyopadhyay, S., Mohan, S., and Corgne, S. (2019). Evaluation of Radarsat-2 quad-pol SAR time-series images for monitoring groundwater irrigation. International Journal of Digital Earth *12*, 1177–1197.

Shi, J., Jiang, L., Zhang, L., Chen, K.S., Wigneron, J.P., Chanzy, A., and Jackson, T.J. (2006). Physically Based Estimation of Bare-Surface Soil Moisture With the Passive Radiometers. IEEE Transactions on Geoscience and Remote Sensing *44*, 3145–3153.

Shi, Y., Hwang, M.-Y., Lei, X., and Sheng, H. (2019). Knowledge Distillation For Recurrent Neural Network Language Modeling With Trust Regularization. ArXiv:1904.04163 [Cs].

Shiklomanov, I.A. (2000). Appraisal and Assessment of World Water Resources. Water International *25*, 11–32.

Siebert, S., and Döll, P. (2010). Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. Journal of Hydrology *384*, 198–217.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., and Feick, S. (2005). Development and validation of the global map of irrigation areas. Hydrol. Earth Syst. Sci. *9*, 535–547.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B.R. (2015). A global data set of the extent of irrigated land from 1900 to 2005. Hydrol. Earth Syst. Sci. *19*, 1521–1545.

Skofronick-Jackson, G., Berg, W., Kidd, C., Kirschbaum, D.B., Petersen, W.A., Huffman, G.J., and Takayabu, Y.N. (2018). Global Precipitation Measurement (GPM): Unified Precipitation Estimation from Space. In Remote Sensing of Clouds and Precipitation, C. Andronache, ed. (Cham: Springer International Publishing), pp. 175–193.

Srivastava, H.S., Patel, P., Sharma, K.P., Krishnamurthy, Y.V.N., and Dadhwal, V.K. (2011). A semi-empirical modelling approach to calculate two-way attenuation in radar backscatter from soil due to crop cover. Current Science 1871–1874.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.-A., and Rahman, A. (2020). Land-use land-cover classification by machine learning classifiers for satellite observations—A review. Remote Sensing *12*, 1135.

Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Introduction to data mining (Boston: Pearson Addison Wesley).

Tebbs, E., Gerard, F., Petrie, A., and De Witte, E. (2016). Emerging and Potential Future Applications of Satellite-Based Soil Moisture Products. In Satellite Soil Moisture Retrieval, (Elsevier), pp. 379–400.

Thenkabail, P., Dheeravath, V., Biradar, C., Gangalakunta, O.R., Noojipady, P., Gurappa, C., Velpuri, M., Gumma, M., and Li, Y. (2009a). Irrigated Area Maps and Statistics of India Using Remote Sensing and National Statistics. Remote Sensing *1*, 50–67.

Thenkabail, P.S., Schull, M., and Turral, H. (2005). Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. Remote Sensing of Environment *95*, 317–341.

Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turral, H., Cai, X., et al. (2009b). Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing *30*, 3679–3733.

Tiercelin, J.-R. (2006). Traité d'irrigation (Paris: Tec & Doc).

Tilman, D., and Clark, M. (2015). Food, Agriculture & the Environment: Can We Feed the World & Save the Earth? Daedalus *144*, 8–23.

Tilman, D., Balzer, C., Hill, J., and Befort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences *108*, 20260–20264.

Ulaby, F.T. (1982). Microwave remote sensing active and passive. Rader Remote Sensing and Surface Scattering and Emission Theory 848–902.

UN Water (2018). Sustainable Development Goal 6: synthesis report 2018 on water and sanitation (New York, New York, United States of America: United Nations).

Van doninck, J., Peters, J., Lievens, H., De Baets, B., and Verhoest, N.E.C. (2012). Accounting for seasonality in a soil moisture change detection algorithm for ASAR Wide Swath time series. Hydrology and Earth System Sciences *16*, 773–786.

Vörösmarty, C.J., Green, P., Salisbury, J., and Lammers, R.B. (2000). Global water resources: vulnerability from climate change and population growth. Science *289*, 284–288.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., et al. (2010). Global threats to human water security and river biodiversity. Nature *467*, 555–561.

Wada, Y., Wisser, D., and Bierkens, M.F.P. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dynam. *5*, 15–40.

Wagner, W., Noll, J., Borgeaud, M., and Rott, H. (1999). Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer. IEEE Trans. Geosci. Remote Sensing *37*, 206–216.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., et al. (2013). The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. Metz *22*, 5–33.

Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., and Wu, X. (2018). Artificial Mangrove Species Mapping Using Pléiades-1: An Evaluation of Pixel-Based and Object-Based Classifications with Selected Machine Learning Algorithms. Remote Sensing *10*, 294.

Wardlow, B.D., and Egbert, S.L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. Remote Sensing of Environment *112*, 1096–1116.

Whitcraft, A., Becker-Reshef, I., and Justice, C. (2015). A Framework for Defining Spatially Explicit Earth Observation Requirements for a Global Agricultural Monitoring Initiative (GEOGLAM). Remote Sensing *7*, 1461–1481.

Whitcraft, A.K., Becker-Reshef, I., Justice, C.O., Gifford, L., Kavvada, A., and Jarvis, I. (2019). No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework. Remote Sensing of Environment *235*, 111470.

White, J.W., Hoogenboom, G., Kimball, B.A., and Wall, G.W. (2011). Methodologies for simulating impacts of climate change on crop production. Field Crops Research *124*, 357–368.

Wulder, M.A., Hall, R.J., Coops, N.C., and Franklin, S.E. (2004). High spatial resolution remotely sensed data for ecosystem characterization. BioScience *54*, 511–521.

Xiang, K., Ma, M., Liu, W., Dong, J., Zhu, X., and Yuan, W. (2019). Mapping Irrigated Areas of Northeast China in Comparison to Natural Vegetation. Remote Sensing *11*, 825.

Yin, X., Yu, X., Sohn, K., Liu, X., and Chandraker, M. (2019). Feature Transfer Learning for Deep Face Recognition with Under-Represented Data. ArXiv:1803.09014 [Cs].

Zhao, B., Huang, B., and Zhong, Y. (2017). Transfer Learning With Fully Pretrained Deep Convolution Networks for Land-Use Classification. IEEE Geosci. Remote Sensing Lett. *14*, 1436–1440.

Zhou, Z., Shin, J., Zhang, L., Gurudu, S., Gotway, M., and Liang, J. (2017). Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Honolulu, HI: IEEE), pp. 4761–4772.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. *5*, 8–36.