Università di Pisa

Satellite crop mapping to better understand agro-ecological zones

Supervisors:
Dr. Nicola Silvestri
Prof. Andrea Taramelli

Co-supervisor:
Dr. Sander Janssen

Candidate:
Laura Paladini

Summary

Introduction

For decades, agro-ecological zonations (AEZs) have been used for a wide range of purposes: from land use planning to the upscaling of crop models, from the estimation of future agricultural land to calculation of yield gaps. The increasing availability of climatic and soil global datasets has led to the creation of several AEZs, built on climatic and pedological parameters. At the same time, researchers have become more aware of the importance of land cover, either described in terms of general categories or with the detailed vegetation type, suggesting that it may be an important element in a zonation. Indeed, especially in the context of agro-ecosystems, the land cover and the agronomic practices, beside soil and climate features, can influence several processes, for instance nutrient cycles.

The case of N pollution deriving from agricultural activities is a perfect example. It is known that climatic variables, such as the amount of precipitations and their distribution across the season, can influence processes like leaching; the same is valid for soil properties, such as texture and porosity. In studying the N cycle or defining fertilizer recommendations, AEZs based on pedo-climatic parameters have often been used. Institutions in charge of defining Nitrate Vulnerable Zones for the Nitrate Directive use this system, as well. The quantity of nitrogen that enters the agro-ecosystem is mainly determined by the farmer, who uses different doses of fertilizer according to the crop cultivated and the customary agronomic practices. Therefore, it is hard to think of an agro-ecological zonation meant for the monitoring and the reduction of N pollution and missing crop type information.

In the process of building an agro-ecological zonation, the choice of the source of data is as important as the data itself. As an example, climatic conditions can change fast in time, so it is essential to rely on a source able to provide always up-to-date information. Satellite remote sensing qualifies as a relevant source of data, hence it has been used for the provision of climatic data for all the existing zonations. Land use and land cover are partly dependent on climate, thus they are expected to change more or less as rapidly. Given the fact that regular field campaigns would be too expensive, both for developed and developing countries, satellite remote sensing has shown great potential also for the contribution of this type of information. Research in the direction of agricultural applications, and more specifically towards crop type mapping, has also been encouraged by the recent development of satellite sensors with spatial resolutions of 10 to 30 meters.

The present research will explore the potential of satellite remote sensing for the provision of information relative to crops, as an auxiliary resource for developing agro-ecological zonations. The techniques used to produce the crop type map will be discussed, in the perspective of classifying at least the most common crops on the territory and distinguishing them from minor crops. Finally, the usefulness of integrating the crop type map into an AEZ will be discussed. The case study is represented by the Muzza area, located in the North of Italy, where N pollution is currently controlled by using agro-ecological zones based on climatic and pedological parameters.

Chapter 1

State of the art

1. State of the art

1.1 Agro-ecological zonations

1.1.1 Introduction to agro-ecological zonations

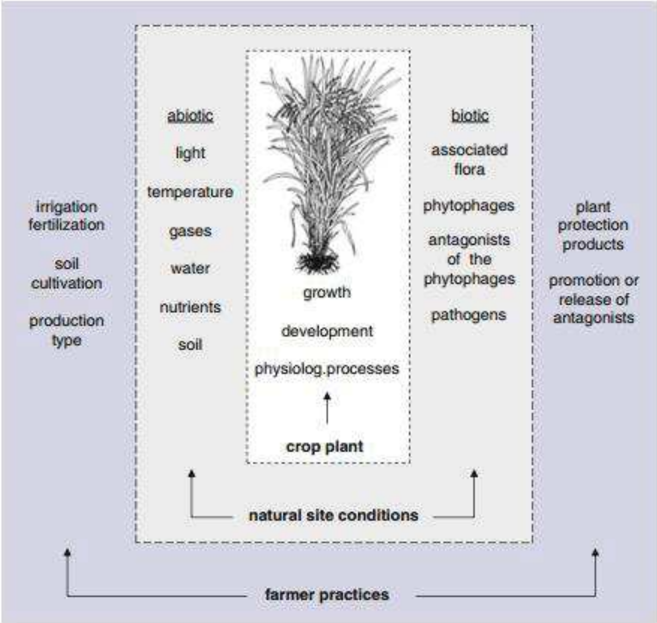
Agro-ecological zones (AEZs), as they were conceived by the FAO, were born for the purpose of evaluating land suitability for cultivation by taking into account at the same time climate and soil parameters, so as to describe the suitability of land for cultivation.

The first official document issued by FAO regarding AEZs was the *Report on the Agro-ecological Zones Projects*, published on the *[FAO] World Soil Resources Report*, 48, in 1978: here AEZs are precisely defined as «zones which have similar combinations of climate and soil characteristics, and similar physical potentials for agricultural production» (FAO, 1996). The same acronym (AEZ) is used to refer to the methodology used to divide an area into agro-ecological zones, defined as the «division of an area of land into smaller units, which have similar characteristics related to land suitability, potential production and environmental impact» (FAO, 1996).

AEZs were conceived as a tool for rural land-use planning and for land resource appraisal (FAO, 1996), but through time they have been used for other purposes, including integrated assessments (Fischer, Shah, Tubiello, & van Velhuizen, 2005) and up-scaling of crop models (Van Wart et al., 2013). Indeed, the term agro-ecological zoning has been associated with a wide range of different activities that are often related yet quite different in scope and objectives (Fischer, van Velthuizen, Shah, & Nachtergaele, 2002). This also implies that a single product can prove useful for a wide range of applications. For instance, the study area of this research could make use of an agro-ecological zonation to define the environmental impact of the agricultural sector.

1.1.2 The meaning of agro-ecological zonations in the context of agro-ecology

Agro-ecology has been defined as «the science of the relationships of organisms in an environment purposely transformed by man for crop or livestock production» (Martin & Sauerborn, 2013). The agroecosystem can be considered as «the environment of the crop» (Martin & Sauerborn, 2013). Since the crop (or the livestock) is the central object of the agroecosystem, it is easy to understand why the broadest concept of agroecosystem takes into account any factor influencing the production, be it biotic or abiotic, physical or human. (Martin & Sauerborn, 2013) use the scheme in Fig. 1 to represent the levels of factors influencing crop production.



Agro-ecological zonations are not meant to represent agroecosystems; however, the principle of growing levels illustrated above has been used to create agro-ecological zones as well. For instance, C. A. Mücher et al. (2003), when listing all the components that characterise an environment or a landscape, use a specific functional hierarchy (Fig. 2).

Fig. 1 – The agroecosystem as the environment of the crop (Martin & Sauerborn, 2013)

In addition to the biotic and abiotic factors, also human interference is considered, since it can affect components on the various hierarchical levels (e.g. geomorphology, soils and vegetation), thus shaping the environment and the landscape (C. A. Mücher et al., 2003).

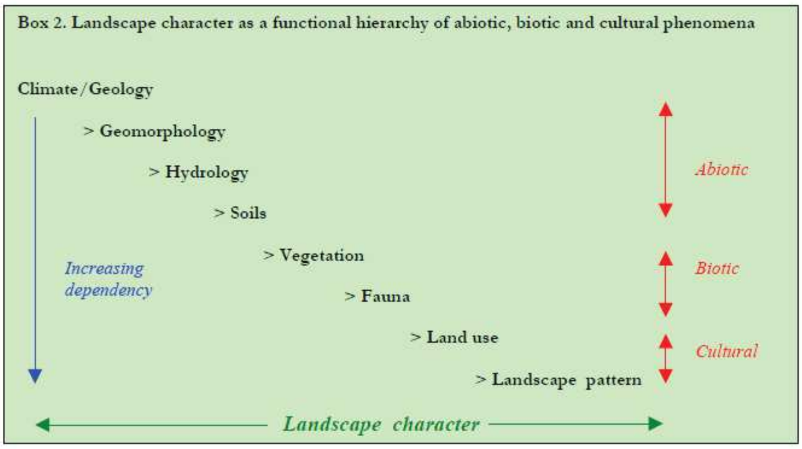


Fig. 2 – Landscape character as a functional hierarchy of abiotic, biotic and cultural phenomena (C. A. Mücher et al., 2003)

Keeping in mind this hierarchical list of components, the spatial limits of an agroecosystem and of an agro-ecological zone are somewhat arbitrary. These are determined by the scale and the number of factors that the researcher chooses. A complete analysis of the agricultural system requires the inclusion of all the parameters, from abiotic to cultural (Fischer et al., 2005); in other studies only abiotic factors are considered; finally, some cases take into account the abiotic and land cover factors. As we will see in the next paragraph, different zonations were created to respond to different purposes.

1.1.3 Existing agro-ecological zonations

In time, researchers have developed the concept of AEZ from FAO, creating zonations with different inputs, results and resolutions. van Wart et al. (2013) have provided a useful review of the climatic part of the models.

The most used agro-ecological zonations are the following (van Wart et al., 2013):

- GAEZ (Global Agro-Ecological Zones): by FAO\IIASA (Fischer et al., 2012);
- SAGE zonation scheme (Center for Sustainability and the Global Environment) (Licker et al., 2010);
- GYGA-ED (Global Yield Gap Atlas Extrapolation Domain) (Atlas, n.d.);
- GAES (Global Agro-Environmental Stratification) (Mücher et al., 2016);
- modifications of GAEZ or SAGE schemes: e.g. HCAEZ (HarvestChoice AEZ);
- GEnS (Global Environmental Stratification) (Metzger et al., 2012).

The HCAEZ and the GEnS are Climatic Zonations (CZs) rather than AEZs; however, they are included in this brief review because they have been used for research in agriculture (Zomer et al., 2014).

Following, in Tab. 1, a brief summary of the variables used in the different methods:

AEZ scheme	Climate variables[1]	Soil variables	Other variables	Spatial resolution	Notes
GAEZ v3.0	Wind run and wind speed; wet day frequency; sunshine duration; day-time and night-time temperatures; reference evapotranspiration; maximum evapotranspiration; actual evapotranspiration; snow balance calculation	Daily soil moisture balance; soil phases; soil drainage; soil texture; organic carbon content; soil acidity (pH); cation exchange capacity of clay; cation exchange capacity of soil; base saturation; total exchangeable bases; calcium carbonate; calcium sulphate; exchange sodium percentage; electrical conductivity	Crop-related: length of growing period; multiple cropping zones for rain-fed production; equivalent length of the growing period; net primary productivity (NPP) Other: land cover; land utilization type; protected area; administrative areas	10 km	It does not provide a stratification of data, but single layers of information. It is crop-specific.
SAGE	growing degree days (GDD; T_{mean}–crop-specific base temperature) and soil moisture	soil pH; soil organic carbon		10 km	Crop-specific

[1] partially from (van Ittersum et al., 2013)

	index				
GYGA-ED	daily max/min temperature, rainfall, humidity	texture; depth of rooting zone; slope	proportion of the harvested area; cropping intensity; some aspects of management (e.g. sowing date and cultivar maturity); water regime	10 km	
GAES	annual mean temperature, annual total precipitation, mean cloud fraction over the growing season, standard deviation of the cloud fraction over the growing season	DTM mean altitude, DTM mean slope	% of irrigated land, Gross Primary Production (GPP), decade when max biomass is reached in growing season 1 and 2, number of growing cycles, number of crop types, field size	1 km	It is an extended version of GYGA-ED.
HCAEZ	length of growing period, mean temperature, elevation			10 km	Only climatic
GEnS	GDD with base temperature of 0°C, an aridity index, evapotranspiration seasonality, temperature seasonality			1 km	Intended for environmental monitoring in general, not specifically for agriculture. Only climatic.

Tab. 1 – Main AEZ schemes and the variables on which they are based

The methods illustrated differ in several crucial points:

- variables used: they can be different with regard to the crop, the climate, the soil and other aspects investigated; they can be a direct (e.g. min\max daily temperature) or indirect (e.g. GDD) measure of the factor of interest;
- source of the variables used: the variables can derived from existing non-spatial datasets (e.g. global census data used in the SAGE), derived from other variables (e.g. potential number of crops derived from the LGP in the GAEZ, or the Aridity Index used in the GEnS) or remotely sensed. The latter is the case of most climate data (e.g. temperature and cloud cover), but in zonations like the GAES also soil and land cover variables are obtained from EO.
- crop specificity: crop-specific models are more accurate in predicting crop productivity, but they cannot be used for the estimation of crop production where rotations are applied; instead, non-crop specific models are less precise (they apply T=0°C as the base temperature in the calculation of GDD

for all crops), but they can be used also in case of crop rotations: this is a big advantage, since much of the world's cropland produces more than one major food crop;

- terrestrial area considered: while some methods make calculations on all the arable land, others take into account only the actual harvested area of major food crops;
- spatial resolution: SAGE, GYGA-ED, GAEZ v3.0 use 10 km grid cells; the highest resolution data is provided by GAES (1kmx1km);
- zone size: this parameter is of extreme importance, since it influences directly the variability within the zones. It is essential to optimise the trade-off between achieving climatic homogeneity within zones and minimising the number of zones necessary to capture large portions of harvested area of target crop (van Wart et al., 2013).This problem is specifically addressed by GAES, which was built by performing a multi-resolution segmentation and which presents 4 hierarchical spatial levels.

1.1.4 Strengths and weaknesses of existing AEZs

Agro-ecological zonations can be used in the framework of different applications; therefore, the evaluation of the best scheme to adopt should be done according to the final objective.

The main characteristics to be considered are:

- variables used: several studies have remarked the importance of having a complete set of information about soil and terrain conditions (D. H. White et al., 2001); moreover, as explained earlier, it would be advisable to include in the zonation not only biophysical variables, but also information on land cover and on the socio-economic context. When the inclusion of this information does not match the final objective of the zonations, it is still advisable to use it as a descriptor.
- crop specificity: the applications of crop-specific zonations are limited to studies focused on one single crop, excluding other types of research (e.g. on yield gaps, yield predictions, integrated assessments); on the other hand, non-crop-specific zonations can be used for a wider range of purposes.
- terrestrial area considered: studies intended for predictions need to be based on the total arable land, while researches meant to assess the state of the art require information on the amount of land which is actually cropped; an agro-ecological zonation including both types of information could be used for both purposes;
- spatial resolution: the quality of the spatial resolution depends on the object of the study and on its physical dimension. A spatial resolution of 1 km or 10 km, such as the ones of the AEZs reviewed, is suitable for studies at the national, continental and global scale; however, a higher degree of spatial resolution is needed for monitoring at the sub-national, local scale.

- zone size: as for spatial resolution, the optimal zone size depends on the scale; again, the zone size of all the zonations reviewed may be too coarse for studies at the national and sub-national levels, since they may not capture the heterogeneity of the studied area; indeed, zone size and within-zone variability are connected (Van Wart et al., 2013). For instance, (van Wart, Kersebaum, Peng, Milner, & Cassman, 2013) in their study conclude that GAZE and HCAEZ are built on zones with a high within-zone heterogeneity, making it impossible to estimate accurately yield gaps. Only the multi-resolution system of the GAES deals specifically with this issue, letting the user choose the most suitable scale, even though the highest resolution offered may not be enough for local studies.
- Presence of land use, land cover or crop type information: when studying phenomena that depend or influence the land cover or the crop type, making this information available within the zonation is very valuable. When referring to nutrient cycles, many works can be found that study N dynamics by focusing on sample fields in each zone (Masvaya et al., 2010, Kaizzi, Ssali, & Vlek, 2006). They couple territory-level data with field-level data, study the chosen phenomenon at the field-level and then generalize the results and the recommendations. In this way, they miss a global view over the territory considered, be it a region or a country, which is particularly important when dealing with processes that occur at the regional scale. To contextualize results and recommendations for farm management, other studies employ statistical information about crop types and\or cropping systems and integrate it into agro-ecological zones. The information can be used in two ways:
  - As an input variable: for instance, (van Beek et al., 2016) use zones in Ethiopia (*woredas*) that are defined on the basis of agro-ecological information, soil type and farming system type. The zones are used to determine the inputs and outputs of nutrient balances that obviously can change considerably according to the farming system.
  - As a descriptor layer: the GAES zonation system is based on 13 input variables; the number of crop types is used as an input variable, but the dominant crop type is given as a descriptor of the stratification. This means that the dominant crop type (land cover) does not influence the clustering of the zones, but simply contributes information about the zones that were built using other variables.

  Among the global AEZs reviewed, only the GAES takes into account land cover information. Hence, the other zonations may be considered only partially useful when studying regional scale-processes.

In conclusion, the pros and cons of each zonation scheme should be evaluated with regard to the specific application. When referring to the monitoring of environmental pollution derived from agricultural activities, wind speed may be less significant than the amount and temporal distribution of precipitations; the depth of the rooting zone may contribute less information than the soil texture; while farm management information should be well-described, with variables like the number and type of crops or the use of irrigation, to be surveyed regularly.

1.2 Remote sensing

1.2.1 Basics of remote sensing

The term *remote sensing* has first been used in the 1960s to describe any means of observing the Earth from afar; it was initially referred to the acquisition of aerial photography, but with time it has been associated to the complete processing chain of remotely sensed products, from image acquisition to the dissemination of the final products (Chuvieco & Huete, 2010). In general, remote sensing can be defined as the acquisition of information about the state and condition of an object through sensors that are not in physical contact with it; the information is transmitted from the object to the sensors in the form of electromagnetic radiation (Chuvieco & Huete, 2010).

The sensor is the tool that acquires remote sensing images. According to the nature of the electromagnetic radiation recorded, sensors can be divided into passive and active. The former record the electromagnetic radiation naturally emitted or reflected from the area of interest. The latter, instead, emit electromagnetic radiation and then record how much of it is reflected back. Many active sensors onboard satellites exist (ESA's Sentinel1 is the most recent one), but the major part of space-based sensors is passive.

Sensor quality is mainly described by its resolution. Resolution is defined as the sensor's ability to discriminate information (Estes & Simonett, 1975), which is also the ability of the sensor to distinguish a specific object from other objects (Chuvieco & Huete, 2010). The discrimination of information is determined by the spatial detail, the number of spectral wavebands and their bandwidth, the spectral range covered, the temporal frequency of observation. Therefore, it is possible to distinguish the following types of resolution:

- Spatial resolution: it is a measure of the fineness of detail of an image (Khorram, Koch, van der Wiele, & Nelson, 2012). The definition of spatial resolution is linked to the Instantaneous Field Of View (IFOV), which is the angular section observed by the sensor, in radians, at a given moment in time; the spatial resolution is commonly defined as the size of the projected IFOV on the ground. The unit used to express spatial resolution is usually the pixel, which is the size of the minimum spatial unit of the image. Sensors with global coverage can have different spatial resolutions: from 0.65 m (IRS's CasrtoSat-2C) to 10m (ESA'S Sentinel-2), from 250 m (e.g. NASDA's ADEOS II) to 1 km (e.g. MODIS). Spatial resolution is one of the most important sensor characteristics, since it affects the level of detail achieved and the accuracy of the analysis: indeed, the smaller the size of the pixel, the smaller the probability that the pixel will be a mix of multiple objects ("mixed pixel"), not easily distinguishable or classifiable (Chuvieco & Huete, 2010). However, the choice of an adequate spatial resolution depends mainly from the scale that has been chosen to study the given problem. There is not a unique definition of scale: (Levin, 1992) calls it the "window of perception", the measuring tool through which a landscape (or any other object) may be viewed or perceived; (Wu & Li, 2009) define the observation scale in remote sensing as the measurement unit at which the data is measured or sampled, thus directly linking it to the concept of spatial resolution. Since different processes occur at different modeling or operational scales (Wu & Li, 2009), different scales reveal different patterns of reality (Marceau & Geoffrey, 1999). Typically, the larger the area covered, the smaller the level of detail (Chuvieco & Huete, 2010); which means, the larger the scale, the lower the spatial resolution. Another concept related to the one of spatial resolution yet different is the Minimum Mapping Unit (MMU), which is the smallest unit of information that can be included in a thematic map (Chuvieco & Huete, 2010).

- Spectral resolution: it is described by the spectral range of the sensor, the number of spectral bands and their bandwidth. It defines the sensor's ability to detect wavelength differences between objects or areas of interest (Khorram et al., 2012). Considering spectral resolution, sensors can be divided into: panchromatic, with one broad spectral band; multispectral, with multiple medium-width spectral bands; and hyperspectral, with many narrow spectral bands, up to hundreds. For instance, the OLI sensor, belonging to Landsat8, has got 9 bands, with bandwidths from 0.02 to 0.18 µm; thus, it classifies as a multispectral sensor. The Hyperion sensor, belonging to EO-1, has got 242 bands, with bandwidths around 0.01 µm.
- Radiometric resolution: it is the sensitivity of the sensor, i.e. its capacity to discriminate small variations in the recorded spectral radiance. In optical-electronic sensors, it refers specifically to the range of values coded by the sensor; most sensors code in 8 bits (256 digital levels per pixel) (Chuvieco & Huete, 2010).
- Temporal resolution: it is the observation frequency, or revisiting period, of the sensor (Chuvieco & Huete, 2010). The temporal resolution of the sensors usually change with their objective, with metereological satellites having short revisiting periods (even 15-30 minutes) and satellites for other applications longer ones (e.g. 16 days for Landsat8, 5 days for Sentinel-2, 1-2 days for MODIS).

It should be remarked that there are not absolute quality standards for sensors, but rather that all the resolutions should be considered according to one's final objective (Chuvieco & Huete, 2010). For instance, fire detection may require higher temporal resolution and smaller spatial resolution than soil mapping. In the light of the specificity of each objective, different countries have designed missions with special focuses: climate, water, vegetation and so on. Among the sensors destined to high-resolution vegetation monitoring, it is important to mention the family of satellites called Landsat, managed by NASA (US National Aeronautics and Space Administration), and the family of the Sentinels, managed by ESA (European Space Agency).

In 1972, the U.S. Geological Survey (USGS) and NASA launched the first satellite of the Landsat project, which is one part of the bigger USGS Land Remote Sensing (LRS) program. The main recipients of Landsat products are people working in agriculture, geology, forestry, regional planning, education, mapping and global change research (USGS, 2016b). The last satellite to be launched in the framework of the mission was Landsat8, which is still operative. The mission objective is to provide images of all landmass and near-coastal areas on the Earth (USGS, 2016a). The satellite carries two passive, multispectral sensors: the Operational Land Manager (OLI) and the TIRS (Thermal Infrared SensorS). The OLI consists of 9 spectral bands, of which bands 2 to 8 are the most useful for vegetation studies; the TIRS present 2 spectral bands, which provide information about surface temperature (Tab. 2) (USGS, 2016c).

	Band	Wavelength (micrometers)	Resolution (meters)
OLI	Band 1 – Ultra Blue (Coastal\Aerosol)	0.43-0.45	30
	Band 2 – Blue	0.45 - 0.51	30
	Band 3 – Green	0.53 - 0.59	30
	Band 4 – Red	0.64 - 0.67	30
	Band 5 – Near Infrared	0.85 - 0.88	30
	Band 6 – Shortwave Infrared (SWIR) 1	1.57 - 1.65	30
	Band 7 – Shortwave Infrared (SWIR) 2	2.11 - 2.29	30
	Band 8 – Panchromatic	0.50 - 0.68	15
	Band 9 – Cirrus	1.36 - 1.38	30
TIRS	Band 10 – Thermal Infrared (TIR) 1	10.60 - 11.19	100 * (30)
	Band 11 – Thermal Infrared (TIR) 2	11.50 - 12.51	100 * (30)

Tab. 2 – Landsat8 spectral bands

Since the mission has operated continuously for more than 40 years, the Landsat archive represents the longest collection of space-based moderate-resolution land remote sensing data.

Following the latest advancements in remote sensing, ESA has scheduled the launch of 6 new satellites, called Sentinels. The objective of the new Earth observation programme is to provide information to improve the management of the environment, understand and mitigate the effect of climate change and ensure civil security (ESA, 2000). The first three satellites are already operative. Sentinel-2, launched in June 2015, has been designed to work in continuity with Landsat and SPOT for some key land services (namely land monitoring, emergency management, security and climate change) (ESA, 2016b). The satellite has got one passive, multispectral sensor, called MSI, which acquires information over 13 bands. The Sentinel-2 mission was specifically design to monitor changes in the vegetation, thus the MSI has got spectral, spatial and temporal resolutions with a higher potential for vegetation monitoring in comparison to Landsat8: indeed, there are 3 bands (B5, B6, B7) in the "red edge" region (ESA, 2016a), which is notably sensitive to plant abiotic stress (Chuvieco & Huete, 2010). Moreover, the bandwidth of most of the bands of interest for vegetation monitoring is narrower (Fig. 3), thus identifying specific absorption features of vegetation; the revisit frequency of 5 days should ensure a reasonable percentage of cloud-free imagery (Whitcraft, Vermote, Becker-Reshef, & Justice, 2015); the spatial resolution of 10-20 m should be able to capture with a higher degree of detail the diverse farming systems found globally.

Band	Central wavelength (micrometers)	Bandwidth (micrometers)	Resolution (meters)
B1	0.443	0.020	60
B2	0.490	0.065	10
B3	0.560	0.035	10
B4	0.665	0.030	10
B5	0.705	0.015	20
B6	0.740	0.015	20
B7	0.783	0.020	20
B8	0.842	1.150	10
B8a	0.865	0.020	20
B9	0.940	0.020	60
B10	1.375	0.030	60
B11	1.610	0.090	20
B12	2.190	1.800	20

Tab. 3 – Sentinel-2 spectral bands

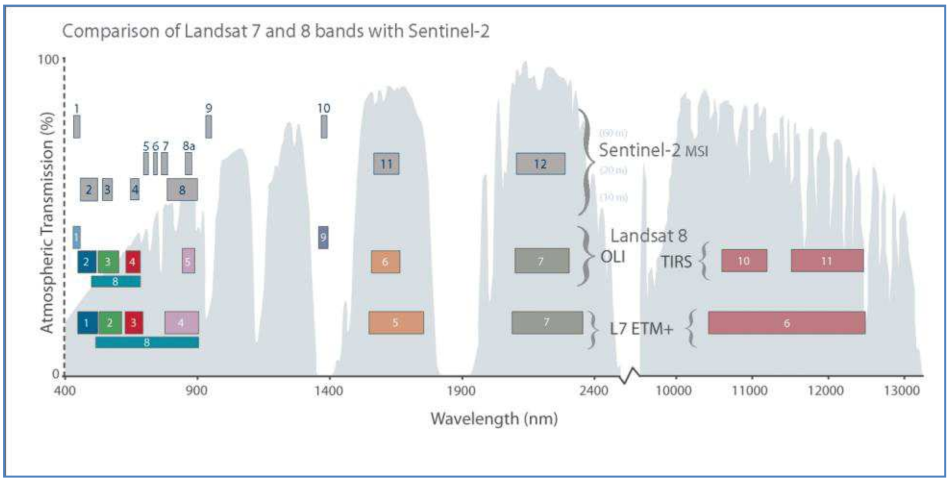


Fig. 3 – Comparison of Landsat 7 and 8 bands with Sentinel-2 (NASA, 2016)

The unprecedented combination of spectral, spatial and temporal resolution offered by Sentinel-2 represents a major step forward compared to the previous multi-spectral missions (Drusch et al., 2012); this will hopefully stimulate the agribusiness sector (Cuca & Tramutoli, 2014), also thanks to its user-oriented approach (Bontemps et al., 2015), while boosting research in remote sensing for agriculture. However, studies regarding the pre-Copernicus period will keep on relying on Landsat-8 data.

1.2.2 The application of remote sensing technologies for the characterization of agricultural systems

It has been shown that agro-ecological zonations should also account for the characteristics of cropland and, where possible, of farm management. However, it is important to remark that monitoring and mapping agricultural systems require clearly defined concepts and objects (Begué et al, 2015). The object of any remotely sensed map can be the cropland, a cropping system or a farming system; these terms have been widely used in literature with different meanings, therefore we will here provide a clear definition, following the extensive review made by (Cochet, 2012) (Fig. 4) : a) cropland: the crop covering a given piece of land at a given time (e.g. maize, wheat, soy); b) cropping system: the agricultural practices and techniques used in a plot or group of plots, in terms of crops cultivated, crop associations, crop successions, level of intensification, according to a specific sequence and pedo-climatic conditions; c) farming system: a group of farms with the same physical resources and technology, in the same socio-economic context (Cochet, 2012). Bégué, Arvor, Lelong, Vintrou, & Simoes (2015) then take the next step by asking which "land maps" to monitor which "agricultural systems"? The authors answer this question by introducing the following concepts: 1) Land cover: it addresses the description of the land surface in terms of soils and vegetation layers, including natural vegetation, crops and human structures (Burley, 1961). The feature describing the land cover is cropland. 2) Land use: it refers to the purpose for which humans exploit the land cover, including land management techniques (Verburg, van de Steeg, Veldkamp, & Willemen, 2009). The described feature is the cropping system. 3) Land use systems: they can be defined as a coupled human-environment system; they describe how land, as an essential resource, is being used and managed (Bégué et al., 2015). They are described by the correspondent farming system.

This distinction must be kept in mind when dealing with remote sensing of agricultural land: mapping the type of crop on the field and irrigated\non-irrigated crops not only respond to different objectives but require also different methods. This work will only refer to crop type mapping.

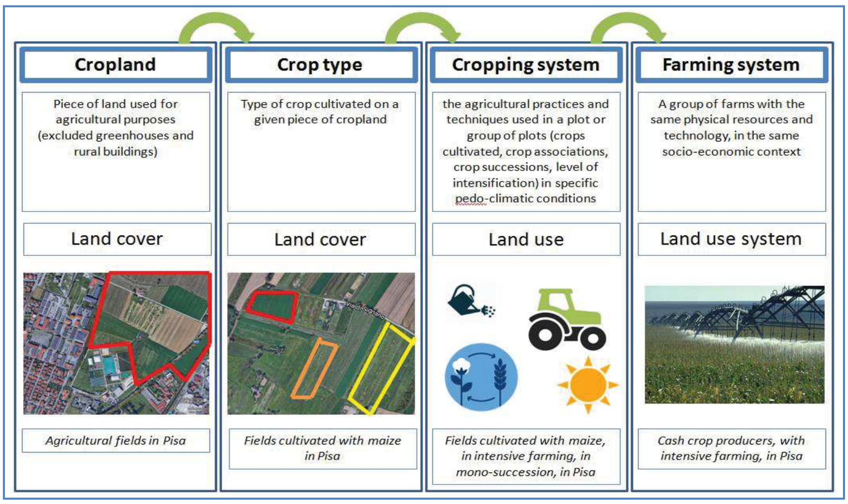


Fig. 4 – The levels of an agricultural system

1.2.3 Crop mapping methods

A wide range of methods has been tested to address the issue of crop type mapping. Crop mapping techniques may be classified on the basis of different criteria, among which time of delivery, classification metrics, approach. Tab. 4 shows some crop mapping methods; it does not represent a complete review of the existing methodologies, but it includes the most common ones.

Criterion	Method		Advantages	Disadvantages	References
Time of delivery	In season		Provides early information.	Lower accuracy than end-of-season products.	(Villa, Stroppiana, Fontanelli, Azar, & Brivio, 2015) (Azar, Villa, Stroppiana, & Crema, 2016)
	End of season		Higher accuracy than in-season products.	Cannot deliver some of the information required by policy-makers.	(Jordi Inglada et al., 2015)
Radiometry	Optical		Easy to understand and use.	Does not work at all with cloud cover.	(Jordi Inglada et al., 2015) (Valero et al., 2016) (Immitzer, Vuolo, & Atzberger, 2016)
	SAR		Can provide information also when there is cloud cover.	More difficult to understand and use.	(Tso & Mather, 1999)
	Optical+SAR		Can provide information also when there is cloud cover.	More difficult to understand and to use.	(J Inglada, Vincent, Arias, & Marais-Sicre, 2016) (Villa et al., 2015)
Spectral resolution	Multispectral		Easier and faster to process.	May not capture all the spectral features needed for specific studies.	(Azar et al., 2016) (Jordi Inglada et al., 2015)
	Hyperspectral		Contains more detailed information about specific absorbance peaks.	Features overload. Long computational time.	(Liu & Bo, 2015)
Classification metrics	Spectral features	Vegetation Indices (VIs)	Easy to compute. Allow feature reduction. Easy to interpret.	May not be useful when studying many crops at the same time. Not directly linked to biophysical variables.	(Azar et al., 2016)
		SAM	High accuracy	Does not perform well when there are mixed pixels.	(Dennison, Roberts, & Peterson, 2007)
		SMA	Can estimate the	Difficult to	(Bannari, Pacheco,

		vegetation fraction within mixed pixels. Can help in sub-pixel analysis, especially needed when dealing with low-resolution data.	transform into thematic information. There is a limitation to the number of endmembers that can be used. Intra- and inter-species variability are difficult to manage.	Staenz, McNairn, & Omari, 2006) (Roth, Dennison, & Roberts, 2012)
	Textural features	Can capture a different type of information.	Usually not enough to classify a high number of crops.	(Balaguer, Ruiz, Hermosilla, & Recio, 2010) (J Inglada et al., 2016)
	Spectral and textural features	Increased classification accuracy.	Possible features overload.	(Murray, Lucieer, & Williams, 2010) (Rodriguez-Galiano, Chica-Olmo, Abarca-Hernandez, Atkinson, & Jeganathan, 2012) (Reis & Taşdemir, 2011) (Peña-Barragán, Ngugi, Plant, & Six, 2011)
Approach (1)	**Pixel-based**	Works well when the size of the studied object equals the pixel size.	Salt-and-pepper effect.	(Valero et al., 2016)
	Object-based	Works well when the size of the studied object is bigger than then pixel size.	Does not work well when the area studied is small (e.g. smallholder agriculture) or heterogeneous (e.g. intercropping).	(B. Schultz et al., 2015) (Li, Wang, Zhang, & Lu, 2015) (Immitzer et al., 2016) (Schmidt, Pringle, Devadas, Denham, & Tindall, 2016) (Peña-Barragán et al., 2011)
Approach (2)	**Single image**	Little load of data to analyse.	Cannot distinguish among crops with similar spectral information.	(Ali, 2002)
	Multi-temporal	Allows the distinction of crops which have similar spectral information (e.g. barley and wheat).	Requires knowledge of the crop calendar. May not give good results if little images are available, because of cloud cover.	(Valero et al., 2016) (Pittman, Hansen, Becker-Reshef, Potapov, & Justice, 2010) (Schmidt et al., 2016)

Tab. 4 – Some of the most common crop mapping methods

The several and diverse crop mapping methods listed above try to deal with different problems that arise when working from the local to the global scale; some of the most important issues that must be taken into account are the following:

- Cloud cover: climatic conditions, especially over certain areas, make it difficult to obtain a series of cloud-free images over an extended period of time, which is needed when working with a multi-temporal approach; in some regions, the highest cloud cover occurs during the critical period for crop identification (Whitcraft et al., 2015);
- Intra-crop variability: the spectral response of any crop varies with the phenological stage, the health status, the climate and soil conditions, the agricultural practices (e.g. irrigation, fertilization); also the temporal response is quite variable, since it is determined by the crop calendar, which in turn is influenced by the climatic conditions of the season, the agricultural practices, the farmer's decision;
- Detection and classification of crops within some specific cropping systems: cropping systems with mixed cropping or intercropping produce pixels that are not pure and therefore not easy to classify;
- Field size: in some regions of the world, the average field size is too small compared to the pixel size, generating again pixels that are not homogeneous (Jordi Inglada et al., 2015);
- Other characteristics of the cropping or farming system: trees scattered in the fields generate shadows that modify in part the spectral response of the vegetation;
- Amount of data to analyse: computational time represents crucial characteristics of the system producing the service, especially when dealing with in-season classifications; feature reduction techniques are not always a solution because they may not capture crop variability (e.g. selecting only some dates of a time series may exclude from the analysis crops with a late sowing).

The issues mentioned represent a challenge to accurate crop type mapping systems and must be considered in relation to one's final purpose when elaborating the processing chain, which is the sequence of steps applied to extract the required information from the imagery. In the following paragraphs we will discuss how certain methodologies deal with some of these problems.

1.2.4 Vegetation Indices (VIs)

Vegetation Indices (VIs) have been extensively used for vegetation studies (Haibo & Yanbo, 2013); they are most commonly calculated as band ratios or combinations. VIs try to capture and enhance spectral characteristics that belong to green vegetation only: the most used bands are the Red and the NIR, since in these spectral regions a sudden increase in reflectance occurs. The behaviour in these regions is called "Red Edge" (Fig. 5) and cannot be found in objects like water or soil.

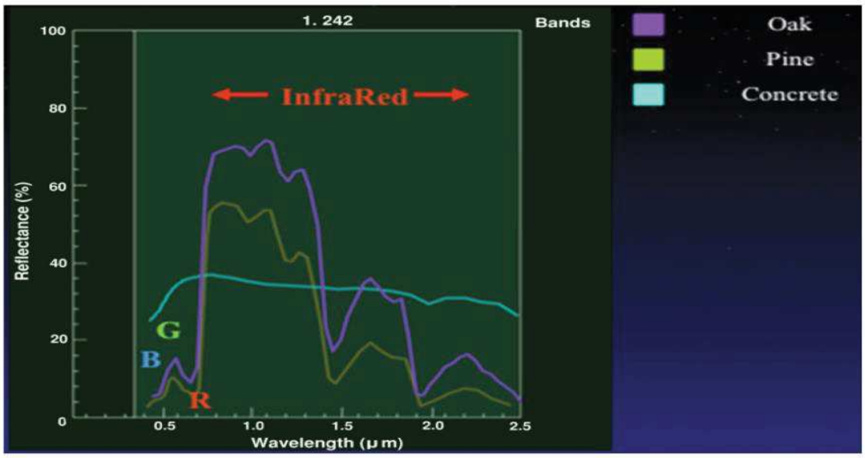


Fig. 5 – The typical spectral signatures of vegetation and soil (Khorram et al., 2012)

The most widely used vegetation index is the Normalized Difference Vegetation Index (NDVI) (Rouse et al, 1973; Taramelli et al., 2013; Azar, Villa, Stroppiana, & Crema, 2016; Schmidt et al., 2016), but many other vegetation indices have been developed over time with different characteristics; some of the most common are listed below in Tab. 5.

Name	Equation	Reference
INTRINSIC VEGETATION INDICES		
Normalized Difference Vegetation Index	$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}$	Rouse et al (1973)
Simple Ratio	$\mathrm{SR} = \frac{\mathrm{NIR}}{\mathrm{Red}}$	Jordan (1969)
Difference Vegetation Index	DVI = NIR – Red	Richardson & Wiegand (1977)
SOIL-ADJUSTED VEGETATION INDICES		
Weighted Difference Vegetation Index	WDVI = NIR – slope*Red	Clevers (1988)
Soil-Adjusted Vegetation Index	$\mathrm{SAVI} = \frac{(1+\mathrm{L})(\mathrm{NIR}-\mathrm{Red})}{\mathrm{NIR}+\mathrm{Red}+\mathrm{L}}$	(Huete, 1988)
Modified Soil-Adjusted Vegetation Index	$\mathrm{MSAVI} = \frac{(1+\mathrm{L})(\mathrm{NIR}-\mathrm{Red})}{\mathrm{NIR}+\mathrm{Red}+\mathrm{L}}$	(Qi, Chehbouni, Huete, Kerr, & Sorooshian, 1994)
WATER-SENSITIVE VEGETATION INDICES		
Normalized Difference Water Index	$\mathrm{NDWI} = \frac{NIR-SWIR}{NIR+SWIR}$	Gao (1996)

ATMOSPHERICALLY CORRECTED VEGETATION INDICES		
Global Environmental Monitoring Index	GEMI = $\eta * (1 - 0.25\eta) - \frac{Red - 0.125}{1 - Red}$	Pinty & Verstraete (1992)
ATMOSPHERE- AND SOIL-ADJUSTED VEGETATION INDICES		
Enhanced Vegetation Index	EVI = $2.5 * \frac{NIR - Red}{NIR + C1 * Red - C2 * Blue + L}$	Liu & Huete (1995)
FEATURE SPACE-BASED VEGETATION INDICES		
Tasselled Cap transform green VI	TCG = $a*Blue + b*Green + c*Red + d*NIR + e*SWIR1 + f*SWIR2$	Kauth & Thomas (1976)
Perpendicular Vegetation Index	PVI = $\frac{NIR - a*Red - b}{\sqrt{1 + a^2}}$	Richardson & Wiegand (1977)

Tab. 5 – Some of the most common vegetation indices

VIs have been said to quantify the "greenness" of a pixel (Huete,2004). Indeed, they have been related to specific plant characteristics, for instance:

- vigour and biomass (Todd, Hoffer & Milchunas, 1998);
- health status (Peters et al, 2002);
- Leaf Area Index (LAI) (Carlson & Ripley, 1997);
- Fractional Vegetation Cover (FVC) (Carlson & Ripley, 1997);
- chlorophyll content (Gamon et al, 1995);
- fraction of Absorbed Photosynthetically Active Radiation (fAPAR);
- canopy structure (Baret & Guyot, 1991);
- plant development and phenology (Vina et al., 2012).

In most cases, the measure of these characteristics is meant to assess the agro-ecosystem functionality; however, VI-derived information on plant phenology has been used also for crop classification.

1.2.5 Phenological Metrics (PMs)

Monitoring crop phenology is an important task for the understanding of the agro-ecosystem, since it can provide much information: among this, how the growing seasons change through space and time (White et al, 2009); how agro-ecosystems react to climate change (White et al, 2009); how much the agro-ecosystem produces (Bolton&Friedl, 2013). Remote sensing techniques have been used to monitor crop phenology, taking advantage of the temporal resolution of some sensors (e.g. MODIS, Landsat, Sentinel); this is done by means of the temporal profile of a chosen Vegetation Index; since each crop has got its own phenological stages, that can occur at different times during the year, observation of crop phenology has also been used for crop type classification: for instance, (Jin et al., 2016) used it to distinguish rainfed from irrigated wheat; (Li et al., 2015) classified six crop types based on NDVI and NDVI Time Series Indices.

On the other hand, it is important to keep in mind two fundamental concepts: first, the signal coming from the surface is not always pure and is often contaminated by the influence of atmosphere and soil; for this reason, (M. A. White et al., 2009) prefers to speak of Land Surface Phenology (LSP), rather than Plant Phenology (PP), thus remarking that the VI temporal profile needs to be interpreted. Secondly, the phenological stages detected from EO cannot represent directly all the crop-specific phenological stages, some of which (e.g. kernel dough stage) are of extreme importance to agronomists (Zeng et al., 2016). Of course it is possible to derive the crop-specific phenological stages from the remotely sensed ones with specific models (e.g. Zeng et al, 2016) or observations (Vina et al., 2012), but these studies are beyond classification purposes. When working with classification, only some phenological metrics can be used, derived from evident features of the VI temporal profile (see Fig. 6 for some examples). For instance, Jin et al. (2016) used the peak of the NDVI temporal profile to distinguish rainfed from irrigated wheat; Li et al., (2015) developed some NDVI Time Series Indices (TSIs) describing the changing pattern of the vegetation index along the season; also Schmidt et al. (2016) used temporal variables, for instance the day in which the NDVI peak occurred, for cropland mapping; Matton et al. (2015) selected the reflectance values to use in the classification deriving them from the days of the maximum and minimum NDVI value, the maximum positive and negative NDVI slopes. The mentioned studies prove that this type of metrics can provide satisfactory results both for cropland and crop type mapping.

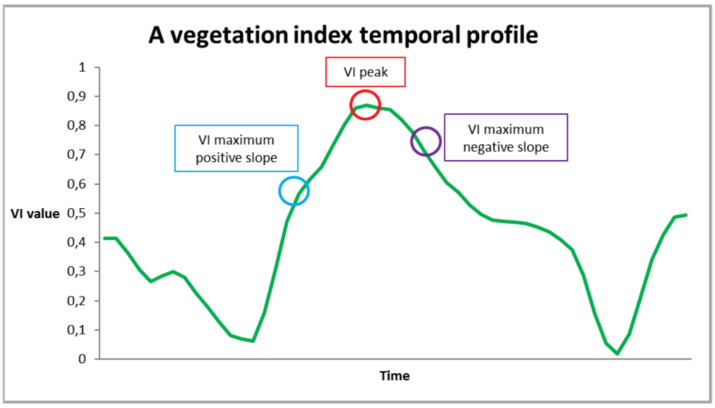


Fig. 6 – Some of the metrics that can be derived from the temporal profile of a vegetation index

1.2.6 Segmentation

Segmentation has been defined as the process of partitioning an image into non-overlapping regions, which are called segments (Schiewe, 2002). This technique has been originally developed to deal with the intra-class spectral variability found in the higher spatial and spectral resolution imagery (Khorram et al., 2012).

Indeed, the segmentation algorithm can produce clusters of pixels that are homogeneous and are therefore simpler to analyse in comparison to the single pixels; moreover, it reduces computational time. Some of the features that can be taken into account to describe the pixels' homogeneity are the following ones (Dey, Zhang, Zhong, & Engineering, 2010):

- spectral characteristics: the pixel values of an image over single or multiple bands;
- texture: the spatial pattern represented by pixel values (Haralick et al, 1973);
- shape and size: two complementary measures used to distinguish objects that have similar spectral features but different form (e.g. a river and a lake);
- context: the relationship of a pixel with its neighbourhood (Thakur & Dikshit, 1997);
- temporal characteristics: they are not directly used as a measure for segmentation, it is rather an approach to the analysis of the above mentioned metrics.

Segmentation algorithms have been used to implement an approach called Object-Based Image Analysis (OBIA): some chosen attributes are used for the segmentation and in the classification process all the pixels belonging to a certain object are assigned to the same class (Blaschke, Lang, & Hay, 2008). This approach is particularly useful in agricultural contexts, where the intra-field variability, caused by mixed pixels at the field borders (De Wit & Clevers, 2004), soil variability or other factors, may lead to undesired salt-and-pepper effects; the application of OBIA, instead, generates objects corresponding to the single fields (Peña-Barragán et al., 2011). The choice of a good segmentation can influence the final results of the classification phase (B. Schultz et al., 2015).

OBIA has been successfully used in Object-based Crop Identification and Mapping (OCIM) (Peña-Barragán et al., 2011), in cropland mapping (Schmidt et al., 2016), with a temporal approach (Li et al., 2015), providing similar or better results in comparison to the pixel-based approach (Valero et al., 2016). In most cases, the quality of the object-based classification is determined by the characteristics of the real fields, and in particular by their dimension (Valero et al., 2016).

1.2.7 Classification

Classification is a process in which each pixel of an image is assigned to a category, among a set of categories of interest (Khorram et al., 2012); this allows the transformation of numerical data into ordinal or thematic information. The classification process consists of three phases (Brivio, Lechi, & Zilioli, 2006): training, allocation and validation.

In the training phase, the analyst defines a set of information in order to help the system distinguish the different classes; in practice, using ancillary data, the analyst selects some pixels, for the training, leaving also a sufficient number of pixels for the validation. Usually, the minimum number of pixels per class accepted is between 10*k* and 30*k*, where k is the feature space dimension (Brivio et al., 2006). The features space dimension is most commonly represented by the spectral bands of the image, but it can also consist of synthetic spectral bands containing other types of information (e.g. Vegetation Indices); the number of bands to be used in the feature space should be accurately chosen in order to avoid the overfitting of the model. There is not a defined training:validation ratio to be used for the division of the pixels; it is possible to find studies using training:validation ratio of 2:1 (Azar et al., 2016), 1:1 (Hao, Zhan, Wang, Niu, & Shakir, 2015), 1:2 (Valero et al., 2016), 80:20 (Brown, Kastens, Coutinho, Victoria, & Bishop, 2013), 70:30 (Oumar, Mutanga, & Ismail, 2012). Several criteria can be used to choose the training and the validation pixels; however, Random Sampling (Hao et al., 2015) or Stratified Random Sampling (Azar et al., 2016) are most commonly used.

The training phase is not always part of the classification process, since it requires ancillary data. Based on the presence of the training phase in the process, classification systems can be divided into:

- supervised systems: the analyst defines the training sites that represent the classes in the chosen classification scheme; then the classification algorithm assigns each pixel to the best matching class (Khorram et al., 2012);
- unsupervised systems: the classification algorithm defines clusters of pixels, on the basis of features selected by the analyst; studying the output clusters, the analyst assigns them to the classes of interest. Unsupervised classification systems can be useful when there is not enough input data, but requires the analyst to define accurately the parameters (e.g. number of classes). The final purpose is indeed to have clusters representing a defined class, but if the parameters are not set properly, the clusters may comprise just mix of classes.

In the allocation phase, the algorithm assigns a label (the class) to each pixel in the image; allocation algorithms can be divided into:

- hard classifiers: they adopt a Boolean logic according to which each pixel strictly belongs or not to a class. They include, for instance, the Maximum Likelihood Classifier (MLC) and the Parallelepiped Algorithm (PA);
- soft classifiers: they admit the possibility that a pixel belongs to more than one class; they include fuzzy logic systems, Artificial Neural Networks (ANN), Spectral Unmixing (SMA) and Multinomial Logistic Regression (MLR).

In general, there are some main issues that the analyst should keep in mind when choosing the algorithm to use (Millard and Richardson, 2015): the Hughes phenomenon, which is the increase in classification accuracy when there is an increase in the input variables, to a point in which the accuracy decreases (overfitting point); non-linearity of variables; imbalanced training samples and noise in training and validation data; computational time. Thus, the pros and cons of each algorithm should be considered.

The main advantage of soft classifiers is the possibility to deal openly with the problem of “mixed pixels”: especially when dealing with low and medium-spatial resolution data, pixels rarely contain only one object within their boundaries. Soft classifiers admit the possibility that one pixel does not exactly correspond to one class. For instance, with SMA the observed radiance is modelled as a mixture of spectrally pure endmember radiances (Taramelli et al., 2013), whose contribution to the total pixel radiance is calculated. In particular, in Linear SMA (LSMA), the contributions of each endmember in a pixel are modelled as a linear combination of endmember spectra weighted by the percentage ground coverage of each endmember (Elmore, Mustard, Manning, & Lobell, 2000). Thus, SMA provides abundance maps of the chosen classes, which show the percent cover of each class in every pixel. This approach has been used for vegetation studies, ranging from habitat type mapping (Valentini, Taramelli, Filipponi, & Giulio, 2015) to urban vegetation abundance (Small & Lu, 2006). There are studies showing that the SMA is a robust method for the estimation of vegetation abundance (e.g. Elmore, Mustard, Manning, & Lobell, 2000) and that therefore it is well suited for monitoring vegetation health and abundance (Small & Lu, 2006), particularly in arid and semi-arid environments (Elmore et al., 2000).

However, maps of continuous variables, such as abundance, are difficult to interpret for a non-experienced user. On the other hand, the translation of this quantitative information into a thematic map poses a challenge even for experienced analysists. In addition to this, analyses conduced with soft classifiers tend to be quite time-consuming. Finally, the sub-pixel analysis obtained with soft classifier is not always necessary: when the size of the object is bigger than the size of the pixel, the performance of hard classifiers is only marginally affected by mixed pixels. For these reasons, the use of hard classifiers is more common: they can quickly provide easy-to-interpret information and are suitable for the monitoring of agricultural systems in many countries.

Among hard classifiers, some of the most used algorithms for land cover classification are:

- Maximum Likelihood Classifier: it is the most used in remote sensing, since it provides a consistent approach to data variability (Chuvieco & Huete, 2010); it works by assigning each pixel to the class for which the membership probability is higher. Its main drawback is the assumption that the pixels’ Digital Levels (DLs) are normally distributed; this also implies that its accuracy decreases when there is overlap between the probability functions. However, good performances have been reported also in comparison to more complex classification algorithms (Azar et al., 2016).

- Classification Trees (CTs): a decision tree is defined as a classification procedure that recursively partitions a dataset into smaller subdivisions on the basis of a set of tests defined at each branch (or node) in the tree (Friedl & Brodley, 1997). The advantages of this classification system are the following (Friedl & Brodley, 1997): they do not require any assumption about the statistical distribution of the data; they can handle input data that have non-linear relations with the respective classes; they can handle data of different nature; they are easily interpretable; they allow the integration of expert-knowledge into the model (Valentini et al., 2015). Because of the several advantages, many automatic algorithms for building the decision tree have been proposed in the last years (Chuvieco & Huete, 2010); for instance, the Classification And Regression Tree (CART), ID3, C4.5, C5.0, J48. CTs have achieved good results also in crop type mapping (Villa et al., 2015).
- Random Forest (RF) (Breiman, 2001): a random forest is a classifier consisting of several small decision trees, instead of a single big one (like in normal CTs), which are tuned and pruned without the analyst's oversight; each tree is grown on a training set created by **B**ootstrap **agg**regat**ing** (also called bagging): this means that every training set is derived by sampling uniformly and with replacement from the original dataset; the use of slightly different training sets for each tree and the use of multiple trees helps in noise reduction. The out-of-bag (*oob*) samples are used to evaluate the error rate of the tree and to select the importance of the classification variables. Each tree node is splitting using a different set of variables, randomly selected, in order to minimize the correlation between the classifiers in the ensemble. During the classification, every tree in the forest, which provides a different result, classifies each pixel; the pixel is then assigned to the class, which was the output of the majority of the trees in the forest. The main characteristics of the Random Forests are: by adding more trees to the forest, the results converge, avoiding the problem of overfitting; the accuracy of the RF depends on the strength of the individual tree classifier and on a measure of the dependence between them; the results are insensitive to the number of features used to split each node; they perform well with high-dimensional data. In addition to this, the RF algorithm requires only two parameters to be set by the user: the number of trees to be grown and the number of features to be used at each node; this makes it easier to use in comparison to other algorithms, for instance Support Vector Machines (SVMs) (Pal, 2005). Finally, it is faster in training than other ensemble methods, can estimate the importance of the variables for the classification and can detect outliers (Gislason, Benediktsson, & Sveinsson, 2006). The advantages of RF have made it the most popular classification algorithm for land cover classification; RF has been compared to Gradient Boosted Trees and Support Vector Machines (Jordi Inglada et al., 2015), Multinomial Logistic Regression and C5.0 decision-tree classifier (Schmidt et al., 2016), J48 Classification Tree (Villa et al., 2015), Adaboost classification tree (Chan & Paelinckx, 2008) always yielding better or comparable results.

The validation phase is meant to determine the capacity of the algorithm to correctly classify the pixels. This capacity is called "accuracy" and it is defined as the degree of agreement between the produced map and the ground truth (Brivio et al., 2006). Indeed, the accuracy is expressed as the percent correctly classified sample sites, as compared to the corresponding reference data (Khorram et al., 2012). The accuracy can be calculated over the single classes of interest or over all the classes (Overall Accuracy, OA). The most used metrics to quantify the accuracy of a map are the following ones (Brivio et al., 2006):

- User's accuracy (UA): the ratio between the correctly classified pixels in a class and the total number of pixels assigned to that class in the produced classification (X_C); it is associated with the Commission Error (CE);
- Producer's Accuracy (PA): the ratio between the correctly classified pixels in a class and the total number of pixels belonging to that class in the reference data (X_R); it is associated with the Omission Error (OE);
- Overall Accuracy (OA): the ratio between the correctly classified pixels and the total number of pixels in the map (X_{TOT});

This data is calculated from a contingency table, called *error matrix* or *confusion matrix*. It is a square matrix where the columns represent the ground truth (reference data) and the rows represent the results of the classification (classified data); in the diagonal of the matrix the correctly classified cases are found.

		Ref.					
		C1	**C2**	**C3**	**Tot**	**UA**	**CE**
Class.	**C1**	X_{11}	X_{12}	X_{13}	X_{C1} ($X_{11}+X_{12}+X_{13}$)	X_{11}/ X_{C1}	$(1- X_{11})/ X_{C1}$
	C2	X_{21}	X_{22}	X_{23}	X_{C2} ($X_{21}+X_{22}+X_{23}$)	X_{22}/ X_{C2}	$(1- X_{22})/ X_{C2}$
	C3	X_{31}	X_{32}	X_{33}	X_{C3} ($X_{31}+X_{32}+X_{33}$)	X_{33}/ X_{C3}	$(1- X_{33})/ X_{C3}$
	Tot	X_{R1} ($X_{11}+X_{21}+X_{31}$)	X_{R2} ($X_{12}+X_{22}+X_{32}$)	X_{R3} ($X_{13}+X_{23}+X_{33}$)	X_{TOT} ($X_{R1}+ X_{R2}+ X_{R3}+ X_{C1}+ X_{C2}+ X_{C3}$)		
	PA	X_{11}/ X_{R1}	X_{22}/ X_{R2}	X_{33}/ X_{R3}			
	OE	$(1- X_{11})/ X_{R1}$	$(1- X_{22})/ X_{R2}$	$(1- X_{33})/ X_{R3}$			

Tab. 6- The general structure of a confusion matrix

1.3 The potential of remote sensing for agro-ecological zoning

Most AEZs use remotely sensed data for their climatic variables (Mücher et al., 2016). The primary driver of change in agro-ecological zones is often considered to be climatic change (Chikodzi, Hardlife, Farai Malvern, & Talent, 2013), thus making remote sensing the perfect source of up-to-date information. Examples like the one of (Mugandani, Wuta, Makarau, & Chipindu, 2012) clearly show that agro-ecological zones are subject to noticeable change even in a short time lapse as 60 years, and therefore need to be constantly monitored. However, Land Use and Land Cover (LULC) are expected to change as rapidly (Olesen & Bindi, 2002). Land cover is an important part of agro-ecological zonations in general, as demonstrated by the case of (Mücher et al., 2016), and it represents a fundamental variable when studying nutrient dynamics and environmental pollution deriving from agricultural activities. Consequently, land cover should be monitored at least as frequently as climatic data. Conducting regular campaigns for the collection of soil and land cover data is often costly and not feasible; remote sensing, instead, may offer up-to-date information with global coverage. Other advantages of including EO-derived information into AEZs are:

- the availability of historical data: missions like Landsat, SPOT or MODIS offer historical data that can be used to track the changes occurred in agro-ecological zones in the past;
- the possibility to continuously cross-validate the available data.

Moreover, issues regarding spatial resolution can be overcome with the newest sensors, offering data at 10 to 30 m.

In conclusion, when land cover or crop types data is required in an agro-ecological zonation, remote sensing can represent the main source of information.

Chapter 2

Description of the study area

2. Description of the study area

2.1 Italy: current state of the agricultural sector

The importance of agriculture in Europe does not lie only in its contribution to the general economy (Europedia, 2011), but also in its impact on land and on the quality of environment, since famers manage almost half of the European land area (EEA, 2011). The European Union measures the impact of agricultural activities on the environment mainly in terms of soil erosion, greenhouse gas emissions, climate change, land use, irrigable area, water pollution, nitrogen surplus, biodiversity, habitat conservation and organic farming (EEA, 2011). In this framework, Italy contributes with one of the highest values of Utilized Agricultural Area (UAA) in the EU-28 (Eurostat, 2012), thus representing a relevant area for environmental issues at the European level.

At the national level, Italian agriculture is an important sector: it represents 2.2% of the total Gross Value Added (GVA) of the country. The products are diverse (Fig. 7), with noticeable differences across regions in farm structure and in production. The average farm size (12 ha) is smaller than the European average (16.1 ha), with more than 50% of the farms consisting of less than 5 ha (European Commission, 2016a); on the other hand, the UAA and the number of farms are among the highest in the EU-27 (Eurostat, 2012). These statistics clearly describe the historical problem of agricultural land pulverization in Italy, even though in the last years there has been a tendency towards the increase of farm size (Eurostat, 2012). Another important issue is represented by regional differences: the Northern regions tend to have larger (Istat, 2013) and more profitable (European Union, 2015) farms. The gap between the North and the South of the country also concerns the environmental problems: for instance, while the North is more subject to nitrate pollution (EEA, 2012a), the South has got more pressure on water sources (EEA, 2016) and it will need to deal with water-limited crop yields in the future (EEA, 2012b).

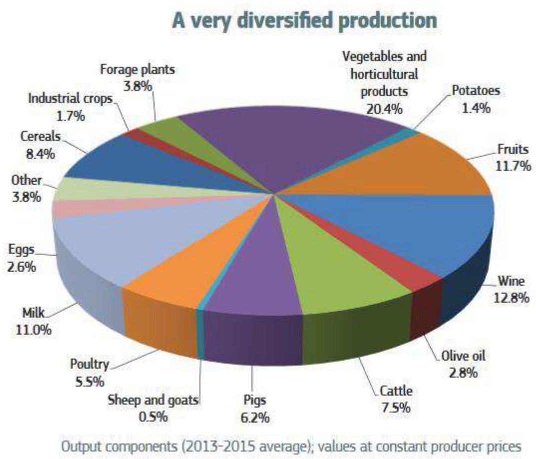


Fig. 7 – Most important agricultural products in Italy (European Commission, 2016b)

2.2 The Muzza area

The study area is located in the North of Italy, in the Lombardy region (Fig. 8); it includes shares of land from the Provinces of Lodi, Milan, Pavia, Bergamo and Cremona, for a total area of approximately 1800 km^2, with homogeneous climatic characteristics. It is an agricultural area that takes advantage of an artificial canal, the Muzza Canal, which makes it one of the biggest irrigated areas of Lombardy; the canal originates from the Adda River, which is the longest tributary of the Po River, the longest river in Italy. Along the Po River, an agricultural area extends, the Padan Plain, of which the Muzza is part.

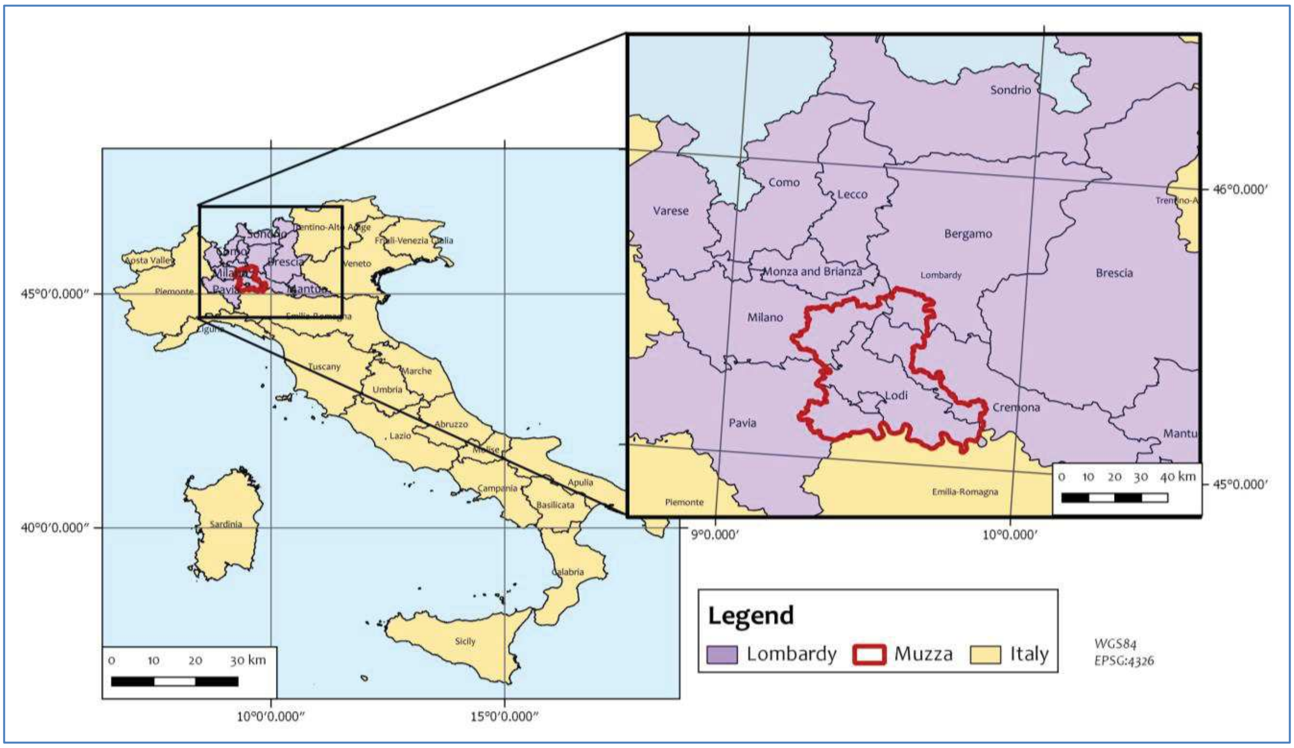


Fig. 8 - The location of Muzza within Italy and the Provinces of the Lombardy Region (Global Administrative Areas, 2012)

A great variety of crops is cultivated in the area (Regione Lombardia, 2015): cereals, cash crops, tree crops, woods, vegetables, pulses and forages are grown in the Province of Lodi; in addition to this, there are greenhouses, nurseries, rural buildings and fallow fields. Annual crops comprise winter crops, cultivated from October to June, and summer crops, cultivated from April to October (Azar et al., 2016). The crop calendar of some of the main cereals, cash crops and forage crops is illustrated in Fig. 9; the calendar of vegetables, pulses, some forages and other crops with a short growing cycle is more flexible and therefore cannot be represented in a similar chart.

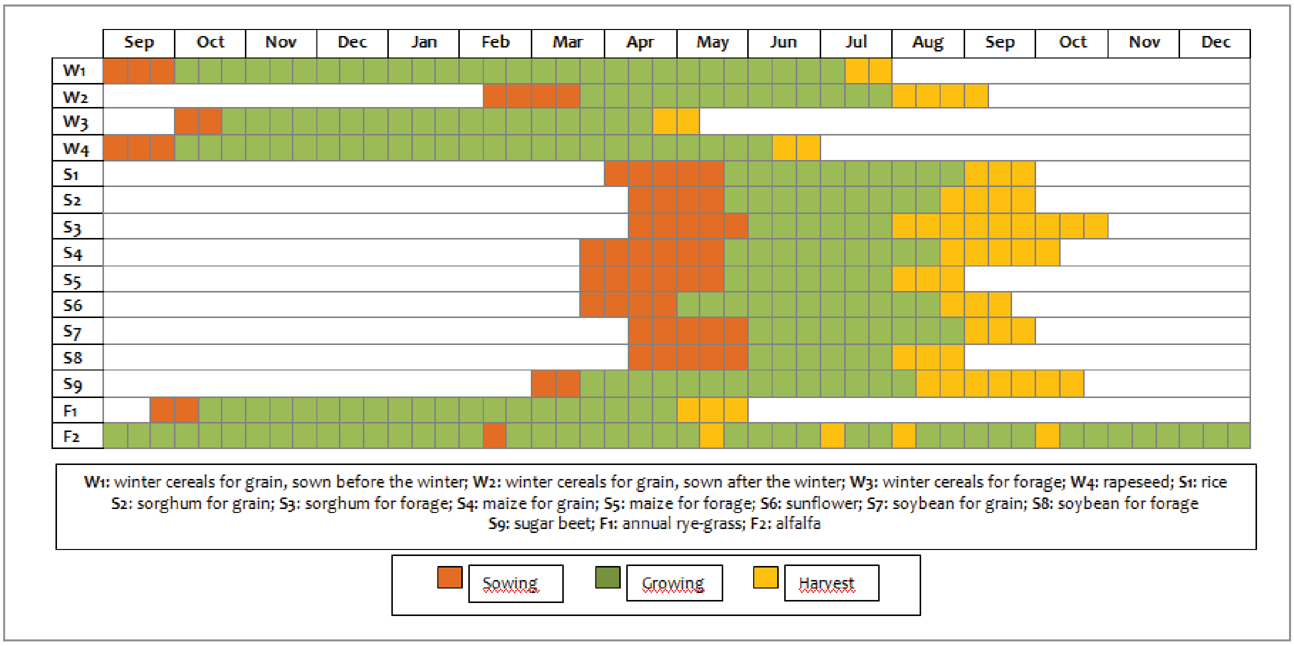


Fig. 9 – Crop calendar of some of the main cereals, cash crops and forage crops found in Muzza (re-elaborated from Baldoni & Giardini, 2001)

According to Regione Lombardia (2015), the most common crops on the territory are maize and winter cereals; in addition to this, also greenhouses and fallow or abandoned land are widespread.

The agricultural sector of the Lombardy Region is among the major ones in Italy, in economic and technological terms. Indeed, Lombardy is the most important Italian region for livestock production (Eurostat, 2012); the presence of a strong livestock production influences the farmers' choices also in terms of crops cultivated, indeed more than 50% of the UAA is destined to forages (herbages, meadows and pastures) (re-elaborated from Istat, 2011). In addition to this, Lombardy is among the top regions for average farm size (18.2 ha) (Eurostat, 2012), use of IT services (Istat, 2010) and average farm economic size (Eurostat, 2012). The economy and the overall structure of the agricultural system can be considered among the most intensive in the country.

The great development of agricultural activities is favoured by the pedo-climatic conditions. The area is flat; the soils have high potential for agricultural production, since they are not very rich in soil organic matter but in they are considered fertile in general (Costantini, Urbano, & L'Abate, 2004) The climate is temperate-suboceanic (Costantini et al., 2004), with 827.6 mm of rain per year (ASP Lombardia, 2010). The yearly amount of precipitation (on average between 670 and 1200 mm) (ARPA Lombardia, 2010), together with the water provided by the canal, makes water abundant.

However, the water resource is subject to pressure both on the quantitative and qualitative side. In the first place, the water demand is very high, coming not only from the agricultural sector but also from the industrial one; moreover, the most used irrigation technique is surface irrigation, which in general does not yield high water use efficiency. The vulnerability of the aquifer to pollution is especially influenced by the superficiality of the aquifer and the texture class of soils. Indeed, the latest report on water quality of the Lodi Province, where the aquifer is very superficial, reports a high content of nitrates and pesticides and attributes it primarily to agricultural activities (ARPA Lombardia, 2015). A study conducted in the framework of the EUCENTRE project SEGUICI studied from EO the texture of soils as a proxy for the vulnerability to pollution (Aa. Vv., personal communication, 2016; Seguici, 2017; Seguici, 2015; Google Play, 2017).

Action has been taken in order to control the use of nitrogen fertilizers and pesticides and decrease the risk of water pollution. In particular, the management of nitrogen fertilizers and of liquid manure is regulated according to the European and Italian law. The legislative references are the EU Nitrates Directive (European Commission, 1991) and the Action Programme of the Lombardy region (ERSAF Lombardia, 2016). The Action Programme defines the areas which are subject to the rules and sets the rules themselves. Beginning from the growing season 2016\2017, the whole territory of the Lombardy Region has been divided into 6 pedoclimatic zones, defined on the basis of the climate and soil (Vulnera & Guida, 2016); a biweekly bulletin ratifies the permission to use nitrogen fertilizers in the single zones. The case of Lombardy thus represents a perfect example of how agro-ecological zonations can help researchers in defining the risk of nitrate pollution from agriculture; in turn, researchers serve as a valuable support for policy makers dealing with environmental legislation.

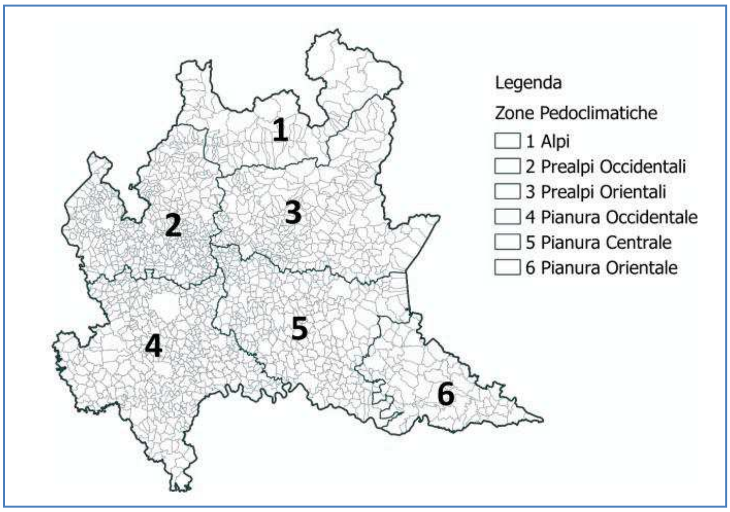


Fig. 10 – The pedoclimatic zones of the Lombardy Region: Alps (1), Western Prealps (2), Eastern Prealps (3), Western Plain (4), Central Plain (5), Eastern Plain (6) (Vulnera & Guida, 2016)

2.3 Location of Muzza within the existing AEZs

According to different zonations, Muzza is divided into a different number and type of zones (Tab. 7 and Fig. 11). The division into pedo-climatic zones provided by the Region Lombardy (from now on called LoPCZ) is also included. See also Appendices 1-5 for the maps of each zonation.

Zoning system	N. of zones	Zone type
GYGA-ED	4	4502 GDD= 3792 – 4829; Aridity Index= 6589 – 7785; Temperature seasonality= 3833 - 8355
		4602 GDD= 3792 – 4829; Aridity Index = 7786 – 8685; Temperature seasonality= 3833 - 8355
		4702 GDD= 3792 – 4829; Aridity Index = 8686 – 10181; Temperature seasonality= 3833 - 8355
		4802 GDD= 3792 – 4829; Aridity Index = 10182 – 12876; Temperature seasonality= 3833 - 8355
GAES	1	Warm temperate and mesic hills dominated by rocks and cropland
GEns	1	Warm temperate and mesic
HCAEZ	2	Temperate\Humid
		Temperate\Sub-humid
LoPCZ	2	Central Plain
		Western Plain

Tab. 7 - The agro-ecological zones into which Muzza is divided, according to different zoning systems

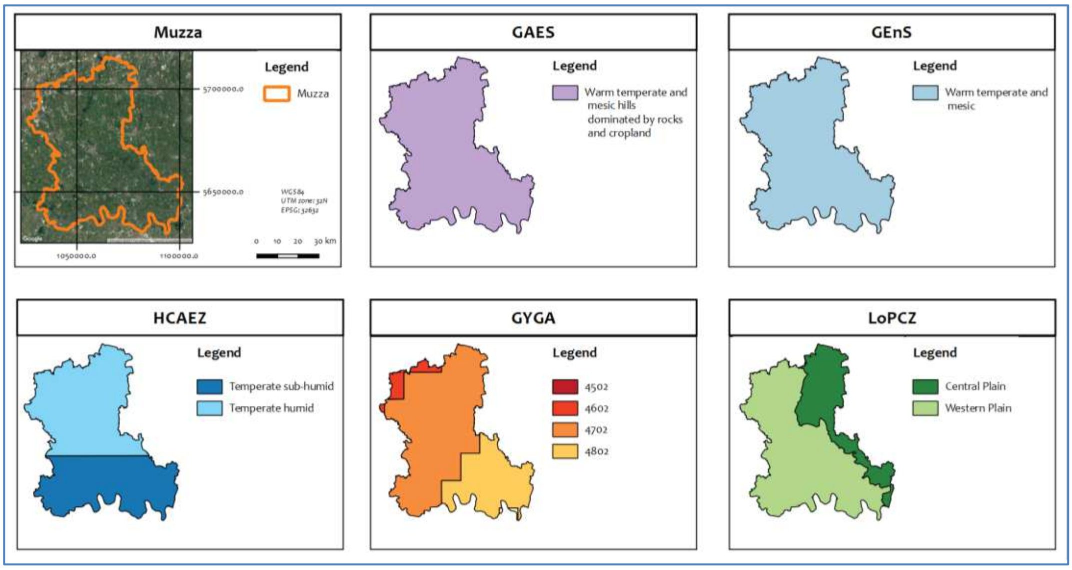


Fig. 11 - The division of Muzza into agro-ecological zones, according to five existing zonations (GAES, GEnS, HCAEZ, GYGA and LoPCZ). Credits for the satellite image at Google Maps, ©2017 TerraMetrics

The number of zones is the most evident difference among the zoning systems. The GAES and the GEnS describe the whole area with one zone only and therefore are too general to be used at the local level; GYGA, HCAEZ and LoPCZ, instead, account for territorial differences.

It can be noted that while LoPCZ divides the area into a Western and an Eastern part, GYGA-ED and HCAEZ show a North-South gradient (Fig. 11). The Region Lombardy did not provide any description of the pedo-climatic zones, therefore it is not possible to understand what are the factors differentiating the two LoPCZ zones found in Muzza.

The four GYGA zones show the same temperature seasonality and the same GDD, but different aridity indices. Aridity indices are mostly used for water and irrigation management, but they can also prove useful for the estimation of the aquifer vulnerability to nitrate pollution, since rainfall influences the percolation of nutrients through the soil profile. Unfortunately, the GYGA-ED only accounts for climate characteristics, thus it cannot provide a complete view over the factors influencing nitrate pollution in the area.

The same comments can be applied to the HCAEZ zones. The Northern part of Muzza has a mean monthly temperature adjusted to sea-level less than 5°C for 1 or more months and has a growing period longer than 270 days; the Southern part of Muzza has a mean monthly temperature adjusted to sea-level less than 18°C for 1 or more months and has a length of the growing period between 180 and 260 days. These characteristics can be considered quite general and only consider the climate features of the area, while pedological conditions are as important when dealing with nutrient cycles and pollution.

Indeed, the Map of the Soils of Italy (Italian National Centre for Soil Mapping, 2012) (Fig. 13) shows that Muzza is located in the region of the soils of Po and associated hills, that includes Cambisols, Calcisols, Luvisols, Vertisols and Fluvisols. Differences among these types are soils are relevant for agricultural activities (Brady & Weil, 2008), in that they have different contents of soil organic matter and micronutrients and that they respond differently to precipitations and irrigation. Also the above mentioned project SEGUICI managed to distinguish soils of the area from EO, detecting two types of soils with different spectral signatures (Aa. Vv., 2016) (Fig. 12). However, almost all the zonations reviewed do not take such soil variability into account.

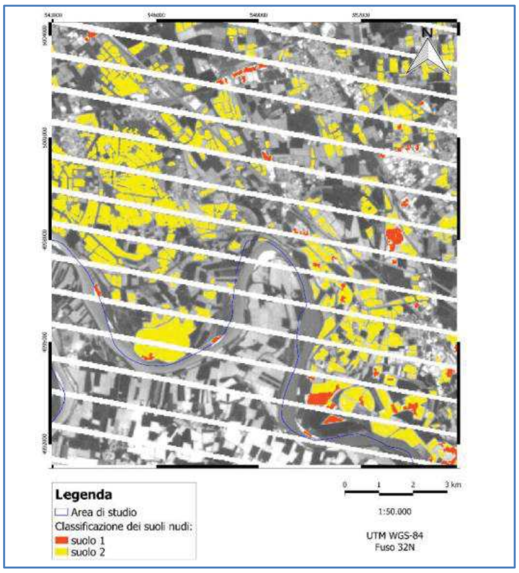


Fig. 12 - A detail of the two types of soil detected in the Muzza for the SEGUICI project (Aa. Vv., personal communication, 2016)

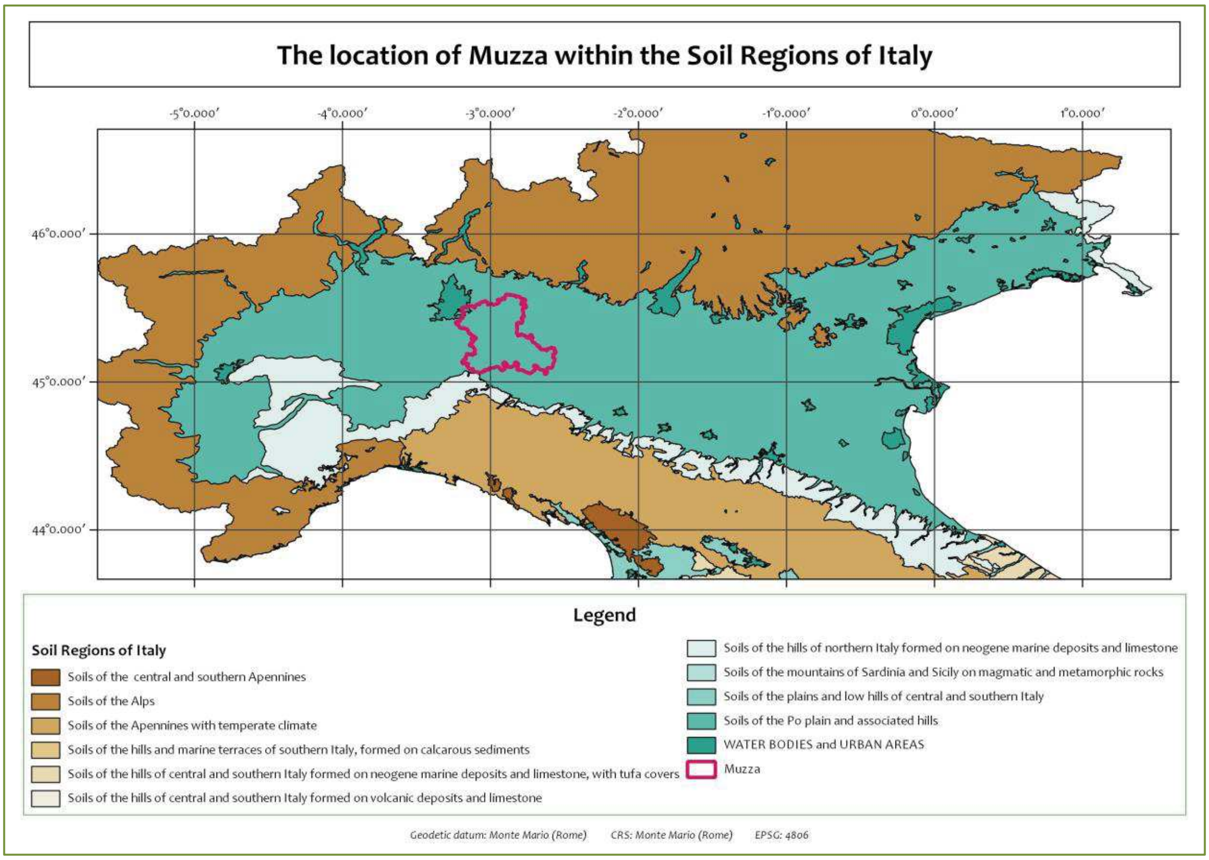


Fig. 13 - The location of Muzza within the Soil Regions of Italy (re-elaborated from Costantini et al., 2004)

The following conclusions can be drawn from the review of existing AEZs for use in the evaluation of nitrate pollution from agricultural activities:

- most of the global zonations do not provide enough detail to be used in territory analysis and in planning at the local level;
- the zonations that describe climatic territorial differences, such as the GYGA-ED and the HCAEZ, lack soil description; given that soil characteristics may have an influence on environmental pollution derived from agricultural activities and that the study area shows soil variability, their contribution to the monitoring of environmental pollution from agricultural activities can only be limited;
- the LoPCZs shows a good degree of detail while accounting for both soil and climate factors, but the description of the parameters used cannot be retrieved from official sources.

Chapter 3

Materials and methods

3. Materials and methods

3.1 Data

3.1.1 Ancillary data

The following ancillary data was used:

- reference data, consisting of a dataset with the description of the agricultural cadastral parcels of the Province of Lodi (Regione Lombardia, 2015): it contains information about the crop(s) cultivated in each parcel (*land_cover*), the total area of the parcel (*parcel_area*), the area of the parcel dedicated to each crop (*field_area*). The information is provided by the farmers owning the parcels; the dataset is referred to the crop year 2014-2015, which spans from the 1st of November 2014 to the 31st of October 2015; only the data of the Province of Lodi was chosen since this administrative region represents almost 50% of the study area;
- a vector file of the cadastral parcels of the Province of Lodi (Lombardia, 2017);
- a mask of the study area, which was produced in the framework of the EUCENTRE project SEGUICI.

The reference data and the vector file are open source and provided by the Region Lombardy through its online portal (Lombardia, 2017). Every file is provided in the UTM32N coordinate system and WGS84 geodetic system.

3.1.2 Imagery

The study area is covered by 2 Landsat8 scenes: Path 194, Row 28 and Path 193, Row 29; this overlap ensures that images of the area are acquired every 8 days. Landsat8 imagery can be freely downloaded from the USGS Earth Explorer, as a Level 2A product, which contains surface reflectance after atmospheric correction(USGS, 2014).

The period ranging from 04-Aug-2014 to 29-Dec-2015 was analysed, for a total of 65 scenes. The period of interest is represented by the crop year 2014-2015, which spans from the 1st of November 2014 to the 31st of October 2015; however, it was decided to extend the analysis up to November 2015 to make sure that late sown crops were also harvested. In addition to this, the months of August, September and October 2014 and December 2015 were used in order to improve the interpolation of the temporal VIs curves. It was decided to analyse the crop year instead of the solar year of the Gregorian calendar (1st January – 31st December) in order to analyse with continuity also the growth of winter crops, which may be sown in October and harvested at the beginning of the summer of the next year. The cloud cover over the area of interest in each image was visually estimated to have a descriptor of the quality of the imagery. The images used are listed in Appendix 7.

Fig. 14 shows that more than half of the images used have got a cloud cover higher than 60%. Usually, this value is adopted as a threshold to determine which images to keep and which to discard. However, not always a high cloud cover affects all the pixels of the image; therefore, it was decided to keep all the images, assigning a quality flag to the single pixels. The quality flags are used in the next stages of the processing chain, to decide if to keep the pixel or to discard it.

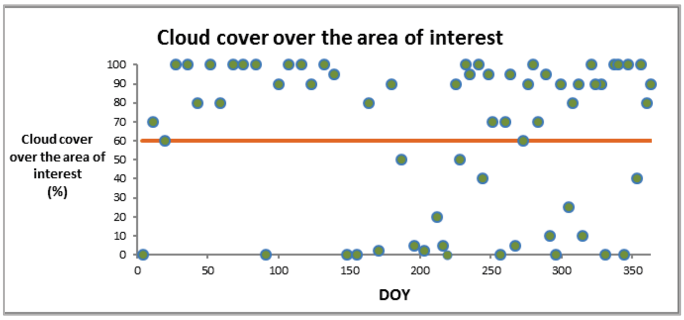


Fig. 14 – Timeline of the cloud cover over the area of interest; the orange line divides the dates with more than 60% cloud cover from the others

3.2 Processing chain

The general workflow is shown in Fig. 15. The single steps of the procedure are explained in the following paragraphs.

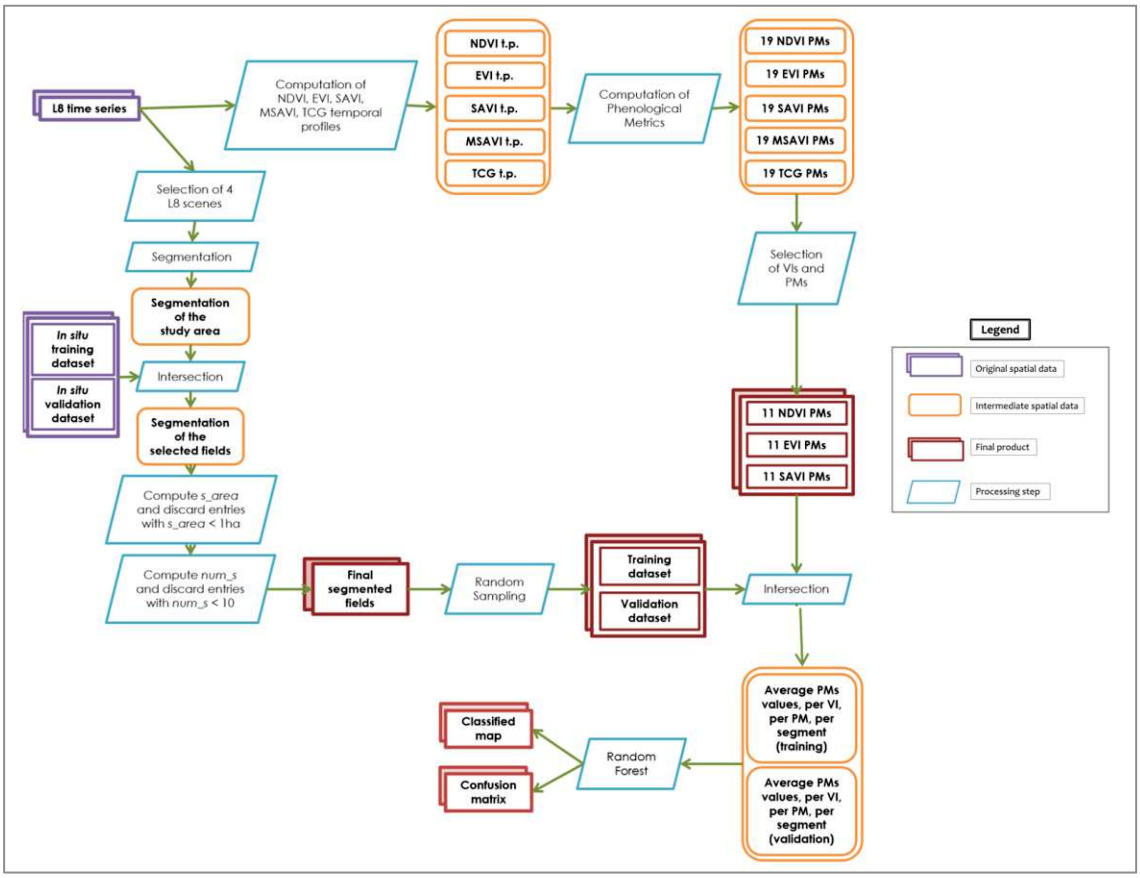


Fig. 15 - The processing chain used in this study;
the construction of the *in situ* training and validation datasets is described in §4.2.2.4

3.2.1 Pre-processing

The pre-processing consisted of atmospheric corrections. Cloud, atmospheric aerosol or other suspended particles in the atmosphere may alter the data recorded by the sensor; atmospheric corrections are meant to fix these atmospheric effects on the image. The Italian Institute for the Research and the Protection of Environment (ISPRA) provided the pre-processed data.

3.2.2 Calculation of VIs and PMs

Calculation of the VIs and PMs was performed by the ISPRA. A general description of the calculated metrics follows.

In order to test VIs with different characteristics, VIs from different categories were calculated: the NDVI for the intrinsic VIs the SAVI and the MSAVI for the soil-adjusted VIs; the EVI for the atmosphere- and soil-adjusted VIs; the TCT for the feature space-based VIs (equations (1), (5), (6), (9), (10) of Tab. 5). The VIs were calculated over all the images using the Landsat8 bands as shown in Tab. 5.

Vegetation Index	Calculation with L8 bands
$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	$\text{NDVI} = \frac{\text{B5} - \text{B4}}{\text{B5} + \text{B4}}$
$\text{SAVI} = \frac{(1+\text{L})(\text{NIR}-\text{Red})}{\text{NIR}+\text{Red}+\text{L}}$	$\text{SAVI} = \frac{(1+\text{L})(\text{B5}-\text{B4})}{\text{B5}+\text{B4}+\text{L}}$
$\text{MSAVI} = \frac{(1+\text{L})(\text{NIR}-\text{Red})}{\text{NIR}+\text{Red}+\text{L}}$	$\text{SAVI} = \frac{(1+\text{L})(\text{B5}-\text{B4})}{\text{B5}+\text{B4}+\text{L}}$
$\text{EVI} = 2.5 * \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{C1}*\text{Red}-\text{C2}*\text{Blue}+\text{L}}$	$\text{EVI} = 2.5 * \frac{\text{B5}-\text{B4}}{\text{B5}+\text{C1}*\text{B4}-\text{C2}*\text{B2}+\text{L}}$
TCG = a*Blue + b*Green + c*Red + d*NIR + e*SWIR1 + f*SWIR2	TCG= (-0.2941*B2) + (-0.2430*B3) + (-0.5424*B4) + (0.7276*B5) + (0.0713*B6) + (-0.1608*B7)

Tab. 8 - The vegetation indices chosen for this study and their calculation with Landsat8 bands; the TCG coefficient were chosen according to (Baig, Zhang, Shuai, & Tong, 2014)

Since the use of Phenological Metrics was proved to be useful for crop type classification, 19 PMs were computed using the temporal profiles of the VIs as the input. Their list and description is shown in Tab. 9, while Fig. 16 shows where the PMs are detected on a sample VI temporal curve. The Number of Growing Seasons was computed once for each pixel, while the remaining 18 metrics were computer for each of the seasons detected.

Phenological metric	Acronym	Description
Number of Growing Seasons	NGS	Number of crops grown.
Start Of Season: VI value	SOS	Minimum value of the VI before the curve begins to increase.
Start Of Season: date	SOS_date	Date on which the SOS occurs.
Start of Growing Season: VI value	SGS	Value of the VI in the point where the curve begins to grow; it represents the moment when the plant begins to grow.
Start of Growing Season: date	SGS_date	Date on which the SGS occurs.
Peak: VI value	Peak	Maximum value of the VI during the growing season.
Peak: date	Peak_date	Date on which the peak occurs.
End of Growing Season: VI value	EGS	Value of the VI in the point where the curve begins to decrease; it represents the moment when the plant begins to turn yellow.
End of Growing Season: date	EGS_date	Date on which the EGS occurs.
End Of Season: VI value	EOS	Value of the VI where the curve reaches a minimum after the peak; it represents the moment of the harvest.
End Of Season: date	EOS_date	Date on which the EOS occurs.
Amplitude	Amp	Difference between the VI value at the peak and the VI value at the SGS.
Maximum Growth Rate	MGS	Maximum positive slope of the curve.
Maximum Growth Rate: date	MGS_date	Date on which the MGS occurs.
Maximum Senescence Rate: date	MSR	Maximum negative slope of the curve.
Maximum Senescence Rate	MSR_date	Date on which the MSR occurs.
Duration Of Season	DOS	Number of days between the SGS and the EGS; it represents the length of the growing cycle of the plant.
Length of Maturity Plateau	LMP	Number of days for which the VI has got a value which is higher than the 7/10ths of the peak.
Seasonal Time Integrated index	STI	The integral of the VI curve calculated between the SGS and the EOS.

Tab. 9 – Phenological metrics tested in this study

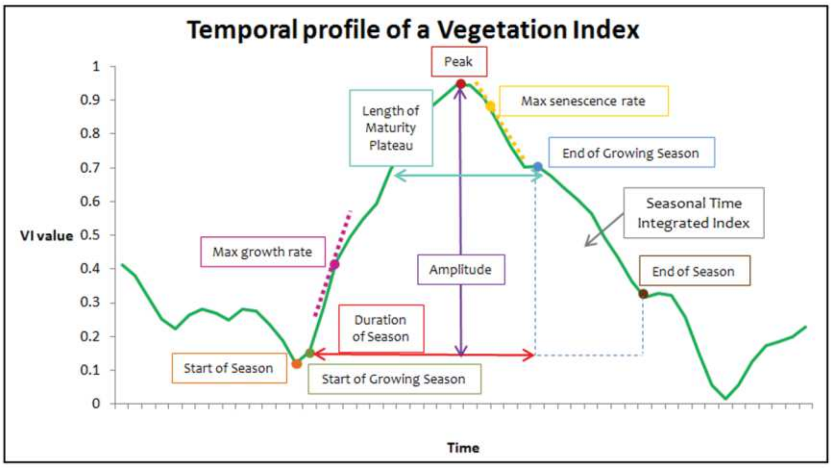


Fig. 16 - Typical location of the calculated phenological metrics on a sample VI temporal curve

3.2.3 Selection of VIs and PMs

After the computation of the VIs and of the PMs, only some of them were selected to be used in the classification. Three types of analysis were done to help in the selection:

- a visual assessment of the performance of the indices: it entailed evaluating their general behaviour, which also affects the detection of the phenological metrics. It is important for the temporal curve to be neither too smooth nor too rough: in the first case, the minimum and maximum values (e.g. peak, end of seasons) would not be detected; in the second case, local minimum or maximum values would be erroneously detected as phenological stages. Consequently, some indices can detect some specific PMs more accurately than others. The visual assessment evaluated the smoothness or roughness of the curves, together with the values of the PMs. VIs that produced too smooth\rough temporal curves were taken out of the analysis; VIs evaluated positively, instead, were kept for the final classification.
- an assessment of the performance of the PMs: by checking their values, it was determined if there were metrics systematically producing any type of error or metrics that did not contribute a reasonable amount of information.
- a statistical assessment of the performance of VIs and PMs together: the following statistical metrics were computed, per PM, per VI, per class: minimum, maximum, first quartile, median, third quartile, frequency distribution. The metrics were then graphically represented in the form of violin plots. These plots are similar to box plots, in that they show the median, the first and the third quartile, with the whiskers representing the maximum and the minimum values; however, they also have a kernel density plot along the sides.

The analysis was done using QGIS 2.18.3 (QGIS Development Team, 2017).

3.2.4 Segmentation

Image segmentation was performed over the whole area, since most of the fields are bigger than 1 pixel (900 m^2). This was confirmed by a visual analysis of the images and by national statistics: indeed, according to the Italian National Institute of Statistics (ISTAT), only 11.5% of the companies of the provinces included in the Muzza area have got a total UAA lower than 1 ha (ISTAT, 2012). Performing the segmentation only on one image would not allow to separate effectively fields which are contiguous and which have got the same type of land cover (green vegetation, irrespective of the crop, or bare soil) at the same time. Using more than one image, instead, increases the probability of separating different fields: indeed, it could be that two contiguous fields are covered with a summer crop at a given moment in the summer; but it’s less likely that they have the same summer crop and the same winter crop in a year. Consequently, images of different periods of the year were chosen to perform the segmentation, compatibly with cloud cover. Only Bands 4 (red) and 5 (NIR) were used, to reduce data redundancy; the following dates were used to enhance the temporal variance in the input segmentation dataset:

Date (dd\mm\yy)	DOY
01\04\2015	91
20\06\2015	171
15\07\2015	196
24\09\2015	267
27\11\2015	331

Tab. 10 - Dates of the Landsat8 images used as input for the segmentation phase

Segmentation was performed with a double purpose: first, to define homogeneous polygons; second, to reduce the computational time, which is one of the objectives of segmentation in general.

Then, the output of the segmentation was intersected with the parcels of the Province of Lodi in QGIS, in order to further increase the level of homogeneity within the polygons.

Following the intersection, polygons with an area lower than 1 ha were discarded; furthermore, classes with a number of fields lower than 10 and classes with a total class area lower than 10 ha were excluded. The crops excluded from this selection were aggregated into the class "Other crops" and used as well for the training and validation of the classification algorithm. The purpose of this latter class is to gather pixels of crops that are not included in the selected ones. If the class "Other crops" is not present, a pixel of sorghum may be classified as any of the selected classes, for instance maize; the same would be for all the pixels of the study area, resulting in a crop map consisting only of some crops. Keeping an extra class, instead, ensures the distinction between the selected classes and all the other crops.

Finally, for each polygon, only the median values of the Phenological Metrics were considered. Indeed, the presence of trenches, natural vegetation or tree rows along the field borders may introduce outliers affecting the training phase; using the median, the problem of outliers can be avoided.

3.2.5 Selection of the training and validation datasets

The ancillary data was organized and divided to build one training and one validation dataset (Fig. 19). Since the retrieved data was not consistent with the objects that can be acquired from EO, the following steps were performed, in order to make it usable for the purposes of the analysis:

1) the vector file of the agricultural parcels of the Province of Lodi was intersected with the mask of the Muzza: this allowed to discard from the analysis some parcels, in the South of the Province, that did not fall within the studied area.
2) In the reference data, the 218 registered crops were grouped in 71 final classes: indeed, some classes were repeated with slightly different names (e.g. "Maize" and "Sweet maize"); other classes were grouped together according to the characteristics of their growing cycle (e.g. silo maize and maize for the production of energy are harvested at the same stage of the growing cycle); finally, some crops were grouped according to their assumed spectral similarity (e.g. barley and wheat were renamed "Winter cereals"). The 71 classes belong to 5 macro-categories, namely cereals & cash crops, forages, pulses & vegetables, trees, other classes (including non-cultivated land, greenhouses, rural buildings, nurseries and buffer areas). Appendix 6 shows the classes found in the dataset and the classes in which they were grouped.
3) The reference data was joined with the table of attributes of the vector file, using the cadastral codes as the common fields in the union. The output was a shapefile in which to each parcel polygon was attributed a label with the corresponding class(es).
4) The entries relative to fields with an area lower than 1 ha were discarded (*field_area*), since hardly detectable from EO.

a) The number of crops registered per parcel (*crop_count*) was computed; the parcels with more than 1 registered crop were discarded. The purpose of this choice was to have a high degree of certainty over the class present in the pixel, needed both for training and validation. Indeed, in the dataset there were also parcels with more than one class: this data may imply that there was a crop succession during the year or that the parcel is divided into multiple fields with different crops. However, the vector file only contains information about the parcels, and not about the single fields; consequently, it is not possible to know with certainty the type and number of crops in the pixels of parcels with multiple classes, which were kept out of the analysis.

5) Two metrics were computed for the remaining entries: the total area of the class, calculated as the sum of the areas of the fields where the class was registered (*class_area*); the number of fields on which that class was registered (*num_fields*).
6) Classes with a *class_area*<10 ha were discarded, since they were considered not enough represented on the territory.
7) Classes with a number of fields lower than 10 *(num_fields*<10*)* were discarded, so as to have at least 7 fields for the training phase and 3 fields for the validation phase.

8) This final dataset was divided into two subsets, one for training and one for validation: within the polygons of each crop class, 70% was destined to training and the remaining 30% to validation. The Random Sampling algorithm was applied.

At the end of step 7), 17 classes were left (Tab. 11) and the total number of entries was reduced from 68,630 to 4,785.

Class	Area (m^2)	Number of fields
Other cereals and cash crops	235034	16
Non-cultivated land	9337868	1642
Greenhouses & rural buildings	2550718	271
Nurseries	273768	18
Woods	4015671	563
Winter cereals for forage	131442	14
Mixed herbages	501855	43
Silo maize	5407821	297
Alfalfa	825456	105
Mixed meadows	2630208	285
Pastures	1008300	89
Rice	866544	89
Maize	10001690	669
Soybean	1796109	194
Tree crops	1799739	215
Tomato	364185	20
Winter cereals	2707555	255

Tab. 11 - The 17 classes selected from the original dataset

The total area and the number of fields per class were calculated on the dataset before and after the cleaning procedure described (Figs. 17-18).

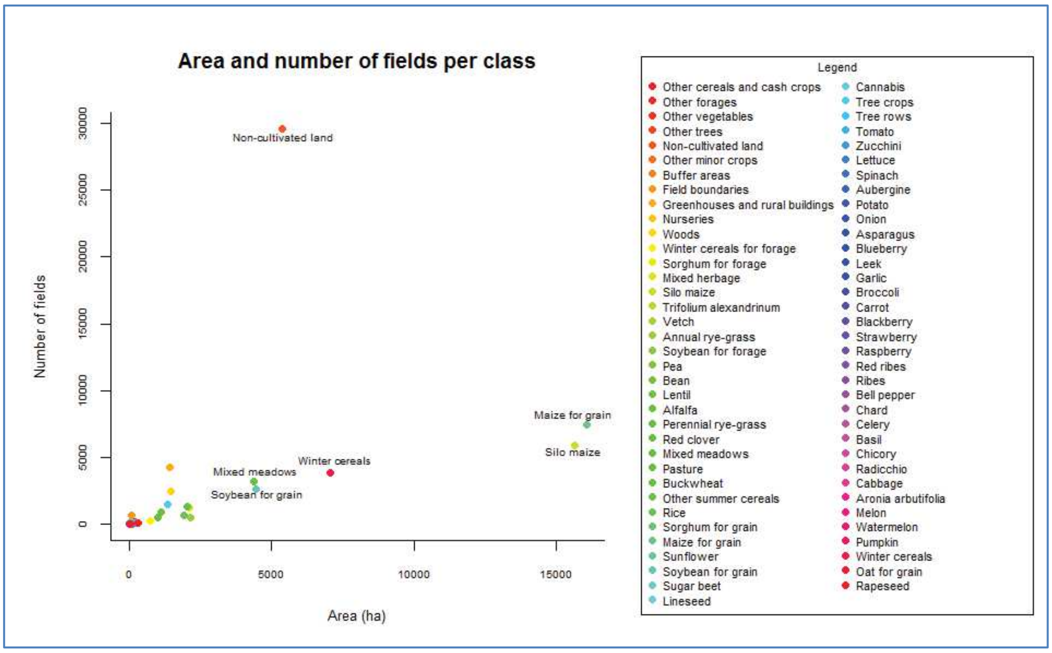


Fig. 17 - Area and number of fields per class; calculation performed on the dataset before the cleaning procedure

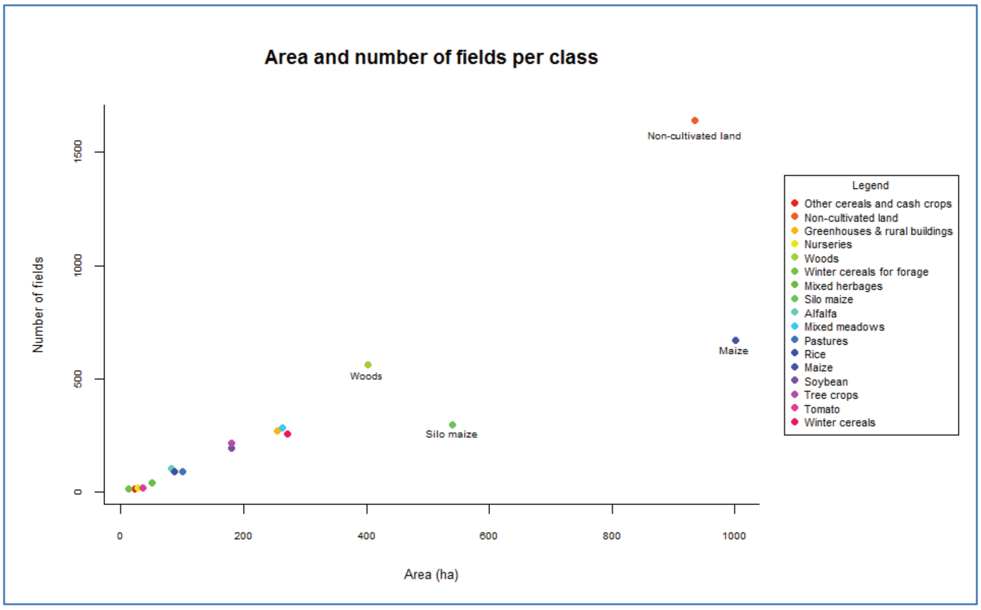


Fig. 18 - The area and number of fields of the 17 classes selected

It is worth noticing that both before and after the cleaning of the dataset, the classes “Maize for grain” and “Silo maize” are among the 4 most important classes for area and number of fields, thus confirming that the agriculture in the area is quite industrial and livestock-oriented; the result is also consistent with the observations of (Azar et al., 2016) and (Villa et al., 2015). The position of “Non-cultivated land” is as well remarkable: the class aggregates land not under production with fallow land, so it is not possible to quantify the problem of farmland abandonment.

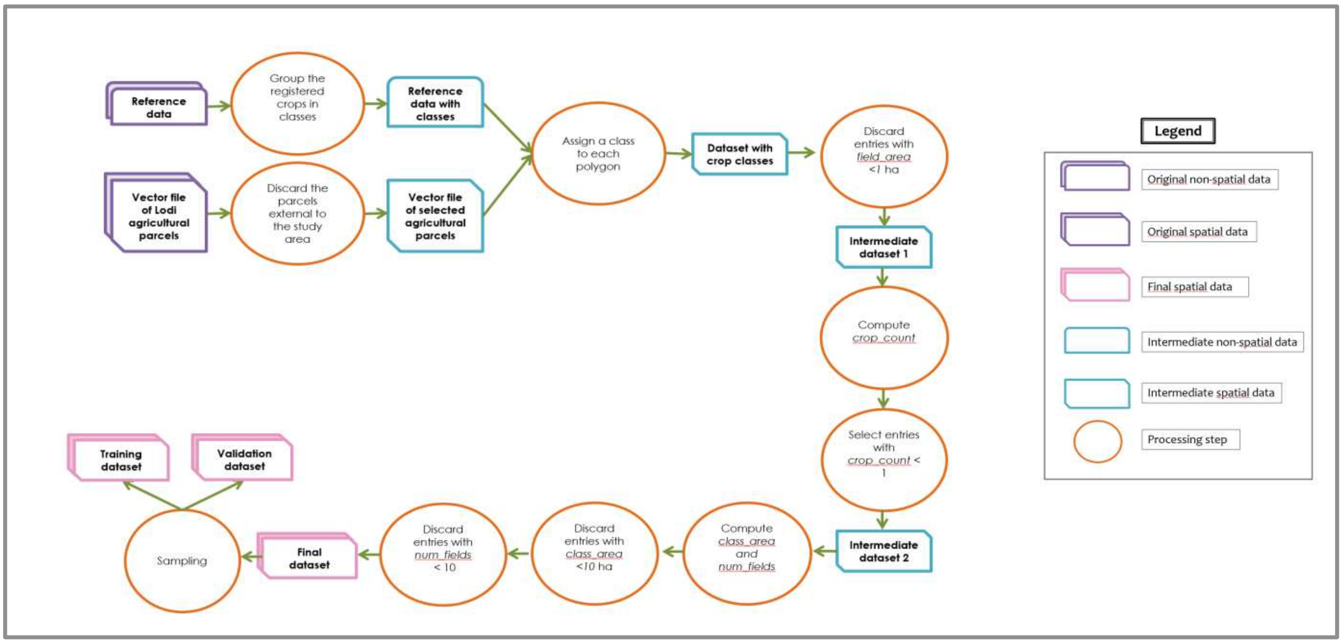


Fig. 19 – Workflow of the procedure used to select the training and the validation dataset starting from the original reference data

3.2.6 Classification

Given the availability of training and validation data, a supervised classification was performed: the study area was classified with the Random Forest (RF) classifier (Breiman, 2001), given its several advantages and the better performances in comparison to other classification algorithms. The RF models were built in R (R Core Team, 2016) with the package “randomForest”(Liaw & Wiener, 2002), using the training and validation datasets as described above.

To compare the performance of the single VIs and of a combination of VIs, different Random Forest models were built: some based on the PMs derived from single VIs; one based on the PMs derived from multiple VIs. Only the VIs and the PMs selected in the previous step, with means of the visual and statistical analyses, were used.

It was not possible to train all the models for all the selected classes: indeed, the PMs were not calculated over all the pixels of the study area, but only on the ones that had a positive quality flag. This decreased the number of pixels and polygons used for the construction of the model, leading also to a slight reduction in the number of classes.

The accuracy of the models was evaluated with a confusion matrix. Omission and Commission Errors were considered in relation to the single classes, while the Overall Accuracy was used to compare the performance of the models: the model yielding the highest Overall Accuracy was used for the classification of the study area, which was done employing the R package “raster” (Hijmans, 2016).

3.2.7 Integration of the classified map in the agro-ecological zonation

Land cover information can be integrated into agro-ecological zonations in two ways:

- As an input variable (van Beek et al., 2016);
- As a descriptor layer (Mücher et al., 2016).

Applying the approach of GAES, the crop map was integrated into the chosen AEZ in the form of a descriptor layer. The LoPCZ was used in this step, since it has a good level of detail, it is built on both climatic and pedological features and it was specifically conceived for the issue of nitrate pollution. The integration was done with these 3 steps:

1. Intersection: the classified map was divided into smaller areas corresponding to the extension of the existing agro-ecological zones;
2. Overlay: the classified map and the agro-ecological zonation were superimposed;
3. Computation of statistics: the percentage cover of each crop type in each zone was calculated and kept as a descriptor of the zone.

The procedure allowed to create an agro-ecological map enriched with crop type information.

Chapter 4

Results and discussion

4. Results and discussion

4.1 Results

4.1.1 Visual assessment of the performance of the VIs

The NDVI, EVI, SAVI and MSAVI were visually tested; Fig. 20 shows the temporal profile of 4 indices sampled on a random pixel of a summer crop.

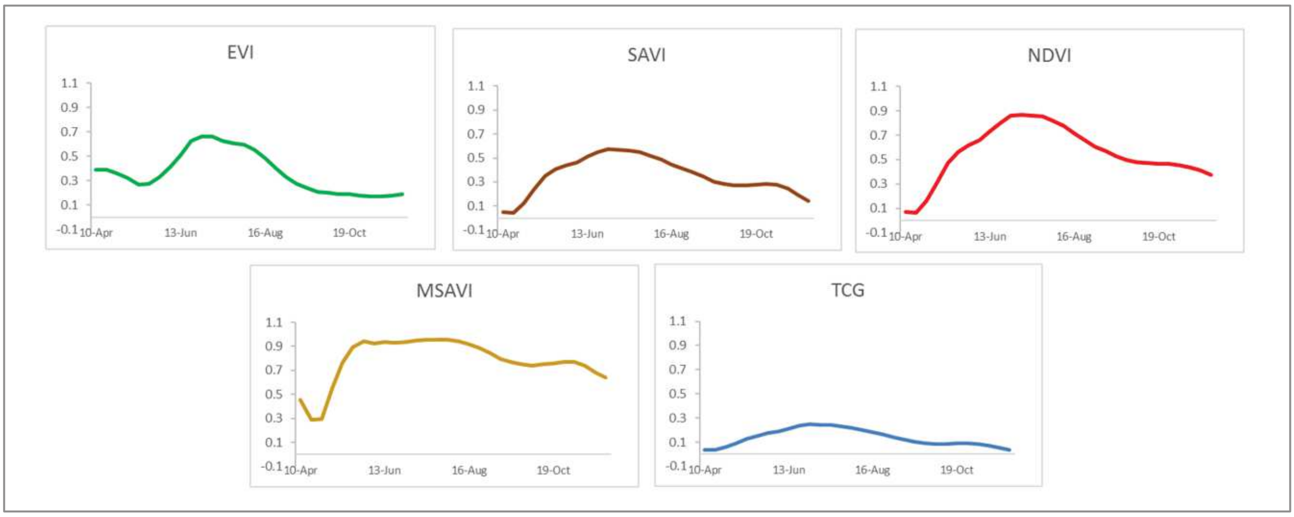


Fig. 20 – Sample temporal profiles of EVI, SAVI, NDVI, MSAVI and TCG

It can be seen that the TCG is extremely smooth, showing a very little amplitude. On the contrary, the MSAVI is quite rough (see for instance the rapid increase around the 12th of May). Moreover, it had in some cases a behaviour difficult to interpret: Fig. 21 shows the case of a pixel in which the peak of the MSAVI is very shifted in comparison to the peak of the other indices.

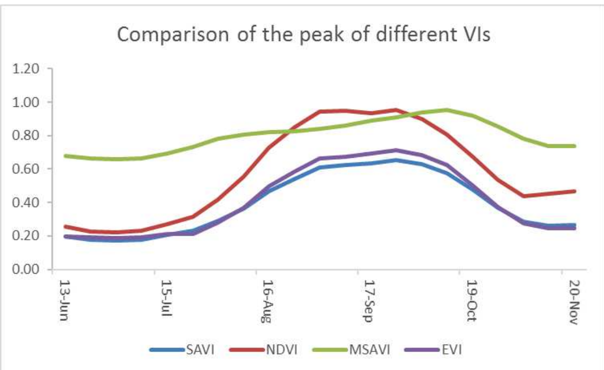


Fig. 21 - Comparison of the peak of different VIs

The three indices that kept the fluctuations in the signal without exaggerating or smoothing them were the NDVI, the EVI and the SAVI. The sensitivity of the 3 indices to phenological changes was similar, as demonstrated by the case of maize and silo maize: all the VIs presented a higher decreasing rate for silo maize than for maize (Fig. 22). This is due to the fact that silo maize is harvested when still green (at the kernel dough stage), while maize for grain production remains on the field for a longer period and it is harvested when already dry.

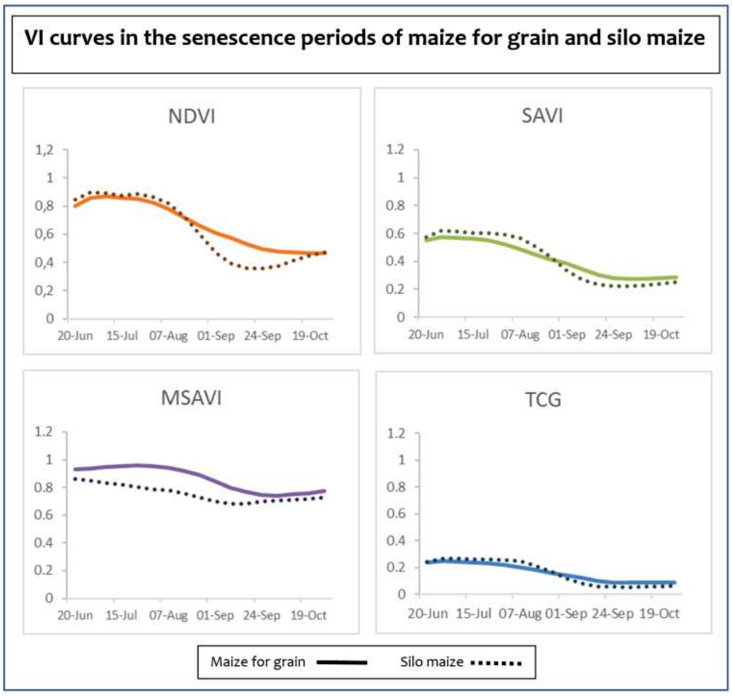


Fig. 22 - Senescence rate of NDVI, SAVI, MSAVI and TCG for maize and silo maize

Among the three indices, the NDVI has always got higher values than the other two.

The behaviour of the EVI is interesting: usually it follows the trend of the other two indices, but in some cases it shows differences. For instance, considering the case of a random temporal profile of a mixed meadow (Fig. 23), it can be seen that it completely misses the second growth cycle and the second cut of the year, between May and June, while the NDVI and the SAVI are able to detect them.

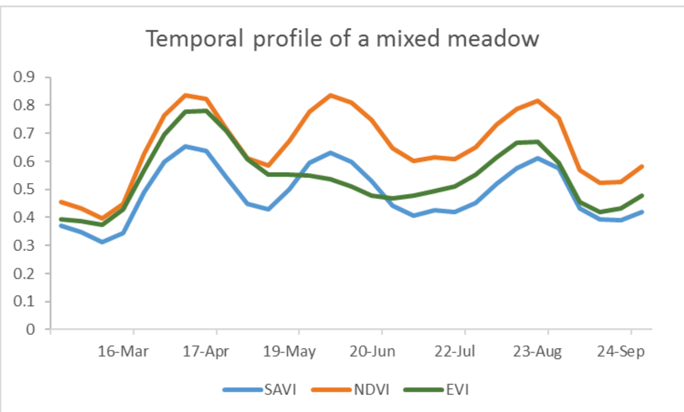


Fig. 23 - The temporal profile of a mixed meadow, calculated with NDVI, SAVI and EVI

Finally, the SAVI behaves similarly to the other indices in the case of medium-high levels of vegetation cover (e.g. temporal profiles of maize or soybean); its capacity to represent the crop growing cycle is more apparent on pixels which do not have a good and uniform vegetation layer, for instance in the case of non-cultivated land.

4.1.2 Assessment of the performance of the PMs

The PMs were calculated over the entire study area. Fig. 24 shows, as an example, the Duration Of Season 1 in Muzza, as calculated with EVI.

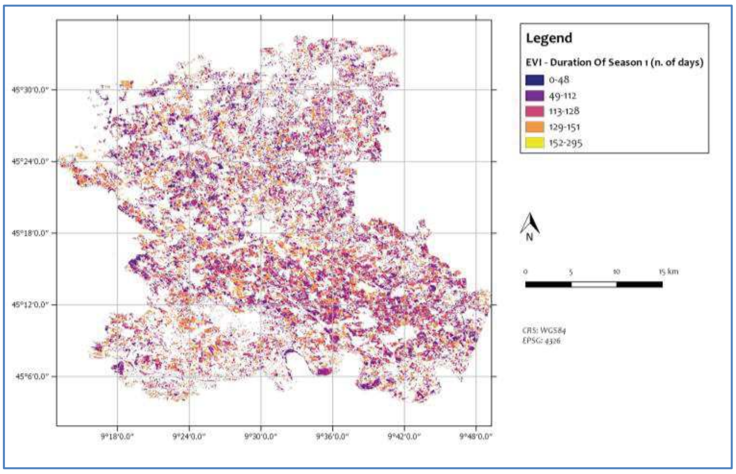


Fig. 24 - The Duration Of Season 1 in Muzza, as calculated with EVI

The PMs were evaluated according to different criteria: their variability, held as an indicator of their ability to distinguish among crops; their meaning in relation the biophysical characteristics of the crops; their detection in relation to other PMs.

In particular, the following observations were made about the calculated PMs:

1. VI value at the Start Of Season (SOS), at the Start of the Growing Season (SGS), at the End Of the Growing Season (EGS) and at the End Of Season (EOS): the intra field variability was found to be quite high. For instance, within one field, the EVI value at the SOS can vary up to 0.74. The SAVI value at the SGS within a single field was found to vary up to 0.64. Similar observations are valid for the NDVI, which has got a maximum intra-field variability of 0.84 at the EGS and of 0.11 at the EOS.
2. Dates of the Start Of Season, Start of Growing Season, End Of Season, End of Growing Season: in many cases, the Start Of Season and the Start of the Growing Season were detected on the same date; the same was observed for the End Of Season and the End of the Growing Season. These metrics are essential to distinguish summer from winter crops.
3. Dates of the Maximum Growth Rate (MGR) and Maximum Senescence Rate (MSR): much intra-field variability was observed for these metrics as well. The range of days in which the MGR and the MSR are detected were calculated per field. Tab. 12 shows the minimum and maximum range found across all the fields, according to the different indices: the NDVI is the one showing highest variability, but also the EVI and the SAVI are very variable.
4. Number of Growing Seasons: it helps in distinguishing fields which have one crop throughout the year, from fields which have a crop succession.
5. VI value in the peak date: it helps in distinguishing crops that produce very high amounts of biomass from crops that do not produce much.

	EVI	NDVI	SAVI
Mean intra-field difference in the detection of the MGR (days)	73	102	84
Mean intra-field difference in the detection of the MSR (days)	74	85	80

Tab. 12 - Maximum and minimum intra-field variability in the detection of the date in which the Maximum Growth Rate (MGR) and the Maximum Senescence Rate (MSR) occur, according to EVI, NDVI and SAVI

As for the remaining metrics, they were found to be quite consistent with the expected crop characteristics.

Following the observations made about the PMs, the following 8 metrics were excluded from the analysis for the explained reasons:

- VI value at the SOS, SGS, EOS and EGS: the information contributed by these metrics was not fundamental, since it was quite variable and very dependent from the correct detection of the date;
- date of the SOS and of the EOS: since the SOS and the SGS were found to be the same in several cases, the SOS was excluded from the analysis; the same reasoning was applied for the EOS;
- date of the MGR and of the MSR: this information was not considered extremely important, because quite variable.

The remaining 11 metrics were kept (Tab. 18), since no negative observations were made.

PMs kept for the classification
Number of Seasons
Start of Growing Season
End of Growing Season
Date of the peak
VI value at the peak
Amplitude
Maximum Growth Rate
Maximum Senescence Rate
Duration of Season
Length of Maturity Plateau
Seasonal Time Integrated Index

Tab. 18 - PMs used in the classification

4.1.3 Statistical analysis of VIs and PMs

The first metric to be checked was the Number of Growing Seasons (NGS) detected in each pixel. The correct detection of the NGS is fundamental, because if the growing cycle of a crop is split into more than one growing cycle, all the PMs will be affected. In turn, the quantification of the NGS is related to the detection of the Start and End of the Growing Seasons. The pixels investigated are supposed to have only one crop, therefore one growing cycle. All the indices miscalculated the NGS in part of the pixels; however, the EVI was the one misclassifying the least number of pixels, while NDVI and SAVI gave worse results, especially for what concerns some specific classes (e.g. tomato).

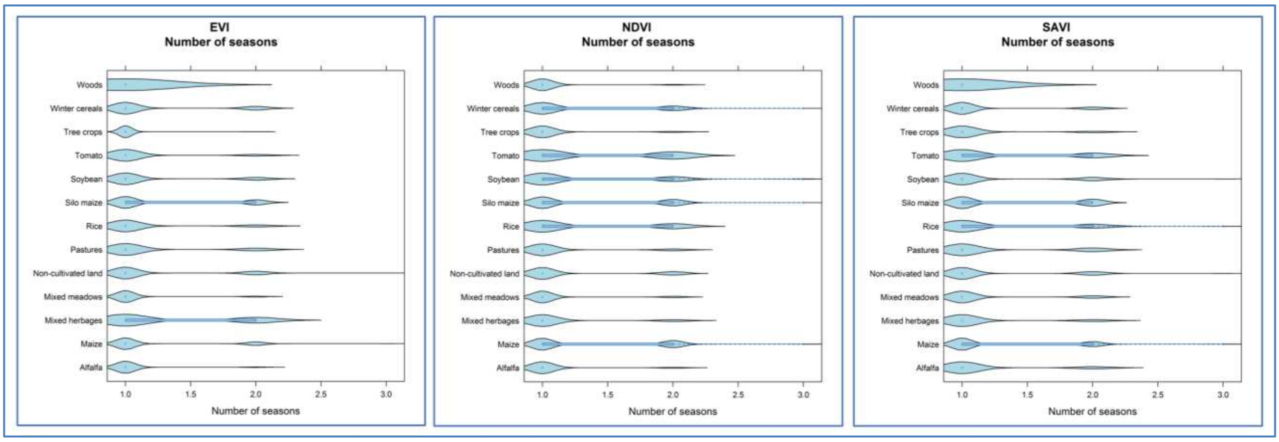


Fig. 25 - Number of growing seasons, per class, as detected by EVI, NDVI and SAVI

The statistical distribution of the peak value shows that the NDVI tends to have higher frequencies around precise values, slightly different for each class; the EVI and the SAVI, instead, do not have such a clear trend (Fig. 26).

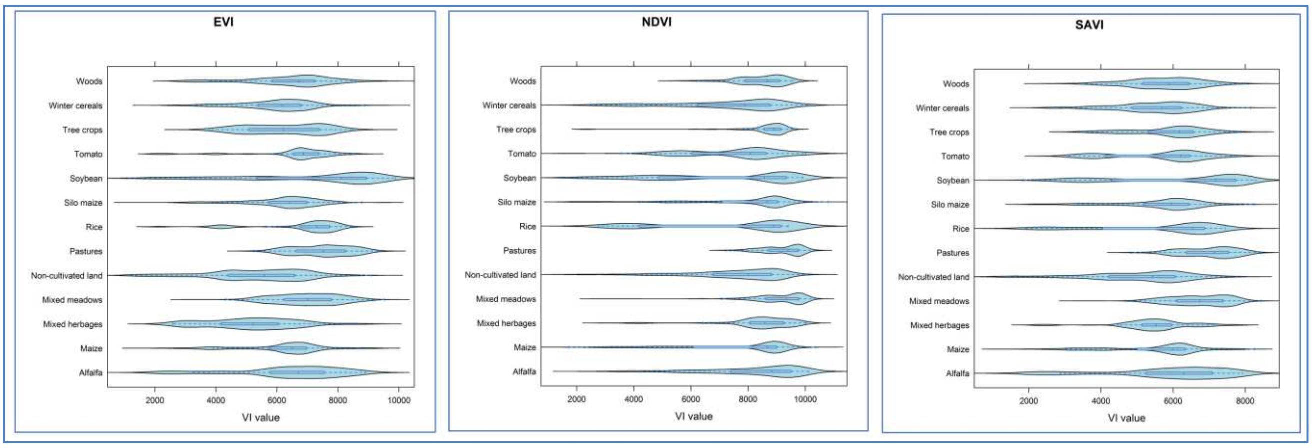


Fig. 26 - Peak values of season 1 calculated with EVI, NDVI and SAVI, per class; all the VIs are on a scale factor of 10000

This is particularly true for some classes, for instance tree crops (Fig. 27). However, when looking at classes with similar amounts of maximum biomass, the NDVI does not seem to differentiate well: for instance, the statistical distributions of the peak value for soybean and for maize or silo-maize are very similar when calculated with the NDVI; the EVI and the SAVI, on the other hand, clearly show a higher peak value for soybean in comparison to maize and silo-maize (Fig. 28).

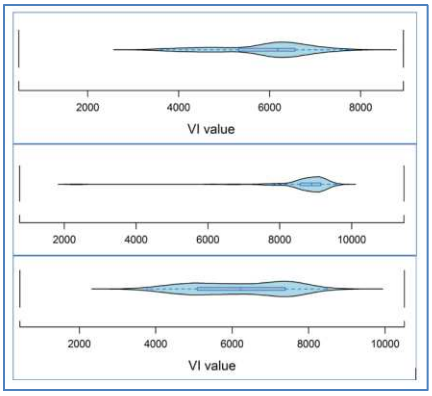


Fig. 27 - The Peak Value for the Tree Crops class, according to different VIs; from above: EVI, NDVI, SAVI. All the VIs are on a scale factor of 10000

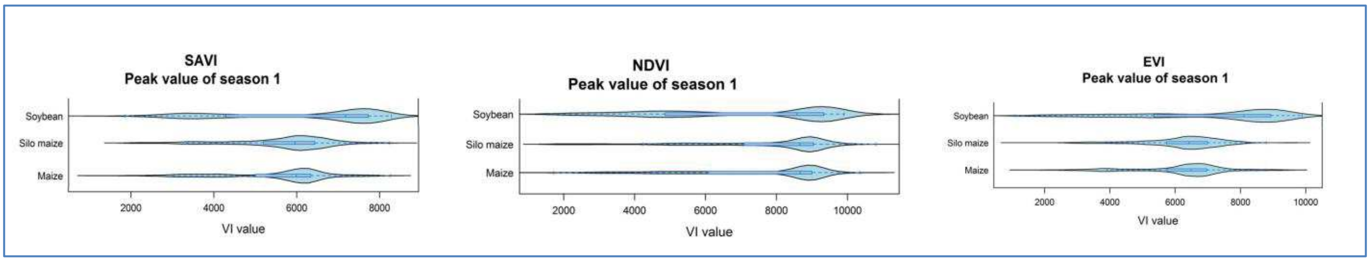


Fig. 28 - Comparison among the peak values of SAVI, NDVI and EVI, for the classes of soybean, silo maize and maize; all the VIs are on a scale factor of 1000

Finally, for some PMs there are no apparent differences among the VIs. When observing the statistical distribution of the Amplitude, for instance, the general trend of the three indices is similar, except for the NDVI showing a bi-modal distribution for some classes (Fig. 29).

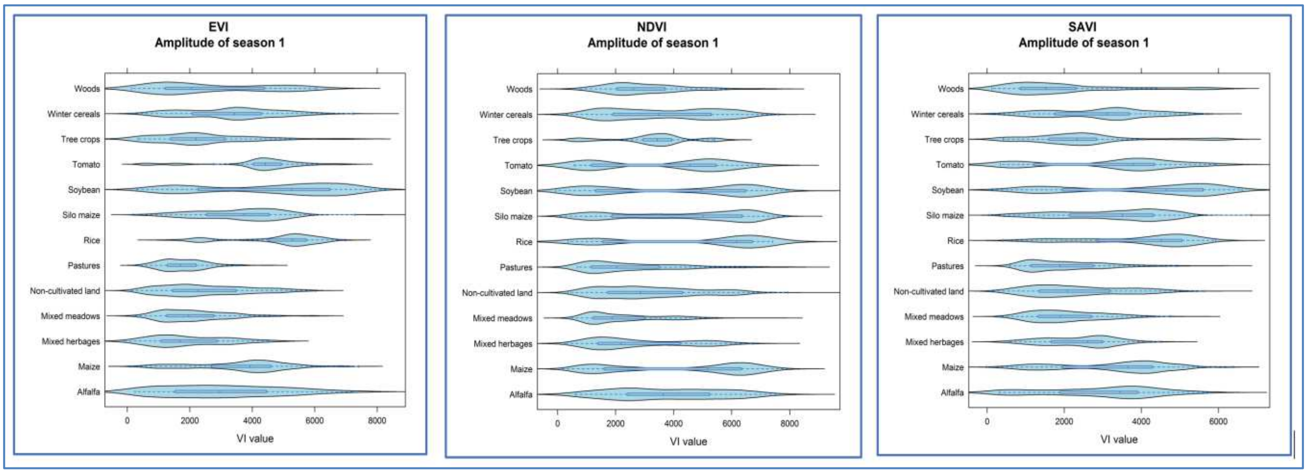
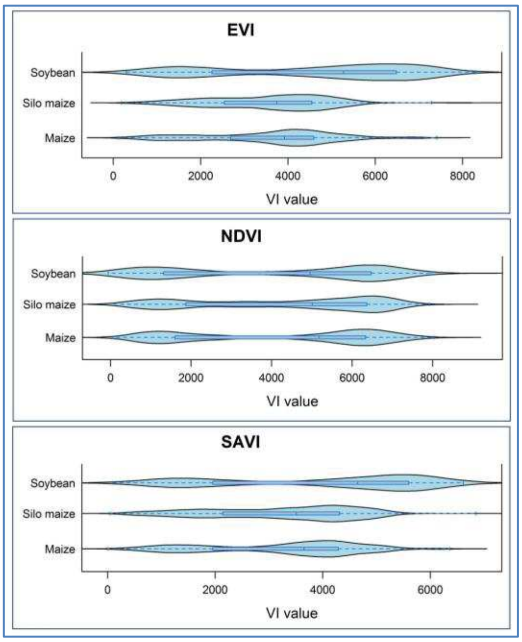


Fig. 29 - The amplitude of season 1, per class, as calculated with EVI, NDVI and SAVI



All the indices can distinguish between two macro-categories: on the one hand, there are herbages, meadows, pastures and non-cultivated land with a small amplitude; on the other hand, there are cash crops with a high amplitude. Distinctions within the macro-categories is more difficult, but some classes are distinguishable: soybean shows an amplitude higher than maize and silo-maize, when calculated both with the EVI and the SAVI.

Fig. 30 - The amplitude of soybean, maize and silo maize as calculated by EVI, NDVI and SAVI; all the VIs are on a 10000 scale

4.1.4 Segmentation and classification

After the intersection of the segmentation and the parcels and the selection of polygons and classes to use, 4 classes were discarded, leaving 13 classes for the classification.

Class	N. of fields	Area (m^2)
Tomato	12	266272
Mixed herbages	11	336282
Alfalfa	20	385383
Rice	16	443009
Pastures	20	555508
Tree crops	36	960139
Soybean	34	1070607
Mixed meadows	56	1275094
Winter cereals	62	1594442
Woods	72	1955710
Non-cultivated land	152	3447477
Silo maize	143	3910557
Maize	211	6599894
Other crops	32047	489623513

Tab. 13 - The final 14 classes used for the classification

The following 4 Random Forest models were built:

1. NDVI PMs-based RF;
2. EVI PMs-based RF;

3. SAVI PMs-based RF;
4. EVI, NDVI and SAVI PMs-based RF (from now on referred to as 'E-N-S PMs-based model').

The number of classes that each model could predict is reported in the table below.

	N. of classes included in the model
EVI	12
NDVI	11
SAVI	11
E-N-S	10

Tab. 14 - The number of classes employed in the construction of the 4 Random Forest models

The results were evaluated with the use of confusion matrices, reported in Appendix 8. The main findings are discussed here.

Tab. 15 shows that the best classified classes in terms of OE were "Other crops", "Tomato", "Soybean" and "Winter cereals"; the best classified classes in terms of CE were "Other crops" and "Pastures". In several cases, the OEs and CEs were as low as 0.

	EVI		NDVI		SAVI		E-N-S	
	OE	CE	OE	CE	OE	CE	OE	CE
Other crops	0.02	0.00	0.02	0.00	0.03	0.01	0.00	0.00
Tomato	0.14	0.00	0.22	0.00	-	-	-	-
Soybean	0.23	0.89	0.21	0.85	0.23	0.88	0.15	0.28
Winter cereals	0.21	0.70	0.21	0.57	0.23	0.81	0.11	0.07
Pastures	0.19	0.13	0.32	0.47	0.17	0.20	0.00	0.00

Tab. 15 - OEs and CEs of the best classified classes, across the 4 RF models tested

The classes with the highest OEs were "Non-cultivated land" and "Silo maize"; the classes with the highest CEs were "Woods", "Mixed herbages", "Soybean" and "Winter cereals" (Tab. 16). Generally, these classes show errors between 0.7 and 0.9, however in many cases the class is completely misclassified.

	EVI		NDVI		SAVI		E-N-S	
	OE	CE	OE	CE	OE	CE	OE	CE
Non-cultivated land	0.30	0.71	0.36	0.64	0.32	0.64	0.03	0.05
Silo maize	0.28	0.74	0.32	0.78	0.27	0.78	0.08	0.04
Woods	-	-	0.00	0.67	1.00	1.00	-	-
Mixed herbages	0.26	0.81	1.00	1.00	1.00	1.00	0.00	0.35
Soybean	0.23	0.89	0.21	0.85	0.23	0.88	0.15	0.28
Winter cereals	0.21	0.70	0.21	0.57	0.23	0.81	0.11	0.07

Tab. 16 - OEs and CEs of the worst classified classes, across the 4 RF models tested

Finally, the differences in OEs and CEs across classes were compared, with almost all the model having extremes of 0 and 1:

	OE			*CE*			*OA*
	Lowest	*Highest*	*Median*	*Lowest*	*Highest*	*Median*	*(%)*
NDVI PMs – based RF	0.00	1.00	0.22	0.00	1.00	0.64	98
EVI PMs – based RF	0.02	1.00	0.23	0.00	1.00	0.72	98
SAVI PMs – based RF	0.03	1.00	0.23	0.01	1.00	0.78	97
E-N-S PMs– based RF	0.00	0.15	0.05	0.00	0.65	0.04	93

Tab. 17 - Highest and lowest OEs and CEs of the 4 RF models tested

All the models presented both good and bad OEs and CEs, while the OAs were similar (97-98%): the only exception is represented by the E-N-S model, which shows a slightly lower OA (93%). The NDVI and the EVI PMs-based RF showed the highest OAs; to choose one model between these two, also the CEs and the OEs were considered, thus leading to the choice of the NDVI PMs-based RF, which was used to produce the final classified map (Fig. 31).

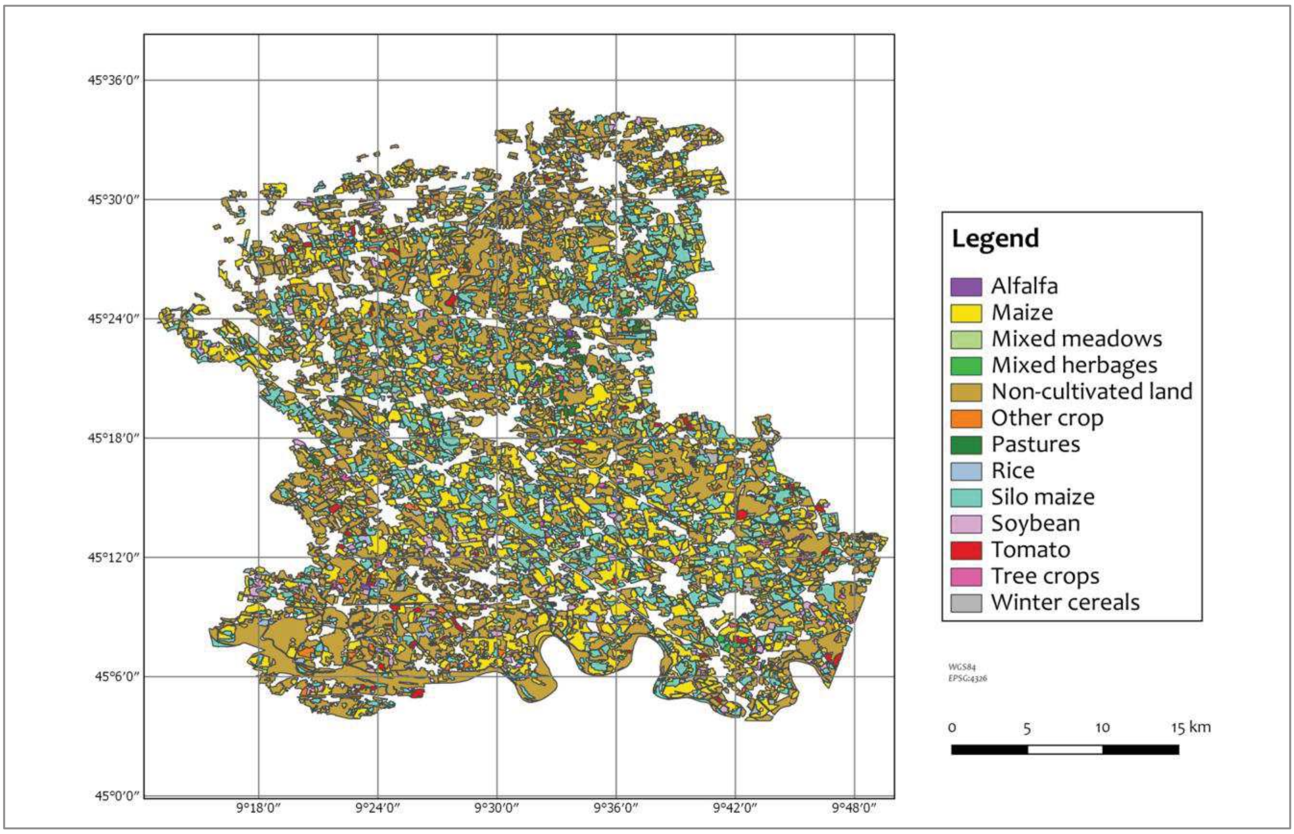


Fig. 31 - Classification of the study area, performed employing the NDVI PMs-based Random Forest classifier

According to the final classification, the greatest part of Muzza is covered by non-cultivated land, maize and silo maize. Other important crops are: mixed meadows, soybean, winter cereals and rice. Non-cultivated land seems to be uniformly distributed across the study area, as well as maize, while rice fields are concentrated in the South-West part of Muzza, between Lodi and Pavia.

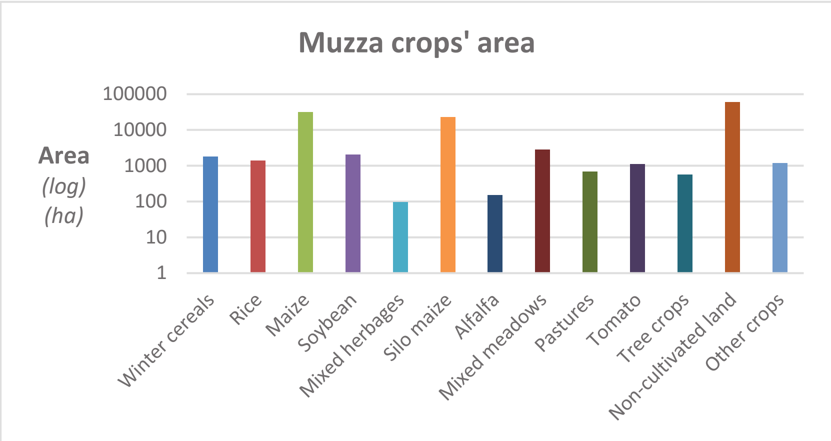


Fig. 32 - Area of the classified crops in Muzza

4.1.5 *Integration of the classified map into the agro-ecological zonation*

The classified map of the crop types in Muzza was integrated into the LoPCZ, as a descriptor layer (Fig. 33).

Fig. 33 - The two agro-ecological zones of Muzza, with the 5 most common crops found in each zone

According the LoPCZ, two agro-ecological zones can be found in Muzza: the Western Plain and the Central Plain. The vegetation cover in the two zones was characterized, by assessing the extension of the crops classified in the previous step, as shown in Figs. – and --. The Central Plain was found to have a higher percentage of non-cultivated land (54%), while in the Western Plain the cultivation of maize is more common (27%, against the 18% of the Central Plain). The distribution of crop requiring low doses of N fertilizer was analysed, aggregating the data of soybean, pastures, mixed meadows, alfalfa and mixed herbages. In both zones the percentage area covered by these crops is around 18%.

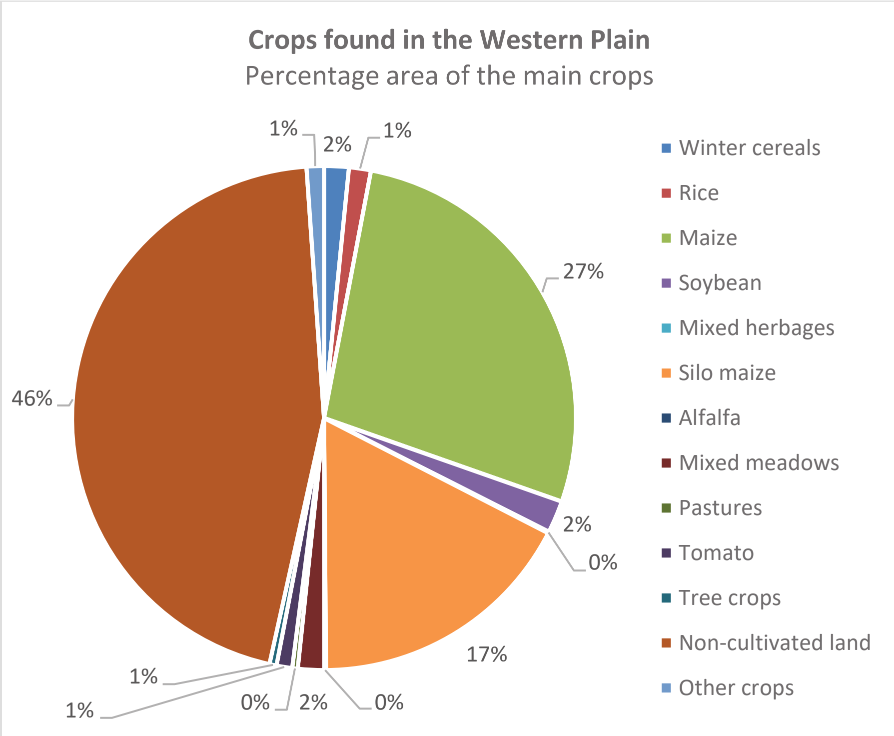


Fig. 34 - Crops found in the Western Plain zone of Muzza according to the classification performed: percentage area of the classified crops

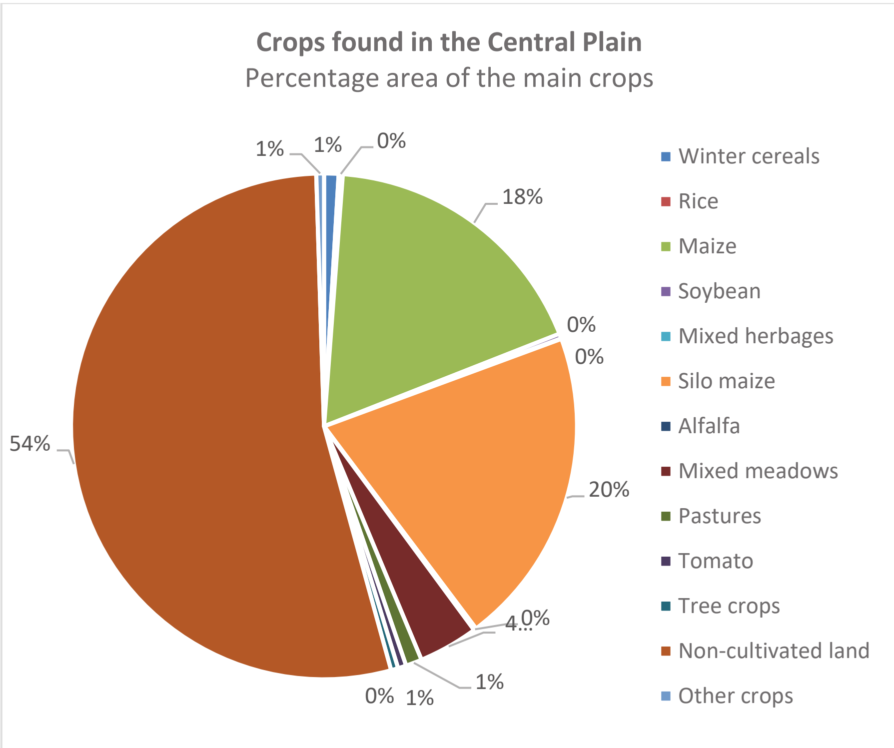


Fig. 35 - Crops found in the Central Plain zone of Muzza according to the classification performed: percentage area of the classified crops

4.2 Discussion

4.2.1 General analysis of VIs and PMs

The smooth behaviour of the TCG observed in this study was reported also by (M. Schultz et al., 2016), who observed over forested areas a rather stable signal with a slight change when deforestation occurred. However, the harvest of a crop is not comparable to deforestation, in terms of magnitude of change; moreover, in some cases crop residues are left on the field, thus making the change in the signal less evident than the change from dense vegetation to bare soil. Even if the same study reported a variable behaviour of the TCG in different ecosystem types, the TCG did not prove well in our case, neither on agricultural fields nor in woodland, being not sensitive enough to the small phenological changes.

The behaviour of the MSAVI can be explained with the presence of a soil parameter, *L*, which is dependent on the soil line, calculated in the NIR-Red space. A unique soil parameter was used to calculate the MSAVI over the whole area, because accounting for the soil variability would make the calculations too complex. However, (Aa. Vv., personal communication, 2016) have shown that Muzza has got at least two types of soil that are quite different in texture and therefore in moisture and spectral reflectance. The spatial variability in soil moisture could explain as well the behaviour of the MSAVI, since VIs are also influenced by this factor (Zhang, Zhang, Shi, & Huang, 2014). Moreover, the temporal interpolation may have not smoothed the normal spiky trend of the index.

The high values of NDVI in comparison to the EVI and the SAVI were to be expected, since the NDVI is known to saturate very quickly (Jackson et al., 2004) (Chen, Fedosejevs, Tiscareno-Lopez, & Arnold, 2006) (Asner, Scurlock, Hicke, Scurlockt, & Hicket, 2003). However, the NDVI was anyway able to detect the most interesting changes throughout the growing cycle, thus confirming the findings of (Vina et al., 2012), who observed that it is sensitive enough to distinguish between the senescence rate of two hybrids of maize.

The behaviour shown by the EVI over the sample meadow pixel is quite unexpected, since other authors found that EVI outperforms NDVI in the detection of meadows (Halabuk, Mojses, Halabuk, & David, 2015). Shen, Chen, Zhu, Tang, & Chen (2010) documented lower values of NDVI and EVI during the flowering period; meadows can be cut around the flowering stage, and this could explain the low value of the EVI, but not the difference between the EVI and the NDVI. In addition to this, the EVI is calculated using the Blue band, which is highly influenced by atmospheric aerosol and clouds: consequently, atmospheric effects can alter the actual EVI value, but they are difficult to account for. In conclusion, the behaviour of the EVI is in some cases difficult to predict and understand; however, the index shows satisfactory results in overall.

Finally, the capacity of the SAVI to represent the growth of wild plant on non-cultivated land can be quite important. Pixels of non-cultivated land represent a challenge for the classification, since rarely they are represented by bare soil, and most often they have a not well-characterized vegetation layer; thus, the LSP and the amount of biomass of these pixels are not similar to the ones of crops. This difference is difficult to detect with the NDVI, because it saturates quickly and does not distinguish between the high amount of biomass of most crops and the medium amount of biomass of the varied vegetation growing on non-cultivated land. Fig. 36 illustrates the behaviour of NDVI in comparison to EVI and SAVI on one pixel of non-cultivated land. The NDVI is the only index showing a regular fluctuation in the signal, more similar to the temporal profile of a crop with a short growing cycle (e.g. tomato). The EVI and the SAVI, instead, present a non-regular temporal profile, more coherent with the expected vegetation cover.

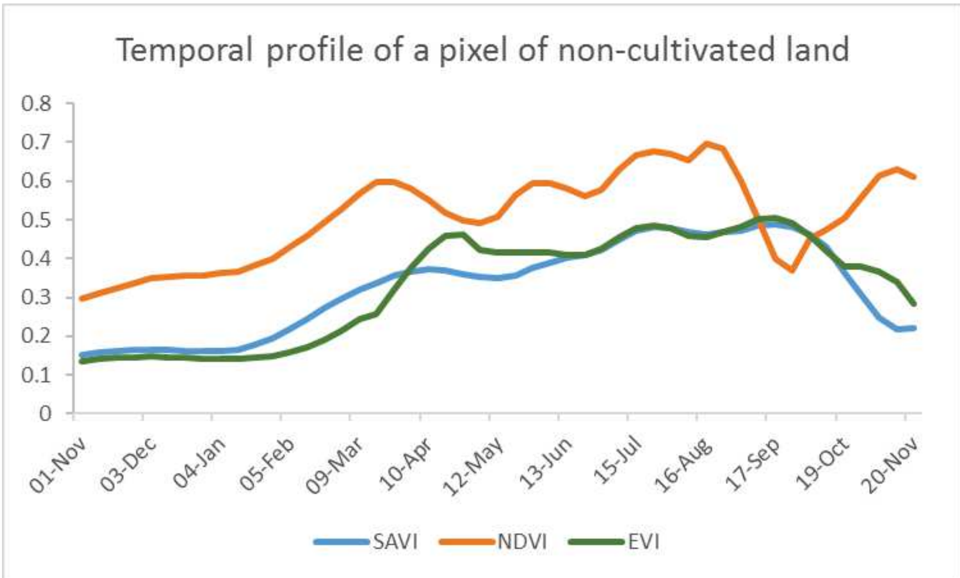


Fig. 36 - Temporal profile of EVI, NDVI and SAVI over a pixel of non-cultivated land

In conclusion, the TCG and the MSAVI were discarded because of their extreme behaviour. The NDVI, the SAVI and the EVI, instead, were included in the following steps of the analysis: even if they have some limitations, they have the potential of overcoming them by complementing each other. Some studies (e.g. (M. Schultz et al., 2016)) suggest that the integration of different VIs leads to a higher classification accuracy.

4.2.2 Statistical analysis of VIs and PMs

The behaviour shown by the NDVI in the detection of the peak, which tends to be concentrated towards high values, is due to the fact that it saturates more easily. This can be an advantage when classifying crops with very different amounts of biomass (e.g. tomato and tree crops), but it does not provide good results when working with crops with similar amounts of maximum biomass.

The difference in Amplitude observed for all the VIs between forages and cash crops is easily explained with the biological characteristics and the agronomic practices related to the two categories. The Amplitude is calculated as the difference between the VI value at the peak and the VI value at the Start of the Growing Season. Herbages have “normal” VI values at the SGS, but they show low peak values. Meadows and pastures have “normal” peak values but tend to have higher VI values at the SGS: this happens because the plants stay on the field for more than one year; they are periodically cut above the root collar, with a frequency of 1-2 months during spring and summe, and then they fast grow back to be cut again. The cultivation of cash crops, instead, follow different rules: before the sowing, the soil is bare, so the VI value at the SGS is low; the maximum amount of biomass of these crops is usually very high, so the VI value at the peak is high as well.

In conclusion, results show that there is not one index always performing better than the others, but rather that each index has a different performance for each phenological metric. This confirms the hypothesis that a combination of VIs might identify crops more accurately than a single VI.

4.2.4 Classification

The best classified classes in terms of OE were “Other crops”, “Tomato”, “Soybean” and “Winter cereals”; the best classified classes in terms of CE were “Other crops”, “Pastures” and “Maize”: this can be explained with features of these classes that are quite unique and not easily confusable. Indeed, winter cereals are almost the only crop, among the selected ones, that grows during the winter. Pastures show a typical oscillatory temporal profile. The temporal behaviour of pastures can also be observed in tomatoes, with some differences in the growth and senescence phases. Thus, also the results achieved for tomatoes are good: some vegetables (e.g. onions, carrots) are difficult to classify, due to their short growing cycle, their flexible calendar and their irregular temporal profile, as observed by Zhong, Hawkins, Biging, & Gong (2011); conversely, tomatoes stay on the field for a longer period because of their progressive ripening. Probably, this characteristic probably makes them more easily recognizable. Soybean shows good OEs but high CEs, meaning that all the fields with soybean were classified as such, but also some non-soybean fields were included in this class: this happened mostly with pixels of non-cultivated land, and in some cases with fields of silo maize, which is reasonable given the similarity of their growing cycles.

“Non-cultivated land” and “Silo maize” showed the highest OEs. Silo maize was mostly classified as “Other crops”, while misclassified pixels of non-cultivated land were in many cases labeled as “Other crops”. The classes with the highest CEs were “Woods”, “Mixed herbages”, “Soybean” and “Winter cereals”: in most cases, all these classes included pixels of non-cultivated land or of other crops.

It is clear that the classes of “Non-cultivated land” and of “Other crops” represent a disturbing element: the cause is most likely the absence of precise phenological features. Indeed, non-cultivated land may include: long-abandoned fields, in which little vegetation grows; fallow land, in which some seeds of previous crops germinated and grew to normal plants; fallow land, on which weeds grow; abandoned land on which natural vegetation grows. Of course, all of them show different Land Surface Phenologies. In addition to this, Prishchepov, Radeloff, Dubinin, & Alcantara (2012) observe that vegetation successions can vary considerably between marginal and non-marginal land and between arable land and grassland, thus producing different temporal profiles. The same is valid for the class “Other crops”, that gathers very different crops, from minor grains to vegetables, thus having much intra-class variability. It is likely for a pixel of sorghum, labeled as “Other crops”, to be classified as “Soybean”, since the two crops have similar growing cycles.

Finally, the differences in OEs and CEs across classes of the same model are noticeable: in the same model, there are always classes that are detected very well and others that show a very low degree of accuracy. This leads to the hypothesis that the characterization of the classes is as important as the classification algorithm. The characterization of the classes occurs through two steps:

1. The choice of the classification variables: for instance, the most important difference between maize and soybean may be the date in which the peak is reached, while the other phenological variables may be very similar; if the classification algorithm is based on all the phenological variables except the peak date, the two crops will be most likely confused. This could explain the poor results obtained by Azar et al. (2016) and Inglada et al. (2015), who employed the temporal profile of reflectances and VIs but no PMs. Hence, the variability in classification accuracy across the classes found in our study could be explained in this way: the variables chosen could characterize well some classes, but were not enough for other classes.
2. The choice of the training samples: both the quality and the size of the training sample are important, since they are related to the intra-class variability. In this study, the training and validation datasets were represented by the data registered by farmers, which are not always reliable. Moreover, it could be argued that the threshold of 10 fields per class, set in this study, was too low to represent the intra-class variability of some classes.

Summarizing, in the classification process some classes can be identified more easily than others, due to distinctive features of their growing cycles that were captured through the chosen variables. However, the accuracy in the detection of these classes can be affected by the presence of other, non-well-defined classes. The high intra-class variability of some classes negatively affects the whole classification, as demonstrated by the case of “Non-cultivated land” and “Other classes”.

Comparing the 4 RF models tested, other observations can be made. Among the models based on single VI PMs, the NDVI PMs - and the EVI PMs-based models performed comparably. The higher OE and CE median values of the EVI indicate that this index, in comparison to the NDVI, has a tendency to underestimate or overestimate the number of fields belonging to certain classes. Halabuk et al. (2015), working with MODIS-derived EVI and NDVI temporal profiles, observed similar results for the classification of cut and uncut meadows: the two indices yielded similar results, but the NDVI performed better. On the contrary, in our case the EVI gave better results than the NDVI for the classification of pastures and mixed meadows, but had slightly worse results considering all the classes. The reasons could be several: in the first place, the spatial and temporal resolutions of the two sensors are different and may affect significantly the results; in the second place, even if in both cases a temporal approach is used, the variables employed are not the same; finally, the different pre-processing steps may influence the temporal profiles tested. This comparison demonstrates that when dealing with classification, the pre-processing steps and the variables employed are as important as the general approach.

The SAVI PMs-based RF shows OEs and CEs higher than those of EVI and NDVI. Unfortunately, no studies were found in which the SAVI was used to monitor crop phenology, thus no comparison is possible. The only possible hypothesis is that the soil component in the index negatively affects its performance in the PMs detection. Indeed, the index may provide good results when the soil characteristics are well-defined, for instance at the field level; but when working over larger areas, soil variability is inevitable and difficult to account for. Thus, the soil variability of the study area may have altered the performance of the index.

Finally, the performance of the E-N-S PMs-based RF may appear strange, since OEs and CEs are very low, but the OA is the lowest among the 4 models. There may be two reasons explaining it: in the first place, the E-N-S PMs-based model was built on a low number of training fields, since the pixels needed to have a positive quality flag for all the 3 VIs. So, the E-N-S PMs-based RF is built on a smaller training dataset then the other 3 models: it is known that the size and quality of the training dataset influences the classification, and this may be the case. Secondly, it is true that increasing the number of variables the classification improves: however, this is true only up to a certain limit, after which the classification accuracy does not increase nor decrease (Lebourgeois et al., 2017). These results suggest that rather than increasing the number of classification variables, a careful choice of the 3 VIs variables should have been done to reduce the redundancy of the information.

According to the final classification, the most common crops in the area are maize, silo maize, mixed meadows, soybean, winter cereals and rice; non-cultivated land is much spread on the territory as well. This data agrees with the statistical analysis performed on the reference data of the Province of Lodi.

4.2.5 Integration of the classified map into the agro-ecological zonation

It has already been shown that AEZs have been used for environmental purposes, for studying nutrient dynamics (Geurts & Berg, 1998) and for fertilizer recommendations (Smaling, 1993). The Nitrate Vulnerable Zones, as defined by the Nitrate Directive, can be considered themselves agro-ecological zones used for N fertilizer recommendations and limitations.

However, none of these zonations include crop type information. Following, the usefulness of integrating crop type information into agro-ecological zonations will be discussed.

From the analysis of the most common crops in the two pedo-climatic zones of Muzza, three major observations came to light:

- non-cultivated land: in the Central Plain the amount of non-cultivated land is higher;
- maize: more maize is cultivated in the Western Plain in comparison to the Central Plain;
- crops with low N requirements (soybean, pastures, mixed meadows, alfalfa, mixed herbages): their area was found to be low in both zones.

This crop type distribution on the territory may have several implications.

Firstly, non-cultivated land seems to be quite widespread. In the classification process, it was not possible to differentiate among abandoned land, fallow land and land covered by cover or catch crops. The presence of cover and catch crops prevents N leaching, that can lead to the aquifer pollution. However, not the same can be said about fallow and abandoned land. If the amount of fallow land is high, it would be advisable to encourage the use of cover and catch crops, to strengthen the efforts made to prevent N pollution. As for abandoned farmland, the situation is not as clear. It has been proved that farmland abandonment influences several environmental processes. (MacDonald et al., 2000) investigated the topic over several sites in Europe, showing the negative effects on biodiversity, landscape, soil quality and natural hazards. On the other hand, (Knops & Tilman, 2000) proved that the N stock in the soil tends to increase after farmland abandonment. So, it is not yet clear what are the effects of farmland abandonment on the quality of soils and on nutrient dynamics. However, the diffusion of non-cultivated soils in the Central Plain of Muzza suggests that more detailed assessments should be made, to avoid negative consequences on N pollution.

Secondly, the diffusion of maize in the Western Plain suggests that this zone is potentially more vulnerable to N pollution than the Central Plain. Indeed, maize requires high doses of N fertilizers (Baldoni & Giardini, 2001): depending on how farmers manage the fertilization process, the potential of N pollution may be very high. Indeed, the use of slow release fertilizers or the division of the fertilizer dose into 2-3 sub-doses distributed across the growing season increases the amount on N absorbed by the plant and decreases the amount of N leached through the soil. However, these practices are expensive and it cannot be assumed that all the farmers apply them. This case shows that the crop type helps in defining the N pollution risk across zones, but that having an additional layer of information regarding farm management could improve the zonation even more.

Thirdly, only a small share of land is dedicated to forage and legume cultivation, suggesting that not many farmers adopt legume-based rotations. It is acknowledged that legumes help in the maintenance of soil fertility (Baldoni & Giardini, 2001) and that their cultivation should regularly be included in rotations: indeed, their residual effects on soil fertility reduce the amount of N fertilizer needed for the next crop in the succession (Mayer, Buegger, Jensen, Schloter, & Heß, 2003). It can be hypothesized that farmers drop legumes cultivation because it's more profitable to cultivate cash crops and buy concentrate feeds for the livestock, instead of self-producing fresh forage. It is desirable that limitations imposed by policies also orient the farmers towards new farm management strategies, but this is not always the case (Macgregor & Warren, 2006) and apparently Muzza farmers did not respond in this way. Moreover, it has been argued that the limitations imposed by the Nitrate Directive may not serve the final scope (Belhouchette et al., 2011). Consequently, decision-makers could consider the option of encouraging the cultivation of more legumes, targeting the agro-ecological zones where the risk of N pollution is higher.

Finally, it must be underlined that to elaborate an efficient strategy for N pollution management, the crop type information should be considered in relation to pedo-climatic parameters. Indeed, variables like the total amount of precipitations, the distribution of precipitations across the growing season or the soil texture influence the movement of N fertilizers along the soil profile up to the aquifer. It is not possible to do such a global reasoning in the present research, since the pedo-climatic characteristics of the chosen zones are not available. Nonetheless, such an approach is advisable.

4.3 Conclusions

In this study, I investigated the usefulness of employing VI-derived Phenological Metrics for crop classification and the meaning of integrating satellite-derived crop type information into agro-ecological zonations. The main conclusions are as follows:

- The integration of crop type information into an AEZ represents a valuable improvement and can be contribute relevant territory information, especially if the interrelations among crop types and pedo-climatic parameters are considered.
- In multi-temporal analysis, the TCG and the MSAVI do not contribute significant information because of their extreme behaviour, which tends to smooth or exaggerate the changes occurring in vegetation. The NDVI, the EVI and the SAVI, instead, appear to keep the characteristics of the most important phenological stages.
- It is known that it's impossible to detect from satellite remote sensing all the phenological stages that are fundamental from an agronomic point of view. Among the PMs that can be detected, differences were found: some of them are quite class-specific, others do not differ much across classes; some are easily detectable, others are very difficult to define with precision. This suggests that more research on PMs is advisable, with the objective of finding easily-detectable PMs able to characterize the growing cycle of different crops.
- VIs perform differently in the detection of the PMs: EVI shows better results in determining the number of seasons, NDVI in the peak value; in the determination of the amplitude, EVI, NDVI and SAVI performed comparably.
- The accuracy in the classification of single crops depends on several factors:
  - classification variables, which should be able to catch the characteristic features of each crop: the peculiar growing cycle of tomato made its classification accurate;
  - variability in the real features of each crop: the different characteristics of the vegetation growing on non-cultivated land made this class not well-characterized and therefore detectable;
  - intra-class variability: the high values of the classes "Other crops" and "Non-cultivated land" affected negatively the Omission Errors of several other classes;
  - quality and quantity of the training samples: the nature of the dataset used, which wasn't validated on the field, may have affected negatively the results.
- The Random Forest classification performed with NDVI and EVI PMs gave similar results, with the former having slightly better results than the latter; the outcome of the SAVI PMs-based classification, instead, was less satisfactory.
- The classification of the study area, overlayed with an agro-ecological zonation, shows that crop types differ with agro-ecological zones.
- Crop type differences across agro-ecological zones suggest that the strategies to handle N pollution deriving from agricultural activities could be diversified.

Conclusions

Through this research, a brief review of AEZs has been done. We have seen that existing global AEZs can differ much in input variables, spatial resolution and zone characteristics. These differences can make one or the other AEZ more suitable for a specific purpose, but they have one fundamental characteristics in common: almost all of them are built on climatic and pedological variables only. This is also true for the LoPCZ (Lombardy Pedo-Climatic Zonation), the agro-ecological zonation adopted by the Region Lombardy to define the Nitrate Vulnerable Zones and control N pollution deriving from agricultural activities.

Indeed, a pedo-climatic AEZ can help in detecting areas which are more vulnerable to N pollution (assuming that the land cover is the same). But observations about the most common crop types in each zone can trigger research and open the way to new inquiries: are crops with high N requirements mostly cultivated in highly vulnerable areas? Is it possible to relieve the environmental pressure by changing the cultivated crops? Is it advisable to re-define the Nitrate Vulnerable Zones, taking into account the effect of the mostly cultivated crops in each area? An AEZ with crop type information is the ideal tool to help experts deal with these questions. Consequently, a crop type map of Muzza was produced and later integrated into the LoPCZ.

The crop type map, referred to the crop year 2014-2015, was elaborated with a multi-temporal approach. Landsat8 imagery was used to calculate the temporal profile of 3 Vegetation Indices (NDVI, EVI and SAVI) and these, in turn, were used to calculate 11 Phenological Metrics (PMs). PMs are metrics able to characterize the growing cycle of the crops by detecting some key stages: for instance, the moment in which the plant reaches the maximum amount of biomass or the moment in which senescence begins. The PMs were used as the input variables for the crop type classification. The key assumption is that even crops with similar growing cycles have got some phenological differences and that the PMs are able to capture them. Using the RandomForest algorithm and different combinations of variables, 4 classifications were performed: 3 using the PMs derived from the single VIs and one combining the PMs derived from all the VIs. The 4 produced classifications yielded high Overall Accuracies, thus supporting the hypothesis that VI-derived PMs are able to characterize the crops growing cycles. In addition to this, the potential of remote sensing for crop type mapping purposes was confirmed. The classification yielding the highest Overall Accuracy was the one produced with the NDVI PMs: this was used for the integration in the LoPCZ.

The integration of the crop type map into the LoPCZ opened new perspectives on the N pollution problem, revealing significant differences between the zones in terms of crop type. Maize, which requires high N fertilization, is more common in one zone, while the other has got more fallow or abandoned land. In both areas, the cultivation of legumes and in general of plants that require low N doses is not very common. These observations, related to the pedo-climatic characteristics of each zone, can potentially serve for multiple purposes: determining the amount on N fertilizer applied in total in each zone; identifying zones with higher risk of N pollution; guiding policy-makers in the choice of the cropping systems to encourage.

In conclusion, this research showed that systems meant to monitor and control N pollution deriving from agricultural activities may greatly benefit from the use of AEZs that include EO-derived crop type information.

Appendix 1 – GYGA zonation of Italy

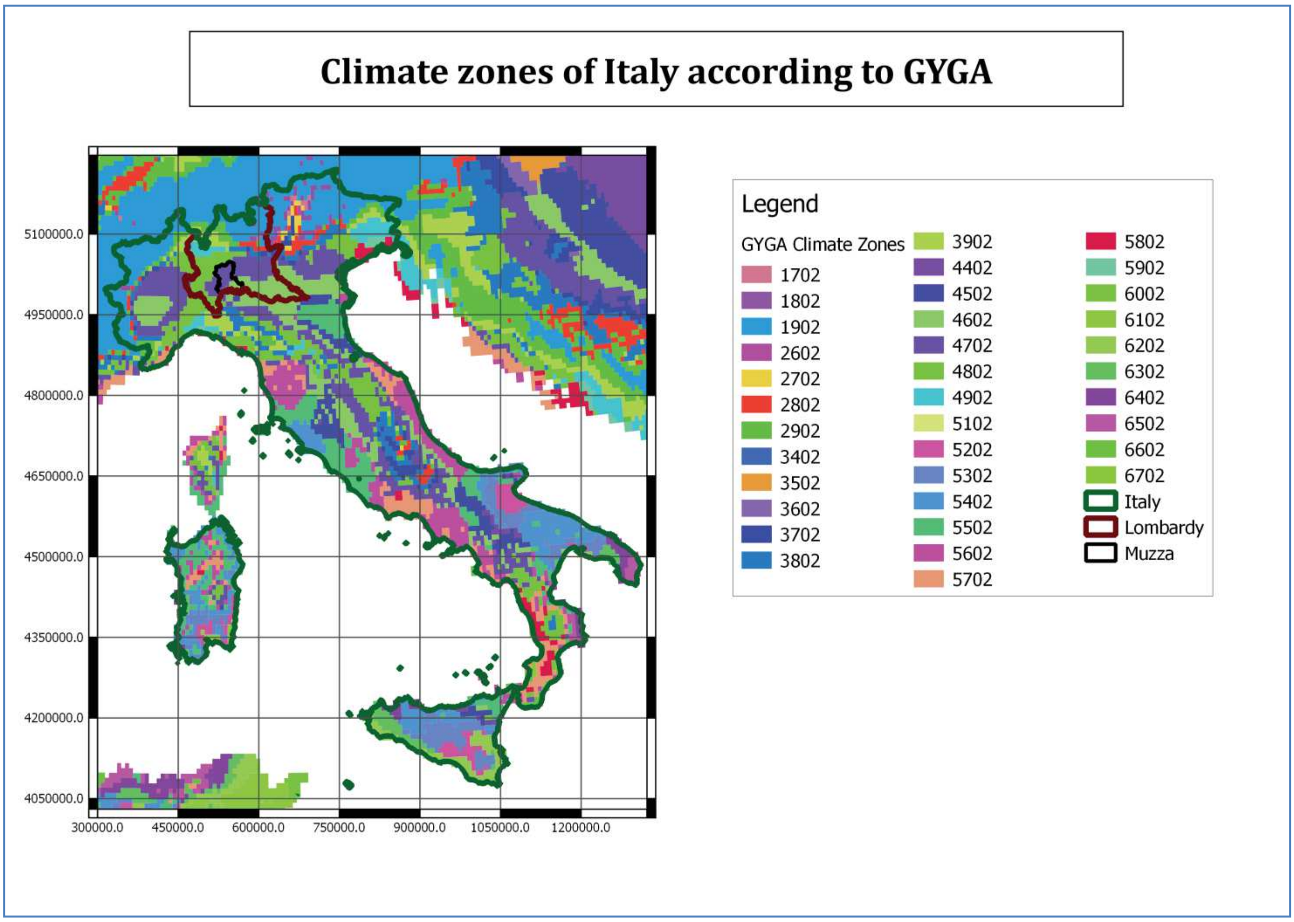


Fig. 37 - Climate zones of Italy according to GYGA

Appendix 2 – GAES zonation of Italy

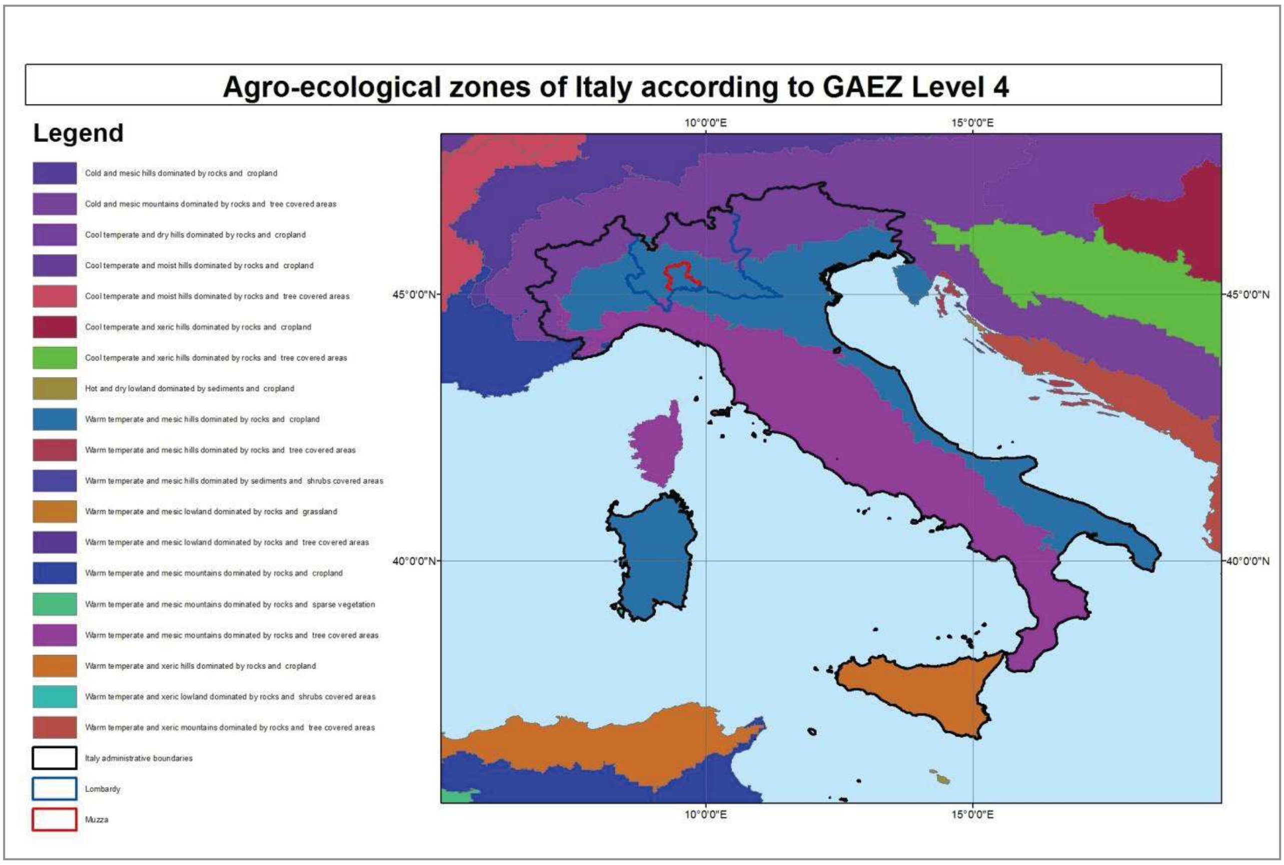


Fig. 38 - Agro-ecological zones of Italy according to GAES

Appendix 3 – GEnS zonation of Italy

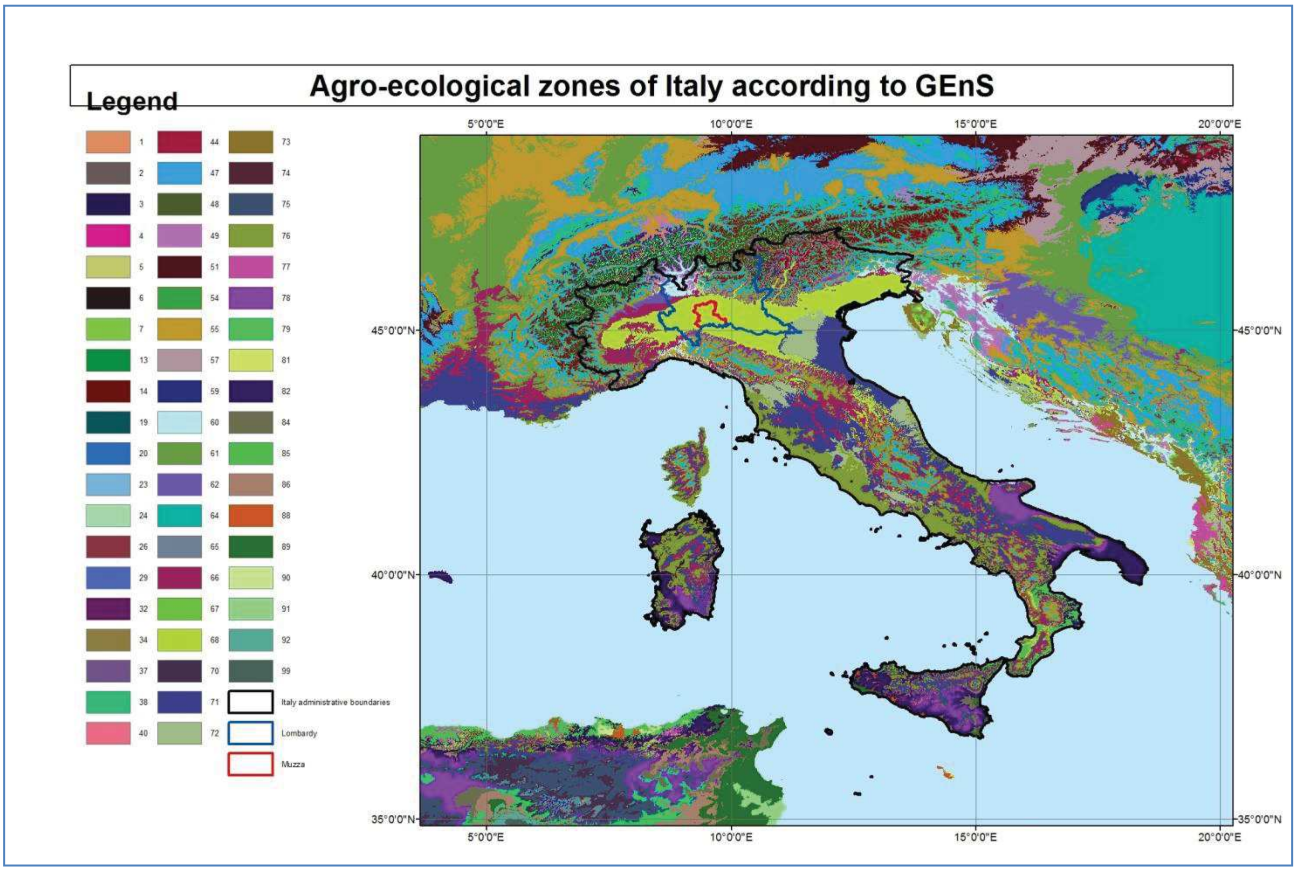


Fig. 39 - Agro-ecological zones of Italy according to GEnS

Appendix 4 – HCAEZ zonation of Italy

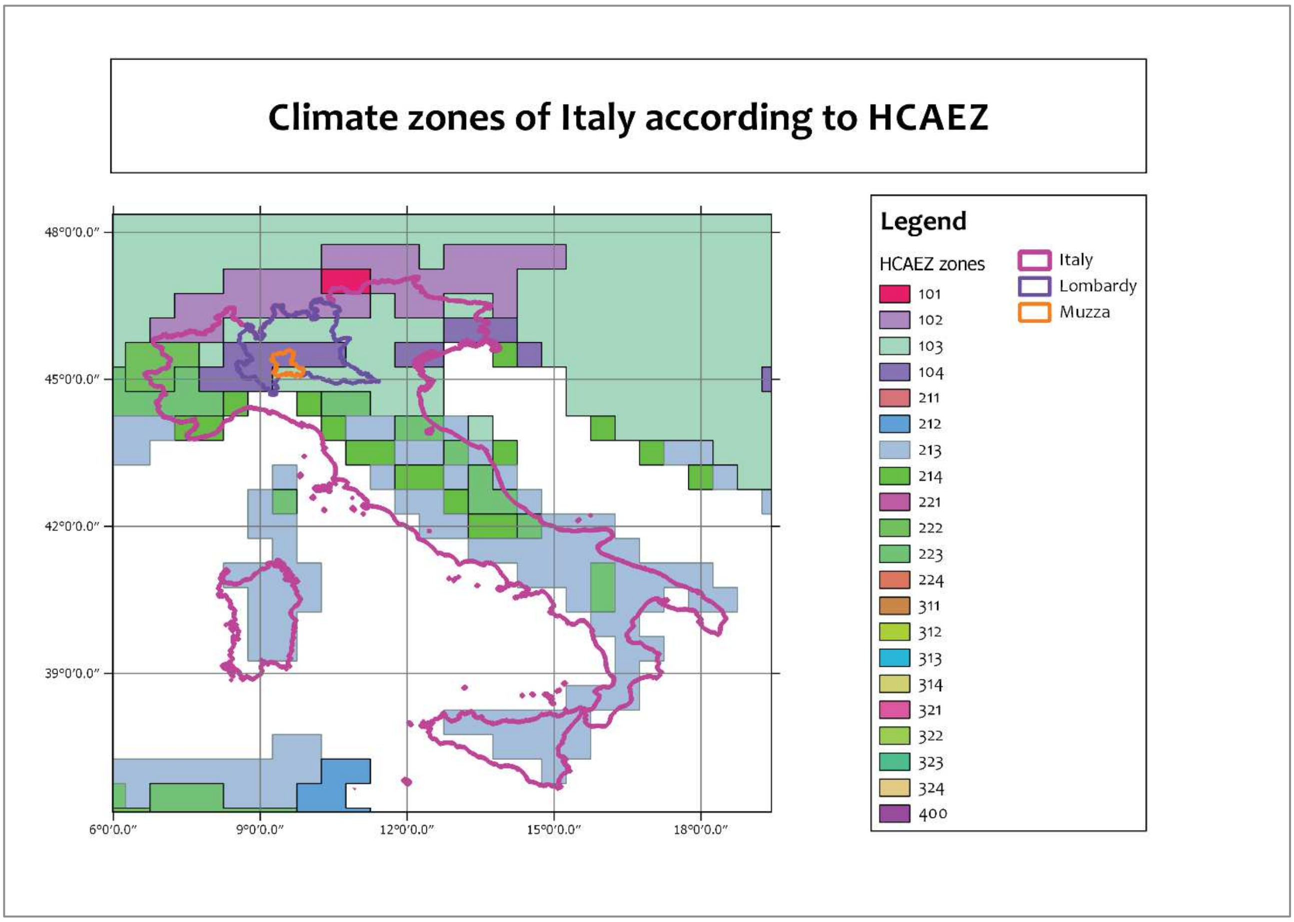


Fig. 40 - Climate zones of Italy according to HCAEZ

Appendix 5 – LoPCZ

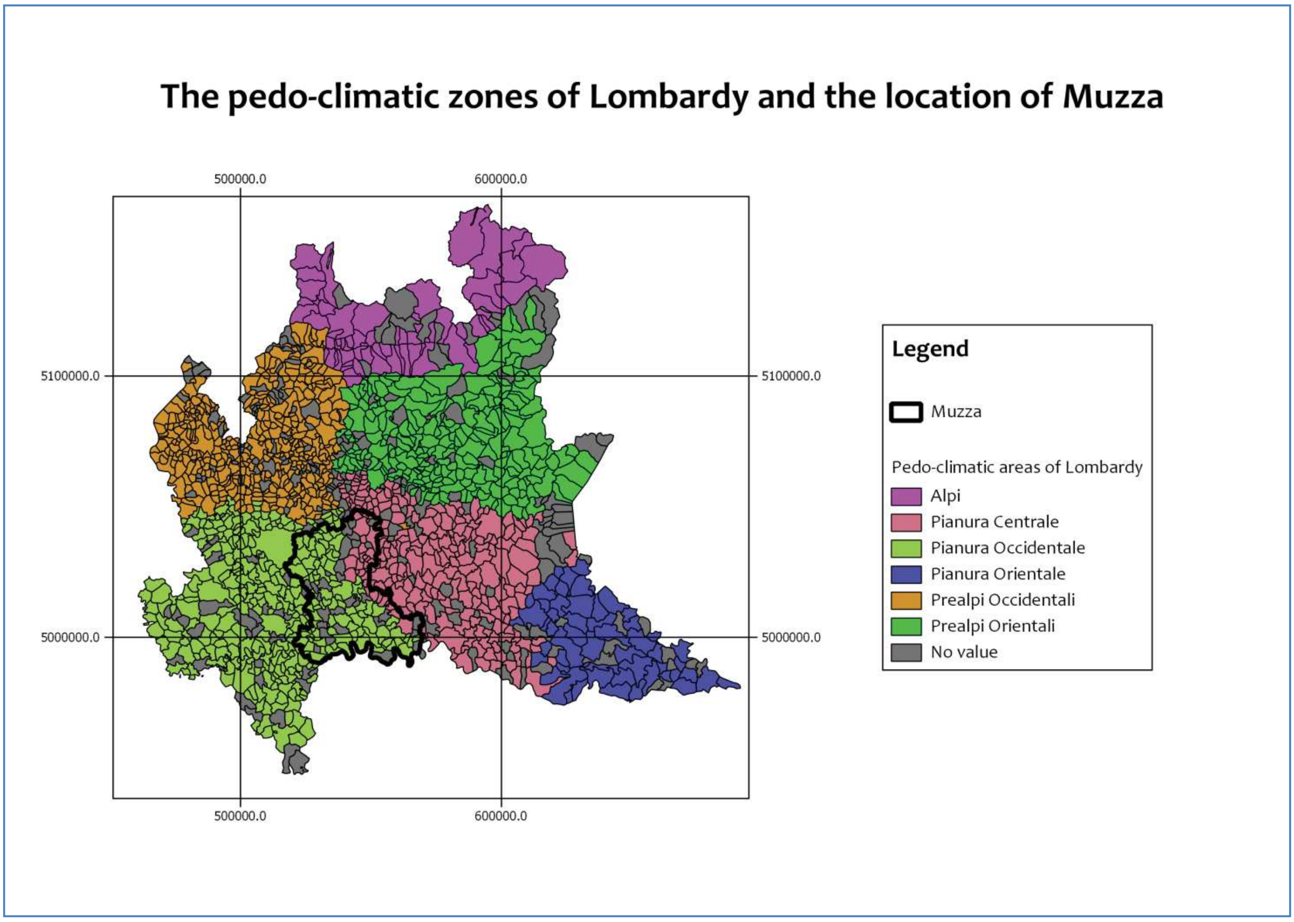


Fig. 41 – Agro-ecological zones of Lombardy according to LoPCZ

Appendix 6 – Crop classes found in the original dataset and crop classes used in this study

Aggregated class	Class found in the original dataset
Cereals and cash crops	
Winter cereals	GRANO (FRUMENTO) TENERO
	GRANO (FRUMENTO) TENERO PER LA PRODUZIONE DI SEME
	GRANO (FRUMENTO) DURO
	ORZO
	FRUMENTO SEGALATO (TRITICALE)
	TRITICALE PER LA PRODUZIONE DI SEME
	FARRO
	SEGALE
Buckwheat	GRANO SARACENO
Other summer cereals	CHENOPODIUM QUINOA
	MIGLIO
	PANICO
Oat	AVENA
Rice	RISO
	RISONE TONDO
	RISONE MEDIO
	RISONE LUNGO B
Sorghum	SORGO DA GRANELLA
	SORGO DA GRANELLA USO ENERGETICO
	SORGO PER LA PRODUZIONE DI SEME
Maize	MAIS DOLCE
	GRANTURCO (MAIS) PER LA PRODUZIONE DI SEME
	GRANTURCO (MAIS) DA GRANELLA USO ENERGETICO
	GRANTURCO (MAIS) DOLCE PER LA PRODUZIONE DI SEME
	MAIS DA GRANELLA
Sunflower	GIRASOLE
	GIRASOLE DA GRANELLA
Soybean	SOIA DA GRANELLA
	SOIA - FAVE EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	SOIA - FAVE USO ENERGETICO-EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
Rapeseed	COLZA E RAVIZZONE DA GRANELLA
	RAVIZZONE - SEMI USO ENERGETICO
	RAVIZZONE - SEMI IBRIDI
	RAVIZZONE - SEMI
Sugar beet	BARBABIETOLA DA ZUCCHERO
Lineseed	LINO NON TESSILE
Cannabis	CANAPA GREGGIA O MACERATA
Other cereals and cash crops	ALTRI SEMINATIVI
Forages	
Herbages (1 year)	
Winter cereals for forage	ORZO USO ENERGETICO
	ORZO - DA FORAGGIO USO ENERGETICO
	ORZO - DA FORAGGIO ERBAIO IN PUREZZA
	TRITICALE - DA FORAGGIO ERBAIO IN PUREZZA
	TRITICALE - DA FORAGGIO USO ENERGETICO
	GRANO (FRUMENTO) TENERO - DA FORAGGIO ERBAIO IN PUREZZA
	GRANO (FRUMENTO) TENERO USO ENERGETICO
	GRANO (FRUMENTO) TENERO - DA FORAGGIO USO ENERGETICO

	AVENA - DA FORAGGIO USO ENERGETICO
	TRITORDEUM - DA FORAGGIO USO ENERGETICO
Sorghum for forage	SORGO DA FORAGGIO
	SORGO DA FORAGGIO ERBAIO IN PUREZZA
Mixed herbages	ERBAIO MISTO
	ERBAIO DI GRAMINACEE
	ERBAIO DI LEGUMINOSE
Silo maize	SILOMAIS E MAIS CEROSO
	MAIS DA FORAGGIO
	GRANTURCO (MAIS) - INSILATO USO ENERGETICO
Trifolium alexandrinum	TRIFOGLIO PER LA PRODUZIONE DI SEME (SP. TRIFOLIUM ALEXANDRINUM L.) EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
Vetch	VECCIA SATIVA EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	VICIA SATIVA L.
	VECCIA
	VECCIA SATIVA ERBAIO IN PUREZZA
	VECCE - DA FORAGGIO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
Annual rye-grass	LOLIUM MULTIFLORUM LAM.
	LOIETTO LOGLIO DA FORAGGIO ERBAIO IN PUREZZA
Soybean for forage	SOIA - DA FORAGGIO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
Meadows (2-5 years)	
Alfalfa	ERBA MEDICA
	ERBA MEDICA - DA FORAGGIO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	ERBA MEDICA - DA FORAGGIO PRATO PASCOLO IN PUREZZA AVVICENDATO - NON PERMANENTE-efa
Perennial rye-grass	LOIETTO LOGLIO DA FORAGGIO PRATO PASCOLO IN PUREZZA AVVICENDATO - NON PERMANENTE
Red clover	TRIFOGLIO PER LA PRODUZIONE DI SEME (SP. TRIFOLIUM PRATENSE L.) EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
Mixed meadow	PRATO POLIFITA DA VICENDA
	PRATO POLIFITA AVVICENDATO - NON PERMANENTE
	PRATO PASCOLO DI GRAMINACEE AVVICENDATO - NON PERMANENTE
	PRATO-PASCOLO
Pastures (>5 years)	
Pastures	PRATO POLIFITA NON AVVICENDATO (PRATO STABILE)
	PRATO POLIFITA NON AVVICENDATO PER ALMENO 5 ANNI (SFALCIATO) - PERMANENTE
	PASCOLO CON ROCCIA AFFIORANTE (TARA 50%)
Other forages	TRIFOGLIO
	TRIFOGLIO - DA FORAGGIO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	ALTRE FORAGGERE
	PRATO IN ROTOLO (TAPPETO ERBOSO)
Pulses and vegetables	
Pea	PISELLI ALLO STATO FRESCO EFA - AREA DI INTERESSE ECOLOGICO - Colture
	PISELLO FRESCO
	PISELLO SECCO
	PISELLI DA ORTO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	PISELLI ALLO STATO SECCO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici

Bean	FAGIOLO
	FAGIOLINO
	FAGIUOLO FRESCO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	FAGIOLO EFA - AREA DI INTERESSE ECOLOGICO - Colture azotofissatrici
	FAGIUOLO FRESCO
Lentil	LENTICCHIA
Tomato	POMODORO
	POMODORO TONDO ALTRE VARIETA'
	POMODORINO DA MENSA
	POMODORINO
Zucchini	ZUCCHINO
	ZUCCHINO PER LA PRODUZIONE DI SEME
Lettuce	LATTUGA
Spinach	SPINACIO
Aubergine	MELANZANA
Potato	PATATA
Onion	CIPOLLA
Asparagus	ASPARAGO
Blueberry	MIRTILLO
Leek	PORRO
Garlic	AGLIO
Broccoli	BROCCOLO
Carrot	CAROTA
Blackberry	MORA
Strawberry	FRAGOLA
Raspberry	LAMPONE
Ribes	RIBES ROSSO
	RIBES
Bell pepper	PEPERONE
Chard	BIETOLA DA ORTO
Celery	SEDANO
Basil	BASILICO
Chicory	CICORIA O RADICCHIO
Radicchio	RADICCHIO
Cabbage	CAVOLO
Aronia arbutifolia	FRUTTI DELLA SPECIE ARONIA ARBUTIFOLIA
Melon	MELONE
Watermelon	COCOMERO
Pumpkin	ZUCCA
Other vegetables	PIANTE ORTICOLE A PIENO CAMPO
	ALTRE ORTICOLE
Trees	
Tree crops	ACTINIDIA
	ALBICOCCO
	ALTRE COLTIVAZIONI LEGNOSE AGRARIE
	ALTRE PIANTE ARBOREE DA FRUTTO
	ALTRE PIANTE ARBOREE DA LEGNO
	CASTAGNO DA MENSA
	CEDRO
	CILIEGIO
	COLTIVAZIONI ARBOREE PROMISCUE (PIU' SPECIE ARBOREE)
	COLTIVAZIONI ARBOREE SPECIALIZZATE
	FICO
	LOTO O KAKI
	LYCIUM BARBARUM (GOJI)
	MELO

	NOCCIOLO
	NOCE
	PERO
	PESCO
	PIOPPETI ED ALTRE COLTIVAZIONI ARBOREE DA LEGNO - SPECIE NON DEFINITA EFA - AREA DI INTERESSE ECOLOGICO - Misure forestali
	PIOPPETO
	PRUGNE DA DESTINARE ALLA TRASFORMAZIONE
	SUSINO
	VITE PER UVA DA AUTOCONSUMO
	VITE PER UVA DA MENSA
	VITE PER UVA DA VINO IN ZONA DOC E/O DOCG
	VITE PER VITIGNI SPERIMENTALI
Tree rows	ALBERI IN FILARE
	ALBERI IN FILARE INCLUSO/ADIACENTE AL PRATO PERMANENTE (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	ALBERI IN FILARE INCLUSO/ADIACENTE AL SEMINATIVO (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	ALBERI ISOLATI INCLUSO/ADIACENTE AL SEMINATIVO (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
Woods	ALBERI DA BOSCO - SUPERFICI IMBOSCHITE AI SENSI DEL REG.(CE) N. 1257/99 MISURA H - ARBORICOLTURA DA LEGNO
	ALBERI DA BOSCO A BREVE ROTAZIONE
	BOSCO
	BOSCO BOSCO MISTO - EFA - AREA DI INTERESSE ECOLOGICO - Misure forestali
	BOSCO DIVERSO DA BOSCO SPONTANEO E/O PREESISTENTE
	BOSCO EFA - AREA DI INTERESSE ECOLOGICO - Misure forestali
	BOSCO MISTO
	CEDUO COMPOSTO
	FUSTAIA MISTA DI CONIFERE E LATIFOGLIE
	GRUPPI DI ALBERI E BOSCHETTI
	GRUPPI DI ALBERI E BOSCHETTI INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	GRUPPI DI ALBERI E BOSCHETTI NON INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
Other trees	ARBORETO CONSOCIABILE (CON COLTIVAZIONI ERBACEE)
Other classes	
Non-cultivated land	ALTRA SUPERFICIE NON UTILIZZATA (TERRENI ABBANDONATI
	RIPOSO VOLONTARIO - COLTURE A PERDERE PER LA FAUNA - MISCUGLIO DI SORGO
	RIPOSO VOLONTARIO - COPERTURA VEGETALE SEMINATA O SPONTANEA
	RIPOSO VOLONTARIO - LAVORAZIONI MECCANICHE INTENZIONE DI SEMINA DOPO IL 15 LUGLIO
	RIPOSO VOLONTARIO - SOVESCIO IN PRESENZA DI SPECIE DA SOVESCIO O DI PIANTE BIOCIDE
	RIPOSO VOLONTARIO - SUPERFICIE INTERESSATA DALL'ESECUZIONE DI INTERVENTI DI MIGLIORAMENTO FONDIARIO
	SUPERFICI AGRICOLE NON SEMINATE – DISATTIVATE
	SUPERFICI AGRICOLE RITIRATE DALLA PRODUZIONE TERRENO COPERTO DA VEGETAZIONE SPONTANEA-COPERTURA VEGETALE SPONTANEA
	SUPERFICI AGRICOLE RITIRATE DALLA PRODUZIONE TERRENO NUDO-LAVORAZIONI FUNZIONALI A INTERVENTI DI MIGLIORAMENTO FONDIARIO
	SUPERFICI AGRICOLE RITIRATE DALLA PRODUZIONE TERRENO

	NUDO-LAVORAZIONI PREPARATORIE DEL TERRENO O PER IL CONTENIMENTO DELLE INFESTANTI
	TARE E INCOLTI
	USO NON AGRICOLO - AREE NON COLTIVABILI
	USO NON AGRICOLO - FABBRICATI PARCHI
Minor crops	ARUNDO DONAX
	BAMBU GIGANTE
	LAVANDA
	ORTO FAMILIARE
	PIANTE
	PIANTE AROMATICHE
	PIANTE AROMATICHE MEDICINALI E SPEZIE
	TARTUFO
	FIORI E PIANTE ORNAMENTALI IN PIENA ARIA
Buffer areas	FASCE TAMPONE
	FASCE TAMPONE NON RIPARIALI ARBOREE E ARBUSTIVE
	FASCE TAMPONE RIPARIALI ARBOREE E ARBUSTIVE-INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	FASCE TAMPONE RIPARIALI ERBACEE-INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	FASCE TAMPONE RIPARIALI ERBACEE-NON INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	SIEPI
	SIEPI E FASCE ALBERATE BARRIERE SCHERMANTI-INCLUSO/ADIACENTE AL SEMINATIVO (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	SIEPI E FASCE ALBERATE INCLUSO/ADIACENTE AL PRATO PERMANENTE (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	SIEPI E FASCE ALBERATE INCLUSO/ADIACENTE AL SEMINATIVO (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	SIEPI E FASCE ALBERATE INCLUSO/ADIACENTE ALLA COLTURA PERMANENTE (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
Field boundaries	FOSSATI E CANALI INCLUSO/ADIACENTE AL PRATO PERMANENTE (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	FOSSATI E CANALI INCLUSO/ADIACENTE AL SEMINATIVO (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	FOSSATI E CANALI INCLUSO/ADIACENTE ALLA COLTURA PERMANENTE (ELEMENTI DEL PAESAGGIO/EFA - AREA DI INTERESSE ECOLOGICO)
	MARGINI DEI CAMPI
	MARGINI DEI CAMPI INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	MARGINI DEI CAMPI NON INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
	MARGINI DEI CAMPI SEMINABILI INCLUSO/ADIACENTE AL SEMINATIVO (EFA - AREA DI INTERESSE ECOLOGICO)
Greenhouses and rural buildings	FIORI E PIANTE ORNAMENTALI PROTETTE IN SERRA
	FIORI E PIANTE ORNAMENTALI PROTETTE IN SERRE O TUNNEL CALDI
	FIORI E PIANTE ORNAMENTALI PROTETTE IN TUNNEL O ALTRO
	SERRE FISSE
	FABBRICATI AGRICOLI
	PIANTE ORTICOLE PROTETTE IN SERRA
Nurseries	VIVAIO FLORICOLI E PIANTE ORNAMENTALI
	VIVAIO FORESTALE
	ALTRI VIVAI

Tab. 19 - Crop classes found in the original dataset and crop classes used in this study

Appendix 7 – List of Landsat8 images used in this study

Path, Row	Year	DOY	Date (day-month)	Cloud cover (%)
194, 28	2014	216	4-8	5
193, 29	2014	225	13-8	90
194, 28	2014	232	20-8	100
193, 29	2014	241	29-8	100
194, 28	2014	248	5-9	95
193, 29	2014	257	14-9	0
194, 28	2014	264	21-9	95
193, 29	2014	273	30-9	60
194, 28	2014	280	7-10	100
193, 29	2014	289	16-10	95
194, 28	2014	296	23-10	0
193, 29	2014	305	1-11	25
194, 28	2014	312	8-11	90
193, 29	2014	321	17-11	100
194, 28	2014	328	24-11	90
193, 29	2014	337	3-12	100
194, 28	2014	344	10-12	0
193, 29	2014	353	19-12	40
194, 28	2014	360	26-12	80
193, 29	2015	004	4-1	0
194, 28	2015	011	11-1	70
193, 29	2015	020	20-1	60
194, 28	2015	027	27-1	100
193, 29	2015	036	5-2	100
194, 28	2015	043	12-2	80
193, 29	2015	052	21-2	100
194, 28	2015	059	28-2	80
193, 29	2015	068	9-3	100
194, 28	2015	075	16-3	100
193, 29	2015	084	25-3	100
194, 28	2015	091	1-4	0
193, 29	2015	100	10-4	90
194, 28	2015	107	17-4	100
193, 29	2015	116	26-4	100
194, 28	2015	123	3-5	90
193, 29	2015	132	12-5	100
194, 28	2015	139	19-5	95
193, 29	2015	148	28-5	0
194, 28	2015	155	4-6	0
193, 29	2015	164	13-6	80
194, 28	2015	171	20-6	2
193, 29	2015	180	29-6	90
194, 28	2015	187	6-7	50
193, 29	2015	196	15-7	5
194, 28	2015	203	22-7	2
193, 29	2015	212	31-7	20

194, 28	2015	219	7-8	0
193, 29	2015	228	16-8	50
194, 28	2015	235	23-8	95
193, 29	2015	244	1-9	40
194, 28	2015	251	8-9	70
193, 29	2015	260	17-9	70
194, 28	2015	267	24-9	5
193, 29	2015	276	3-10	90
194, 28	2015	283	10-10	70
193, 29	2015	292	19-10	10
194, 28	2015	299	26-10	90
193, 29	2015	308	4-11	80
194, 28	2015	315	11-11	10
193, 29	2015	324	20-11	90
194, 28	2015	331	27-11	0
193, 29	2015	340	6-12	100
194, 28	2015	347	13-12	100
193, 29	2015	356	22-12	100
194, 28	2015	363	29-12	90

Tab. 20 – List of Landsat8 scenes used in this study, with the estimation of the cloud cover over the area of interest

Appendix 8 – Confusion matrices of the 4 Random Forest models tested

Random Forest performed with EVI phenological metrics
Confusion matrix

	Winter cereals	Non-cult. land	Woods	Mixed herb.	Silo maize	Alfalfa	Mixed mead.	Past.	Rice	Soy bean	Tomato	Other crops	Tot	CE
Winter cereals	213	0	0	0	0	0	0	0	0	0	0	491	704	0.70
Non-cult. land	0	218	0	0	0	0	0	0	0	0	0	535	753	0.71
Woods	0	0	0	0	0	0	0	0	0	0	0	53	53	1.00
Mixed herb.	0	0	0	0	0	0	0	0	0	0	0	71	71	1.00
Silo maize	0	0	0	0	531	0	0	0	0	0	0	1474	2005	0.74
Alfalfa	0	0	0	0	0	0	0	0	0	0	0	35	35	1.00
Mixed mead.	0	0	0	0	0	0	82	0	0	0	0	359	441	0.81
Past.	0	0	0	0	0	0	0	136	0	0	0	20	156	0.13
Rice	0	0	0	0	0	0	0	0	87	0	0	123	210	0.59
Soybean	0	25	0	0	0	0	0	0	0	31	0	220	276	0.89
Tomato	0	0	0	0	0	0	0	0	0	0	134	0	134	0.00
Other crops	56	70	0	0	209	0	29	31	40	9	22	181990	182456	0.00
Tot	269	313	0	0	740	0	111	167	127	40	156	185371	187294	
OE	0.21	0.30	1.00	1.00	0.28	1.00	0.26	0.19	0.31	0.23	0.14	0.02		
OA	0.979													

Tab. 21 – Confusion matrix of the classification performed with the EVI PMs-based RF

Random Forest performed with NDVI phenological metrics
Confusion matrix

	Winter cereals	Non-cult. land	Woods	Mixed herb.	Silo maize	Mixed mead.	Past.	Rice	Soy bean	Tomato	Other crops	Tot	CE
Winter cereals	213	0	0	0	0	0	0	0	0	0	282	495	0.57
Non-cult. Land	0	391	0	0	0	0	0	0	0	0	709	1100	0.64
Woods	0	0	20	0	0	0	0	0	0	0	40	60	0.67
Mixed herb.	0	0	0	0	0	0	0	0	0	0	71	71	1.00
Silo maize	0	0	0	0	513	0	0	0	0	0	1811	2324	0.78
Mixed mead.	0	0	0	0	0	53	0	0	0	0	431	484	0.89
Past.	0	0	0	0	0	0	42	0	0	0	37	79	0.47
Rice	0	0	0	0	0	0	0	87	0	0	138	225	0.61
Soybean	0	25	0	0	0	0	0	0	44	0	228	297	0.85
Tomato	0	0	0	0	0	0	0	0	0	184	0	184	0.00
Other crops	56	198	0	0	239	16	20	40	12	51	214561	215193	0.00
Tot	269	614	20	0	752	69	62	127	56	235	218308	220512	
OE	0.21	0.36	0.00	1.00	0.32	0.23	0.32	0.31	0.21	0.22	0.02		
OA	0.980												

Tab. 22 – Confusion matrix of the classification performed with the NDVI PMs-based RF

Random Forest performed with SAVI phenological metrics Confusion matrix													
	Winter cereals	**Maize**	**Non-cult. land**	**Woods**	**Mixed herb.**	**Silo maize**	**Mixed mead.**	**Past.**	**Rice**	**Soy bean**	**Other crops**	Tot	CE
Winter cereals	88	27	0	0	0	0	0	0	0	0	346	461	0.81
Maize	0	1578	0	0	0	0	0	0	0	0	2135	3713	0.58
Non-cult. Land	0	0	299	0	0	0	0	0	0	0	533	832	0.64
Woods	0	0	0	0	0	0	0	0	0	0	40	40	1.00
Mixed herb.	0	0	0	0	0	0	0	0	0	0	71	71	1.00
Silo maize	0	16	0	0	0	479	0	0	0	0	1715	2210	0.78
Mixed mead.	0	46	0	0	0	0	57	0	0	0	441	544	0.90
Past.	0	0	0	0	0	0	0	148	0	0	37	185	0.20
Rice	0	0	0	0	0	0	0	0	87	0	98	185	0.53
Soybean	0	0	25	0	0	0	0	0	0	31	202	258	0.88
Other crops	27	678	113	0	0	175	7	31	40	9	192508	193588	0.01
Tot	115	2345	437	0	0	654	64	179	127	40	198126	202087	
OE	0.23	0.33	0.32	1.00	1.00	0.27	0.11	0.17	0.31	0.23	0.03		
OA	**0.97**												

Tab. 23 – Confusion matrix of the classification performed with the SAVI PMs-based RF

Random Forest performed with EVI, NDVI and SAVI phenological metrics Confusion matrix												
	Winter cereals	**Maize**	**Non-cult. land**	**Mixed herb.**	**Silo maize**	**Mixed mead.**	**Past.**	**Rice**	**Soy bean**	**Other crops**	Tot	CE
Winter cereals	513	38	0	0	0	0	0	0	0	0	551	0.07
Maize	26	2979	0	0	55	4	0	0	0	0	3064	0.03
Non-cult. Land	0	27	799	0	12	0	0	0	0	0	838	0.05
Mixed herb.	0	0	0	46	25	0	0	0	0	0	71	0.35
Silo maize	0	38	0	0	1732	0	0	0	27	0	1797	0.04
Mixed mead.	39	89	0	0	28	90	0	14	0	0	260	0.65
Past.	0	0	0	0	0	0	82	0	0	0	82	0.00
Rice	0	0	0	0	0	0	0	173	0	0	173	0.00
Soybean	0	0	25	0	33	0	0	0	151	0	209	0.28
Other crops	0	0	0	0	0	0	0	0	0	51	51	0.00
Tot	578	3171	824	46	1885	94	82	187	178	51	7096	
OE	0.11	0.06	0.03	0.00	0.08	0.04	0.00	0.07	0.15	0.00		
OA	**0.93**											

Tab. 24 – Confusion matrix of the classification performed with the EVI-NDVI-SAVI PMs-based RF

List of tables

List of figures

List of abbreviations

AEZ: Agro-Ecological Zoning; Agro-Ecological Zones

CE: Commission Error

CZ: Climatic Zonation; Climatic Zone

EMD: Euclidean Minimum Distance

EO: Earth Observation

EVI: Enhanced Vegetation Index

FAO: Food and Agriculture Organization

FAPAR: Fraction of Absorbed Photosynthetically Active Radiation

FVC: Fractional Vegetation Cover

GAES: Global Agro-Environmental Stratification

GAEZ: Global Agro-Ecological Zonation

GEnS: Global Environmental Stratification

GYGA-ED: Global Yield Gap Atlas Extrapolation Domain

LAI: Leaf Area Index

LoPCZ: Lombardy Pedo-Climatic Zonation

LSP: Land Surface Phenology

LUC: Land Use Change

MCL: Maximum Likelihood Classification

MSAVI: Modified Soil-Ajusted Vegetation Index

NDVI: Normalized Difference Vegetation Index

OA: Overall Accuracy

OBIA: Object-Based Image Analysis

OE: Omission Error

OLI: Operational Land Manager

PA: Producer’s Accuracy

PM: Phenological Metric

RS: Remote Sensing

SAM: Spectral Angle Mapper

SAR: Synthetic Aperture Radar

SAVI: Soil-Adjusted Vegetation Index

SMA: Spectral Unmixing

TCG: Tasselled Cap Greenness

UA: User's Accuracy

UAA: Utilized Agricultural Area

VI: Vegetation Index

Definitions

Agro-ecology: «the science of the relationships of organisms in an environment purposely transformed by man for crop or livestock production» (Martin & Sauerborn, 2013).

Agro-ecological zones: «zones which have similar combinations of climate and soil characteristics, and similar physical potentials for agricultural production» (FAO, 1996).

Agro-ecological zoning: the «division of an area of land into smaller units, which have similar characteristics related to land suitability, potential production and environmental impact» (FAO, 1996).

Classification: «a process in which each pixel of an image is assigned to a category, among a set of categories of interest» (Khorram et al., 2012).

Land cover: «the description of the land surface in terms of soils and vegetation layers, including natural vegetation, crops and human structures» (Burley, 1961); «the observed (bio)physical cover on the earth's surface» (FAO, 2000).

Land use: «it refers to the purpose for which humans exploit the land cover, including land management techniques» (Lambin, Geist, & Rindfass, 2006).

Land use systems: «they can be defined as a coupled human-environment system; they describe how land, as an essential resource, is being used and managed» (Bégué et al., 2015).

Radiometric resolution: «it is the sensitivity of the sensor, i.e. its capacity to discriminate small variations in the recorded spectral radiance» (Chuvieco & Huete, 2010).

Remote sensing: «the acquisition of information about the state and condition of an object through sensors that are not in physical contact with it; the information is transmitted from the object to the sensors in the form of electromagnetic radiation» (Chuvieco & Huete, 2010).

Resolution: «the sensor's ability to discriminate information» (Estes & Simonett, 1975); «the ability of the sensor to distinguish a specific object from other objects» (Chuvieco & Huete, 2010).

Segmentation: the process of partitioning an image into non-overlapping regions, which are called segments (Schiewe, 2002).

Spatial resolution: «it is a measure of the fineness of detail of an image» (Khorram et al., 2012).

Spectral resolution: «it defines the sensor's ability to detect wavelength differences between objects or areas of interest» (Khorram et al., 2012).

Temporal resolution: «it is the observation frequency, or revisiting period, of the sensor» (Chuvieco & Huete, 2010).

Vegetation Indices: «techniques to extract quantitative information on the amount of vegetation, or greenness, for every pixel in an image. They typically involve spectral transformations of two or more bands […]» (Chuvieco & Huete, 2010).

References

1. Aa., Vv. (2016). SEGUICI - Elaborazione di prodotti da Earth Observation per lo studio della vulnerabilità degli acquiferi in aree agricole.
2. Ali, S. (2002). *Use of spectral and temporal unmixing for crop identification using multi-spectral data*. Retrieved from http://www.itc.nl/library/papers/msc_2002/gfm/samia_ali.pdf
3. ARPA Lombardia. (2010). SINTESI CLIMATICA DEL 2010 IN LOMBARDIA.
4. ARPA Lombardia. (2015). *STATO DELLE ACQUE SOTTERRANEE. AREA IDROGEOLOGICA ADDA-OGLIO. RAPPORTO ANNUALE 2014*.
5. Asner, G. P., Scurlock, J. M. O., Hicke, J. A., Scurlockt, J. M. O., & Hicket, J. A. (2003). Global Synthesis of Leaf Area Index Observations: Implications for Ecological and Remote Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography Global Ecology & Biogeography*, *12*(1), 191–205. Retrieved from http://www.jstor.org/stable/3697502%5Cnhttp://about.jstor.org/terms
6. ASP Lombardia. (2010). Ambiente e territorio. QUANTITÀ DI PRECIPITAZIONI TOTALI ANNUE E MEDIA DECENNALE LOMBARDIA E PROVINCE LOMBARDE. ANNO 2009.
7. Azar, R., Villa, P., Stroppiana, D., & Crema, A. (2016). Assessing in-season crop classification performance using satellite data : a test case in Northern Italy, 361–380. https://doi.org/10.5721/EuJRS20164920
8. Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, 1. https://doi.org/10.1080/01431160110106113
9. Balaguer, A., Ruiz, L. A., Hermosilla, T., & Recio, J. A. (2010). Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Computers and Geosciences*, *36*(2), 231–240. https://doi.org/10.1016/j.cageo.2009.05.003
10. Baldoni, R., & Giardini, L. (2001). *Coltivazioni Erbacee*. Patron Editore.
11. Bannari, A., Pacheco, A., Staenz, K., McNairn, H., & Omari, K. (2006). Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data. *Remote Sensing of Environment*, *104*(4), 447–459. https://doi.org/10.1016/j.rse.2006.05.018
12. Bégué, A., Arvor, D., Lelong, C., Vintrou, E., & Simoes, M. (2015). Agricultural Systems Studies Using Remote Sensing. In P. S. . Thenkabail (Ed.), *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. CRC Press. Retrieved from http://www.crcnetbase.com/doi/book/10.1201/b19322
13. Belhouchette, H., Louhichi, K., Therond, O., Mouratiadou, I., Wery, J., Ittersum, M. van, & Flichman, G. (2011). Assessing the impact of the Nitrate Directive on farming systems using a bio-economic modelling Chain. *Agricultural Systems*, *104*(2), 135–145. https://doi.org/10.1016/j.agsy.2010.09.003
14. Blaschke, T., Lang, S., & Hay, G. J. (2008). *Object-Based Image Analysis*. (T. Blaschke, S. Lang, & G. J. Hay, Eds.). Springer-Verlag Berlin Heidelberg.
15. Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., … Defourny, P. (2015). Building a Data Set over 12 Globally Distributed Sites to Support the Development of Agriculture Monitoring Applications with Sentinel-2. *Remote Sensing*, *7*(12), 16062–16090. https://doi.org/10.3390/rs71215815
16. Brady, N., & Weil, R. (2008). *The nature and properties of soils*. Pearson.
17. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
18. Brivio, P. A., Lechi, G., & Zilioli, E. (2006). Classificazione delle immagini. In *Principi e metodi di Telerilevamento*.
19. Brown, J. C., Kastens, J. H., Coutinho, A. C., Victoria, D. de C., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment*, *130*, 39–50. https://doi.org/10.1016/j.rse.2012.11.009
20. Burley, T. M. (1961). Land use or land utilization? *The Professional Geographer*, *13*, 18–20.
21. Chan, J. C. W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne

hyperspectral imagery. *Remote Sensing of Environment*, *112*(6), 2999–3011. https://doi.org/10.1016/j.rse.2008.02.011
22. Chen, P.-Y., Fedosejevs, G., Tiscareno-Lopez, M., & Arnold, J. G. (2006). Assessment of MODIS-EVI and vegetation-NDVI composite data using agricultural measurements: an example at corn fields in Western Mexico. *Environmental Monitoring and Assessment*, *119*(1–3), 69–82. https://doi.org/10.1007/sl0661-005-9006-7
23. Chikodzi, D., Hardlife, Z., Farai Malvern, S., & Talent, M. (2013). Reclassification of agro-ecological zones in Zimbabwe - the rationale, methods and expected benefits: The case of masvingo province. *Journal of Sustainable Development in Africa*, *15*, no.*1*(1), 104–116.
24. Chuvieco, E., & Huete, A. (2010). *Fundamentals of Satellite Remote Sesnsing*. CRC Press, Taylor and Francis Group, LLC.
25. Cochet, H. (2012). The systeme agraire concept in francophone peasant studies. *Geoforum*, *43*(1), 128–136. https://doi.org/10.1016/j.geoforum.2011.04.002
26. Costantini, E., Urbano, F., & L'Abate, G. (2004). Soil regions of Italy. *CRA-ISSDS, Firenze*.
27. Cuca, B., & Tramutoli, V. (2014). *Open Document How Can Space Make a Difference for Agriculture Sector*.
28. De Wit, a. J. W., & Clevers, J. G. P. W. (2004). Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, *25*(20), 4091–4112. https://doi.org/10.1080/01431160310001619580
29. Dennison, P. E., Roberts, D. A., & Peterson, S. H. (2007). Spectral shape-based temporal compositing algorithms for MODIS surface reflectance data. *Remote Sensing of Environment*, *109*(4), 510–522. https://doi.org/10.1016/j.rse.2007.02.009
30. Dey, V., Zhang, Y., Zhong, M., & Engineering, G. (2010). A Review on Image Segmentation Techniques With Remote Sensing Perspective. In *ISPRS TC VII Symposium – 100 Years ISPRS* (Vol. XXXVIII, pp. 31–42).
31. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., … Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026
32. EEA. (2011). Agriculture.
33. EEA. (2012a). Percentage of groundwater body area not achieving good chemical status due to nitrate (a) and total nitrogen input from organic and inorganic fertilisers (b).
34. EEA. (2012b). Projected changes in water-limited crop yield.
35. EEA. (2016). Water exploitation index plus for river basin districts.
36. Elmore, A. J., Mustard, J. F., Manning, S. J., & Lobell, D. B. (2000). Quantifying Vegetation Change in Semiarid Environments: Precision and Accuracy of Spectral Mixture Analysis and the Normalized Difference Vegetation Index. *Remote Sensing of Environment*, *73*, 16.
37. ERSAF Lombardia. (2016). Allegato A - " Programma d ' Azione regionale per la protezione delle acque dall ' inquinamento provocato dai nitrati provenienti da fonti agricole nelle zone vulnerabili ai sensi della Direttiva nitrati 91 / 676 / CEE ."
38. ESA. (2000). Copernicus Overview. Retrieved March 22, 2017, from http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3
39. ESA. (2016a). Sentinel-2, Plant health. Retrieved December 28, 2016, from http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Plant_health
40. ESA. (2016b). Sentinel-2 Thematic Areas and Services. Retrieved December 28, 2016, from https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/thematic-areas-and-services
41. Estes, J. E., & Simonett, D. S. (1975). Fundamentals of Image Interpretation. In R. G. Reeves (Ed.), *Manual of Remote Sensing*. Falls Church, VA: American Society of Phtogrammetry.
42. European Commission. (1991). Directive 91/676/EEC. *Official Journal of the European Communities*, *L 375*(31.12.1991), 1–8.
43. European Commission. (2016a). *Italy - CAP in your country*. *Quality Progress*.
44. European Commission. (2016b). *Italy - Statistical factsheet*.

45. European Union. (2015). *Confronto dei redditi degli agricoltori negli Stati membri dell'Unione Europea*.
46. Europedia. (2011). EU agricultural policy.
47. Eurostat. (2012). Agricultural census in Italy. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Agricultural_census_in_Italy#Main_statistical_findings
48. FAO. (1996). Agroecological zoning. Guidelines. *FAO Soils Bulletin*, *76*.
49. Fischer, G., Shah, M., Tubiello, F. N., & van Velhuizen, H. (2005). Socio-economic and climate change impacts on agriculture: an integrated assessment, 1990-2080. *Philos Trans R Soc Lond B Biol Sci*, *360*(1463), 2067–2083. https://doi.org/10.1098/rstb.2005.1744
50. Friedl, M. a. M. A., & Brodley, C. E. C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409. https://doi.org/10.1016/S0034-4257(97)00049-7
51. Geurts, P. M. H., & Berg, M. Van Den. (1998). A simple agro-ecological zonation for fertilizer recommendations in Mozambique. *Soil Use and Management*, *14*, 136–141.
52. Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011
53. Global Administrative Areas. (2012). GADM database of Global Administrative Areas, version 2.8. Retrieved March 28, 2017, from http://www.gadm.org/
54. Google Play. (2017). SEGUICI vegetation report. Retrieved March 22, 2017, from https://play.google.com/store/apps/details?id=it.eucentre.vegetationreport&hl=it
55. Haibo, Y., & Yanbo, H. (2013). Remote Sensing Applications to Precision Farming. In *Remote Sensing of Natural Resources* (pp. 333–352). https://doi.org/doi:10.1201/b15159-24\r10.1201/b15159-24
56. Halabuk, A., Mojses, M., Halabuk, M., & David, S. (2015). Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. *Remote Sensing*, *7*(5), 6107–6132. https://doi.org/10.3390/rs70506107
57. Hao, P., Zhan, Y., Wang, L., Niu, Z., & Shakir, M. (2015). Feature Selection of Time Series MODIS Data for Early Crop Classification Using Random Forest: A Case Study in Kansas, USA, 5347–5369. https://doi.org/10.3390/rs70505347
58. Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. R package version 2.5-8. Retrieved from https://cran.r-project.org/package=raster
59. Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X
60. Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, *8*(3), 166. https://doi.org/10.3390/rs8030166
61. Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., … Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, *7*(9), 12356–12379. https://doi.org/10.3390/rs70912356
62. Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). Improved Early Crop Type Identification By Joint Use of HighTemporal Resolution SAR And Optical Image Time Series. *Remote Sensing*, *8*(362), 21. https://doi.org/10.3390/rs8050362
63. Istat. (2010). Lombardia, 6° censimento generale dell'agricoltura risultati definitivi, 1–5.
64. Istat. (2011). Lombardia - Istat. Retrieved from https://www.istat.it/it/lombardia/dati?qt=gettable&dataset=DCSP_COLTIVAZ&dim=21,1,9,0,0&lang=2&tr=0&te=0
65. Istat. (2013). *6 Censimento generale dell'Agricoltura: Atlante dell'agricoltura italiana*.
66. ISTAT. (2012). 6 Censimento Agricoltura 2010 - Data warehouse.
67. Italian National Centre for Soil Mapping. (2012). Soil maps of Italy. Retrieved March 24, 2017, from http://www.soilmaps.it/en/home.html
68. Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., … Hunt, E. R. (2004).

Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, *92*(4), 475–482. https://doi.org/10.1016/j.rse.2003.10.021
69. Jin, N., Tao, B., Ren, W., Feng, M., Sun, R., He, L., … Yu, Q. (2016). Mapping Irrigated and Rainfed Wheat Areas Using Multi-Temporal Satellite Data. *Remote Sensing*, *8*(3), 207. https://doi.org/10.3390/rs8030207
70. Kaizzi, C. K., Ssali, H., & Vlek, P. L. G. (2006). Differential use and benefits of Velvet bean (Mucuna pruriens var. utilis) and N fertilizers in maize production in contrasting agro-ecological zones of E. Uganda. *Agricultural Systems*, *88*(1), 44–60. https://doi.org/10.1016/j.agsy.2005.06.003
71. Khorram, S., Koch, F. H., van der Wiele, C. F., & Nelson, S. A. C. (2012). *Remote Sensing*. *Remote sensing.* https://doi.org/10.1007/978-1-4614-3103-9
72. Knops, J. M. H., & Tilman, D. (2000). Dynamics of Soil Nitrogen and Carbon Accumulation for 61 Years after Agricultural Abandonment Author (s): Johannes M . H . Knops and David Tilman Reviewed work (s): Source : Ecology , Vol . 81 , No . 1 (Jan ., 2000), pp . 88-98 Published by : Ecolog. *Ecology*, *81*(1), 88–98.
73. Lambin, E. F., Geist, H. J., & Rindfass, R. R. (2006). Introduction: Local processes with global impacts. In *Land-Use and Land-Cover Change: Local Processes and Global Impacts* (pp. 1–8). Springer-Verlag Berlin Heidelberg.
74. Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., & Bégué, A. (2017). A combined Random Forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sensing*, Under Revision. https://doi.org/10.3390/rs9030259
75. Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology : The Robert H . MacArthur Award Lecture. *Ecology*, *73*(6), 1943–1967.
76. Li, Q., Wang, C., Zhang, B., & Lu, L. (2015). Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data. *Remote Sensing*, *7*(12), 16091–16107. https://doi.org/10.3390/rs71215820
77. Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 18-22.
78. Liu, X., & Bo, Y. (2015). Object-based crop species classification based on the combination of airborne hyperspectral images and LiDAR data. *Remote Sensing*, *7*(1), 922–950. https://doi.org/10.3390/rs70100922
79. Lombardia, R. (2017). Geoportale Regione Lombardia - Servizio Download Dati. Retrieved March 22, 2017, from http://www.geoportale.regione.lombardia.it/download-ricerca
80. MacDonald, D., Crabtree, J. ., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., … Gibon, a. (2000). Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. *Journal of Environmental Management*, *59*(1), 47–69. https://doi.org/10.1006/jema.1999.0335
81. Macgregor, C. J., & Warren, C. R. (2006). Adopting sustainable farm management practices within a Nitrate Vulnerable Zone in Scotland: The view from the farm. *Agriculture, Ecosystems and Environment*, *113*(1–4), 108–119. https://doi.org/10.1016/j.agee.2005.09.003
82. Marceau, D. J., & Geoffrey, J. H. (1999). Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, 25, n.4, 10.
83. Martin, K., & Sauerborn, J. (2013). Introduction. In *Agroecology*. Springer Science+Business Media.
84. Masvaya, E. N., Nyamangara, J., Nyawasha, R. W., Zingore, S., Delve, R. J., & Giller, K. E. (2010). Effect of farmer management strategies on spatial variability of soil fertility and crop nutrient uptake in contrasting agro-ecological zones in Zimbabwe. *Nutrient Cycling in Agroecosystems*, *88*(1), 111–120. https://doi.org/10.1007/s10705-009-9262-y
85. Matton, N., Canto, G. S., Waldner, F., Valero, S., Morin, D., Inglada, J., … Defourny, P. (2015). An automated method for annual cropland mapping along the season for various gobally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sensing*, *7*, 13208–13232. https://doi.org/10.3390/rs71013208
86. Mayer, J., Buegger, F., Jensen, E. S., Schloter, M., & Heß, J. (2003). Residual nitrogen

contribution from grain legumes to succeeding wheat and rape and related microbial process. *Plant and Soil*, *255*(2), 541–554. https://doi.org/10.1023/A:1026081015076
87. Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Sayre, R., Trabucco, A., & Zomer, R. (2012). A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, *22*(5), 630–638. https://doi.org/10.1111/geb.12022
88. Mücher, S., De Simone, L., Kramer, H., de Wit, A., Roupioz, L., Hazeu, G., … Cormont, A. (2016). A new Global Agro-Environmental Stratification (GAES). *Wageningen Environmental Research*, *Report 276*. https://doi.org/10.18174/400815
89. Mugandani, R., Wuta, M., Makarau, A., & Chipindu, B. (2012). Re-classification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. *African Crop Science Journal*, *20*(Supplement s2), 9.
90. Murray, H., Lucieer, A., & Williams, R. (2010). Texture-based classification of sub-Antarctic vegetation communities on Heard Island. *International Journal of Applied Earth Observation and Geoinformation*, *12*(3), 138–149. https://doi.org/10.1016/j.jag.2010.01.006
91. NASA. (2016). Sentinel-2A Launches—Our Compliments & Our Complements. Retrieved December 28, 2016, from http://landsat.gsfc.nasa.gov/sentinel-2a-launches-our-compliments-our-complements/
92. Olesen, J. E., & Bindi, M. (2002). Consequences of climate change for European agricultural productivity, land use and policy. *European Journal of Agronomy*, *16*(4), 239–262. https://doi.org/10.1016/S1161-0301(02)00004-7
93. Oumar, Z., Mutanga, O., & Ismail, R. (2012). Predicting thaumastocoris peregrinus damage using narrow band normalized indices and hyperspectral indices using field spectra resampled to the hyperion sensor. *International Journal of Applied Earth Observation and Geoinformation*, *21*(1), 113–121. https://doi.org/10.1016/j.jag.2012.08.006
94. Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698
95. Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, *115*(6), 1301–1316. https://doi.org/10.1016/j.rse.2011.01.009
96. Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V., & Justice, C. O. (2010). Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, *2*(7), 1844–1863. https://doi.org/10.3390/rs2071844
97. Prishchepov, A. V., Radeloff, V. C., Dubinin, M., & Alcantara, C. (2012). The effect of Landsat ETM/ETM+ image acquisition dates on the detection of agricultural land abandonment in Eastern Europe. *Remote Sensing of Environment*, *126*, 195–209. https://doi.org/10.1016/j.rse.2012.08.017
98. QGIS Development Team. (2017). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from http://www.qgis.org/
99. Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1
100. R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/
101. Regione Lombardia. (2015). Particelle agricole Provincia di Lodi(dati generali). Retrieved September 13, 2016, from https://www.dati.lombardia.it/Agricoltura/Particelle-agricole-Provincia-di-Lodi-dati-general/hvwv-fgj3Opens in new window.
102. Reis, S., & Taşdemir, K. (2011). Identification of hazelnut fields using spectral and gabor textural features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(5), 652–661. https://doi.org/10.1016/j.isprsjprs.2011.04.006
103. Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, *121*, 93–107.

https://doi.org/10.1016/j.rse.2011.12.003
104. Roth, K. L., Dennison, P. E., & Roberts, D. A. (2012). Comparing endmember selection techniques for accurate mapping of plant species and land cover using imaging spectrometer data. *Remote Sensing of Environment*, *127*, 139–152. https://doi.org/10.1016/j.rse.2012.08.030
105. Schiewe, J. (2002). SEGMENTATION OF HIGH-RESOLUTION REMOTELY SENSED DATA - CONCEPTS , APPLICATIONS AND PROBLEMS. In *Symposium on Geospatial Theory, Processing and Applications, Ottawa*.
106. Schmidt, M., Pringle, M., Devadas, R., Denham, R., & Tindall, D. (2016). A Framework for Large-Area Mapping of Past and Present Cropping Activity Using Seasonal Landsat Images and Time Series Metrics. *Remote Sensing*, *8*(4), 312. https://doi.org/10.3390/rs8040312
107. Schultz, B., Immitzer, M., Formaggio, A. R., Sanches, I. D. A., Luiz, A. J. B., & Atzberger, C. (2015). Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sensing*, *7*(11), 14482–14508. https://doi.org/10.3390/rs71114482
108. Schultz, M., Clevers, J. G. P. W., Carter, S., Verbesselt, J., Avitabile, V., Quang, H. V., & Herold, M. (2016). Performance of vegetation indices from Landsat time series in deforestation monitoring. *International Journal of Applied Earth Observation and Geoinformation*, *52*(May 2012), 318–327. https://doi.org/10.1016/j.jag.2016.06.020
109. Seguici. (2015). Seguici - Geoportale per uso agricolo. Retrieved March 22, 2017, from http://servizi.terraria.com/seguici_webgis/index.jsp
110. Seguici. (2017). Seguici. Retrieved March 22, 2017, from http://www.seguici.eu/index.php?option=com_content&view=featured&Itemid=101
111. Shen, M., Chen, J., Zhu, X., Tang, Y., & Chen, X. (2010). Do flowers affect biomass estimate accuracy from NDVI and EVI? *International Journal of Remote Sensing*, *31*(8), 2139–2149. https://doi.org/10.1080/01431160903578812
112. Smaling, E. (1993). An agro-ecological framework for integrated nutrient management, with special reference to Kenya, 250.
113. Small, C., & Lu, J. W. T. (2006). Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis. *Remote Sensing of Environment*, *100*(4), 441–456. https://doi.org/10.1016/j.rse.2005.10.023
114. Taramelli, A., Pasqui, M., Barbour, J., Kirschbaum, D., Bottai, L., Busillo, C., … Small, C. (2013). Spatial and temporal dust source variability in northern China identified using advanced remote sensing analysis. *Earth Surface Processes and Landforms*, *38*(8), 793–809. https://doi.org/10.1002/esp.3321
115. Tso, B., & Mather, M. (1999). Crop discrimination using multi-temporal SAR imagery. *International Journal of Remote Sensing*, *20*(12). Retrieved from http://www.tandfonline.com/doi/abs/10.1080/014311699212119
116. USGS. (2014). Earth Explorer. Retrieved March 24, 2017, from https://earthexplorer.usgs.gov/
117. USGS. (2016a). Landsat 8 (L8) Data Users Handbook - Section 1. Retrieved December 28, 2016, from https://landsat.usgs.gov/landsat-8-l8-data-users-handbook-section-1
118. USGS. (2016b). Landsat Project Description. Retrieved December 28, 2016, from https://landsat.usgs.gov/landsat-project-description
119. USGS. (2016c). What are the band designations for the Landsat satellites? Retrieved December 28, 2016, from https://landsat.usgs.gov/what-are-band-designations-landsat-satellites
120. Valentini, E., Taramelli, A., Filipponi, F., & Giulio, S. (2015). An effective procedure for EUNIS and Natura 2000 habitat type mapping in estuarine ecosystems integrating ecological knowledge and remote sensing analysis. *Ocean and Coastal Management*, *108*, 52–64. https://doi.org/10.1016/j.ocecoaman.2014.07.015
121. Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., … Koetz, B. (2016). Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. *Remote Sensing*, *8*(1), 1–21. https://doi.org/10.3390/rs8010055

122. van Beek, C. L., Elias, E., Yihenew, G. S., Heesmans, H., Tsegaye, A., Feyisa, H., … Mengist, S. (2016). Soil nutrient balances under diverse agro-ecological settings in Ethiopia. *Nutrient Cycling in Agroecosystems*, *106*(3), 257–274. https://doi.org/10.1007/s10705-016-9803-0
123. van Wart, J., Kersebaum, K. C., Peng, S., Milner, M., & Cassman, K. G. (2013). Estimating crop yield potential at regional to national scales. *Field Crops Research*, *143*, 34–43. https://doi.org/10.1016/j.fcr.2012.11.018
124. Van Wart, J., van Bussel, L. G. J., Wolf, J., Licker, R., Grassini, P., Nelson, A., … Cassman, K. G. (2013). Use of agro-climatic zones to upscale simulated crop yield potential. *Field Crops Research*, *143*, 44–55. https://doi.org/10.1016/j.fcr.2012.11.023
125. Verburg, P. H., van de Steeg, J., Veldkamp, A., & Willemen, L. (2009). From land cover change to land function dynamics: A major challenge to improve land characterization. *Journal of Environmental Management*, *90*, 1327–1335.
126. Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., & Brivio, P. A. (2015). In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features. *Remote Sensing*, *7*(10), 12859–12886. https://doi.org/10.3390/rs71012859
127. Vina, A., Gitelson, A. G., Rundquist, D. C., Keydan, G., Leavitt, B., & Schepers, J. (2012). Monitoring maize (Zea mays L.) phenology with remote sensing. *Agronomy Journal*, *96*. https://doi.org/10.3390/rs4082477
128. Vulnera, Z., & Guida, L. (2016). Allegato A - Bollettino nitrati.
129. Whitcraft, A. K., Vermote, E. F., Becker-Reshef, I., & Justice, C. O. (2015). Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sensing of Environment*, *156*, 438–447. https://doi.org/10.1016/j.rse.2014.10.009
130. White, D. H., Lubulwa, G. A., Menz, K., Zuo, H., Wint, W., & Slingenbergh, J. (2001). Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities. *Environment International*, *27*(2–3), 181–187. https://doi.org/10.1016/S0160-4120(01)00080-0
131. White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., … Lauenroth, W. K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, *15*(10), 2335–2359. https://doi.org/10.1111/j.1365-2486.2009.01910.x
132. Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors*, *9*(3), 1768–1793. https://doi.org/10.3390/s90301768
133. Zeng, L., Wardlow, B. D., Wang, R., Shan, J., Tadesse, T., Hayes, M. J., & Li, D. (2016). Remote Sensing of Environment A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. *Remote Sensing of Environment*, *181*, 237–250. https://doi.org/10.1016/j.rse.2016.03.039
134. Zhang, F., Zhang, L. W., Shi, J. J., & Huang, J. F. (2014). Soil moisture monitoring based on land surface temperature-vegetation index space derived from MODIS data. *Pedosphere*, *24*(4), 450–460. https://doi.org/10.1016/S1002-0160(14)60031-X
135. Zhong, L., Hawkins, T., Biging, G., & Gong, P. (2011). A phenology-based approach to map crop types in the San Joaquin Valley, California. *International Journal of Remote Sensing*, *32*(22), 7777–7804. https://doi.org/10.1080/01431161.2010.527397
136. Zomer, R. J., Trabucco, A., Wang, M., Lang, R., Chen, H., Metzger, M. J., … Xu, J. (2014). Environmental stratification to model climate change impacts on biodiversity and rubber production in Xishuangbanna, Yunnan, China. *Biological Conservation*, *170*, 264–273. https://doi.org/10.1016/j.biocon.2013.11.028

Acknowledgments

I would like to thank ISPRA for providing the data on which I worked. In particular, I thank Andrea Taramelli, Emiliana Valentini, Federico Filipponi, Alessandra Nguyen Xuan and Antonella Tornato for giving me the support needed to accomplish this work.

This research was also partially supported by the Alterra Institute of the Wageningen University. I express my sincere gratitude to Sander Janssen, who has wisely guided me through this complex route, and to Sander Mücher, for sharing his knowledge with me: you have set for me a model of researcher that I hope to reach one day.

I am also grateful to Prof. Enrico Bonari and Prof. M. Enrico Pè, for giving me precious advice and for helping me preserving my position at the School.

Finally… there are a lot more people to thank!

First, my family, for supporting me in every decision, even the bad ones.

Thanks to Edgar, for being my side and comforting me in every moment.

Thanks to Francesca, for being a supportive friend and a wonderful person: our ways are going in different directions, but our friendship won't.

Thanks to Giovanni, for being with me every day (how could you do that?) and helping me with the exams: now we are even.

Thanks to Giulia, Daniel, Marchetti, Giovanni Macchi, Ciurli and Giacomo: you have cheered these 5 years up!

Thanks to Michiel, Christina, Lieven, Roeland, Eline, Janna, Alvaro and Moises, for making me feel at home in Wageningen and lending me a shoulder when I needed it.

Lastly, thanks to the School: I received many bad and many good things, but above all I received the opportunity to do all that I have done.