

Masters Program in **Geospatial Technologies**



OPTIMIZING THE ALLOCATION OF FIELD SAMPLES FOR REMOTE SENSING AND MACHINE LEARNING-BASED CROP CLASSIFICATION MODELS BY ASSESSING THE AREA OF APPLICABILITY.

Lucas Casuccio

Dissertation submitted in partial fulfillment of the requirements
for the degree of *Master of Science in Geospatial Technologies*

**OPTIMIZING THE ALLOCATION OF FIELD SAMPLES FOR
REMOTE SENSING AND MACHINE LEARNING-BASED CROP
CLASSIFICATION MODELS BY ASSESSING THE AREA OF
APPLICABILITY.**

Dissertation supervised by:

**Dr. Christian Knoth
Associate Researcher
Westfälische Wilhelms-Universität Münster**

Co-supervised by:

**Prof. Dr. Hanna Meyer
Associate Professor
Westfälische Wilhelms-Universität Münster**

**Prof. Dr. Filiberto Pla
Full Professor
Universitat Jaume I**

February, 2023

| Acknowledgements

I would like to express my deep gratitude to my supervisor and my co-supervisors for accepting to guide me through this endeavor with great patience and most valuable feedback, without which I could not start thinking of accomplishing this work. Thanks should also go to the professors and assistants involved in this program, from whom I had the chance to take lessons that impacted and inspired me.

Special thanks to Prof. Marco Painho, whose kind and undeniable support was there for us all the way from the beginning, and to Dr. Christian Knoth, Prof. Hanna Meyer, Dr. Christoph Brox, Dr. Tessio Novack, Agr. Eng. Eduardo Rollero and Dr. Marvin Ludwig, who generously provided knowledge, expertise and support during crucial stages of this journey.

Also, I could not have undertaken this challenge without my classmates, whom I would like to thank for their companionship and help, moral and spiritual support, and for settling this journey into a great and memorable milestone.

Lastly, I would be remiss in not mentioning my parents, Alejandra and Tito, my brothers Pablin, Mariano and Paolo, and my friends Gonzalo, Luis, Carlos, André, Guillem, Julia, Jaskaran, Mara, Ayman; their support and belief in me kept my spirits and motivation high during this process. Thank you folks, for your friendship, patience and kindness.

OPTIMIZING THE ALLOCATION OF FIELD SAMPLES FOR REMOTE SENSING AND MACHINE LEARNING-BASED CROP CLASSIFICATION MODELS BY ASSESSING THE AREA OF APPLICABILITY.

| Abstract

Reliable and accurate crop classification information can provide key insights for food security, production sustainability, agronomic decisions, logistics, and yield estimations; serving several different sectors and industries. Since decades ago, crop classification has been escalated by the use of different remote sensing sources and machine learning techniques, creating prediction models that could retrieve information from continuous areas and classify them according to the ground-truth data used as references to identify and map the targeted crops. However, the process of ground-truth sampling is nowadays the most challenging part resources-wise, since its obtaining implies gathering reference information through field surveys and crowdsourcing. Alternatively, efforts to reduce or guide ground-truth necessity have been studied, by using limited or transferred reference data, which implies pre-existent databases availability and state-of-the-art data extraction workflows. This study explores the Area of Applicability (AOA) assessment for validating crop classification model predictions by identifying areas outside the estimated prediction limits and providing different levels of confidence for the prediction made based on the training data and the variables presented to the model. The outcomes could provide a variant for areas that do not hold or provide open access to historical ground-truth data collections, to help optimize ground-truth samples allocation for crop classifications, by offering a potential guide for indicating areas that need further samples collection for presenting reliable predictions over these extensions, and could potentially bring further insights for validation of practices involving prediction models' spatiotemporal transferability.

Keywords: Area of Applicability, spatial analysis, crop classification, remote sensing, machine learning, ground-truth samples.

| Contents

List of Figures	viii
List of Tables	x
1. Introduction	1
1.1 Crop Production	1
1.2 Crop Classification Models	2
1.2.1 Remote sensing	2
1.2.2 Machine Learning algorithms.	4
1.2.3 Ground-truth samples	5
1.3 Increasing efficiency of ground-truth collection	6
1.4 Area of Applicability (AOA)	7
1.5 Objectives and research questions	9
2. Methodology	10
2.1 Study Area	10
2.2 Methodology Outline	11
2.3 Ground-truth data	13
2.4 Remote sensing variables	15
2.4.1 Reflectance predictors (bands)	15
2.4.2 Index predictors (vegetation indices)	17
2.5 Machine Learning classification algorithm	19
2.6 Traditional accuracy assessments	20
2.7 Area of Applicability	20
2.7.1 Cross validation strategy	21
2.7.2 Dissimilarity index	23

2.7.3 Area of Applicability assessment	24
2.7.4 Area of Applicability Calibration	25
3. Results	27
3.1 Alto Paraná Department prediction and assessment	27
3.1.1 Classification	27
3.1.2 Area of Applicability - Alto Paraná 2020/2021	28
3.2 ROI prediction (Spatial transfer validation)	31
3.2.1 Classification	31
3.2.2 Area of Applicability - ROI 2020/2021	32
3.2.3 Classification - ROI 2020/2021 + extra samples	35
3.2.4 Area of Applicability - ROI 2020/2021 + extra samples	36
3.3 Alto Paraná Department prediction with different year variables (Time transfer validation)	40
3.3.1 Classification	40
3.3.2 Area of Applicability - Alto Paraná (Time transfer validation)	42
4. Discussion	45
4.1 AOA assessment and ground-truth allocation.	45
4.2 Other methodologies considerations	46
4.3 Limitations	50
4.4 Future Scope	51
5. Conclusion	53
Bibliography	54

| List of Figures

1	Study area departments located in Paraguay and details of their national and departmental roads distribution, these departments compound the ROI.	11
2	Methodology outline workflow	12
3	Ground-truth samples represented as polygons in the study area	13
4	Comparison of the performance during the prediction task contemplating different cross-validation strategies including (a) random cross-validation and (b) spatial cross-validation	22
5	AOA calculation plotting the AOA threshold and the relation between the prediction DI and the train DI, showing how they adjust to each other and the parts of them that fall outside or inside the AOA, according to the maximum tier of the DI calculated over the training data	24
6	AOA calibration result plotting the Accuracy and the DI for the different cross-validations applied to the model prediction.	26
7	Classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and remote sensing variables	28
8	Area of Applicability (AOA) of the classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and remote sensing variables	29
9	Calibrated AOA over Alto Parana department	30
10	Classification model prediction for the ROI departments (Alto Paraná, Canindeyú and Itapúa) with 2020/2021 summer cycle ground-truth samples from Alto Paraná and remote sensing variables from the entire ROI	32
11	Area of Applicability (AOA) of the classification model prediction for the ROI with 2020/2021 summer cycle ground-truth samples and remote sensing variables	33
12	Calibrated AOA over the ROI (Alto Parana and neighboring departments) with Alto Paraná 2020/2021 summer cycle ground-truth	34
13	Classification model prediction for the ROI departments (Alto Paraná, Canindeyú and Itapúa) with 2020/2021 summer cycle ground-truth samples from Alto Paraná and from the neighboring departments as well and remote sensing variables from the entire ROI	35
14	Comparison of ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas.	36

15 Area of Applicability (AOA) of the classification model prediction for the ROI with 2020/2021 summer cycle Alto Paraná + neighboring departments ground-truth samples and remote sensing variables.	37
16 Comparison of the AOA assessment results over the ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas	38
17 Calibrated AOA over the ROI (Alto Parana and neighboring departments) including Alto Paraná 2020/2021 summer cycle ground-truth, and added samples of neighboring areas	39
18 Comparison of the AOA calibration results over the ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas	40
19 Classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables	41
20 Area of Applicability (AOA) of the classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables.....	42
21 Calibrated AOA results for the classification model prediction of Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables	43

| List of Tables

1	Population distribution in pixels of ground-truth sampling references from Alto Paraná department - 2020/2021 summer cycle	14
2	Population distribution in pixels of ground-truth sampling references from Alto Paraná department + neighboring areas (complete ROI) - 2020/2021 summer cycle	14
3	Sentinel 2 bands included as predictors for the classification model	16
4	Vegetation indices included as predictors for the classification model	17

Introduction

1.1 Crop production

Crop production is a key source of food for human and animal nutrition, as well as a provenance of different raw materials for several industrial products such as biofuel, textiles, and paper, representing a foremost role in economic development in many countries, especially in developing nations (FAO, 2022). These benefits offered by crop production are in constant growth for demand, as the global human population reached an unprecedented number of above 8 billion inhabitants, as estimated by the United Nations in their last summary of World Population Prospects (UN DESA, 2022). By also considering the percentage of this total world population that is estimated to be living without daily access to basic nutrition supplies (FAOSTAT, 2022), and in addition to a combination of other major challenges that include climate change, it becomes essential to keep looking forward to improvements in the efficiency and sustainability of crop production worldwide. It is crucial to maintain and enhance the effectiveness of practices that can provide a benefit or advantage to the vast domain of crop production, which is composed of different parts that interact and which impact transcend to several different industries. Accordingly, crop production is a global-scale need that claims for adaptive solutions and sustainable practices that include the efficient use of resources as a premise.

Crop monitoring is a field of studies and practices that serves benefit to crop production by providing spatial information for analyzing agricultural fields (Wu et al, 2014), involving the use of a variety of techniques that can include remote sensing data provided by aerial imagery, satellite imagery, drone-based imagery, and ground-truth measurements to gather data about the crops locally for interpretation and validation of

their status (Boryan et al, 2011). This data can be then analyzed to identify any issues such as pest infestations, soil nutrient deficiencies, and irrigation problems, which can then be addressed to improve crop health and productivity; contributing to the evolution of production systems that through constant survey can provide data for identifying potential threats and solutions.

1.2 Crop Classification Models

Inside crop monitoring different tasks, the process of detecting and quantifying crops over surface extensions is normally referred to as crop classification. Reliable and accurate crop classification information can provide key insights for food security, production sustainability, agronomic decisions, logistics, and yield estimations (Xie et al, 2008; Liu, 2015; Anderson et al, 2017); serving several different sectors and industries that find an advantage by integrating this information to their considerations for decision making (Hendricks & Er, 2015; Waldner et al, 2015)

1.2.1 Remote sensing

Since decades ago, and along with other types of land use classification mapping, crop classification estimations have been impulse by using models based on remote sensing (Bauer, 1975; Odenweller & Johnson, 1984) providing the advantage of registering and analyzing broad extensions of continuous space, which characteristics are assessed and registered by passive or active sensors using energy emissions reflected from the surface of the targeted area and measuring the radiation, absorption, back-scatter or reflectance of electromagnetic, force or acoustic waves information (Quihao, 2013). In case of optical sensors, the information retrieved by the measurements is stored and classified as different bands, regarding the different wavelengths registered by the sensors. including imagery collected by different sources such as handheld sensors, sensors mounted on aircraft or spacecraft (Prasad et al. 2011). This optical data is normally represented as pixels that contain the measured data. Along with the bands used to gather and classify the optical spatial data, there are indices that can be calculated by combining these different bands

in search of highlighting features on the analyzed covers depending on the target to evaluate. In case of vegetation monitoring, indices are mostly calculated based on the differential response of vegetation in near infrared (NIR) and red (R) portions of the electromagnetic spectrum, or other portions of it that correspond to the visible (RGB) part (Bannari et al, 1995; Nijland et al, 2014). Integrating these different remote sensing techniques into crop monitoring brought an exponential growth in terms of data compilation and analysis capacity (Ma et al, 2015; Comber & Wulder, 2019). Considering the use of time-series is also of value when it comes to selecting remote sensing variables, as Chang et al 2007 discussed, taking measurements that consider not only spectral but also temporal information from crops during the growing seasons enable models to extract features from time-series data to retrieve valuable information, utilizing the data produced from detecting the characteristically seasonal patterns that different crops can display.

Different types and sources of remotely sensed information can be integrated into crop classification models as variables to perform predictions (Fawwaz, 1982; Park & Park, 2015; Viskovic et al, 2019). A very distributed source of imagery for crop classification models is satellite imagery, due these space borne systems' capacity to cover and revisit large extensions of the globe regularly, and also due the possibility of accessing some open sources for downloading and manipulating this data without representing necessarily a cost for the user. Satellites can hold different types of sensors that gather data from the surface of the globe with different time, space and spectral resolutions or frequencies, depending on the mission. Registers indicate that crop classifications can be assessed by different satellites with their different sensors, including the use of satellites with optical multispectral sensors such as Sentinel 2, Landsat, MODIS, Planet, among others (Kobayashi et al, 2020; Gilbertson et al, 2017; Dell'Acqua et al, 2018; Xang et al, 2021), Radio Detection And Ranging (RADAR) sensors, such as Sentinel 1 (Son et al, 2021), and also by combining these different sources (Orynbaikyzy et al, 2019, Rao et al, 2021) and including other less common combination variations including airborne based remote sensing with sensors that provide data such as hyperspectral optical imagery (Yang et al, 2013; Zhang et al, 2016) or Light Detection and Ranging (LiDAR) imagery (Jahan & Awrangjeb, 2017, Prins & van Niekert, 2020).

1.2.2 Machine Learning algorithms

Along with remote sensing, machine learning is another technique which by being included into geospatial analysis has significantly increased the capacity of exponentially escalating the reach of general remotely sensed-based analysis and classifications (Lary et al, 2016, Maxwell et al, 2018). The use of different machine learning supervised, semi-supervised or unsupervised classification algorithms techniques including Random Forest (Breiman, 2001), Support Vector Machine (Cortes & Vapnik, 1995), Artificial Neural Networks (McCulloch & Pitts, 1943) and their variations, allowed the construction of sophisticated workflows that escalated predictions reach, allowing to include several layers of different data sets and finding relations within them. In the case of crop classification, where crop covers are the aim for prediction, techniques have been largely explored through the integration of these machine learning algorithms enabled the use of models that could retrieve information over continuous areas, to then classify this extensions according to the matching ground-truth references to identify and map the targeted crop areas extensions (Perales et al, 2003; Mathur & Foody, 2008; Ok et al, 2011; Kang et al, 2018; Campos-Taberner et al, 2020; Chakhar, 2020; Siesto et al, 2021; Luo et al, 2023; Gallo et al, 2023). Random Forest is the machine learning method selected for this work due its ability of fitting non-linear relations and its robustness when it comes to dealing with noise on the data input, maintaining stability even when the amount of error introduced in the classification is high (Su et al, 2019).

As described, crop classifications can use different methods and obtain different results, varying with the type of imagery used and their different combinations, the machine learning techniques used for predictions, which accuracies are then measured by different types of assessments and validations. In case of this work, the traditional accuracy assessment will be performed by calculating the overall accuracy, which is estimated by the amount of predictions that were classified correctly and it has been a measure for accuracy assessments in remote sensing classification since decades ago (Congalton, 1991), along with the kappa index (Cohen, 1960) calculation which compares the overall accuracy with random chances, adding another

layer of assessment to the prediction. The combination of these two statistical tools for accuracy assessment of a classification model based on remote sensing data is studied by Fitzgerald & Lees, 1994; where they demonstrate that overall accuracy by itself could mislead interpretations on prediction model performances.

1.2.3 Ground-truth samples

Despite the mentioned importance and contribution that remote sensing and machine learning have as fundamental parts of classification models, these workflows are highly dependent on the reference data used as samples for training and/or validating the model. These samples are commonly known as ground-truth references, which as Hoffer, 1972 defined, involves the collection of relevant observations in the frame of the concerning properties that are intended to be used in order to guide or validate the classification algorithms, to enable the model to learn about the different data that these references hold to then analyze the features from the rest of the predictor variables and extend the estimations to areas that contain continuous information but are not referenced and, are therefore unknown.

Sparse and insufficient ground-truth referenced data is still a detrimental factor for crop classification modeling (Song et al, 2017), and the process of collecting this data for most classification tasks is arguably the most challenging part of the workflow process regarding time and resources (Ali et al, 2022). In addition to the mentioned importance of ground-truth, Rußwurm et al, 2020 mentioned that the bottleneck for producing high accuracy classification maps shifted from a lack of remote sensing data or processing capacity towards the lack of ground truth references. Thus, it is valid to study the possibilities of generating new results based on different years' samples or areas to eliminate (or reduce as much as possible) the on-field surveys to collect the reference data or fill gaps between non-sampled cycles or areas.

The training data available for this study was collected in Paraguay, which is one of the main producers of commodities crops in South America (CAPECO, 2021) where crop and livestock production reaches 17%

of the GDP of the nation. Sampling efforts were focused in the easternmost region of the country, particularly in the Alto Paraná department.

1.3 Increasing efficiency of ground-truth collection

The collection of ground-truth references is commonly accomplished through field surveys or crowdsourcing. Collecting these sampling data implies resources destined for that task to reach and monitor fields year after year, ideally including remote areas. This approaches could represent a challenge that raises special concern regarding the consideration of the resources and time destined to conduct the collection of samples accurately (Zhong et al, 2012; Fowler et al. 2020; Kluger et al, 2021), especially in regions where the accessibility to the rural areas is not particularly well developed in terms of infrastructure and consequently increases the difficulty of the surveying tasks, considering that these areas are normally where most of the crop production is based among countries,.

In most recent years, studies were developed based on crop classification model predictions application while including the analysis and validation of techniques that are meant to reduce the need for new in-situ sampling of ground-truth, or guiding the efforts to collect these samples. These studies were based on methodologies that used different resources for generating or guiding the collection of ground-truth to serve as training samples, including approaches that use crop-phenology-based pixel identification (Massei et al, 2017; Belgiu et al, 2021), or previously collected and validated data sets of samples that are transferred through space (Pengyu et al, 2020; Huang et al, 2020;), by using samples of areas that do not correspond to the assessed study area or, transferred through time (Ji et al, 2018; Li et al, 2021; Liu et al, 2022; Hu et al, 2022), which is based on including training data from the same study area but corresponding to different years (or cycles, in case of crops) as an alternative to yearly in situ registers of samples and by using information from other models and exploring classification predictions training and validation using different years' or different areas' training data as ground-truth references to search for an alternative

increase efficiency of the use of ground-truth samples by generating or guiding ground-truth collection or identification by using existing sources, including methods based on both time and space ground-truth samples transferring (Wang et al, 2019; Lin et al, 2022; Pott et al, 2022).

However, these techniques and approaches are made possible by architectures of unsupervised classifiers, including neural networks, which are fed by pre-existent databases of ground-truth collected by private (Bingfang, 2020) or government-funded projects, such as the EU for its Common Agricultural Policy (CAP) (Sitokonstantinou et al, 2018; Defourny et al, 2019; Koetz et al, 2019), or the Crop Detection Layer (CDL) supported by the US Department of Agriculture (Boryan et al, 2011). These databases are not always exempt of limitations due to errors and noise (Arias et al. 2020) since some were referenced by declarations on surveys.

Thus with these efforts, and considering the dependence of the mentioned methodologies on large previously existing datasets and complex architectures of neural networks and other classifiers, it is valid to analyze methodologies that could offer alternatives for reducing or directing sampling efforts, particularly in cases of regions that do not hold organized registers of locally assessed ground-truth databases by different crop cycles or where such information is not openly available.

1.4 Area of Applicability (AOA)

The Area of Applicability (AOA) (Meyer & Pebesma, 2021) is defined by its authors as the area where we enabled a model to learn about relationships based on the presented training data, and where the estimated cross-validation performance holds. This spatial statistics analysis tool is available in the CAST package (Meyer et al, 2023) of the environment and programming language R (R Core Team, 2023) and provides a limit for predictions to be defined inside areas where through relations of calculated dissimilarity between the samples and cross-validations, assess a map for determining where the model is confident and where it is not among the analyzed extension of area covered by the classification. The AOA assessment was

developed and tested on global scale model predictions (Meyer & Pebesma, 2022; Ludwig et al, 2023), and there is also registers of its usability for classification model prediction restricted to a certain study area at regional or smaller scales (Dolan et al, 2021), as a spatial accuracy validation option that could provide information on how applicable and where is the analyzed classification model prediction.

Crop classification models are useful for providing important information while predicting over continuous areas, since these techniques enable the use of complex relations of different data to provide results based on the presented variables and ground-truth reference data. However, there is always a level of uncertainty on these classification predictions, since they are made over continuous areas that are not completely referenced with ground-truth data, and which features could be unknown to the classification model. Commonly, the reliability of these predictions are expressed by traditional accuracy assessment validations as the ones mentioned previously, which provide a general validation of the model, normally with no spatial reference to interpret the obtained values. Hence, crop classification model predictions usability could benefit from more precise and explicit spatial validation tools that could improve the interpretation of classification models outcomes, allowing not only a more insightful view of the model performance, but also potentially revealing detailed geospatial information regarding the aptitude of the training data presented to the classification model and provide a performance confidence measure in relation to the extension of area that the model attempts to predict over.

The present work intends to approach the previously mentioned ground-truth efficiency gap by attempting to generate results that could potentially detect the necessity of sample collection by calculating the Area of Applicability (AOA) of supervised learning model predictions trained with limited ground-truth, by identifying areas where the model is not reliable in the analyzed space and potentially indicate extensions within the study area that hold information that is presumably unknown to the model and would need further ground-truth sample references to provide a reliable classification over these continuous space knowledge gaps.

1.5 Objectives and research questions

This work has the objectives of assessing the reliability of a crop classification model which was based on a limited amount of samples, assessing the transferability of a crop classification model while extrapolating the prediction to neighboring areas, assess the transferability of a crop classification model in a different year, and prove if the methodology has potential to guide collection of new ground-truth samples based on a previously detected necessity of sampling.

Leading to the following research questions:

- Can the AOA assessment help to identify areas where training samples are needed?
- How transferable can a trained model be for guiding new sampling data collection on neighboring non-sampled areas?
- How transferable can a model be through different years?

The outcomes should present indicators on how the predictions of the model perform over a originally sampled study area, then spatially transferring the model expanding the study area by including neighboring areas without samples, to then include ground-truth samples that do correspond to those new areas and see how the classification prediction and the performance assessments varies. A third model will be tested over the originally sampled study area but transferring the samples through time for a new classification, to then again evaluate performances of the prediction and the impact on the AOA.

Methodology

2.1 Study Area

The study area or region of interest (ROI) covers the easternmost part of Paraguay, including the departments of Alto Paraná, Itapúa and Canindeyú (Figure 1). With a size of approximately 4.506.400 hectares combined, these three departments are located in the Oriental Region of Paraguay and share borders with Brazil and Argentina. Entirely located between 23.5° and 27.3° South and between 54.3° and 55.3° West. The climate is moderate, with the fresh season running from May to July and the warm season from October to March. These departments hold over 50% of the extensive crop production areas in the country and during the summer cycle they are mostly destined to soybeans, corn, among other crops such as rice in Itapúa and sugarcane in the limits of Alto Paraná and Canindeyú, also including and pasture areas, forest reserves, and natural high and lowlands.

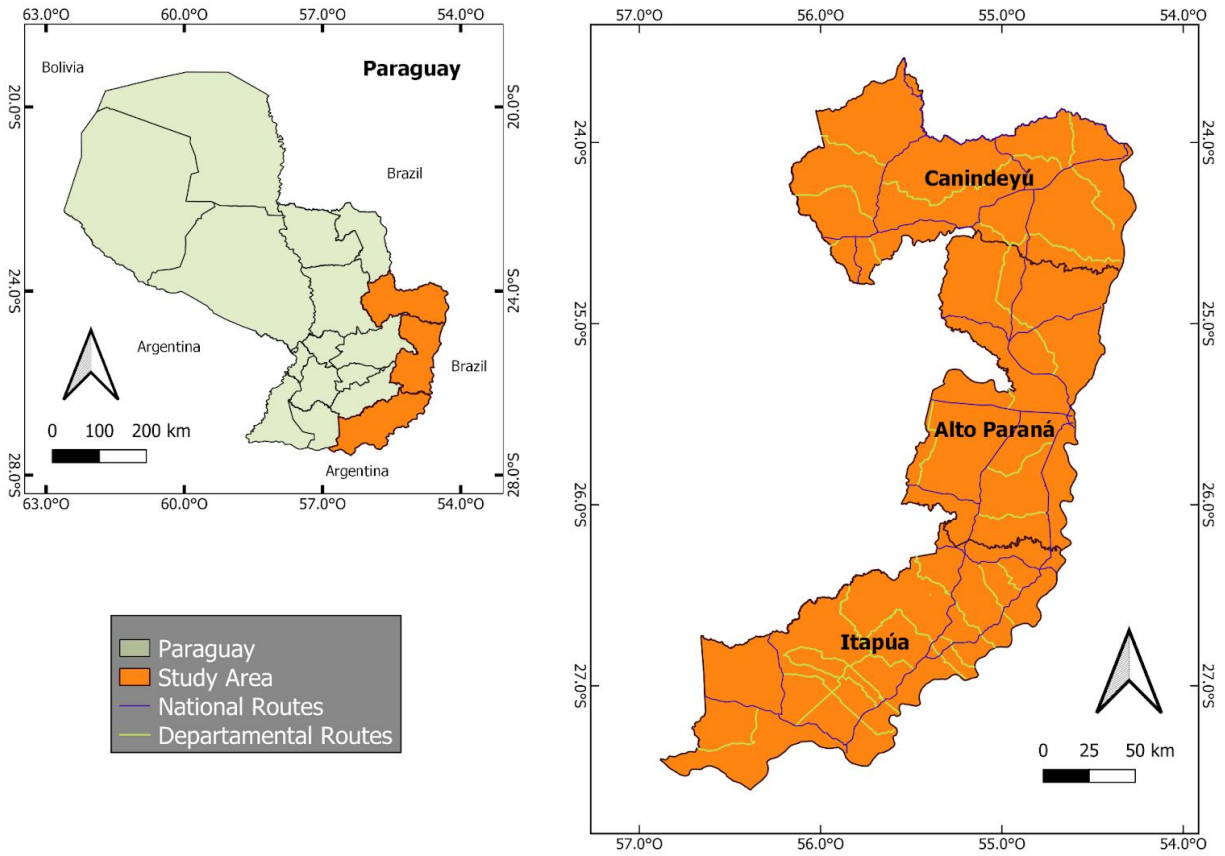


Figure 1. Study area departments located in Paraguay and details of their national and departamental roads distribution, these departments compound the ROI.

2.2 Methodology Outline

An overview of the methodology can be separated into three different stages (Figure 2). The initial stage starts with the variables, which are the ground-truth samples and the remote sensing predictors. Three different sets of predictors were independently prepared to attempt to answer each research question. Data extracted from these variables will be introduced in a machine learning-based classification model as training data for the second stage of the outline, while also performing cross-validations over these training data. After obtaining each classification predictions, these will be validated by using traditional accuracy

measurements along with the AOA assessment, which outcomes will be with an AOA layer to indicate where the model is applicable over the predicted area and a AOA calibration layer for showing levels of confidence regarding the model predictions applicability over the predicted area.

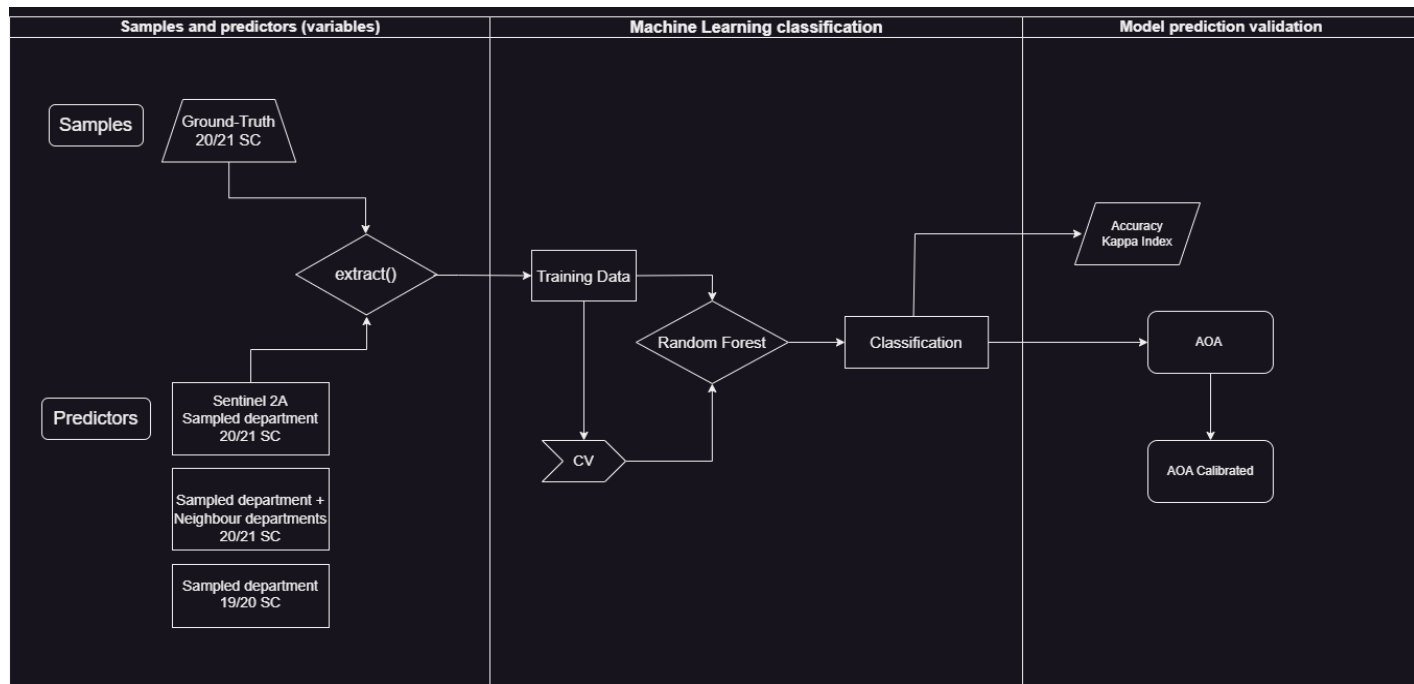


Figure 2. Methodology outline workflow.

AOA and AOA calibration assessments over the prediction outcomes will be the final step to provide results which will be determined by the dissimilarity between the training data provided to the model as training data and their cross-validations, to set an AOA threshold and therefore margin or differentiate the areas where that hold information that differs beyond the dissimilarity index. These areas will be represented on maps as validation assessments that could identify potential areas that would need further sampling for offering a confident prediction over them.

2.3 Ground-truth data

The ground-truth data samples used to train the model were collected from the region of interest (ROI) during the months of October 2020 and January 2021, by registering and labeling field data on site. This collection of samples, represented as black polygons over the map in Figure 3, was adjusted to the extension of the National or Departmental routes due to the simplification of time and resources that this simplified way of sampling collection meant considering also the infrastructural availability to access some areas among the ROI.

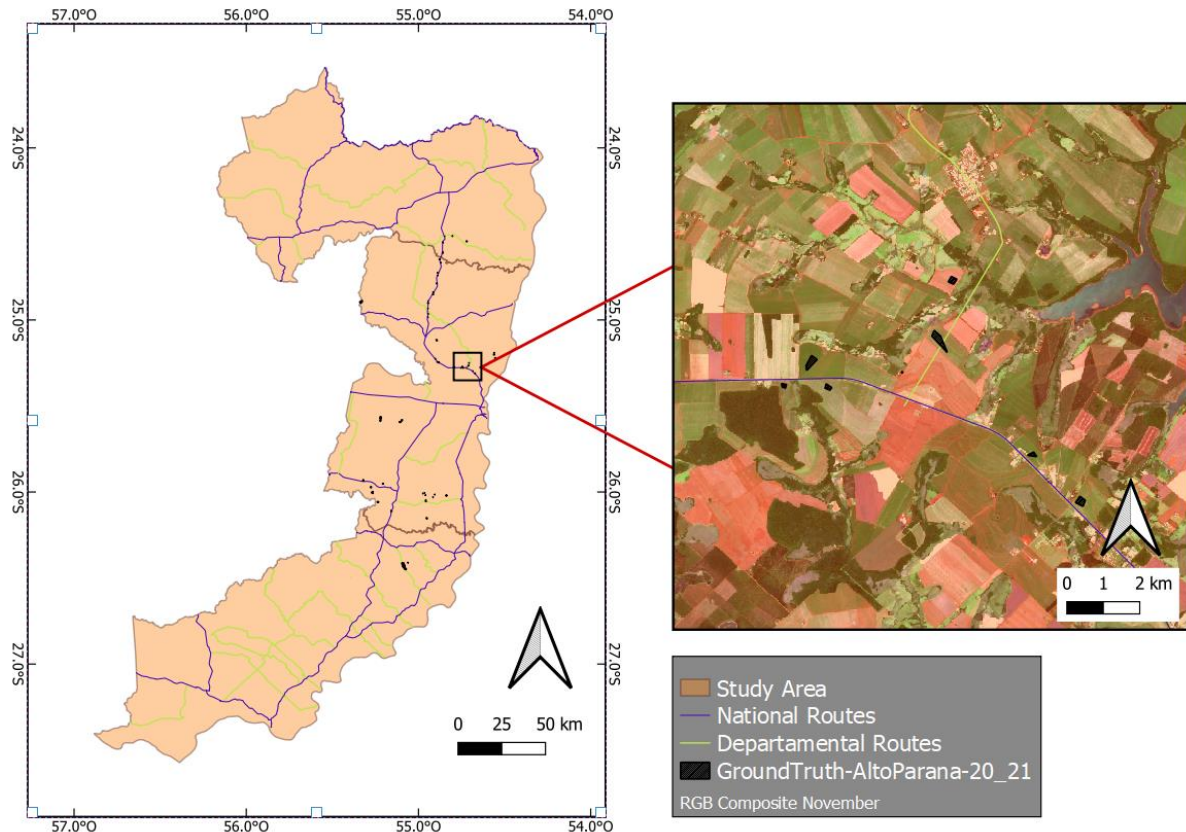


Figure 3. Ground-truth samples represented as polygons in the study area.

The population of samples was distributed as described in Table 1. and the samples used as ground-truth references were collected in the Alto Paraná department. The samples showed in Table 1. represent the samples used for training the classification model in the predictions made over Alto Paraná, including the

classifications that sought to study how the prediction and validations vary when spatially transferring by expanding the classification including neighboring areas (Canindeyú and Itapúa departments), that along with Alto Paraná department compound the ROI. Only 3 classes were categorized including soybean, maize and other, which represents natural fields and lowlands.

Pixels quantity distribution by Class		
Soybean	Maize	Other (natural/lowlands)
183.000	109.000	147.000

Table 1. Population distribution in pixels of ground-truth sampling references from Alto Paraná department - 2020/2021 summer cycle.

By excluding the samples from these neighboring areas and using only the samples inside Alto Paraná, we can then contrast the classification and validation results by adding samples from the neighboring areas as well. After including the samples from the neighboring areas, the population distribution is altered (Table 2).

Pixels quantity distribution by Class		
Soybean	Maize	Other (natural/lowlands)
202.000	312.000	147.000

Table 2. Population distribution in pixels of ground-truth sampling references from Alto Paraná department + neighboring areas (complete ROI) - 2020/2021 summer cycle.

The ground-truth sample distribution used in this work is limited and does not follow any previously assessed distribution structure or collection plan besides being adjusted to accessibility of road infrastructure connections.

2.4 Remote sensing variables

All remote sensing variables selected as predictors are derived from Sentinel 2 imagery including bands and indices clipped to the study area. The Sentinel 2A satellite was launched on June 23, 2015 and Sentinel 2B on March 7, 2017, both have the same specifications and collect multispectral data including 13 bands, covering from visible wavelength regions to shortwave infrared bands (SWIR) wavelength regions.

The selection of bands and indices is based on a previously existing crop classification model of private use. The predictors are separated in three different sets and presented independently to the model for each of the attempted validations. The first set contemplates only the Alto Paraná department, the second set is for the ROI (Alto Paraná + neighboring departments of Itapúa and Canindeyú) and the third set is only Alto Paraná but with images corresponding to the 2019/2020 summer cycle.

2.4.1 Reflectance predictors (bands)

The Sentinel 2 imagery to generate the predictors were acquired and stacked within 7 different time frames or intervals, regarding the availability of images with no more than 15% of cloud coverage. A less permissive cloud filter would have marginalized an important amount of data for covering the complete ROI in the different time frames. The decision of gathering these time series within different time frames is based on avoiding confusion while labeling samples with the same crops but different phenology among the study area, given that the summer cycle crops are sowed mostly during late August/early September and harvested in late February/early March of the contiguous year. The 7 time frames were fixed as '2020-

10-01' - '2020-10-21', '2020-10-21' - '2020-11-02', '2020-11-03' - '2020-11-15', '2020-11-16' - '2020-12-27', '2020-12-20' - '2021-02-12', '2021-02-15' - '2021-02-20'

The previously mentioned time frames of imagery result as one image per interval. By merging them, they are grouped in order to ensure the generation of a complete and usable image (due to partial coverage regarding satellite path and cloud presence) to be used as predictors. From each composite image representing each of the 7 time frames, the bands 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8A', 'B8', 'B11', 'B12' are selected (Table 3). The composite images are then stacked and clipped to the study area and exported as 7 different time intervals.

Sentinel 2 Bands	Central Wavelength (µm)	Resolution (m)
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - VRE	0.705	20
Band 6 - VRE	0.740	20
Band 8 - NIR	0.842	10
Band 8A - VR	0.865	20
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Table 3. Sentinel 2 bands included as predictors for the classification model.

The selection of bands was made based on a previously available model, including RGB but also several bands from the different infrared parts of the spectrum, commonly used for monitoring vegetation.

2.4.2 Index predictors (vegetation indices)

To include more variables, vegetation indices (Table 4) are generated for each time series and merged together into one image per index. The index predictors are then not assessed by interval but by index, where every index is compiled separately by merging each image generated that represents a time interval in the previous step. This method implies that every individual image exported compiles the merged and averaged values of each time interval determined for each different index calculated and clipped to the study area.

These index predictors are compounded by 8 different vegetation indices (Table 4)

Name	Acronym	Sentinel 2A Bands combination	Author
Enhanced Vegetation Index	EVI	$2.5 * (B81 - B4) / ((B8 + 6 * B4) - 7.5 * B2) + 1$	Huete et al, 2002
Green Chlorophyll Vegetation Index	GCVI	$(B8 / B3) - 1$	Gitelson et al, 1996
Green Normalized Difference Vegetation Index	GNDVI	$(B8 - B3) / (B8 + B3)$	Gitelson et al, 1996

Modified Soil Adjusted Vegetation Index	MSAVI	$(2 * B8 + 1 - \sqrt{\text{pow}((2 * B8 + 1), 2) - 8 * (B8 - B4)}) / 2$	Qi et al, 1994
Normalized Difference Vegetation Index	NDVI	$(B8 - B4) / (B8 + B4)$	Rouse et al, 1974
Red Edge Normalized Difference Vegetation Index	reNDVI	$(B8a - B5) / (B8a + B5)$	Gitelson et al, 1994
Normalized Difference Water Index	NDWI	$(B3 - B8) / (B3 + B8)$	Gao et al, 1996
Transformed Chlorophyll Absorption in Reflectance Index	TCARI	$(3.0 * (B5 - B4) - 0.2 * (B5 - B3) * B5 / B4) / ((1.0 + 0.16) * (B8 - B4) / (B8 + B4 + 0.16))$	Haboudane et al, 2002

Table 4. Vegetation indices included as predictors for the classification model.

The selection of the vegetation indices, similar to the selection of the bands, was made considering a previously existent crop classification model.

The same approach was used for gathering images to compose the predictor variables for the 2019/2020 summer cycle, collected between late August and early March, to try to ensure including early sowing

parcels and late ones as well due the characteristics of this particular summer cycle. The 7 time frames to collect the images for the time validation dataset where '2019-08-24' - '2019-09-23', '2019-09-24' - '2019-10-23', '2019-10-24' - '2019-11-20', '2019-11-24' - '2019-12-30', '2019-12-31' - '2020-01-25', '2020-01-26' - '2020-02-08', '2020-02-09' - '2020-03-06'. After gathering these images, the vegetation indices were calculated and included as variables as well.

All remote sensing imagery predictors are down-scaled to 30 meter/pixel spatial resolution for reducing the size of the files to be processed and fitted into the model.

2.5 Machine Learning classification algorithm

Random Forest is the machine learning classification algorithm selected for the model since the characteristics of the training data samples are limited in terms of quantity and spatial distribution among the ROI. As mentioned, the ground-truth samples collection was highly influenced by the accessibility in terms of proximity to main roads and is unbalanced due the dominance of soybean production in the region within the summer production cycle. Another contributing reason for considering Random Forest appropriate for this case is that there will be attempts to transfer the model predictions spatially and also through time, all based on the same ground-truth data from a Alto Paraná which will only match the area considering time and space for one of the predictions, therefore, there are expectations of noise presence and surfaces that are not represented by the training data due possible feature space differences that are not met with the available samples. Also considering the prediction to be performed as a comparison of the originally spatially transferred model with a new prediction after adding samples that are outside the limits of Alto Paraná and fall into the neighboring areas, which mostly correspond to maize.

The Random Forest algorithm is applied through the CARET R package, configured with 100 trees and a distribution rate of 75/25 for training and testing. For tuning the model and guaranteeing independence

between the training and testing data, samples were split in 10 folds for cross-validation using the Spatial CV function of the CAST R package and introduced to the classifier.

2.6 Traditional accuracy assessments

Overall Accuracy and Kappa Index (Eq. 1) are the accuracy assessments selected for a first validation following the methodology of traditionally applied for evaluating crop classification model predictions based on machine learning, remote sensing and ground-truth data.

$$K = \text{Pagree} - \text{Pchance} \quad (1)$$

Where,

K = Kappa index

Pagree = Proportion of trials which in judges agree

Pchance = Proportion of trials in which agreement would be expected due to chance

The Overall Accuracy will indicate the percentage of labels classified correctly and the Kappa Index will compare the Overall Accuracy with random chance to establish a percentage value of confidence. These validations will be applied to all classification predictions and displayed along the results for traditional accuracy assessment validation.

2.7 Area of Applicability

An analysis and validation will be applied for each spatial prediction by assessing the Area of Applicability (AOA), as the final part of the workflow and for attempting to respond to the research questions, the AOA of each prediction will be assessed by first, defining a spatial cross-validation strategy for the classification

model to then calculate the dissimilarity index (DI) of the training data by determining the maximum distance between samples presented to the model, which will set a threshold for assessing the prediction areas that can be considered inside or outside the area of applicability, depending on how different is the presented data to the most dissimilar sample, to then apply also a final calibration of the AOA for attempting to provide a clearer vision of where the model is confident and, to which level of accuracy.

The AOA is applied by using the `aoa()` function of the CAST R package.

2.7.1 Cross validation strategy

As a first step of this validation implementation and since the AOA assessment is highly dependent on the cross-validation strategy selected for the training data.

After analyzing the characteristics of the samples and testing the impact on the prediction results comparing the outcomes generated by applying spatial cross-validation strategy with including a random cross-validation (Figure 4) to the training dataset, the contrast showed how the random CV model adjusts very well to the data that is not strange to the model, and the how the spatial CV model increases the ability to predict over new locations, further away from the reference data.

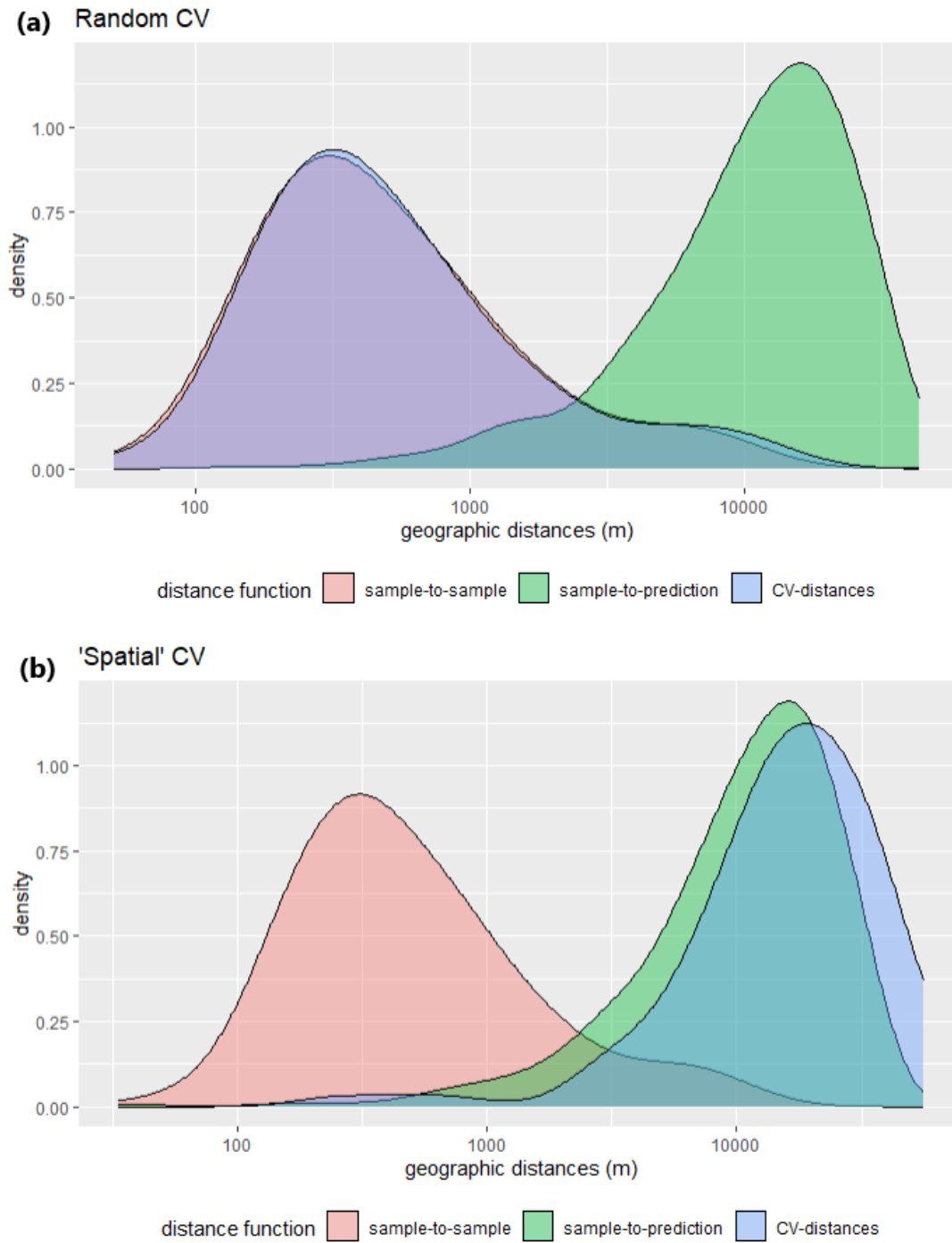


Figure 4. Comparison of the performance during the prediction task contemplating different cross-validation strategies including (a) random cross-validation and (b) spatial cross-validation

The spatial cross-validation leave-one-out is the strategy selected for performing the classifications, as an attempt to reduce effects of spatial dependency of the samples and to increase its reliability, given the limited and sparse conditions of the ground-truth samples dataset.

The cross-validation strategy for the model's different predictions will be performed using the `CreateSpacetimeFolds()` function available on the CAST R package with 10 folds.

2.7.2 Dissimilarity index

The calculation of the Dissimilarity Index (DI) is done by using the `aoa()` function, available on the CAST R package.

The training data and the prediction model are used as inputs to the `aoa()` function, which calculates the DI, ideally after the previous cross-validation strategy definition. The DI is obtained by analyzing the training data for measuring the distance of each data point presented to the function to the nearest training data (if a cross-validation strategy is applied, this distance is calculated between points that belong to the same cross-validation fold). The nearest distance for each point is then divided by the average of the mean distances between all training data points providing the DI, which is calculated for each training data point. The highest DI among the whole assessed training data points will set the AOA threshold, indicating the maximum dissimilarity among training data.

The DI is calculated to set a threshold for determining which data falls inside or outside the AOA, depending on how dissimilar is the new data presented to the predictor.

2.7.3 Area of Applicability assessment

The calculation of the AOA calibration is done by using the `aoa()` function, available on the CAST R package.

Based on the previous steps of cross-validation strategy and DI calculation, the AOA determines the areas where the prediction model is expected to predict with an error comparable with the cross-validation error of the model by analyzing the relation between the prediction and the training data to then be classified inside or outside the AOA based on their dissimilarity and if its value lie above or below the AOA threshold (Figure 5.).

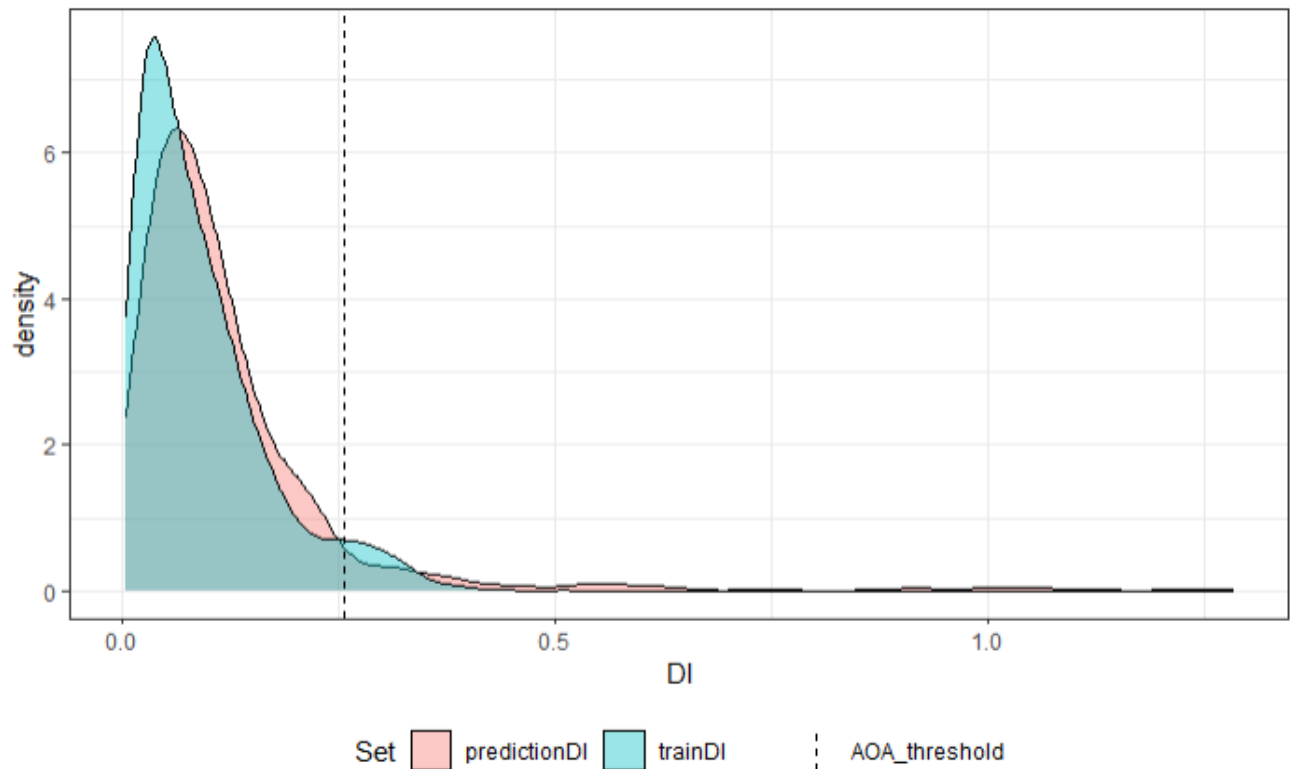


Figure 5. AOA calculation plotting the AOA threshold and the relation between the prediction DI and the train DI, showing how they adjust to each other and the parts of them that fall outside or inside the AOA, according to the maximum tier of the DI calculated over the training data.

Using the training data and the predictor model as inputs, the AOA has only two values: 0 and 1. 0 means that the prediction over a determined area is outside the area of applicability, 1 means that the prediction over a determined area is inside the area of applicability.

This assessment provides an output that covers the same area that the prediction takes into account to perform its extrapolations, analyzing each spatially referenced data for assessing their applicability to the prediction model based on the presented training data.

2.7.4 Area of Applicability Calibration

The calculation of the AOA calibration is done by using the `calibrate_aoa()` function, available on the CAST R package.

This assessment takes the previously calculated AOA result as input along with the prediction model and allows the setting of multiple cross-validations to be applied on the analysis.

The calibration of the AOA consists of a moving windows analysis that uses DI values which, through multiple cross-validations, allows to calculate the relation of the accuracy of the prediction model and the DI (Figure 6).

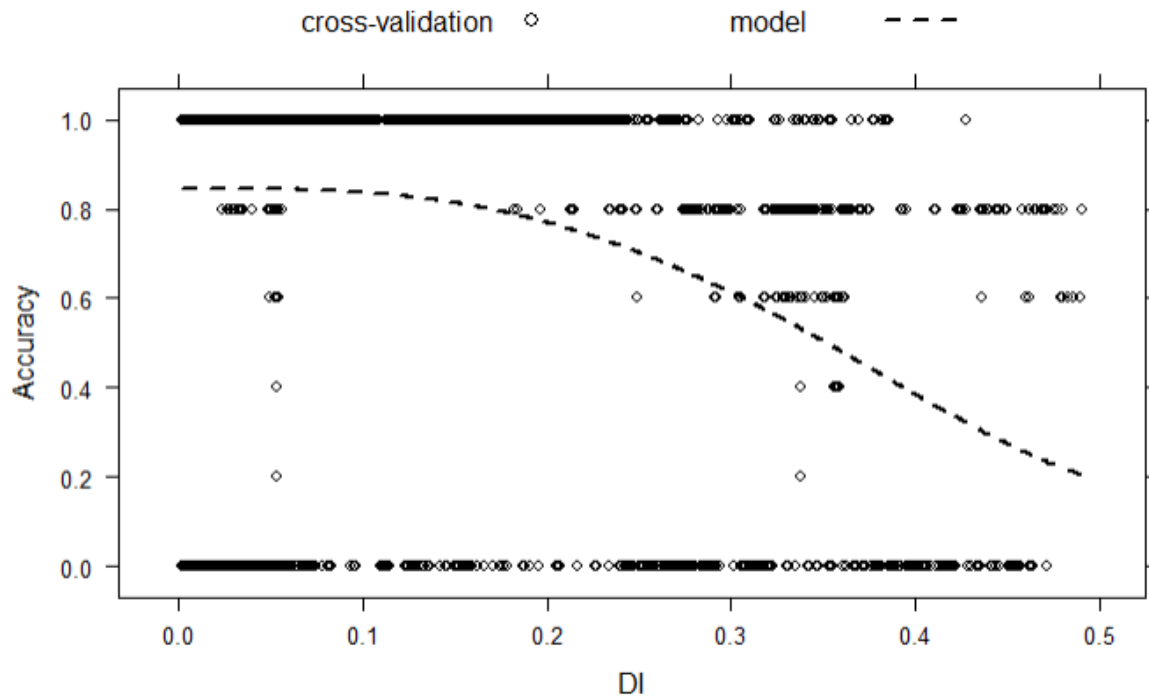


Figure 6. AOA calibration result plotting the Accuracy and the DI for the different cross-validations applied to the model prediction.

The result of the calibration provides a performance measurement that quantitatively represents the prediction levels of confidence in the different analyzed areas, for a deeper understanding of the extensions where the prediction is confident or not and how dissimilar is this confidence within the assessed prediction accuracy.

Results

After applying the spatial cross-validation strategy and running the different models, the results can be presented organized in the following order, corresponding each sub-title to each attempt to respond each research question, by using the same methodology but altering the remote sensing predictor variables while maintaining the same ground-truth data references, only altering this for the spatially transferred model second part, where the original prediction is compared to a new one that includes samples from the contiguous areas as well.

Along with the model prediction results, the traditional accuracy assessments results will be displayed along with the AOA validation for each prediction.

3.1 Alto Paraná Department prediction and assessment

3.1.1 Classification

The classification performed over Alto Paraná department including information of the summer cycle of 2020/2021, with the ground-truth samples matching the same year of the imagery variables predictors used for this classification, reflected the high dominance of soybean crops in the analyzed area. In figure 7 we can see with more detail the crop classification prediction, along with the samples and the road infrastructure.

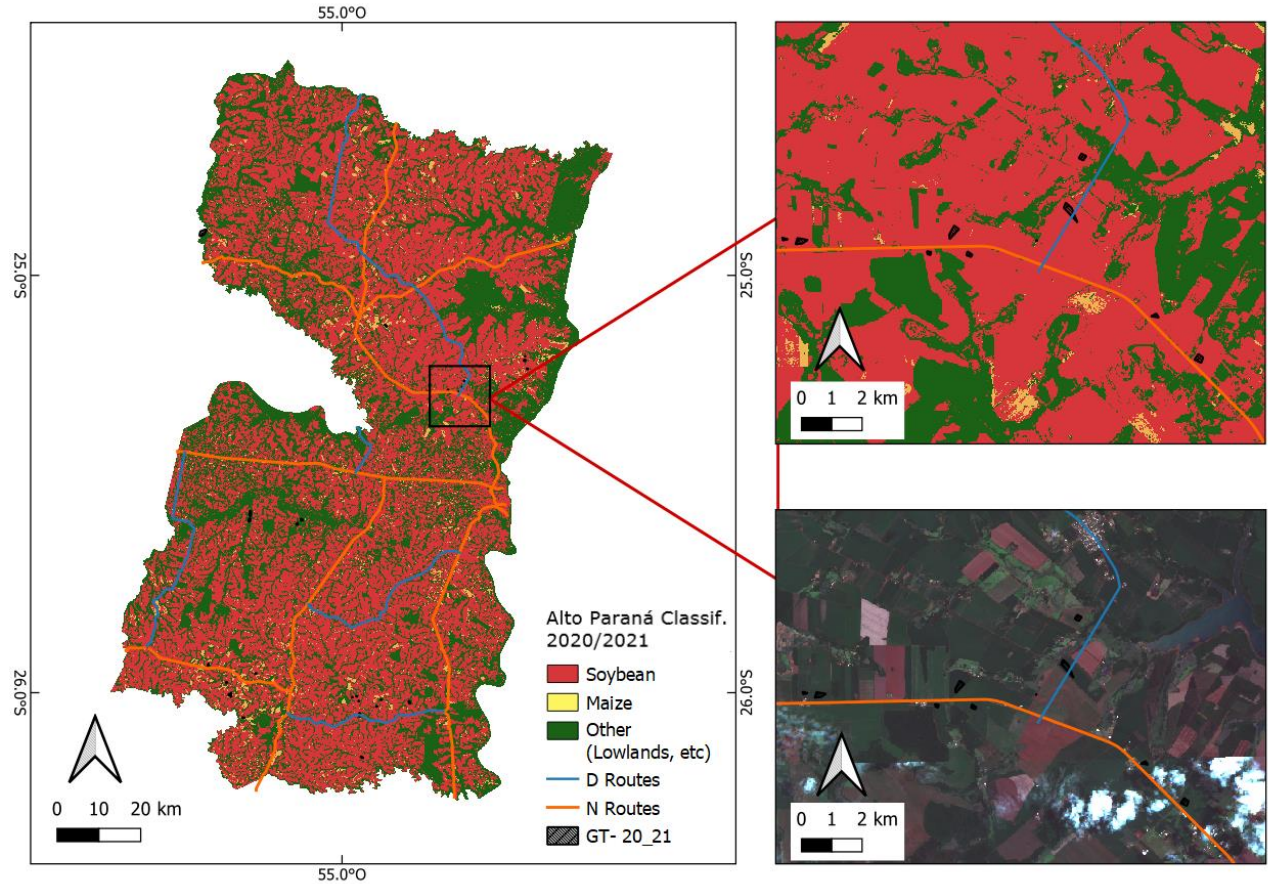


Figure 7. Classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and remote sensing variables

There is also an RGB image that is a composite image from January 2021, where we can appreciate the same zoomed in area and detect the clouds that made it through the filter.

The accuracy assessment for this classification was 91.8% for Overall Accuracy and 86.6% for Kappa index.

3.1.2 Area of Applicability - Alto Paraná 2020/2021

After generating the prediction and the accuracy assessment, the AOA was calculated by determining the Dissimilarity Index to set a threshold of maximum distance among the training data cross-validation independent folds. The results of the AOA are displayed in Figure 8, where it is visible in purple, the areas

that fall inside the AOA and therefore are classified as not strange to the model regarding the presented training data.

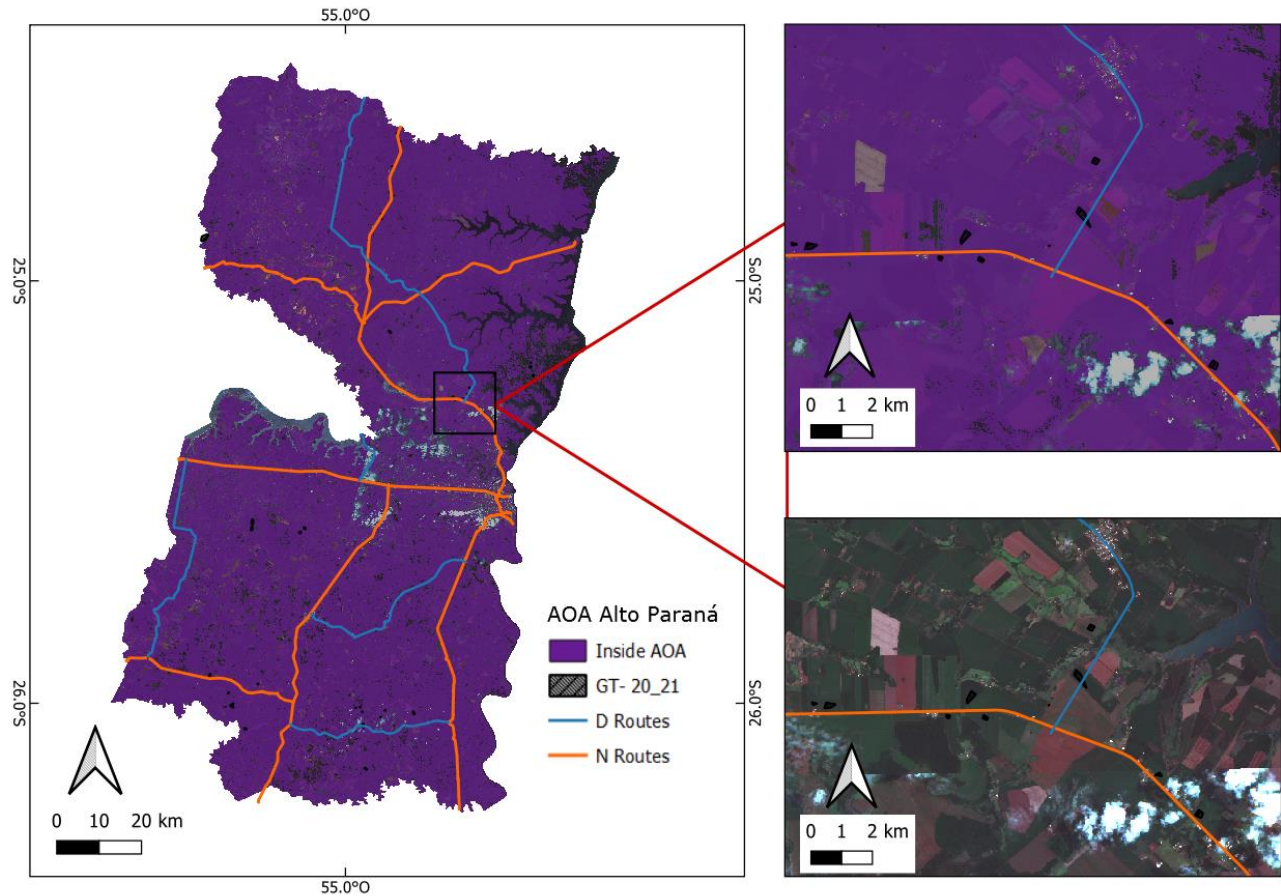


Figure 8. Area of Applicability (AOA) of the classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and remote sensing variables.

For a direct comparison, the same zoomed in area displayed in the prediction is now shown again to check how the AOA adjusts over it in more detail. We can see that clouds, water bodies but also a few parcels are not covered by the AOA, contrasted by the RGB composite of January 2021.

To grab more insights on the assessments made over the prediction, the AOA is calibrated to categorize the areas in a visually explicit map that classified the areas on a scale of confidence (Figure 9).

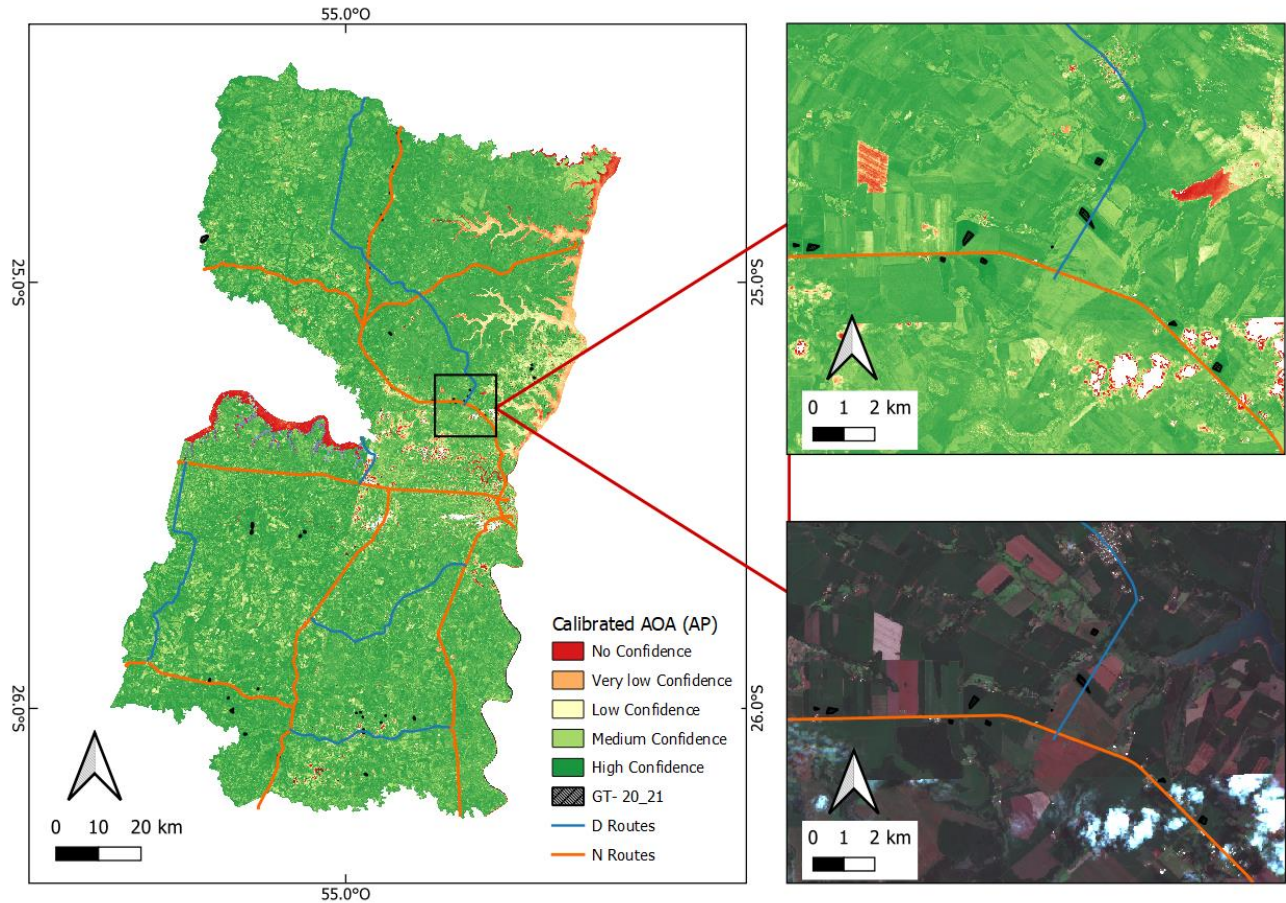


Figure 9. Calibrated AOA over Alto Parana department.

The AOA calibration analysis brings further detail to the 0 - 1 scale of the first AOA output. We can see now how the confidence of the model extrapolates over the analyzed areas, to distinguish by category where the model is confident and where it is not. Again, there is a clear visual of the impact of clouds, water bodies, urban settlements and parcels that hold a presumably non-previously-identified Class.

3.2 ROI prediction (Spatial transfer validation)

3.2.1 Classification

The classification performed over the ROI (Alto Paraná, Canindeyú and Itapúa departments) including information of the summer cycle of 2020/2021, with the ground-truth samples restricted to Alto Paraná but matching the same year of the imagery variables predictors used for this classification, again it reflects the high dominance of soybean crops in the analyzed area, but also a considerable presence of maize specially in Itapúa department. In figure 10 we can see with more detail the crop classification prediction, along with the road infrastructure.

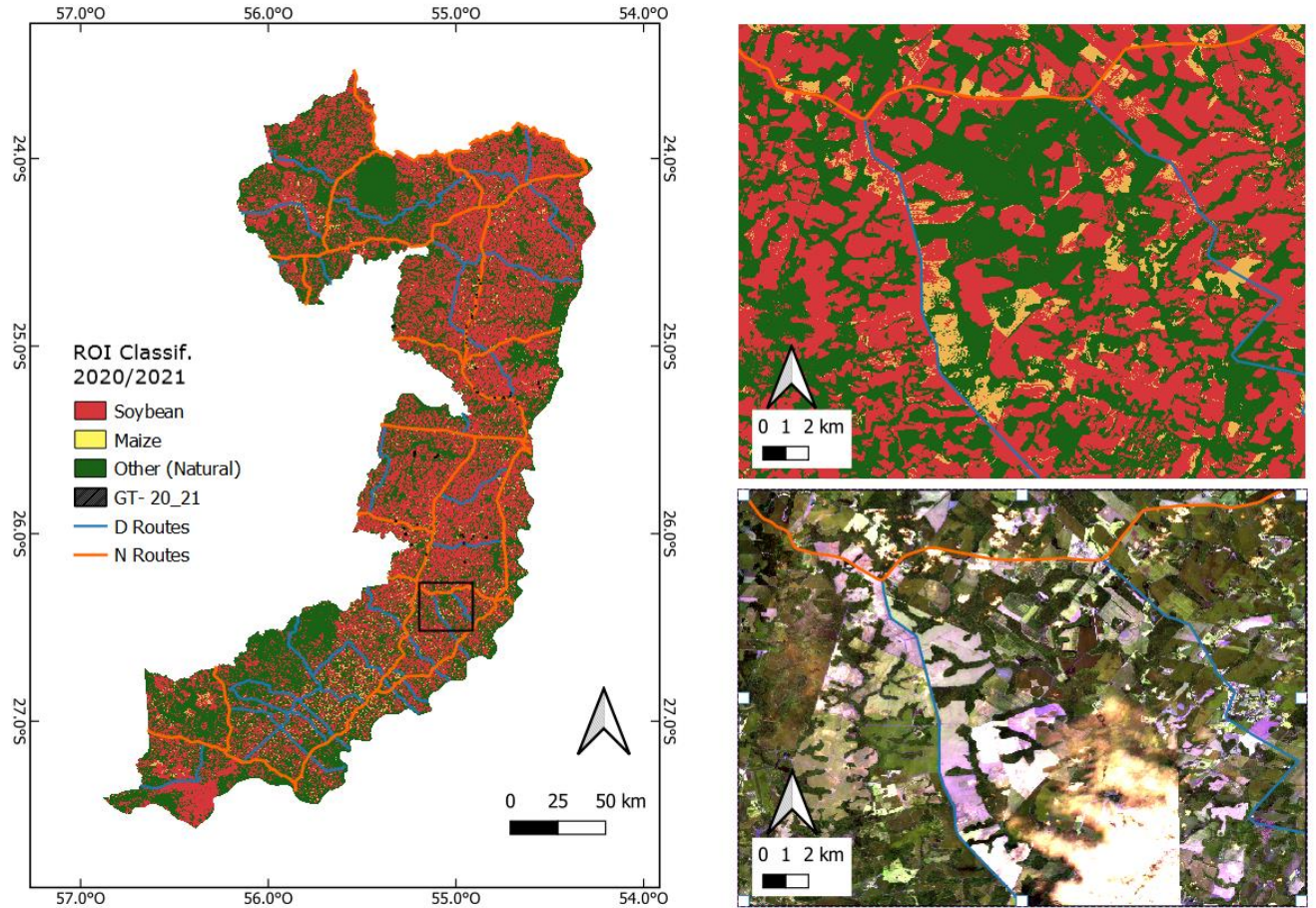


Figure 10. Classification model prediction for the ROI departments (Alto Paraná, Canindeyú and Itapúa) with 2020/2021 summer cycle ground-truth samples from Alto Paraná and remote sensing variables from the entire ROI.

There is also an RGB image that is a composite image from January 2021, where we can appreciate the same zoomed in area as in the prediction layer to see how the classified fields look like in this area and how clouds made it through the filter too.

The accuracy assessment for this classification was 93.6% for Overall Accuracy and 87.5% for Kappa index, increasing slightly in both assessments compared with the Alto Paraná prediction validation.

3.2.2 Area of Applicability - ROI 2020/2021

After generating the prediction and the accuracy assessment, the AOA was calculated by determining the Dissimilarity Index to set a threshold of maximum distance among the training data cross-validation

independent folds. The results of the AOA are displayed in Figure 11, where it is visible in purple, the areas that fall inside the AOA and therefore are classified as not strange to the model regarding the presented training data.

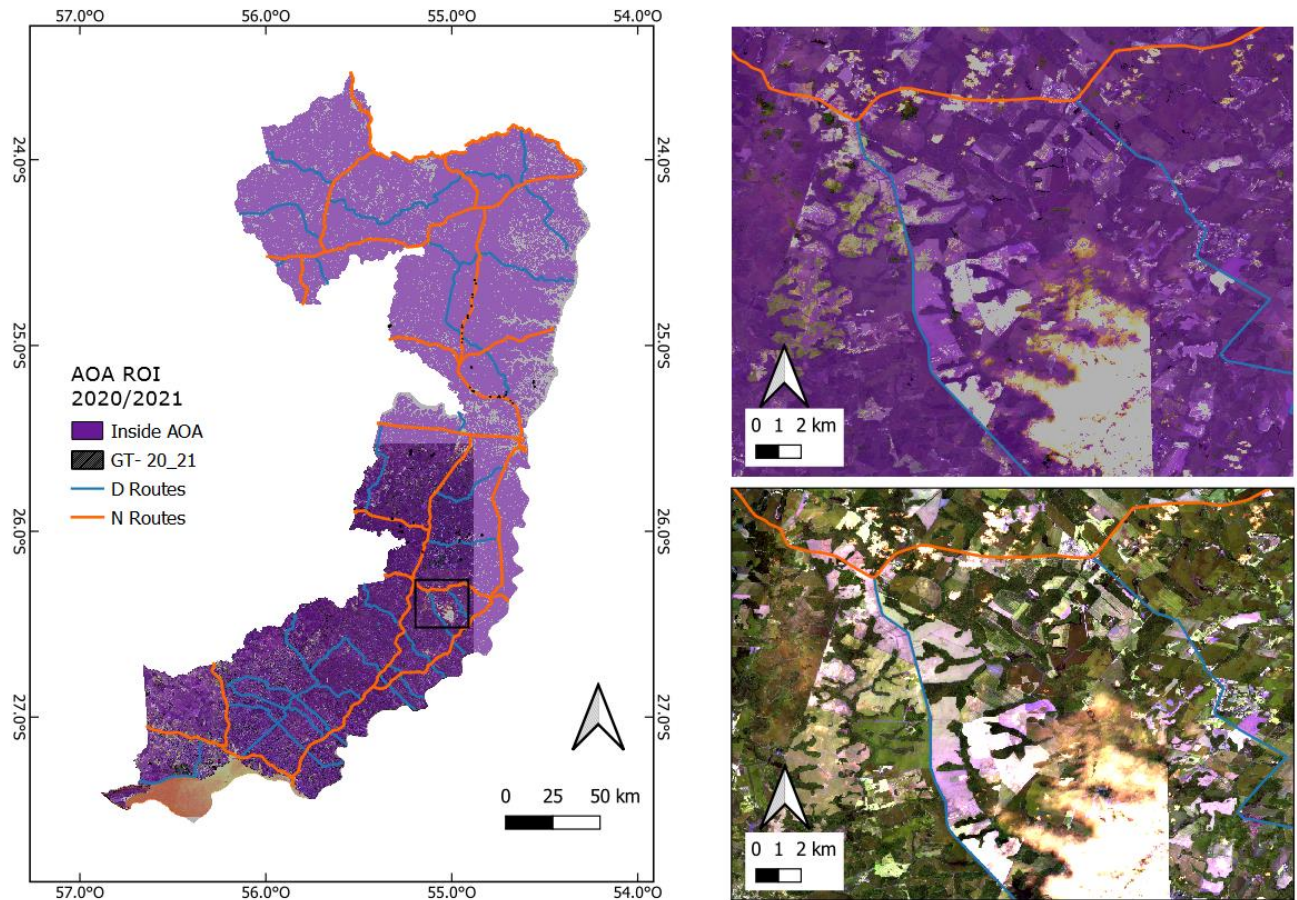


Figure 11. Area of Applicability (AOA) of the classification model prediction for the ROI with 2020/2021 summer cycle ground-truth samples and remote sensing variables.

In the background, there is a partial cover of the south section of the ROI with an RGB image composite from January 2021 for adding contrast in the zoomed in part for a direct comparison, the same zoomed in area displayed in the prediction shown again to check how the AOA adjusts over it in more detail. We can see that clouds, water bodies but also a few parcels are not covered by the AOA.

To grab more insights on the assessments made over the prediction, the AOA is calibrated to categorize the areas in a visually explicit map that classified the areas on a scale of confidence (Figure 12).

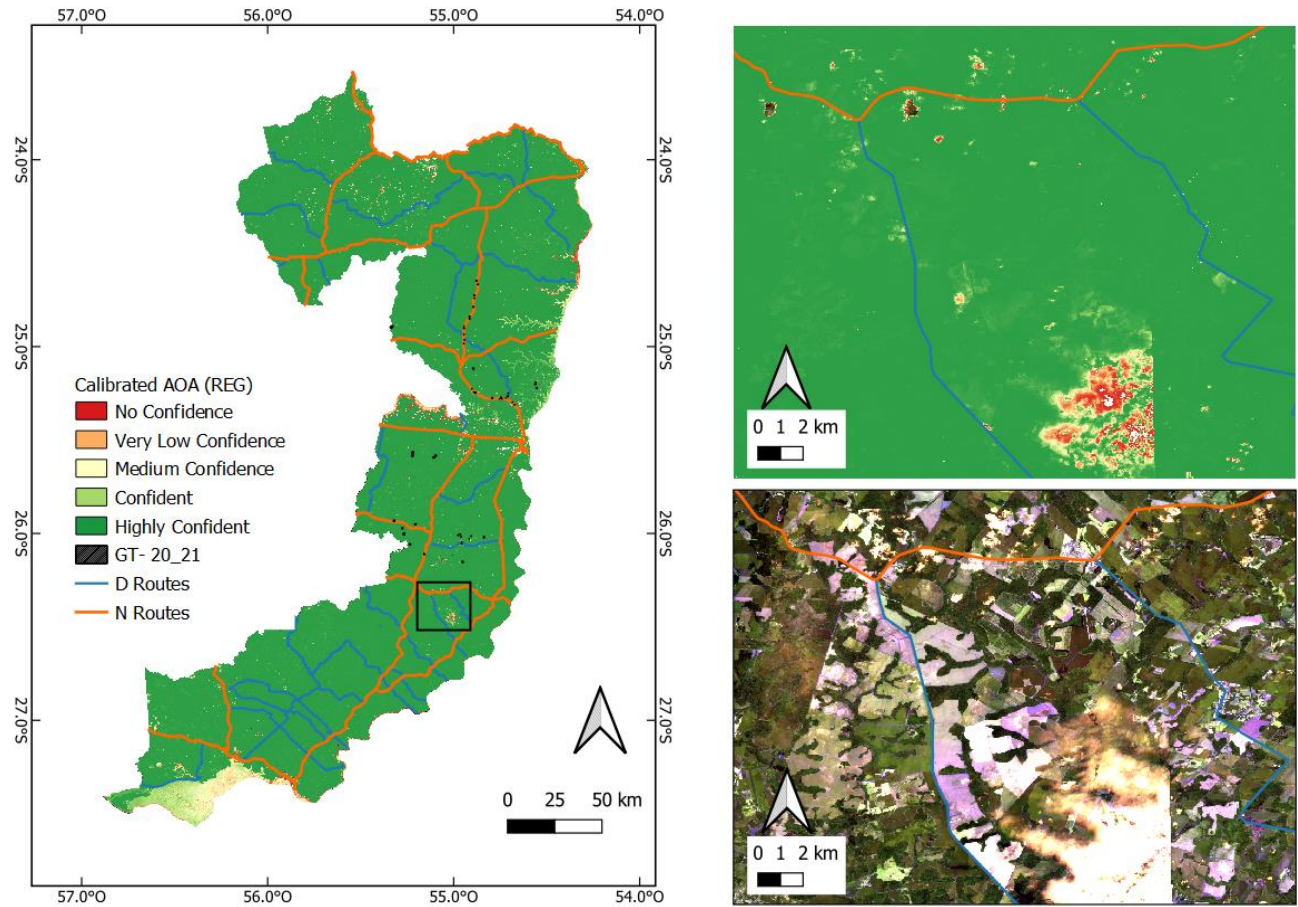


Figure 12. Calibrated AOA over the ROI (Alto Parana and neighboring departments) with Alto Parana 2020/2021 summer cycle ground-truth.

The AOA calibration analysis brings further detail to the 0 - 1 scale of the first AOA output. We can see now how the confidence of the model extrapolates over the analyzed areas, to distinguish by category where the model is confident and where it is not. Again, there is a clear visual of the impact of clouds, water bodies, and parcels that have a slight drop in confidence (dropping to a yellowish color).

3.2.3 Classification - ROI 2020/2021 + extra samples

For adding a further analysis on the spatially transferred model prediction approach, a new classification was performed by also adding samples from the neighboring areas collected in the same time period of 2020/2021 and following the same methodology of picking areas near the main roads infrastructure. The result (Figure 13) shows the effect over the ROI and the zoomed in area was maintained following the previous displayed prediction map.

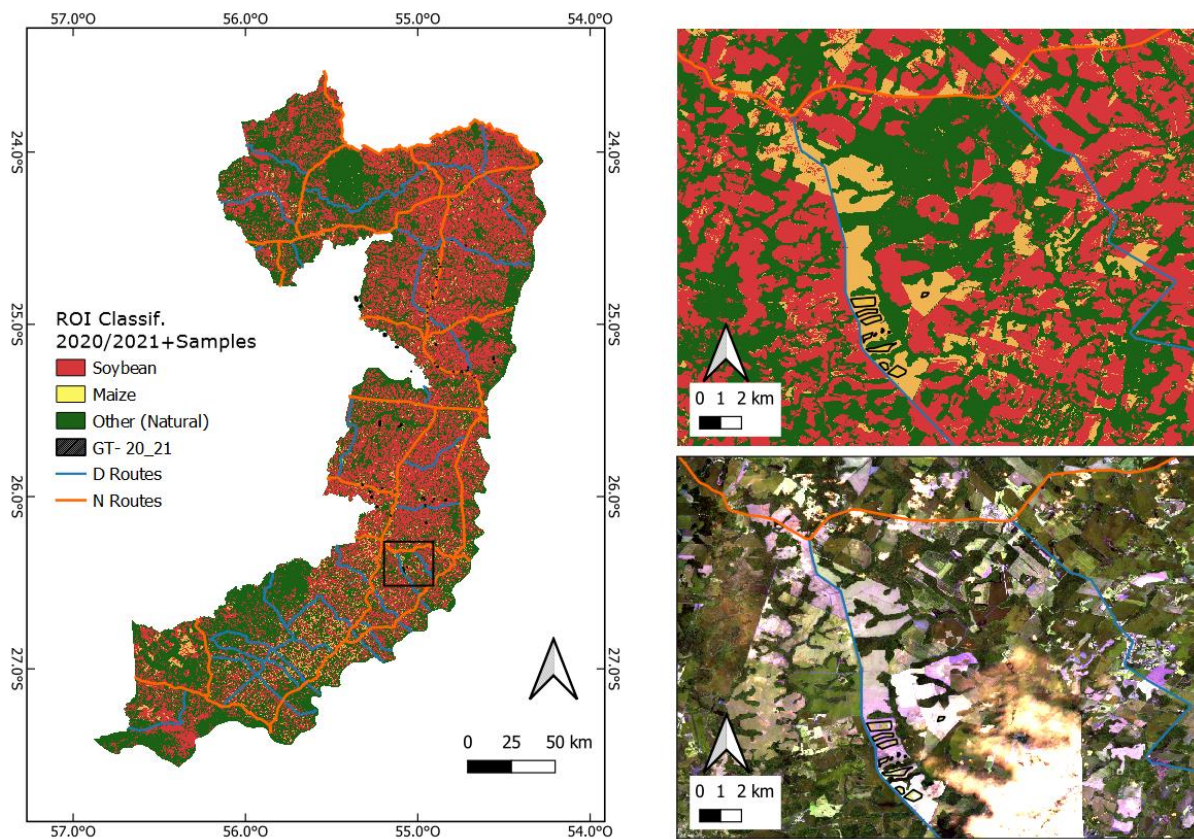


Figure 13. Classification model prediction for the ROI departments (Alto Paraná, Canindeyú and Itapúa) with 2020/2021 summer cycle ground-truth samples from Alto Paraná and from the neighboring departments as well and remote sensing variables from the entire ROI.

In comparison with the previous classification (Figure 14), we can see how the maize areas increase disregarding the high dominance of soybean in the ROI. Fields with an “salt and pepper” effect regarding

mixing classes are now more solid for representing one class. However, this impact does not seem to affect the Other class, which was not visibly altered regarding the comparison.

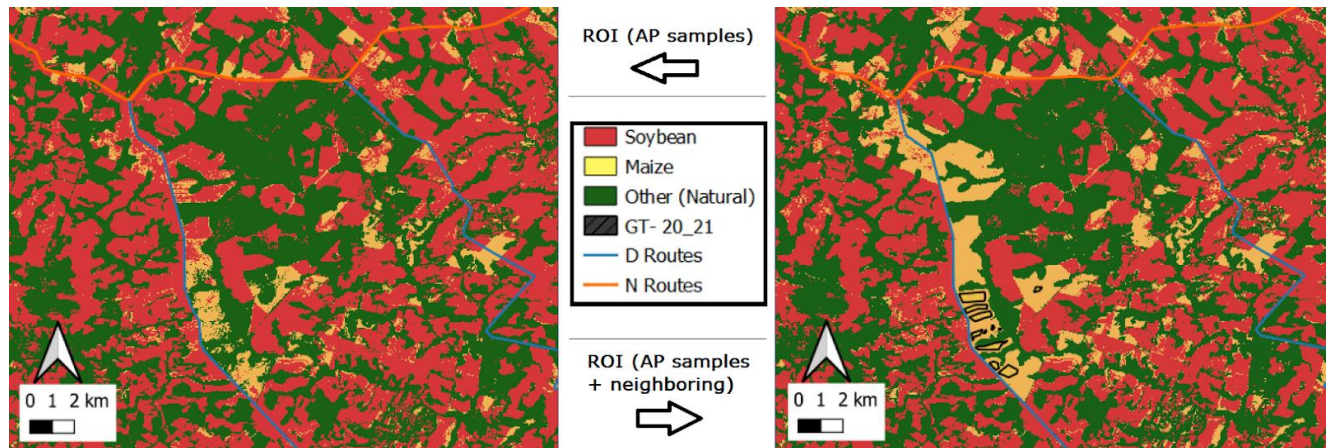


Figure 14. Comparison of ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas.

The accuracy assessment for this classification was 91.5% for Overall Accuracy and 87.1% for Kappa index, decreasing slightly in both assessments compared with the ROI prediction validation with only Alto Paraná ground-truth samples, but mostly in the case of Overall Accuracy.

3.2.4 Area of Applicability - ROI 2020/2021 + extra samples

To assess the impact of the new samples on the AOA assessment of the new prediction, we can see on Figure 15, how the AOA increases in the areas that were previously margined despite not having cloud interference over them. This effect is not restricted to the parcels where the ground-truth are overlapping but also there is an increase in the surrounding parcels.

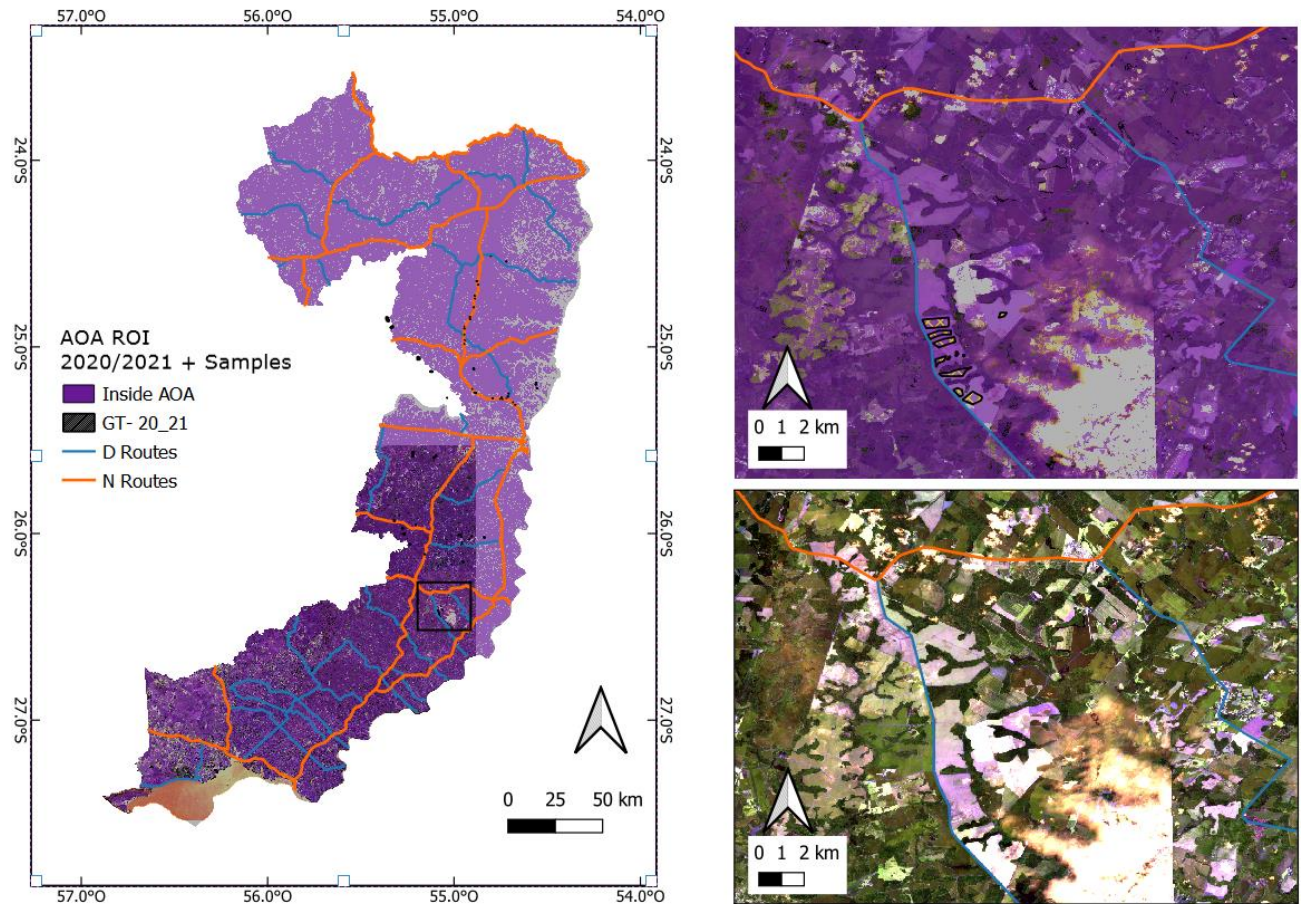


Figure 15. Area of Applicability (AOA) of the classification model prediction for the ROI with 2020/2021 summer cycle Alto Paraná + neighboring departments ground-truth samples and remote sensing variables.

As in the previous analysis, the RGB image from January 2021 is used as a contrast again, which allows to better compare the effects of the new samples over the AOA analysis, covering now more parcels and showing that the areas outside the AOA are mostly surfaces with cloud interference.

In a direct comparison (Figure 16) over the zoomed in area is visible with more detail how this mentioned effect of the new samples over the prediction expands the AOA after applying the analysis. Is also noticeable how the AOA does not cover the clouds as expected but also there are what seem to be agricultural parcels with a different class from the ones presented as training data.

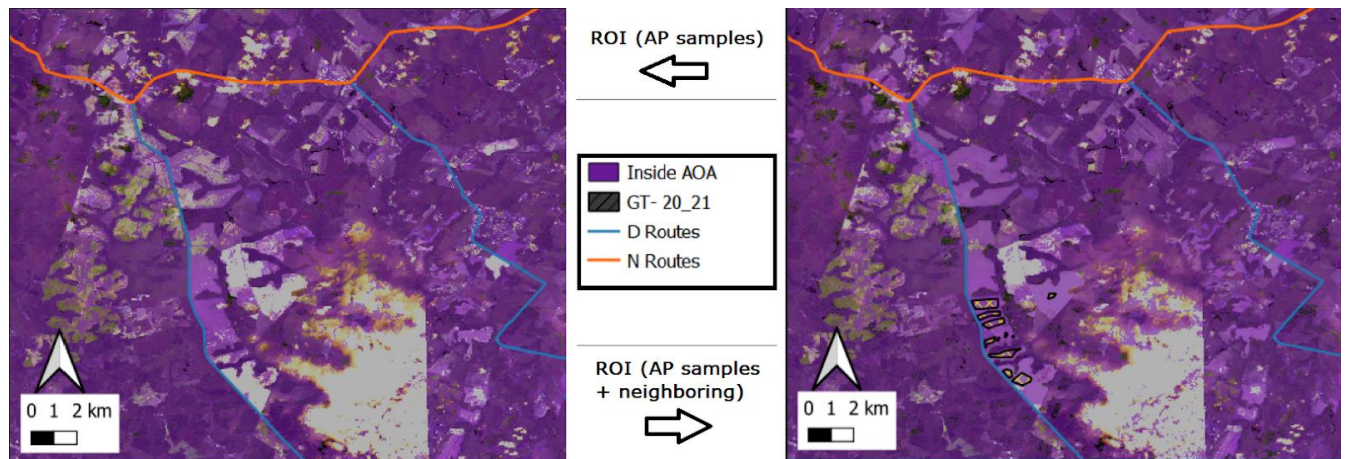


Figure 16. Comparison of the AOA assessment results over the ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas.

To complete the comparison, the AOA is calibrated as well as the in the previous prediction analysis (Figure 17). There is evidence of a general decrease in confidence over the region but also an increase in previously low-confidence areas.

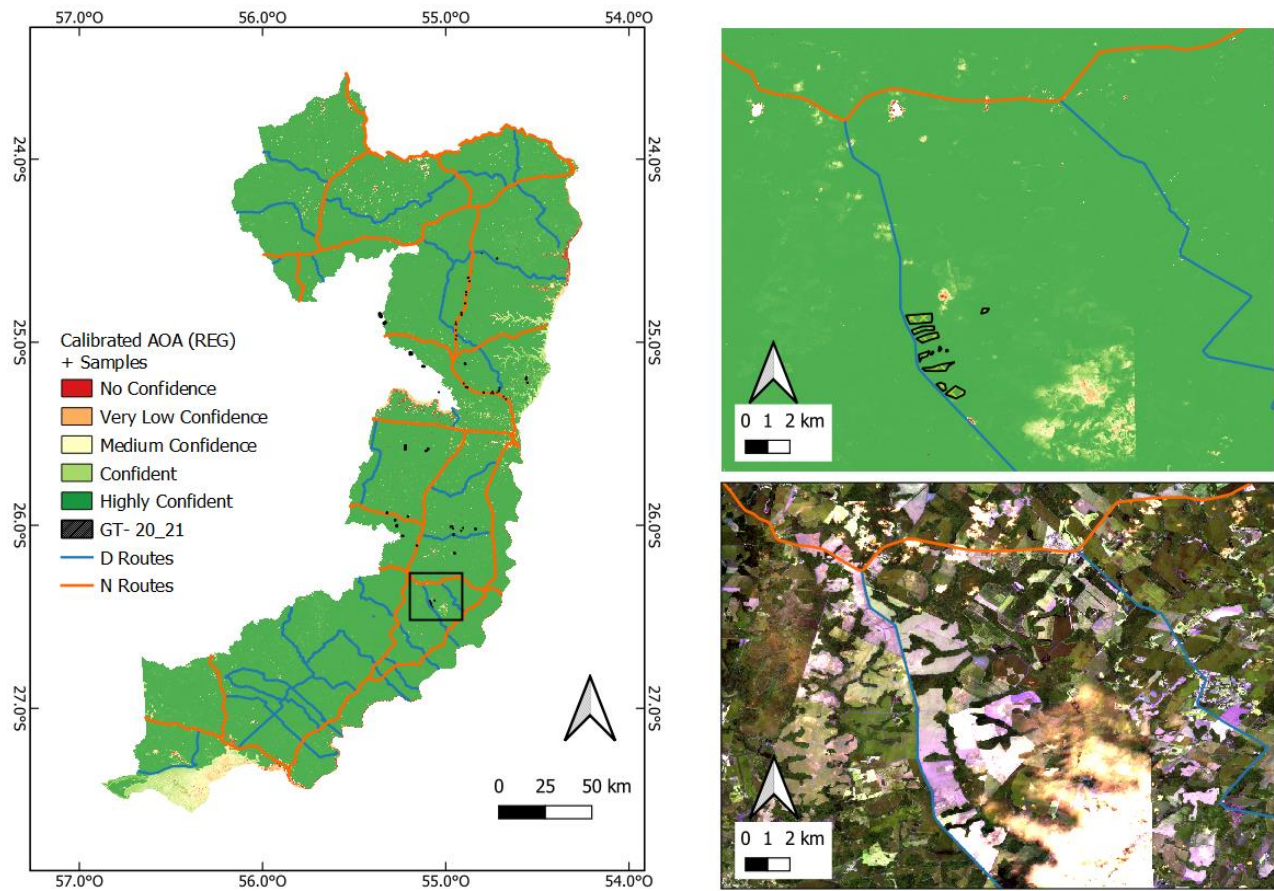


Figure 17. Calibrated AOA over the ROI (Alto Paraná and neighboring departments) including Alto Paraná 2020/2021 summer cycle ground-truth, and added samples of neighboring areas.

The general decrease of confidence is visible more clearly in Figure 18 where the greenness becomes more light and yellow tones accentuate over the areas that are not covered by the AOA assessment in the previous step.

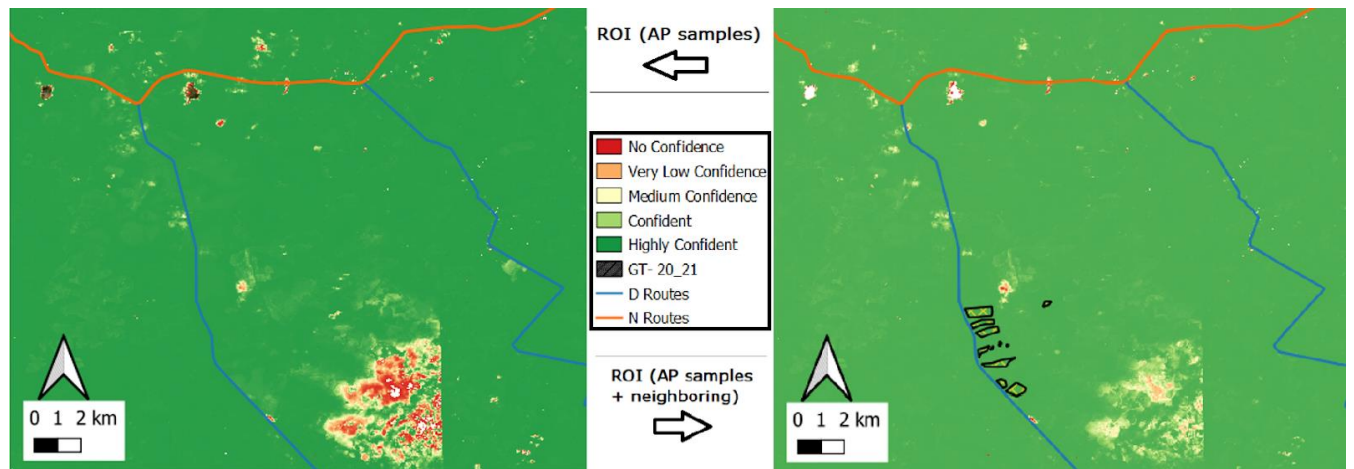


Figure 18. Comparison of the AOA calibration results over the ROI classification predictions before (left) and after (right) adding new samples of the neighboring areas.

3.3 Alto Paraná Department prediction with different year variables (Time transfer validation)

3.3.1 Classification

For assessing the transferability of the classification model, a prediction was made over Alto Paraná department while using the ground-truth samples corresponding to the summer cycle of 2020/2021, but with imagery variables predictors from the 2019/2020 summer cycle. As a result, we can see in Figure 19 how the model predicted over these areas based on the data presented to the model.

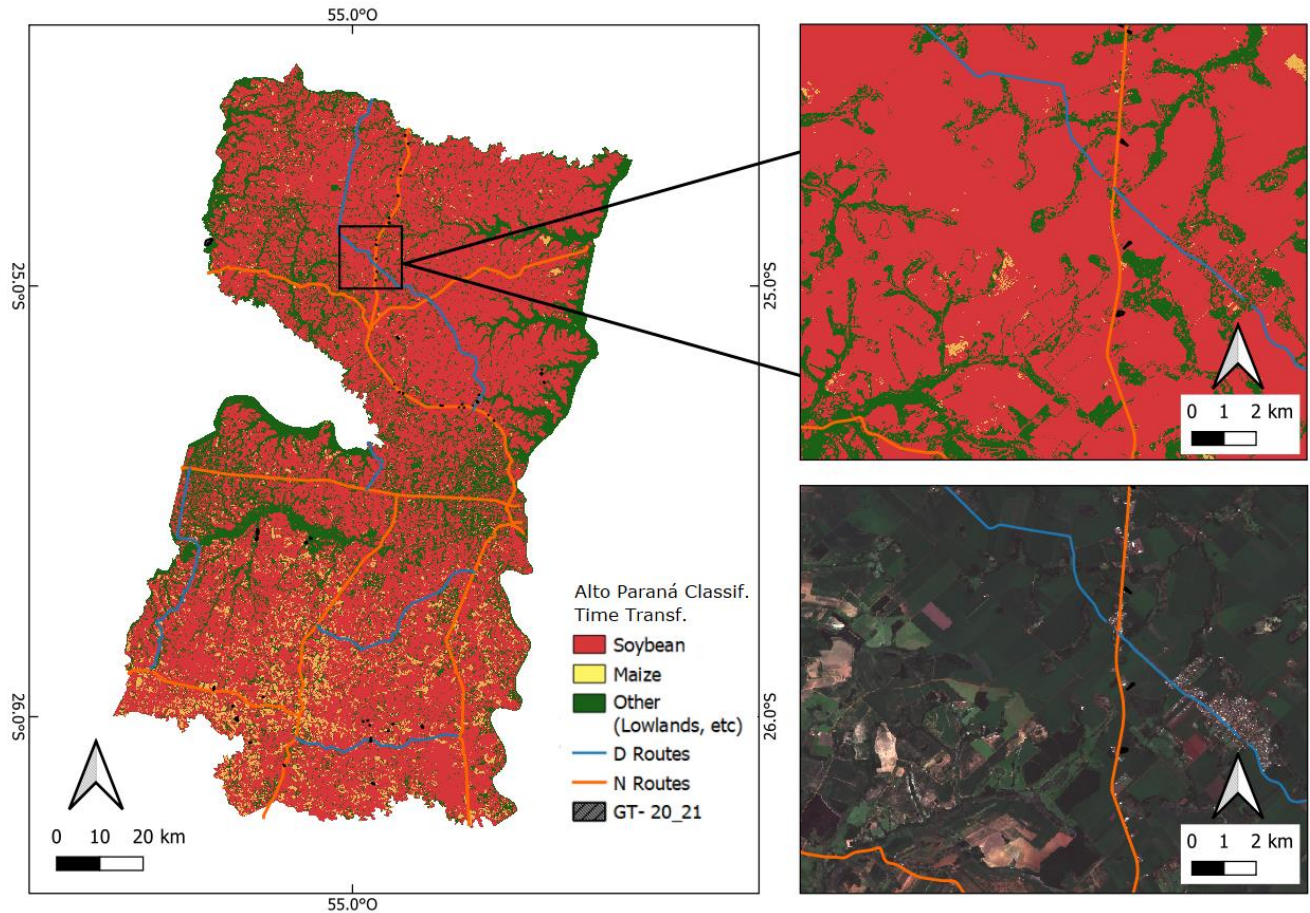


Figure 19. Classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables.

There is also an RGB image that is a composite image from January 2021, where we can appreciate the same zoomed in area and detect the clouds that made it through the filter.

The accuracy assessment for this classification was 75.9% for Overall Accuracy and 62.4% for Kappa index. Evidencing a decrease in the accuracy compared with the previous predictions, as expected, specially by interpreting the Kappa index results, which add insightful information to a presumably overoptimistic Overall Accuracy assessment.

3.3.2 Area of Applicability - Alto Paraná (Time transfer validation)

By assessing the AOA over the time transferred classification prediction results over Alto Paraná department, we can see in Figure 20 how the AOA decreases its extension compared to the classification model trained with Alto Paraná 2020/2021 summer cycle ground-truth and also remote sensing imagery predictors.

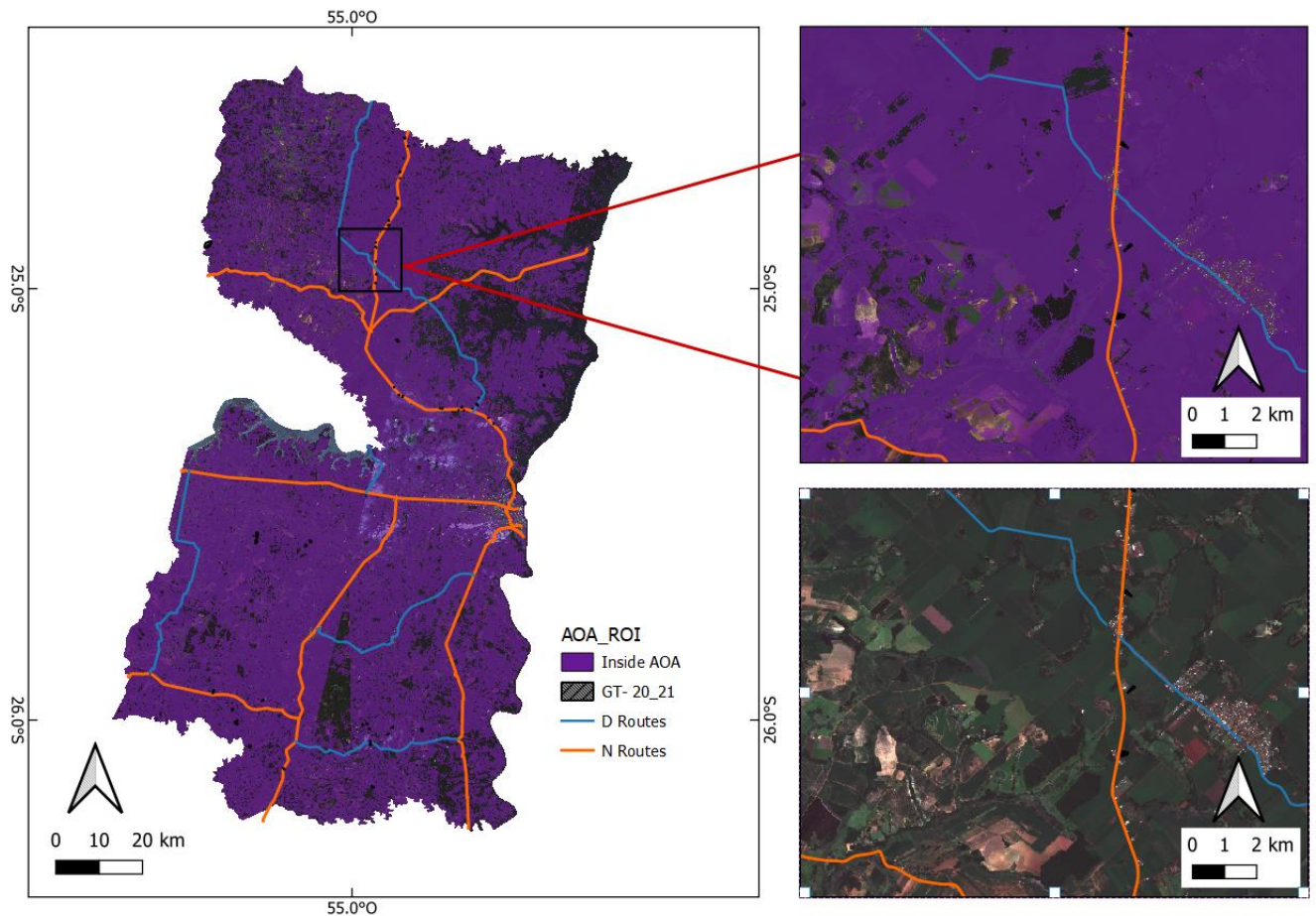


Figure 20. Area of Applicability (AOA) of the classification model prediction for Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables.

This time the zoomed in area displays how the AOA looks in detail, contrasted by an RGB image of January 2020. We can see that some parcels are not covered by the AOA, along with urban areas that are also left outside the AOA. This gives a clear visual of where the model is applicable or not regarding the training

data presented as ground-truth. There is also a higher impact of clouds and even missing predictors, like it is visible on the center-south part of the main map of Alto Paraná, where there is a triangle-shaped area that is not covered by the AOA, presumably due not enough images covering that specific among the collected imagery since is where satellite paths overlap too.

The decrease of the AOA is more visually clear when the AOA calibration is done (Figure 21). The confidence drops in general in comparison with previous assessments, increasing the noticeability of empty parts (below red) and yellow/light green areas. All of these less confident parts fall outside the AOA in different levels of confidence.

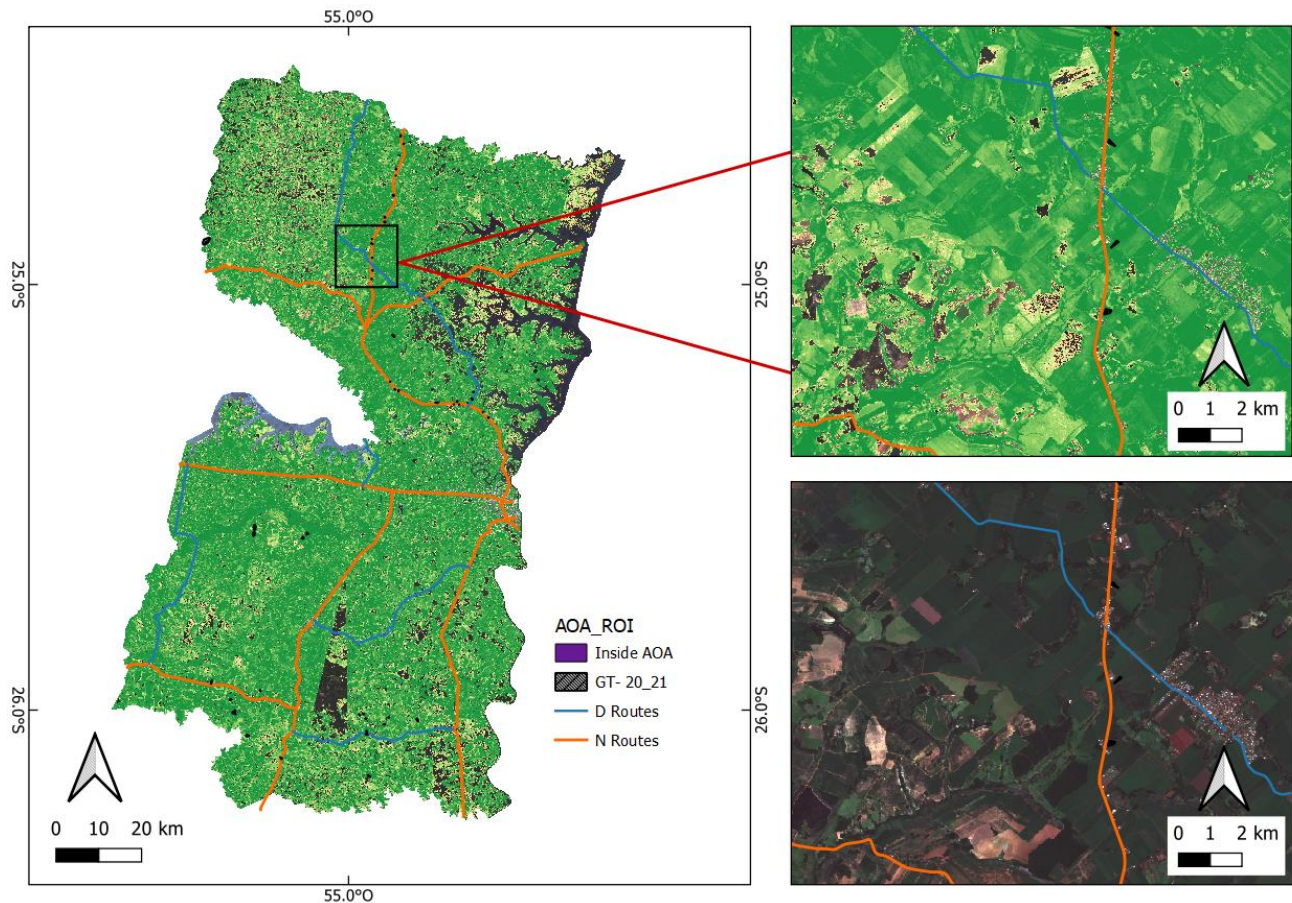


Figure 21. Calibrated AOA results for the classification model prediction of Alto Paraná department with 2020/2021 summer cycle ground-truth samples and 2019/2020 summer cycle remote sensing variables.

These variances over the agricultural fields could mean that the areas that present very different information from what the training data set presented to the model could mean that different agronomic decisions (irrigation, non-planted fields, different crops, crop species varieties, etc.) were made over those fields and that further sampling is needed to potentially increase the confidence of the prediction, assessed its AOA.

Discussion

4.1 AOA assessment and ground-truth allocation

The presented methodology approached the premise of increasing ground-truth sample gathering efficiency and guidance, by generating results that could potentially serve as a source of spatially informed guide for decision making, based on the results that succeeded in the detection of areas where sampling efforts could be needed to be directed in order to increase the reliability of classification model predictions are extrapolated to areas that are beyond the dissimilarity index of the presented ground-truth references used to build training datasets. Hence, the methodology holds potential to improve the efficiency of crop classification tasks, while also bringing spatial information-based support for further analysis or data generation.

The assessment of the AOA of crop classification prediction model results indicate which areas of the presented models are showing less confidence and thus, reveal a site on the analyzed area that could be interpreted as a potential category where the predictors information used as variables and samples presented to the model do not match. The areas not covered by the AOA could be interpreted as a gap in the prediction that needs to be filled with labeling information in order to be classified accurately and, for a further step the calibration layer could be analyzed to interpret the areas that are less confident regarding the dissimilarity index and the cross-validations made over the training and how remotely they are settled from the main road infrastructure, helping to optimize the allocation of ground-truth samples.

Considering the limited samples over the region of interest and their spatial distribution and unbalanced population, the model was not trained in optimal conditions and the results of the classification predictions might be biased. Since high-quality sampling demands an intensive labor that can represent substantial investment of resources, difficult to afford by individual researchers or even organizations (Wang et al

2022), it is valid to test methodologies such as the use of AOA assessment over these types of predictions that are made on limited and scarce ground-truth samples, and also in other predictions based on broader ground-truth samples availability with their previous validation, to add a further analysis onto predictions that could potentially lead to a more efficient use of ground-truth sampling efforts and resources.

The road infrastructure of this particular region restricts the accessibility to fields within distances that could be considered as remote, which is the reason because the samples are distributed conveniently along the areas that represent a better option in terms of accessibility. The roads that are not represented in the displayed route nets example of the study area are internal dirt roads that are susceptible to weather conditions such as precipitation. The challenges that limited accessibility represents in rural areas are not restricted to the analyzed study area, since there are conditions of less infrastructure among the same country and this replicates as well in other countries or regions, and its impact over the deployment of ground-truth collection scouting tasks could be reduced by using a method that could improve resource efficiency by using crop classification model predictions made with limited samples to then indicate potential areas where samples are needed by assessing the area of applicability of the predictions. Implementing ground-truth samples collection surveys without relying on methods that could lower the necessity of sampling or serve as a guide for directing these collection tasks, can represent a demand of resources that might jeopardize the accomplishment of reliable, dynamic and extensive crop classification prediction models.

4.2 Other methodologies considerations

Disregarding the efforts to improve crop classification models and processes while optimizing time and financial resources for approaching the ground-truth sampling challenge, and although there are registers of several different methods, there have been no findings of projects using the same methodology to optimize ground-truth necessity.

There are, however, methodologies that approach the premise of increasing efficiency of resources while obtaining ground-truth generation or guidance for sampling. Commonly, these approaches involve analyzing historical crop cover maps to assess crop information over time and space, and using remote sensing data machine learning to map crop cover out of the information generated from the historical data to generate training data and proposing new work frames for validations while reducing the need of traditional ground-truth sampling efforts. An example could be the research made by Belgiu et al, 2021 suggests by applying crop-phenology-based sampling generation, by compiling and applying analysis using unsupervised learning through datasets of historical classification data based on ground-truth on field surveys, the classification predictions accuracies are assessed by the Overall Accuracy analysis, reaching percentages of over 95%. Along with these examples, there are other studies that attempted new classifications while extracting information from Crop Data Layer historical information by using unsupervised learning techniques, to combine the training data generated it with local surveys as did by Pengyu et al, 2020 who in different assessments over the classification model predictions obtained Overall Accuracies ranging between 86% and 97% being less accurate in cases where the model was spatially transferred as well as timely, due to the variations of crops species and their different cycles, and management decisions, attached to the weather and soil conditions of each region; also discussed by Waldner et al, 2016, where they used historical land cover maps to extract training data and referred to the challenge of these methodologies that rely on historical data to cope with marginal years, where natural hazards or others have an impact over the crops growth. This condition adds to the base fact discussed by Foody et al, 1988, who mentioned the considerable variability of crop production and hence the difficulties of accurate crop classification generation. With the objective of testing the feasibility and reliability of a trained model for crop type identification by inter-annual transfer, Hu et al, 2022, while discussing that improving the efficiency of ground-truth references collection and data sharing were methods to overcome the insufficient crop reference sample problem, tested a method that using existing ground-truth samples references from previous years combined with Sentinel 1 and Sentinel 2 imagery, and using Random Forest as selected machine learning classifier, mentioning that as well as many previous studies did because of its

mentioned condition of robustness to deal with scarce, noisy or unbalanced training data, while obtaining Overall Accuracy assessments over 90%, they concluded that regarding their results, ground-truth samples do not need to be collected on a regular year basis and that crop classification maps can be predicted or retraced by the inter-annual transfer of the trained model. On another example of use of databases constructed on previous years' field inspections, Arias et al, 2020 used CAP declarations as ground-truth data and by analyzing spectral signatures on Sentinel 1 images of the labeled fields, they generated a pattern to identify other non-sampled areas to perform classifications with Random Forest, reaching Overall Accuracies over 70%.

In other studies, as an attempt of presenting an automated workflow on Google Earth Engine (Gorelick et al, 2017), Zhang et al, 2022 uses information of historical Crop Data Layer time series, to extract “trusted pixels” through spectral signatures using R^2 coefficient between the reference data and the predicted result, to then use the extracted pixels as training samples for a supervised classification combined with present year Sentinel 2 imagery, which prediction is further validated by the in-season Crop Data Layer used as ground-truth data, assessing the accuracy of the prediction by Overall Accuracy, Kappa Index and Confusion Matrix analysis. Tang et al, 2022 used a methodology that through the use of limited ground-truth samples previously collected, intended to verify different deep learning models in large-scale using Sentinel 2 as remote sensing variables to conduct unsupervised learning based models accuracy comparison also including Random Forest, assessing the accuracy of the models with Overall Accuracies from 70%-95%, concluding to provide a technical reference for achieving the automatic national crop census.

Although the approaches are different, in terms of increasing the efficiency of the available data for crop classification predictions while considering resources efficiency, the presented AOA assessment methodology for optimizing ground-truth sampling allocation proved to have an advantage compared with other methodologies that also approached the efficiency of ground-truth sampling efforts. By using the area of applicability as a validation of classification model predictions there is an alternative to workflows relying exclusively on different models that include convolutional neural networks architectures and/or the

availability of historical and on season extensive databases (Crop Layer Data, CAP crop map data) to extract information for training or validating classification models or transferring models that were previously trained and validated with broad availability of ground-truth data. Datasets that can also contain a certain number of misclassified pixels, mixed pixels, and noisy pixels (Lin et al, 2022), and could potentially mislead the interpretation of results, reducing their accuracy or reliability. In addition, as far as these methodologies display, they do not explicitly offer layers to be used as a potential guide for spatially informed decisions regarding new ground-truth allocation efforts, since their focus is to reach automation by relying on unsupervised classification and extensive historical validations to transfer models as a source of training and or validation data extraction, which in most cases was cataloged as successful based on the traditional accuracy assessments performed over their predictions using spatiotemporally transferred information.

Regarding the accuracy assessments, taking into account traditional methods like Overall Accuracy, Producer's Accuracy, Kappa index, Confusion matrix, Root Mean Square Error, and other statistical validations well distributed among classification model performance measurements; although used to validate the accuracy of predictions results based on spatiotemporally transferred information, it is noticeable that these tools alone fell short in providing the complete level of visual insight for interpreting predictions reliability and confidence as the outcomes that could be reached when the AOA assessment is included as a classification model prediction validation. Disregarding this observation, the mentioned accuracy assessments could be considered to be enough to determine the validation of different model predictions, which results could be interpreted over optimistically and without a spatial notion on where the model prediction is not adjusting to the presented reference data and to which level.

4.3 Limitations

Although not demanding on extensive quantities of well-distributed ground-truth samples availability, the methodology does rely on the availability of ground-truth samples for initial predictions. However, the approach proved useful even when the references are limited since the analysis could start on a restricted area to then be expanded to surrounding areas or different years for further validations.

The AOA assessments over the analyzed classification predictions presented the challenge of being demanding regarding computational capacity and time for assessing the analysis, with over 8 hours for the `calibrate_aoa()` function over the Alto Paraná predictions. The assessments made over the study area including the three departments could not be done by a 12 gigabytes RAM laptop due lack of capacity and functions such as `aoa()` and `calibrate_aoa()` were done in alternative equipment with higher RAM capacity to obtain the final outcomes.

Up to the date of this study, the AOA analysis is only possible within R and is not available to be used in other environments or work frames.

The AOA methodology suggests to perform a predictor variables analysis to adjust the prediction results before applying the `aoa()` and `calibrate_aoa()` functions by weighting the variables used to run the model's prediction and assess which ones are the most influential for determining the prediction result and in that way, limit the prediction to the use of those variables. Partial analysis was practiced to accomplish this step but due time and the computational capacity restrictions to complete the workflow after this step, the variables weighting could not be included in the results. Although this step could not be integrated completely into the presented workflow for the reasons given above, to see the assessment of predictions that did not contemplate any refinements regarding variable selection could still be considered as relevant.

Cloud presence proved to be detrimental since its interferences would not allow it to express the predictors values on determined surfaces affected by these interferences, which is translated to areas that fall outside

the AOA and prevents the surface below from being assessed. A suggestion to improve this would be the integration of radar imagery such as the ones provided by Sentinel-1, due the immunity to cloud presence that this type of remote sensing source provides. Also, the cloud presence in combination with the different satellite footprints that compound the entire study area can cause differences between the same image composite by acquiring images from the entire area but at different dates for each footprint, which increases the negative effect of cloud presence as well.

The study did not contemplate any pre-process to clip out urban or protected areas from the remote sensing predictor variables.

4.4 Future scope

The expansion of the methodology by including other regions and other years' information for further assessment could be interesting to observe. Also, including information that could increase the validation of results applying guided ground-truth collection and re-train models would help to explore the potential of the methodology for being a reliable alternative for ground-truth sample allocation and detection of necessity.

Since the model contemplated predictor variables that included several different bands compiled into different images regarding several time frames, as well as several different vegetation indices, the use of a weighted analysis for determining the incidence of the different remote sensing variables on the classification result could be interesting to elucidate, in order to better adjust the model predictions before continuing with further validations.

Integrate Sentinel 1 radar imagery to avoid cloud interference, which although visible on the variables, are not evidenced on the classification prediction but are completely margined by the AOA assessment. This

integration could potentially help to allow predictions to be made without any interference that would not let the AOA be assessed over the actual surfaces below the clouds captured in the remote sensing variables

Although Random Forest seems the most appropriate machine learning algorithm option for integrating it into the classification model, based on the training data used and the spatiotemporal transferring of data, it could be interesting to evaluate other classification algorithms and assess the AOA over these predictions as a comparison.

Conclusion

The Area of Applicability assessment can be used to validate crop classification models by identifying areas outside the estimated limit and providing different levels of confidence for a prediction, which could help optimize ground-truth sampling gathering by taking into account the information revealed by this assessment and the accessibility to the extensions revealed by the different analysis for making decisions on where to direct the efforts and resources for further training on the region analyzed, disregarding the quality of the samples available in terms of distribution and population balance. Transferring models to other areas while assessing the area of applicability not only help to obtain an insightful vision on how and where is the transferred information used for the model is confident for making a prediction, and for having an auxiliary validation for interpret traditional accuracy assessments but also, the assessment of the AOA over predictions that are extrapolated to non-sampled areas showed how this technique could potentially improve the efficiency of resources for ground-truth samples collection, by identifying extensions that are in different levels of confidence, similar or not to the training data. Finally, transferring a model through time carries risks and is a questionable resort to be applied directly but, by assessing the area of applicability of the time transferred model predictions, the outcome could provide an additional validation source to traditional methods for a deeper evaluation that could be an initial step for generating and validating predictions over past years which do not have reference data or is restricted or limited. This methodology could potentially benefit studies that do not have access to historical datasets or cannot apply state-of-the-art methodologies to generate training data, as an alternative to reduce costs of surveys by offering an alternative to optimize ground-truth sampling necessity, providing information that could serve other considerations to guide the allocation of new samples.

Bibliography

- Ali, U., Esau, T.J., Farooque, A.A., Zaman, Q.U., Abbas, F. and Bilodeau, M.F. 2022. Limiting the Collection of Ground Truth Data for Land Use and Land Cover Maps with Machine Learning Algorithms. *ISPRS Int. J. Geo-Inf.*, 11, 333. <https://doi.org/10.3390/ijgi11060333>
- Anderson, K., Ryan, B., Sonntag, W., Kavvada, A. and Friedl, L. 2017. Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-Spatial Inf. Sci.*, 20, 77–96.
- Arias, M., Campo-Bescós, M.Á. and Álvarez-Mozos, J. 2020. Crop Classification Based on Temporal Signatures of Sentinel-1 Observations over Navarre Province, Spain. *Remote Sens.*, 12, 278. <https://doi.org/10.3390/rs12020278>
- Bannari, A., Morin, D., Bonn, F. and Huete, A. R. 1995. A review of vegetation indices. *Remote Sensing Reviews*, 13:1-2, 95-120, DOI: 10.1080/02757259509532298
- Bauer, M. E. 1975. *Advances in Agronomy. Volume 27: The Role of Remote Sensing In Determining The Distribution and Yield of Crops*, 271–304. doi:10.1016/s0065-2113(08)70012-9
- Belgiu, M., Bijker, W., Csillik, O. and Stein, A. 2021. *Phenology-based sample generation for supervised crop type classification. International Journal of Applied Earth Observation and Geoinformation*, 95(), 102264– doi: 10.1016/j.jag.2020.102264
- Bingfang, W. 2020. Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. *Geogr. Sustain.*, 1, 25–32.
- Boryan, C., Yang, Z., Mueller, R. and Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* 26 (5), 341–358.
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), pp.5-32.
- Cámara Paraguaya de Exportadores y Comercializadores de Cereales y Oleaginosas – CAPECO. 2021. Área de siembra, producción y rendimiento. Disponible en <http://capeco.org.py/areade-siembra-produccion-y-rendimiento/>
- Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G. and Gilabert, M.A., 2020. Understanding deep learning in land use classification based on Sentinel-2A time series. *Scientific reports*, 10(1), pp.1-12.
- Chakhar, A., Ortega-Terol, D., Hernández-López, D., Ballesteros, R., Ortega, J. F., Moreno, M. A. 2020. Assessing the Accuracy of Multiple Classification Algorithms for Crop Classification Using Landsat-8 and Sentinel-2 Data. *Remote Sensing*, 12(11), 1735–. doi:10.3390/rs12111735
- Chang, J., Hansen, M.C., Pittman, K., Carroll, M. and DiMiceli, C. 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. *Agron. J.*, 99, 1654–1664.
- Cohen, J. A. 1960. Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.*, 20, 37–46.
- Comber, A. and Wulder, M. 2019. Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, (), tgis.12559–. doi:10.1111/tgis.12559
- Congalton, R. G. 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sens. Environ.* 37:35-46.
- Cortes, C. and Vapnik, V. 1995. Support-vector networks. *Machine Learning*, 20(3), 273–297.

- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L. and Rabaute, T. 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.*, 221, 551–568.
- Dell’Acqua, F., Iannelli, G.C., Torres M.A. and Martina, M.L.V. 2018. “A Novel Strategy for Very-Large-Scale Cash-Crop Mapping in the Context of Weather-Related Risk Assessment, Combining Global Satellite Multispectral Datasets, Environmental Constraints, and in Situ Acquisition of Geospatial Data.” *Sensors (Switzerland)* 18 (2): 2. doi:10.3390/s18020591.
- Dolan M.F.J., Ross, R.E., Albretsen, J., Skarðhamar, J., Gonzalez-Mirelis, G., Bellec, V.K., Buhl-Mortensen, P. and Bjarnadóttir, L.R. 2021. Using Spatial Validity and Uncertainty Metrics to Determine the Relative Suitability of Alternative Suites of Oceanographic Data for Seabed Biotope Prediction. A Case Study from the Barents Sea, Norway. *Geosciences*, 11, 48. <https://doi.org/10.3390/geosciences11020048>
- FAO. 2022. FAOSTAT: Macro Indicators. In: FAO. Rome. Cited May 2022. <http://www.fao.org/faostat/en/#data/MK>
- FAO, IFAD, UNICEF, WFP and WHO. 2022. The State of Food Security and Nutrition in the World 2022. Repurposing food and agricultural policies to make healthy diets more affordable. Data available on FAOSTAT. <https://www.fao.org/faostat/en/#data/FDs>
- Fawwaz U.T., Li, R.Y. and Shanmugan, K. S. 1982. Crop Classification Using Airborne Radar and Landsat Data. *GE-20*(1), 0–51. doi:10.1109/tgrs.1982.4307519
- Fitzgerald R.W. and Lees B.G. 1994. Assessing the classification accuracy of multisource remote sensing data. , 47(3), 362–368. doi:10.1016/0034-4257(94)90103-1
- Footy G.M., Curran P.J., Groom G.B. and Munro D.C. 1988. Crop Classification with Multi-Temporal X-Band SAR data. *Proceedings of IGARSS ’88 Symposium*. Edinburgh, Scotland
- Fowler, J., Waldner, F. and Hochman, Z. 2020. All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods, *International Journal of Applied Earth Observation and Geoinformation*, Volume 91, 102114, ISSN 1569-8432, <https://doi.org/10.1016/j.jag.2020.102114>.
- Gallo, I., Ranghetti, L., Landro, N., La Grassa, R., Boschetti, M. 2023. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 195, Pages 335-352, ISSN 0924-2716, <https://doi.org/10.1016/j.isprsjprs.2022.12.005>.
- Gao, B. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. , 58(3), 257–266. doi:10.1016/s0034-4257(96)00067-3
- Gilbertson, J. K., Kemp, J. and Van Niekerk, A. 2017. “Effect of Pan-Sharpening Multi-Temporal Landsat 8 Imagery for Crop Type Differentiation Using Different Classification Techniques.” *Computers and Electronics in Agriculture* 134: 151–159. doi:10.1016/j.compag.2016.12.006.
- Gitelson, A.A., Kaufman, Y.J. and Merzlyak, M.N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. , 58(3), 289–298. doi:10.1016/s0034-4257(96)00072-7
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R. 2016 Google Earth Engine: Planetary-scale geospatial analysis for everyone, *Remote Sensing of Environment*, <http://dx.doi.org/10.1016/j.rse.2017.06.031>
- Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J. and Dextraze, L. 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. , 81(2-3), 416–426. doi:10.1016/s0034-4257(02)00018-4

- Hendricks, N. P. and Er, E. 2018. Changes in cropland area in the United States and the role of CRP. *Food Policy*, 75, 15-23.
- Hoffer, R. M. 1972. The importance of ground truth data in remote sensing. UN Panel Meeting on the Estab. and Implementation of Res. Programs in Remote Sensing (No. NASA-CR-130020).
- Hu, Y., Zeng, H., Tian, F., Zhang, M., Wu, B., Gilliams, S., Li, S., Li, Y., Lu, Y. and Yang, H. 2022. "An Interannual Transfer Learning Approach for Crop Classification in the Hetao Irrigation District, China" *Remote Sensing* 14, no. 5: 1208. <https://doi.org/10.3390/rs14051208>
- Huang, H., Wang, J., Liu, C., Liang, L., Li, C. and Gong, P. 2020. The migration of training samples towards dynamic global land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161(), 27–36. doi:10.1016/j.isprsjprs.2020.01.01
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. , 83(1-2), 195–213. doi:10.1016/s0034-4257(02)00096-2
- Jahan, F., and M. Awrangjeb. 2017. "Pixel-Based Land Cover Classification by Fusing Hyperspectral and LiDAR Data." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42 (2W7): 711–718. doi:10.5194/isprs-archives-XLII-2-W7-711-2017.
- Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y. 2018. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* 10, 75. <https://doi.org/10.3390/rs10010075>
- Kang, J., Zhang, H., Yang, H. and Zhang, L., 2018. Support Vector Machine Classification of Crop Lands Using Sentinel-2 Imagery. 1-6. 10.1109/Agro-Geoinformatics.2018.8476101.
- Kluger, D.M., Wang, S., and Lobell, D.B. 2021. Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. *Remote Sensing of Environment*, -. doi:10.1016/j.rse.2021.112488
- Kobayashi, N., Tani, H., Wang, X. and Sonobe, R. 2020. Crop classification using spectral indices derived from Sentinel-2A imagery, *Journal of Information and Telecommunication*, 4:1, 67-90, DOI: 10.1080/24751839.2019.1694765
- Koetz, B., Defourny, P., Bontemps, S., Bajec, K., Cara, C., Kucera, L., Malcorps, P. and Milcinski, G. 2019. SEN4CAP Sentinels for CAP monitoring approach. In *Proceedings of the 2019 JRC IACSWorkshop*, Valladolid, Spain, 10–11 April 2019
- Lary, D. J., Alavi, A. H., Gandomi, A. H. and A. L. Walker. 2016. "Machine Learning in Geosciences and Remote Sensing." *Geoscience Frontiers* 7 (1): 3–10. doi:10.1016/j.gsf.2015.07.003
- Li, H., Wang, G., Dong, Z., Wei, X., Wu, M., Song, H. and Amankwah, S.O.Y. 2021. Identifying Cotton Fields from Remote Sensing Images Using Multiple Deep Learning Networks. *Agronomy*, 11, 174. <https://doi.org/10.3390/agronomy11010174>
- Lin, C., Zhong, L., Song, X. P., Dong, J., Lobell, D. B. and Jin, Z. 2022. Early- and in-season crop type mapping without current-year ground truth: Generating labels from historical information via a topology-based approach. *Remote Sensing of Environment*, 274, [112994]. <https://doi.org/10.1016/j.rse.2022.112994>
- Liu, P. 2015. A survey of remote-sensing big data. *Front. Environ. Sci.*, 3, 1–6.
- Liu, Z., Zhang, L., Yu, Y., Xi, X., Ren, T., Zhao, Y., Zhu, D. and Zhu, A. 2022. Cross-Year Reuse of Historical Samples for Crop Mapping Based on Environmental Similarity. *Front. Plant Sci.* 12:761148. doi: 10.3389/fpls.2021.76114

- Luo, K., Lu, L., Xie, Y., Chen, F., Yin, F. and Li, Q. 2023. Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning. *Computers and Electronics in Agriculture*, Volume 205, 107577, ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2022.107577>.
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A. and Jie, W. 2015. Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51, 47– 60.
- Massey, R., Sankey, T. T., Congalton, R. G., Yadav, K., Thenkabail, P. S., Ozdogan, M. and Meador, A. J. S. 2017. MODIS phenology-derived, multi-year distribution of conterminous US crop types. *Remote Sensing of Environment*, 198, 490-503.
- Mathur, A. and Foody, G.M., 2008. Crop classification by support vector machine with intelligently selected training data for an operational application, *International Journal of Remote Sensing*, 29:8, 2227-2240, DOI: 10.1080/01431160701395203
- Maxwell, A.E., Warner, T.A., and Fang, F. 2018. Implementation of machine-learning classification in remote sensing: an applied review, *International Journal of Remote Sensing*, 39:9, 2784-2817, DOI: 10.1080/01431161.2018.1433343
- McCulloch, W. S. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133
- Meyer H. and Pebesma. E. 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution* –. doi:10.1111/2041-210x.13650
- Meyer, H. and Pebesma, E. 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*. 13.
- Meyer, H., Milà, C., Ludwig, M. and Linnenbrink, J. 2023. CAST: 'caret' Applications for Spatial-Temporal Models. <https://github.com/HannaMeyer/CAST>, <https://hannameyer.github.io/CAST/>.
- Nijland, W., de Jong, R., de Jong, S. M, Wulder, Michael, A., Bater, Chris, W. and Coops, N. C. 2014. Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras. *Agricultural and Forest Meteorology*, 184(), 98–106. doi:10.1016/j.agrformet.2013.09.007
- Odenweller, J.B. and Johnson, K.L. 1984. Crop identification using Landsat temporal-spectral profiles. 14(1-3), 39–54. doi:10.1016/0034-4257(84)90006-3
- Ok, A.O., Akar, O. and Gungor, O. 2012- Evaluation of random forest method for agricultural crop classification, *European Journal of Remote Sensing*, 45:1, 421-432, DOI: 10.5721/EuJRS20124535
- Orynbaikyzy, A., Gessner, U. and Conrad, C. 2019. Crop type classification using a combination of optical and radar remote sensing data: A review. *international journal of remote sensing*, 40(17), 6553-6595.
- Park, J.K. and Park, J.H. 2015. Crops Classification Using Imagery of Unmanned Aerial Vehicle (UAV). *Journal of The Korean Society of Agricultural Engineers*, 57(6), 91–97. <https://doi.org/10.5389/KSAE.2015.57.6.091>
- Pengyu, H., Liping, D., Chen, Z. and Liying, G. 2020 . Transfer learning for crop classification with Cropland Data Layer data (CDL) as training samples, *Science of the Total Environment*, <https://doi.org/10.1016/j.scitotenv.2020.138869>
- Perales, F.J., Campilho, A.J.C., de la Blanca, N.P. and Sanfeliu, A. 2003. [Lecture Notes in Computer Science] Pattern Recognition and Image Analysis Volume 2652 || Support Vector Machines for Crop Classification Using Hyperspectral Data. , 10.1007/b12122(Chapter 16), 134–141. doi:10.1007/978-3-540-44871-6_16

- Pott L.P., Amado T.J.C., Schwalbert R.A., Corassa G.M., Ciampitti I.A. 2022. Crop type classification in Southern Brazil: Integrating remote sensing, crop modeling and machine learning. *Computers and Electronics in Agriculture*. Volume 201. 107320, ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2022.107320>.
- Prasad, S., Bruce, L.M. and Chanussot, J. 2011. Introduction. *Optical Remote Sensing*, 1–8. doi:10.1007/978-3-642-14212-3_1
- Prins, A.J., and van Niekerk, A. 2020. Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms. *Geo-spatial Information Science*, 1–13. doi:10.1080/10095020.2020.1782776
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. and Sorooshian, S. 1994. A modified soil adjusted vegetation index. , 48(2), 119–126. doi:10.1016/0034-4257(94)90134-1
- Qihao, W. 2013. "Introduction to Remote Sensing Systems, Data, and Applications". doi: 10.1201/b15159-3
- R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rao, P., Zhou, W., Bhattarai, N., Srivastava, A.K., Singh, B., Poonia, S., Lobell, D.B. and Jain, M. 2021. Using Sentinel-1, Sentinel-2, and Planet Imagery to Map Crop Type of Smallholder Farms. *Remote Sens.*, 13, 1870. <https://doi.org/10.3390/rs13101870>
- Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. 1974 Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS-1 Symposium NASA, NASA SP-351, Washington DC, 309-317.
- Rußwurm, M., Wang, S., Korner, M. and Lobell, D., 2020. Meta-learning for few-shot landcover classification. In: *Proceedings of the IEEE/Cvf Conference on Computer Vision and Pattern Recognition Workshops*, pp. 200–201.
- Siesto, G., Fernández-Sellers, M. and Lozano-Tello, A. 2021. Crop Classification of Satellite Imagery Using Synthetic Multitemporal and Multispectral Images in Convolutional Neural Networks. *Remote Sens.* 13, 3378. <https://doi.org/10.3390/rs13173378>
- Sitokonstantinou, V., Papoutsis, I., Kontoes, C., Arnal, A.L., Andrés, A.P.A. and Zurbano, J.A.G. 2018. Scalable parcel-based crop identification scheme using Sentinel-2 data time-series for the monitoring of the common agricultural policy. *Remote Sens.*, 10, 911
- Son, N.T., Chen, C.F., Chen, C.R., Toscano, P., Cheng, Y.S., Guo, H.Y. and Syu, C.H. 2021 A phenological object-based approach for rice crop classification using time-series Sentinel-1 Synthetic Aperture Radar (SAR) data in Taiwan, *International Journal of Remote Sensing*, 42:7, 2722-2739, DOI: 10.1080/01431161.2020.1862440
- Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V. and Hansen, M.C. 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sens. Environ.*, 190, 383–395.
- Su, T. and Zhang, S., 2021. Object-based crop classification in Hetao plain using random forest. *Earth Sci Inform* 14, 119–131. <https://doi.org/10.1007/s12145-020-00531-z>
- Tang, J., Zhang, X., Chen, Z. and Bai, Y. 2022. Crop Identification and Analysis in Typical Cultivated Areas of Inner Mongolia with Single-Phase Sentinel-2 Images. *Sustainability*, 14, 12789. <https://doi.org/10.3390/su141912789>
- United Nations Department of Economic and Social Affairs, Population Division. 2022. World Population Prospects 2022: Summary of Results. UN DESA/POP/2022/TR/NO. 3.
- Viskovic, L., Kosovic, I.N. and Mastelic, T. 2019. IEEE 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) - Split, Croatia (2019.9.19-2019.9.21). 2019 International

- Conference on Software, Telecommunications and Computer Networks (SoftCOM) - Crop Classification using Multi-spectral and Multitemporal Satellite Imagery with Machine Learning. 1–5. doi:10.23919/SOFTCOM.2019.8903738
- Waldner, F., Fritz, S., Di Gregorio, A., Plotnikov, D., Bartalev, S., Kussul, N., Gong, P., Thenkabail, P., Hazeu, G., Klein, I., Löw, F., Miettinen, J., Dadhwal, V., Lamarche, C., Bontemps, S. and Defourny, P. 2016. *A Unified Cropland Layer at 250 m for Global Agriculture Monitoring. Data*, 1(1), 3-doi:10.3390/data1010003
- Wang, S., Azzari, G. and Lobell, D. B. 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222(), 303–317. doi:10.1016/j.rse.2018.12.026
- Wang, H., Chang, W., Yao, Y., Liu, D., Zhao, Y., Li, S., Liu, Z., Zhang, X. 2022. "CC-SSL: A Self-Supervised Learning Framework for Crop Classification With Few Labeled Samples" in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Vol. 15. pp. 8704-8718. doi: 10.1109/JSTARS.2022.3211994.
- Xie, Y., Sha, Z. and Yu, M. 2008. Remote sensing imagery in vegetation mapping: A review. *J. Plant. Ecol.*, 1, 9–23.
- Zhang, C., Di, L., Lin, H., Li, L., Guo, Z., Yang, G.Y., Eugene, Y. and Di, A. Yang. 2022. Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data. *Agric. Syst.*, 201, Article 103462
- Zhang, X., Sun, Y., Shang, K., Zhang L. and Wang, S. 2016. "Crop Classification Based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4117-4128, doi: 10.1109/JSTARS.2016.2577339.
- Zhang, X., Yang, G., Xu, X., Yao, X., Zheng, H., Zhu, Y., Cao, W. and Cheng, T. 2021. An assessment of Planet satellite imagery for county-wide mapping of rice planting areas in Jiangsu Province, China with one-class classification approaches, *International Journal of Remote Sensing*, 42:19, 7610-7635, DOI: 10.1080/01431161.2021.1964710
- Zhong, L.H., Gong, P. and Biging, G.S. 2012. Phenology-based Crop Classification Algorithm and its Implications on Agricultural Water Use Assessments in California's Central Valley. *Photogrammetric Engineering and Remote Sensing*; 78: 799-813.