

POLITECNICO
MILANO 1863

Rice cropland monitoring and mapping in the Po river catchment based on Sentinel-1-derived time-series phenology parameters.

TESI DI LAUREA MAGISTRALE IN
GEOINFORMATICS ENGINEERING - INGEGNERIA GEOINFOR-
MATICA

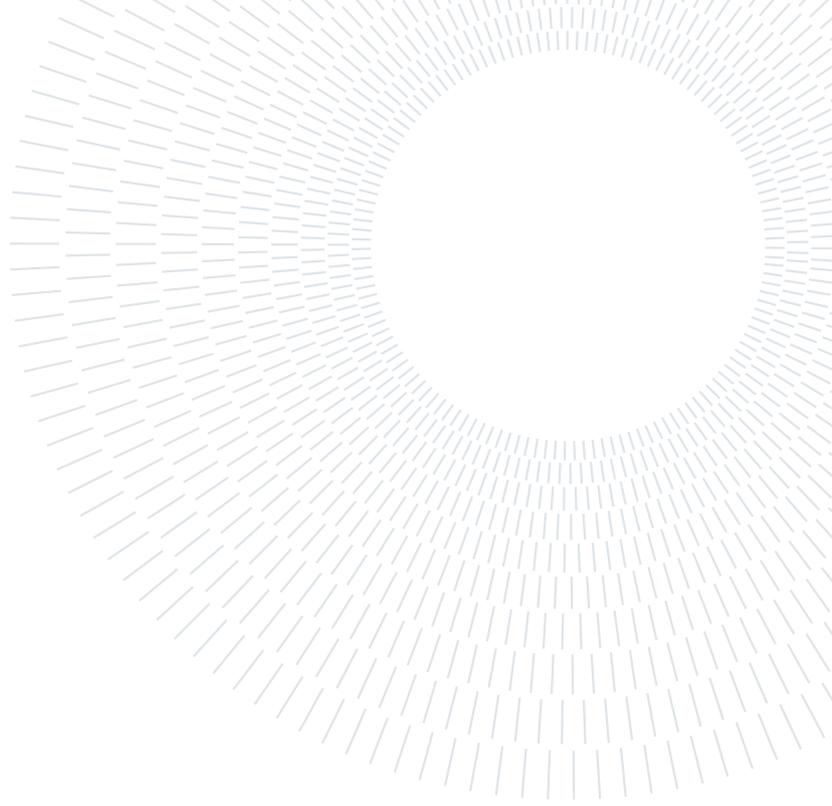
Author: **Katayoun Fakherifard**

Student ID: 10662650

Advisor: Prof. Wolfgang Wagner and Prof. Giovanna Venuti

Co-advisors: Isabella Greimeister-Pfeil

Academic Year: 2022-2021



To My Beloved Wonderful Parents

Laya and Mahmood

**without whom I could not be able to
fly towards my dreams**

Abstract

Rice cultivation is one of the most vital agricultural activities in the world for human beings because it is a staple food for us. One of the ways to monitor the distribution of this valuable grain and collect cultivation statistics is through rice-based detection of cultivation patterns around the world or in smaller regions by government agency authorities, by applying remote sensing-based rice detection algorithms to the region of interest. SAR-based rice monitoring has shown remarkable results over time in distinguishing rice cropland from other types of land cover because of its high sensitivity to flooded surfaces. For this study, a dataset with high spatial and temporal resolution was used. Multi-temporal SAR data analysis is the best known approach for rice area analysis and detection. The annual backscatter variation of rice acreage is higher than other crops during the rice growing season, and this is the key principle behind the rice monitoring algorithms.

In this thesis, we applied the algorithm developed by "Nguyen and Wagner" [40] to SAR multi-temporal VV and VH-polarised datasets from 2019 and 2020, aiming to distinguish rice acreage distribution from the other crop types. We improved the parameterization of the algorithm for the Po river catchment and obtained a rice map with about 88% precision, 99% overall accuracy, 77% F1-score and 0.78 Kappa. Moreover, we demonstrated that the VH-polarisation time series is more efficient than the VV-polarisation in rice seasonality detection and mapping. We also developed further steps by analyzing the local extrema by segmenting the rice growing season into subsections and analyzing them separately to improve the original algorithm. This method was applied in the valleys of the Po River located in northern Italy, which are rich in rice cultivation with a high distribution of rice paddies.

Keywords: rice, paddy, remote sensing, remote sensing, SAR, Sentinel-1, polarization, VV and VH polarization, temporal and spatial resolution, multi temporal, backscatter

Abstract in lingua italiana

La coltivazione del riso è una delle attività agricole più vitali al mondo per l'uomo, perché è un alimento di base per noi. Uno dei modi per monitorare la distribuzione di questo prezioso cereale e raccogliere statistiche sulla coltivazione è il rilevamento basato sul riso dei modelli di coltivazione in tutto il mondo o in regioni più piccole da parte delle autorità governative, applicando algoritmi di rilevamento del riso basati sul telerilevamento alla regione di interesse. Il monitoraggio del riso basato sul SAR ha mostrato nel tempo risultati notevoli nel distinguere le coltivazioni di riso da altri tipi di copertura del suolo, grazie alla sua elevata sensibilità alle superfici allagate. Per questo studio è stato utilizzato un set di dati ad alta risoluzione spaziale e temporale. L'analisi dei dati SAR multitemporali è l'approccio più conosciuto per l'analisi e il rilevamento delle aree risicole. La variazione annuale della retrodiffusione della superficie coltivata a riso è più elevata rispetto alle altre colture durante la stagione di crescita del riso, e questo è il principio chiave alla base degli algoritmi di monitoraggio del riso.

In questa tesi, abbiamo applicato l'algoritmo sviluppato da "Nguyen e Wagner" [40] ai dataset SAR multi-temporali VV e VH-polarizzati del 2019 e 2020, con l'obiettivo di distinguere la distribuzione della superficie coltivata a riso dagli altri tipi di colture. Abbiamo migliorato la parametrizzazione dell'algoritmo per il bacino idrografico del Po e abbiamo ottenuto una mappa del riso con una precisione dell'88% circa, un'accuratezza complessiva del 99%, un F1-score del 77% e un Kappa di 0,78. Inoltre, abbiamo dimostrato che la serie temporale con polarizzazione VH è più efficiente della polarizzazione VV nel rilevamento e nella mappatura della stagionalità del riso. Abbiamo anche sviluppato ulteriori passi analizzando gli estremi locali, segmentando la stagione di coltivazione del riso in sottosezioni e analizzandole separatamente per migliorare l'algoritmo originale. Questo metodo è stato applicato nelle valli del fiume Po, situate nell'Italia settentrionale, ricche di coltivazioni di riso e con un'elevata distribuzione di risaie.

Keywords: riso, risaia, telerilevamento, telerilevamento, SAR, Sentinel-1, polarizzazione, polarizzazione VV e VH, risoluzione temporale e spaziale, multi temporale, backscatter

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
1 Introduction	1
1.1 Statement of the Research Problem	4
1.2 Physical Mechanism and Equations of Scattering	5
1.2.1 Radar-based remote sensing	5
1.2.2 Radar Equation	6
1.3 Rice Growing Characteristics	7
1.4 Remote Sensing of Rice Cropland	8
1.5 Literature Review	9
1.6 Research Objective and Outline	15
1.6.1 Research questions	15
1.6.2 Research Hypothesis or Propositions	16
1.7 Scientific Impact	17
2 Data and Study Area	19
2.1 Sentinel-1 Timeseries Dataset	19
2.2 Reference Data (existing survey data)	20
2.3 Study Area	21
2.4 Rice Cultivation Calendar in Study Area	24
2.5 Software and Tools	27
2.5.1 Methods and data analysis	27
3 Methodology	31
3.1 Approach Methodology	32

3.2	Preprocessing	33
3.2.1	Apply Orbit File	33
3.2.2	Radiometric Calibration	33
3.2.3	Range Doppler Terrain Correction (Geocoding)	34
3.2.4	Resampling (Speckle Filtering)	34
3.2.5	Conversion to dB	35
3.3	Orbit correction	35
3.4	Time Series Smoothing	36
3.5	Extraction of the statistical parameters of the Sentinel-1 backscatter	41
3.6	Parameters Definition and Temporal Rice Classification	48
3.7	Threshold Optimization (decision-rule setting)	52
3.8	Validation and accuracy assessment	54
4	Results	55
4.1	Final Results	55
4.2	Comparison to Nguyen and Wagner's Approach	60
4.3	Improving the Classification Approach by the Analysis of Local Extrema	61
5	Discussion	63
5.1	Logic behind the algorithm	63
5.2	Rice Cropland Mapping Final Results	63
5.3	Flaws, Failures and Limitations of this Decision-Rule-Based Algorithm	67
5.4	Polarization Comparison and Selection	68
5.5	Comparing 2 years time series (2019 & 2020)	70
6	Conclusions and future developments	73
	Bibliography	75
	List of Figures	83
	List of Tables	87
	List of Symbols	90
	List of Abbreviations	92

1 | Introduction

Rice is one of the extensively consumed food resources for humans. Accordingly, rice cultivation occupies one of the first positions as an important agricultural activity in the agricultural sector to produce a staple food for the growing world population in all continents, based on their cultural and dietary habits. Nowadays, the agricultural sector is facing great challenges due to various factors such as global demand for food, increasing price competition due to the impact of globalization of markets and volatility of food prices [45], and the need for more environmentally and economically sustainable farming system in developed countries [46].

Stable and sustainable crop production requires information on where, when, how, and what type of crops are grown. Therefore, the development of an up-to-date and accurate information system on the distribution of agricultural fields and their species is not only a study case for food security and agricultural planning, but is also important in the context of other scientific disciplines, including water resource management, marketing, methane emissions relevant to the greenhouse effect, and climate change.

Italy is the largest rice producer in Europe, accounting for more than half of the total rice production in the European Union (EU) [54]. Therefore, the preliminary and precise monitoring of rice is nowadays critical due to its importance for the governmental and private interdisciplinary sectors directly or indirectly associated to rice cultivation from an agronomic point of view. For this reason, nowadays many agricultural or cultivation information systems have been developed by different organs at continental or national level, a system that includes the subsystem of soil, plants, crops, pesticides and insects subsystems, which assumingly converts solar energy, water, nutrients, labor and all other variables into food, feed and fuel (FAO 1996 [13]). Accordingly, any type of record that takes into account the temporal and spatial location of rice acreage is important information for the economy and agricultural facets in all directions. Due to this reason, the Food and Agriculture Organization of the United Nations (FAO) publishes annual and monthly statistics on the production, import and export of various agricultural crops and other food resources, as well as other useful statistics that can be used for decision-making

and management (www.fao.org/faostat/). Another reference layer for rice acreage is the classified maps of the Corine Land Cover (CLC) product with a spatial resolution of 100 m, which are updated every 4 years. Because of its low temporal and spatial resolution, in most cases it does not meet the requirements of many users who desire very precise and high quality analysis.

Most of the countries have developed their own customized information system to monitor rice production, including Italy, by assessing the spatial distribution of rice acreage. IRES (Italian Rice Experiment Station: www.ires.online/), AGEA or SIAN (Italian Informative Agricultural System: www.sian.it/portale-sian/home.jsp), and ersaf (Regional Agency for Agricultural and Forestry Services [9] - is an institution of the Lombardy Region) have been developed. The information system of these organizations has the mission to carry out technical and promotional actions for the development of the agricultural and forestry sector and rural territory, promoting transversality, multifunctionality and integration. In addition, the Earth Observation Global Agricultural Monitoring, GEOGLAM (<https://earthobservations.org/index.php>) was established to develop and strengthen global crop production to increase market transparency and improve food security by producing and disseminating relevant, annually updated, and actionable information on agricultural conditions and production prospects at the national, regional, and global levels through the coordination of satellite observation systems in different regions of the world.

Earth observation has proven to be a useful method for rice mapping at larger scales by using geospatial data, e.g., remote sensing data, weather data, and in particular Synthetic Aperture Radar (SAR) data, which is one of the most powerful remote sensing tools for a variety of applications, including agricultural purposes and rice cropland monitoring and detection. Unlike optical technology, synthetic aperture radar can make observations through darkness, clouds, and rain to detect changes in habitat, water and moisture levels, effects of natural or human disturbances, and changes in the Earth's surface after events such as earthquakes or the opening of holes in the Earth. Due to the day-night and weather-independent data collection capabilities of the SAR sensors, they became one of the most popular data sources to be consulted and used by researchers for a variety of research topics. The launch of the European Space Agency (ESA) Sentinel-1A and B missions with the SAR C-band sensor mounted on it, on April 3, 2014, and April 25, 2016, with a revisit period of 6 days, was a turning point in most studies and industries, as it provided data with improved spatial and temporal resolution (images with a spatial resolution of 10m approximately every 1 to 4 days over Europe). This is one of the reasons why the data from SAR are best suited to monitor and map the distribution of rice fields

and accelerate the progress by adopting appropriate strategies and algorithms. The main advantage of SAR is the high temporal resolution (because clouds are no problem) and the very characteristic low backscatter signal of flooded surfaces. This open-source data is available free of charge through the Copernicus website.

In the context of remote sensing of rice croplands, many studies and papers have confirmed that the use of dense SAR time series has very efficient impacts and results. To this end, using multitemporal SAR datasets is a tactical idea to retrieve classified rice maps or the rice growing cycle consistently at a larger scale, according to the temporal variation of the SAR backscatter signal, Sigma-naught (σ_0) [22]. Due to the fact that the temporal variation of σ_0 throughout the rice growing season for rice cropland is very high compared to the other crops or land covers [20] i.e., the standing water typically contains a very low backscatter value, which increases quickly as soon as there is vegetation. Many classification algorithms have been introduced for rice identification so far. The main principle of these classifiers is based on time series analysis of SAR backscatter values for both cross-polarized (VH) and co-polarized (VV) or even the ratio of both (HH/VV, HH/HV, or VV /VH) [51]. Most studies agree that cross-polarized (VH) backscatter values are more strongly correlated with rice growth than the others.

The main objective of this thesis is to monitor and classify the distribution of rice fields in the Po river catchment by applying a phenology-based classification algorithm developed by "Ngyuen and Wagner" [41] [40] [39] [38], The functionality and transferability of the original algorithm to our study area is evaluated by defining a specified and customized rice growing calendar over this region. The final aim is to adjust the parameterization of the original algorithm for optimal results in the Po river catchment. To achieve this goal, we used a dense time series of SAR backscatter images for 2019 and 2020 in both VV and VH polarization mode over the Po catchment, where contains very large area of rice paddies in both Lombardy and Piedmont regions, with the highest rice production rates in Italy Figure 2.3.

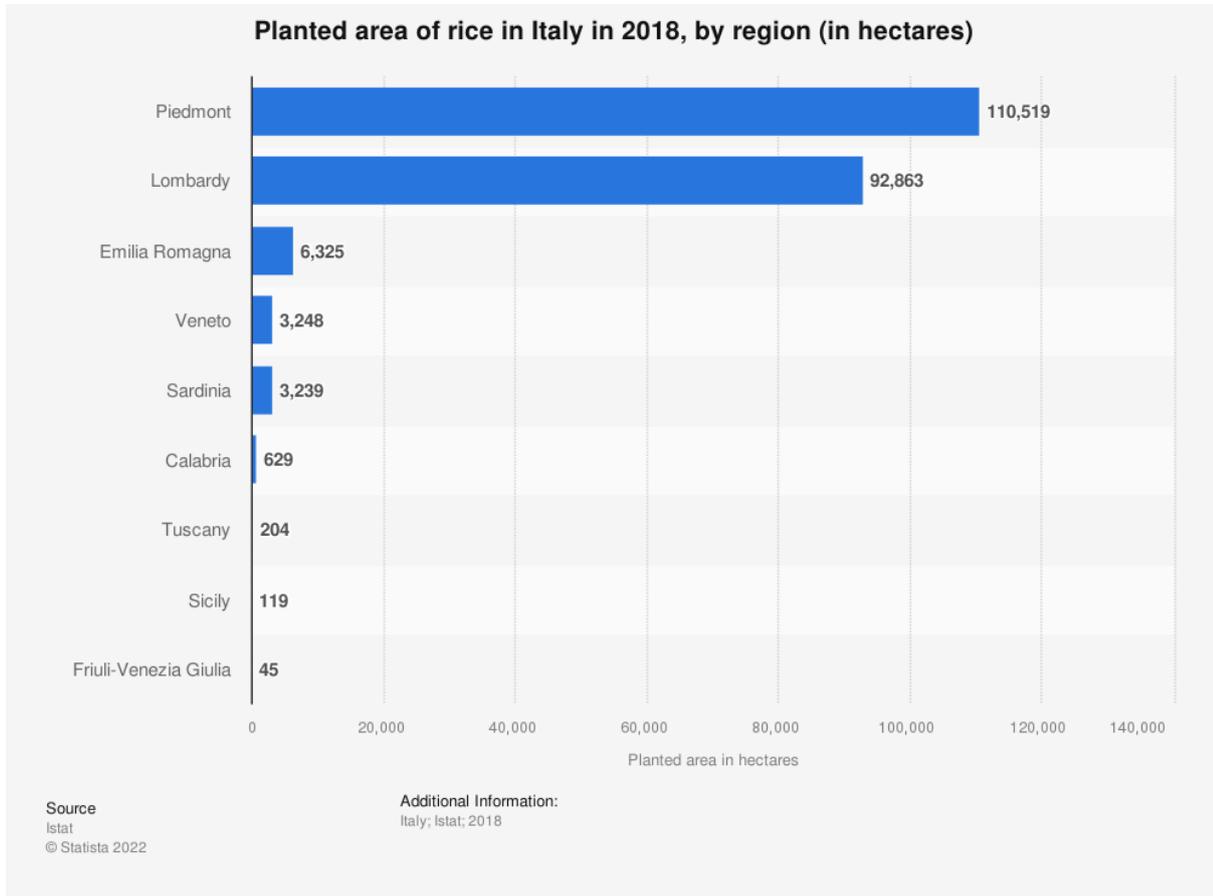


Figure 1.1: Planted area of rice in Italy in 2018 according to istat website.

1.1. Statement of the Research Problem

One of the ultimate goals of this thesis is to introduce an already proposed algorithm for the classification of rice by reapplying it to a particular area in Italy, important for us, in order to make further decisions. By using C-band SAR time series as input and evaluating the applicability of the propounded phenology-based rice classification algorithm (first proposed by "Nguyen et al" [39]) and readjusting it with our observed variables based on the conditions over the study area. And finally, improving the algorithm by adding extra steps in order to achieve higher reliability of the final results, using some validation techniques with specific reference data for the rice growing layer. At the beginning of the research, numerous questions initially emerged, including the following:

- 1) What type of dataset is more proficient for rice monitoring and mapping?
- 2) Is the choice of dataset dependent on the region, its characteristics, and scale?

- 3) Why are SAR time series suitable for monitoring rice cultivation in this particular area?
- 4) How can we process the SAR time series data in order to ensure the reliability of the final results with high accuracy level?
- 5) What useful information or parameters can we extract from the stacked SAR dataset?
- 6) Why is a phenology-based algorithm applicable for rice paddy classification and analysis?
- 7) Which phenological parameters are suitable for monitoring and classification of rice?
- 8) Does this algorithm have shortcomings with the proposed methodology or what are the main aspects that lead to typical errors in the final results?
- 9) Is this algorithm applicable at larger scales or is it better to limit to local or minuscule scales.

There were many questions we asked ourselves during this research, and tried to find reasonable answers to most of them from a researcher's point of view, based on the experience and information we gained during the conduction procedure, not only using the already published works, but also by putting together all the gained knowledge and reasoning in a rational way.

1.2. Physical Mechanism and Equations of Scattering

1.2.1. Radar-based remote sensing

A satellite-borne Synthetic Aperture Radar (SAR) scans the Earth's surface using microwave radiation from a great height above the Earth (500-1000Km). The SAR antenna emits electromagnetic pulses and receives the backscattered echo from the Earth's surface. Based on these echoes, high-resolution image data and other data products can be generated.

Microwave remote sensing enables the use of electromagnetic radiation with a wavelength between 1cm-1m. This application gives us the ability to discriminate between different targets on the ground because the signal intensity is different for each specific object with exclusive physical and chemical intrinsic properties. The time between the transmitted signal and the received signal is used to measure the distance of the target, which is

used in many analyses, for example, in the context of forest monitoring to monitor the density of the canopy of different tree species. One of the most important measurements is the strength of the backscattered signal, which plays the most important role in remote sensing-based analyses. A system that operates in this way is called RADAR (the name stands for RADio Detection And Ranging) and can enable a "microwave image" of the observed scene to be obtained. The received backscatter signal can be considered as a combination of basic scattering mechanisms [38]. Different applications to different targets within our field of interest require their own interpretation of the characterization of the scattering mechanism. In our case study, rice acreage monitoring, we will use a very specific scattering mechanism that will be discussed in more detail in the following chapters.

Due to their wavelength, the interaction of microwaves with the target object, such as soil, water bodies, and plants, is very different from the other frequencies of electromagnetic radiation, including infrared and optical waves.

1.2.2. Radar Equation

Theoretically, the radar equation describes the physical power budget of the signal scattered by the satellite's transmitters and received by the receivers, and their correlation with the characteristics of the radar.

$$P_r = \frac{P_t G_t A_r}{(4\pi R_t R_r)^2} G_s A_s (1 - K_a) \quad (1.1)$$

Here P_t and P_r are the transmitted and received signal intensities of the antenna with gain G_t . A_r is the effective aperture of the receiving antenna and A_s is the effective area of the incident beam intercepted by the transmitter. K_a is the fraction absorbed by the target with gain G_t . R_t is the distance between the radar transmitter and the target. R_r is the distance between the target and the radar receiver [38]. After some more signal processing steps on the transmitted and received signals, the formula for the backscatter cross-section (the most important formula for processing the raw satellite data) is expressed as follows:

$$\sigma_0(dB) = 10 \log \sigma_0(m^2 m^{-2}) \quad (1.2)$$

The coefficient σ_0 is the scattering coefficient, the conventional measure of the strength

of radar signals reflected from a distributed scatterer, usually expressed in dB, the last value used in remote sensing data for further interpretation, and it expresses the reflective backscattering properties of the surface.

In addition, the ratio of normalised and corrected signal intensity coefficient of a pixel, also known as "Digital Number (DN)", is a dimensionless number that compares the observed strength with that expected for an area of one square metre. Sigma-naught (σ_0) is defined with respect to the nominally horizontal plane and generally varies significantly with the incidence angle, wavelength, and polarisation, as well as with the properties of the scattering surface itself for further uses, using the following simplified formula:

$$\sigma_0(dB) = 10 \log(DN) \quad (1.3)$$

which depends on the physical characteristics of the target surface, such as geometric structure, surface roughness, and orientation; and on the electrical characteristics of the surface, such as dielectric constant, conductivity of the matter, and moisture content; and on radar characteristics, such as frequency, polarization (VV, VH, HH, and HV), and the emitted incidence angle [56]. Finally, this σ_0 is the very last parameter used to obtain the required geophysical and biophysical parameters.

1.3. Rice Growing Characteristics

Each agricultural crop has a characteristic cropping pattern and many different algorithms have been introduced to distinguish one from another, including rice. Rice cropland is the most common agricultural land cover in the Lombardy and Piedmont regions in Italy. In general, rice cultivation includes 3 main cultivation phases: vegetative, reproductive, and ripening (maturity) or harvesting phases [40]. The cultivated rice plant is an annual grass and grows to a height of about 1.2 metres. Figure 1.2 briefly depicts the phases that a rice plant goes through. The first phase, the vegetative stage, includes the sowing (seeding), flooding, tillage, germination, and transplanting stages. The duration of the vegetative stage varies by cultivar and depends on many factors such as temperature, soil conditions, and rice variety, but usually ranges from 40 to 100 days or more. The second phase that follows, the reproductive phase, includes panicle instigation (a panicle is a cluster of flowers that grows at the end of a branch or shoot), head formation (heading), and flowering, which takes approximately 35 days to complete. Irrigation continues during this phase, but as the rice plants begin to flower and their stems elongate, the water surface

disappears from view. Ripening is the final phase when the final product of rice, the white grain, can be harvested. During this phase, irrigation is discontinued and water completely disappears from the soil during harvest. This stage usually lasts 30 days depending on different variables "Nelson et al, 2014 [37]" (Figure 1.2).

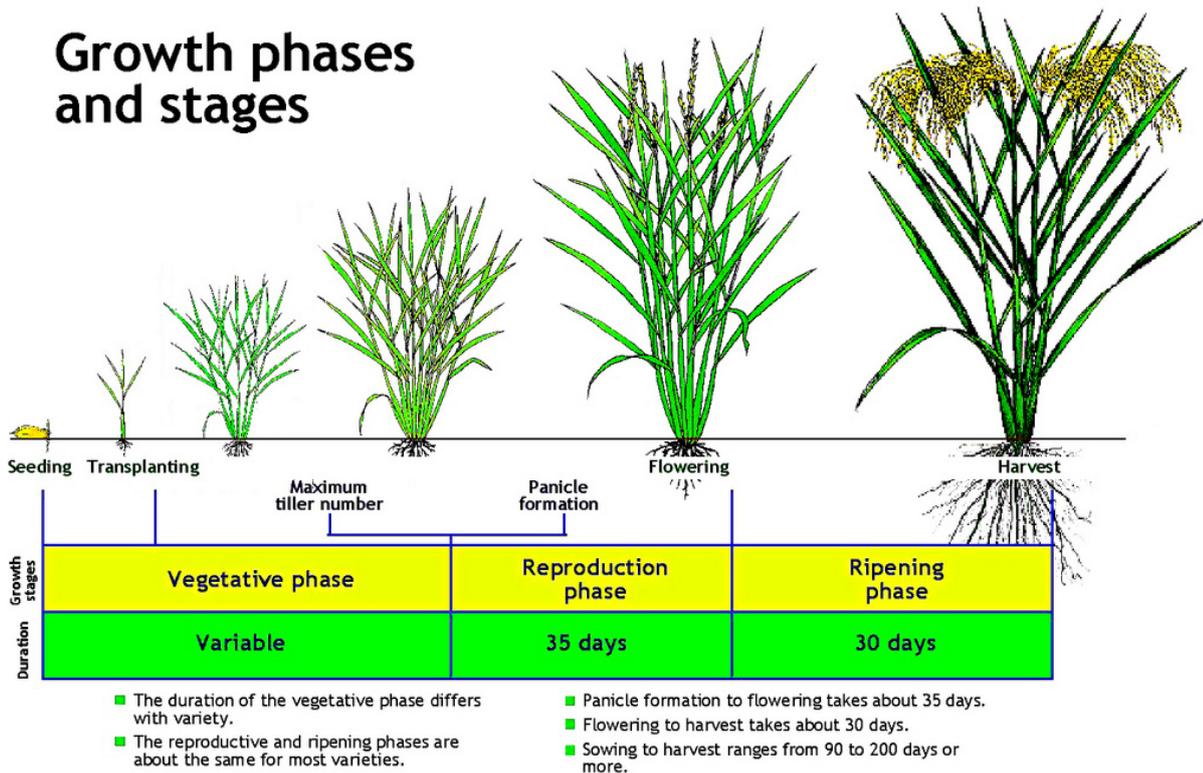


Figure 1.2: Rice growth phases (source: www.flickr.com). This scheme is the outline of different physiological phases that the rice grain goes through during its growing season. for the most part, the entire procedure takes 90 to 200 days or more, from sowing to harvest. However it might vary from time to time and region by region (depending on so many factors).

1.4. Remote Sensing of Rice Cropland

Rice is one of the staple food resources in Italy, economically, culturally and agriculturally. Therefore, mapping and monitoring the distribution of this momentous crop is of vital importance to the Italian government. This valuable grain goes through different physiological phases during its cultivation, which makes it exclusive from other crops. Determining these phases and detecting and mapping paddy fields requires analysis over a dense time series, and the time series data from SAR accomplish this task. On the

grounds that, Sentinel-1 has a low revisit time and weather-independent property, which is exactly what we are looking for. One of the very first studies using SAR sensors to map rice distribution was performed with the European Remote Sensing Satellite (ERS-1) concluding highly accurate and promising results [1] [23]. The mechanism of active remote sensing-based rice mapping works with the signal-plant-water interactions at different phases of rice cultivation, i.e., large dynamics of microwave backscatter values can be observed and mapped in rice crops. These backscatter values change at different phases from seeding and flooding to preharvest and harvest stages. The signal backscatter over a flooded rice growing area is very low compared to a vegetated area, due to the fact that water is an absorber of microwave signals and backscatter coefficients are very low thanks to the specular reflection of the flooded area. With the evolving of the vegetative, reproduction, ripening (maturity) and harvesting stages, a rapid increase in radar backscatter values was observed, which can be attributed to the increasing of height and density of rice plants in the rice paddies. These different stages of rice cultivation process coincide with changes in the interaction of plants with radar backscatter and can therefore be observed and mapped with an appropriate set of remote sensing data.

By putting all the pieces of the puzzle together [59] [58], we can reshape a general model to explain and map rice acreage via backscatter measurements over time. The fact that the annual backscatter dynamic range (i.e. MAX minus MIN backscatter) of rice is higher than for other crops, is the theory and basis of most rice detection algorithms, including our algorithm based on "Nguyen and Wagner" algorithm [39]. This is another reason why we need a dense time series SAR data to capture this annual variation. After defining the principles of the active remote sensing-based rice mapping model, the other important step is the selection of an efficient dataset which is compatible with the study area and the algorithm to be applied on.

1.5. Literature Review

Considering the vitality of rice from various aspects, monitoring of rice cultivation fields over time by acquiring significant cropping patterns of this grain is indispensable to have a well organized agricultural information system. There are many different methods that have been developed long ago until today, including different types of data collection, which is traditionally done by ground surveying the fields, which is labor intensive and costly, covering a large area [44] [62]. These methods are not very convenient and the areas

are not compatible with current market and information system requirements to monitor the distribution of rice fields over time. With the development of complex remote sensing-based algorithms and techniques, crop monitoring has been modernized over time and has become more cost-effective with more reliable results over even larger continental scales. More specifically, with various hyper-temporal and diverse remote sensing data, including SAR, hyper-spectral, multicopter UAV and RGB digital cameras, and RGB-based remote sensing through the collection of various indices over time [16], it became more automated and efficient to monitor rice dynamics in a given area.

Abundant water is needed for rice cultivation, so everywhere in the world rice cultivation starts with direct seeding into the flooded soil or with flooding the soil prepared by sowing (wet sowing and dry sowing). This important stage throughout the process is the key principle of most remote sensing-based rice monitoring algorithms, including "Nguyen and Wagner [39]" classifier. Rice fields are inundated at the beginning of the growing season and remain so until the reproductive stage. However, for other crops such as olives, wheat, and vineyards, the process is not the same and fields are not flooded as frequently during the vegetative and reproductive phases. Therefore, by monitoring the phases and identifying the flooding periods, it is possible to distinguish rice from other crops. Rice growing phases coincide with spatial and temporal changes in plant growth. Two main remote sensing data can be used to detect the cultivation phases: active remote sensing and passive remote sensing. Changes in plant morphology during these phases can be easily detected using different remote sensing methods. The noteworthy rule for this circumstance is that the state of interaction between the plant and the signal (microwaves or light) can be observed and processed and mapped by logical principles. With this main principle, we can demonstrate that each plant has its own exclusive pattern and signature from the point of view of the algorithm. More specifically, here we can capture different cultivation phases for different agricultural crops and plants, which would lead to distinguish them from each other. To achieve this, it is not possible to capture the cropping signature with a single imagery, since we consider time in our algorithms and cropping patterns. For this reason, for an accurate and reliable rice detection, or even for any kind of crop monitoring, we need a dense time series of remote sensing data to dynamically monitor the cropping stages.

In optical remote sensing, one of the common methods is to obtain different indices (such as NDVI and NDVWI) from the time series and discriminating different classes by applying crop signature detection to them. However, optical remote sensing has some

disadvantages compared to temporal use, e.g., it is not weather independent, is distorted by the particles in the atmosphere, cannot be acquired during the night (there is no energy source, sun), etc. Numerous studies and researches have been conducted worldwide on the application of optical sensors in monitoring a variety of land areas (water bodies, agricultural crops, forestry, and various vegetation types) and their mapping. For example, "Nguyen et al 2012 [43]" mapped the irrigated rice patterns of the Mekong Delta in Vietnam through hyper-temporal SPOT NDVI image analysis. "Immerzee et al 2005 [17]" used harmonic analysis of SPOT VGT-S10 NDVI time series to understand precipitation patterns and land use interactions in Tibet. "Gumma et al 2014 [15]" used MODIS 500 m data for 2010 to map the seasonal extent of rice cropland in Bangladesh with high cropping intensity. "Fang et al 1998 [11]" estimated rice area year by year using NOAA AVHRR and Landsat TM. "Martinez et al 2005 [30]" Mapped perennial cropping patterns in small irrigation districts using time series analysis of Landsat TM imagery. Almost all of these studies have used data with high temporal resolution for their studies and most of them have not been able to identify the smaller paddy fields due to two main problems: firstly, the low temporal-spatial resolution of the optical data used, and the atmospheric effects due to existent of suspended particulate matter in the atmosphere and the cloud cover. Secondly, the focus on a large area of study, because as the study area grows, the variety of the land covers increase as well, and the ground might contain more classes which makes it harder to distinguish between them.

Unlike the optical remote sensing, the microwave remote sensing techniques such as SAR (Synthetic Aperture Radar) have tremendous ability to collect data regardless of weather and day-night, which is a great advantage of electromagnetic microwave signals, and they can detect the temporal backscatter pattern of different plants and crops. Backscatter is the portion of the outgoing radar signal that the target (in this case, vegetation or agricultural crops) redirects directly back toward the radar antenna. The received backscatter is a function of the characteristics of the radar system, the topography, and the properties of the crops. Canopy structure of crop and water content vary depending on crop type, growth stage, and its condition. These variations can be detected by SAR sensors to distinguish between different crop types [10] [32] [33] [31]. In order to monitor a particular crop and discriminate it from other classes of crops or ground covers, it is beneficial to have a high level of knowledge and perception regarding the crop growing calendar and its different phases during the growing season and use it in a rational way to extract the classified map of our desired crop. "Nguyen et al 2015 [39]" proposed to use a dense SAR stack of images throughout the entire rice growing season and extract

the phenological parameters to classify the rice fields and achieve the best discrimination between the different types of plants in 8 different areas. "Xiao et al 2021 [60]" implemented a similar algorithm in Mapping paddy rice with Sentinel-1/2 and phenology, object-based algorithm in Hangjiahu Plain in China using GEE platform. "Busetto and Casteleyn et al 2017 [5]" took the same direction in monitoring rice areas in Italy, Spain and Greece in the ERMIS project "Downstream services for rice crop monitoring in Europe: from regional to local scale". In another study by "Clauss et al 2018 [7]", they performed rice mapping with Sentinel-1 time series using superpixel segmentation, where they have segmented the study area into superpixels and applying almost the same algorithm to classify rice fields. "Stroppiana et al 2019 [55]" mapped rice acreage and inundation dynamics during the season using both optical and SAR satellite data. Their approach relies on a priori knowledge of cropping dynamics to adjust TSD (Temporal Spectra Descriptors) calculation time horizons and thresholds to local conditions. The output products consist of maps of rice acreage, rice seeding practices (dry and flooded rice), and flooding practices. They applied a more convoluted algorithm compared to the other algorithms in other studies, but the idea behind it is still the same. "Planque et al 2021 [49]" applied a knowledge-based descriptive algorithm using Sentinel-1 time series for a national crop mapping. "Onojeghuo et al 2018 [47]" has introduced another algorithm for rice crop mapping by applying machine learning methods to multi-temporal Sentinel-1 and Landsat data. They also introduced a rice cultivation calendar for their algorithm. "Mansaray et al [29]" mixed both active and passive remote sensing methods for mapping rice fields in urban Shanghai city in southeast China, using Sentinel-1A and Landsat 8 datasets, which showed good results over a larger scaled area. Almost all the algorithms we presented in this section from different publications used the inundation of rice fields before crop establishment which is different from other types of land covers grown with other types of crops. Backscatter values are very low during flooding in the vegetative and transplanting phases, and higher during the early growth stages of the rice plant in the reproductive and maturity stages. In summary, the rapid changes and high correlations of backscatter values in the temporal SAR backscatter profiles provide a specific rice pattern that we can use as a rice signature to determine rice distribution in our study area.

Sentinel-1 consists of a constellation of two satellites, Sentinel-1A and Sentinel-1B, which together have a revisit time of 6 days. The Sentinel-1 constellation provides us with high spatial and temporal resolution, making this mission superior to other radar missions. The C-band Sentinel-1 SAR provides land services and applications, such as crop monitoring, land cover mapping, and change monitoring, by offering an interferometric Wide

(IW) swath, dual-polarization (horizontal (H) and vertical (V)) acquisition mode, and fine spatial resolution data (20 m). Free data products are available in single-polarization (VV or HH) and in cross-polarization (VV +VH or HH+HV) from the Copernicus platform and can therefore be used to map different cropping patterns. To map different cropping patterns, we have the possibility to use different polarizations that have different interactions with different plants and crops due to multiple scattering caused by structures (stems, leaves, fruits) within the crop volume [31].

Some studies use single polarization of backscatter data and others use multi-polarization for mapping, depending on the plant types, the temporal density of the available dataset, and the study area. "Nguyen et al [39]" utilized multi-polarization backscatter time series to map rice fields and concluded that the best polarization for rice monitoring is VH-polarization. "Bouvet et al 2009 [2]" algorithm recorded the change of intensity backscattering coefficient of vertically or horizontally co-polarized data (VV or HH), they evaluated the use of HH/VV polarization ratio and they had higher accuracy and required less temporal coverage compared to single polarization for rice mapping. "Lasko et al 2018 [21]" used VV-only, VH-only, and both polarizations together and concluded by comparing the final results that the data for VV/VH had the highest overall accuracy based on the accuracy assessment, while VV had the lowest overall accuracy. "Nguyen et al 2016 [42]" used Sentinel 1A time series in dual-polarization (VV/VH) during the rice growing season in the Mekong Delta to analyze the relationship between the growth cycle of rice plants and the temporal variation of SAR backscatter at different polarizations. Their results show that VH-backscatter is more sensitive to rice growth than VV-backscatter. Furthermore, "Mansaray et al 2017 [29]" and "Nguyen et al 2017 [40]" also demonstrated that VH-backscatter is more sensitive to rice growth pattern than VV-backscatter with higher accuracy. "Clauss et al 2017 [7]" chose VH polarized data over VV due to the higher dynamic range of backscatter over rice areas caused by agronomic flooding followed by vertical plant growth, and by applying the algorithm to both VV and VV-backscatter over six study sites in different rice growing regions. Last but not least, based on the literature review conducted, we concluded that VH-polarization is more sensitive to rice growth and classification more than the other polarizations.

Another vital point in remote sensing-based crop detection is the selection of an appropriate algorithm that is applicable to the selected study area and compatible with the dataset used, taking into account the temporal density and the essence of the dataset. So far, several remote sensing-based crop classifiers have been introduced. Some of the

algorithms applied Deep Learning and Machine Learning algorithms to the Sentinel-1 time series to extract rice acreage from a wide range of agricultural crops or other land covers, including Deep Recurrent Neural Network (RNN) by "Ndikumana et al [36]" and "Zhang et al 2020 [61]", Support Vector Machine (SVM) and Random Forest (RF) Machine Learning classification algorithms by "Onojeghuo et al [47]", "Zheng et al 2015 [63]" and "Gislason et al 2006 [14]", K-nearest neighbours (KNN) by "Myint et al [35]" and the application of a Decision Tree approach (DTs) by "Waheed et al 2006 [57]" were applied with different results. The supervised classification algorithms require a set of training data for classification, which is one of the limitations and difficulties of this type of classifiers. After training and testing the algorithm, the classifier may even be overfitted or underfitted in the final classification product, so there is the challenge of choosing the effective parameters as well.

The other types of remote sensing-based crop classification algorithms adopt a crop phenology-based classifier by setting the decision nodes for the decision tree (a decision rule-based classification) to detect and map rice patches using Sentinel-1A time series. "Nguyen et al 2017 [40]" investigated a phenology-based approach to map rice crop at a continental scale by using space-borne C-band SAR data. Mode S-1A IW time series with a spatial resolution of 20m and VH-polarization were used, covering eight sites in the Mediterranean region. The results of the publication show that the proposed approach is efficient and operationally feasible to extract rice cropland areas with high accuracy at a spatial resolution of 20m (single polarization) using S-1A time series.

A decision tree approach was used to map rice-growing areas in the Mekong Delta region by extracting rice phenological parameters. This extraction by "Nguyen et al 2016 [42]" study is the first Sentinel-1A-based analysis to map rice-growing areas and estimate phenological parameters for rice at the regional scale, as reported by "Nguyen et al 2016". Similarly, "Nguyen et al 2015 [39]" applied a crop phenology-based classifier to detect and map single-, double- and triple-cropped rice areas with multi-year Envisat ASAR data, also applied it to map rice fields using Sentinel-1A time series. Another publication by "Nelson et al 2014 [37]" contributed to rice mapping and monitoring methods based on multi temporal SAR data. It presented a new rule-based algorithm based on agronomically relevant rules and parameters that classifies rice area with consistently high accuracy across different rice environments and cropping practices (wet season and dry season, irrigated and rainfed, direct-sown and transplanted, short- or long-duration varieties). The study demonstrates that regularly acquired X-band imagery (HH SAR) is

suitable for monitoring rice plants in major rice growing areas in South and Southeast Asia. Many similar researches have applied the similar phenology-based algorithm to distinguish rice area from other crops with high accuracy, such as "Lopez et al [27]", "Inoue et al 2013[18]", "Inoue et al 2002 [19]", "Salehi et al 2017 [50]", "Son et al [53]", "Dong et al [8]", "Lopez et al [28]".

In this thesis we aim to monitor and detect rice acreage over northern Italy using Sentinel-1A and B, VH and VV polarization time series. To do this, we use an existing phenology-based classification approach by defining the decision-tree decision-nodes by extracting the statistical backscatter parameters of rice, which is proposed by "Nguyen and Wagner" [39] [40].

1.6. Research Objective and Outline

1.6.1. Research questions

This work aims to investigate the use of C-band SAR time series for monitoring rice acreage and classification. It is motivated by some questions that have been revisited, argued, and acknowledged during this thesis:

- The advantages of sentinel-1 time series in rice monitoring (pros and cons).
- Why is this algorithm applicable in our study region?
- Is Sentinel-1 one of the best types of data to map regional and continental rice production and cultivation? And why?
- Which type of polarization provides the best results for rice monitoring and classification? VH or VV?
- What kind of rice phenology parameters can be retrieved from SAR time series? And why are these specific parameters able to distinguish rice from other types of land cover?
- Is this method - extraction of statistical backscatter parameters - suitable for monitoring rice cropland? Why does this algorithm work well for rice?
- Are the results consistent with the provided reference data?
- Why can the SAR -based time series approach be used for large scale rice monitor-

ing?

- Where does this algorithm usually tend to fail? And why?
- How and in what sense we improve the algorithm?
- What are the other parameters extracted from the dataset, and are they beneficial enough to increase the accuracy of the monitoring? Why are they effective?

1.6.2. Research Hypothesis or Propositions

There are three main ultimate objectives for this study, with the following propositions:

objective 1:

Proposition 1.1. *Which polarization(s), or combinations thereof, are superior to the other polarizations in discriminating the temporal pattern or signature for different types of crops, especially rice?*

Hypothesis: VH-backscatter indicates the best polarization in our studies with the most discriminatory temporal signatures of different cropping patterns when the different results of the different polarizations (VV and VH, VH/VV) are evaluated and compared.

objective 2:

Proposition 1.2. *Does this specific classifier, "Nguyen and Wagner"[39] method [40] provides us with convincing results of accuracy?*

Hypothesis: This classifier is one of the appropriate rule-based algorithms in terms of good discriminative ability with high precision and without the need for on-site data collection.

objective 3:

Proposition 1.3. *Besides this classifier's course of actions, is there any other strategy to improve the current classifier to achieve a higher accuracy with reliable results?*

Hypothesis: This affair was one of the most challenging phases throughout the study, with much trial and error course of actions. The most recent hypothesis for this proposition is the local extrema approach with a window size of 80 days for the entire rice growing season.

1.7. Scientific Impact

Any kind of research in any sake has some influence on the direction of solving a problem, even though if it is negligible. With our research in this area, we hope to contribute to further research that would directly or indirectly initiate others in the scientific communities and interdisciplinary fields. The significant outcome that could contribute to a brief review of this work from a scientific point of view are the following:

A SAR-based analysis over the entire rice growing season using preprocessed backscatter time series (stack of $\sigma_{normalized}^0$) in both VV and VH polarization modes by extracting statistical backscatter parameters of rice, i.e., we call them statistical parameters because these quantity maps tend to summarize or describe a certain aspect of the population over time (temporally). And creating the decision nodes of our decision tree by defining a rice growth calendar (i.e., beginning of the season, end of the season, and length of the season of rice cultivation) compatible with our study area using the collected statistical backscatter parameters. In deriving the final results and analyzing them through different evaluation techniques, we found that the VH-backscatter coefficient is more sensitive to rice growth than VV-polarization. This work is expected to provide reliable results even in large areas with continental scales containing varied land covers, farms and croplands, but will lead to errors in mountainous regions and forests with complex climatic and weather situations. In addition, by comparing the captured results of two consecutive years 2019 and 2020, we achieved to almost similar results for both years in general. This indicates e.g. that the parameters required for the decision tree can be robustly derived from one year of data.

2 | Data and Study Area

2.1. Sentinel-1 Timeseries Dataset

Deciding on the type of dataset to use in this study was a substantial step. Analyzing the different types of data and their objective applicability, the Sentinel-1 C-band SAR sensor proved to be the best option for mapping rice. The L-Band is typically used for environmental, positioning and communication purposes or for satellite navigation and telecommunications with a frequency range of 1-2 GHz. The X-Band is generally used by civilian, military and government agencies for weather monitoring, air traffic control, maritime vessel traffic control, defense tracking and vehicle speed detection for law enforcement with a frequency range of 8.0-12 GHz. Finally, the C-Band with a frequency range of 4-8 GHz is primarily used for Earth observation and global mapping, change detection, monitoring of areas with low to moderate penetration, higher coherence, ice, and maritime navigation. Given these three options when choosing the data band, we decided to proceed with the C-band in our investigations.

As mentioned earlier, dense time series datasets are best suited for rice monitoring. For this reason, both Sentinel-1 A and B multi temporal images were used for this work. In general, the SAR sensor generates weather- and time-independent data series every 6 replicate days for both missions and 12 revisit days for a single mission. For this reason, monitoring time series is interesting and beneficial to us without the need for atmospheric correction, preprocessing and the spatial resolution (20 m) is very good. The Sentinel-1 constellation with the SAR sensor has several polarization modes, including VV, VH, HH, and HV. For our purposes, we can choose data with co-polarization (VV or HH) and with cross-polarization (VH or HV) or a combination of them (VV /VH, HH/HV and etc).

In summary, the data from Sentinel-1 SAR provide the opportunity to obtain information on plants with spatial resolution appropriate for individual rice fields and sufficient temporal resolution to capture the growth profiles of different crop species for further analysis.

SAR Data Properties

	dataset property
Satellite	Sentinel-1 A and B
Sensor	SAR C-Level
Product level	GRD
Frequency	5.045 GHz
Polarization	VV and VH
Mode	Descending and Ascending
Image mode	IW
Temporal resolution	6 days
Spatial resolution	20m
Incidence Angles of A	15, 88, 117, 161
Incidence Angles of B	66, 95, 139, 168

Table 2.1: Properties of the Sentinel-1 time series data used in this research.

2.2. Reference Data (existing survey data)

The reference data mainly used in this study is the rice fields layer of the European Corine Land Cover2018 (CLC2018) product. In order to compare the classification results in our study, over 9tiles are used (Figure 2.1: 9 segments of squares). This product is compiled and assembled every 4-6 years in 44 classes and has a spatial resolution of 100 m for the European continent. The latest version CLC2018, funded by Copernicus, was produced in less than a year. The CLC product has a wide range of applications supporting various community policies related to the domain of environment, agriculture, transport, etc. Due to its coarse resolution, it may not meet all the users' requirements seeking results with very high accuracy. In our case study, we downscaled the CLC rice map from a resolution of 100 m to 20 m. However, some details and features with a width of less than 50m are lost, such as the channels between rice paddies or rivers with a width of less than 50 m. In addition, during the optimization of the thresholds and in the evaluation phase, errors

can occur which are considerable and cannot be turn a blind eye and ignored. Another problem of this product that we found during our analysis is that it contains some areas with a wrong classification of rice when compared to the local rice layer downloaded from ersaf (Ente Regionale Per I Servizi All'agricoltura e alle Foreste), the website for the analysis and monitoring of land use in the Lombardy region (www.ersaf.lombardia.it), a data sort with high accuracy and reliability. Unfortunately, when searching for similar websites, we could not find similar reference data for the Piedmont region. In order to have consistent reference data for the entire Po catchment, we use the CLC2018 as main reference despite the aforementioned drawbacks. Another source of discrepancies could be the gap between the CLC2018 acquisition date and the SAR time series we have used for 2019 and 2020.

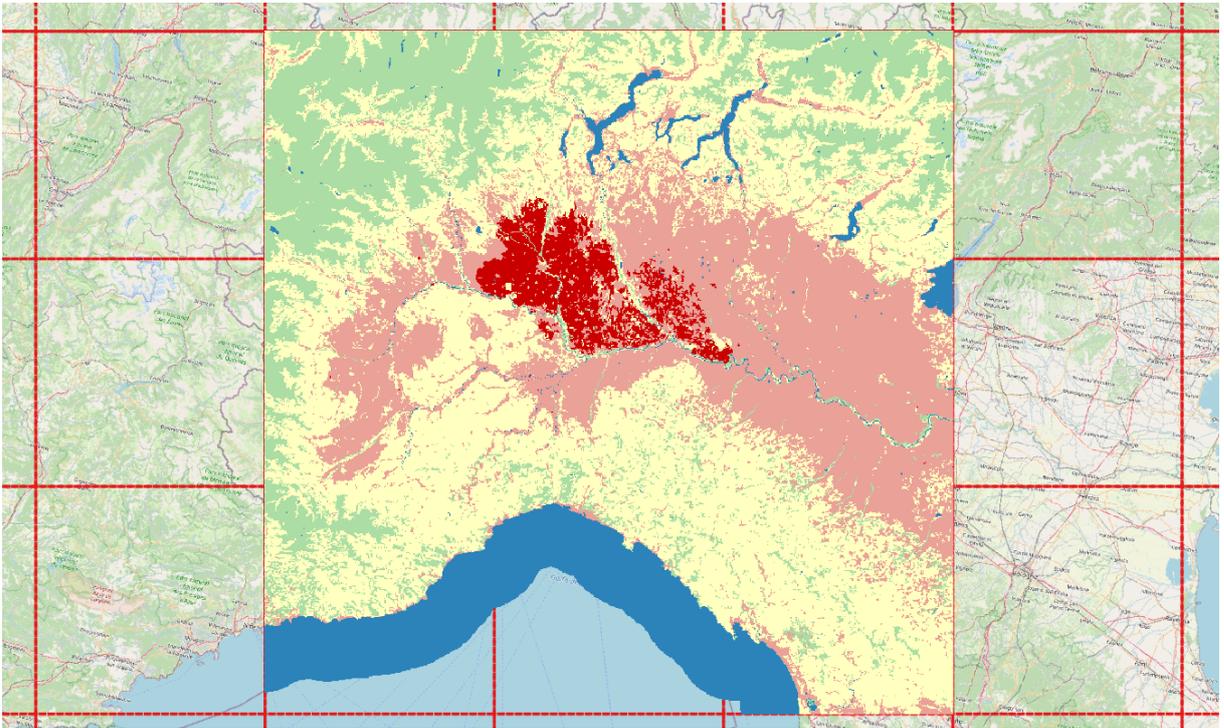


Figure 2.1: According to CLC2018 classification the red area is the class of rice fields in north part of Italy located in Lombardy and Piedmont regions (9 tiles of data).

2.3. Study Area

Our study area is located in the northern part of Italy (Po Valley), crossing Lombardy and Piedmont regions (Figure 2.2). These two regions are the most important rice-growing and producing areas in Italy, with a high percentage of rice paddies. This territory is

an intensive agricultural region with an area of approximately 11,200 Km² where various crops such as corn, olives, soybeans and rice are grown (Figure 2.3). According to the Eurostat website (Figure 2.4), rice cultivation in Italy (Figure 2.4 refers to all of Italy) covers more than 2000 km². Rice fields are distributed around the Po River between the provinces of Milan, Vercelli, Mortara, Novara and Pavia, representing about 90% of the total area under rice cultivation in Italy and 40% in all of Europe. The Po Valley is one of the prominent and momentous agricultural and subsequently industrial areas in Europe. The particular reason for this circumstance is existence of flat lands and the abundant water supplies from the Po River for various purposes, mainly for the irrigation of agricultural crops, especially rice, which requires a lot of water.



Figure 2.2: Location of the study area. The inset shows the Po river and the agricultural lands around it in yellow

Annual temperature varies between 10 and 30°C during the rice growing season, i.e., from April to November, and average precipitation exceeds 950 mm/year, with peaks in spring and autumn in this region. The altitude of the valley through which the Po River flows, excluding its tributaries, varies from about 4m to approximately 2100m, i.e. there is a huge difference in altitude, which gives rise to a different climate and vegetation. Italy is well known for its different varieties of rice, especially Japonica and Indica, which are cultivated in this area. Unlike many other countries where rice cultivation is scattered, in Italy it is denser in a particular region.

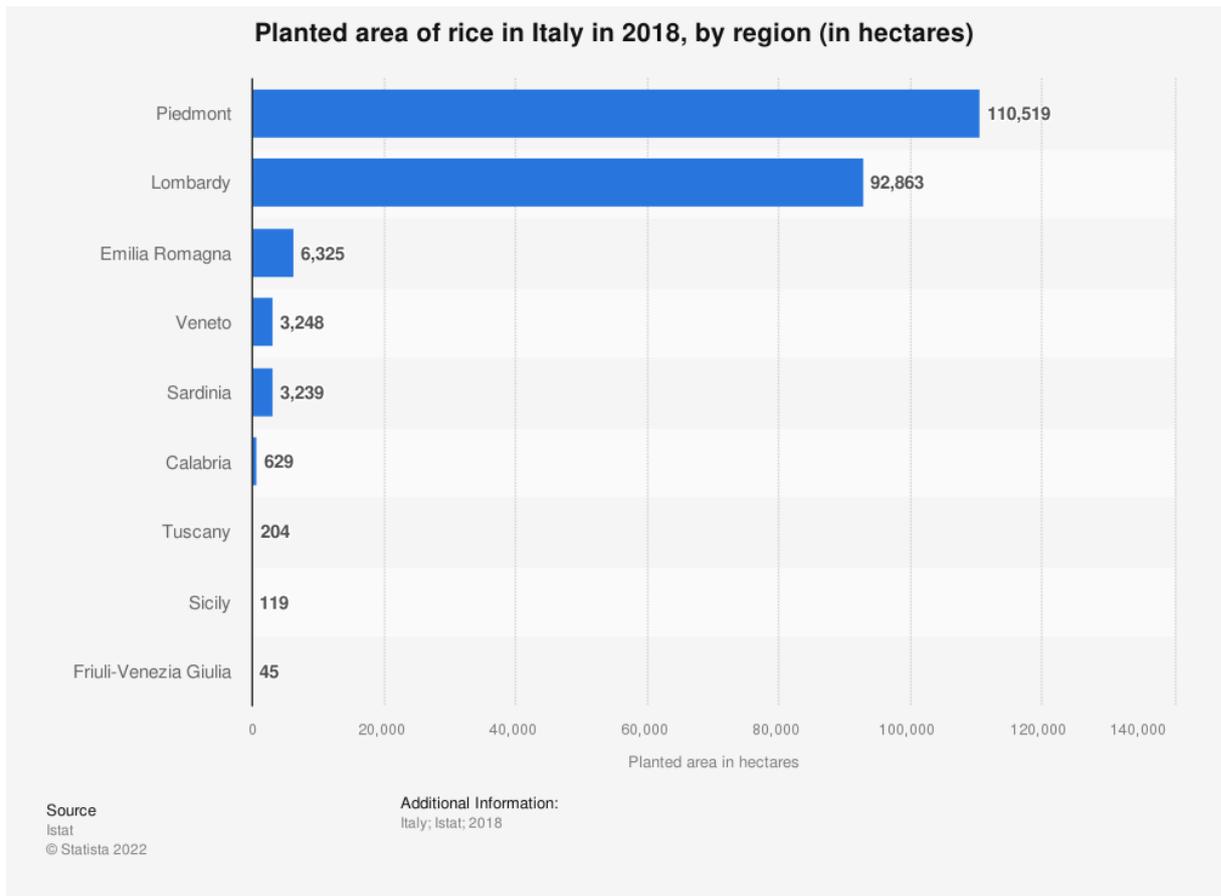


Figure 2.3: Planted area of rice in Italy in 2018 according to istat website

The Italian rice cultivation system exerts the flooding technique for paddy irrigation at the beginning of the rice growth season. In the study area, there are three different phases of cultivation: Planting (vegetative), or more precisely sowing, begins in early April and lasts until mid/late May, i.e. when the rice fields are flooded with abundant water using a sophisticated irrigation system to prepare them for seeding subsequently, the flooded water is maintained until a certain time in different levels. After the rice plants are firmly established in the water and soil, precipitation regularly maintains to a certain level. This phase is important, among other reasons, because this is the time when the rice reproduction and ripening happens approximately around early June to mid/late September. During this phase, the fields begin to turn green and the water slowly and leisurely disappears, and then the harvesting phase starts. The last phase, known as harvesting begins roughly around September/October and lasts until November, which depends on many factors, including temperature (during this phase the soils gets dried and the rice turns into a white grain as an end product). In Italy there is only one harvest per year, which typically takes place in September/October.

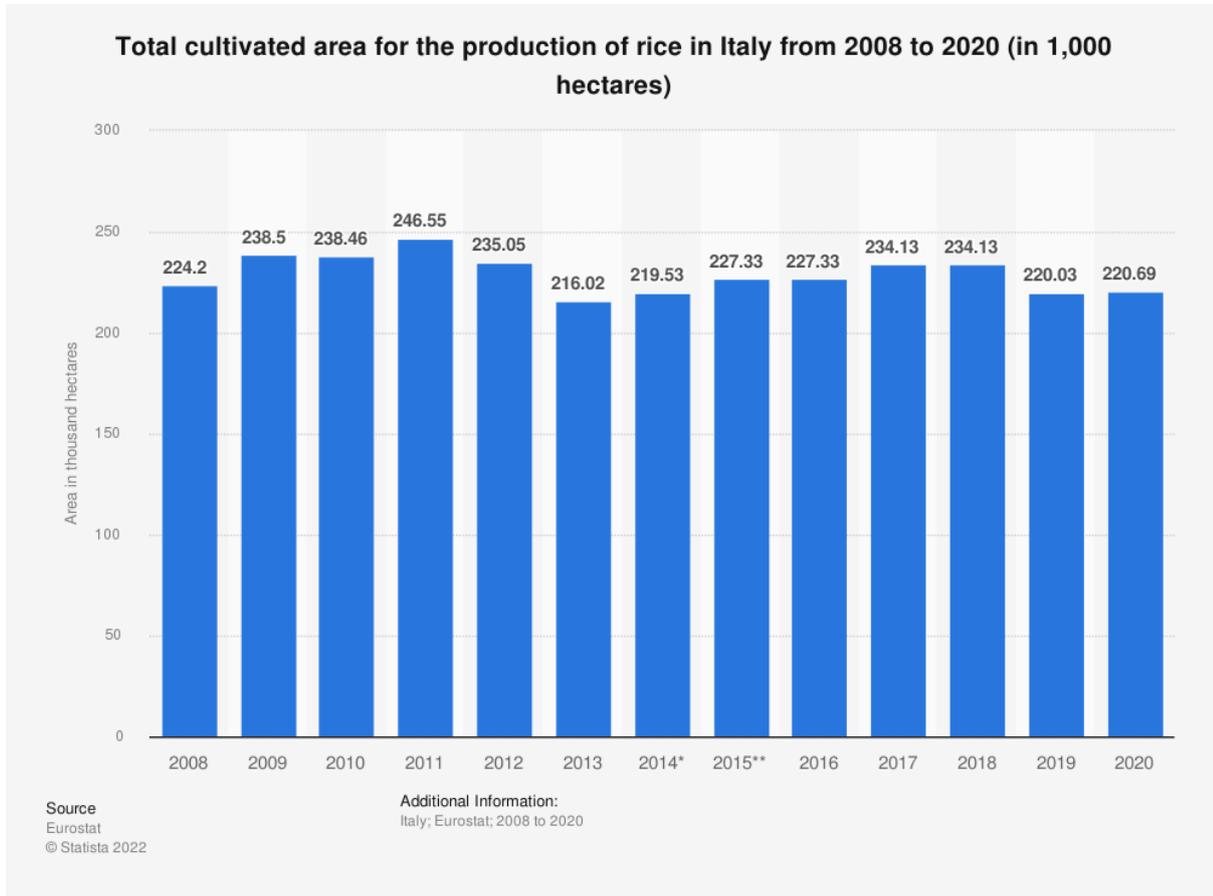


Figure 2.4: Total cultivated area for the production of rice in Italy from 2008 to 2020 according to Eurostat website (ec.europa.eu)

This rice cultivation pattern is exclusive in this area and very compatible with the algorithm presented in this work, which is elaborated in more detail in the following chapters through various arguments.

2.4. Rice Cultivation Calendar in Study Area

Rice cultivation has different phases similar to those of most rice-growing regions. Particularly, Italian rice cropping system applies the flooding technique for paddy irrigation at the beginning of the rice growing season by sowing directly into the flooded areas or sowing directly before the flooding phase. Looking at the rice cultivation in detail on the Sentinel-hub website (www.sentinel-hub.com/explore/sentinelplayground/) and in the local references [5], referring to the Sentinel-2 L1C mission and the NDWI (based on the combination of bands B3 - B8 / B3 + B8) and SWIR (based on bands 12, 8A, 4) indices we can recognize that in the Po River basin, the normal growing season begins in early/mid-

April, when the rice fields are submerged with the help of a sophisticated irrigation system towards the channels between the rice fields. Seeding takes place accordingly in early or mid-April and lasts until late May or early July. Once the rice plants are planted, the water starts to disappear from the irrigation system and the rice seeds enter the germination phase and become green plants. Conventionally, we have chosen to call this phase the vegetative or pre-harvest phase, during vegetative irrigation continuously, but as the rice plants begin to grow and turn green, we will have the specular backscatter effect that causes the higher backscatter values during data collection than the flooding phase. This phase begins in this region in June/July and lasts until late-August/September. The next important phase for rice field mapping is the rice harvest phase when the soils are dry and the rice plants have turned into white grains as the final product, which is exactly when the backscatter values (σ_0) reach their high values (Figure 2.6).

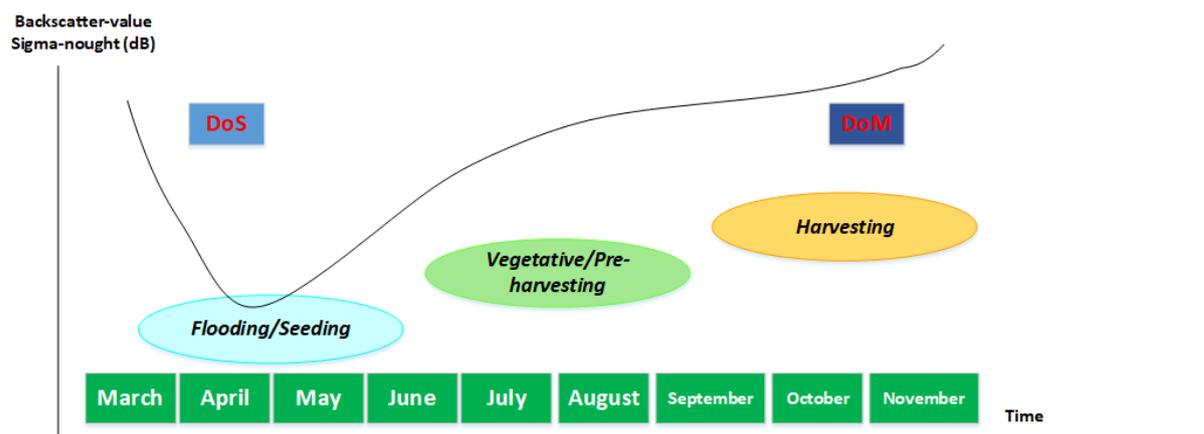


Figure 2.5: Temporal backscatter variation of rice during the growing season. This scheme facilitates the gap of rice backscatter dynamics at each stage. During the flooding phase, the backscatter values of the rice pixels are very low due to the irrigation and presence of abundant water on the surface. Backscatter levels begin to increase as the rice harvest reaches its final stage and the water vanishes from the soil.

The harvest phase takes place at the end of September/October or October/November, depending on when the floods began, and varies in the small regions of the Po River basin. In Europe, rice cultivation usually begins with direct seeding into the flooded soil (hydroseeding) or into the dry soil (dry seeding), but in the Po River basin the first method was usually used. In both cropping methods, the inundation water is retained until harvest time (the flooding phase contains minimum σ_0 values, and harvest contains maximum σ_0 values) Figure 2.6. Another reason why our dataset and algorithm are applicable in this region.

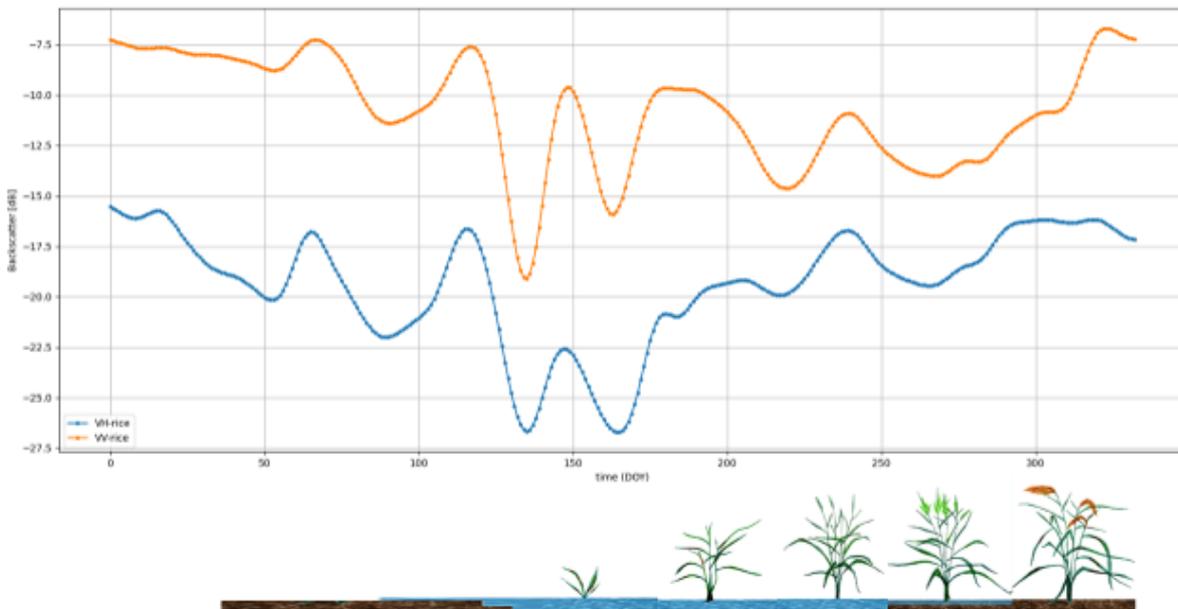


Figure 2.6: Derived rice cultivation calendar by obtaining the mean backscatter values for a rice classified pixel.

The harvesting phase takes place at the end of September/October or October/November, depending on when the floods began, and varies in the small regions of the Po River basin. In Europe, rice cultivation usually begins with direct seeding into the flooded soil (hydroseeding) or into the dry soil (dry seeding), but in the Po River basin the first method was usually used. In both cropping methods, the inundation water is retained until harvest season (the flooding phase contains minimum σ_0 values, and harvest contains maximum σ_0 values) Figure 2.6. Another reason why our dataset and algorithm are applicable in this region. In April and May, there may be problems with emergence due to lower temperatures that limit germination of rice seeds.

During the vegetative phase in the hottest months of the year around July and August, precipitation increases, which increases the risk of rice blast infections, so rice fields are constantly irrigated. At the end of July/beginning of August, the rice grains begin to flower, so that the harvest of the plants begins in September/October/November. The average timing of the phases and yield is influenced by variety of variables, and may vary each year depending on cultivation and irrigation conditions. In this study, we analysed everything in detail and tried to announce a logical rice cultivation calendar compatible with the Po River basin, cultivation techniques and climatic conditions.

2.5. Software and Tools

Software and tools used for managing big and heavy data, data processing and completing thesis conduction are as follows:

Tools used during the thesis

Tool/Software	Version	Purpose
Python	3.6	SAR imagery processing, data cube construction, obtaining the statistical backscatter parameters, threshold optimization, validation, classification, graphing, statistical analysis
SNAP Tool	8.0.0	SAR time series preprocessing
Qgis	3.18.2	Threshold defining, results visualization, resampling, etc
Sentinel-hub	Online	Defining the consistent rice growing calendar to the study area
LaTeX (Overleaf)	Online	Writing the manuscript of thesis

Table 2.2: Software and tools used while completing entire thesis conduction.

2.5.1. Methods and data analysis

Preprocessing was performed using the SNAP sentinel-1 toolbox, including application of the orbit file, radiometric correction, geocoding, resampling, and dB conversion.

The entire process of structuring the data cube, data smoothing, orbit correction, phenological parameters (σ_{MIN}^0 , σ_{MAX}^0 , σ_{MMD}^0 , DoS, DoM, LoS), acquisition, threshold optimization, validation, classification, time series analysis, and plots was performed and completed in Python, using different packages for each process:

- Yeoda: (your earth observation data access) a well designed package by Geo-department of TUwien, for lower and higher level data cube classes to work with well defined and structured earth observation data.
- Gdal
- Numpy
- Pandas
- Geo-pandas
- Matplotlib
- Geopathfinder
- Xarray and Data-array
- Sklearn

Qgis was used to visualize the results during each procedure, especially for the decision nodes that define or evaluate the thresholds we set.

Sentinel-hub playground is a very useful online tool that uses the Sentinel Hub technology to facilitate the discovery and exploration of Sentinel-1, Sentinel-2, Landsat-8, DEM, and MODIS imagery in full resolution, as well as access to Earth observation (EO) data products. It is a graphical interface to a complete and daily updated Sentinel-2 archive, a comprehensive resource for anyone interested in the Earth's changing surface, whether natural or man-made. With visualization of various indices and insight and inspect into time and rice growing season, we have defined a comprehensive and informative rice growing calendar for further applications.

3 | Methodology

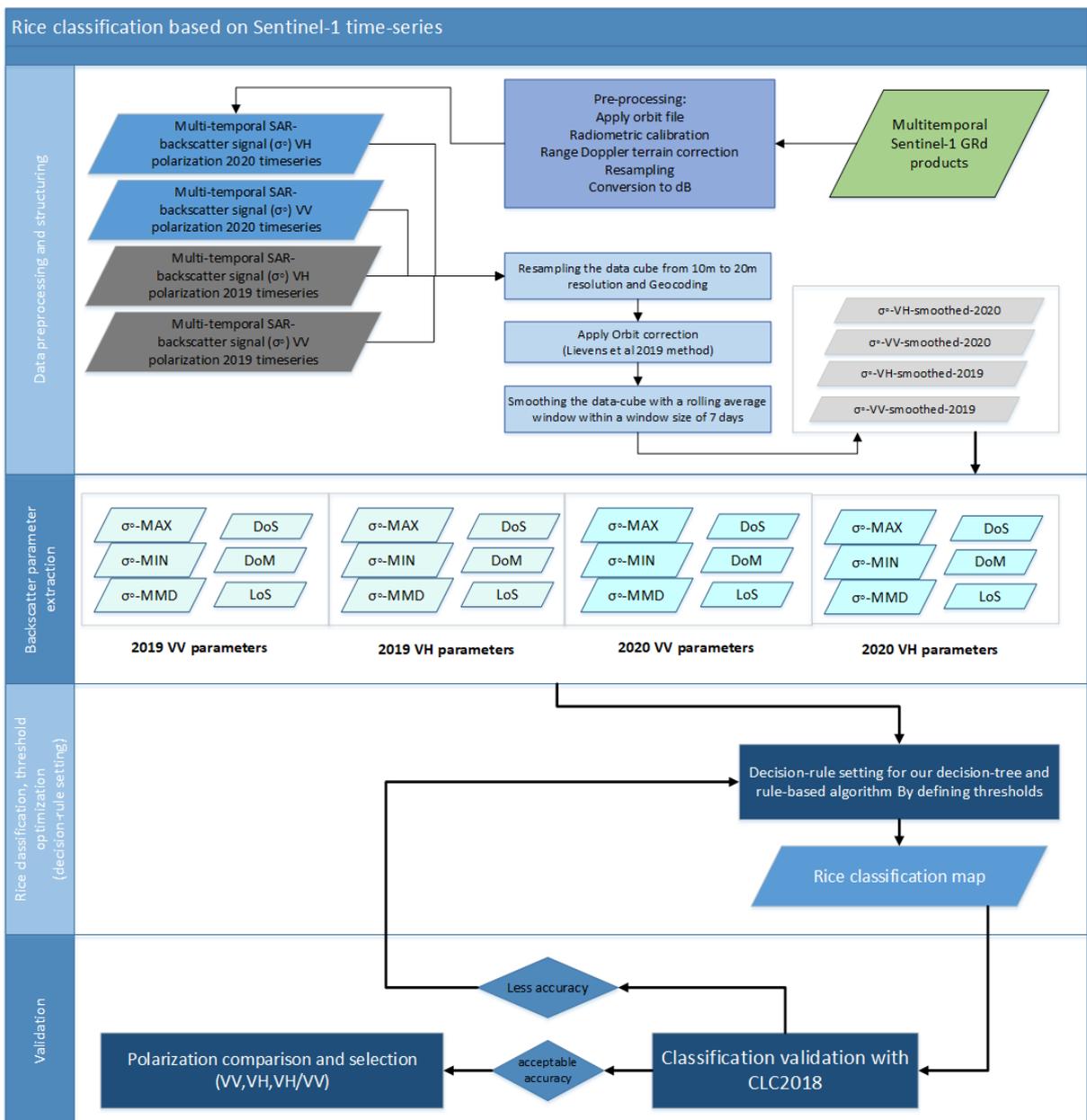


Figure 3.1: Methodology flowchart, the initial sketch.

In the Figure 3.1, parameters σ_{MAX}^0 and DoM refer to the backscatter values and the date when the backscatter reaches a maximum value and is defined as the global maxima in σ_{smooth}^0 time series. The σ_{MIN}^0 and DoS values are the backscatter values and the date of the beginning of the season, and defined as the global minima in σ_{smooth}^0 time series. σ_{MMD}^0 refers to the correlation between the global minima and maxima during the rice growing season, which is also called the sensitivity. LoS is the length of the season, defined as the number of days difference between DoM and DoS.

3.1. Approach Methodology

Various classification algorithms have been introduced in different studies so far. Most of them have utilized machine learning methods or index-based classification techniques to classify rice acreage from other classes. The time series algorithm used in this work, unlike the other algorithms, deals with the inherent characteristics of the cultivation of a particular crop by analyzing it during its growing season, and the modality of interaction of microwave radiation with different targets, i.e., mainly rice and water. Moreover, this algorithm was introduced by "Nguyen et al [39]" and we applied the similar procedures in the workflow with some additional steps.

The flowchart of the algorithm is shown above Figure 3.1. The algorithm of "Nguyen et al" includes 5 main modules as follows:

1. Preprocessing of Sentinel-1 data and organization of data.
2. Temporal filtering and orbit correction.
3. Extraction of statistical parameters from the backscatter coefficients.
4. Identification of rice-fields (rice classification) using a knowledge-based decision-tree approach.
5. Validation and evaluation of accuracy using the reference data (CLC2018).

Sentinel-1A and B Synthetic Aperture Radar (SAR) in IW (Interferometric Wide swath) acquisition mode were imaged over the study area in C-band for rice classification purposes and further analysis. We used time series of images covering years 2019 and 2020 with two available polarizations over land; vertically transmitted, horizontally received (VH) and vertically transmitted and received (VV) for both.

Beforehand, to obtain a usable homogeneous dataset, the imagery was preprocessed in Level-1 GRD format and converted to a geocoded backscatter coefficient for the terrain.

The GRD product is a multi-looked product and is projected onto the ground range. The goal of preprocessing is to enhance the image data, suppressing unwanted noise and distortion and enhancing some image features that are important for further applications in the processing phase. The steps of the preprocessing section (3.2) were applied by TUWien colleagues and provided by them for this work. From (3.2) onward the entire analysis was implemented by the author. The preprocessing workflow includes the following steps:

3.2. Preprocessing

3.2.1. Apply Orbit File

The orbit of a satellites can deviate over time, no matter how well the satellite is adjusted. There are many reasons for this, such as the imperfect sphere of the Earth, other massive objects in the solar system that perturb their orbits with their gravity, solar winds, atmospheric drag, etc. Consequently, when these discrepancies occur, either the dataset or the satellite mission itself must be periodically adjusted. Therefore, it is indispensable to have amended orbital files that notify where the satellite is exactly located. The orbital positions are processed by the Copernicus Precise Determination (POD) service, and the files are delivered within 20 days of data acquisition in space with an accuracy of 5cm [3]. Sentinel-1 orbit information can be found on the ESA website.

Normally we use the orbit state vectors provided in the metadata information of the products from SAR, which are not accurate enough. The exact satellite orbit is calculated only after a few days after the product is generated. We update the orbit metadata with a restituted precise orbit file which is available in the software SNAP. This software automates the downloading and updating of the orbit state vectors for each SAR scene in its metadata by providing accurate satellite position and velocity information and improving geocoding [12].

3.2.2. Radiometric Calibration

The objective of radiometric calibration is to convert the digital pixel values (DN) stored in the SAR image products into calibrated SAR backscatter values. The Look Up Tables (LUTs) included in the Sentinel-1 Level-1 products can be used for this conversion. Again, a vector called the calibration vector is embedded within the metadata of the product that performs the conversion of the digital numbers (DNs) of the SAR images into sigma-naught values σ_0 . This mechanism is automated in SNAP. It automatically determines

what kind of product we insert as input and what conversion to apply to the product[24] product. The value of sigma-naught depends on many factors, including wavelength, angle of incidence, target properties, and polarization.

3.2.3. Range Doppler Terrain Correction (Geocoding)

SAR products are usually captured from different viewing angles which causes some distortions in the end product. By applying the terrain corrections we can compensate the appeared distortions in our product, so we can achieve products with high conformity to the real world. customarily Range Doppler terrain correction is used to negate the effect of the geometric distortion caused by topography [34]. The geometric distortions are caused by the shadows, angular orientation with respect to the ground, altitude during data acquisition, foreshortening, layover, and using the digital elevation model to correct the location of each pixel [12]. The Range Doppler approach is the most appropriate way to perform the geometric correction. The method needs information about the topography (normally provided by a DEM) as well as orbit satellite information to correct the topographic distortions and derive precise geolocation for each pixel of the image. The Range Doppler correction operator in SNAP implements the Range Doppler orthorectification method for geocoding SAR scenes from images in radar geometry [52].

Geometric correction and radiometric calibration are mandatory to compensate the effect of different variables such as incidence angle, acquisition with different sensors, temporal ramifications, or effects that causes the appearance of different values for the same location that restricts us in comparing the products.

3.2.4. Resampling (Speckle Filtering)

Resampling is a way to manipulate the digital image and transform its pixel values to other values due to many reasons including terrain correction in our case the reason was mainly related to the high volume of the dataset and intensity of the calculation with huge Sentinel-1 dataset which was time-consuming. We can also consider resampling as one of the sections of the terrain correction, but here we tend to separate it from the geometric correction phase because by resampling our dataset we could able to reduce the speckle noise in a way by going from 10m resolution to 20m resolution. In SNAP software while doing the geometric correction the operator allows us the selection of the image resampling method and the target pixel spacing in the target CRS. This step has to be

done after geocoding in order to be able to reach satisfactory results. The main denoising step has done in the next section named filtering in order to acquire spotless products.

3.2.5. Conversion to dB

Finally, after applying all the above processes on the whole dataset with two different polarization (VV and VH) separately, in the last step of preprocessing by exerting a logarithmic transformation, the unitless SAR backscatter coefficients are converted to dB (decibel).

$$\sigma_0(dB) = 10 \log(DN) \quad (3.1)$$

This last phase generates the end product for further classification and analysis steps.

3.3. Orbit correction

As part of Sentinel-1, two solar-synchronous, polar-orbiting satellites (S-1A and S-1B) provide day and night radar images regardless of the weather. A cycle for each satellite lasts 12 days, with 175 orbits per cycle for each satellite. Both Sentinel-1A and Sentinel-1B orbit on the same orbital plane, but with a phase difference of 180°. For each cycle of the Sentinel-1 satellites, we would have observations with different orbits, likewise the observations will represent different incidence angles as well [26]. As far as this work is concerned, the orbits are in two general modes: ascending and descending with corresponding incidence angles of 15, 88, 117, and 161 for the ascending mode and 66, 95, 139, and 168 for the descending mode, for both VV and VH polarization. The key point to elaborate on is that the data taken at different incidence angles have different strengths of the radar backscatter signals. Similarly, the consideration of the influence of the incidence angle on the backscatter values in our studies and analyzes is noteworthy.

There is a certain behavior that can be decoded by examining the same area but using different images with different incidence angles, and that is that the strength of the radar backscatter signal gradually decreases as the incidence angle increases. The influence of the incidence angle may not be as great in the flat-lands or arable and agricultural lands as it is in mountainous areas, but it is beneficial to take it into account. In addition, it is practical to point out that the strength of the backscatter, especially in a co-polarized

situation, depends to a large extent on the specular double-bounce effect. Ergo, by considering the incidence angle effect we could somewhat neglect the double-bounce effect. Two main solutions have been proposed to account for the incidence angle effect. One of them was the optimal solution, which was used for the rest of the calculations.

The first and simplest solution was to separate the dataset according to its incidence angle and the acquisition mode. Comprehensively, we proposed not only the use of a single decision-rule tree for the entire database cube, but also the application of different decision-rule trees with different thresholds for the entire scene and for each incidence angle separately. As well as extracting the phenology parameters for each incidence angle separately. In such a case, instead of 6 backscatter parameters, we would have 48 parameters (6 parameters for 8 orbits). Theoretically, this method seems very easy to understand and apply, but when it comes to the computational part, it is very time consuming and tedious to prosecute the further steps. To be more precise, implementing threshold optimization with 48 parameters manually and then validating all the results and rerunning the threshold calculation is a very intensive and lengthy procedure that requires a lot of energy, effort and time. Therefore, we have found another solution to correct this effect using a simple normalization technique.

To mitigate the effect of the incidence angle, we used a method introduced by "Lievens et al, 2019". To minimize the effect of incidence angle on the backscatter values (σ_0), we separated the backscatter values by their orbits in a 12 days cycle and repeated this for all current cycles with identical incidence angles for each of the orbits. Alongside, We calculated the mean of the backscatter over the total σ_0 values and the static bias between the σ_0 values from different orbits. The static bias was then removed by rescaling the mean σ_0 of each orbit to the overall mean and applying this mean correction to the individual σ_0 measurements [25]. This method helped us attenuate and normalize the orbital effect on the entire Sentinel-1 dataset.

3.4. Time Series Smoothing

The main purpose of filtering the Sentinel-1 backscatter time series is to remove the effects of various types of noise in the collection of the data, such as speckle noise, in both the spatial and temporal dimensions, and to reduce the short-term effects of the environmental conditions [40]. A rolling average window was applied to the data cube with a window size of 7 days (e.g., a window size of 25 days is considered to form the averages for smoothing). The rolling function is used to look at the moving average line of a pixel. In order to find the optimal value of the window size, we examined the time series diagrams using the

trial-and-error method and concluded that the appropriate value for temporal smoothing is 7. For the selection of the window type, we decided to use the "Hamming window". The Hamming window function has a sinusoidal shape and results in a broad peak with small side lobes and does not reach zero, so the signal still has a slight discontinuity. The Hamming window suppresses the nearest side lobe better, but is worse at suppressing all others. These window functions are useful for noise measurements (lumen.ni.com). At the end of this process, we obtain a smoothed backscatter time series ($\sigma_{VH-smoothed}^0$ and $\sigma_{VV-smoothed}^0$), which is used as the main input for further calculations, including the extraction of the backscatter parameters and subsequent analysis. As a result of this process, we selected some random control points, one inside the Mediterranean Sea and the other in the rice field using CLC2018, to plot the differences in temporal backscatter values before (Figure 3.2) and after (Figure 3.3) the application of time series filtering. As evident in the graph, the results were normalized and we obtained quite prosperous outcomes at the end of the smoothing process.

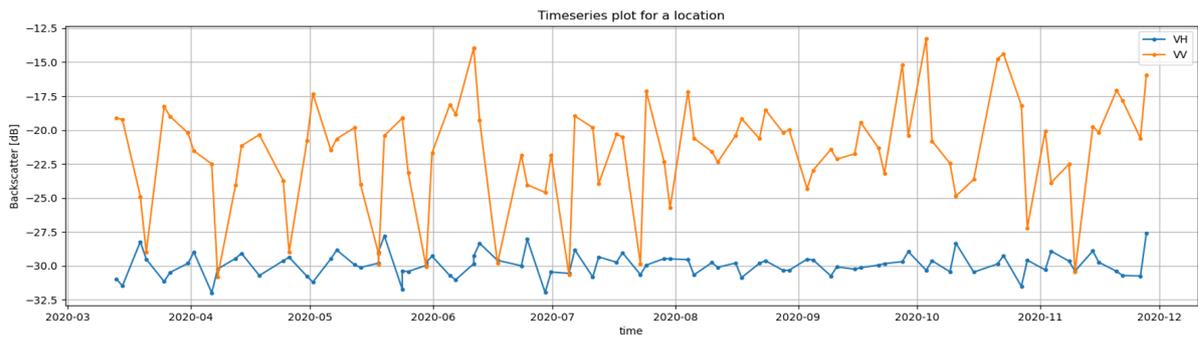


Figure 3.2: $\sigma_{VH/VV}^0$ time series graph of a random point inside the ocean, before temporal-filtering and orbit-correction.

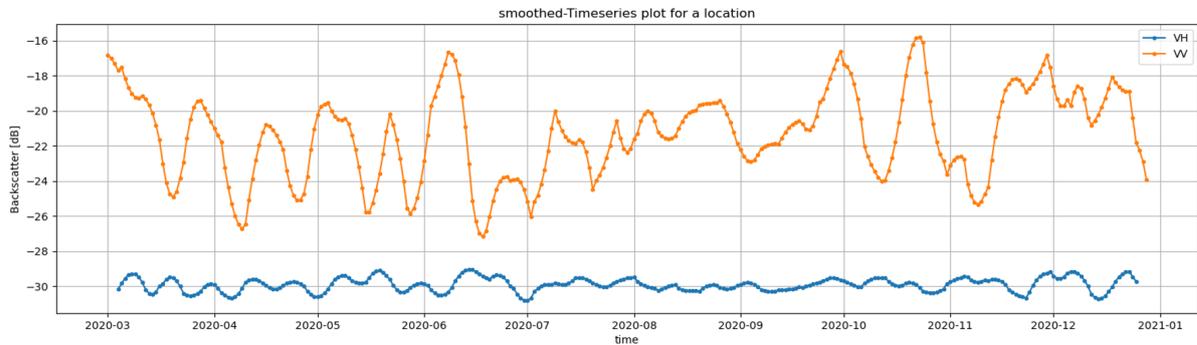


Figure 3.3: $\sigma_{VH/VV-smoothed}^0$ time series graph of a random point inside the ocean, after temporal-filtering and orbit-correction.

As it is delineated, the preprocessing and processing results for the VH-backscatter (σ_{VH}^0) are intriguing. The fluctuations of the blue line could be due to many changing factors in the environment, including temperature, humidity, and other variables. Nevertheless, since this is an open-water, it is most likely the effect of wind and waves that causes the dynamics. By exploiting these graphs (Figure 3.3), it has been concluded that the VH-polarization time series is the most advisable and sensitive kind of data for the sake of dynamics analysis and classification of a surface containing water.



Figure 3.4: $\sigma_{VH/VV}^0$ time series graph for a random rice pixel, before temporal-filtering and orbit-correction.

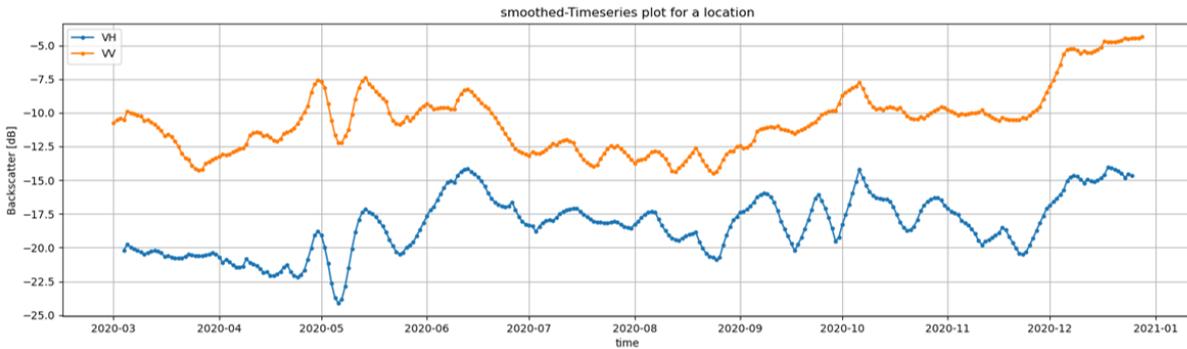


Figure 3.5: $\sigma_{VH/VV-smoothed}^0$ graph time series for a random rice pixel, after temporal-filtering and orbit-correction.

It is intuitive that the backscatter values and the time series trend of a rice pixel is totally different than a spot inside the open water because of many variables, including the level and volume of water on the surface. In general, VV-backscatter values are higher than VH-backscatter values, but the trend of the dynamics and the fluctuations are almost similar. By observing the $\sigma_{VH-smoothed}^0$ time series (the blue line), we can perceive that the soils of the rice paddies are dried at the beginning of the season. Following that, the

first abrupt drop of the $\sigma_{VH-smoothed}^0$ line, which is equivalent to the global minima, implies that the first flooding of this rice field occurs by early-May. As complementary, another local minima is the second irrigation of the field by late-May. Despite the fact that the $\sigma_{VV-smoothed}^0$ backscatter values produce completely different results, the global minima of VV-backscatter values transpire by late-March (Figure 3.5). This result is not realistic, in accordance with the Italian rice cultivation calendar, especially in the Lombardy region. We were not convinced and did not find it sufficient to draw a conclusion by just looking at these results, so we inquired more about rice pixels by comparing their VH and VV backscatter values by setting them side by side. However, we hypothesized that the classification with the VV-backscatter values engenders a considerable number of false positives. Therefore, we came up with the idea of not reckoning with the VV-polarised data at the core of our conclusions, but just as an ancillary contributing component. The main goal of plotting these graphs is to capture a distinctive pattern or a signature for the rice pixels from the time series in different polarizations and incidence angles in different locations, in order to be able to classify the rice pixels.

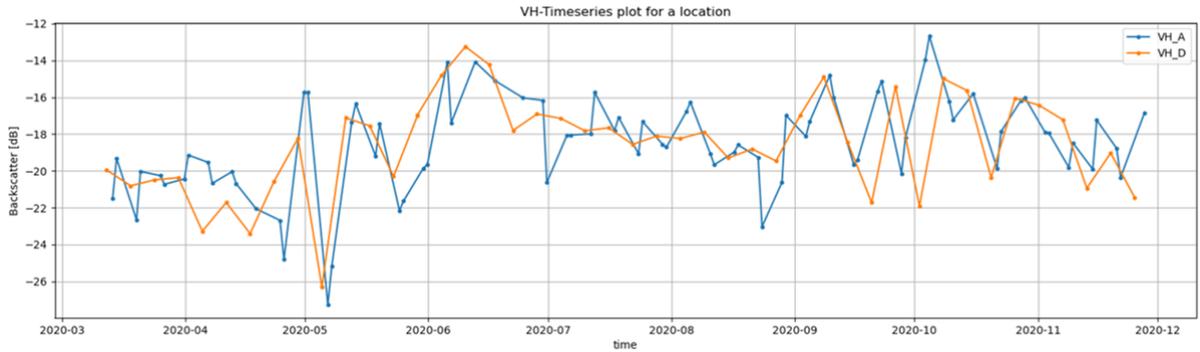


Figure 3.6: Graph of $\sigma_{VH-A/D}^0$ time series of a random rice pixel for separated ascending and descending modes, before temporal-filtering and orbit-correction.

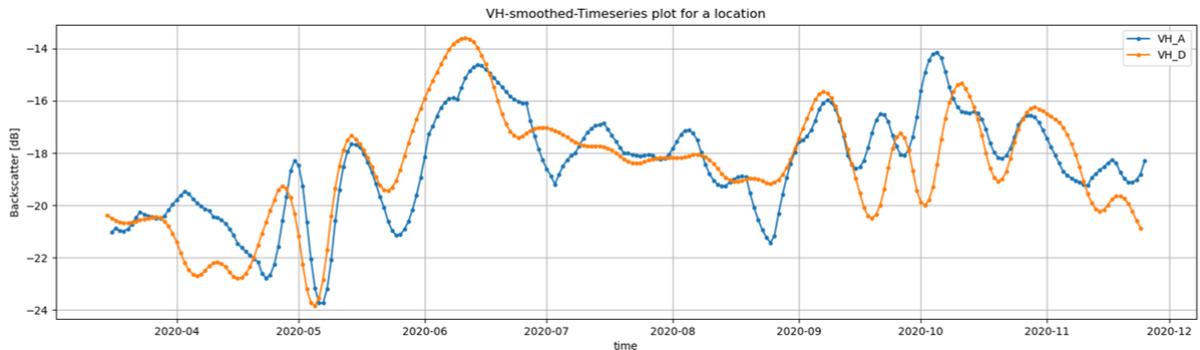


Figure 3.7: Graph of $\sigma_{VH-A/D-smoothed}^0$ time series of a random rice pixel for separated ascending and descending modes, before temporal-filtering and orbit-correction.

The subsequent proposal made by the author was to treat the backscatter values individually, separated by their incidence angles' mode in descending and ascending modes, to generate a unique temporal signature for the rice (Figure 3.7) in different locations. By deeply analysing the time series graph of both the descending and ascending modes, we can perceive that the trend of the dynamics of both lines is similar during most of the timestamps. However, there are still some vertex points in which the opening directions of the parabolas are opposite to each other (e.g., in figure 3.5, at the beginning of October and by late-September, where the red and orange parabolic lines show different directions). These abrupt changes in the magnitude of VH-backscatter has a direct relation with the nature of VH-backscatter values ($\sigma_{VH-A/D}^0$) and the SAR geometry e.g., incidence angle. Consequently, for a deeper analysis, we decided to examine the backscatter values separated by their orbits or incidence angles.

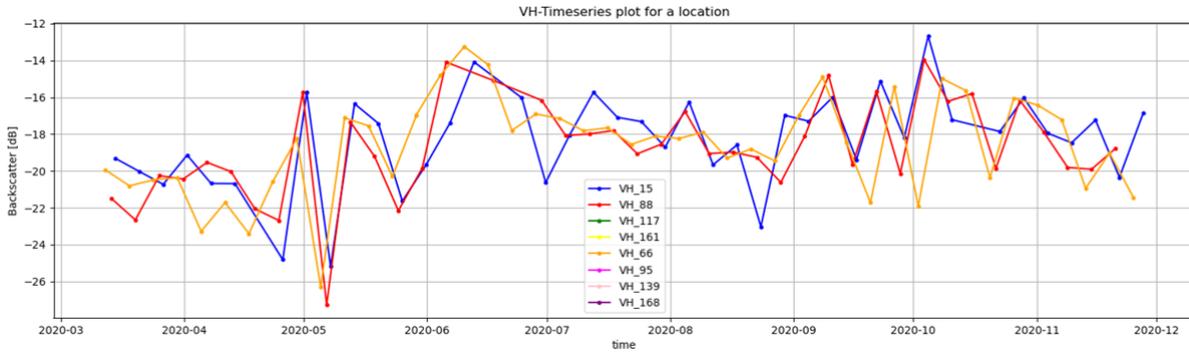


Figure 3.8: Graph of $\sigma_{VH-A/D}^0$ time series of a random rice pixel for separated orbits, before temporal-filtering.

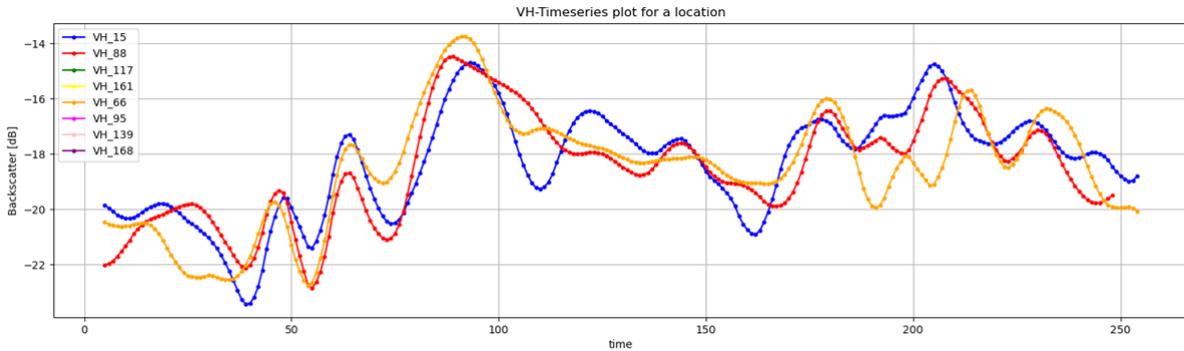


Figure 3.9: Graph of $\sigma_{VH-A/D-smoothed}^0$ time series of a random rice pixel for separated orbits, before temporal-filtering.

Sentinel-1 data are influenced by their incidence angle, as the incidence angle increases, the strength of the radar backscatter signal decreases gradually. The dynamic range over

different locations varies due to different incidence angles and farming activities. During the reproductive phase in August/September/early-November, backscatter values continuously increase until they achieve the maximum value in September/November. During the reproductive phase, VH-backscatter values of $\sigma_{VH-smoothed}^0$ oscillate between -18 and -12 dB, which is influenced by variations in incidence angle, water level in the fields, different cultivation activities, or rice varieties. The temporal sampling of S-1A and S-1B together is 5-6 days (the revisit time). Despite the temporal gaps introduced in Sentinel-1 data acquisitions, in order to provide convincing spatial/temporal information, to compensate different agricultural effects, incidence angle and acquisition mode effect and the regions with complex conditions such as the mountainous regions, we recommend to use the temporal-smoothing and orbit-correction in tandem over the entire VH-backscatter values (σ_{VH}^0).

3.5. Extraction of the statistical parameters of the Sentinel-1 backscatter

The statistical parameters of rice were defined according to the paper "Nguyen et al, 2017". Previously determined dataset over the entire study area for 2019 and 2020 was used to determine these parameters. These parameters are defined as follows:

1 - σ_{MAX}^0 : Maximum backscatter value during the rice growing season for each pixel. This parameter was acquired by applying the maximum function over the study area, pixel by pixel along the temporal dimension of the smoothed dataset $\sigma_{smoothed}^0$ Figure 3.10. In the C-band images from Sentinel-1, the vegetation pixels have higher backscatter values compared the water body pixels. This is due to the fact that water reflects microwave radiation in specular direction, i.e., away from the sensor, while vegetation as well as the soil surface scatters a proportion of the microwave radiation back to the sensor. It is expected that the σ_{MAX}^0 values are most likely, occur at the end of the vegetation period with higher values. The particular reason for the circumstance is that at the end of the growing season, all the flooded water from the irrigation disappears from the rice fields and the soils are dry and the rice plants begin to appear on the surface as a final product. In this case, the $\sigma_{smoothed}^0$ values begin to increase. The vegetative or agricultural surfaces are assumed to have their maximum backscatter values. From Figure 3.10, which shows the maximum backscatter values for each pixel during the growing season, we can demonstrate that the regions with yellow color are the areas that do not contain water or at least less amount of water, with moderate $\sigma_{smoothed}^0$ values. Consequently, these regions have a high possibility to be considered as forestry or other croplands. The areas with blue

color are water bodies with very low $\sigma_{smoothed}^0$ values, and the areas with red color are urban areas with a very high $\sigma_{smoothed}^0$ value. Our target rice is included inside the green area. This map is informative, but still we cannot distinguish between different types of vegetation and crops. The only useful information we can extract is the elimination of the urban areas and the water bodies.

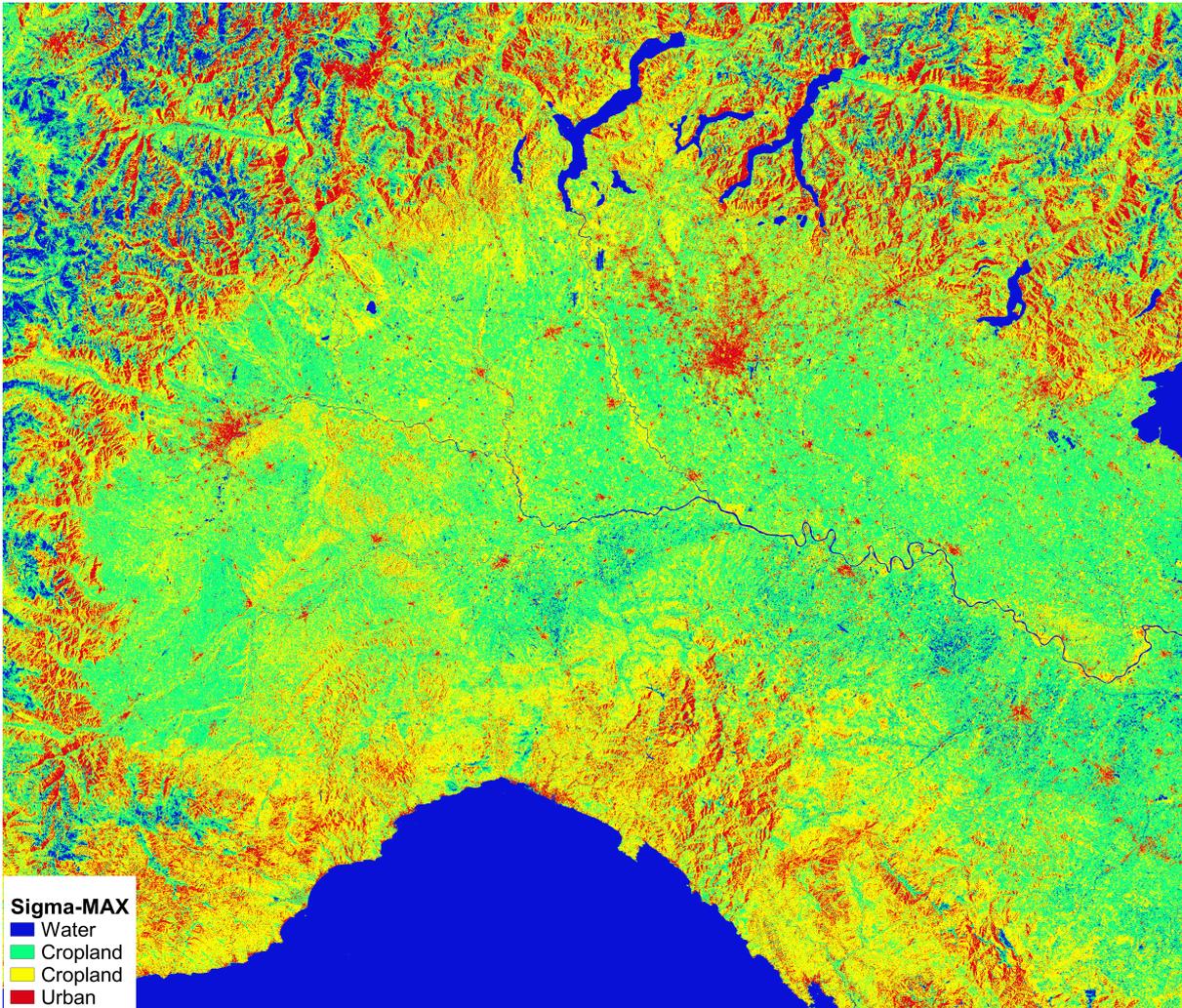


Figure 3.10: σ_{MAX}^0 or Maximum backscatter value during the rice growth season.

2 - σ_{MIN}^0 : Minimum backscatter value of each pixel during the rice growing season (backscatter at the beginning of the growing season). This parameter was collected by applying a minimum function over the temporal dimension of the smoothed dataset Figure 3.11. As we mentioned earlier, water has a very low scattering power and absorbs a large amount of microwave radiation. This actively demonstrates that the σ_{MIN}^0 values are expected to be very low σ_0 at the beginning of the rice growing season, since this is the time when the rice growing areas are flooded, thus there is a very high amount of

water on the surface of the rice fields. As time passes, the water level decreases so the σ_0 values begin to increase dramatically. Intuitively in the Figure 3.11, the blue area has higher σ_{MIN}^0 , which means that red and yellow areas have very low σ_{MIN}^0 values. Therefore, these areas are most likely the places that contain water on the surface, e.g., rice paddies, arable lands, irrigated farmland, rivers, ocean and the mountainous regions filled with snow. This map provides us with insightful information to distinguish general agricultural lands from the other classes. In the next steps, we are going to be able to make more detailed distinctions in order to extract the rice class.

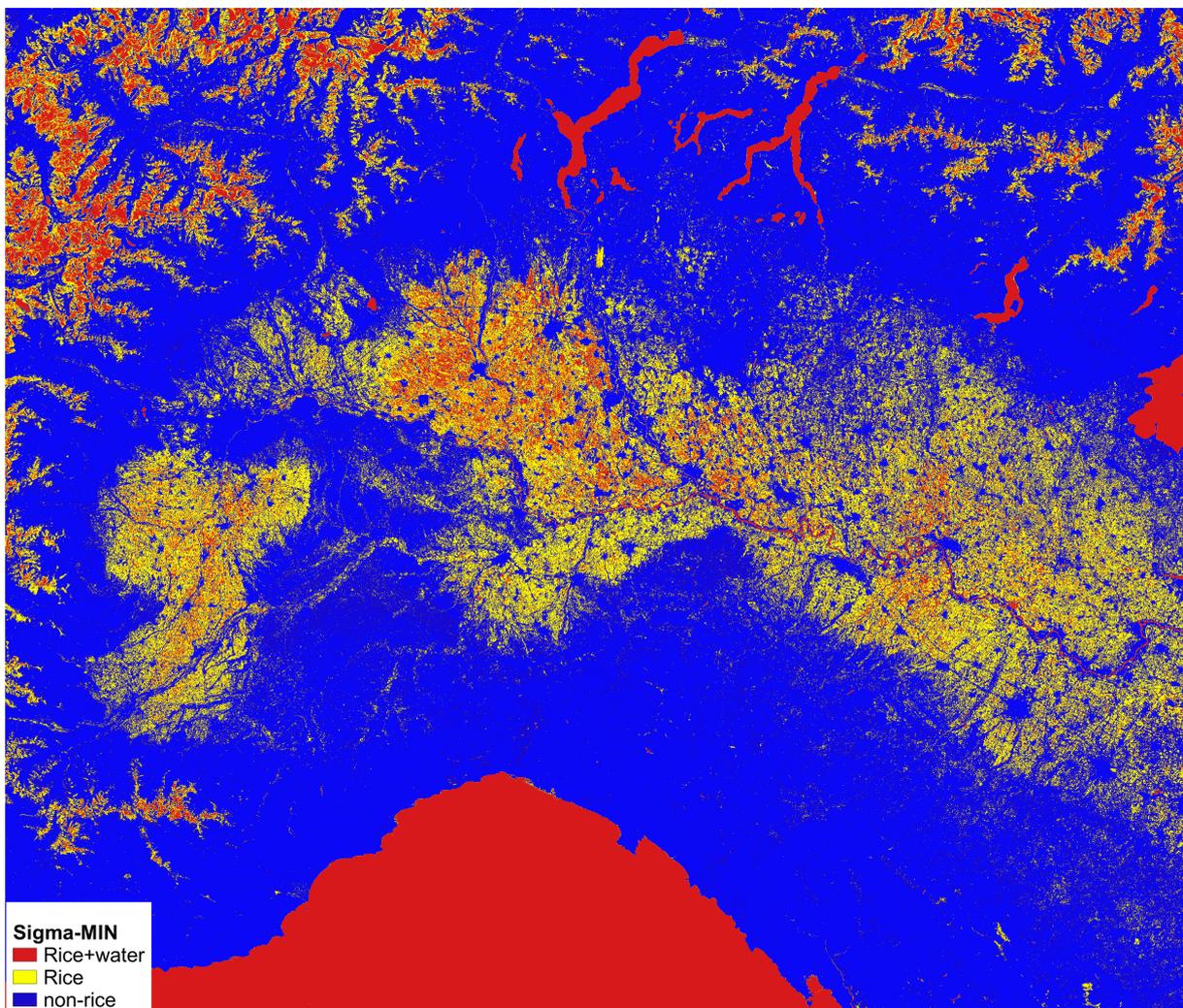


Figure 3.11: σ_{MIN}^0 or Maximum backscatter value during the rice growth season.

3 - σ_{MMD}^0 : Maximum Minimum Difference or the amplitude backscatter, is the deviation parameter, the difference between maximum and minimum backscatter values during the rice growing season Figure 3.12. This map is a key parameter in the classification process to extract useful information. It is expected that the sensitivity of rice regions is very

high, because the backscatter values vary greatly from the flooding phase to the post-harvest/harvest phase (end of the rice growing season), i.e. there is a larger range and discrepancy during the growing season in the rice regions, so the difference will be high. The blue colored regions are assumed to be the areas with higher sensitivity or higher deviation, with higher possibility to be classified as rice. The red, green and yellow regions have lower sensitivity or deviation of $\sigma_{smoothed}^0$ values, for this reason they will be classified as non-rice class.

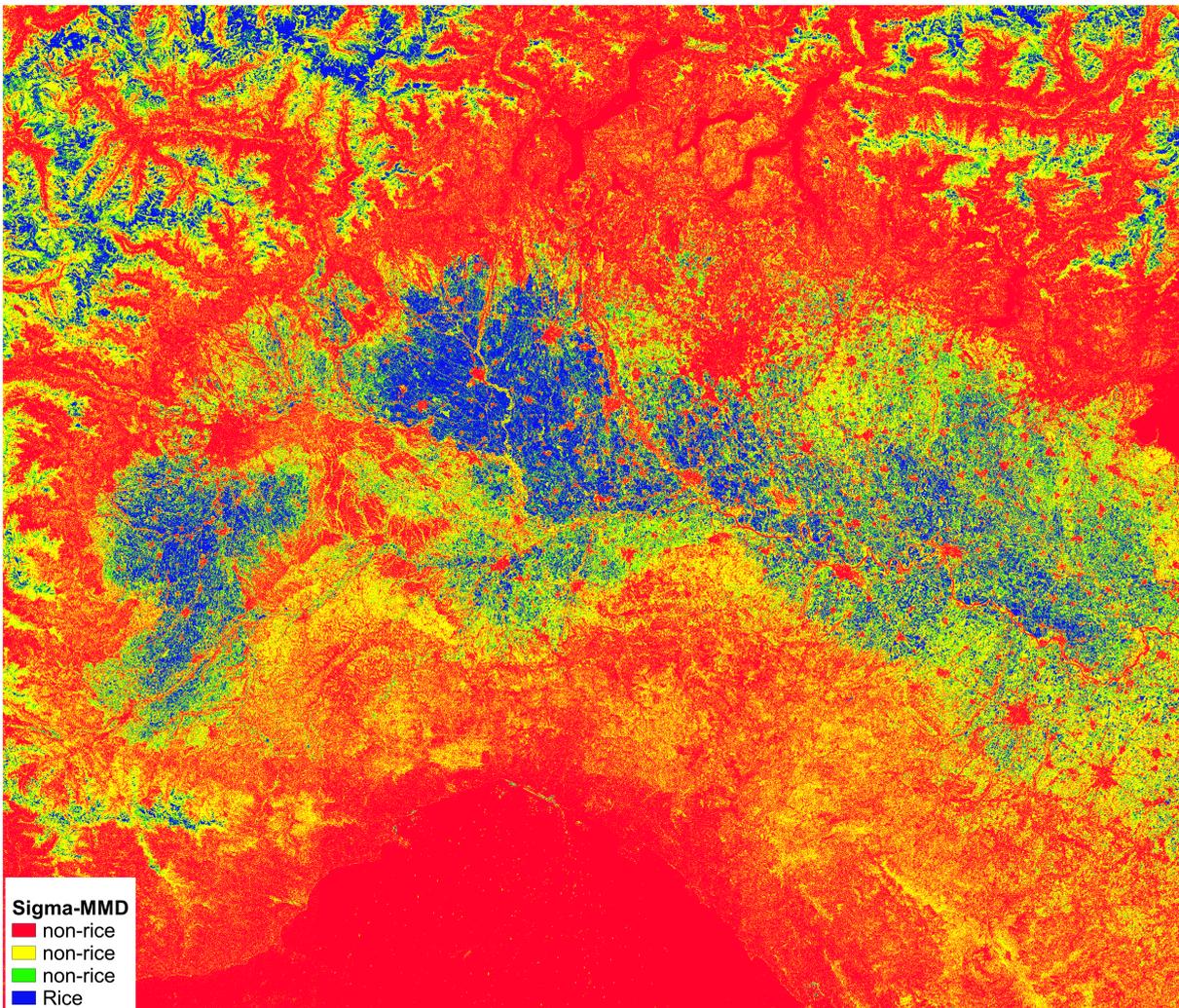


Figure 3.12: σ_{MMD}^0 or the deviation map. All agricultural regions have high σ_{MMD}^0 than e.g. forest, water or urban areas.

4 - **DoM**: Date of Maximum backscatter at the end of the rice growing season, when the rice plant reaches its final stage post-harvest/harvest Figure 3.13. This map depicts the day of the year (DOY) with the maximum $\sigma_{smoothed}^0$ value. Thus, we expect the peak of $\sigma_{smoothed}^0$ rice fields (within one rice-growing cycle) to occur at the end of the growing

season by September/October or October/November. The blue regions in the map have a high DoM value, so they are more likely to be classified as rice. Red, green and yellow areas have lower DoM, meaning that their σ_{MAX}^0 occur earlier than expected date during the start of rice growing season, which contradicts our rules, so we will classify them as non-rice. The queries in this map are intuitive, so we cannot rely only on this map to extract the rice pixels. We need to combine the other information and maps with this map to achieve a successful classification.

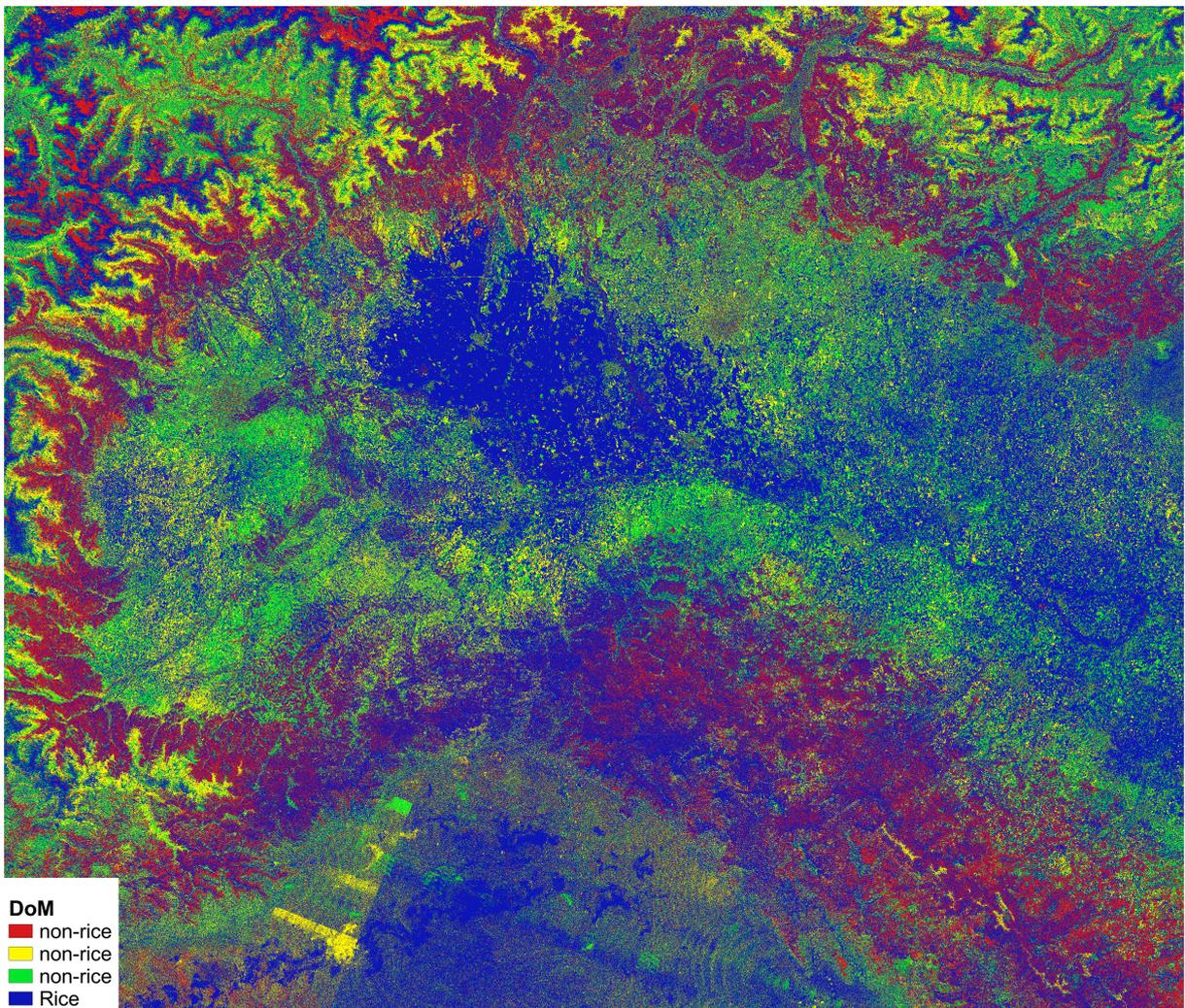


Figure 3.13: The date when the backscatter coefficient reaches a maximum value is defined as the global maxima in $\sigma_{smoothed}^0$ time series, during the rice growing season. This date must come after the date of the start of the season, where it reaches its global minima.

5 - **DoS**: Date of the beginning of rice growing season (flooding and seeding) or the date within the rice growing season when the pixels have their minimum backscatter values Figure 3.14. This map represents the day of the year (DOY) with the minimum $\sigma_{smoothed}^0$

for each pixel. We assume that the rice pixels have their minimum values at the beginning of the growing season by April/May/June due to the flooding and presence of water in the rice fields. The green and yellow areas on the map have very low DoS and are potentially rice growing areas. The red and blue areas have their minimum values at the middle or the end of the growing season, so they are classified as non-rice. therefore, by setting a threshold, we separated the rice growing areas from the rest of the land covers.

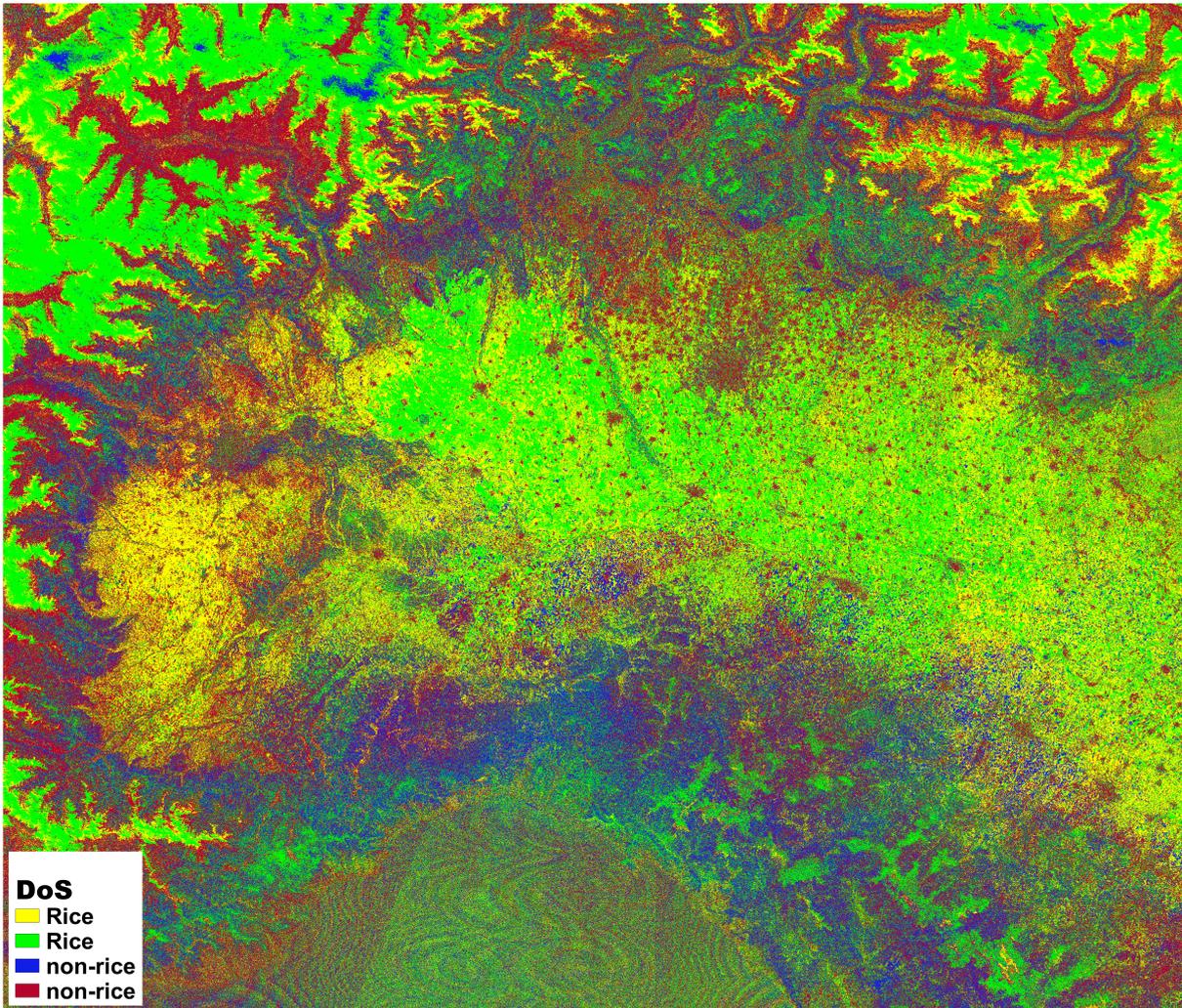


Figure 3.14: During the growing season, the date of the beginning of season is defined as the global minima in $\sigma_{smoothed}^0$ time series.

6 - **LoS**: Length of the Season, is the difference between the start of rice cultivation and the end of the season for each pixel, i.e. the number of days between DoM and DoS (Figure 3.15). This map shows the number of days from the beginning of the cultivation (flooding/seeding) till the harvest phase, i.e., it shows the past days between two main deviations of backscatter values or days between the 2 global extremas. According to the

rice cultivation calendar, the rice area have a lower seasonality compared to the other crops (the minimum is 50 days and the maximum is 3 months). In the figure 3.15, the areas with blue and green colors are meant to be rice, with a lower LoS (Length of the Season) number compared to the yellow and red regions, with a higher length of the season, so we classify it as non-rice.

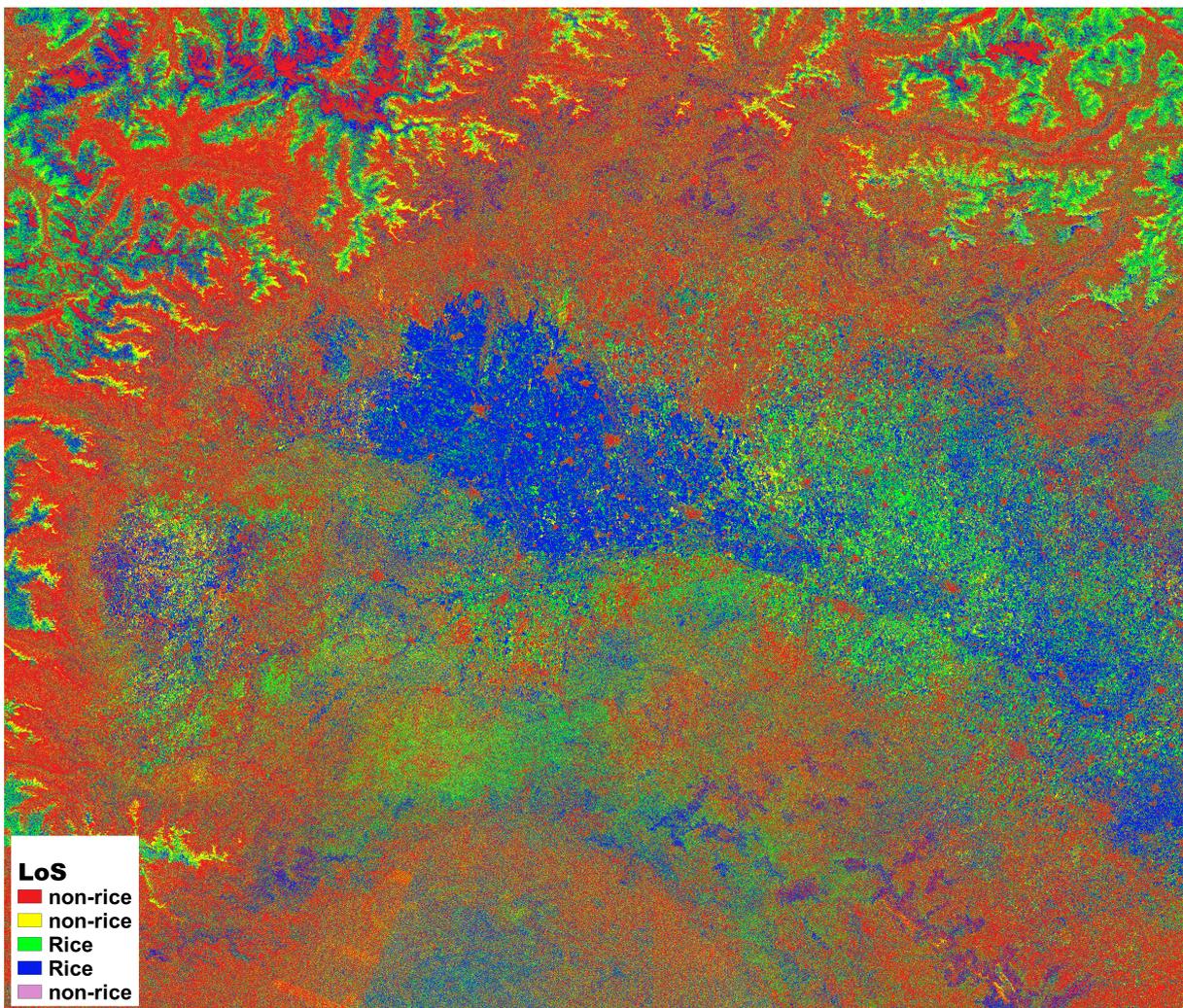


Figure 3.15: The length of the season is defined as the number of days difference between DoM and DoS.

Finally, after calculation of all the 6 parameters for both VH and VV polarized datasets and for 2019 and 2020 separately, we can delineate the "rice" area which meets all of our conditions and set of rules according to the defined rice cultivation calendar and expertise knowledge in combination with the "Nguyen and Wagner" algorithm [40], and classify the rest of the regions in a general "non-rice" class. To eliminate unrealistic peaks, a threshold for VH-backscatter is required.

3.6. Parameters Definition and Temporal Rice Classification

After extracting the corresponding statistical parameters and delineating a precise rice cultivation calendar for both years 2019 and 2020, we determined the primary threshold values for identifying the potential rice cultivation areas. Here, we perform a pixel-based classification with moderate thresholds from an agronomic point of view, without optimising the thresholds (this stage was performed in the next steps to achieve higher levels of accuracy, here we only want to detect the potential rice growing areas).

We start by using the same parameter thresholds as suggested by "Nguyen et al, 2017" and then update them according to our estimated rice cultivation calendar and obtained performance metrics for 2 subsequent years. The value of the thresholds may change due to the different irrigation methods in different locations, different geographical location, different growing seasons, the accuracy we want to achieve in the end, the reference data (their resolutions and accuracy, i.e. how reliable they are, which has many shortcomings in our case), the smoothing method and the parameters we choose for smoothing the dataset and orbit correction method, and so on. Since rice may have different seasonality, the established thresholds can change each year. This is a crucial point to consider by providing different rice growing calendars for different years and updating the thresholds of the corresponding parameters accordingly. Based on the visual interpretation of the parameters (done via Qgis) and the basic preliminary information that we have, the thresholds that we established considering the article of "Nguyen et al, 2017" and our rice cultivation calendar are presented in the following table 3.1.

Initial thresholds chosen from the knowledge-based decision rules, customised with the study area (Po catchment) and "Nguyen & Wagner"

	Thresholds	description
σ_{MAX}^0	$-17 < \sigma_{MAX}^0 < -13$	The peak of VH-backscatter (global maxima) at the end of the growing season.
σ_{MIN}^0	$\sigma_{MIN}^0 < -23$	The valley of VH-backscatter (global minima) at the begin of the growing season.
σ_{MMD}^0	$8/8.5/9 < \sigma_{MMD}^0$	Deviation map or amplitude backscatter, the difference between the date of maximum backscatter and the begin of the growing season.
<i>DoM</i>	$210 < DoS < 330$	Day of Maximum backscatter, The rice harvesting phase in September/October or October/November.
<i>DoS</i>	$90 < DoS < 180$	Day of the beginning of the season, when the flooding and seeding starts. In Italy it usually starts from early/mid April till June.
<i>LoS</i>	$50 < LoS < 210$	Temporal distance or number of days difference, has to be greater than the shortest possible rice growing cycle and smaller than the longest possible rice growing cycle.
Base-Line	20 dB	Temporal evolution of the backscattering coefficients derived from VH-polarization (where, σ_{VH}^0 is base line)

Table 3.1: Defined threshold values by "Nguyen et al, 2015, 2016, 2017, 2018". The units of the thresholds are mentioned in dB and DOY (Day Of Year).

By applying the generalized thresholds as components of our decision-rule tree to the 6 extracted statistical parameters, we obtain the primary results of the classification map with "rice" and "non-rice" classes (see the following Figure 3.16).

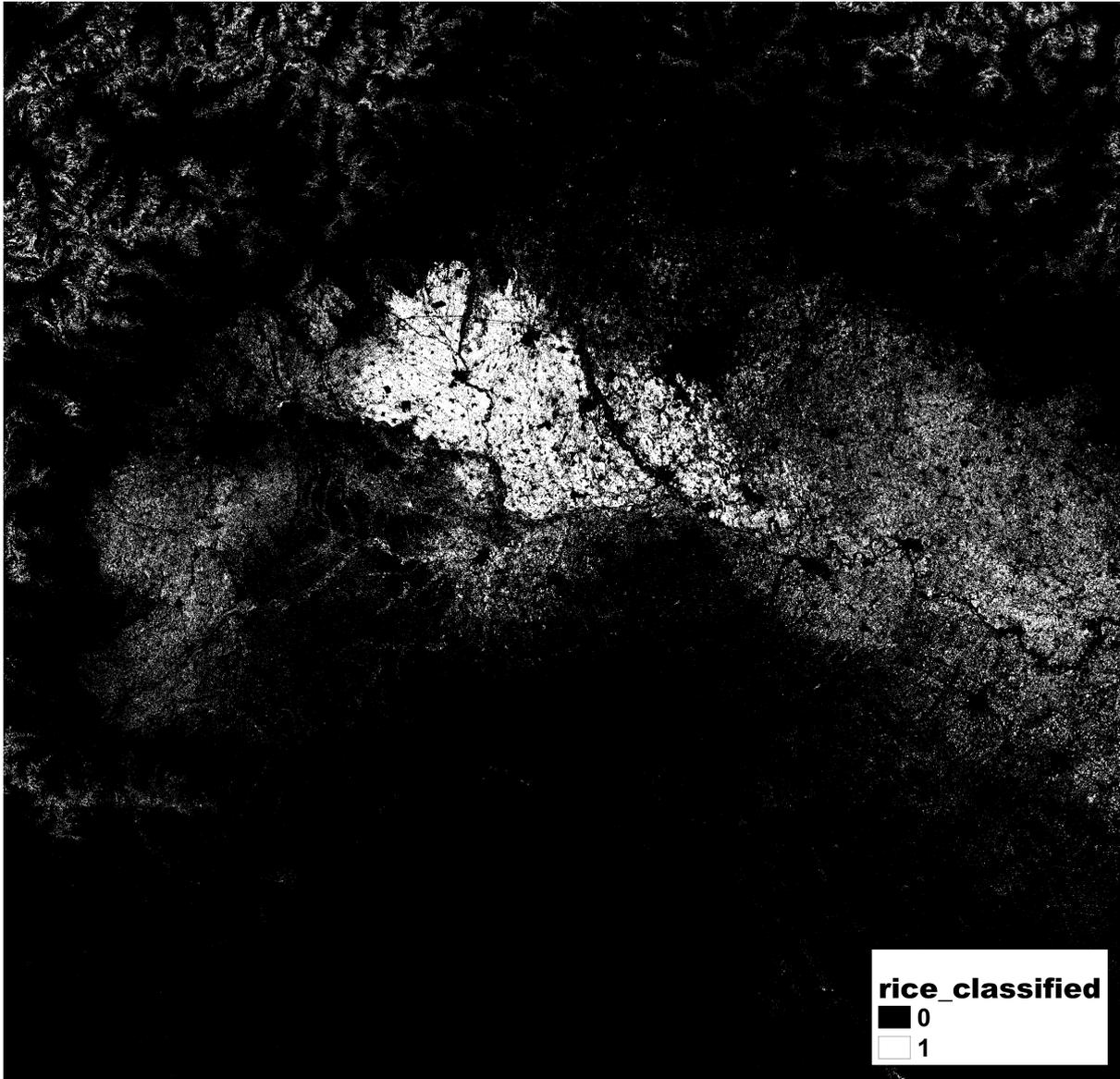


Figure 3.16: first generated rice map by defining the thresholds on the statistical parameters of rice.

The first thresholding step was done according to our expert knowledge and referring to "Nguyen et al, 2017 [40]". "Nguyen et al, 2017 [40]" introduces the selection of generalized threshold values for the identification of potential rice growing areas over eight different study areas on a small scale. In their case study, according to the high backscatter signal

variability in 8 different areas, they were trying to introduce the best generalized threshold values that could classify the rice croplands with the highest possible precision in all the study areas. In our case study, due to the high temporal variability in the SAR backscatter signal across the big-scaled data, the raw output from the general thresholding of phenological parameters contained some noisy pixels. This implies that most fields were not fully classified as rice or non-rice. In this case, to reduce the number of false positive pixels, the thresholds must be optimized considering the data acquisition and the constraints of the local rice cultivation calendar. The following step is known as "threshold optimisation".

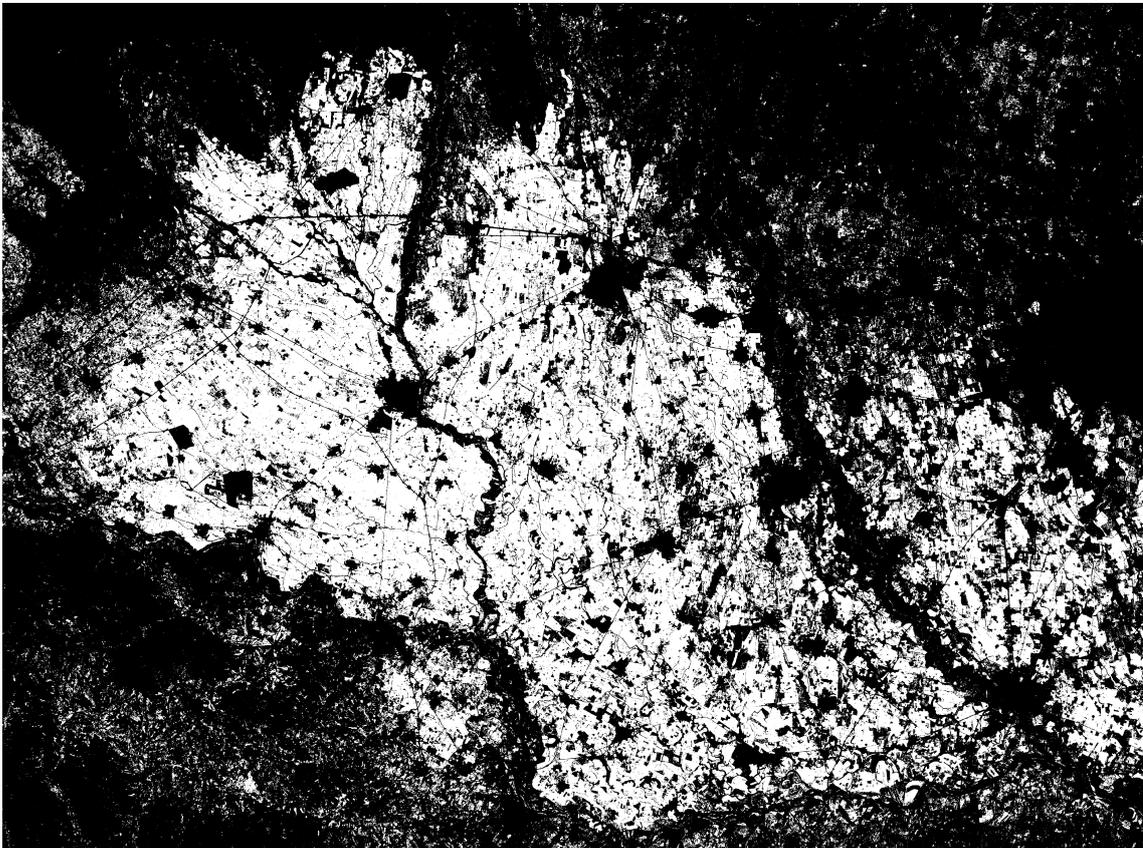


Figure 3.17: Zoomed in rice map with the details.

Extracting the statistical parameters of a random rice pixel

σ_{MAX}^0	σ_{MIN}^0	σ_{MMD}^0	DoM	DoS	LoS
16.8	-25.3	9.6	187	162	109

Table 3.2: A rice pixel that complies with all the threshold-based conditions set out.

For validation of potential rice fields after a general thresholding of the dynamic range backscatter parameters to identify image pixels that change more than the defined threshold, we compared the results with the Corine Land Cover 2018 map, which is our reference data. While comparing the acquired rice map with the reference data (CLC2018), we found some inconsistencies between the two maps. For example, some pixels in the mountain region were misclassified as rice this could be due to the complex weather situation in the alpine mountains. The other misclassified pixels refer to arable cropland and other agricultural fields due to the regular and seasonal irrigation in these regions, so our algorithm was not able to distinguish them from rice fields because they behaved similarly at the previously defined thresholds. To remove these misclassified pixels, we need to be more precise and go through the details of the rice growing calendar. In addition, we can use a threshold optimization phase to distinguish rice paddies from all other classes. According to "Nguyen et al", a generalized threshold for rice fields can only be determined if the optimal SAR data collection is guaranteed, otherwise the thresholds should be optimized by considering the data and the rice cultivation calendar [38] [39] [40].

Validation results of the generalized threshold phase:

Precision-score = 28.20%

Recall-score = 73.80%

Accuracy = 92%

F1-score = 40.80%

The results are not satisfactory because the accuracy is very low (the Precision-score is about 28% and this is not the accuracy we want). In the next steps we will see that the results improve as well.

3.7. Threshold Optimization (decision-rule setting)

The thresholds we have determined must be optimized by taking into account the data acquisition and the constraints of the local cultivation calendar for rice, by attempting to minimize the error and increase the overall accuracy. Optimum decision-rules setting employs the same parameters and thresholds as in the previous section are used at the beginning. Then the thresholds are changed and updated in each iteration, trying to converge the error set to zero. The way we can distinguish rice may vary. We tried to analyze 6 backscatter statistic parameters separately to structure the decision-rule tree by adding rules at each analysis. By manually setting the thresholds and calculating

the precision value for each trial, we were able to obtain the following results with the thresholds. All these steps were done separately for both 2019 and 2020 and VV and VH polarization, so we would end up introducing 4 different rice maps with different thresholds.

Original optimized thresholds by "Nguyen & Wagner" in eight region

	VH - Thresholds
σ_{MAX}^0	$-19 < \sigma_{MAX}^0 < -13$ (dB)
σ_{MIN}^0	$\sigma_{MIN}^0 < -20$ (dB)
σ_{MMD}^0	$8.5 < \sigma_{MMD}^0$
DoM	$210 < \text{DoM} < 330$ (DOY)
DoS	$90 < \text{DoS} < 180$
LoS	$50 < \text{LoS} < 120$

Table 3.3: Defined threshold optimization vales for $\sigma_{VH-smoothed}^0$ by Nguyen et al [40].

Thresholds after the optimization phase in Po catchment

	VH - Thresholds
σ_{MAX}^0	$-18 < \sigma_{MAX}^0 < -11.5$
σ_{MIN}^0	$\sigma_{MIN}^0 > -22.5$ (dB)
σ_{MMD}^0	$9 < \sigma_{MMD}^0 < 18$
DoM	$220 < \text{DoM} < 333$ (DOY)
DoS	$95 < \text{DoS} < 180$
LoS	$55 < \text{LoS} < 220$

Table 3.4: Defined threshold optimization vales for $\sigma_{VH-smoothed}^0$ backscatter coefficients.

Thresholds after the optimization phase in Po catchment

	VV - Thresholds
σ_{MAX}^0	$-15 < \sigma_{MAX}^0 < -5.5$ (dB)
σ_{MIN}^0	$-25 < \sigma_{MIN}^0 < -12$
σ_{MMD}^0	$6 < \sigma_{MMD}^0$
DoM	$92 < \text{DoM} < 333$ (DOY)
DoS	$92 < \text{DoS} < 180$
LoS	$92 < \text{LoS} < 333$

Table 3.5: Defined threshold optimization vales for $\sigma_{VV-smoothed}^0$ backscatter coefficient.

The CLC2018 map is used to validate and express the compartment boundaries of the pixels' changes below and above the thresholds, which are classified as potential rice crop-land areas.

3.8. Validation and accuracy assessment

After building the classification model in order to increase the evaluation of our model in terms of accuracy in the threshold optimization phase, and for accuracy evaluation and validation of the final classification results (rice and non-rice), standard accuracy metrics were used. We calculated the standard confusion matrix to obtain the values of other accuracy metrics, including precision-score, recall-score, F1-score, commission error, omission error, and Kappa coefficient.

Precision: Appropriate when minimizing the false positives is the focus.

Recall: Appropriate when minimizing the false negatives is the focus.

F1-score: Appropriate when minimizing the total accuracy is the focus.

4 | Results

4.1. Final Results

Rice cultivation is one of the most fateful sectors amongst other agricultural branches in Italy for the entire European Union because rice is a primary source of food for more than half of the world's population. Accordingly, the evaluation and monitoring of rice distribution with up-to-date and accurate results is essential and ubiquitous not only for rice monitoring purposes but also for other related research topics and study areas. Every year, many projects and research are carried out on rice monitoring and analysis in different regions, introducing new algorithms or updating the previous ones to obtain better results with higher accuracy. To this end, we decided to conduct our thesis on rice acreage monitoring in one of the largest European rice producing countries, Italy. Through the temporal assessment of two consecutive years using data acquired by the Sentinel-1A B satellites with a C-band SAR (Synthetic Aperture Radar) sensors mounted on it, launched by ESA (European Space Agency). Thanks to the high spatial, temporal resolution (20 meters and 6 days, respectively) and cross-polarized (VH) data collected by the Sentinel-1 satellites (which has a high capability in rice classification), we are able to investigate and analyze the whole rice growing season with an acceptable accuracy and finally improve and strengthen a predefined classification algorithm by "Nguyen et al, 2017" to preferentially distinguish the rice growing areas from the other classes.

The method of "Nguyen et al" adopts a phenology or rule-based plant classifier by using phenological parameters derived from Sentinel-1 VH backscatter time series as input to the decision-rule based classifier [39] [40]. Here, we renamed the phenological parameters to backscatter statistics and recorded them as mentioned in Chapter3. We preprocessed the entire dataset for two consecutive years (2019 and 2020) separately, applied orbit correction and temporal filtering, and structured the dataset in a way that makes it easier to run further processes and perform various analyzes using Python libraries such as numpy, pandas, geopandas, gdal, xarray, netCDF, and yeoda, etc. In the end, we have obtained a neat and proper set of data with a clear and concise structure that is easy

to maneuver. To scrutinize the results, our first analysis step was the interpretation of selected pixel time series that are certainly located in rice fields as well as in the other agricultural fields. The depicted plots show that we have achieved the desired results. The graphs are shown as follows.

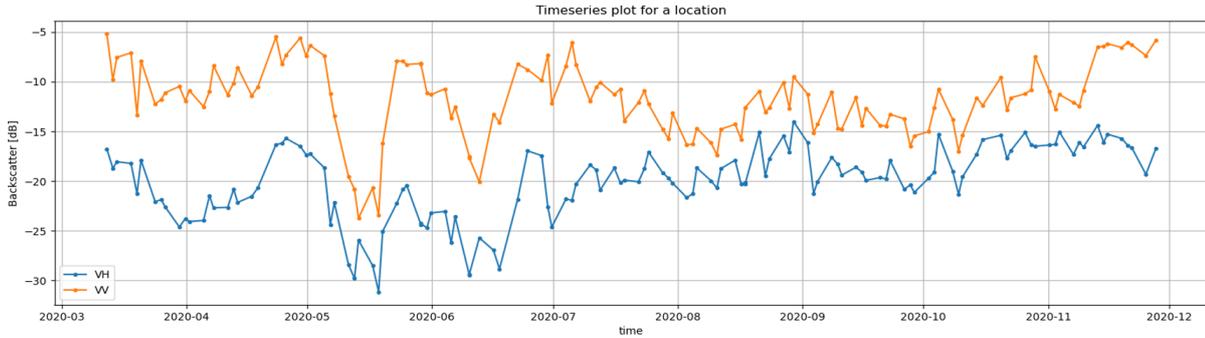


Figure 4.1: σ_{VH}^0 -preprocessed and σ_{VV}^0 -preprocessed values of a single random pixel to be classified as rice before applying orbit correction and temporal filtering.

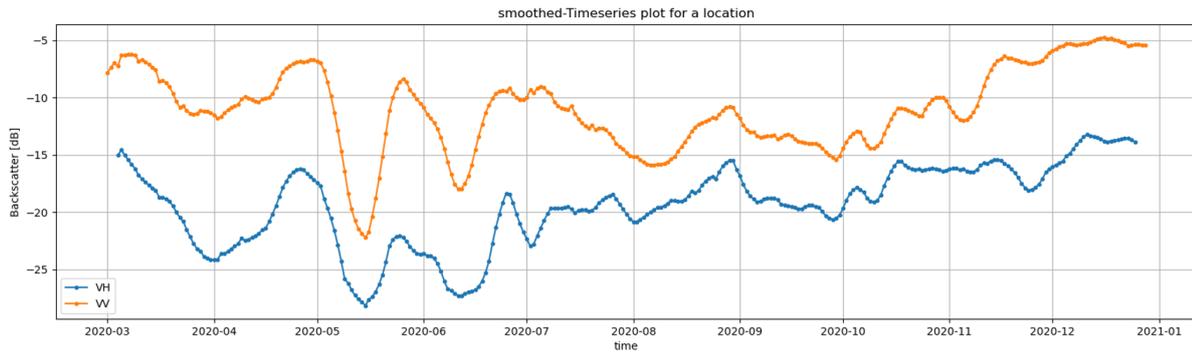


Figure 4.2: σ_{VH}^0 -smoothed and σ_{VV}^0 -smoothed values of the same pixel to be classified as rice after applying the entire preprocessing and processing phases. As you can see, the time series values are smoothed and do not contain keen fluctuations anymore.

Figure 4.2 shows the temporal-orbital smoothed backscattering values ($\sigma_{smoothed}^0$) of a pixel classified as rice. By analysing this graph, we capture two important anomalies that represent one of the key points of our analysis. The two valleys (local minima) of the time series graph, one during mid-May and the other around mid-June, which corresponds to the inundation/seeding phase in our rice growing calendar. The second local minima is most likely due to the second irrigation of the fields. Then, the backscatter values begin to gradually increase until it reaches the vegetative phase, which occurs in August/September (first peak) according to the rice cultivation calendar. Finally, the harvesting phase during October/November or sometimes in December, when the backscatter values reach their

maximum values, as shown in the graphs. In the vegetative phase, backscatter value increases with the growth of vegetation and finally the images from SAR sensor show no significant difference between rice and other agricultural fields or other vegetated lands. Furthermore, during the post-harvest or harvest season, the growth of rice plants in rice fields and the absence of water minimize the double-bounce effect of SAR signals, which explains the decrease in backscatter values during this phase. During this period, VH-backscatter values fluctuate between the range of -17 and -13dB according to "Nguyen et al". This phenomenon can also be observed in our VH curve Figure4.2.

The reference dataset we used is Corine Land Cover2018 (CLC2018), which has a coarse resolution compared to our main dataset and therefore does not include details related to the channels between rice paddies or the gaps between two widely separated paddies. In addition, we found that CLC2018 has some inconsistencies with the real world when we analyze the data in Sentinel Hub pixel by pixel (referring to the Sentinel-2 L1C mission and the NDWI (based on the combination of bands B3 - B8 / B3 + B8) and SWIR (based on bands 12, 8A, 4) indices). And by acquiring more data from the local websites of the Piedmont and Lombardy municipalities (www.ersaf.lombardia.it), we understood that there are a lot of discrepancies. And as long as the data acquired from the local municipalities were fragmented and had some shortcomings, since some localities did not cover our whole study area, we could not use them as reference maps for further evaluation. But we referred to these maps in our analysis to identify and justify the places where our algorithm fails compared to CLC2018 by further reasoning. In this case we continued working with the same CLC2018 reference data. Nevertheless, we already knew that the final results we would collect would be less accurate because of the problems mentioned above. The threshold optimization phase is closely related to the reference data (CLC2018) that we present, and it is almost the phase where we introduce the final thresholds by evaluating them at each iteration with the CLC2018 rice map. consequently, we change the thresholds when the rice map we get matches CLC2018 or not, in order to achieve higher accuracy. In other words, this reference map plays a vital role in the whole process. If it is problematic for some reason, we might lose part of the accuracy by ending up with different values for threshold optimization.

The classification results were evaluated using the CLC2018 product with an overall accuracy ($\text{accuracy} = (\text{tp} + \text{tn}) / (\text{tp} + \text{tn} + \text{fp} + \text{fn})$) more than 84% for the extracted rice cultivated regions, and 97% for the entire study area. Respectively, recall score ($\text{recall} = \text{tp} / (\text{tp} + \text{fn})$) equal to, 76% and 52%. As predicted before, the final result of the evaluation might be less accurate than the situation if we had more reliable reference

data with the same spatial resolution (20m). Overall, Sentinel-1 based classification maps reveal much more details compared to the rice field class acquired from CLC2018 data [40]. However, our final results are convincing and acceptable as follows.

class	Rice	Non-rice	Total	Producer accuracy
Rice	5028265	1849127	6877392	73.2%
Non-rice	1594310	12248298	13842608	90%
Total	6622575	14097425	20720000	
User accuracy	76%	87%		Kappa=0.72

Figure 4.3: Classification results of $\sigma_{VH-smoothed}^0$ after masking the mountains and limiting the area by cropping the dataset to a smaller region where contains rice fields rather than other crops for comparison purposes (Figure 3.17)

class	Rice	Non-rice	Total	Producer accuracy
Rice	3720494	2846745	6567239	57%
Non-rice	3508389	214924372	218432761	98.5%
Total	7228883	217771117	225000000	
User accuracy	52%	99%		Kappa = 0.6

Figure 4.4: Classification results of $\sigma_{VH-smoothed}^0$ of the entire region with 225M pixels over 9 tiles (Figure 2.1).

class	Rice	Non-rice	Total	Producer accuracy
Rice	4503631	21683153	26186784	20%
Non-rice	2136469	196676747	198813216	99%
Total	6640100	218359900	225000000	
User accuracy	68%	90%		kappa = 0.25

Figure 4.5: Classification results of $\sigma_{VV-smoothed}^0$ of the entire region.

		actual	
		Rice	Non-Rice
predicted	Rice	True Rice (TP)	false Rice (FP)
	Non-Rice	False non-Rice (FN)	True non-Rice (TN)

Figure 4.6: The defined confusion matrix for our case study.

Tables 4.3, 4.4, 4.5 represent the confusion matrix, which includes the values of True-Positive (true rice), False-Positive (false rice), False-Negative (false non-rice) and True-Negative (true non-rice) 4.6, by counting the number related to each component of our rice-map. The rows refer to the actual values and the columns are related to the predicted ones. The goal is to make a trade off between recall and precision or increase the amount of F1-score in general. In addition, the producer accuracy and the user accuracy are computed from the confusion matrix values. The final accuracy of each table is computed as follows:

Table 4.3 : F1-score = 75%	Accuracy = 84%	Precision = 73.2%	Recall = 76%
Table 4.4 : F1-score = 54%	Accuracy = 97%	Precision = 57.0%	Recall = 52%
Table 4.5 : F1-score = 30%	Accuracy = 90%	Precision = 20.0%	Recall = 68%

According to the final evaluation results (Figures 4.4 and 4.5), the cross-polarised (VH) data were used for the further classification approaches due to the superior rice classification abilities of the SAR cross-polarised measurement, compared to the co-polarised (VV) data with promising accuracy, with Kappa equal to 0.6 and 0.72. To assess the representativeness of the presented rice cultivation pattern and algorithm, the classification accuracy obtained from the VV-polarized data was meager and did not meet our conditions and expectations in the evaluation context, with a Kappa equal to 0.25, which is negligible, and we would like to achieve a higher kappa and precision. We strongly illustrate that the Sentinel-1 time series with VH-polarisation will provide satisfactory results for the development of an operational continental-scale rice monitoring and mapping system.

4.2. Comparison to Nguyen and Wagner's Approach

In this thesis, a method developed by "Nguyen and Wagner, [40]" for the Po River basin in northern Italy was applied and further developed to accurately distinguish rice fields from fields with other crop types. The study by "Nguyen and Wagner" was conducted for multiple, relatively small test areas in the Mediterranean region. We focus on northern Italy, more specifically the Po River basin. "Nguyen and Wagner" used Sentinel-1A data from 2015, and here we will apply the algorithm to Sentinel-1A and Sentinel-1B data from two years (2019 and 2020). We start with the same phenological parameters and thresholds as proposed by "Nguyen and Wagner", but update them according to our rice growing calendar and obtained performance metrics. By applying the final algorithm to data from two seasons, it will be possible to evaluate the robustness of the classification and interpret any differences in the results of the two seasons. Finally, a local extrema approach was implemented and tested to further improve the classification accuracy (this further step will be discussed in more detail in the next chapters). The collected results of the confusion matrix for Italy only and the accuracy according to the method of "Nguyen and Wagner" are presented in the following figure 4.8.

class	Rice	Non-rice	Total	Producer accuracy
Rice	1404	56	1460	86.2%
Non-rice	356	3392	3748	90.5%
Total	1760	3448	5208	
User accuracy	79.8%	98.4%		Kappa=0.70

Figure 4.7: The confusion matrix of the classification obtained by "Nguyen & Wagner".

Accuracy = 83.9% Recall = 79.8% Precision = 98.4% F1-score = 86.2%

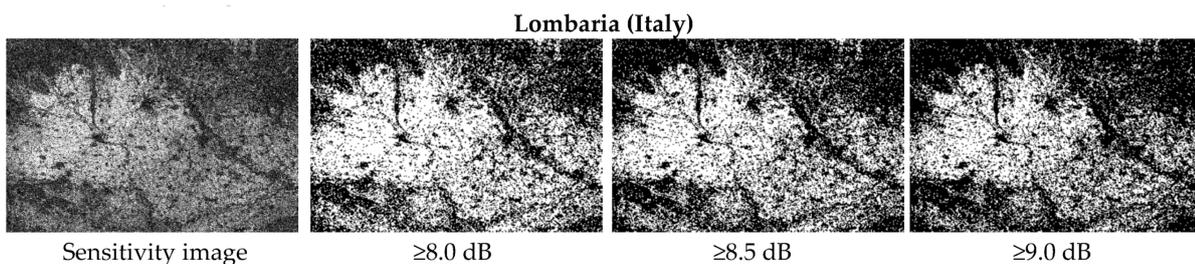


Figure 4.8: Rice classification map obtained by "Nguyen & Wagner".

In the algorithm of "Nguyen and Wagner", they have defined the threshold parameters using the seasonal VH time series and the results of the smoothed backscatter profiles of Italy. And we defined the threshold by analysing different indices of various dynamic maps from the sentinel-hub playground website (www.sentinel-hub.com/explore/sentinelplayground/), including the NDWI index of the Sentinel-2 L2A mission. For this purpose, we analysed the different days during the rice growing season in Italy by inspecting the region around the Po Valley, and we were able to define a rice cultivation calendar that we previously introduced 2.3. When comparing the study area of "Nguyen et al, 2017" with ours, they have employed a small part of northern Italy in relation to the area where we conducted the work, which includes nine tiles, as shown in the Figure 4.9 with yellow and white colours, respectively.

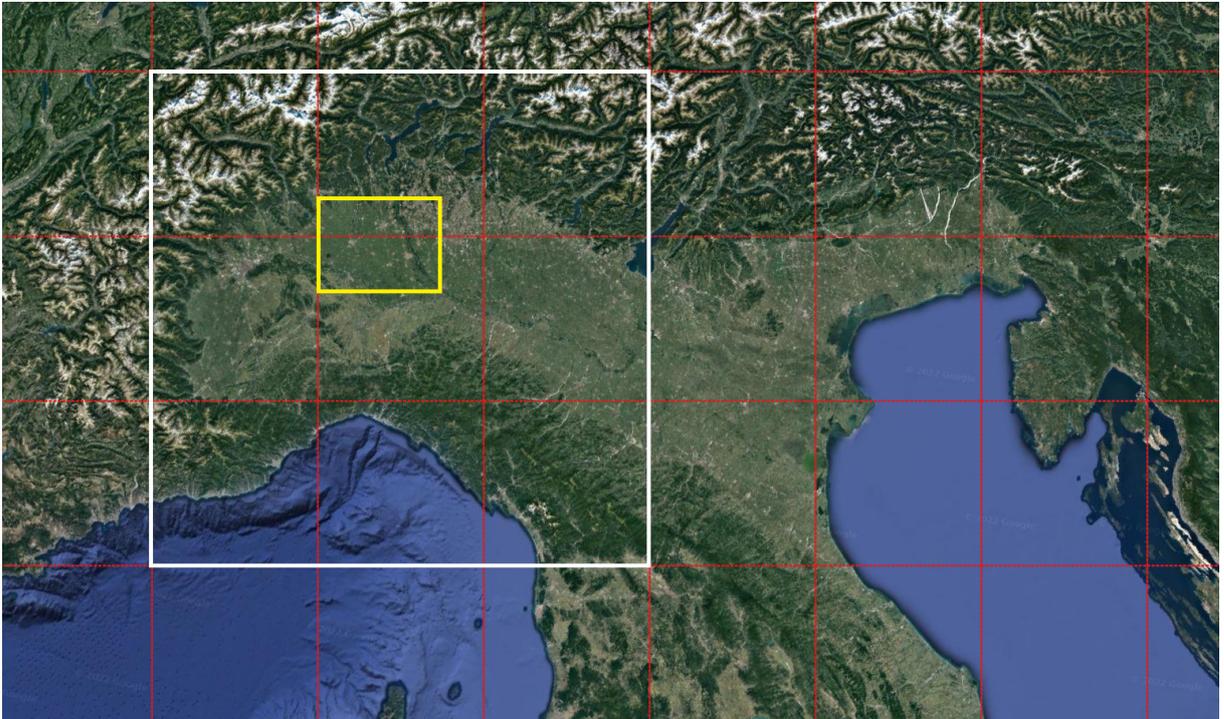


Figure 4.9: Comparison of the study area. The yellow rectangle refers to the area in the Nguyen et al. paper and the white rectangle refers to our study area.

4.3. Improving the Classification Approach by the Analysis of Local Extrema

plants within the regions of our study sites (by inspection at the sentinel hub playground). Local minima and their subsequent local maxima were grouped into pairs for further calculation. For each pair, we considered the backscatter difference between a local maxima

and the preceding local minima. The difference between the local extrema must exceed 6.5 dB to be considered as a rice area. The last classification criterion is satisfied by the threshold between the local minima and the local maxima. If these three decision-tree nodes were fulfilled by any pair of extremes within the time series, the object was classified as rice [6].

At the beginning of the main methodology, our decision tree consists of six decision nodes (annual minimum or global minima, maxima or global maxima, and the deviation between global minima and maxima). In the previous stages, we realized that these 6 decision nodes are suitable to distinguish seasonal vegetation, such as crops and deciduous forests, from the land covers with a constant low backscatter, such as water bodies and permanently irrigated areas, or high backscatter, such as urban areas [40] [64] [48].

Then we proceeded with the improvement phase by defining the local extrema parameters by applying the exact same functions not on the whole data cube but segmenting it into three subsets of 80 days each and performing the same procedures for each subset separately. In this way, we were able to determine six different local parameters for three segments. After a more detailed visual and quantitative analysis of each parameter, we decided to use the local maxima and local minima of the first two segments for our calculations, since we could recognize some distinct pattern in the case of rice plots, as the flooding and heading stages of rice plants can be clearly determined in smaller time periods with higher deviations. We calculated the local extrema of our time series within 80 windows, because the temporal interval was chosen based on the shortest possible growing season for rice plants within our study area according to the Italian rice growing calendar. The final classification criterion is met if either the local minima is less than -24.6 / -25 dB or the local maxima is between -20.1 and 16 dB / -19 and -15.4 dB in two consecutive windows. If these three decision tree nodes were satisfied by any extrema pair within the time series, the object was classified as rice because at least one crop in the year could be considered as rice.

5 | Discussion

5.1. Logic behind the algorithm

This classification method introduced by "Nguyen et al" uses the time series images of SAR to map rice growing areas based on their temporal variations, taking into account the specific and unique behaviour of rice plants. This algorithm focuses mainly on rice cultivation in Europe over 8 different regions located in 8 different countries, under the flooding conditions and the period when the flooding stops and, reciprocally, the rice plants start to grow until the harvesting season. This actively demonstrates that the annual variation of σ_0 of rice fields is greater than that of other vegetative land covers or agricultural crops. The backscatter coefficients of the cross-polarised data show a significant correlation with rice plant development. "Schmitt and Brisco, 2013" also found that the cross-polarised data has produced the best relationship with the rice age after planting [4]. This algorithm is compatible with our study area because of the rice cultivation technique exerted in northern Italy, which is direct seeding in flooded soil (water seeding). So, in this case, we have captured the seasonal variations of the backscatter values with the SAR imagery, which was very satisfactory during this study.

5.2. Rice Cropland Mapping Final Results

After applying the local extrema approach in addition to the previous steps and results, we mapped the rice distribution with an overall accuracy, precision and recall of 99%, 88% and 68.2% for the whole Po catchment in Lombardy and Piedmont region, table 5.1. For the rice class, the producer's accuracy is 88% and this is a quite satisfactory performance. For the non-rice land cover class, a value of 99% is calculated, which indicates how often the real land features are correctly represented on the classified map, or the probability that a certain land cover of an area (here rice and non-rice) is classified as such on the ground. The producer's accuracy is the complement of the omission error, so our omission error for the rice and non-rice is 12% and 1%, respectively. The user's accuracy of the study area for the rice and non-rice classes is 68.2% and 99%, respectively. In other

words, the user's accuracy essentially tells how often the class on the map will actually be presented on the ground and it is a complement to the commission error.

class	Rice	Non-rice	Total	Producer accuracy
Rice	2746323	372090	3118413	88%
Non-rice	1282560	220599027	221881587	99%
Total	4028883	220971117	225000000	
User accuracy	68.2%	99%		Kappa=0.78

Figure 5.1: Confusion matrix of accuracy assessment related to 2020 VH-backscatter data after applying the local extrema approach for further enhancement.

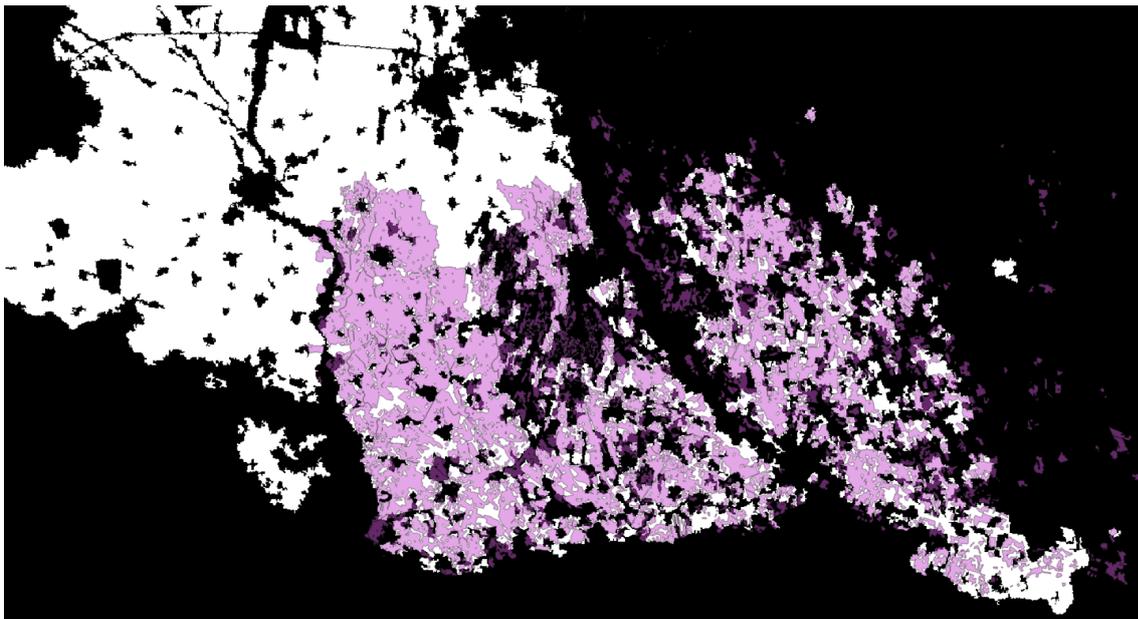


Figure 5.2: Comparing the CLC2018 map (white) to the local rice reference data (purple) we can see many differences between two maps that make CLC2018 less reliable.

The user's accuracy for the rice class is lower than we expected. By examining the whole procedures and the datasets used, we suspected that this low rice user's accuracy is related to our reference data CLC2018, as we demonstrate that it has a very low spatial resolution (100m) and has many inconsistencies with respect to the real world (Figure 5.2), by analysing and comparing it with the local classified agricultural data collected from the municipalities of each region. This fact leads us to obtain many false negative

pixels (false non-rice pixels), that contributes to the misclassified pixels as rice (Figure 5.3). Had we had more accurate reference data than CLC2018 with better resolution, the user's would have increased dramatically to approximately 90-80% (Figure 5.2).

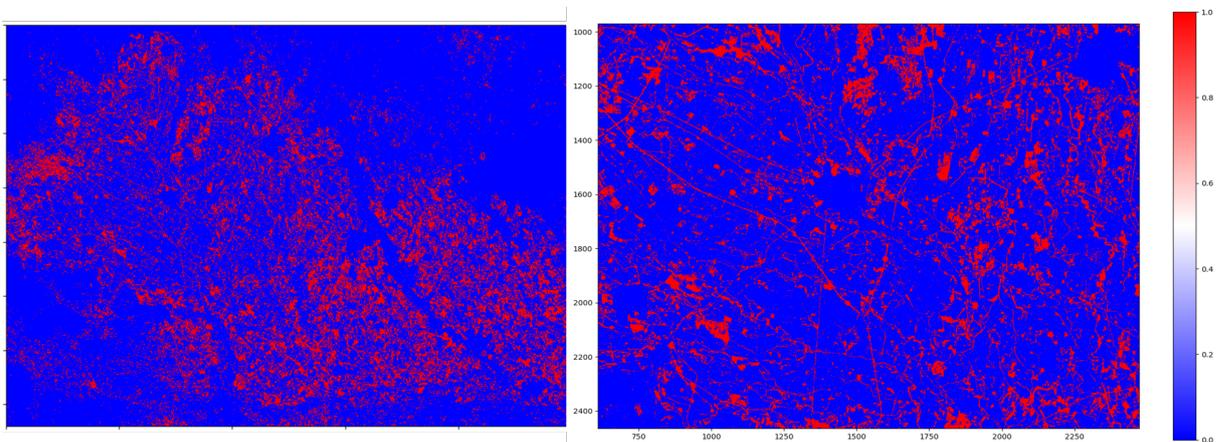


Figure 5.3: Map of false negatives. The blue areas are the truly classified regions, and the red spots are the misclassified regions.



Figure 5.4: Rice-paddy classification based on 2020 Sentinel-1 time series in Po catchment. The white areas are classified as rice, the black areas as non-rice.

The map of rice fields throughout the northern region of Italy from the Sentinel-1 time series is shown in a larger scale in Figure 5.4, and the detailed map in Figure 5.5. As it has been depicted in This figure, the details such as the channels between the rice paddies are misclassified. Our classifier classified these details as non-rice, which is in fact correct, but based on the CLC2018 map those areas must have been classified as rice, which contradicts our assumption.

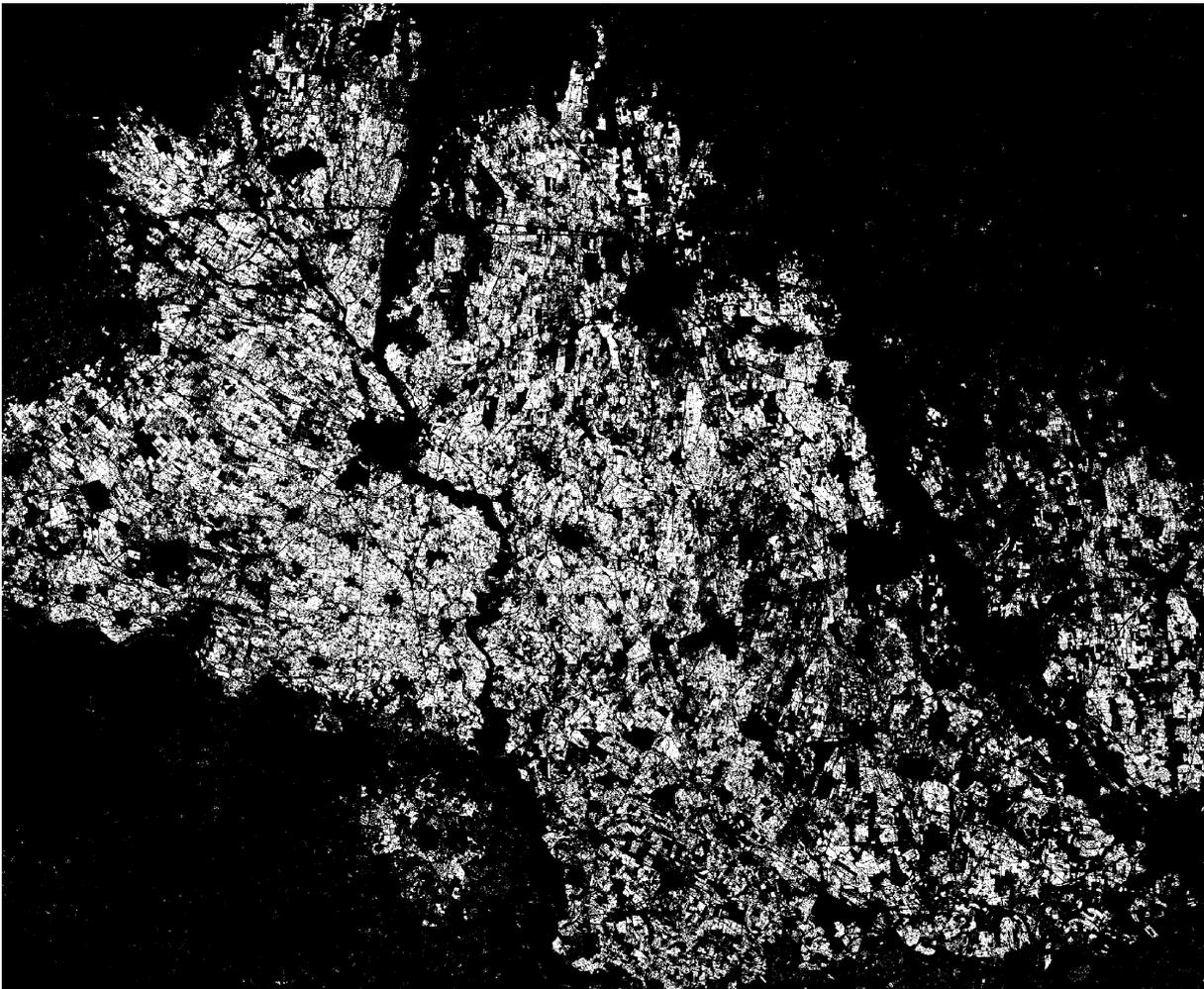


Figure 5.5: The classified rice fields in more details at a lower scale.

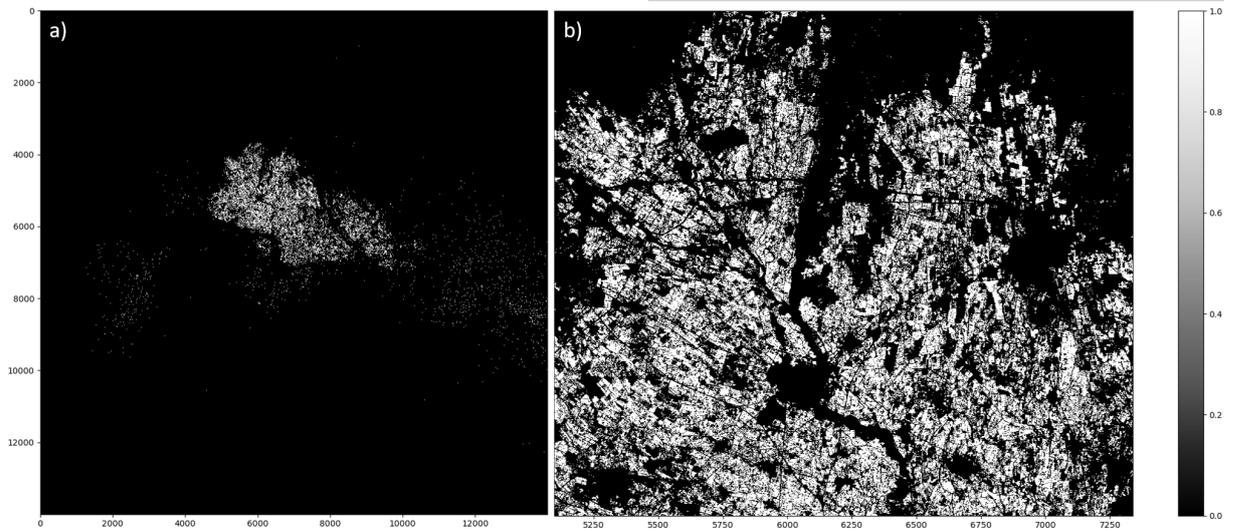


Figure 5.6: (a) rice-paddy classification based on 2019 Sentinel-1 time series in Po catchment. The white areas are classified as rice, the black areas as non-rice. (b) The rice fields in details at a lower scale.

By analysing and comparing the final maps, Figure 5.4 with the one we acquired previously, Figures 3.16 and 3.17, we can obviously see the misclassified pixels in the other agricultural fields have been removed and fixed. Our method achieved the highest classification accuracy after applying the local extrema approach on this specific study area. Conversely this also means that our method introduced some limitations applied on different areas with less prominent field boundaries but by adding the ultimate step at the end of the procedures we could succeed to some extent disregarding other mentioned limitations during this work.

5.3. Flaws, Failures and Limitations of this Decision-Rule-Based Algorithm

In this study, we aim to present an existing automatic method (Nguyen et al) for identifying rice growing areas by extracting statistical parameters after performing an orbit correction to VH-backscatter time series. The results are expected to show that the application of a decision-rule based classification approach using smoothed VH-backscatter time series and optimal incidence angle normalization can accurately classify rice and non-rice fields. This algorithm fails at classifying the rice and non-rice pixels from mountainous regions (alpine ranges) due to the complex conditions at higher elevations during

different seasons. Different agricultural activities during the growing season (rice varieties, water level in the fields, density of rice plants in the fields), SAR acquisition period can lead to this kind of misclassification. Also, the northern part of Italy is rich in a variety of agricultural fields, different vegetation and permanently irrigated lands, so misclassification may occur due to temporary or permanent irrigation. Another problem that can affect and interfere the collection of statistical parameters, thresholds, and rice maps is unexpected rainfall, which leads to misclassification because the non-rice pixels are considered rice due to the water supply caused by rainfall on the soil surface.

We used a conventional classification technique with manually optimized thresholds to classify the rice fields. The challenge of the threshold-based approach is to find appropriate thresholds for our dataset, and this was one of the difficulties we faced.

The Sentinel-1 data are affected by the incidence angle, whereas the strength of the SAR backscatter signal gradually decreasing as the incidence angle increases. For adjusting this problem, we used 2 different methods. The first one is to apply an orbit correction algorithm (Lievens et al, 2019 method) to our dataset to compensate the orbital effect, which showed better results compared to the second solution. The second method is to separate the dataset by its orbits and visualize the time series graphs separately, which was computationally intensive and time consuming.

We chose this method as the classification algorithm because it has high accuracy and is relatively easy to implement. This algorithm does not require a training dataset or in situ data collection, so it is not computationally intensive in itself. The only challenge we faced was dealing with a large amount of data, which resulted in long processing times due to the high spatial and temporal resolution and the size of the study area.

5.4. Polarization Comparison and Selection

In this study, different sets of data with two polarizations, VV and VH, were used and the results were compared to each other. Another goal of this work is to decide which polarisation (VH or VV or both at the same time, VV/VH) gives the best results and is suitable for rice detection and classification. The polarisation comparison was done based on the backscatter coefficients over time during the rice growing season over sample plots and by comparing the final results of the evaluation phase.

One of the comparison methods we used was to extract the preprocessed temporal backscatter coefficients of some random points for VV and VH separately and record them for

further analysis.

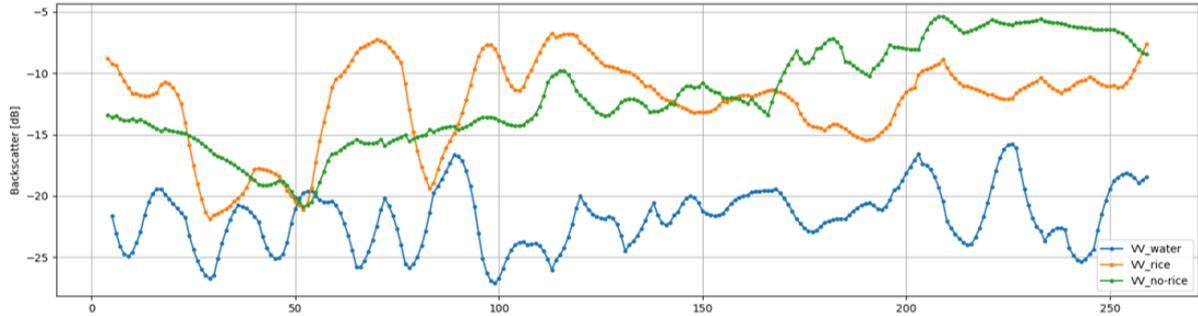


Figure 5.7: Sample plot of σ_{VV}^0 -smoothed for three random points in 3 different classes: water, rice and non-rice.

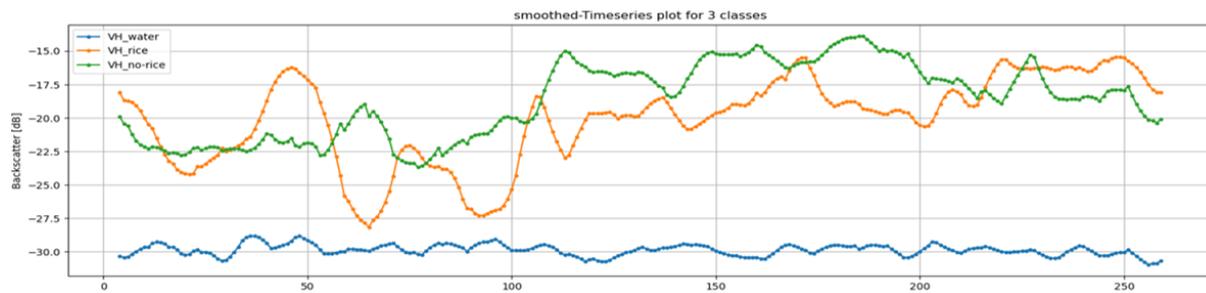


Figure 5.8: Sample plot of σ_{VH}^0 -smoothed for three random points in 3 different classes: water, rice and non-rice.

As shown in the graphs above, figure 5.7 is plotted with the temporal backscatter coefficients for three points in three different classes, it exhibits very high dynamics with many fluctuations that by analysing more points we were not able to capture a singular pattern unique to the Rice class. Instead, by analysing the graph of VH-backscatter, we recognize that each class we have defined exhibits specific behaviour. For example, in the case of water bodies, the blue graph line is almost the same throughout the year, but with small fluctuations around the same mean value, which is due to the different backscatter values caused by different incidence angles. In the orange graph line related to a rice pixel, we can simply show that the flooding of the rice field occurs in May and the second one in June, by the local minima in the curve.

After the preliminary results of the rice classification by applying the threshold optimization on VV and VH polarizations separately, the obtained confusion matrices, producer's

accuracy, user's accuracy, and Kappa were presented in the tables 4.5 and 4.4. Consequently, the producer accuracy for the rice class on VV-polarization was very low compared to what it was on VH-polarization. This means that by using the VV-backscatter coefficients for classification, we have introduced a large number of false-positives and a smaller number of true-positives than the expected number, which is very important for us. The Kappa value for VV-polarization was calculated to be 0.25, which is very low. On the other hand, the calculated Kappa value for VH-polarization was 0.6, which was accepted in this case to take further steps, such as the mentioned local extrema approach. In the bargain, according to "Schmitt and Brisco, 2013", the backscatter coefficients of the cross-polarized data have a significant correlation with the development of the rice plants and give the best relationship with the rice age after the transplanting stage [51].

Last but not least, due to the better rice classification abilities of cross-polarized (VH) measurements of SAR compared to the co-polarized (VV) data, a dense Sentinel-1 cross-polarized (VH) dataset was used for the rest of the calculations and for further considerations and proceeds, because it has more promising significance in discriminating rice from other land types.

5.5. Comparing 2 years time series (2019 & 2020)

By analysing the time series plot of a random rice pixel for both time series of 2019 and 2020 (Figure 5.9) we could depict that the backscatter values are approximately similar to each other during the year, and both plots align each other.

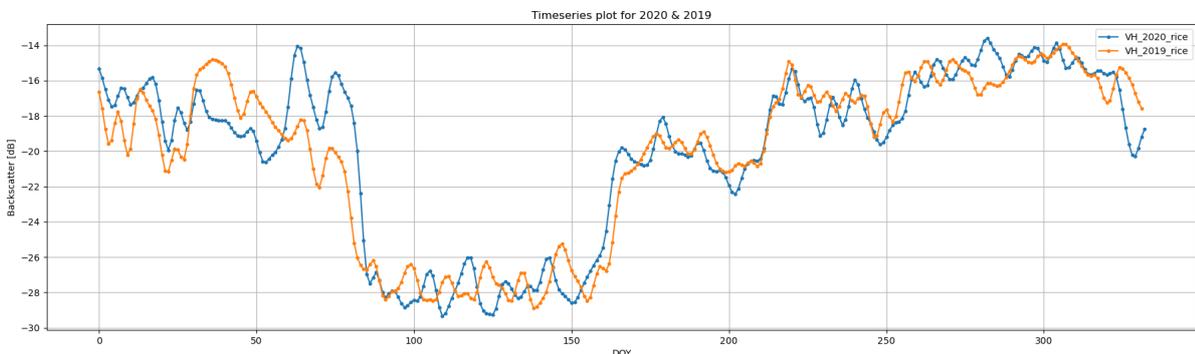


Figure 5.9: Time series plot of two consecutive years 2019 and 2020

The same procedures applied to the 2020 data were also applied to the 2019 time series, the only difference being the thresholds. During the preparation of the rice growing

calendar, we noticed that the rice cultivation stages were slightly different from those for 2020. This difference was negligible, and the thresholds we used after the optimization step only had decimal differences, that was something we could ignore because it did not dramatically affect the final results of the Kappa coefficient, 0.74 for the entire study area and 0.7 for just the rice region.

class	Rice	Non-rice	Total	Producer accuracy
Rice	4984397	2103762	7088159	71%
Non-rice	1638178	11993663	13631841	88%
Total	6622575	14097425	20720000	
User accuracy	75%	85.1%		Kappa=0.7

Figure 5.10: confusion matrix of derived rice region from $\sigma_{VH-smoothed}^0$ in a lower scale for 2019.

class	Rice	Non-rice	Total	Producer accuracy
Rice	4021270	697488	4718758	86%
Non-rice	2618830	217662412	220281242	99%
Total	6640100	218359900	225000000	
User accuracy	61%	99.7%		Kappa=0.74

Figure 5.11: confusion matrix of the entire study area from $\sigma_{VH-smoothed}^0$ for 2019.

When we look at the results of both time series and compare them, we can conclude that this unique algorithm is very well suited to map rice and that it can be used to differentiate rice from other agricultural crops. And ultimately, from the comparison of 2019 and 2020 tables 5.11 and 5.1, indicating that the parameters required for the decision tree can be robustly derived from one year of data, due to the similar retrieved accuracy.

$$\text{Kappa}_{2019} = 0.74$$

$$\text{Kappa}_{2020} = 0.78$$

6 | Conclusions and future developments

This study was mainly conducted based on remote sensing data, an archive of Sentinel-1A and B time series with a spatial resolution of 20 meters with VH and VV polarization in 2019 and 2020, without using an in situ dataset, to detect and map the distribution of rice cultivation areas in the Po catchment in the northern part of Italy at a large scale. We have demonstrated that rice-growing areas with a given agronomic inundation and seeding can be mapped by determining the statistical backscatter parameters using a decision rule-based technique. We can apply this method to different areas by adjusting and customizing the thresholds using the rice cropping calendar of each region and visually examining the dataset and parameters from an agronomic perspective. Defining the decision-tree nodes and their thresholds was one of the most important parts of the whole procedure. Reliable and accurate reference data with high resolution is essential to optimize the selected thresholds. In other words, the final optimized rice map and the chosen thresholds are highly correlated with the reference data, from which we do the temporal validation at each iteration. We proved that less accurate reference data can lead to misclassified pixels and reduce the classification accuracy by producing many false positives. To validate the performance of the decision trees and calculate the accuracy of the classifier, we used the overall accuracy, precision, F1-score, and kappa values. In each case, the calculated metrics are 84.06%, 97.98%, 52.33% and 0.78 respectively, for the entire study area in 9 tiles.

Another objective of this study was to find out which polarization is the most appropriate to distinguish rice paddies from other classes. In other words, which of the VV and VH polarizations show discriminative behavior in the case of rice? We proved that VH is the most suitable to classify rice due to its lower fluctuations and oscillations in the water bodies (VH shows a more robust effect on water despite VV, which shows sudden fluctuations and oscillations). We do not say that VV is not a suitable signature for classifying rice. Other studies have also used VV-polarization either for enhancement or classification, but with different methods and algorithms.

Our method not only proved the robustness of the algorithm proposed by "Nguyen et al". but also increased the accuracy of the final results by adding an additional procedure, using the local extrema approach, and defining new thresholds for the newly calculated statistical parameters based on the rice growing calendar. This step is particularly efficient in distinguishing different types of cropland from each other.

Despite the mentioned deficiencies and limitations, this study has shown that the proposed approach is capable of detecting and mapping the distribution of rice cropping areas using Sentinel-1 time series images with high resolution and the corresponding rice cropping calendar by obtaining feasible results at the end. Our results have shown that the SAR data cube with an appropriate temporal density over time plays a crucial role in this study for mapping rice fields or detecting and monitoring rice cultivation in the agricultural sectors or beyond.

In conclusion, it is noteworthy to mention that the recommendations proposed in this study could be considered in future studies to generalize and adapt the algorithm to achieve better accuracy and be applicable to larger continental scales. And collecting more information regarding the exact timing of irrigation, water supply, and soil preparation are also key factors for improving our rice growing calendar and, consequently, the classification results.

Bibliography

- [1] J. Aschbacher, A. Pongsrihadulchai, S. Karnchanasutham, C. Rodprom, D. R. Paudyal, and T. Le Toan. Assessment of ERS-1 SAR data for rice crop mapping and monitoring. 2002.
- [2] A. Bouvet, T. Le Toan, and N. Lam-Dao. Monitoring of the rice cropping system in the mekong delta using envisat/asar dual polarization data. *IEEE transactions on geoscience and remote sensing*, 47(2):517–526, 2009.
- [3] B. BRISCO, R. J. BROWN, B. SNIDER, G. J. SOFKO, J. A. KOEHLER, and A. G. WACKER. Tillage effects on the radar backscattering coefficient of grain stubble fields. *International Journal of Remote Sensing*, 12(11):2283–2298, 1991. doi: 10.1080/01431169108955258. URL <https://doi.org/10.1080/01431169108955258>.
- [4] B. Brisco, A. Schmitt, K. Murnaghan, S. Kaya, and A. Roth. Sar polarimetric change detection for flooded vegetation. *International Journal of Digital Earth*, 6(2):103–114, 2013. doi: 10.1080/17538947.2011.608813. URL <https://doi.org/10.1080/17538947.2011.608813>.
- [5] L. Busetto, S. Casteleyn, C. Granell, M. Pepe, M. Barbieri, M. Campos-Taberner, R. Casa, F. Collivignarelli, R. Confalonieri, A. Crema, et al. Downstream services for rice crop monitoring in europe: From regional to local scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12):5423–5441, 2017.
- [6] K. Clauss, M. Ottinger, and C. Kuenzer. Mapping rice areas with sentinel-1 time series and superpixel segmentation. *International Journal of Remote Sensing*, 39(5): 1399–1420, 2018. doi: 10.1080/01431161.2017.1404162. URL <https://doi.org/10.1080/01431161.2017.1404162>.
- [7] K. Clauss, M. Ottinger, and C. Künzer. Mapping rice areas with sentinel-1 time series and superpixel segmentation. *International journal of remote sensing*, 39(5): 1399–1420, 2018.
- [8] J. Dong, X. Xiao, M. A. Menarguez, G. Zhang, Y. Qin, D. Thau, C. Biradar, and B. Moore III. Mapping paddy rice planting area in northeastern asia with landsat

- 8 images, phenology-based algorithm and google earth engine. *Remote sensing of environment*, 185:142–154, 2016.
- [9] ersaf.lombardia. Ente regionale per i servizi all’agricoltura e alle foreste. URL <https://www.ersaf.lombardia.it/it/>.
- [10] E. S. A. (ESA). *ESA ASAR Product Handbook*. 2007. URL <https://earth.esa.int/eogateway/documents/20142/37627/ASAR-Product-Handbook.pdf>.
- [11] H. Fang, B. Wu, H. Liu, and X. Huang. Using noaa avhrr and landsat tm to estimate rice area year-by-year. *International Journal of Remote Sensing*, 19(3):521–525, 1998.
- [12] F. Filippini. Sentinel-1 grd preprocessing workflow. In *Multidisciplinary digital publishing institute proceedings*, volume 18, page 11, 2019.
- [13] Food and A. O. of the United Nations. *Food aid for food security 2011*. 2011. URL <https://www.fao.org/3/a0800e/a0800e.pdf>.
- [14] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson. Random forests for land cover classification. *Pattern recognition letters*, 27(4):294–300, 2006.
- [15] M. K. Gumma, P. S. Thenkabail, A. Maunahan, S. Islam, and A. Nelson. Mapping seasonal rice cropland extent and area in the high cropping intensity environment of bangladesh using modis 500m data for the year 2010. *ISPRS Journal of Photogrammetry and Remote Sensing*, 91:98–113, 2014. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2014.02.007>. URL <https://www.sciencedirect.com/science/article/pii/S0924271614000446>.
- [16] W. Guo, T. Fukatsu, and S. Ninomiya. Automated characterization of flowering dynamics in rice using field-acquired time-series rgb images. *Plant methods*, 11(1): 1–15, 2015.
- [17] W. Immerzeel, R. Quiroz, and S. De Jong. Understanding precipitation patterns and land use interaction in tibet using harmonic analysis of spot vgt-s10 ndvi time series. *International Journal of Remote Sensing*, 26(11):2281–2296, 2005.
- [18] Y. Inoue and E. Sakaiya. Relationship between x-band backscattering coefficients from high-resolution satellite sar and biophysical variables in paddy rice. *Remote Sensing Letters*, 4(3):288–295, 2013.
- [19] Y. Inoue, T. Kurosu, H. Maeno, S. Uratsuka, T. Kozu, K. Dabrowska-Zielinska, and J. Qi. Season-long daily measurements of multifrequency (ka, ku, x, c, and l) and

- full-polarization backscatter signatures over paddy rice field and their relationship with biological variables. *Remote Sensing of Environment*, 81(2-3):194–204, 2002.
- [20] T. Kurosu, M. Fujita, and K. Chiba. Monitoring of rice crop growth from space using the ers-1 c-band sar. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4): 1092–1096, 1995. doi: 10.1109/36.406698.
- [21] K. Lasko, K. P. Vadrevu, V. T. Tran, and C. Justice. Mapping double and single crop paddy rice with sentinel-1a at varying spatial scales and polarizations in hanoi, vietnam. *IEEE journal of selected topics in applied earth observations and remote sensing*, 11(2):498–512, 2018.
- [22] T. Le Toan, F. Ribbes, L.-F. Wang, N. Floury, K.-H. Ding, J. A. Kong, M. Fujita, and T. Kurosu. Rice crop mapping and monitoring using ers-1 data based on experiment and modeling results. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1): 41–56, 1997. doi: 10.1109/36.551933.
- [23] T. Le Toan, F. Ribbes, L.-F. Wang, N. Floury, K.-H. Ding, J. A. Kong, M. Fujita, and T. Kurosu. Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Trans. Geosci. Remote Sens.*, 35(1):41–56, 1997.
- [24] D. Lemp and B. Koch. Forest monitoring using terrasar-x data – evaluation of processing methods and first results. 11 2008.
- [25] H. Lievens, M. Demuzere, H.-P. Marshall, R. H. Reichle, L. Brucker, I. Brangers, P. de Rosnay, M. Dumont, M. Girotto, W. W. Immerzeel, T. Jonas, E. J. Kim, I. Koch, C. Marty, T. Saloranta, J. Schöber, and G. J. M. De Lannoy. Snow depth variability in the northern hemisphere mountains observed from space. *Nature Communications*, 10(1):4629, Oct 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12566-y. URL <https://doi.org/10.1038/s41467-019-12566-y>.
- [26] H. Lievens, I. Brangers, H.-P. Marshall, T. Jonas, M. Olefs, and G. De Lannoy. Sentinel-1 snow depth retrieval at sub-kilometer resolution over the european alps. *The Cryosphere*, 16(1):159–177, 2022. doi: 10.5194/tc-16-159-2022. URL <https://tc.copernicus.org/articles/16/159/2022/>.
- [27] J. M. Lopez-Sanchez, S. R. Cloude, and J. D. Ballester-Berman. Rice phenology monitoring by means of sar polarimetry at x-band. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7):2695–2709, 2011.
- [28] J. M. Lopez-Sanchez, F. Vicente-Guijalba, J. D. Ballester-Berman, and S. R. Cloude.

- Polarimetric response of rice fields at c-band: Analysis and phenology retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2977–2993, 2013.
- [29] L. R. Mansaray, W. Huang, D. Zhang, J. Huang, and J. Li. Mapping rice fields in urban shanghai, southeast china, using sentinel-1a and landsat 8 datasets. *Remote Sensing*, 9(3):257, 2017.
- [30] J. A. Martínez-Casasnovas, A. Martín-Montero, and M. A. Casterad. Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of landsat tm images. *European Journal of Agronomy*, 23(2):159–169, 2005.
- [31] H. McNairn and J. Shang. A review of multitemporal synthetic aperture radar (sar) for crop monitoring. *Multitemporal remote sensing*, pages 317–340, 2016.
- [32] H. McNairn, C. Champagne, J. Shang, D. Holmstrom, and G. Reichert. Integration of optical and synthetic aperture radar (sar) imagery for delivering operational annual crop inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5):434–449, 2009.
- [33] H. McNairn, J. Shang, X. Jiao, and C. Champagne. The contribution of alos palsar multipolarization and polarimetric data to crop classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12):3981–3992, 2009.
- [34] E. Meier, U. Frei, and D. Nüesch. Precise terrain corrected geocoded images. *SAR Geocoding: Data and Systems*, pages 173–185, 1993.
- [35] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, 115(5):1145–1161, 2011.
- [36] E. Ndikumana, D. Ho Tong Minh, N. Baghdadi, D. Courault, and L. Hossard. Deep recurrent neural network for agricultural classification using multitemporal sar sentinel-1 for camargue, france. *Remote Sensing*, 10(8):1217, 2018.
- [37] A. Nelson, T. Setiyono, A. B. Rala, E. D. Quicho, J. V. Raviz, P. J. Abonete, A. A. Maunahan, C. A. Garcia, H. Z. M. Bhatti, L. S. Villano, et al. Towards an operational sar-based rice monitoring system in asia: Examples from 13 demonstration sites across asia in the riice project. *Remote Sensing*, 6(11):10773–10812, 2014.
- [38] B. D. Nguyen. *The use of SAR backscatter time series for characterising rice phenology*. PhD thesis, Wien, 2018.
- [39] D. Nguyen, W. Wagner, V. Naeimi, and S. Cao. Rice-planted area extraction by

- time series analysis of envisat asar ws data using a phenology-based classification approach: A case study for red river delta, vietnam. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(7):77, 2015.
- [40] D. B. Nguyen and W. Wagner. European rice cropland mapping with sentinel-1 data: The mediterranean region case study. *Water*, 9(6):392, 2017.
- [41] D. B. Nguyen, A. Gruber, and W. Wagner. Mapping rice extent and cropping scheme in the mekong delta using sentinel-1a data. *Remote Sensing Letters*, 7(12):1209–1218, 2016. doi: 10.1080/2150704X.2016.1225172. URL <https://doi.org/10.1080/2150704X.2016.1225172>.
- [42] D. B. Nguyen, A. Gruber, and W. Wagner. Mapping rice extent and cropping scheme in the mekong delta using sentinel-1a data. *Remote Sensing Letters*, 7(12):1209–1218, 2016.
- [43] T. T. H. Nguyen, C. De Bie, A. Ali, E. Smaling, and T. H. Chu. Mapping the irrigated rice cropping patterns of the mekong delta, vietnam, through hyper-temporal spot ndvi image analysis. *International journal of remote sensing*, 33(2):415–434, 2012.
- [44] C. Norasma, M. A. Sari, M. Fadzilah, M. Ismail, M. Omar, B. Zulkarami, Y. Hassim, and Z. Tarmidi. Rice crop monitoring using multicopter uav and rgb digital camera at early stage of growth. In *IOP Conference Series: Earth and Environmental Science*, volume 169, page 012095. IOP Publishing, 2018.
- [45] G. of Twenty. *Ministerial Declaration: Action Plan on Food Price Volatility and Agriculture : Meeting of G20 Agriculture Ministers, Paris, 22 and 23 June 2011*. 2011. URL <http://www.g20.utoronto.ca/2011/2011-agriculture-plan-en.pdf>.
- [46] G. of Twenty. *G20, “Action Plan on Food Security and Sustainable Food Systems,” 2015*. 2015. URL <http://www.mofa.go.jp/files/000111212.pdf>.
- [47] A. O. Onojeghuo, G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred, and Y. Miao. Mapping paddy rice fields by applying machine learning algorithms to multi-temporal sentinel-1a and landsat data. *International journal of remote sensing*, 39(4):1042–1067, 2018.
- [48] M. Ottinger, C. Kuenzer, G. Liu, S. Wang, and S. Dech. Monitoring land cover dynamics in the yellow river delta from 1995 to 2010 based on landsat 5 tm. *Applied Geography*, 44:53–68, 2013. ISSN 0143-6228. doi: <https://doi.org/10.1016/j.apgeog.2013.07.003>. URL <https://www.sciencedirect.com/science/article/pii/S0143622813001641>.

- [49] C. Planque, R. Lucas, S. Punalekar, S. Chognard, C. Hurford, C. Owers, C. Horton, P. Guest, S. King, S. Williams, et al. National crop mapping using sentinel-1 time series: A knowledge-based descriptive algorithm. *Remote Sensing*, 13(5):846, 2021.
- [50] B. Salehi, B. Daneshfar, and A. M. Davidson. Accurate crop-type classification using multi-temporal optical and multi-polarization sar data in an object-based image analysis framework. *International Journal of Remote Sensing*, 38(14):4130–4155, 2017.
- [51] A. Schmitt and B. Brisco. Wetland monitoring using the curvelet-based change detection method on polarimetric sar imagery. *Water*, 5(3):1036–1051, 2013. ISSN 2073-4441. doi: 10.3390/w5031036. URL <https://www.mdpi.com/2073-4441/5/3/1036>.
- [52] A. Schubert, D. Small, N. Miranda, D. Geudtner, and E. Meier. Sentinel-1a product geolocation accuracy: Commissioning phase results. *Remote Sensing*, 7(7):9431–9449, 2015. ISSN 2072-4292. doi: 10.3390/rs70709431. URL <https://www.mdpi.com/2072-4292/7/7/9431>.
- [53] N.-T. Son, C.-F. Chen, C.-R. Chen, H.-N. Duc, and L.-Y. Chang. A phenology-based classification of time-series modis data for rice crop monitoring in mekong delta, vietnam. *Remote Sensing*, 6(1):135–156, 2014.
- [54] D. Stroppiana, M. Boschetti, R. Azar, M. Barbieri, F. Collivignarelli, L. Gatti, G. Fontanelli, L. Busetto, and F. Holecz. In-season early mapping of rice area and flooding dynamics from optical and sar satellite data. *European Journal of Remote Sensing*, 52(1):206–220, 2019. doi: 10.1080/22797254.2019.1581583. URL <https://doi.org/10.1080/22797254.2019.1581583>.
- [55] D. Stroppiana, M. Boschetti, R. Azar, M. Barbieri, F. Collivignarelli, L. Gatti, G. Fontanelli, L. Busetto, and F. Holecz. In-season early mapping of rice area and flooding dynamics from optical and sar satellite data. *European Journal of Remote Sensing*, 52(1):206–220, 2019.
- [56] F. T. Ulaby, A. Aslam, and M. C. Dobson. Effects of vegetation cover on the radar sensitivity to soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, (4):476–481, 1982.
- [57] T. Waheed, R. Bonnell, S. O. Prasher, and E. Paulet. Measuring performance in precision agriculture: Cart—a decision tree approach. *Agricultural water management*, 84(1-2):173–185, 2006.

- [58] C. Wang, J. Wu, Y. Zhang, G. Pan, J. Qi, and W. A. Salas. Characterizing l-band scattering of paddy rice in southeast china with radiative transfer model and multitemporal alos/palsar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4):988–998, 2009.
- [59] L.-F. Wang, J. A. Kong, K. H. Ding, T. Le Toan, F. Ribbes, and N. Floury. Electromagnetic scattering model for rice canopy based on monte carlo simulation. *Electromagn. Waves (Camb.)*, 52:153–171, 2005.
- [60] W. Xiao, S. Xu, and T. He. Mapping paddy rice with sentinel-1/2 and phenology, object-based algorithm—a implementation in hangjiahu plain in china using gee platform. *Remote Sensing*, 13(5):990, 2021.
- [61] W. Zhang, H. Liu, W. Wu, L. Zhan, and J. Wei. Mapping rice paddy based on machine learning with sentinel-2 multi-temporal data: Model comparison and transferability. *Remote Sensing*, 12(10):1620, 2020.
- [62] Y. Zhang, C. Wang, J. Wu, J. Qi, and W. A. Salas. Mapping paddy rice with multitemporal alos/palsar imagery in southeast china. *International journal of Remote sensing*, 30(23):6301–6315, 2009.
- [63] B. Zheng, S. W. Myint, P. S. Thenkabail, and R. M. Aggarwal. A support vector machine to identify irrigated crop types using time-series landsat ndvi data. *International Journal of Applied Earth Observation and Geoinformation*, 34:103–112, 2015.
- [64] Z. Zhu, C. E. Woodcock, J. Rogan, and J. Kellndorfer. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using landsat and sar data. *Remote Sensing of Environment*, 117:72–82, 2012. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2011.07.020>. URL <https://www.sciencedirect.com/science/article/pii/S0034425711002823>. Remote Sensing of Urban Environments.

List of Figures

1.1	Planted area of rice in Italy in 2018 according to istat website.	4
1.2	Rice growth phases (source: www.flickr.com). This scheme is the outline of different physiological phases that the rice grain goes through during its growing season. for the most part, the entire procedure takes 90 to 200 days or more, from sowing to harvest. However it might vary from time to time and region by region (depending on so many factors).	8
2.1	According to CLC2018 classification the red area is the class of rice fields in north part of Italy located in Lombardy and Piedmont regions (9 tiles of data).	21
2.2	Location of the study area. The inset shows the Po river and the agricultural lands around it in yellow	22
2.3	Planted area of rice in Italy in 2018 according to istat website	23
2.4	Total cultivated area for the production of rice in Italy from 2008 to 2020 according to Eurostat website (ec.europa.eu)	24
2.5	Temporal backscatter variation of rice during the growing season. This scheme facilitates the gasp of rice backscatter dynamics at each stage. During the flooding phase, the backscatter values of the rice pixels are very low due to the irrigation and presence of abundant water on the surface. Backscatter levels begin to increase as the rice harvest reaches its final stage and the water vanishes from the soil.	25
2.6	Derived rice cultivation calendar by obtaining the mean backscatter values for a rice classified pixel.	26
3.1	Methodology flowchart, the initial sketch.	31
3.2	$\sigma_{VH/VV}^0$ time series graph of a random point inside the ocean, before temporal-filtering and orbit-correction.	37
3.3	$\sigma_{VH/VV-smoothed}^0$ time series graph of a random point inside the ocean, after temporal-filtering and orbit-correction.	37

3.4	$\sigma_{VH/VV}^0$ time series graph for a random rice pixel, before temporal-filtering and orbit-correction.	38
3.5	$\sigma_{VH/VV-smoothed}^0$ graph time series for a random rice pixel, after temporal-filtering and orbit-correction.	38
3.6	Graph of $\sigma_{VH-A/D}^0$ time series of a random rice pixel for separated ascending and descending modes, before temporal-filtering and orbit-correction.	39
3.7	Graph of $\sigma_{VH-A/D-smoothed}^0$ time series of a random rice pixel for separated ascending and descending modes, before temporal-filtering and orbit-correction.	39
3.8	Graph of $\sigma_{VH-A/D}^0$ time series of a random rice pixel for separated orbits, before temporal-filtering.	40
3.9	Graph of $\sigma_{VH-A/D-smoothed}^0$ time series of a random rice pixel for separated orbits, before temporal-filtering.	40
3.10	σ_{MAX}^0 or Maximum backscatter value during the rice growth season.	42
3.11	σ_{MIN}^0 or Minimum backscatter value during the rice growth season.	43
3.12	σ_{MMD}^0 or the deviation map. All agricultural regions have high σ_{MMD}^0 than e.g. forest, water or urban areas.	44
3.13	The date when the backscatter coefficient reaches a maximum value is defined as the global maxima in $\sigma_{smoothed}^0$ time series, during the rice growing season. This date must come after the date of the start of the season, where it reaches its global minima.	45
3.14	During the growing season, the date of the beginning of season is defined as the global minima in $\sigma_{smoothed}^0$ time series.	46
3.15	The length of the season is defined as the number of days difference between DoM and DoS.	47
3.16	first generated rice map by defining the thresholds on the statistical parameters of rice.	50
3.17	Zoomed in rice map with the details.	51
4.1	$\sigma_{VH-preprocessed}^0$ and $\sigma_{VV-preprocessed}^0$ values of a single random pixel to be classified as rice before applying orbit correction and temporal filtering.	56
4.2	$\sigma_{VH-smoothed}^0$ and $\sigma_{VV-smoothed}^0$ values of the same pixel to be classified as rice after applying the entire preprocessing and processing phases. As you can see, the time series values are smoothed and do not contain keen fluctuations anymore.	56

4.3 Classification results of $\sigma_{VH-smoothed}^0$ after masking the mountains and limiting the area by cropping the dataset to a smaller region where contains rice fields rather than other crops for comparison purposes (Figure 3.17) . . . 58

4.4 Classification results of $\sigma_{VH-smoothed}^0$ of the entire region with 225M pixels over 9 tiles (Figure 2.1). 58

4.5 Classification results of $\sigma_{VV-smoothed}^0$ of the entire region. 58

4.6 The defined confusion matrix for our case study. 59

4.7 The confusion matrix of the classification obtained by "Nguyen & Wagner". 60

4.8 Rice classification map obtained by "Nguyen & Wagner". 60

4.9 Comparison of the study area. The yellow rectangle refers to the area in the Nguyen et al. paper and the white rectangle refers to our study area. . 61

5.1 Confusion matrix of accuracy assessment related to 2020 VH-backscatter data after applying the local extrema approach for further enhancement. . 64

5.2 Comparing the CLC2018 map (white) to the local rice reference data (purple) we can see many differences between two maps that make CLC2018 less reliable. 64

5.3 Map of false negatives. The blue areas are the truly classified regions, and the red spots are the misclassified regions. 65

5.4 Rice-paddy classification based on 2020 Sentinel-1 time series in Po catchment. The white areas are classified as rice, the black areas as non-rice. . . 65

5.5 The classified rice fields in more details at a lower scale. 66

5.6 (a) rice-paddy classification based on 2019 Sentinel-1 time series in Po catchment. The white areas are classified as rice, the black areas as non-rice. (b) The rice fields in details at a lower scale. 67

5.7 Sample plot of $\sigma_{VV-smoothed}^0$ for three random points in 3 different classes: water, rice and non-rice. 69

5.8 Sample plot of $\sigma_{VH-smoothed}^0$ for three random points in 3 different classes: water, rice and non-rice. 69

5.9 Time series plot of two consecutive years 2019 and 2020 70

5.10 confusion matrix of derived rice region from $\sigma_{VH-smoothed}^0$ in a lower scale for 2019. 71

5.11 confusion matrix of the entire study area from $\sigma_{VH-smoothed}^0$ for 2019. . . . 71

List of Tables

2.1	Properties of the Sentinel-1 time series data used in this research.	20
2.2	Software and tools used while completing entire thesis conduction.	27
3.1	Defined threshold values by "Nguyen et al, 2015, 2016, 2017, 2018". The units of the thresholds are mentioned in dB and DOY (Day Of Year). . . .	49
3.2	A rice pixel that complies with all the threshold-based conditions set out. .	51
3.3	Defined threshold optimization vales for $\sigma_{VH-smoothed}^0$ by Nguyen et al [40].	53
3.4	Defined threshold optimization vales for $\sigma_{VH-smoothed}^0$ backscatter coefficients.	53
3.5	Defined threshold optimization vales for $\sigma_{VV-smoothed}^0$ backscatter coefficient.	54

List of Symbols

Variable	Description
P_t	Transmitted Signal
P_r	Received Signal
G_t	Gain
A_r	Effective Aperture of the Receiving Antenna
A_s	Effective Area of the Incident Beam
K_a	Fraction Absorbed by Target with Gain G_t
R_t	Distance Between Radar Transmitter and Target
R_r	Distance Between Target and Radar Receiver
σ_0	The Scattering Coefficient, decibel (dB)
DN	Ratio of Normalised and Corrected Signal Intensity Coefficient of a Pixel
σ_{VH}^0	Stack of VH-Backscatter Time-Series
σ_{VV}^0	Stack of VV-Backscatter Backscatter Time-Series
$\sigma_{normalized}^0$	Stack of Preprocessed Backscatter Time-Series
$\sigma_{smoothed}^0$	Stack of Smoothed and Orbit-corrected Backscatter Time-Series
$\sigma_{VH-smoothed}^0$	Stack of Smoothed and Orbit-corrected VH-Backscatter Time-Series
$\sigma_{VV-smoothed}^0$	Stack of Smoothed and Orbit-corrected VV-Backscatter Time-Series
σ_{MAX}^0	Maximum Backscatter Value
σ_{MIN}^0	Minimum Backscatter Value
σ_{MMD}^0	Maximum Minimum Difference or Amplitude Backscatter
DoM	Date of Maximum Backscatter
DoS	Date of Minimum Backscatter
LoS	Length of Rice Growing Season
TP	True-Positive or True-Rice
TN	True-Negative or True Non-rice
FP	False-Positive or False Rice
FN	False-Negative or False Non-Rice

List of Abbreviations

abbreviation	Description
<i>GEOGLAM</i>	Group on Earth Observation Global agricultural Monitoring Initiative
<i>IRES</i>	Italian Rice Experiment Station
<i>SIAN</i>	Italian Informative Agricultural System
<i>EU</i>	European Union
<i>FAO</i>	Food and Agriculture Organization
<i>ERSAF</i>	Ente Regionale Per I Servizi All'agricoltura e alle Foreste
<i>ESA</i>	European space Agency
<i>MARS</i>	Monitoring Agriculture with Remote Sensing
<i>NDVI</i>	Normalized Difference Vegetation Index
<i>CRS</i>	Coordinate Reference System
<i>SWIR</i>	Short Wave Infrared
<i>EO</i>	Earth Observation
<i>SAR</i>	Synthetic Aperture Radar
<i>GRD</i>	Ground Resolved Distance
<i>S – 1A</i>	Sentinel-1 Ascending
<i>S – 1D</i>	Sentinel-1 Descending
<i>POD</i>	Copernicus Precise Determination
<i>CLC</i>	Corine Land Cover
<i>DOY</i>	Day Of Year
<i>DoM</i>	Day of Maximum backscatter
<i>DoS</i>	Day of Minimum backscatter
<i>LoS</i>	Length of the Season
<i>DEM</i>	Digital Elevation Model

Acknowledgements

I would like to express my profound gratitude to:

The Faculty of Geodesy and Geoinformation of TUwien for giving me the opportunity of conducting my entire master thesis in this department by providing everything I needed to conclude this amazing work. My special thanks goes to my dear supervisor Univ.Prof. Dipl.-Ing.Dr.techn.Wolfgang Wagner for his advices and my adviser dear Dipl.-Ing. Isabella Pfeil for all of her unique guidance, support and supervision with her valuable comments and input that helped me whilst working on the thesis during my start in Vienna. Also, my gratitude and recognition is with to the entire GEO Remote Sensing Research Group for the unique working environment and extraordinary teamwork, to all the meetings of the Department of Photogrammetry and Remote Sensing. Here I had the opportunity to work with interesting people from all over the world, who kindled my enthusiasm more about pursuing my studies within an international context and I could discover the importance of team spirit to give much more efficiency and precision to a task which is very precious for me.

In addition, Special thanks to my second supervisor Univ.Prof.Dr.Giovanna Venuti for her supervision and guidelines during my master studies in Politecnico di Milano university. I would like to thank to everyone that I have not mentioned individually for their valuable input and helpful instructions academically and mentally. The teamwork on conferences, workshops, meetings is greatly appreciated. Furthermore, thanks to all scientists whom I have not had the opportunity to meet in person but inspired and guided me on the right path with my research.

Last but not the least, my sincere thanks to my unconditionally beloved parents Laya Talati and Mahmoud Fakherifard and my sister Sadaf, for their continuing mental support and patience during this period. Thanks for their encouragement and believing in me, which helped me to pursue my dreams and studies till this point and for the future. And special thanks to my dear friends Parastuo Moradi, Hairaw Al-Barwari and Golnaz Memari for being there next to me during my hard time and unconditional supports.

