

# **DEVELOPMENT OF A SPATIALLY EXPLICIT ACTIVE LEARNING METHOD FOR CROP TYPE MAPPING FROM SATELLITE IMAGE TIME SERIES**

BEATRICE ANTHONY KAIJAGE

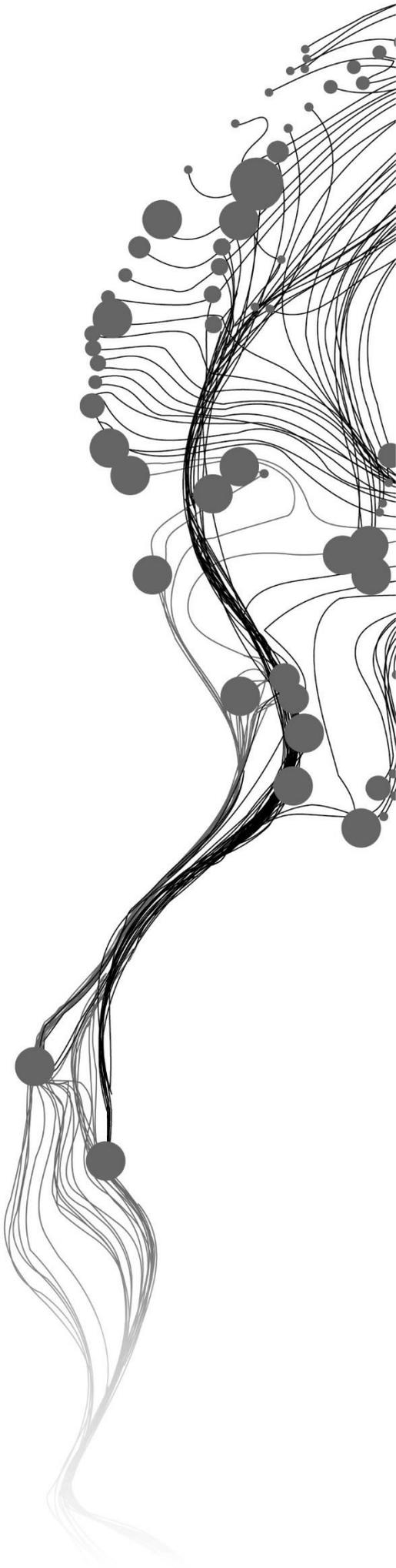
July, 2021

SUPERVISORS:

Dr M. Belgiu

Dr. ir.W. Bijker





# **DEVELOPMENT OF A SPATIALLY EXPLICIT ACTIVE LEARNING METHOD FOR CROP TYPE MAPPING FROM SATELLITE IMAGE TIME SERIES**

**BEATRICE ANTHONY KAIJAGE**

Enschede, The Netherlands, July, 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Geoinformatics

**SUPERVISORS:**

Dr M. Belgiu

Dr. ir.W. Bijker

**THESIS ASSESSMENT BOARD:**

Prof.dr.ir. A. Stein (Chair)

Dr.ir. T.A. Groen (External Examiner, NRS-ITC)

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

## ABSTRACT

Insufficient training samples for effective classification is one of the drawbacks of existing supervised classification methods. Collecting training samples via field campaigns is time-consuming and costly, especially when gathering data from a vast area. As a result, Remote Sensing(RS) requires approaches that can work with a small number of training samples while still providing excellent accuracy. Active Learning (AL) is one of these approaches. AL is a Machine Learning(ML) method whose purpose is to attain satisfactory classification results with a small number of training datasets, resulting in accurate information extraction at low annotation costs. AL reduces the training sample size required for training a classifier up to tenfold by identifying the most informative and diverse samples from a set of unlabeled samples. Informative samples are those for which a classifier has difficulty classifying or labeling them, and sample diversity refers to how dissimilar the selected samples are from one another. Most of the existing AL approaches are dedicated to querying informative samples based on their spectral characteristics, neglecting spatial information. This research aims to develop a spatially explicit AL method for crop type mapping using Satellite Image Time Series(SITS) and assesses its performance compared to the existing AL techniques that ignore the spatial component in the selection of informative samples. The developed AL method that includes the spatial component and the AL technique that excludes the spatial component were both evaluated using crop data and Sentinel-2 time-series images collected in 2019. The two AL techniques were compared to the classification performance obtained utilizing the whole training dataset. The AL method with the spatial component used 27% of the entire training sample dataset and 57% of the informative training samples acquired from the AL method that excludes the spatial component to achieve an overall accuracy of 80%. This accuracy is almost identical to the overall accuracy of the AL method without the spatial component (82%) and when using the entire data set (84%). Comparisons were made using other metrics like Kappa statistic, user's and producer's accuracy and quality of the sample design. The developed spatially explicit AL method showed a good performance with a low number of samples. In addition, it performed better in the case of crop types with high interclass similarities like potatoes and maize. A challenge was faced in classifying mixed classes consisting of different land cover classes.

Given these findings, adding the spatial component in AL is a critical contribution to the field of agriculture, especially in developing countries where we do not have access to a large number of samples required for accurate crop mapping and monitoring due to the high cost of sample acquisition.

Keywords: Active Learning, machine learning, satellite image time series, variogram, crop types.

## ACKNOWLEDGEMENTS

First and foremost, I want to express my gratitude to God Almighty for His protection, direction, and unconditional love during my time in the Netherlands.

Heartfelt gratitude is extended to my supervisors, Dr Mariana Belgiu and Dr. ir. Wietske Bijker for their endless support, guidance, encouragement, and constructive recommendations throughout my research period. Amid Corona, their efforts to have regular discussions and their deep concern were pillars of my research.

Apart from my supervisors, I'd like to express my gratitude to the remainder of the thesis committee, particularly prof.dr.ir. A. Stein, Drs. J.P.G. Bakx (Wan), and dr. ir. T.A. Groen for their interesting remarks and questions.

I appreciate the GFM staff's lectures, plus all ITC staff who shared their knowledge with me through practical teachings and the provision of extensive experience in the field of geospatial sciences. I was able to complete this research using the knowledge I acquired from them.

Special thanks go to my late dad, Anthony Rwebangira Kaijage, who passed away in January this year. He has been a strong pillar, significant supporter and motivation to my life and my research too. May his soul rest in peace.

Deep gratitude is extended to my precious daughter, Gloria Sekela, and my mother, Telephosa Anthony, for their continuous prayers and encouragement during my research. Heartfelt gratitude is extended to my dear brothers (Victor, Robert, Christopher, Emmanuel, Adolph and Martin) and sisters (Redempter, Mariadolla and Adolphina) for their moral support and love.

In a nutshell, I would like to express my gratitude to all of my friends at ITC, particularly Catherine Nabukulu, Stephen Akinremi, Dawit Beyene, Hassan Oladapo, Patience Musila and Vasudha Chaturvedi, as well as all friends back home, for their continued presence in my life and unwavering support during my research period.

---

## TABLE OF CONTENTS

---

1.	INTRODUCTION.....	1
1.1.	Motivation & problem statement.....	1
1.2.	Research objective and questions.....	2
1.3.	Organization of the thesis.....	3
2.	LITERATURE REVIEW.....	4
2.1.	Introduction.....	4
2.2.	Active Learning.....	4
2.3.	Active Learning scenarios.....	5
2.4.	Spectral-domain heuristics and metrics in AL.....	5
2.5.	Incorporating the spatial component in AL.....	7
2.6.	Spatial-domain heuristics and metrics in AL.....	7
2.7.	Classification.....	8
3.	STUDY AREA DESCRIPTION, DATA ACQUISITION AND PREPROCESSING.....	10
3.1.	Study area description.....	10
3.2.	Data description.....	11
3.3.	Ground truth data acquisition and preprocessing.....	11
3.4.	Image acquisition and preprocessing.....	13
4.	METHODOLOGY.....	18
4.1.	Workflow of the thesis.....	18
4.2.	Active learning components.....	19
4.3.	Active Learning considering the spectral domain only.....	20
4.4.	Incorporating the spatial component in AL.....	21
5.	RESULTS.....	27
5.1.	Classification using the entire dataset of training samples.....	27
5.2.	Classification using training samples generated from spectral-domain AL.....	29
5.3.	Classification using the AL algorithm with the spatial component.....	34
5.4.	Comparison of the AL algorithms.....	38
6.	DISCUSSION.....	42
6.1.	Summary of findings.....	42
6.2.	Strengths of the developed algorithm.....	42
6.3.	Weaknesses of the algorithm.....	43
6.4.	Opportunities.....	43
6.5.	Threat.....	44
7.	CONCLUSION AND RECOMMENDATIONS.....	45
7.1.	Conclusion.....	45
7.2.	Limitations.....	46
7.3.	Recommendations.....	46
	LIST OF REFERENCES.....	47
	APPENDIX A.....	52
	APPENDIX B.....	53

## LIST OF FIGURES

Figure 3-1 Location map of the study area (Source: Google Earth).....	10
Figure 3-2 Agriculture plot distribution over the study area. The other areas are other types of landcover present in the study area.....	12
Figure 3-3 Number of plots per crop type class .....	12
Figure 3-4 Training and validation sample locations for the nine target classes.....	13
Figure 3-5 Illustration of the Earth Engine App used to get an overview of the cloud percentage values for all images in each month.....	14
Figure 3-6 Image acquisition dates with respect to cloud cover percentage. ....	15
Figure 3-7 Temporal profiles for a) alfalfa, b) beets, c) cereals, d) maize, e) orchard, f) onions, g) potatoes, h) a combination of all target classes.....	16
Figure 4-1 Flow chart of the proposed methodology of the research.....	18
Figure 4-2 Active Learning components .....	19
Figure 4-3 Distribution of target classes in feature space before applying Active Learning.....	19
Figure 4-4 AL workflow involving the spatial component in querying informative samples.....	21
Figure 4-5 Crop type spatial distribution in geographical space before AL .....	21
Figure 4-6 Variogram model parameters, range, sill, nugget, and partial sill. ....	23
Figure 4-7 Gaussian model fitted from the data for one month.....	24
Figure 4-8 Exponential model fitted from the data for one month.....	24
Figure 4-9 Spherical model fitted from the data for one month.....	25
Figure 5-1 Assessment of the sample classification results obtained by using the entire dataset of training samples. UA – User’s Accuracy, PA – Producer’s Accuracy, OA – Overall Accuracy .....	27
Figure 5-2 Crop type map generated from classification using all training samples (350 samples).....	28
Figure 5-3 Initial predictions per individual learner. Y and X axis stand for principal component analysis components 2 and 1 respectively as created as a result of reducing the spectral features to two dimensions .....	29
Figure 5-4 Committee predictions by initial 40 training samples. ....	29
Figure 5-5 Prediction accuracy with an increment of training samples. Y axis is the increment in classification accuracy and X axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.....	30
Figure 5-6 committee prediction accuracy after 129 queries. Y axis is the increment in classification accuracy and X axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples. ....	30
Figure 5-7 Crop sample distribution before AL .Redundancy is illustrated by the colored oval shapes potatoes(purple oval), cereals(green oval), alfalfa and beets(dark blue oval), potatoes and onions(red oval) and the spectral mixture of potatoes, beets, and maize(yellow oval).....	31
Figure 5-8 Crop sample distribution after AL in the spectral domain. Redundancy reduction shown by the colored oval shapes potatoes(purple oval), cereals(green oval), alfalfa and beets(dark blue oval), potatoes and onions(red oval) and the spectral mixture of potatoes, beets, and maize(yellow oval). ....	31
Figure 5-9 Assessment of the sample classification results obtained by applying the AL that considers the spectral domain only. UA – User’s Accuracy, PA – Producer’s Accuracy, OA – Overall Accuracy .....	32
Figure 5-10 Crop type map generated from training samples selected using the AL algorithm that considers the spectral domain only .....	33
Figure 5-11 Prediction accuracy with an increment of training samples obtained from AL algorithm with the spatial component. Y- axis is the increment in classification accuracy and X- axis displays the number	

of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.....	34
Figure 5-12 Committee prediction accuracy after 129 queries considering the spectral and spatial domain. Y- axis is the increment is classification accuracy and X- axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.....	35
Figure 5-13 Crop sample spatial distribution before AL. Redundancy illustrated by the colored oval shapes orchard (red oval), alfalfa(dark blue oval), maize(light green oval), cereals(yellow oval) and potatoes(purple oval). .....	36
Figure 5-14 Crop sample distribution after AL in the spectral and spatial domain. Redundancy reduction is shown by the colored oval shapes orchard(red oval), alfalfa(dark blue oval), maize(light green oval), cereals(yellow oval) and potatoes(purple oval). .....	36
Figure 5-15 Assessment of the sample classification results obtained by applying the AL that considers the spectral and spatial domain. UA – User’s Accuracy, PA – Producer’s Accuracy, OA – Overall Accuracy	37
Figure 5-16 Crop type map generated from training samples selected using the AL algorithm that considers the spectral and spatial domain.....	37
Figure 5-17 Comparison of the user’s accuracies for the AL algorithm with the spatial component(gray), AL algorithm without the spatial component (orange) to the reference accuracy obtained using the entire dataset.....	40
Figure 5-18 Misclassification of cereals class by the AL algorithm that includes the spatial component (first row),orchard class by AL algorithm that excludes the spatial component(second row) and a misclassification by both AL algorithms(third row). .....	41

## LIST OF TABLES

---

Table 2-1 Active Learning scenarios description .....	5
Table 3-1 Data description .....	11
Table 3-2 Band properties of acquired Sentinel images.....	14
Table 4-1 Estimated variogram parameters, $np$ (the number of point pairs per lag), $dist$ (lag distance) and gamma (mean of the variogram cloud points in the bin).....	22
Table 4-2 Sum of Square Error values for monthly variograms considering all variogram model types (Exponential, Spherical and Gaussian) .....	24
Table 4-3 Variation of the ranges with their respective nugget values for the variograms fitted for each month.....	26
Table 5-1 Training sample class distribution before AL and after AL considering the spectral domain only .....	32
Table 5-2 Training sample class distribution before AL and after AL considering the spectral and spatial domain.....	35
Table 5-3 Overall accuracies and Kappa statistics for the classification run using all training samples, samples from the AL method that excludes the spatial component and samples from the AL method that includes the spatial component.....	39
Table 5-4 Comparison of the user's accuracies for the three scenarios per crop type class. ....	39
Table 5-5 Comparison of the misclassified samples for the AL method that excludes the spatial component (169) and samples from the AL that includes the spatial component (97) .....	40
Table 5-6 Variation of the number of samples used for classification for the AL with spatial component and the AL algorithm that excludes the spatial component.....	41

# 1. INTRODUCTION

## 1.1. Motivation & problem statement

Hunger is one of the most sensitive issues that affect human life globally. Previous studies revealed that 11% of people worldwide suffer from intense hunger and malnutrition (Cervantes-Godoy et al., 2014). The Sustainable Development Goal 2 (SDG2), "Zero Hunger", deals with all matters relating to food security and efficient food production to combat hunger. One of the strategies to attain food security is the proper management of food sources, for example, establishing resilient agricultural systems (Wu, Ho, Nah, & Chau, 2014). Efficient agricultural systems are maintained through monitoring, proper planning, and management.

Remote sensing (RS) has made it possible to acquire information that is useful for the efficient management of agricultural systems through various data sources and satisfactory information processing methods (Treitz & Rogan, 2004). In the field of agriculture, RS has been used for crop mapping, crop monitoring, assessing crop risks, and crop yield prediction, among others (Khanal, Kc, Fulton, Shearer, & Ozkan, 2020). RS has been applied in large scale crop mapping based on multisource datasets (X. Liu et al., 2020), crop productivity assessment (Richter, Agostini, Barker, Costomiris, & Qi, 2016), active and fallow lands determination (Xie, Tian, Granillo, & Keller, 2007), crop quantity monitoring (Khan, de Bie, van Keulen, Smaling, & Real, 2010), crop yield assessment (Shanmugapriya, Rathika, Ramesh, & Janaki, 2019) and crop classification (Belgiu & Csillik, 2018).

Image classification is one of the RS techniques suitable for crop mapping since it helps recognize patterns that exist in the real world from images (Radhika, 2016). Several image classification approaches exist, namely unsupervised, semi-supervised and supervised methods (D. Lu & Weng, 2007). Unsupervised methods create classes in an image using the spectral characteristics inherent in the image solely. The user is not required to provide training data (Naghdy, Todd, Olaode, & Naghdy, 2014). Semi-supervised classification falls between unsupervised and supervised classification approaches. It is suitable when the training samples are insufficient, whereby informative unlabeled samples are identified, assigned to the target classes and further used to classify the image iteratively (Xiong, Zhang, & Jiang, 2010). Semi-supervised classification and Active Learning (AL) are machine learning methods used for classification (Bruzzone & Persello, 2010). The fundamental distinction between these two methods is that the semi-supervised approach labels the informative pixels iteratively automatically without user interaction whereas, the AL approach involves interaction between the user and the system (Bruzzone & Persello, 2009). The supervised approach relies on training data to learn the characteristics of the specific classes of interest from the remote sensing dataset (Miranda, Mutiara, Ernastuti, & Wibowo, 2018).

One of the drawbacks of existing supervised classification methods is insufficient training samples for effective classification (Stumpf, Lachiche, Malet, Kerle, & Puissant, 2014). Collecting training samples through field campaigns is a time-consuming and expensive task, especially when we want to extract information from a large area. Therefore, RS requires methods that can operate with a low number of training samples and yet give high accuracy (Ball, Anderson, & Wei, 2018). Various techniques like Transfer Learning (TL), AL (Cao, Yao, Xu, & Meng, 2020), among others, have been used to address the issue of insufficient training samples. Transfer learning involves using information from already acquired images known as the source domain to classify newly acquired images (target domain) whose information

is not available (Begüm Demir, Bovolo, & Bruzzone, 2013b). The target domain might be an image of the same area taken at different times or an area with a different but related distribution to the source domain (Pan & Yang, 2010). The AL method enriches the training sample set repeatedly by selecting informative samples from the unlabeled samples set, labelling them, adding them to the labelled sample set and then performing image classification (Tuia, Ratle, Pacifici, Kanevski, & Emery, 2009). As a result of the classification of images, various maps are generated depending on the purpose of the study. In this research, AL will be used for crop type mapping using Satellite Image Time Series (SITS). This will aid continuous monitoring, planning, and proper management of crop areas leading to efficient food production, hence addressing hunger.

As mentioned above, AL heuristics rely on selecting the most informative samples from the unlabeled sample pool iteratively (Begüm Demir, Bovolo, & Bruzzone, 2012). Informative samples are the samples for which a classifier has a hard time predicting their classes or labels. The great majority of previous studies have used AL methods considering only the spectral characteristics in the feature space when selecting unlabeled informative samples (Rajan, Ghosh, & Crawford, 2008). However, they may be spatially close to each other, reflecting a high probability of being similar, leading to redundancy in their selection. The inclusion of spatial information in AL might prevent the selection of redundant samples. Unfortunately, most AL heuristics do not account for spatial information when selecting the most informative samples (Xue, Zhou, & Zhao, 2018). Previous studies proved that incorporating both the spectral and spatial information in AL gives more robust results than those that rely on spectral measurements only (Patra, Bhardwaj, & Bruzzone, 2017). Therefore, the use of both spectral and spatial components as criteria for selecting informative unlabeled samples improves the efficiency of AL methods. In Pasolli et al.(2011), a criterion based on spatial information was proposed and combined with the spectral criteria in selecting informative samples for AL. The first heuristic calculated the distance of the samples to the support vectors in the feature space, and the second heuristic calculated the distance of the samples to the nearest support vector in the spatial domain. The two combined heuristics generated highly uncertain samples in the feature space but far from support vectors spatially. The classification accounting for the combined heuristics gave higher accuracy than the classification that excluded the spatial component. Our research is dedicated to investigating different solutions to integrate spatial information in AL. Thus, we aim to develop an AL method that incorporates both the spatial and spectral domain for application in crop type mapping from satellite image time series.

## 1.2. Research objective and questions

The main goal of the research topic is to develop a spatially explicit AL method for crop type mapping using satellite image time series. Thus the research aims to address the following specific objectives and the related questions:

**Objective 1:** To systematically investigate different spatial metrics that can be used to improve state-of-the-art AL methods.

Research question 1.1: What metrics can be used to assess spatial autocorrelation between the labels in the spatial domain?

Research question 1.2. : What criteria should be considered in choosing the best metrics for assessing the spatial autocorrelation between the labels in the spatial domain?

**Objective 2:** To test the developed AL method's performance and assess its effectiveness in crop type mapping for satellite image time series.

Research question 2.1: How does the developed AL method perform in comparison to the AL algorithm that excludes the spatial component?

### **1.3. Organization of the thesis**

The rest of the thesis comprises six chapters. The literature review follows, which delves deeper into the research concepts and terminologies employed in the study. The study area description, data acquisition, and data preprocessing steps are covered in the third chapter. The methodological workflow of the research is described in detail in the fourth chapter. The fifth chapter displays the results attained in this research, the sixth chapter discusses the research findings, and the final presents the conclusions drawn from the preceding chapter's discussions. This chapter also reflects how the research questions were addressed and gives recommendations of the thesis based on the findings.

## 2. LITERATURE REVIEW

### 2.1. Introduction

Most machine learning techniques need large training data sets to perform well in their tasks (Konyushkova, Raphael, & Fua, 2017). Methods like the use of a learning curve have been used to determine the required amount of training datasets for performing a machine learning task by looking at the performance of a model with respect to the amount of training data (Beleites, Neugebauer, Bocklitz, Krafft, & Popp, 2013). However, insufficient training data is currently identified as one of the challenges in RS data classification using machine learning methods (Li, Martin, & Estival, 2019). This challenge has been highlighted in different studies dedicated to hyperspectral image classification (Willis, 2004), multispectral image time series classification (Begüm Demir, Bovolo, & Bruzzone, 2013a), among others. The problem of insufficient training samples exists, especially when applying deep learning techniques to perform different remote sensing tasks like semantic segmentation or object detection (Ball, Anderson, & Chan, 2017; Matsuoka, Hayasaka, Fukushima, & Honda, 2007; Milan et al., 2018). This is because deep learning requires large datasets to capture target features efficiently. Previous literature suggests methods that could be used to address the problem of insufficient training data availability. In Ball et al. (2018), for example, Transfer Learning, Generative Adversarial Networks (GANs) and Unsupervised Learning were suggested as possible ways to address the issue of having small amounts of training data in deep learning. AL is another solution to address the issue of insufficient training data that state-of-the-art supervised classification methods are currently confronted with (Tuia et al., 2009).

### 2.2. Active Learning

AL is a field in machine learning which is sometimes referred to as query learning. In this field, the learner can choose the data from which it learns to perform classification. Its goal is to attain satisfactory classification results with few training datasets, leading to accurate information extraction at low annotation costs (Settles, 2009). AL reduces the training sample size required for training a classifier up to tenfold by identifying the most informative and diverse samples from the pool of unlabeled samples (Sugiyama & Nakajima, 2009). Informative samples are the samples for which a classifier has a hard time predicting their classes or labels. The sample diversity refers to how different the selected samples are from each other. AL has been applied in various image analysis tasks such as multispectral image segmentation (Mitra, Uma Shankar, & Pal, 2004), hyperspectral image classification (Rajan et al., 2008), object-based image classification (Ma, Fu, & Li, 2018), or regression (Pasolli, Melgani, Alajlan, & Bazi, 2012).

An Active Learner consists of five components, a classifier or set of classifiers  $C$ , trained using a labelled dataset  $L$ , a query  $Q$ , for selecting informative labels from a pool of unlabeled samples  $U$  and a supervisor,  $S$ , who assigns labels to the unlabeled samples (M. Li & Sethi, 2006). This makes a quintuple structure with components  $C$ ,  $L$ ,  $Q$ ,  $U$  and  $S$ . The initial labelled training samples are used to train the classifier, and a classification task is performed. The query  $Q$  uses a particular criterion to select informative unlabeled samples from the unlabeled sample set  $U$  while accommodating the classification output. Then, the supervisor labels these unlabeled samples, which are later added to the training dataset, and the classifier is retrained for classification. This process is iterative until a specified stopping criterion is achieved, for example, looking at the confidence level of the classifier (Vlachos, 2008).

### 2.3. Active Learning scenarios

Based on the AL literature survey, learners query samples in a variety of contexts, including stream-based, pool-based, and membership query synthesis (Settles, 2009). These scenarios are described in Table 2-1. Based on the fact that AL aims to be cost-effective, pool-based AL is considered for this research over stream-based AL since stream-based AL assumes label acquisition cost is free and thus selects each sample for querying to decide the informativeness instead of choosing only informative ones only. In the case of Membership Query Synthesis, labelling the artificially generated instances is a challenging task. Therefore, pool-based sampling is the preferred AL.

Table 2-1 Active Learning scenarios description

<b>Pool based AL</b>	<b>Stream-based AL</b>	<b>Membership Query Synthesis</b>
The cost of acquiring an unlabeled instance is considered.	Assumes that the cost of acquiring an unlabeled instance is free.	AL algorithm generates a new unlabeled instance within the input space and queries supervisor(labeler) for labelling.
Instances are drawn from the pool according to a user-defined informativeness measure. Only the informative ones are drawn from the pool.	Selects each unlabeled instance in the pool, and the Active Learner has to decide whether to ask the supervisor to label the current data sample or not based on a query strategy.	The supervisor labels the artificially generated instances, which is difficult because some textual instances are incomprehensible to human annotators.
Focuses on more than one data sample at a time.	It focuses on only one data sample at a time.	Focuses on more than one data sample at a time.
Assumes the presence of a large pool of unlabeled data	Assumes the presence of a large pool of unlabeled data	Generates artificial AL instances from the region of uncertainty of the classifier
The distribution of the samples is considered	The distribution of the samples is considered	Sample distribution is not considered since the Active Learner may request labels for any unlabeled instances, including the new instances it generates.

AL workflows depend on three main components: the model (learner) chosen, the uncertainty measure, and the query strategy used to select informative samples (He et al., 2014). The query selection criteria are the root of the AL algorithm since they decide which samples are informative based on various uncertainty measures and entirely depend on the classification output (Crawford, Tuia, & Yang, 2013). Most previous approaches account for the spectral domain when querying informative samples while ignoring the spatial one.

### 2.4. Spectral-domain heuristics and metrics in AL

Many AL heuristics that query samples based on their spectral characteristics in the feature space exist, for example, the uncertainty sampling-based and Committee-based heuristics (Adla, Group, Engineering, & Lafayette, 2014).

Under the uncertainty sampling query technique, the learner selects instances for which it is least certain on how to label (Breiman, 2001). Uncertainty sampling has several measures of uncertainty. One of them relies on models that use posterior probability to decide the class to which an instance belongs. For example, in a binary classification scenario, instances that give a probability close to 0.5 are informative. An alternative measure of uncertainty sampling is the Least Confidence (LC) measure, in which the learner chooses the instance for which it has the least confidence in its most likely label. This approach considers only the most probable label but ignores other label probabilities. To overcome this, the Margin Sampling approach as a measure of uncertainty selects the instance for which the difference between the first and second most probable labels is the smallest. To utilize all the possible label probabilities, the uncertainty measure used is entropy, sometimes known as Shannon entropy (Shannon, 1948). Application of the entropy formula is made to each instance, and the one with the largest value is queried. The higher the entropy, the more the uncertainty.

$$H(x) = - \sum_{y \in Y} (P(y|x) \log P(y|x)) \quad (1)$$

where  $P(y|x)$  is the a posteriori probability

$y \in Y = \{y_1, y_2, \dots, y_k\}$  denotes the output class

$H(x)$  is the uncertainty measurement function based on the entropy estimation of the classifier's posterior distribution.

Uncertainty sampling can also be used by non-probabilistic models such as Support Vector Machine, which assumes that instances near the decision boundaries are the most informative.

The Query By Committee framework can also be used as a query technique in AL as an alternative to uncertainty sampling. In this approach, a committee of learners are all trained with the labelled training dataset, and each learner is allowed to vote for which label an instance belongs to (X. Li, Zaïane, & Li, 2006). The most informative instance is the one for which the majority of the learners disagree. This method is said to be less computational than other active learning approaches and has the advantage of being independent of the classifier since several committee members (learners) are involved in selecting the most informative sample (Stumpf et al., 2014). For this reason, it was chosen for this research. In addition, the model used in this research is a Random Forest, which is an ensemble of classifiers and, therefore, getting votes from the ensemble classifiers will ease the detection of informative samples. There should be a construction of a committee of learners and a measure of disagreement between the committee members to perform a QBC algorithm of selecting informative samples. Query by bagging, which executes bootstrap aggregation by randomly sampling instances with replacement, thus conserving the sample distribution, and query by boosting, which performs sampling without replacement, thus changing the sample distribution, is used to construct committee members. The query by boosting and query by bagging techniques create randomness in samples' choice, making this approach more robust. The commonly used metrics as a measure of disagreement between the committee members are the vote entropy and the Kullback-Leibler (KL) divergence, also referred to as relative entropy. Entropy is a measure of sample uncertainty (Crawford et al., 2013). For vote entropy, each committee member votes on the labels of query candidates and the most informative query is the instance about which they most disagree. Instances with high vote entropy will be considered for labelling and later added to the training samples. KL divergence estimates the difference between two probability distributions. It measures how one probability distribution differs from the reference probability distribution: the observed and actual probability distribution. The larger the KL divergence between a committee member's posterior label distribution, the more informative the query is.

$$KL(P || Q) = - \sum_{x \in X} P(x) * \log(Q(x) / P(x)) \quad (2)$$

where  $KL(P || Q)$  is the L divergence between two distributions  $Q$  and  $P$   
 “ $||$ ” operator indicates “divergence” or divergence of  $P$  from  $Q$ , which are probability distributions.

KL divergence is calculated as the product of the negative sum of probability of each event in  $P$  and the log of the probability of the event in  $Q$  over the probability of the event in  $P$ . A drawback of this approach is its tendency to leave out some cases in which committee members disagree, making it uncertain (Settles, 2009). For this reason, vote entropy will be used as an uncertainty measure in our research.

#### 2.4.1. Weakness of Spectral-domain based AL

Most of the literature deals with AL heuristics that consider spectral data only. This approach solves the uncertainty of the samples in feature space. Still, the samples may be similar or relative to each other, leading to redundancy in selecting training samples (Crawford et al., 2013). Retraining a classifier based on a single most informative sample drawn from the unlabeled pool for each iteration is computationally costly and time-consuming. Some selected samples might be similar and do not bring important changes to the model (Stumpf et al., 2014). Methods on reducing the sample labelling time have been proposed in different literature (Stumpf et al., 2014). An example is batch mode AL, which considers sample uncertainty and diversity (Begm Demir, Persello, & Bruzzone, 2011). Uncertainty refers to how informative a sample is in the classification process, and sample diversity refers to how dissimilar the selected samples are with each other in the spectral domain. This research proposes a method that incorporates the spatial component in selecting informative samples with the hopes of lowering computing costs, saving time, and eliminating redundant samples while achieving comparable results to using many training samples.

#### 2.5. Incorporating the spatial component in AL

Just as the batch mode AL heuristics are included to avoid redundancy in the selection of spectrally similar samples in the feature space, it is possible to extend the same idea to the geographic space whereby pixels that are geographically close to the initial training samples are more likely to give similar information to the model (Crawford et al., 2013). This idea has been used in previous studies for determining the selection of spatially collocated samples (Munoz-Mari, Tuia, & Camps-Valls, 2012), coherent clusters (Volpi, Tuia, & Kanevski, 2012) and in Liu et al. (2008), who considered minimization of the cost function of travel time to minimize the travel distances between sample locations in geographic space using the Traveling Salesman Problem. Our research aims to develop an AL method that incorporates both the spatial and spectral domain for application in crop type mapping from satellite image time series.

#### 2.6. Spatial-domain heuristics and metrics in AL

Previous studies have successfully incorporated the spatial domain in AL (Patra et al., 2017; Crawford, Tuia, & Yang, 2013; Q. Lu, Ma, & Xia, 2017). The spatial component was used to select informative samples considering the geographic space. The metrics used for the spatial component included the distance minimization using the Travelling Sales Man approach to minimize label acquisition costs (A. Liu, Jun, & Ghosh, 2009), minimization of spatially collocated samples, characterized by a decrease in model efficiency if the selected pixels are similar to the chosen ones in the previous iteration (Volpi et al., 2012) and the level of clustering for each class (Lee & Crawford, 2005). This research incorporates the spatial component by making use of the spatial distribution of the samples. Spatial distribution analysis has been accounted for in crop mapping studies using the idea of spatial autocorrelation (Mathur, 2015). Spatial autocorrelation measures the similarity of objects within an area, the level of dependence between

variables, and dependence strength. The quantification of spatial autocorrelation has been categorized into local measures, global measures and variogram, which is a geostatistical approach.

### **2.6.1. Global and Local measures of spatial autocorrelation**

Global measures look at the level of clustering across the entire area of interest. They generally answer the question if there is a clustering pattern or not. Examples include Moran's I (Moran, 1950), Geary's C ratio (Geary, 1954), and joint count statistics. Moran's I index has been used in different applications, including plant population study (Mathur, 2015). It measures the overall spatial autocorrelation of an area and returns a single index to display the pattern. Moran's I ranges from -1 to 1, whereby -1 is perfect clustering of dissimilar values (perfect dispersion), 0 is no autocorrelation (perfect randomness), and +1 indicates perfect clustering of similar values (it's the opposite of dispersion). Moran's I and other Global measures of spatial autocorrelation yield only one single statistic that describes the spatial autocorrelation of the entire region without identifying where the similarity (or dissimilarity) occurs. This makes it unsuitable for this research since it is vital to understand where the dissimilarity or similarity in sample distribution occurs to select informative samples. Local measures such as Local Indicator of Spatial Association (LISA), local Moran's I, among others, zoom to a local extent to identify the clusters that are explained in the overall global pattern. They locate the location of clusters. These also give a single statistic for each locality. In this research, the aim is to look at the spatial autocorrelation to the point level, that is, the point-to-point relation in terms of the spatial distribution. A variogram fits this task since it models the spatial autocorrelation and distance dependence between observations (point locations).

### **2.6.2. Variogram**

A variogram is one of the geostatistical tools used to show spatial correlation by plotting the variance of point pairs with increasing distance between them (Curran, 1988). It is used to visualize and model spatial variation. The variogram increases as a function of the distance between point location pairs to portray that points close to each other may have similar values to those far from each other. In other words, it quantifies the spatial autocorrelation and shows how spatial variation changes as a function of the distance between point location pairs. The variogram has three main parameters: the sill, range, and nugget (Liu, Xie, & Xia, 2013). The nugget represents the non-spatial variability, and the sill is the total variability. The variance between point pairs increases until the sill and then becomes constant after the sill. The distance until which the semivariance levels off at the sill is called the range. The samples that appear above the range are not spatially correlated and hence can be selected for labelling. The variogram has been used in RS for various applications like textual image classification (Jakomulska & Clarke, 2001), structural and statistical analysis of textural images (Pham, 2016) and classification of SAR images (Tonye et al., 2011). In this research, the variogram is used for querying uncertain samples based on their degree of spatial correlation as a function of distance. Thus, the variogram model is used to incorporate the spatial component in the AL algorithm by assessing the lag or distance between training sample pairs. The uncertainty in the spatial domain will be considered by looking at the points whose in-between distance lies above the range since they are spatially uncorrelated and are dissimilar based on Tobler's law, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

## **2.7. Classification**

AL uses a supervised approach in classification since training samples are required to train a model initially. Different supervised classification methods exist, for example, Random Forest (RF), Maximum Likelihood Classifier (MLC) and Support Vector Machine (SVM). However, a good classifier should have the ability to handle feature nonlinearity, address the insufficiency of training samples due to a large number of features, also known as the "curse of dimensionality" or Hughes phenomenon (Chutia,

Bhattacharyya, Sarma, Kalita, & Sudhakar, 2016), handle imbalanced training samples and reduce the computational time (Gislason, Benediktsson, & Sveinsson, 2006). The MLC assumes that the data follows a normal distribution (parametric), which is not true with remote sensing data, while SVM and RF are non-parametric classifiers. RF has proved to be an efficient classifier for classification using a small number of training samples (Han, Jiang, Zhao, Wang, & Yin, 2018). SVM and RF are mostly used interchangeably in most RS tasks though RF has been observed to perform better when many input variables are available such as in the case of the hyperspectral image classification scenarios (Abdel-Rahman, Mutanga, Adam, & Ismail, 2014).

### **2.7.1. Random Forest classifier**

A RF is a classifier consisting of an ensemble of decision trees built using a subset of features and training samples (Breiman, 2001). It uses bagging and bootstrapping approaches to select subset training samples and features and generate trees that are used for prediction. The randomness in selecting training samples and features for splitting the decision tree nodes minimizes the correlation between the tree hence addressing the overfitting problem (Gislason et al., 2006). Almost two-thirds of the data are used for training, and the remaining third, also known as the Out Of Bag (OOB) samples, is used for internal cross-validation of the model (Breiman, 2001). RF is a computationally light algorithm and has only a few parameters to adjust, namely the number of trees (*n<sub>tree</sub>*) and the number of features to be considered for splitting the decision tree nodes (*m<sub>try</sub>* parameter). The assessment of the model's accuracy is done using independent validation samples, and the overall accuracy of the classified area is determined based on the confusion matrix. This research will use the RF classifier.

### 3. STUDY AREA DESCRIPTION, DATA ACQUISITION AND PREPROCESSING

#### 3.1. Study area description

The study area is Noord Beveland, situated in the Southwestern Netherlands. It is one of the Municipalities in the province of Zeeland. Different municipalities border Noord-Beveland: Schouwen-Duiveland in the north, Veere in the west, Middleburg in the southwest, Goes in the south, Kapelle to the southeast, and *Tholen* in the east. The municipality has a population of 7,308 inhabitants. It has an area of 121.6 km<sup>2</sup>, whereby 85.96 km<sup>2</sup> is land, and 35.62 km<sup>2</sup> is water. Various activities are practised in this area, for instance, recreation activities, but the main one is the agricultural activities with different crops grown such as wheat, potatoes, corn etc.(Koks, de, & Koome, 2012). Figure 3-1 shows the study area.

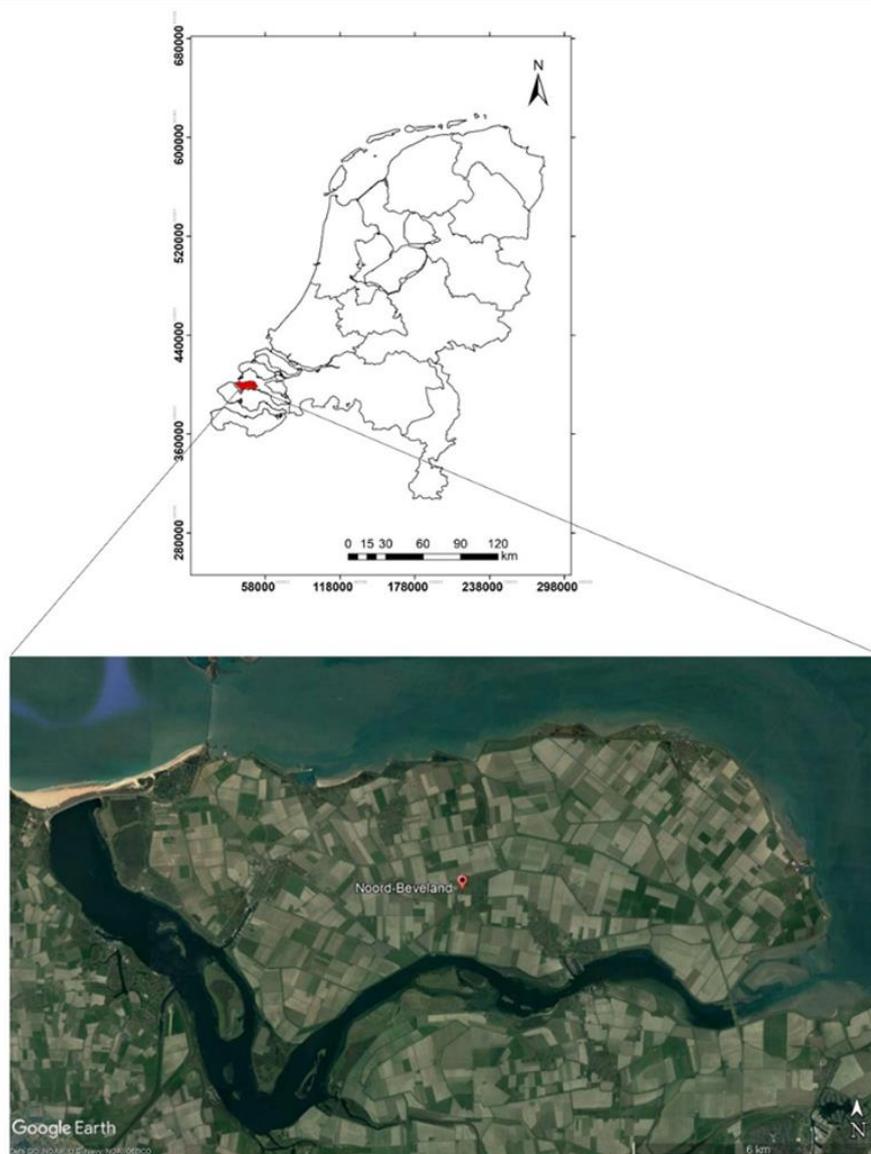


Figure 3-1 Location map of the study area (Source: Google Earth).

### 3.2. Data description

Table 3-1 describes the data used in our proposed research. Further details about the data and data management plan for this research are presented in Appendix A.

Table 3-1 Data description

Data	Temporal resolution	Spatial resolution	Type
Sentinel 2 Satellite image time series	5 days	10m, 20m, 60m	Raster
Crop parcel data (Ground Truth)	1 Year(2019)		Shapefile
The boundary of the study area			Shapefile

### 3.3. Ground truth data acquisition and preprocessing

The crop data consists of the location of agricultural parcels in the Netherlands with the cultivated crop linked to it. This data is available nationwide and is updated every year. It is a selection of information from the Basisregistratie Parcelen (BRP) of the Netherlands Enterprise Agency, and the parcel boundaries are based on the boundaries from the AAN file (Agricultural Area of the Netherlands) (Esri, 2018).

#### 3.3.1. Ground truth data acquisition

A total of 1584 crops cultivated in the study area were grouped into four different crop categories, namely:

- Arable land/farm land/ploughland (Bouwland)
- Grassland (Grasland)
- Nature area (Natuurterrein)
- Fallow land (Braakland)

Given the focus of our research on crop type mapping, only arable land was considered for the research, resulting in a total of 54 crop classes. The classes were filtered into seven crop classes based on the fact that some crop classes had very few plots (less than five plots). The resulting classes to be used for the research were cereals, maize, potatoes, alfalfa, beets, onions and orchard. Figure 3-2 shows the distribution of these crop classes in the study area, and Figure 3-3 shows the number of plots used per crop class for this research. The cereals class had the highest number of plots, followed by potatoes, beets, onions, orchard, maize, and the alfalfa class. Two additional classes for water and 'other' areas were added to distinguish crop areas from non-crop regions, making a total of nine classes.

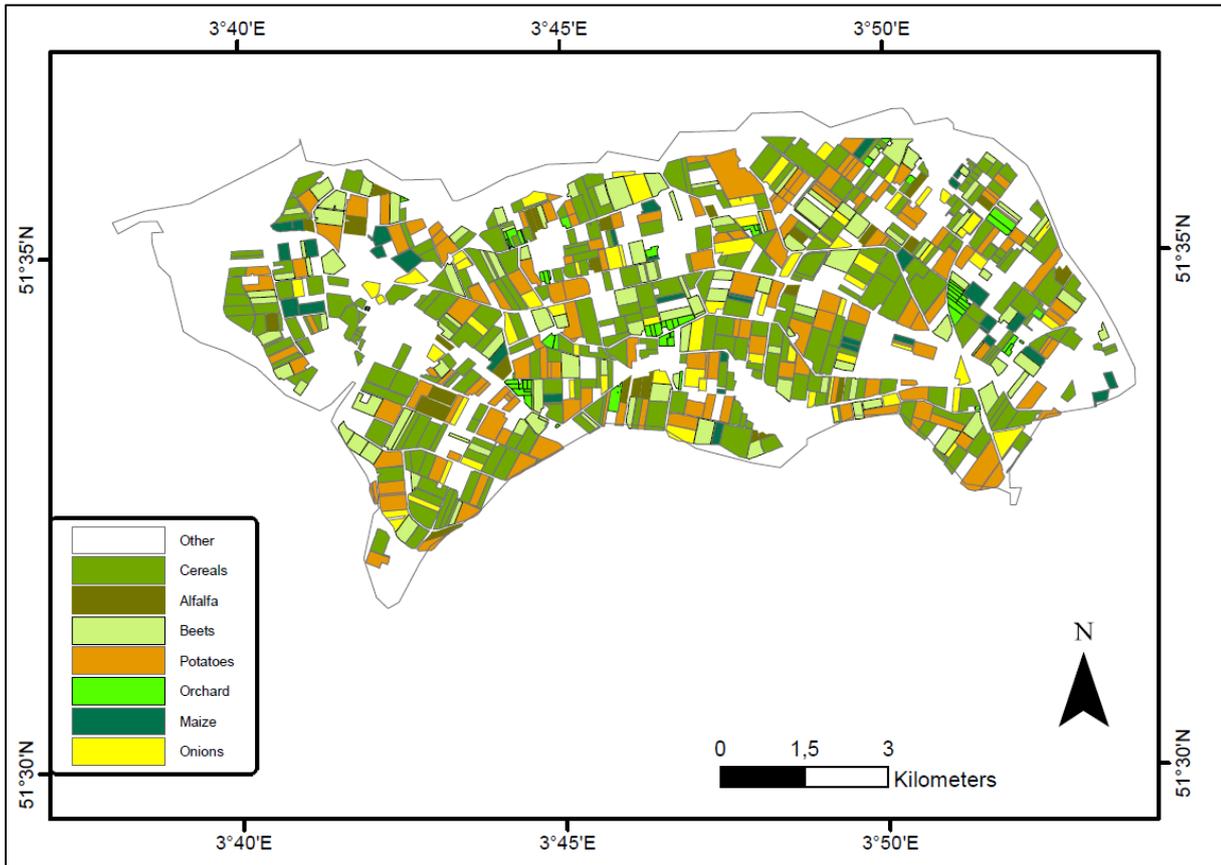


Figure 3-2 Agriculture plot distribution over the study area. The other areas are other types of landcover present in the study area

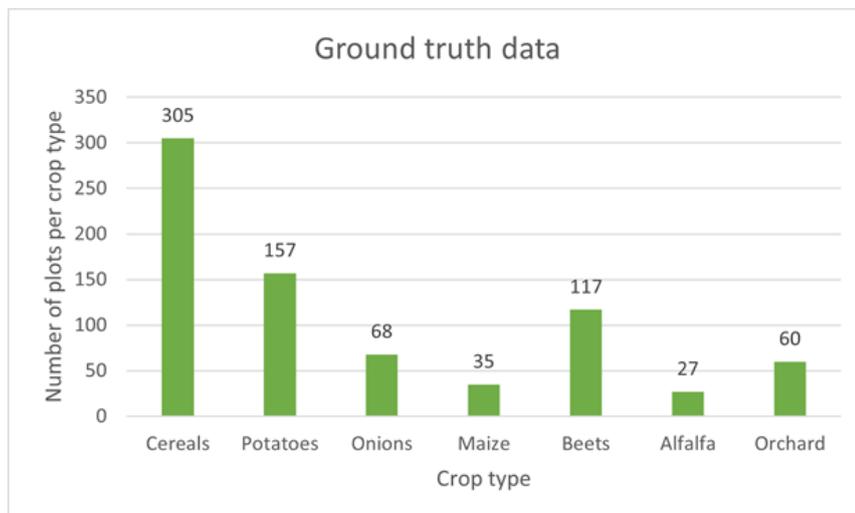


Figure 3-3 Number of plots per crop type class

### 3.3.2. Training sample preparation

A total of 630 samples were obtained for the nine target classes (70 samples per class). The Sampling Design tool in ArcMap was used to sample the points using the stratified sampling technique (Buja & Menza, 2013). Raster values (NDVI values) at point locations of the training samples were extracted using the **Extract Multi Values to Points** tool in ArcMap. The dataset was lastly divided into (70%) and

validation samples (30%). Therefore, we generated a total of 450 training samples (50 samples per class) and 180 validation samples (20 samples per class). The training and validation data were sampled from different plots ensuring the spatial dependence of training and validation samples. Figure 3-4 shows the training and validation sample distribution within the study area.

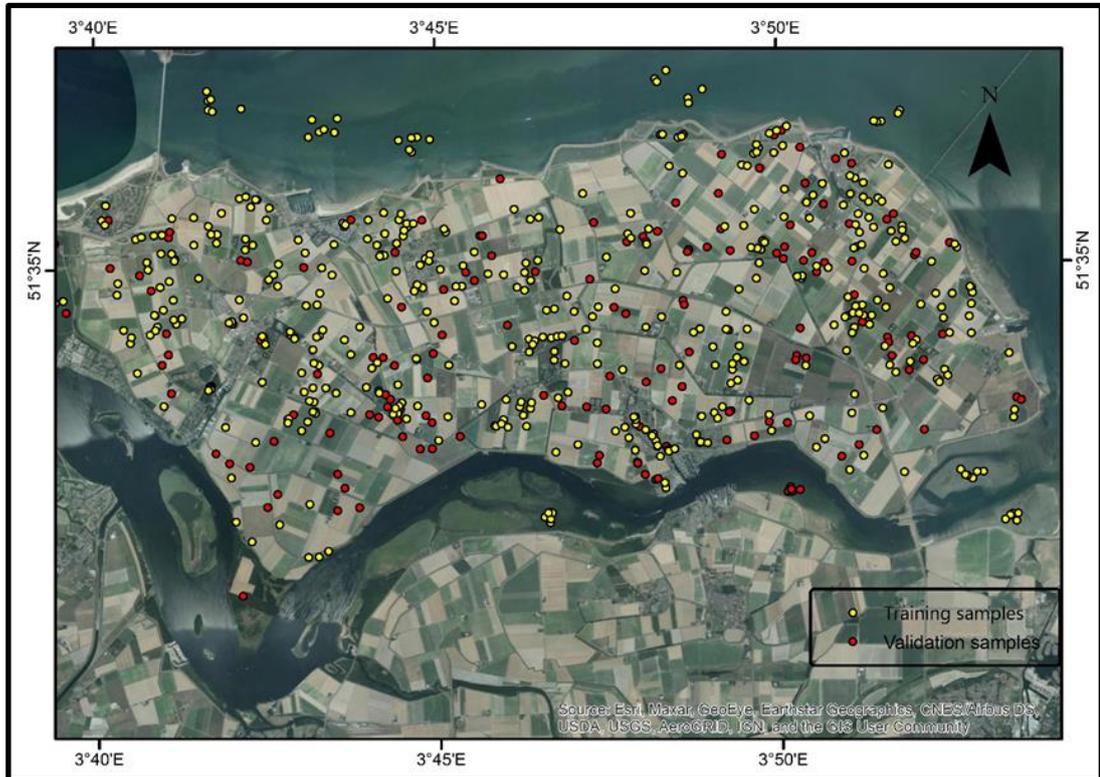


Figure 3-4 Training and validation sample locations for the nine target classes

### 3.4. Image acquisition and preprocessing

For this proposed research, a monthly time-series of Sentinel-2 mission (Bottom of Atmosphere product) images were acquired from the European Space Agency. The acquired images were for the year 2019, and their acquisition was done using Google Earth Engine. They were a total of 12 images. The images had 13 bands with 10m, 20m and 60m resolution, respectively. Table 3-2 shows the band properties of the images acquired. The images were acquired by first loading the image collection and then filtering the images using different criteria such as date range, area of interest and cloud cover percentages.

Table 3-2 Band properties of acquired Sentinel images

Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Coastal aerosol
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Red-Edge
B6	20m	740 nm	Red-Edge
B7	20m	783 nm	Red-Edge
B8	10m	842 nm	NIR
B8a	20m	865 nm	NIR narrow
B9	60m	940 nm	Water Vapour
B10	20 m	1375 nm	Cirrus
B11	20 m	1610 nm	Short Wave Infrared (SWIR 1)
B12	20 m	2190 nm	Short Wave Infrared (SWIR 2)

### 3.4.1. Cloud masking

Since the target was to get at least one image per month, the cloud percentage value was set in such a way that it was not too low to avoid excluding too many images or too high to get cloudy images. The Earth Engine App was used to identify the cloud percentages for the images monthly. Figure 3-5 shows a cross-section of the Google Earth engine app that helped in viewing the cloud cover percentages for all images in each month. The marked area (purple point) is the study area.

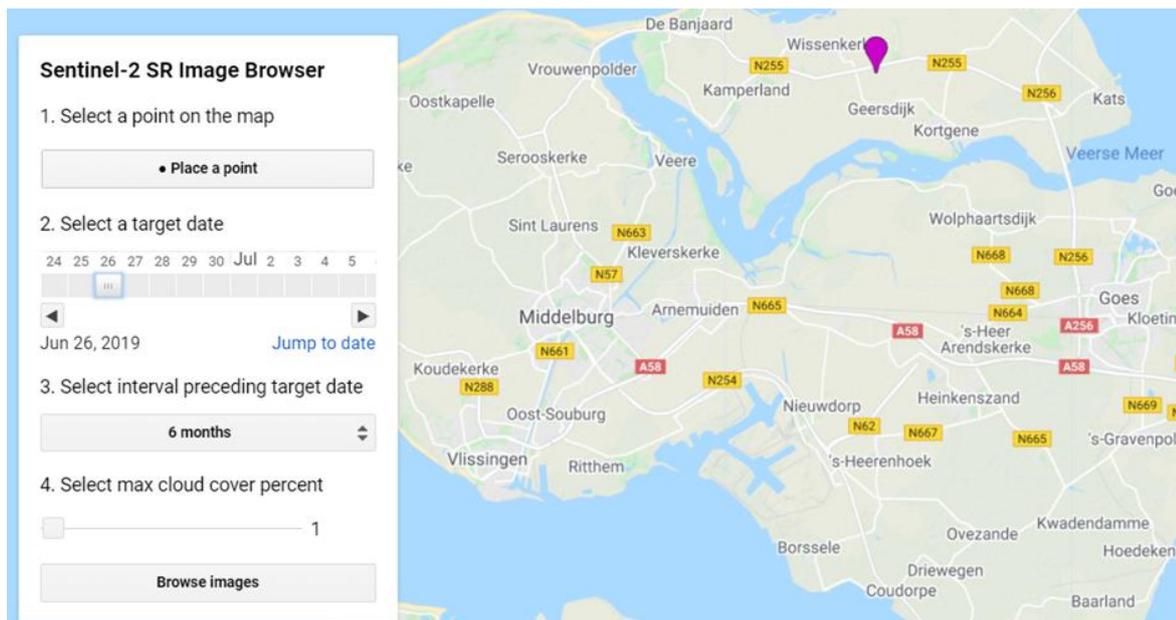


Figure 3-5 Illustration of the Earth Engine App used to get an overview of the cloud percentage values for all images in each month.

Images with the least cloud percentage per month were selected monthly. Figure 3-6 shows the cloud percentages for each image acquired with their respective dates of acquisition. Images of May, October and November had high cloud cover percentages, while February, July, August, and September had the least cloud cover.

Three QA bands were present in each image, where QA60 is a bitmask band containing cloud mask information. QA stands for 'Quality Assessment' while 60 stands for the spatial resolution in meters. Bit 10 stands for opaque clouds. When the value of Bit 10 is 0, there are no opaque clouds present, and when the value is 1, there are opaque clouds present. Bit 11 stands for cirrus clouds. If the value is 0, there are no cirrus clouds, but if the value is 1, then cirrus clouds are present. Both bit values were set to 0 to indicate clear conditions. The *maskS2clouds* function in Google Earth Engine was used for this task.

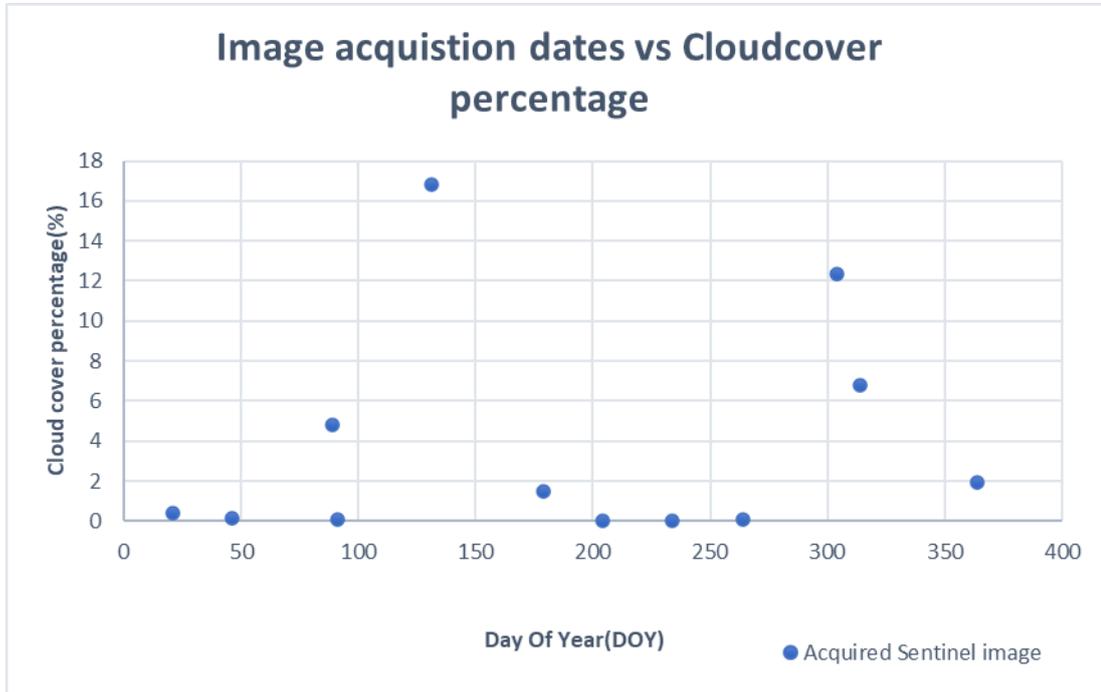


Figure 3-6 Image acquisition dates with respect to cloud cover percentage.

The increase in spatial and temporal resolution in satellite data calls for analysis that considers the time component (Belgiu & Csillik, 2018). Some researches in crop mapping demonstrate that time-series images perform better than single date images (Xiong et al., 2017). The assessment of vegetation productivity, health, and monitoring uses various vegetation indices in RS (Nirbhay Bhuyar, 2020). Monitoring also involves the time component, looking at different stages of the crops as they progress in their growth. Some of the vegetation indices used are the Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), among others. To get an overview of the crop stages and track the changes, it is essential to have images acquired at different times (C. Sun, Y. Bian, T. Zhou, 2019). This research uses NDVI (Rouse, Jr., Haas, Schell, & Deering, 1973) as an index to track crop changes with time. NDVI is calculated using the Red and Near Infrared bands. Equation 3 illustrates the calculation of the NDVI index.

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

### 3.4.2. NDVI calculation

NDVI values range from -1.0 to +1.0, whereby very low NDVI values indicate bare soil (0.1 or less), moderate NDVI values (0.2 to 0.5) express sparse vegetation or season of emergence in vegetation growth and high values (0.6 and above) indicate crops being at the peak of their growth stages (USGS, n.d.). In this research, the NDVI indices were calculated for each image. Figure 3-7 shows NDVI temporal profiles for

crop classes used in this research. In the end, an NDVI image composite was attained by stacking all the NDVI images calculated for each month. NDVI image products were exported from GEE for further processing in R software.

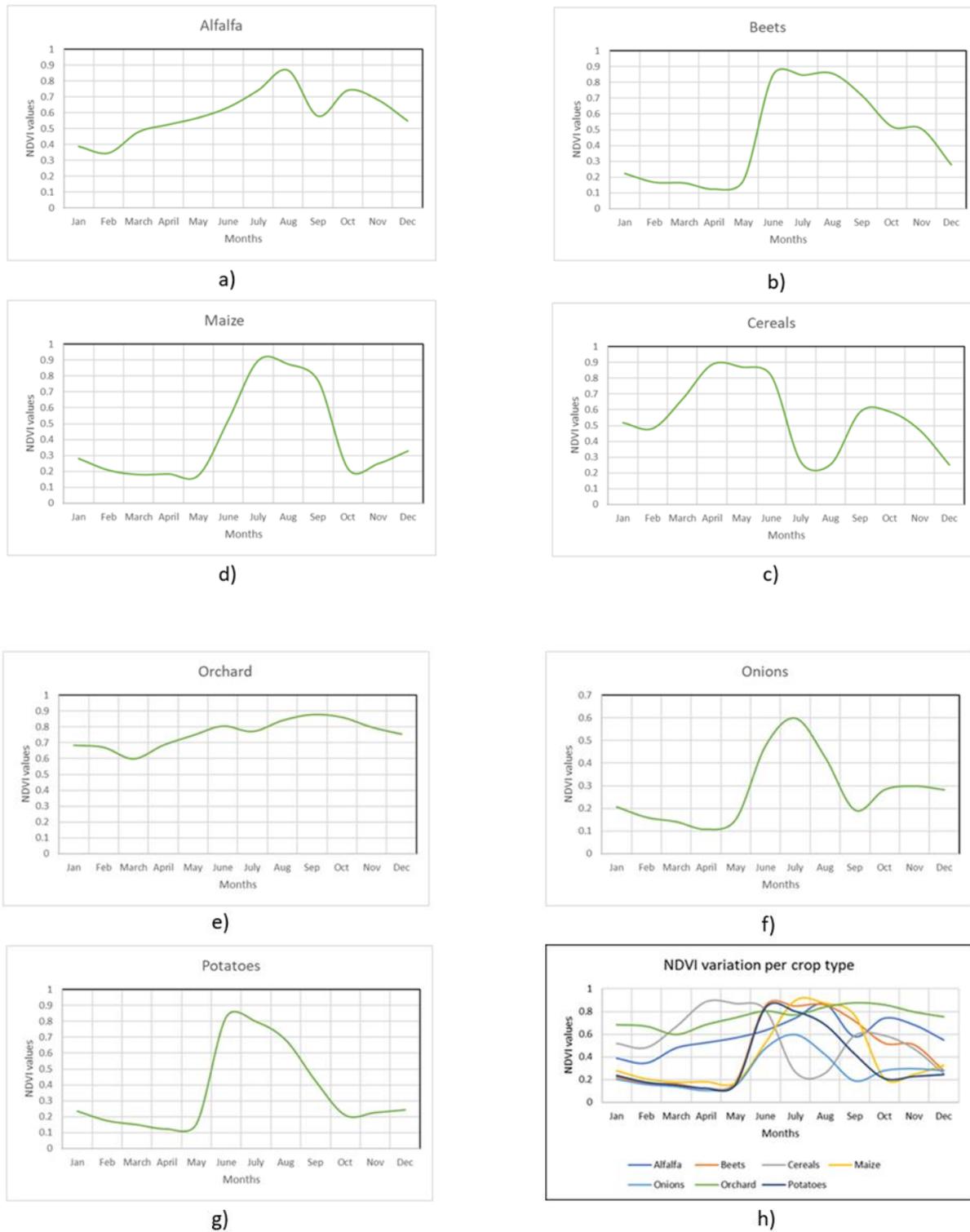


Figure 3-7 Temporal profiles for a) alfalfa, b) beets, c) cereals, d) maize, e) orchard, f) onions, g) potatoes, h) a combination of all target classes

The temporal profiles generated using the NDVI values of the target classes reflect the crop growth stages. The NDVI peaks indicate the stage at which the crops are at the peak of growth. The alfalfa and cereals target classes had two peaks. This means these crop had two cropping cycles. The beets class were at their peak growth from June to August, maize crops in July and August, onions in July and potatoes in June. The low NDVI values in the graphs represent a stage when the crops are sown or harvested. The land appears almost bare in these stages hence the low NDVI values. An example is the potatoes class, whereby the crops start growing(emergence stage) in April and May, and in October and November, the potatoes are harvested.

### **3.4.3. Image co-registration**

Image co-registration is required to ensure the spatial alignment of the images. A base or master image was selected, and other images were aligned to it. The master image selection was made by taking into consideration the cloud cover percentage. The image with the lowest cloud cover percentage was considered as the master image. Based on the available images, the image acquired on the 23<sup>rd</sup> of July 2019 was the master image, with a cloud cover percentage of 0.04. Common features such as road corners and plot corners were identified in the master image, and they were checked for alignment. This was done in ArcMap by overlaying the images on each other. In our case, the images were all aligned. After all the above steps, the twelve NDVI images were stacked to form an NDVI composite image.

## 4. METHODOLOGY

This chapter describes how AL has been used to generate samples required to perform crop mapping for time-series satellite images. The following workflow provides a general overview of the steps of this research in figure 4-1.

### 4.1. Workflow of the thesis

The initial steps of the workflow were the acquisition and preprocessing of ground truth crop data and the satellite image time series required for this research. Section 3.4 in the previous chapter explains the image acquisition and preprocessing steps carried out in detail, as shown in figure 4-1 (highlighted in red). Section 3.3 in the previous chapter also explains the ground truth data acquisition and preprocessing, as seen in figure 4-1 (highlighted in green). Next was the use of AL in the selection of informative training samples out of all the available training samples. This was done using an AL algorithm that excludes the spatial component (considers spectral domain only) and later on developing an algorithm that incorporates the spatial component in selecting informative samples. Classification was then performed using all training samples and then performed using informative samples generated from both AL algorithms independently, yielding crop type maps. Analysis of the results was done by comparing the outputs of the two AL algorithms to the outputs generated from using the entire training sample dataset, and later on, a comparison between the two AL algorithms. This was done through accuracy assessment and performance evaluation of the AL algorithms.

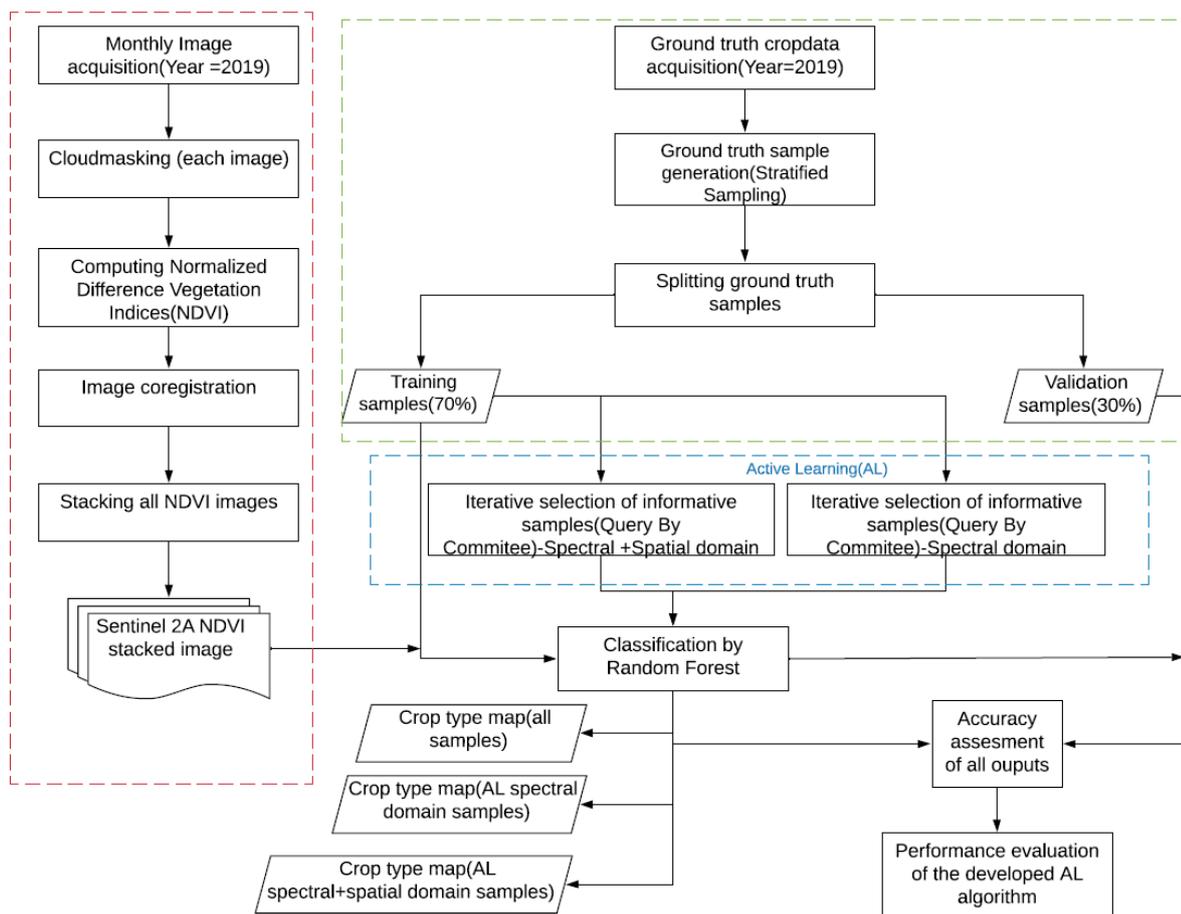


Figure 4-1 Flow chart of the proposed methodology of the research.

#### 4.2. Active learning components

The AL technique has a quintuple structure with five elements which are a classifier or set of classifiers  $C$ , a labelled dataset  $L$ , a query  $Q$ , for selecting informative labels from a pool of unlabeled samples  $U$  and a supervisor,  $S$ , who assigns labels to the unlabeled samples (M. Li & Sethi, 2006). This AL structure is displayed in figure 4-2. AL workflows are dependent on three main components: the model (learner) chosen, the uncertainty measure, and the query strategy used to select informative samples (He et al., 2014).

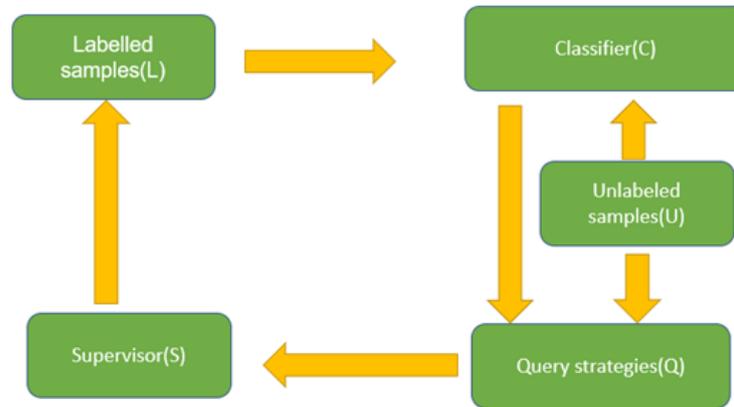


Figure 4-2 Active Learning components

In the AL part, only the seven crop classes were used. The remaining two classes (water and 'other') were used in the final steps of mapping to distinguish the crop classes from other classes outside the scope of this research. For the seven crop classes, there were 350 total samples (50 samples per class). The distribution of the target classes in the feature space is depicted in figure 4-3. Principal Component Analysis (PCA) transformation was applied to the dataset for visualization purposes.

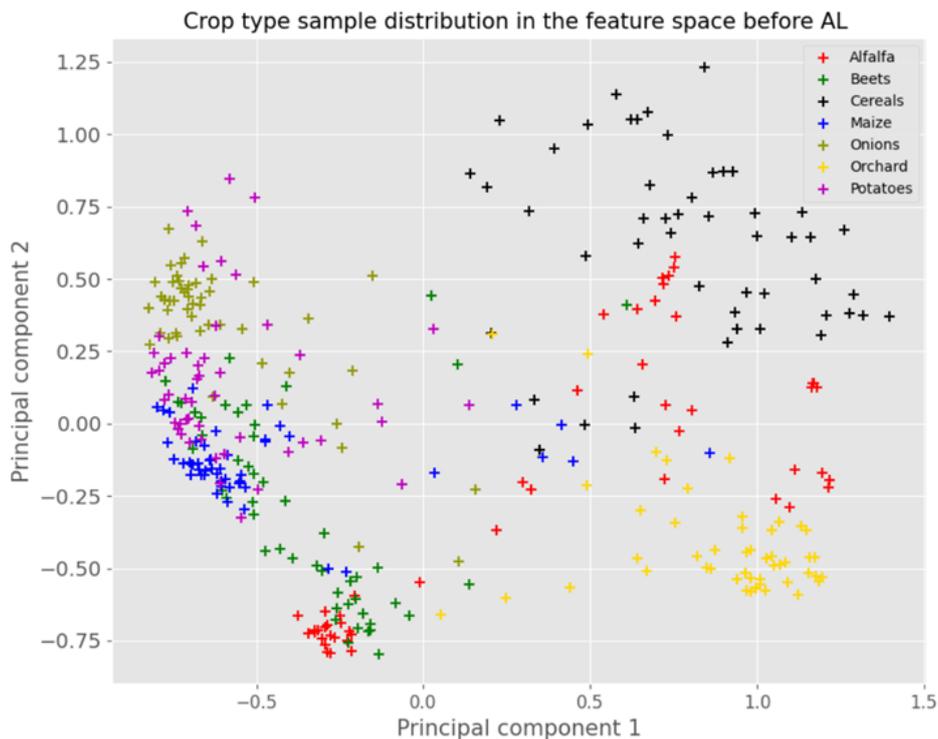


Figure 4-3 Distribution of target classes in feature space before applying Active Learning.

It is evident in figure 4-3 that there is a high inter-class overlap in the feature space between maize, potatoes, beets, onions, and alfalfa classes. This could be due to the existence of several crops on the same field (mixed cropping). Classification in such scenarios is a challenge and may affect the information derived from the samples since one class may be wrongly assigned to the other class due to close resemblance (Gebbinck, 1998). Furthermore, some classes in a specific dataset may have significant spectral overlap, which means that these classes cannot be discriminated by image classification. The following sections describe how AL was applied to meet the research objectives considering both the spectral and spatial domain.

### **4.3. Active Learning considering the spectral domain only**

In this approach, the informative samples were queried based on their spectral characteristics in the feature space, as explained in the steps below:

#### **4.3.1. Pool generation**

The AL scenario used for this research was pool-based AL described in Table:2-1 in the second chapter. A pool refers to the set of unlabeled samples from which the Active Learner draws informative samples. To create a pool, the 350 training samples were divided into a training set  $L$  of 40 samples and an unlabeled set  $U$  with 310 samples that form the pool. Instances were to be drawn from the pool according to a user-defined informativeness measure. The labels for the 40 initial training samples were known, while for the 310 samples forming the pool, the labels were assumed to be unknown.

#### **4.3.2. Initializing a committee(Model)**

Referring to section 2.4, which describes the spectral domain heuristics used for querying informative samples from the pool, the query strategy to be used was the Query By Committee (QBC) strategy. This required generation of a committee of AL members. Different studies state that there is no specific number of committee members that should generally be used. A small number of committee members have worked well in various studies (Seung, Opper, & Sompolinsky, 1992; Settles, 2009). Based on the previous studies, the committee used for the selection of informative samples had two members. The committee members consisted of the estimator, which is the RF classifier, and the uncertainty measure, the vote entropy discussed in chapter two. These committee members(model) were trained using the labeled dataset and later used to predict the labels of the samples in the pool.

#### **4.3.3. Iterative selection of informative samples, labelling and prediction**

Using the QBC strategy, informative samples were queried from the pool while assessing the prediction accuracy of the committee with an increment of the training samples used to train the model. The samples were queried based on their NDVI values, and the informativeness was determined using vote entropy as a metric. One of the stopping criteria in AL would be to measure the performance of the trained classifier on an annotated dataset and stop when the performance of the model increases at a non-satisfactory rate or stops improving. After 129 queries, the classification accuracy increased at a very slow rate and later stopped improving. This happened every time the classification was run. These samples were exported and saved to be used for crop type mapping. Using the 'water' and 'others' class samples and the samples generated from AL based on the spectral component only, classification was performed on the composite NDVI image to get a crop type map. RF classifier was used for this classification purpose. The code for the above processes is displayed in GitHub(<https://github.com/beatrice327/Active-Learning-thesis>).

#### 4.4. Incorporating the spatial component in AL

In this approach, the informative samples were queried based on their spectral and spatial characteristics in the feature space as explained in the steps below: Several spatial domain heuristics were surveyed as described in section 2.6 and in this research, the variogram was chosen to be used in the selection of informative samples in the spatial domain. In the classic AL workflow (Figure 4-2), the query criteria used are spectral based. This step involves the spatial component in the selection of informative samples. Informative samples, in this case, are the samples from the pool that fulfil the spectral and spatial criteria when queried. The learner iterates through all the samples in the pool to select informative samples. Figure 4-4 shows this AL procedure.

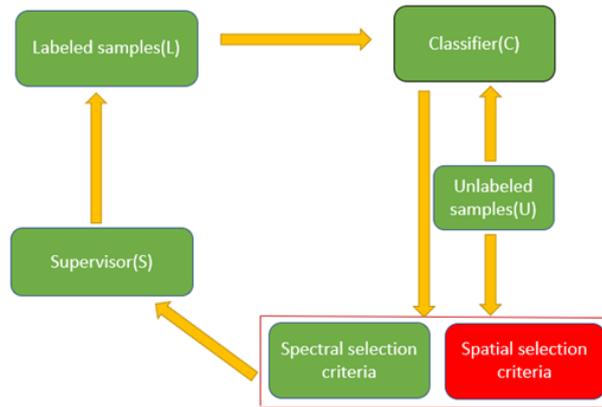


Figure 4-4 AL workflow involving the spatial component in querying informative samples

The spatial distribution of the 350 crop samples before AL is as shown in Figure 4-5. The red-highlighted regions show that the same crop class is sampled many times at close locations. Having many samples of a particular class taken many times at the same location or near locations surrounded by the same class is redundant and time-consuming. It is more efficient to obtain more samples of the same class at a reasonable distance that is not near the already sampled area.

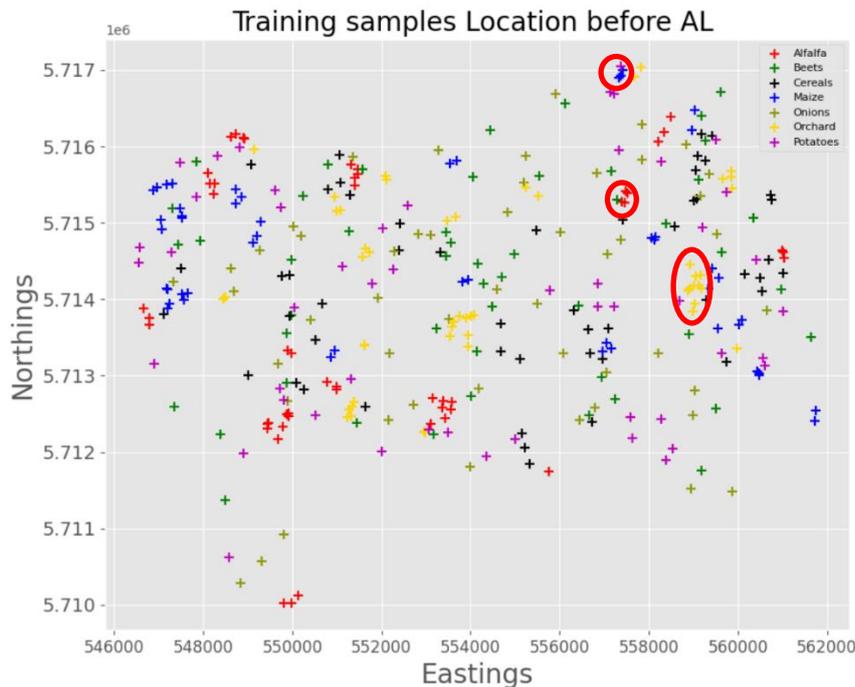


Figure 4-5 Crop type spatial distribution in geographical space before AL

Introducing the spatial component in the selection of informative samples can minimize this redundancy. The variogram is used for this purpose since it characterizes the spatial structure of an area. As explained in section 2.6.2, the variogram has a parameter called range which is the distance at which a variogram achieves a plateau (levels off). Sample pairs whose in-between distance lies above the range are said to be spatially uncorrelated. This means these samples are informative spatially since they are far from each other and are likely to differ in properties with reference to Tobler's law of Geography which states that "Everything is related to everything else, but near things are more related than distant things"(Tobler, 1970). Choosing points below the range is not recommended since the samples are correlated, and they may likely share similar characteristics. Therefore, the spatial component was incorporated in AL by taking points above the range of a variogram. The learner queries an informative sample spectrally from the pool, then checks the Euclidean distance between this chosen point and all points in the training set. If any of the Euclidean distances between the chosen point and the labelled points in the training set is below the range, the point does not qualify to be informative, but if it is above the range, the point is considered informative. By combining spatial and spectral dimensions, the AL strategy selected informative samples representative of all classes spectrally and at different spatial locations. The following steps explain the stages in this task.

#### 4.4.1. Variogram estimation

Variograms were estimated monthly for all NDVI images considering values for the 350 training samples. Variogram estimation was done monthly to highlight the differences in crop stages. Parameter values like the number of point pairs per lag distance, cutoff, among others, were considered. By default, the *gstat* package in R calculates the sample variogram of 0.33 of the maximum possible lag and a default of 15 bins. At least 30 pairs of points are needed for a reliable estimate of a sample variogram for each lag distance (Esri, 2016). An example is shown in Table 4-1, in which all bins or lags have a number of point pairs greater than 30. Parameters *np* show the number of pairs of points used to estimate the sample variogram for a given distance (lag), *dist* stands for the lag (distance), which is the mean distance between the point pairs in a bin, *gamma* is the sample variogram value at that lag which is attained by calculating the mean of the variogram cloud points in the bin.

Table 4-1 Estimated variogram parameters, *np* (the number of point pairs per lag), *dist* (lag distance) and gamma (mean of the variogram cloud points in the bin)

np	dist(m)	gamma
569	232.86	0.09
1138	570.18	0.12
1578	940.47	0.11
1950	1305.25	0.12
2390	1674.59	0.12
2798	2045.96	0.12
2919	2412.35	0.12
2853	2789.36	0.12
3073	3156.93	0.12
2972	3534.45	0.11
2992	3898.13	0.13
2470	4269.12	0.12
2233	4646.14	0.12
2218	5019.31	0.12
2105	5391.97	0.12

#### 4.4.2. Fitting a variogram model to the data

Different variogram models are normally used in geostatistics to model spatial variability; however, the choice of the variogram model to use depends on the parameters such as the sill, nugget, and range. These parameters were used as a metric to select informative samples in the new AL algorithm that considers the spatial component. These parameters are illustrated in figure 4-6. The distance at which the model levels out is called the range. It is the distance up until which the regionalized variable (NDVI in our case) is autocorrelated. At greater distances than the range, the regionalized variable is uncorrelated. The value at which the semivariogram model attains the range is called the sill (also known as the total variability in the data). The nugget refers to the non-spatial variability, which are measurement errors or spatial sources of variation at distances that are less than the sampling interval or both. It is some value greater than 0, and it intercepts at the y-axis of the variogram. For example, if the semivariogram model intercepts the y-axis at 1, then the nugget is 1. Partial sill refers to the difference between the sill and the nugget values.

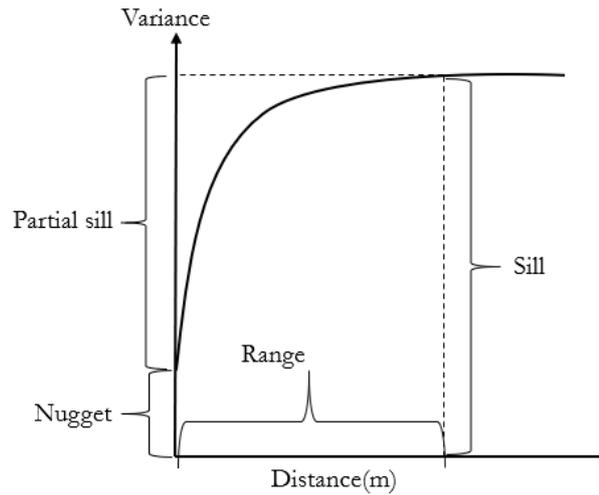


Figure 4-6 Variogram model parameters, range, sill, nugget, and partial sill.

The fitted models were the spherical, exponential, and Gaussian models since they are the most used models in variogram modelling (Mohebzadeh, 2018). The optimal model was chosen by considering the Sum of Square Error (SSErr) of the estimates of the variograms fitted from all models. The SSErr is considered since it gives the "goodness of fit". The higher the SSErr, the less the goodness of fit of the variogram. Different ranges were set while examining the SSErr for each model type, as shown in Table 4-2. After several runs, the spherical model gave the least SSErr for most of the variograms modelled for each month. Figures 4-7, 4-8, and 4-9 show the model types as fitted on the variogram.

Table 4-2 Sum of Square Error values for monthly variograms considering all variogram model types (Exponential, Spherical and Gaussian)

Month	SSErr(Exponential model)	SSErr(Spherical model)	SSErr(Gaussian model)
January	0.238	0.22	0.22
February	0.168	0.157	0.155
March	0.379	0.374	0.528
April	0.976	0.916	0.914
May	0.488	0.465	0.619
June	0.251	0.237	0.244
July	0.321	0.32	0.321
August	0.199	0.198	0.198
September	0.577	0.589	0.643
October	0.6	0.579	0.598
November	0.434	0.408	0.513
December	0.389	0.379	0.392

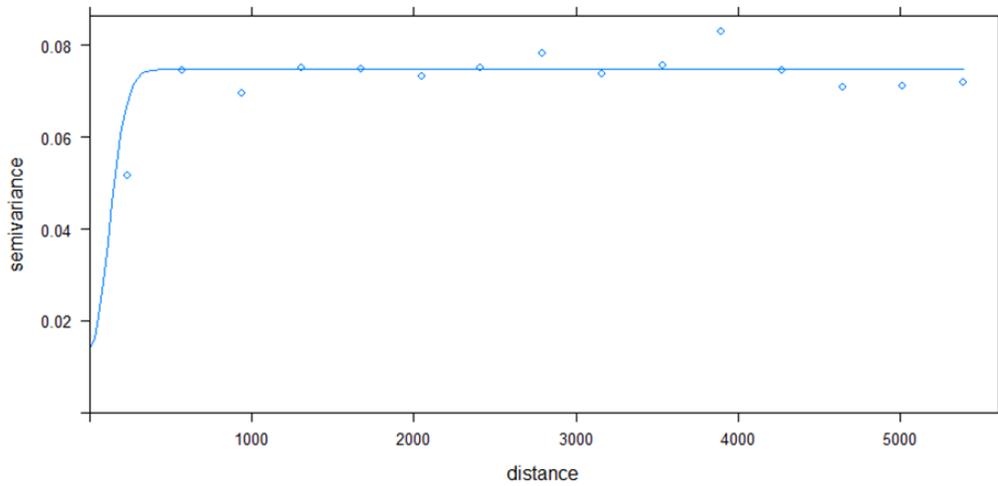


Figure 4-7 Gaussian model fitted from the data for one month

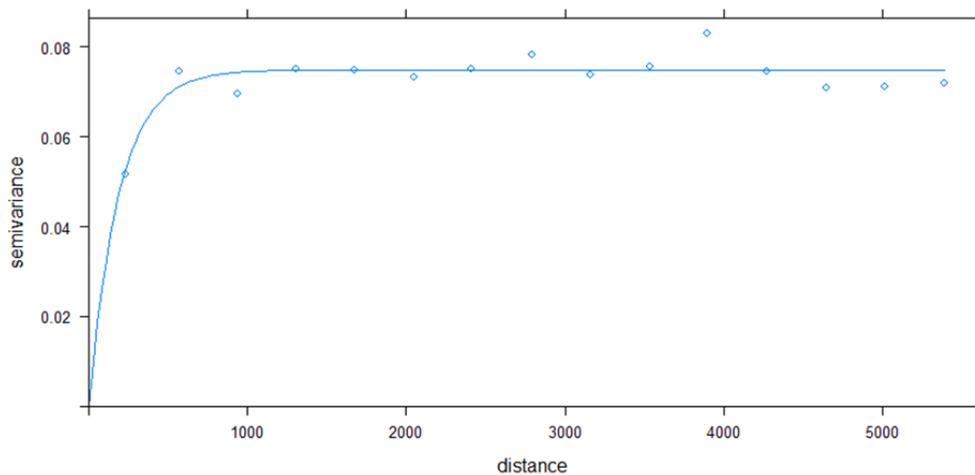


Figure 4-8 Exponential model fitted from the data for one month

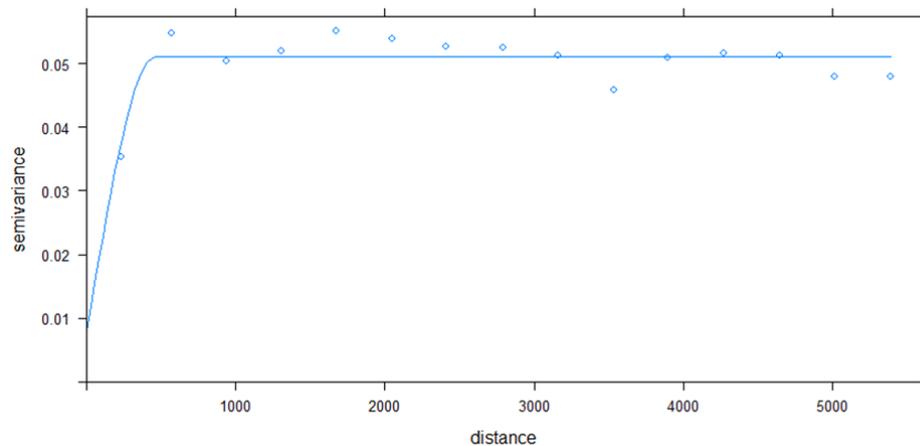


Figure 4-9 Spherical model fitted from the data for one month

In the above scenarios, Figures 4-7,4-8 and 4-9, the curve fit for the spherical and exponential models appear better than the Gaussian model since they are close to the data. On the other hand, the Exponential and Gaussian variograms approach the sill asymptotically, and so they do not have strict ranges. The ranges for these models are normally taken as the lag distances at which the variance reaches 95% of the total sill (Blöschl, 2002). This means these two models estimate the range, not the exact range like the spherical model. For this reason, too, the spherical variogram was chosen.

#### 4.4.3. Selection of range value

Different ranges were applied in modelling the variograms using the spherical model for all the 12 NDVI images while considering the 350 samples. The nugget value(variance) was used as a determinant for each month to choose the ranges, whereby variogram models with low nugget values were chosen. The nugget variance is a value for variance when the lag distance ( $h$ ) = 0 on the variogram. It explains the randomness of the data (Diggle, Tawn, & Moyeed, 1998). The higher the nugget, the higher the uncertainty or randomness of the data. The nugget accounts for different error sources, such as measurement errors at distances that are smaller than the sampling interval. Measurement errors can be errors from instruments, observation errors, among others. Therefore, the nugget value is an estimate of the sum of all residual variance contributions from all measurement steps such as sampling, observation, and analysis, among others(Engström & Esbensen, 2018).In selecting the best ranges, the variograms chosen for each month were the ones with the least nugget values. The variograms of the 12 months are attached in Appendix B of this research.

As mentioned earlier, the variograms were estimated monthly to account for the crop growth stages. Figure 3.7 illustrates the NDVI temporal files for the seven crop classes. It is observed that the crops displayed different patterns, which portray different stages of crop growth. The months to be considered for the range choice were the ones for which the crops appear separable. From graph 3.7 (h), all months showed separability for some crops classes, but not all. It was therefore challenging to select a range based on crop separability per month. Table 4-3 shows that the smallest range is 417.6m, and the largest is 905.4m. To choose a range to be considered for selecting informative samples spatially, the smallest range was considered. This is because it is the shortest distance above which spatial dependency of all the 350 samples is no longer present considering all bands(12 NDVI images in this case). The larger range values were not chosen because taking a larger value means spatially uncorrelated points will also be considered to be spatially correlated. In addition, the in-between distance between the chosen points will be large, hence having few points and a chance of leaving out informative samples. For this reason, the chosen range was 417.6m. Therefore, all samples whose in-between distance is above this range will be considered informative spatially for the AL part.

Table 4-3 Variation of the ranges with their respective nugget values for the variograms fitted for each month

Month	Range(m)	Nugget
January	462.3	0.008
February	477.4	0.005
March	611.2	0.029
April	513	0.029
May	417.6	0.025
June	587.3	0.008
July	905.4	0.043
August	721.1	0.050
September	804.9	0.043
October	733.6	0.033
November	735.3	0.028
December	794	0.018

#### 4.4.4. Iterative selection of informative samples, labelling and prediction considering both the spectral and spatial domain

The AL algorithm was run to select informative samples that fulfil both the spectral and spatial criteria. After iterating through the pool of 310 samples (excluding 40 samples that were initially used to train the model), a total of 57 samples qualified to fulfil both the spatial and spectral criteria. For every iteration, the learner queried an informative sample spectrally from the pool  $U$ , then checked the Euclidean distance between this queried sample and all points in the training set  $L$ . If any of the Euclidean distances between the chosen point and the labelled points in the training set was below the range, the point did not qualify to be informative, but if it was above the range (417.6 m), the point was considered informative. These 57 samples were exported together with the initial 40 samples used to train the model and stored to be used for crop type mapping. Using a RF classifier, classification was performed on the composite NDVI image. The code for the above processes is displayed in GitHub (<https://github.com/beatrice327/Active-Learning-thesis>).

#### 4.4.5. Assessment of the classification outputs and evaluation of the AL algorithms

As elaborated in section 4.1, a RF classification was run on the NDVI image composite before the AL task was done. The outputs from this classification were used as a reference to evaluate the performance of the AL algorithm that excludes the spatial component and the one that includes the spatial component. Comparisons were made using the confusion matrices for all approaches. We considered the Kappa coefficient, overall accuracy, user's accuracy and producer's accuracy for assessing the classification results. The overall accuracy describes the proportion of the reference samples that were classified correctly. Producer's accuracy, also known as sensitivity, is calculated by the number of correct positive predictions divided by the total number of positives. It depicts how often real features in the area of interest are correctly shown on the classification map. User's accuracy, sometimes referred to as precision, shows how often the classes predicted on the map are really present on the ground. It is calculated by taking the total number of correct classifications for a particular class and dividing it by the row total. The Kappa coefficient uses the overall model accuracy and per class accuracies considering both the predictive and reference aspects to correct for agreement between the classes (Cohen, 1960). It compares an observed accuracy with an expected accuracy obtained from a random choice of samples. It ranges from 0 to 1. Values that are close to 1 reflect the agreement between the observed and expected accuracy values.

## 5. RESULTS

This chapter presents the results of the workflow described in the previous chapter. Section 5.1 presents the output of RF classification of all samples before AL. The RF outputs of samples generated by the AL that only considers the spectral domain are shown in Section 5.2. The findings of RF classification from AL samples generated considering the spectral and spatial domains are presented in Section 5.3. Section 5.4 describes the performance of the developed AL method and assesses its effectiveness in crop type mapping for satellite image time series.

### 5.1. Classification using the entire dataset of training samples

After classification using all 450 training samples, the prediction accuracy was assessed using 180 validation samples. Figure 5-1 illustrates the confusion matrix and other statistical metrics of the classification considering all training samples. Figure 5-2 shows the crop type map generated from this classification task.

Confusion Matrix and Statistics  
Kappa = 0.8188

	Alfalfa	Beets	Cereals	Maize	Onions	Orchard	Potatoes	Water	Other	Total	UA(%)
Alfalfa	19	0	0	0	0	0	0	0	0	19	100
Beets	0	15	0	3	0	0	1	0	1	20	75
Cereals	0	0	17	0	1	0	1	0	0	19	89.47
Maize	0	0	0	14	0	0	1	0	0	15	93.33
Onions	0	0	0	1	15	0	0	0	0	16	93.75
Orchard	0	1	0	2	0	18	0	0	3	24	75
Potatoes	0	4	1	0	0	0	17	0	0	22	77.27
Water	0	0	0	0	0	0	0	20	0	20	100
Other	1	0	2	0	4	2	0	0	16	25	64
Total	20	20	20	20	20	20	20	20	20	180	
PA(%)	95	75	85	70	75	90	85	100	80		
OA(%)											83.89

Figure 5-1 Assessment of the sample classification results obtained by using the entire dataset of training samples. UA – User’s Accuracy, PA – Producer’s Accuracy, OA – Overall Accuracy

#### 5.1.1. Overall accuracy of the classification performed using all available training samples

The final overall accuracy obtained after running 50 iterations of classification was 84%. The final overall accuracy was obtained by taking the mean of the overall accuracies obtained at each iteration. The range of the overall accuracies was from 82% to 86%.

#### 5.1.2. Producer’s Accuracy(Sensitivity) of the classification performed using all available training samples

The Producer’s Accuracy was calculated for each class, as seen in Figure 5-1. All values were close to 100%, indicating good model prediction performance. The water class had all its samples correctly classified (100%), followed by alfalfa (95%), orchard (90%), potatoes (85%), cereals (85%), other (80%), beets (75%), onions (75%) and lastly the maize class (70%). The maize class being the least predicted efficiently makes sense, as seen in Figure 4-3, since it is spread out in the feature space and has high spectral overlap with other classes like potatoes, beets and cereals, confusing the classification process.

### 5.1.3. User's accuracy (Positive predicted values) of the classification performed using all available training samples

Figure 5-1 above shows that the classes water and alfalfa had the highest user's accuracy, namely 100%. This means all areas classified as water and alfalfa on the map were actually water and alfalfa on the ground, respectively. 93.75% of the samples classified as onions were actually onions on the ground for the onions class. A maize sample(1) was classified wrongly as onions class. For the maize class, 93.33% of the samples classified as maize were actually maize on the ground, and a potatoes sample(1) was wrongly classified as maize. Considering the within-class variation of the maize class discussed previously, the classifier performed well. 89.47% of samples classified as cereals were actually cereals on the ground. A potato sample(1) and an onion sample(1) were misclassified as cereals. 77.27% of samples classified as potatoes were actually potatoes on the ground. Beets samples(4) and cereals (1) were misclassified as potatoes. For the beets class, 75% of the samples classified as beets were actually beets on the ground. Potatoes class(1 sample), maize samples (3) and 'other' class sample(1) were wrongly classified as beets. The orchard class also had the same user's accuracy as beets. A beets sample(1), maize samples(2) and 'other' class sample(3) were wrongly classified as orchard. The 'other' class had the least user's accuracy. This could be because the samples places had a mixture of other classes like built-up areas and vegetation mixture.

### 5.1.4. Kappa statistic of the classification performed using all available training samples

The classification using all the samples had a kappa value of 82%, as per the results displayed in figure 5-1.

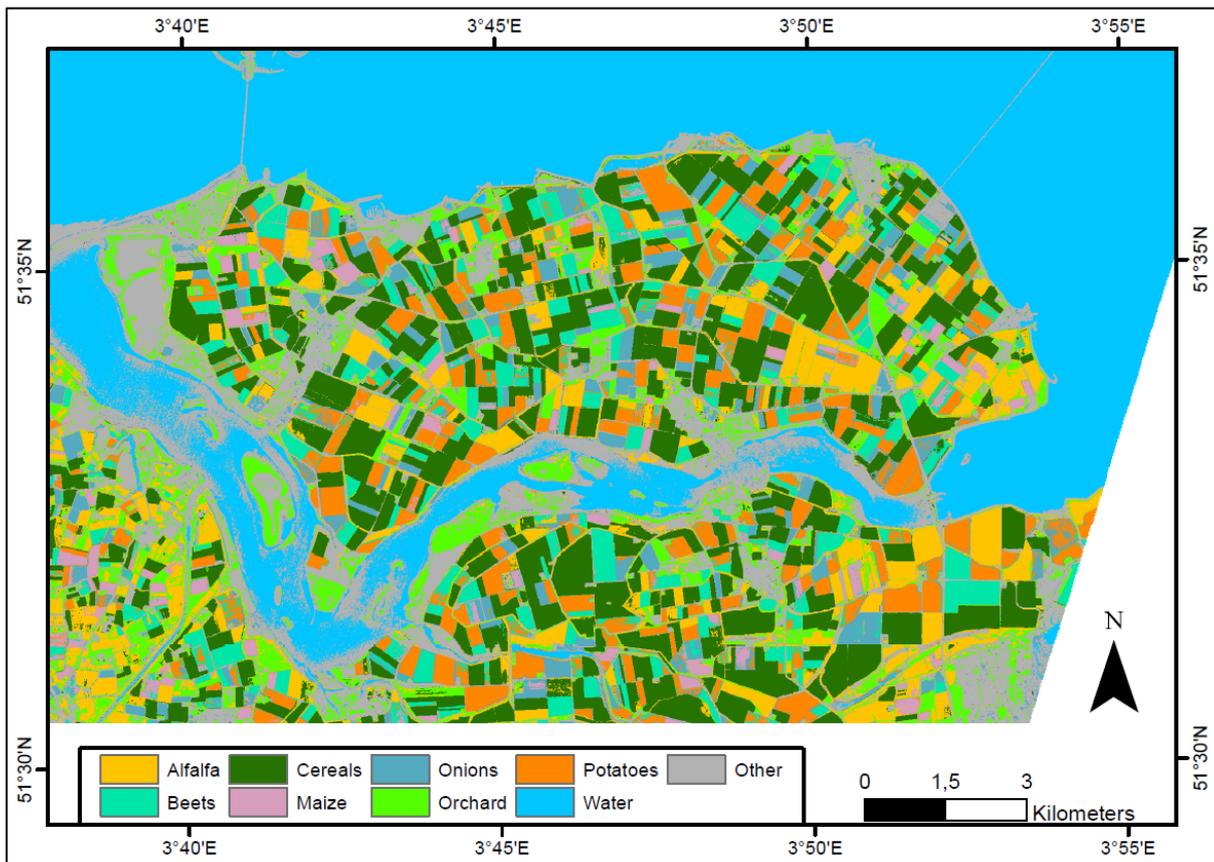


Figure 5-2 Crop type map generated from classification using all training samples (350 samples)

## 5.2. Classification using training samples generated from spectral-domain AL

### 5.2.1. Initial committee predictions before AL

Figures 5-3 and 5-4 illustrate the prediction of each learner independently and as a committee, respectively. Both learners used the same number of training samples (20 samples), summing up to 40 samples. Their predictions are similar, as seen in Figure 5-3. The initial committee prediction accuracy was 69.1%.

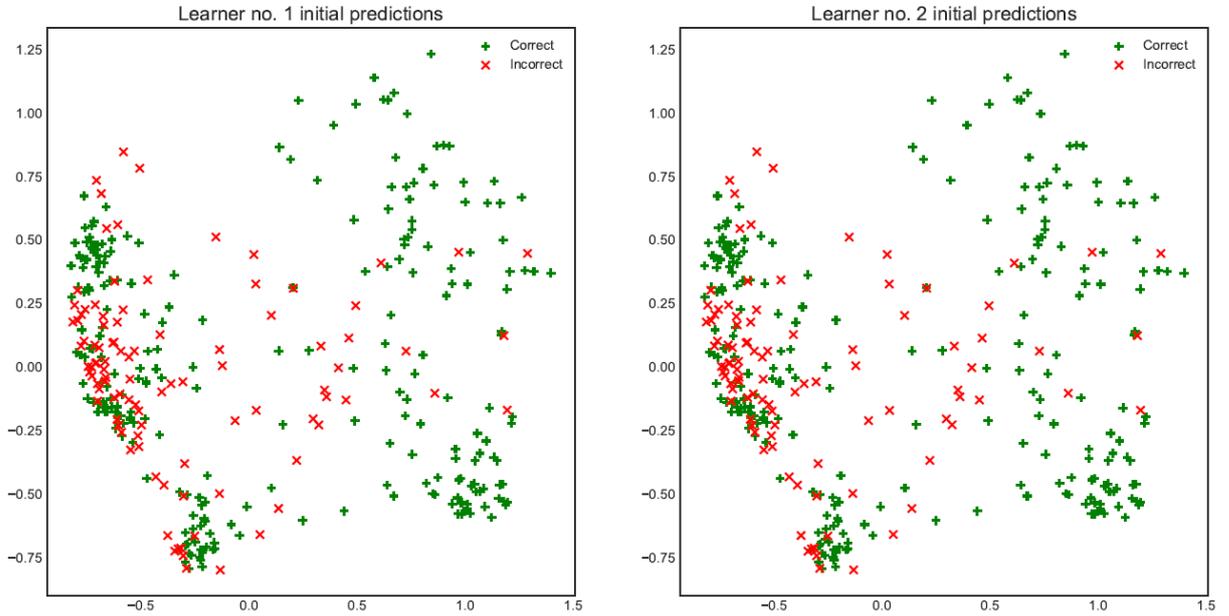


Figure 5-3 Initial predictions per individual learner. Y and X axis stand for principal component analysis components 2 and 1 respectively as created as a result of reducing the spectral features to two dimensions

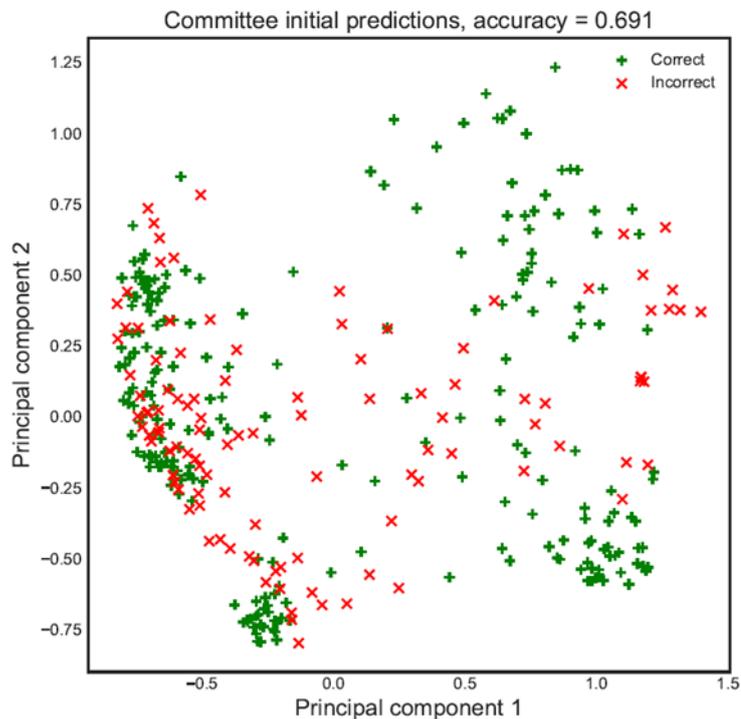


Figure 5-4 Committee predictions by initial 40 training samples.

### 5.2.2. Selection of informative samples considering the spectral domain only

The model's performance (accuracy) improved at an unsatisfactory rate after 129 queries, as evidenced by the curve flattening, with a committee prediction accuracy of 99.43 % (figure 5-6). This was the stopping criterion. Figure 5-5 depicts the increase in prediction accuracy when informative samples queried from the pool are added to the training set.

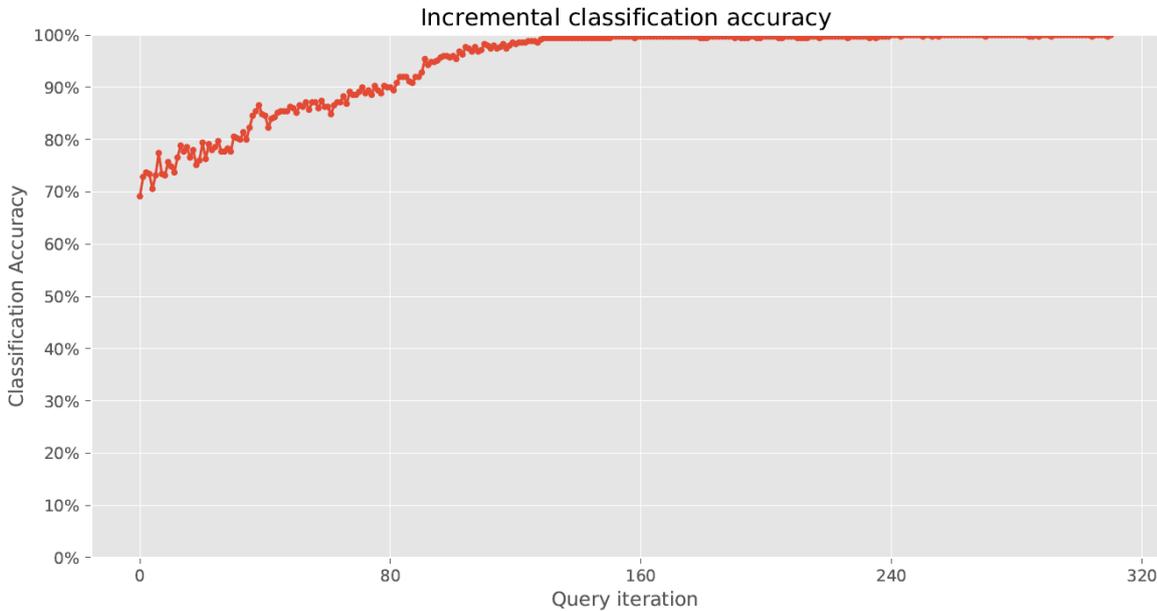


Figure 5-5 Prediction accuracy with an increment of training samples. Y axis is the increment in classification accuracy and X axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.

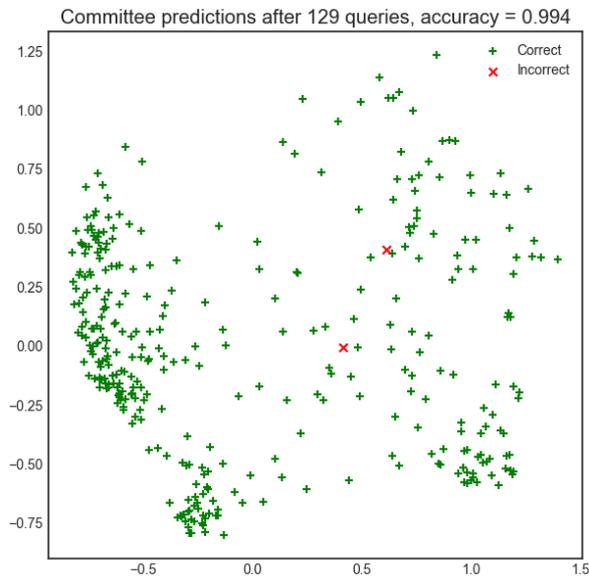


Figure 5-6 committee prediction accuracy after 129 queries. Y axis is the increment in classification accuracy and X axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.

The 129 samples plus the initial 40 samples used to train the model give a total of 169 samples. This is 48% of all the samples used for classification initially, excluding water and other classes that are not crop classes (350 samples). Table 5-1 shows the class distribution before AL(350) and after selecting informative samples using AL that considers only the spectral domain(169). The classes are not well balanced, with the potato class having the largest number of samples(41). It could be because of difficulty in the classification due to its within-class variation, as seen in Figure 4-3. Figure 5-7 and 5-8 show the distribution of the samples in the feature space before(350 samples) and after the AL algorithm (169 samples), respectively. The distributions are different in the sense that the redundancy of samples has been eliminated as shown by the colored oval shapes potatoes(purple oval), cereals(green oval), alfalfa and beets(dark blue oval), potatoes and onions(red oval) and the spectral mixture of potatoes, beets, and maize(yellow oval).

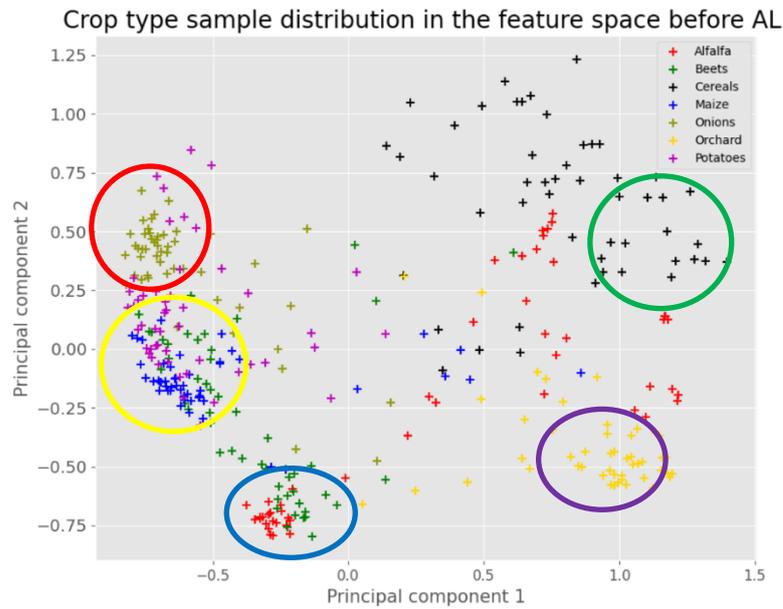


Figure 5-7 Crop sample distribution before AL .Redundancy is illustrated by the colored oval shapes potatoes(purple oval), cereals(green oval), alfalfa and beets(dark blue oval), potatoes and onions(red oval) and the spectral mixture of potatoes, beets, and maize(yellow oval).

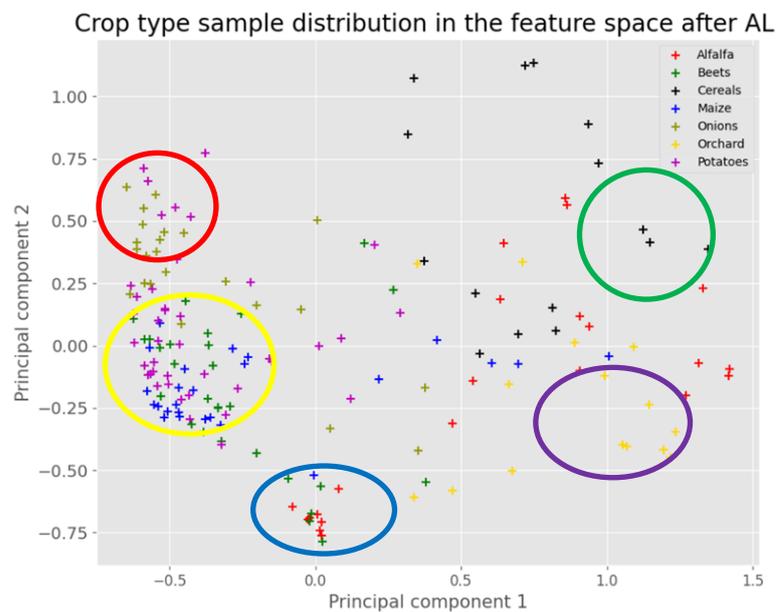


Figure 5-8 Crop sample distribution after AL in the spectral domain. Redundancy reduction shown by the colored oval shapes potatoes(purple oval), cereals(green oval), alfalfa and beets(dark blue oval), potatoes and onions(red oval) and the spectral mixture of potatoes, beets, and maize(yellow oval).

Table 5-1 Training sample class distribution before AL and after AL considering the spectral domain only

Crop type	Training sample number before AL	Training sample number after AL
Alfalfa	50	22
Beets	50	28
Cereals	50	15
Maize	50	25
Onions	50	23
Orchard	50	15
Potatoes	50	41
Total	350	169

### 5.2.3. Classification using samples obtained from the algorithm that considers the spectral domain only

The NDVI image composite was classified using the 169 samples described in the previous section plus the water and 'other' class samples. The prediction accuracy was assessed using the 180 validation samples mentioned in section 5.1. The confusion matrix and the Kappa statistic were used for assessment. Figure 5-9 illustrates the confusion matrix and other statistical metrics of the classification run. Figure 5-10 shows the crop type map generated from this classification task.

#### Confusion Matrix and Statistics

Kappa = 0.7938

	Alfalfa	Beets	Cereals	Maize	Onions	Orchard	Potatoes	Water	Other	Total	UA(%)
Alfalfa	17	0	0	0	0	0	0	0	0	17	100
Beets	0	15	0	4	0	0	3	0	1	23	65.22
Cereals	2	0	19	0	2	0	1	0	0	24	79.17
Maize	0	0	0	14	0	0	1	0	0	15	93.33
Onions	0	0	0	0	15	0	0	0	0	15	100
Orchard	0	1	0	1	0	19	0	0	6	27	70.37
Potatoes	1	4	1	1	0	0	15	0	0	22	68.18
Water	0	0	0	0	0	0	0	20	0	20	100
Other	0	0	0	0	3	1	0	0	13	17	76.47
Total	20	20	20	20	20	20	20	20	20	180	
PA(%)	85	75	95	70	75	95	75	100	65		
OA(%)											82

Figure 5-9 Assessment of the sample classification results obtained by applying the AL that considers the spectral domain only. UA – User's Accuracy, PA – Producer's Accuracy, OA – Overall Accuracy

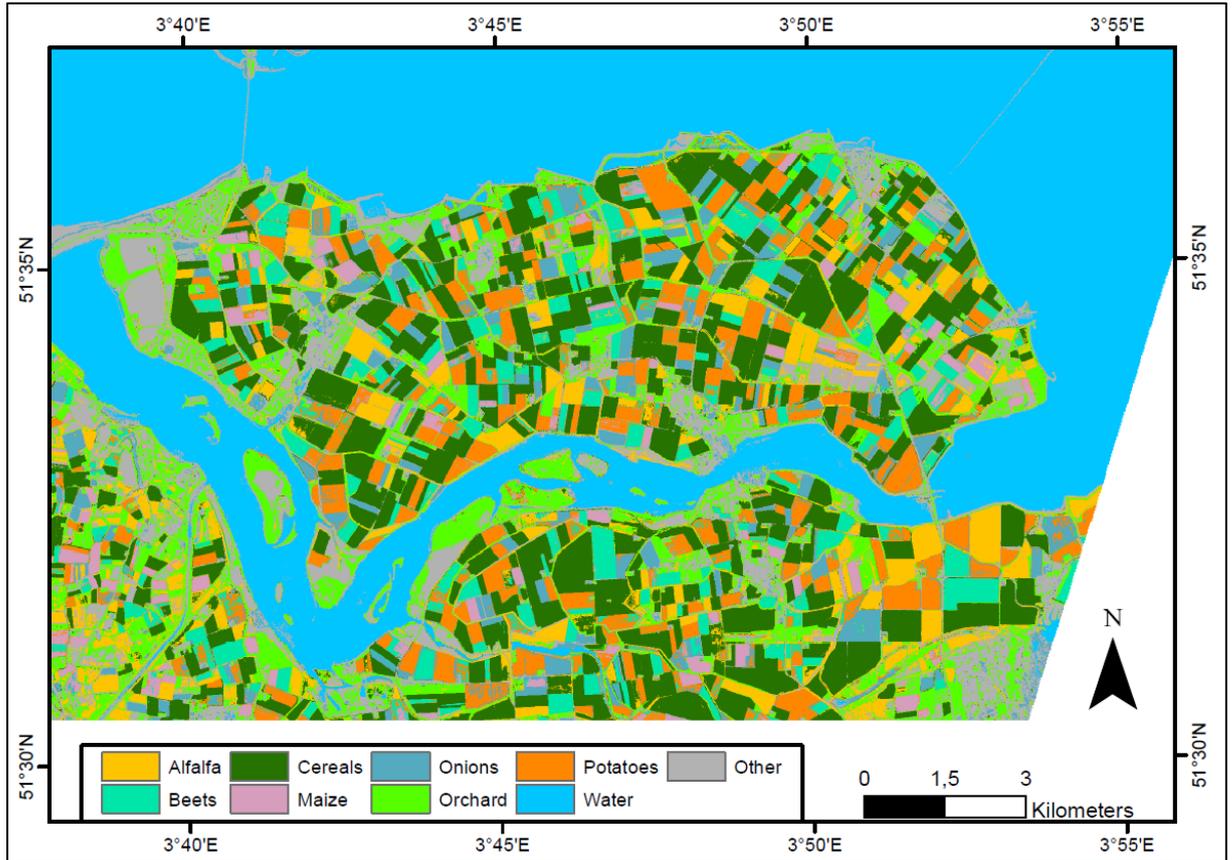


Figure 5-10 Crop type map generated from training samples selected using the AL algorithm that considers the spectral domain only

#### 5.2.4. Overall accuracy from samples resulting from AL in the spectral domain

The final overall accuracy obtained after running 50 iterations of classification was 82%. The final overall accuracy was obtained by taking the mean of the overall accuracies obtained at each iteration. The range of the overall accuracies was from 81% to 85%. The previous accuracy attained using all training samples was 84%. This means that fewer samples could be used to attain similar accuracy as many samples when carefully chosen, hence fulfilling the purpose of AL which is to operate with a low number of training samples to achieve high accuracy.

#### 5.2.5. Producer's Accuracy from samples resulting from AL in the spectral domain

The producer's accuracy was calculated for each class, as seen in Figure 5-9. All values are close to 100%, which shows good model prediction performance. The water class had all its samples correctly classified as in the previous classification, followed by orchard(95%) and cereals(95%), whose producer accuracy is higher than the previous producer accuracies for the orchard (90%) and cereals (85%), respectively. The same producer's accuracies were observed in the beets, maize, and onions classes as in the prior classification that used all samples. However, the producer's accuracy dropped from 95% to 85% for the alfalfa class, from 85% to 75% for the potatoes class and from 80% to 65% for the 'other' class.

#### 5.2.6. User's accuracy from samples resulting from AL in the spectral domain

The user's accuracy per class is shown in Figure 5-9 by the row that shows the positive predicted values. The classes water, alfalfa and onions had the highest user accuracy, namely 100%. The onions had misclassifications in the previous classification using all training samples. Therefore, an improvement was

observed in the classification of onions. For all the three classes, all samples classified as water, alfalfa and onions on the map were, therefore, water, alfalfa, and onions on the ground, respectively. The user's accuracy also increased for the others class from 64% in the previous classification to 76%. The maize class maintained the same user's accuracy as in the previous classification. 93 % of the samples classified as maize were actually maize on the ground, and a potatoes sample(1) was wrongly classified as maize as in the previous task. The user's accuracy decreased for the beets class 75% to 65%, cereals 89% to 79%, orchard 75% to 70% and potatoes 77% to 68%.

### 5.2.7. Kappa statistic of spectral domain-based classification

The AL samples' classification (169) yielded a kappa value of 79%, as per the results displayed in figure 5-9. This is close to the Kappa value in the previous classification involving all the 350 training samples (82%).

## 5.3. Classification using the AL algorithm with the spatial component

### 5.3.1. Selection of informative samples in the spectral and spatial domain

After 57 queries, the classification performance of the model increased at a non-satisfactory rate with a committee prediction accuracy of 90.90% (figure 5-12). This is less than the prediction accuracy of the previous task using the AL algorithm that excludes the spatial component(99.43%). Figure 5-11 shows the increment in prediction accuracy with the addition of informative samples queried from the pool to the training set. It is observed that after 57 queries, the curve flattens.

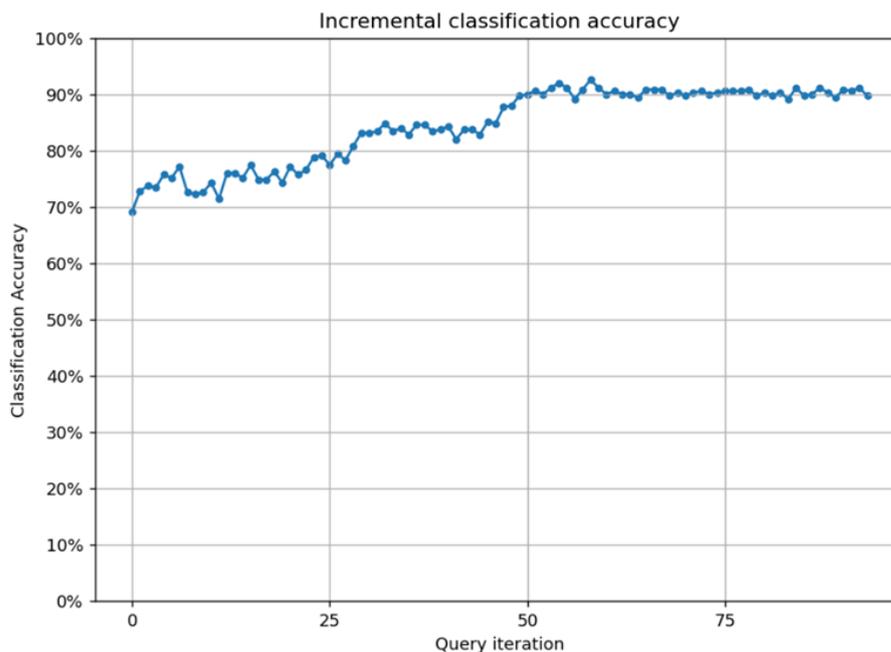


Figure 5-11 Prediction accuracy with an increment of training samples obtained from AL algorithm with the spatial component. Y- axis is the increment in classification accuracy and X- axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.

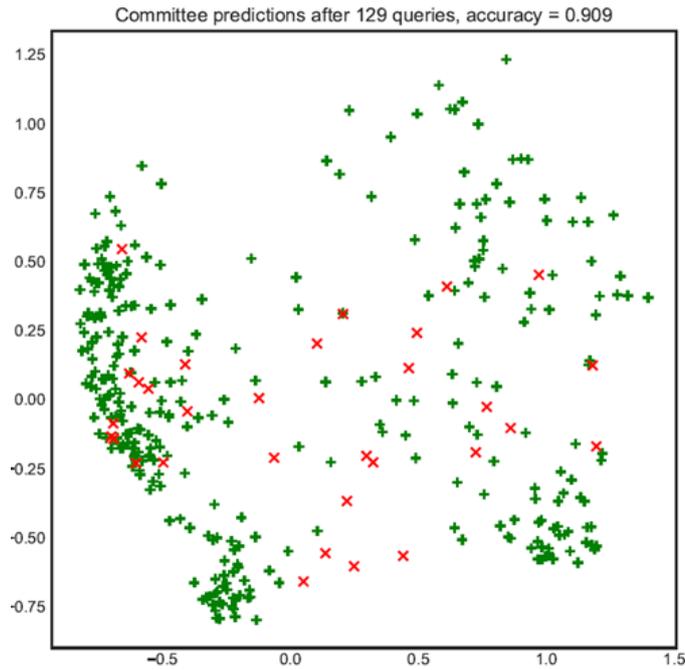


Figure 5-12 Committee prediction accuracy after 129 queries considering the spectral and spatial domain. Y- axis is the increment is classification accuracy and X- axis displays the number of iterations performed whereby in each iteration, one informative sample is chosen. The number of query iterations=the number of queried samples.

The 57 samples, together with the 40 initial samples used to train the model, give a total of 97 informative samples. This is 57.39% of the samples used for AL considering the spectral domain and 27.71% of the entire training sample dataset (350 samples). Table 5-2 shows the class distribution after selecting informative samples using the new AL algorithm that considers both the spatial and spectral domain. Classes with high interclass variation were given preference in the selection, the potato class having the largest number of samples(24) as in the previous task. Figures 5-13 and 5-14 show the spatial distribution of the samples before(350 samples) and after the AL algorithm with the spatial component (97samples), respectively. The distributions are different in the sense that the redundancy of samples has been eliminated, as shown by the colored oval shapes orchard(red oval), alfalfa(dark blue oval), maize(light green oval), cereals(yellow oval) and potatoes(purple oval).

Table 5-2 Training sample class distribution before AL and after AL considering the spectral and spatial domain

Crop type	Training sample number AL with a spatial component
Alfalfa	11
Beets	16
Cereals	8
Maize	15
Onions	16
Orchard	7
Potatoes	24
Total	97

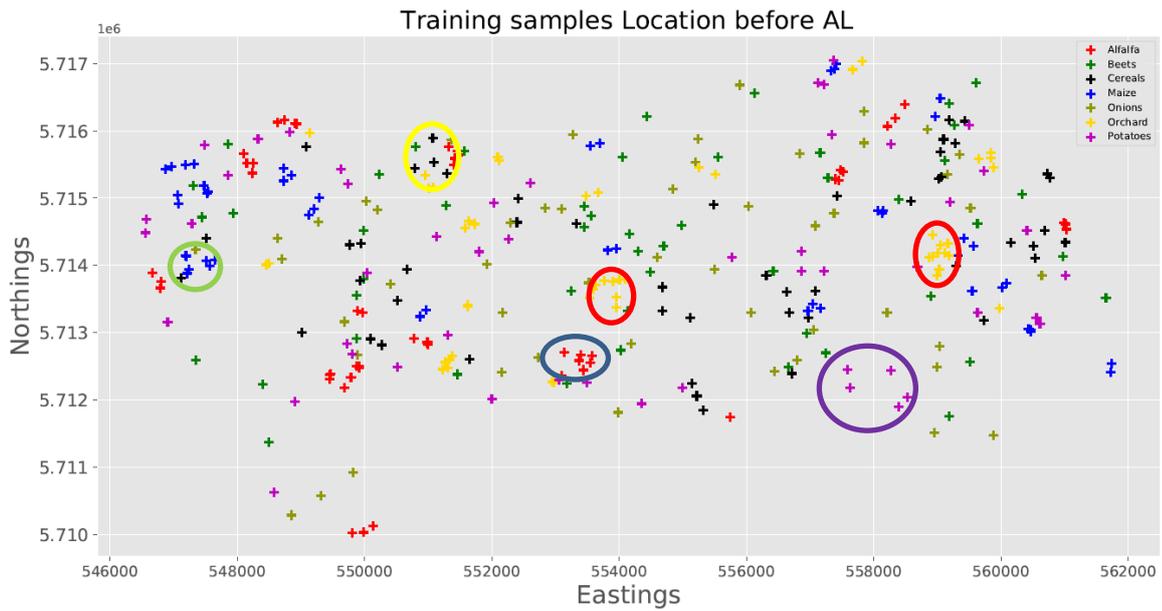


Figure 5-13 Crop sample spatial distribution before AL. Redundancy illustrated by the colored oval shapes orchard (red oval), alfalfa (dark blue oval), maize (light green oval), cereals (yellow oval) and potatoes (purple oval).

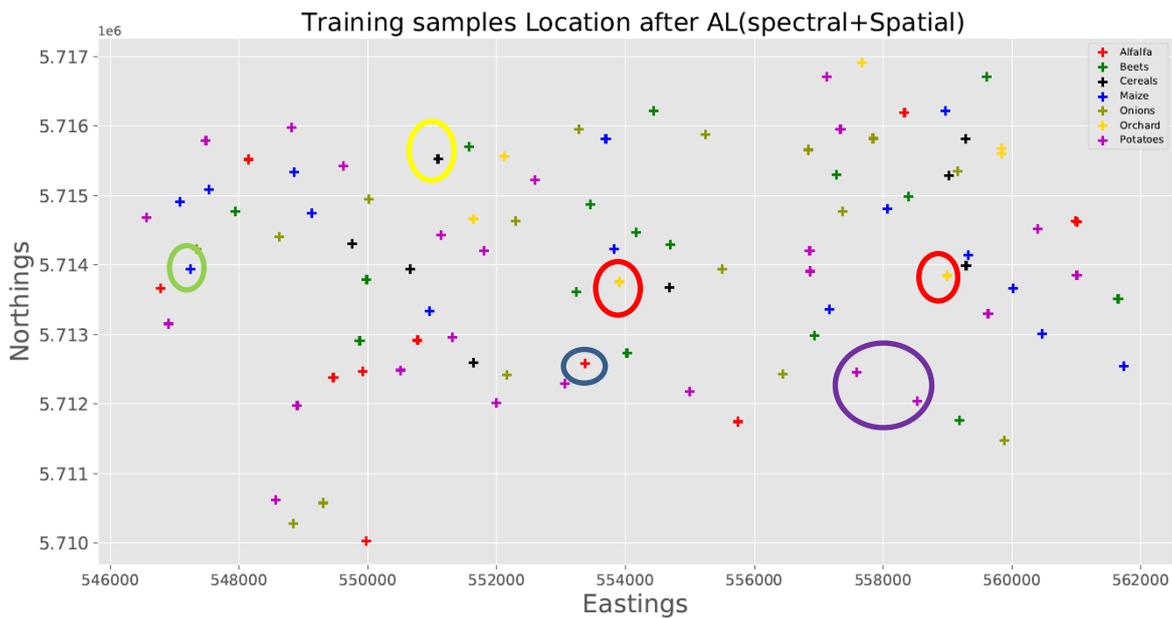


Figure 5-14 Crop sample distribution after AL in the spectral and spatial domain. Redundancy reduction is shown by the colored oval shapes orchard (red oval), alfalfa (dark blue oval), maize (light green oval), cereals (yellow oval) and potatoes (purple oval).

### 5.3.2. Classification using samples obtained from the AL with a spatial component

Figure 5-15 illustrates the confusion matrix and other statistical metrics of the classification run using the 97 samples plus the water and ‘other’ class samples, described in section 5.3.1. Figure 5-16 shows the crop type map generated from this classification task.

Confusion Matrix and Statistics  
Kappa = 0.775

	Alfalfa	Beets	Cereals	Maize	Onions	Orchard	Potatoes	Water	Other	Total	UA(%)
Alfalfa	19	0	0	0	0	0	0	0	3	22	86.36
Beets	0	15	1	1	0	0	3	0	1	21	71.43
Cereals	0	0	19	0	1	0	1	0	0	21	90.48
Maize	0	0	0	15	0	0	1	0	0	16	93.75
Onions	0	0	0	1	16	0	0	0	2	19	84.21
Orchard	0	1	0	1	0	19	0	0	8	29	65.52
Potatoes	1	4	0	1	0	0	15	0	0	21	71.43
Water	0	0	0	0	0	0	0	20	0	20	100
Other	0	0	0	1	3	1	0	0	6	11	54.55
Total	20	20	20	20	20	20	20	20	20	180	
PA(%)	95	75	95	75	80	95	75	100	30		
OA(%)											80

Figure 5-15 Assessment of the sample classification results obtained by applying the AL that considers the spectral and spatial domain. UA – User’s Accuracy, PA – Producer’s Accuracy, OA – Overall Accuracy

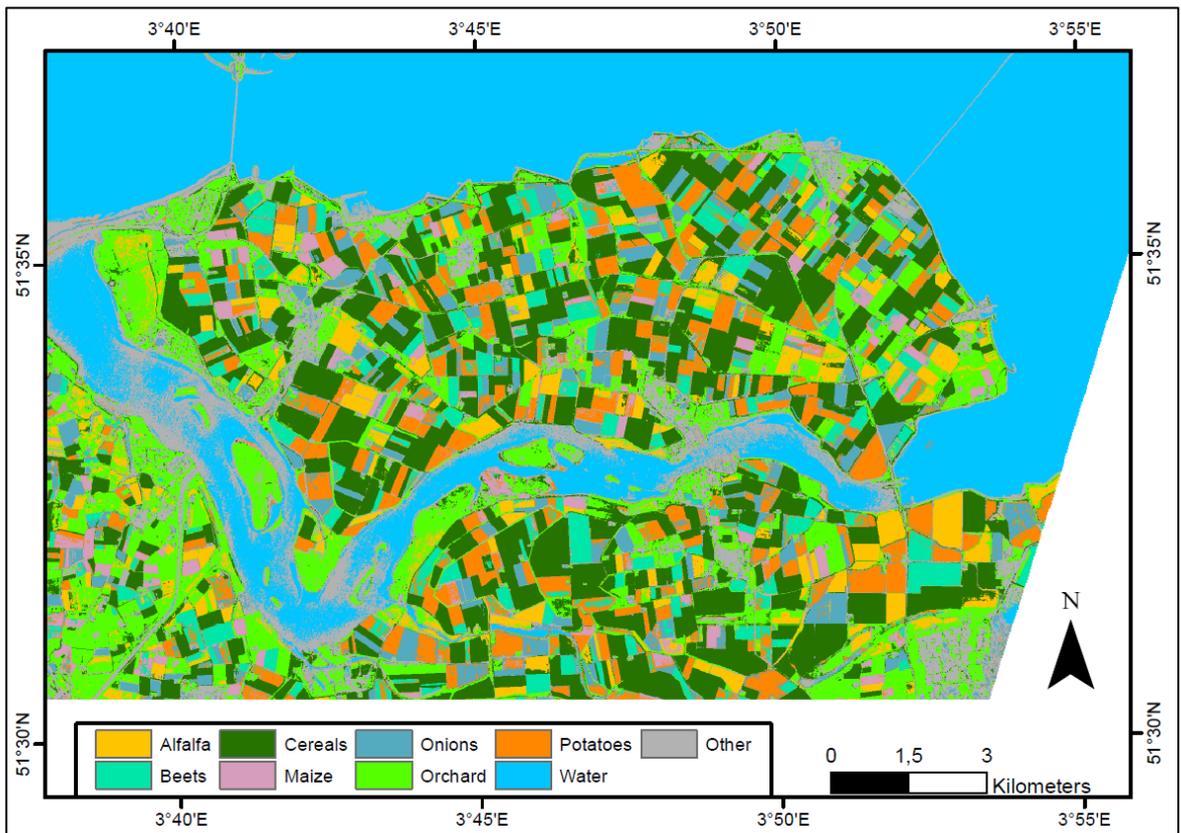


Figure 5-16 Crop type map generated from training samples selected using the AL algorithm that considers the spectral and spatial domain

### 5.3.3. Overall accuracy for the AL algorithm with the spatial component

The final overall accuracy obtained after running 50 iterations of classification was 80%. The overall accuracy range was from 78% to 82%, hence fulfilling the purpose of AL, which is to operate with a low number of training samples to attain high accuracy.

### 5.3.4. Producer's Accuracy for the AL algorithm with the spatial component

All the Producer's Accuracy values except the 'other' class are close to 100%, which shows good model prediction performance. The classes alfalfa(95%), beets (75%) and water(100%) maintained the same producer's accuracy as the one achieved when using the entire dataset of training samples. An increment in producer accuracy was observed in the classes cereals (85% to 95%), maize (70% to 75%), onions (75% to 80%) and orchard (90% to 95%). A decline in producer accuracy was observed in the potatoes class(85% to 75%) and the 'other' class(80% to 30%). The producer's accuracy improved for most classes, and a big decline is observed in the 'other' class.

### 5.3.5. User's accuracy for the AL algorithm with the spatial component

The user's accuracy per class is shown in Figure 5-15. The class water had the highest user accuracy, which was 100%. The maize class showed improvement in the classification compared to the classification that uses all training samples, 93% to 94%. All samples classified as maize were actually maize on the ground except for one potato sample, misclassified as maize. The class cereals also showed an improvement in classification from 89% to 91%. The user's accuracy decreased for all the remaining classes ;beets class 75% to 71% ,onions 94% to 84%,orchard 75% to 66%, potatoes 77% to 71% and others class 64% to 54%.

### 5.3.6. Kappa statistic for the AL algorithm with the spatial component

The classification using all the AL samples (97) yielded a kappa value of 78%, as per the iteration displayed in figure 5-15. This is close to the Kappa value in the previous classification involving all 129 training samples (79%).

## 5.4. Comparison of the AL algorithms

The AL algorithm that considers the spectral domain only and excludes the spatial component was compared to the AL that includes the spatial component. The comparison was made by looking at the overall classification accuracy of the two methods, Kappa statistic, number of samples used, user's accuracy, and misclassified samples' quantity. The outputs from the classification using all samples were used as a reference to compare the two AL algorithms.

### 5.4.1. Kappa statistic comparison of the two AL algorithms

Table 5-3 shows that the Kappa values for all the two AL approaches are close to the reference Kappa value, which is the Kappa value calculated from the entire data set of training samples. The difference in Kappa values between the reference classification, which uses the whole training sample dataset, and the AL classifications, which excludes the spatial component, is 3%(82%-79%) and 4%(82%-78%) for the AL, including the spatial element, which is a small difference. The difference in Kappa values for the two AL algorithms is 1%(79%-78%), even lesser, making them similar. In addition, the Kappa value varies for all the 50 iterations, so there may be cases when there is an insignificant difference in the Kappa values. Basing on this fact, the new AL that includes the spatial component can be relied on despite using few samples for classification.

#### 5.4.2. Overall accuracy comparison of the two AL algorithms

Table 5-3 also shows the overall accuracies attained after classification using samples selected using the AL algorithms with and without the spatial component. The overall accuracy variations with iterations that were run were similar to some extent, from 82% to 86% when all samples were considered, 81% to 85% when excluding the spatial component and 78% to 82% when the spatial component was included. From the above outcomes, the three approaches achieve similar accuracies. Despite having fewer samples, the AL algorithm with the spatial component attained a difference of 4% and, the AL algorithm without the spatial component achieved a difference of 2% when compared to the classification accuracy attained using the whole training sample dataset. A difference of 2% was observed between the overall accuracies of the two AL algorithms, which is small. The algorithm that includes the spatial component has a greater advantage since it attained almost similar overall accuracy despite using 57 % of the samples used for AL considering the spectral domain only and 28% of the entire training dataset (350 samples).

Table 5-3 Overall accuracies and Kappa statistics for the classification run using all training samples, samples from the AL method that excludes the spatial component and samples from the AL method that includes the spatial component.

Scenario	Sample number	Kappa statistic	Overall accuracy(%)
Classification using all samples(used as reference data)	350	0.82	84
Classification using AL samples excluding spatial component	169	0.79	82
Classification using AL samples including spatial component	97	0.78	80

#### 5.4.3. User's accuracy comparison of the two AL algorithms

Table 5-4 shows the user's accuracies attained per class after classification for all the three scenarios in tabular format, and figure 5-17 shows them visually. The difference in the user's accuracy values is small for both AL algorithms compared to the user's accuracy attained using all samples(reference). In the Alfalfa class, the AL excluding the spatial component performs better than the new algorithm that includes the spatial component by having the same user's accuracy as the reference approach. It also performs better for the onions and orchard classes. However, the AL with spatial component performs better in classifying the classes beets, cereals, maize, and potatoes. This means the new AL algorithm that considers the spatial component performs well compared to the AL excluding the spatial component since it correctly classifies the majority of the classes. Generally, there is not much difference between the two AL algorithms in terms of proper classification of classes despite having different training sample number.

Table 5-4 Comparison of the user's accuracies for the three scenarios per crop type class.

Landcover class	All 350 samples(%)	AL without a spatial component 169 samples (%)	AL with a spatial component 97 samples(%)
Alfalfa	100%	100	86
Beets	75%	65	71
Cereals	89	79	90
Maize	93	93	94
Onions	94	100	84
Orchard	75	70	66
Potatoes	77	68	71

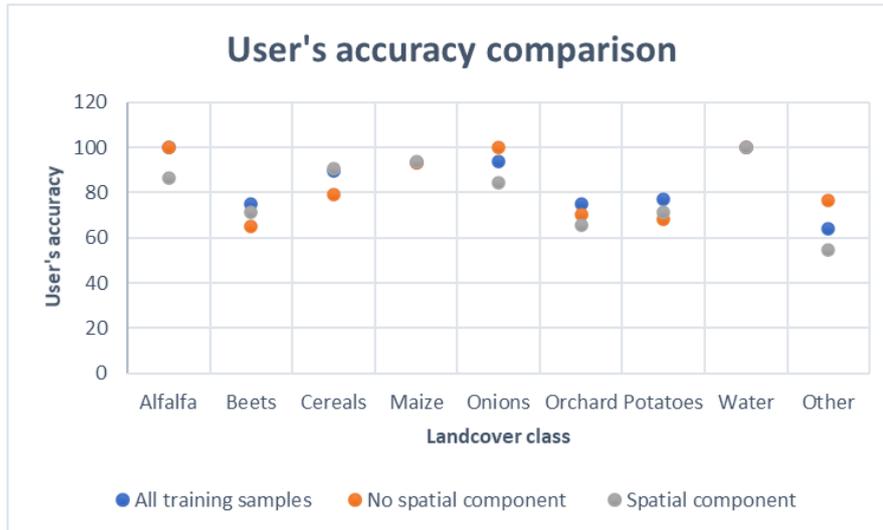


Figure 5-17 Comparison of the user's accuracies for the AL algorithm with the spatial component(gray), AL algorithm without the spatial component (orange) to the reference accuracy obtained using the entire dataset.

#### 5.4.4. Comparison in the quantity of misclassification of samples

The misclassified samples per class were also analyzed for both algorithms to evaluate both techniques. The overall number of misclassified samples in the classification results obtained by the AL algorithm with the spatial component was lower than in the classification results obtained by the AL algorithm without the spatial component, as illustrated by Table 5-5. Regarding the number of misclassifications, the AL algorithm with the spatial component performs similarly to the classification using all samples. Figure 5-18 shows a visual assessment of the misclassification. The ground truth label, classified label by the AL algorithm that excludes the spatial component, and classified label by the AL method that includes the spatial component are shown by columns 1,2 and 3, respectively. The first row shows a cereals plot(red) misclassification by the AL algorithm that includes the spatial component. The second row illustrates orchard class misclassification(green) by the AL algorithm that excludes the spatial component. The third row illustrates misclassification of the alfalfa class (orange) by both AL algorithms. The AL with the spatial component was mainly affected by the 'other' class since it is a mixed class.

Table 5-5 Comparison of the misclassified samples for the AL method that excludes the spatial component (169) and samples from the AL that includes the spatial component (97)

Landcover class	Misclassified samples from approach that uses all training data	Misclassified samples from AL algorithm with no spatial component	Misclassified samples from AL algorithm with a spatial component
Alfalfa	0	0	0
Beets	4	7	5
Cereals	2	5	2
Maize	1	1	1
Onions	1	0	1
Orchard	3	2	2
Potatoes	5	7	6
Total	16	22	17

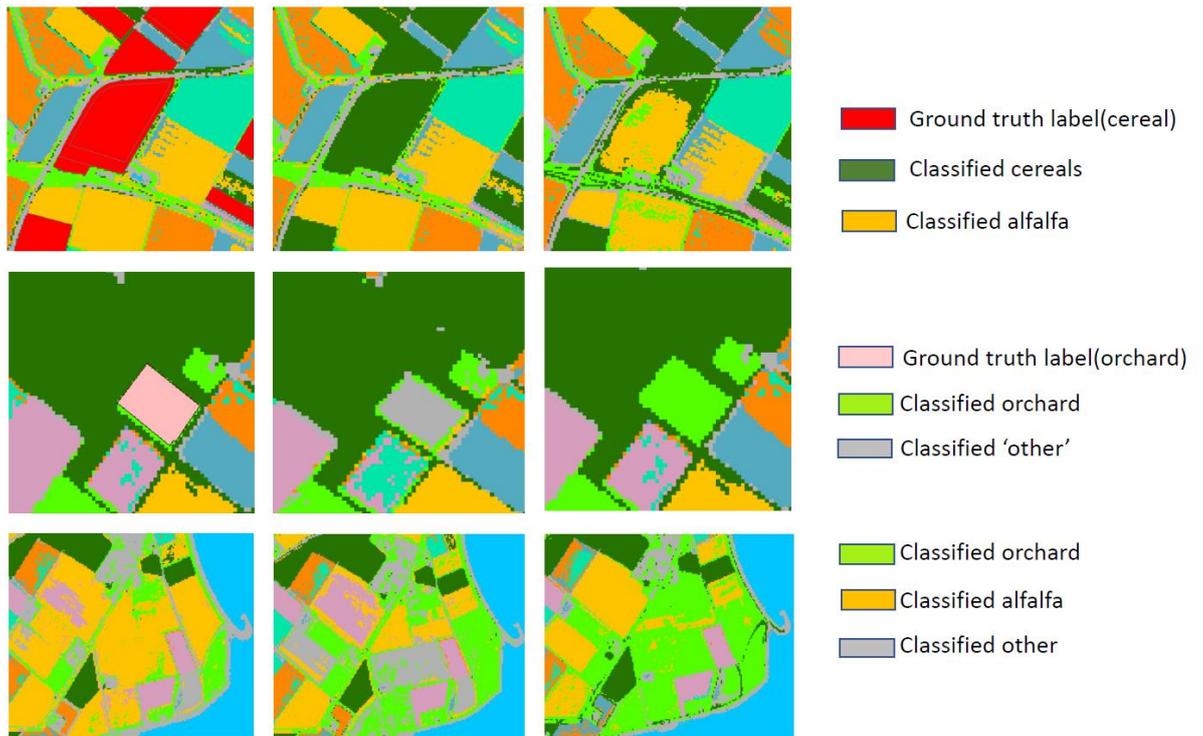


Figure 5-18 Misclassification of cereals class by the AL algorithm that includes the spatial component (first row), orchard class by AL algorithm that excludes the spatial component (second row) and a misclassification by both AL algorithms (third row).

#### 5.4.5. Comparison of the sample number and distribution of the two AL algorithms

The number of samples per class generated from the AL with the spatial component was almost half the number of samples from the AL algorithm that excludes the spatial component; however, a similarity in the pattern of class number distribution was observed in both methods. From table 5-6, it is seen that for AL generated training samples, the number of samples per class was not the same as the case that uses all training samples for classification (50 samples per class). Preference was given to the classes with a high interclass spectral variation.

Table 5-6 Variation of the number of samples used for classification for the AL with spatial component and the AL algorithm that excludes the spatial component

Crop type	All training samples	Training sample number (spatial component included)	Training sample number (spatial component excluded)
Alfalfa	50	11	22
Beets	50	16	28
Cereals	50	8	15
Maize	50	15	25
Onions	50	16	23
Orchard	50	7	15
Potatoes	50	24	41
Total	50	97	169

## 6. DISCUSSION

This chapter contains a discussion of the research findings. The performance of the developed AL algorithm that includes the spatial component is compared to the performance of the AL algorithm that excludes the spatial component. The findings are summarized in section 6.1, followed by sections 6.2, 6.3, 6.4, and 6.5, which analyze the created AL method's strengths, weaknesses, opportunities, and threats.

### 6.1. Summary of findings

This research assessed the performance of the AL method that incorporates the spatial component in the selection of informative samples for crop type mapping using Sentinel-2 NDVI time series. This method was applied in a study area called Noord Beveland in the southwest Netherlands. The samples' performance using this method was compared to the performance of the informative samples selected using the AL algorithm that excludes the spatial component. The number of samples used, overall classification accuracy, Kappa statistic, user's accuracy, and misclassification rate were all considered in the performance evaluation. The original training sample dataset and its outputs served as the benchmark against which both AL techniques were compared. The method with the spectral and spatial components reached an accuracy of 80% using 97 samples, while the method using only the spectral characteristics attained an overall accuracy of 82% using 169 samples against a baseline of 84% accuracy when using the entire training dataset of 350 samples. The method that used both spectral and spatial components had a Kappa statistic of 78%, while the method that just used spectral characteristics had a Kappa statistic of 79%, compared to a Kappa statistic of 82% when the whole training dataset was used. Compared to the AL technique without the spatial component, the AL technique with the spatial component showed a higher user accuracy for most of the crop classes and also scored well in classifying classes with within-class variation in spectral features such as maize, beets and potatoes.

### 6.2. Strengths of the developed algorithm

This research showed that by including both the spectral and the spatial component in sample selection, almost the same overall accuracy and Kappa statistic can be achieved using only a quarter of the entire training dataset. Including only the spectral component in sample selection needed only half the samples, compared to an approach without considering spectral or spatial diversity in sample selection. This implies that similar overall accuracies and Kappa statistics were attained with a smaller number of training samples, lowering the amount of manual work required for sample labeling and the computational time required to train the classifier. Redundancy in terms of selecting similar samples was also omitted. In this aspect, the AL algorithm with the spatial component performs well compared to the AL that excludes the spatial component in the sense of using a small number of training samples to attain similar classification accuracies and Kappa statistics. The aspect of having fewer samples when the spatial component is involved in AL is similar to research by Pasolli et al.(2011), albeit in their study, integrating the spatial component resulted in higher overall accuracy than the AL method using only spectral information. User's accuracy per class was computed in this study, as shown in Table 5-4. Compared to AL without the spatial component, AL with the spatial component resulted in higher user's accuracy for most of the classes. This was observed in the classes beets, cereals, maize and potatoes. For the classes, alfalfa, orchard, and onions, including the spatial component, did not increase accuracy. For the case of producers' accuracy, including the spatial component in the AL algorithm gave an improvement in the classes alfalfa, maize, and onions. The rest of the classes maintained the same producers' accuracy in both AL methods. Despite using 27% of the entire training dataset for classification, the developed AL algorithm achieved good classification

results in terms of the overall accuracy, Kappa coefficient, the user's accuracy and producer's accuracy are considered.

Several classes in our study showed inter-class similarities, as illustrated in Figure 5-7. The beets, maize, and potatoes exhibited a lot of spectral overlap. Similarly, onions overlapped with the potatoes class, and alfalfa overlapped with cereals, beets, and orchard classes. In such cases, classification is difficult and may have an impact on the information derived from the samples since one class may be mistakenly assigned to the other due to a close resemblance (Gebbinck, 1998). High spectral overlap between classes can lead to sample misclassification; limiting the sample size makes it much more difficult to distinguish spectrally similar classes. Despite using 27% of the total training sample dataset, the developed AL method favoured high-uncertainty classes when selecting informative samples, as they may improve classification accuracy. Improving classification accuracy for these classes required a high number of samples due to their uncertainty; therefore, the algorithm favoured classes with inter-class similarity in our scenario. Table 5-6, which shows the number of samples per class chosen by the method, verifies this. Classes like potatoes, beets, onions, maize, and alfalfa which have high spectral overlap, were given more weight in the selection of informative samples, while the orchard and cereals classes whose spectral distribution is satisfactory were given less weight. In terms of user's and producer's accuracy of the classes, the proposed AL methodology yielded promising results for classes with a high interclass similarity such as maize (user's accuracy, UA, of 94% and producer's accuracy, PA, of 75%), onions (UA of 84% and PA of 80%), potatoes (UA of 71% and PA of 75%), beets (UA of 71% and PA of 75%) and alfalfa (UA of 86% and PA of 95%). Results from other studies illustrated that classes with high interclass similarity like maize and potatoes have proved to be a challenge in the classification process and hence cause poor classification performance (Belgiu, Bijker, Csillik, & Stein, 2021). The same challenge was reported in research about mapping complex farming areas (Gella, Bijker, & Belgiu, 2021). Therefore, the developed AL method demonstrates good results in classifying classes with high interclass similarity. Classes with a small interclass spectral similarity such as cereals performed well too (UA of 90% and PA of 95%) except for the orchard class (UA of 66% and PA of 95%), which due to confusion induced by the 'other' class which was not included in the AL process. The confusion between these two classes was because they were mixed classes. The 'other' class consisted of all other classes except the seven crop classes used for this research and the water class. This caused confusion with the orchard class, which had the pears and apple trees. Usually, the trees are planted in rows with grass between them, and the 'other' class also contains grass, so this may also increase confusion. In general, the algorithm with the spatial component performs good classification even with crop classes that have high spectral overlap.

### 6.3. Weaknesses of the algorithm

The developed AL algorithm is sensitive to the choice of range. Sample point pairs whose in-between distance is greater than the range value are not spatially correlated. Poor variogram estimation means a poor description of the spatial dependency of the area, hence affecting the choice of informative samples. When the range is too low, there is a possibility of the algorithm selecting samples that are redundant in terms of informativeness, with reference to Tobler's law of Geography that says near things are most likely similar. On the other hand, ranges that are too high may also cause the algorithm to leave out some informative samples and yield a sample structure that may be less informative. Therefore, a careful choice of the variogram that yields an optimal sample design must be made. A solution to this could be optimizing variogram estimation using Spatial Simulated Annealing (SSA) (Groenigen & Stein, 1998).

### 6.4. Opportunities

The developed spatially explicit AL method will impact future applications when a low number of training samples is available. It generates a lower number of samples that can still achieve good classification

results. This reduces the computational cost and saves time. One of the drawbacks highlighted by the existing supervised classification methods is insufficient training samples for effective classification (Stumpf et al., 2014; Ball et al., 2018). However, similar to Tuia et al.(2009), our study revealed that we should not focus solely on the number of samples but also on the representativeness of the samples. The developed algorithm performs well in this aspect by selecting a low number of informative samples that are highly representative of the crop classes. This means the developed spatially explicit AL method addresses the problem of training sample insufficiency in the Machine Learning field by achieving good classification results using few informative samples at a low computational cost and using less time.

### **6.5. Threat**

Challenges might be faced when applying the developed spatially explicit AL method to crop areas with different spatial patterns. Having an area with varying spatial patterns means creating many variograms to capture the spatial variability, which is practically challenging. Having many variograms means having various ranges. This makes it difficult to make a choice of a suitable range that is suitable for selecting informative samples spatially.

## 7. CONCLUSION AND RECOMMENDATIONS

This chapter concludes this research by stating the findings in relation to the main goal, specific objectives, and research questions. Limitations of the study and recommendations based on this research are also made.

### 7.1. Conclusion

The main goal of the research topic was to develop a spatially explicit AL method for crop type mapping using satellite image time series. The data acquired and used for this research was Sentinel-2 image time series and crop parcel data for the year 2019 in the study area called Noord Beveland in the Netherlands. This research managed to come up with an AL method that incorporates the spatial component in the selection of informative samples which were used for crop mapping. This goal was met by addressing the specific objectives of the research through answering the related research questions as explained in the following sections:

#### 7.1.1. Addressing the research questions

**Objective 1:** To systematically investigate different spatial metrics that can be used to improve state-of-the-art AL methods.

**Research question 1.1: What metrics can be used to assess spatial autocorrelation between the labels in the spatial domain?**

Several spatial autocorrelation metrics can be used to assess spatial autocorrelation as described in (Amgalan, Mujica-Parodi, & Skiena, 2020). As elaborated in section 2.6, Global spatial autocorrelation measures like Moran's I, Geary's C ratio, Joint count statistics, among others, are used to quantify the level of clustering across an area of interest. However, they yield only one single statistic that describes the spatial autocorrelation of the entire region without identifying where the similarity (or dissimilarity) occurs. The global spatial autocorrelation measures determine whether there is clustering or not but not where the clusters are. To locate where the clusters are, Local Indicator of Spatial Association (LISA), local Moran's I, among others, are used as metrics. They locate the clusters and determine the association between them. They also give a single statistic to quantify the spatial autocorrelation of the clusters. Another metric is a variogram, which is a geostatistical approach. It is one of the geostatistical tools used to show spatial correlation by plotting the variance of point pairs with increasing distance between them (Curran, 1988). It is used to visualize and model spatial variation.

**Research question 1.2. : What criteria should be considered in choosing the best metrics for assessing the spatial autocorrelation between the labels in the spatial domain?**

One of the criteria considered was the sample distribution pattern. In this research, it was vital to understand where the similarity in sample distribution occurs in order to select informative samples. The Global measures of spatial autocorrelation were not suitable for this research since they only show whether there is clustering(similarity) and not where the clusters occur. They just give a single statistic to quantify the similarity(or dissimilarity). Local Indicator of Spatial Association (LISA) can identify where the similarity(clusters) and dissimilarity occur and also give a single statistic to describe the spatial association between these clusters. In addition, both the Local Indicators of Spatial Association and Global measures of spatial autocorrelation are area-based methods of spatial association (Guo, Du, Haining, & Zhang, 2012); however, this research uses point samples. Therefore, a point-based spatial autocorrelation metric was to be used to determine the point-to-point relation in terms of the spatial distribution. Since a variogram quantifies the spatial autocorrelation and shows how spatial variation changes as a function of the distance between point location pairs, it was considered appropriate for this

research. The range parameter of a variogram was used as a value to select informative samples that were used for crop mapping. Section 4.4 explains how the range was used in the selection of informative samples in the AL process.

Another factor to look at was the diversity and representativeness of points. Caution was taken when choosing the range value among the range values obtained from the 12 NDVI images. This is because choosing a high range could cause the algorithm to leave out some informative samples and yield a sample structure that may be less informative, and when the range is too low, there is a possibility of the algorithm selecting samples that are redundant in terms of informativeness.

**Objective 2:** To test the performance of the developed AL method and assessing its effectiveness in crop type mapping for satellite image time series.

**Research question 2.1: How does the developed AL method perform in comparison to the AL algorithm that excludes the spatial component?**

The developed AL method performed well in the study area despite using half the number of informative samples as the AL algorithm that excludes the spatial component and almost a quarter of the entire dataset of training samples. It yielded an overall classification accuracy of 80%, which is lower by 2% from the AL algorithm that excludes the spatial component, a Kappa statistic of 78%, which is lower by 1%, and its user's accuracy was higher for the majority of the crop classes.

## 7.2. Limitations

The following limitations affected the study:

- The lack of ground truth data for the 'other' class affected the classification results. This class represented all of the remaining landcover classes with the exception of the water class and the seven crop classes employed in this study.
- Using time-series satellite images to account for crop development across the growing seasons was constrained by high cloud coverage. As a result, one NDVI image was acquired per month.

## 7.3. Recommendations

- Other ways in which the variogram could be used for the purpose of sample design, such as Simulated Annealing (Srivastava, Stein, Rossiter, Garg, & Garg, 2016; Shetty, Gupta, Belgiu, & Srivastav, 2021), could be used as an intermediate step in the selection of samples with respect to the spatial distribution of each class.
- Future research could be done in applying the variogram in the sample design for areas with different spatial variations.

Therefore, the developed AL method demonstrates that integrating spatial information in AL is a useful strategy in crop mapping since it allows for efficient crop mapping with a low number of informative training samples at a cheap annotation and computing cost. Furthermore, this approach could be extended to other forms of landcover mapping besides crop mapping for future research.

## LIST OF REFERENCES

- A. Liu, Jun, G., & Ghosh, J. (2008). Active learning with spatially sensitive labeling costs. *NIPS Workshop on Cost-Sensitive Learn., Whistler, Canada*, 1–8.
- Abdel-Rahman, E. M., Mutanga, O., Adam, E., & Ismail, R. (2014). Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88(November), 48–59. <https://doi.org/10.1016/j.isprsjprs.2013.11.013>
- Adla, R., Group, G., Engineering, C., & Lafayette, W. (2014). Review of Active Learning Algorithms in Remote Sensing , exploring scope for parallelising heuristics in nEQB and MS-cSV algorithms Review of Active Learning Algorithms in Remote Sensing , exploring scope for parallelising heuristics in nEQB and MS-cSV al, (June).
- Amgalan, A., Mujica-Parodi, L. R., & Skiena, S. S. (2020). Fast spatial autocorrelation. *Proceedings - IEEE International Conference on Data Mining, ICDM, 2020-Novem*, 12–21. <https://doi.org/10.1109/ICDM50108.2020.00010>
- Ball, J. E., Anderson, D. T., & Chan, C. S. (2017). Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(04), 1. <https://doi.org/10.1117/1.jrs.11.042609>
- Ball, J. E., Anderson, D. T., & Wei, P. (2018). State-of-the-art and gaps for deep learning on limited training data in remote sensing. *International Geoscience and Remote Sensing Symposium (IGARSS), 2018-July*, 4119–4122. <https://doi.org/10.1109/IGARSS.2018.8518681>
- Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta*, 760(June 2012), 25–33. <https://doi.org/10.1016/j.aca.2012.11.007>
- Belgiu, M., Bijker, W., Csillik, O., & Stein, A. (2021). Phenology-based sample generation for supervised crop type classification. *International Journal of Applied Earth Observation and Geoinformation*, 95(July 2020), 102264. <https://doi.org/10.1016/j.jag.2020.102264>
- Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204(September 2017), 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>
- Blöschl, G. (2002). *Geostatistics for Environmental Scientists*. *Vadose Zone Journal* (Vol. 1). <https://doi.org/10.2136/vzj2002.3210>
- Breiman, L. (2001). *Random Forests* (Vol. 45).
- Bruzzone, L., & Persello, C. (2009). Active learning for classification of remote sensing images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3, 693–696. <https://doi.org/10.1109/IGARSS.2009.5417857>
- Bruzzone, L., & Persello, C. (2010). Recent trends in classification of remote sensing data: Active and semisupervised machine learning paradigms. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3720–3723. <https://doi.org/10.1109/IGARSS.2010.5651236>
- Buja, K., & Menza, C. (2013). Sampling Design Tool for ArcGIS - Instruction Manual. NOAA's *Biogeography Branch*, 1–16. Retrieved from <http://www.arcgis.com/home/item.html?id=ecbe1fc44f35465f9dea42ef9b63e785>
- C. Sun, Y. Bian, T. Zhou, and J. P. (2019). Using of Multi-Source and Multi-Temporal Remote. *Sensors*, 19, 1–23. Retrieved from <https://www.mdpi.com/1424-8220/19/10/2401/htm>
- Cao, X., Yao, J., Xu, Z., & Meng, D. (2020). Hyperspectral Image Classification with Convolutional Neural Network and Active Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), 4604–4616. <https://doi.org/10.1109/TGRS.2020.2964627>
- Cervantes-Godoy, D., Dewbre, J., PIN, Amegnaglo, C. J., Soglo, Y. Y., Akpa, A. F., ... Swanson, B. E. (2014). *The future of food and agriculture: trends and challenges*. *The future of food and agriculture: trends and challenges* (Vol. 4). <https://doi.org/10.2307/4356839>
- Chutia, D., Bhattacharyya, D. K., Sarma, K. K., Kalita, R., & Sudhakar, S. (2016). Hyperspectral Remote Sensing Classifications: A Perspective Survey. *Transactions in GIS*, 20(4), 463–490.

- <https://doi.org/10.1111/tgis.12164>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Crawford, M. M., Tuia, D., & Yang, H. L. (2013). Active learning: Any value for classification of remotely sensed data? *Proceedings of the IEEE*, 101(3), 593–608. <https://doi.org/10.1109/JPROC.2012.2231951>
- Curran, P. J. (1988). The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, 24(3), 493–507. [https://doi.org/10.1016/0034-4257\(88\)90021-1](https://doi.org/10.1016/0034-4257(88)90021-1)
- Demir, Begm, Persello, C., & Bruzzone, L. (2011). Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3), 1014–1031. <https://doi.org/10.1109/TGRS.2010.2072929>
- Demir, Begüm, Bovolo, F., & Bruzzone, L. (2012). Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5 PART 2), 1930–1941. <https://doi.org/10.1109/TGRS.2011.2168534>
- Demir, Begüm, Bovolo, F., & Bruzzone, L. (2013a). Classification of time series of multispectral images with limited training data. *IEEE Transactions on Image Processing*, 22(8), 3219–3233. <https://doi.org/10.1109/TIP.2013.2259838>
- Demir, Begüm, Bovolo, F., & Bruzzone, L. (2013b). Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1), 300–312. <https://doi.org/10.1109/TGRS.2012.2195727>
- Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). *Model-based geostatistics*. *Journal of the Royal Statistical Society. Series C: Applied Statistics* (Vol. 47). <https://doi.org/10.1111/1467-9876.00113>
- Engström, K., & Esbensen, K. H. (2018). Variographic assessment of total process measurement system performance for a complete ore-to-shiping value Chain. *Minerals*, 8(7). <https://doi.org/10.3390/min8070310>
- Esri. (2016). FAQ: How can one determine the lag size to use in Geostatistical Analyst? Retrieved from <https://support.esri.com/en/technical-article/000004923>
- Esri. (2018). Basic Registration of Crop Plots. Retrieved from <https://www.arcgis.com/home/group.html?id=2c19a595e2c54bc7a49dd14bbb690e95#overview>
- Gebbinck, M. S. . (1998). *Decomposition of Mixed Pixels in Remote Sensing Images to Improve the Area Estimation of Agricultural Fields*.
- Gella, G. W., Bijker, W., & Belgiu, M. (2021). Mapping crop types in complex farming areas using SAR imagery with dynamic time warping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175(March), 171–183. <https://doi.org/10.1016/j.isprsjprs.2021.03.004>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Groenigen, J. W., & Stein, A. (1998). Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing. *Journal of Environmental Quality*, 27(5), 1078–1086. <https://doi.org/10.2134/jeq1998.00472425002700050013x>
- Guo, L., Du, S., Haining, R., & Zhang, L. (2012). Global and local indicators of spatial association between points and polygons: A study of land use change. *International Journal of Applied Earth Observation and Geoinformation*, 21(1), 384–396. <https://doi.org/10.1016/j.jag.2011.11.003>
- Han, T., Jiang, D., Zhao, Q., Wang, L., & Yin, K. (2018). Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*, 40(8), 2681–2693. <https://doi.org/10.1177/0142331217708242>
- He, T., Zhang, S., Xin, J., Zhao, P., Wu, J., Xian, X., ... Cui, Z. (2014). An active learning approach with uncertainty, representativeness, and diversity. *Scientific World Journal*, 2014(October). <https://doi.org/10.1155/2014/827586>
- Jakomulska, A., & Clarke, K. C. (2001). Variogram-Derived Measures of Textural Image Classification, 345–355. [https://doi.org/10.1007/978-94-010-0810-5\\_30](https://doi.org/10.1007/978-94-010-0810-5_30)
- Khan, M. R., de Bie, C. A. J. M., van Keulen, H., Smaling, E. M. A., & Real, R. (2010). Disaggregating and mapping crop statistics using hypertemporal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 12(1), 36–46. <https://doi.org/10.1016/j.jag.2009.09.010>
- Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote Sensing in Agriculture — Accomplishments , Limitations , and Opportunities.
- Koks, E. E., de, H., & Koome, E. (2012). Comparing Extreme Rainfall and Large-Scale Flooding Induced

- Inundation Risk – Evidence from a Dutch Case-Study. *Studies on Water Management Issues*, (May 2014), 2–26. <https://doi.org/10.5772/30378>
- Konyushkova, K., Raphael, S., & Fua, P. (2017). Learning active learning from data. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 4226–4236.
- Lee, S., & Crawford, M. M. (2005). Unsupervised multistage image classification using hierarchical clustering with a bayesian similarity measure. *IEEE Transactions on Image Processing*, 14(3), 312–320. <https://doi.org/10.1109/TIP.2004.841195>
- Li, M., & Sethi, I. K. (2006). Confidence-Based Active Learning, 28(8), 1251–1261.
- Li, N., Martin, A., & Estival, R. (2019). Combination of Supervised Learning and Unsupervised Learning Based on Object Association for Land Cover Classification. *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*. <https://doi.org/10.1109/DICTA.2018.8615871>
- Li, X., Zaïane, O. R., & Li, Z. (2006). *LNAI 4093 - Advanced Data Mining and Applications. Second International Conference, ADMA 2006*.
- Liu, A., Jun, G., & Ghosh, J. (2009). Active learning of hyperspectral data with spatially dependent label acquisition costs. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5(May 2014). <https://doi.org/10.1109/IGARSS.2009.5417684>
- Liu, Q., Xie, W. J., & Xia, J. B. (2013). Using Semivariogram and Moran's I Techniques to Evaluate Spatial Distribution of Soil Micronutrients. *Communications in Soil Science and Plant Analysis*, 44(7), 1182–1192. <https://doi.org/10.1080/00103624.2012.755999>
- Liu, X., Zhai, H., Shen, Y., Lou, B., Jiang, C., Li, T., ... Shen, G. (2020). Large-Scale Crop Mapping from Multisource Remote Sensing Images in Google Earth Engine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 414–427. <https://doi.org/10.1109/JSTARS.2019.2963539>
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823–870. <https://doi.org/10.1080/01431160600746456>
- Lu, Q., Ma, Y., & Xia, G. S. (2017). Active learning for training sample selection in remote sensing image classification using spatial information. *Remote Sensing Letters*, 8(12), 1210–1219. <https://doi.org/10.1080/2150704X.2017.1375610>
- Ma, L., Fu, T., & Li, M. (2018). Active learning for object-based image classification using predefined training objects. *International Journal of Remote Sensing*, 39(9), 2746–2765. <https://doi.org/10.1080/01431161.2018.1430398>
- Mathur, M. (2015). Spatial autocorrelation analysis in plant population: An overview. *Journal of Applied and Natural Science*, 7(1), 501–513. <https://doi.org/10.31018/jans.v7i1.639>
- Matsuoka, M., Hayasaka, T., Fukushima, Y., & Honda, Y. (2007). Land cover in East Asia classified using terra MODIS and DMSP OLS products. *International Journal of Remote Sensing*, 28(2), 221–248. <https://doi.org/10.1080/01431160600675911>
- Milan, A., Pham, T., Vijay, K., Morrison, D., Tow, A. W., Liu, L., ... Leitner, J. (2018). Semantic Segmentation from Limited Training Data. *Proceedings - IEEE International Conference on Robotics and Automation*, 1908–1915. <https://doi.org/10.1109/ICRA.2018.8461082>
- Miranda, E., Mutiara, A. B., Ernastuti, & Wibowo, W. C. (2018). Classification of Land Cover from Sentinel-2 Imagery Using Supervised Classification Technique (Preliminary Study). *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, (September), 69–74. <https://doi.org/10.1109/ICIMTech.2018.8528122>
- Mitra, P., Uma Shankar, B., & Pal, S. K. (2004). Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognition Letters*, 25(9), 1067–1074. <https://doi.org/10.1016/j.patrec.2004.03.004>
- Mohebzadeh, H. (2018). Comparison of Methods for Fitting the Theoretical Variogram to the Experimental Variogram for Estimation of Depth to Groundwater and its Temporal and Spatial Variations. *American-Eurasian Journal of Agricultural and Environmental Science*, 18(2), 64–76. <https://doi.org/10.5829/idosi.aejas.2018.64.76>
- Moran, P. (1950). Notes on Continuous Stochastic Phenomena Published by : Biometrika Trust Stable URL : <http://www.jstor.org/stable/2332142>. *Biometrika*, 37(1), 17–23.
- Munoz-Mari, J., Tuia, D., & Camps-Valls, G. (2012). Semisupervised classification of remote sensing images with active queries. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10 PART1), 3751–3763. <https://doi.org/10.1109/TGRS.2012.2185504>

- Naghdy, G. A., Todd, C., Olaode, A., & Naghdy, G. (2014). Unsupervised Classification of Images: A Review. *International Journal of Image Processing (IJIP)*, 8(5), 325–342.
- Nirbhay Bhuyar. (2020). Crop Classification with Multi-Temporal Satellite Image Data. *International Journal of Engineering Research And*, V9(06), 221–225. <https://doi.org/10.17577/ijertv9is060208>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Pasolli, E., Melgani, F., Alajlan, N., & Bazi, Y. (2012). Active learning methods for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10 PART2), 4071–4084. <https://doi.org/10.1109/TGRS.2012.2187906>
- Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., & Emery, W. J. (2011). Improving active learning methods using spatial information. *International Geoscience and Remote Sensing Symposium (IGARSS)*, (May 2014), 3923–3926. <https://doi.org/10.1109/IGARSS.2011.6050089>
- Patra, S., Bhardwaj, K., & Bruzzone, L. (2017). A spectral-spatial multicriteria active learning technique for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12), 5213–5227. <https://doi.org/10.1109/JSTARS.2017.2747600>
- Pham, T. D. (2016). The Semi-Variogram and Spectral Distortion Measures for Image Texture Retrieval. *IEEE Transactions on Image Processing*, 25(4), 1556–1565. <https://doi.org/10.1109/TIP.2016.2526902>
- R . C . Geary. (1954). The Contiguity Ratio and Statistical Mapping Author ( s ): R . C . Geary Source : The Incorporated Statistician , Vol . 5 , No . 3 ( Nov . , 1954 ) , pp . 115-127 + 129-146 Published by : Wiley for the Royal Statistical Society Stable URL : <http://www.jstor>. *Royal Statistical Statistician, The Incorporated*, 5(3), 115–127.
- Radhika, K. (2016). A tutorial on classification of remote sensing data 1, (August), 881–885.
- Rajan, S., Ghosh, J., & Crawford, M. M. (2008). An active learning approach to hyperspectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4), 1231–1242. <https://doi.org/10.1109/TGRS.2007.910220>
- Richter, G. M., Agostini, F., Barker, A., Costomiris, D., & Qi, A. (2016). Assessing on-farm productivity of Miscanthus crops by combining soil mapping, yield modelling and remote sensing. *Biomass and Bioenergy*, 85, 252–261. <https://doi.org/10.1016/j.biombioe.2015.12.024>
- Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third ERTS Symposium*.
- Settles, B. (2009). Computer Sciences Active Learning Literature Survey, (January).
- Seung, H. S., Opper, M., & Sompolinsky, H. (1992). Query by committee. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, (January), 287–294. <https://doi.org/10.1145/130385.130417>
- Shanmugapriya, P., Rathika, S., Ramesh, T., & Janaki, P. (2019). Applications of Remote Sensing in Agriculture - A Review. *International Journal of Current Microbiology and Applied Sciences*, 8(01), 2270–2283. <https://doi.org/10.20546/ijcmas.2019.801.238>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(4), 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Shetty, S., Gupta, P. K., Belgiu, M., & Srivastav, S. K. (2021). Assessing the effect of training sampling design on the performance of machine learning classifiers for land cover mapping using multi-temporal remote sensing data and google earth engine. *Remote Sensing*, 13(8). <https://doi.org/10.3390/rs13081433>
- Srivastava, V., Stein, A., Rossiter, D. G., Garg, P. K., & Garg, R. D. (2016). Simulated Annealing with Variogram-Based Optimization to Quantify Spatial Patterns of Trees Extracted from High-Resolution Images. *IEEE Geoscience and Remote Sensing Letters*, 13(8), 1084–1088. <https://doi.org/10.1109/LGRS.2016.2565743>
- Stumpf, A., Lachiche, N., Malet, J. P., Kerle, N., & Puissant, A. (2014). Active learning in the spatial domain for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), 2492–2507. <https://doi.org/10.1109/TGRS.2013.2262052>
- Sugiyama, M., & Nakajima, S. (2009). Pool-based active learning in approximate linear regression. *Machine Learning*, 75(3), 249–274. <https://doi.org/10.1007/s10994-009-5100-3>
- Tobler, R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region Author ( s ): W . R . Tobler Source : Economic Geography , Jun . , 1970 , Vol . 46 , Supplement : Proceedings . International Geographical Union . Commission on Quantitative Methods ( Jun . , 19. *Economic Geography*, 46, 234–240.
- Tonye, E., Fotsing, J., Essimbi Zobo, B., Talla Tankam, N., Kanaa, T. F. N., & Rudant, J. P. (2011).

- Contribution of variogram and feature vector of texture for the classification of big size SAR images. *Proceedings - 7th International Conference on Signal Image Technology and Internet-Based Systems, SITIS 2011*, (March), 382–389. <https://doi.org/10.1109/SITIS.2011.67>
- Treitz, P., & Rogan, J. (2004). Remote sensing for mapping and monitoring land-cover and land-use change-an introduction. *Progress in Planning*, 61(4), 269–279. [https://doi.org/10.1016/S0305-9006\(03\)00064-3](https://doi.org/10.1016/S0305-9006(03)00064-3)
- Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., & Emery, W. J. (2009). *Active learning methods for remote sensing image classification*. *IEEE Transactions on Geoscience and Remote Sensing* (Vol. 47). <https://doi.org/10.1109/TGRS.2008.2010404>
- USGS. (n.d.). NDVI, the Foundation for Remote Sensing Phenology. Retrieved from [https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science_center_objects=0#qt-science_center_objects)
- Vlachos, A. (2008). A stopping criterion for active learning. *Computer Speech and Language*, 22(3), 295–312. <https://doi.org/10.1016/j.csl.2007.12.001>
- Volpi, M., Tuia, D., & Kanevski, M. (2012). Memory-based cluster sampling for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8), 3096–3106. <https://doi.org/10.1109/TGRS.2011.2179661>
- Willis, C. J. (2004). Hyperspectral image classification with limited training data samples using feature subspaces. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, 5425(August 2004), 170. <https://doi.org/10.1117/12.543622>
- Wu, S. H., Ho, C. T., Nah, S. L., & Chau, C. F. (2014). Global Hunger: A Challenge to Agricultural, Food, and Nutritional Sciences. *Critical Reviews in Food Science and Nutrition*, 54(2), 151–162. <https://doi.org/10.1080/10408398.2011.578764>
- Xie, H., Tian, Y. Q., Granillo, J. A., & Keller, G. R. (2007). Suitable remote sensing method and data for mapping and measuring active crop fields. *International Journal of Remote Sensing*, 28(2), 395–411. <https://doi.org/10.1080/01431160600702673>
- Xiong, B., Zhang, X. J., & Jiang, W. S. (2010). Semi-Supervised Classification based on Gaussian Mixture Model for remote imagery. *Science China Technological Sciences*, 53(1 SUPPL.), 85–90. <https://doi.org/10.1007/s11431-010-3211-5>
- Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., ... Thau, D. (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>
- Xue, Z., Zhou, S., & Zhao, P. (2018). Active Learning Improved by Neighborhoods and Superpixels for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 15(3), 469–473. <https://doi.org/10.1109/LGRS.2018.2794980>

# APPENDIX A

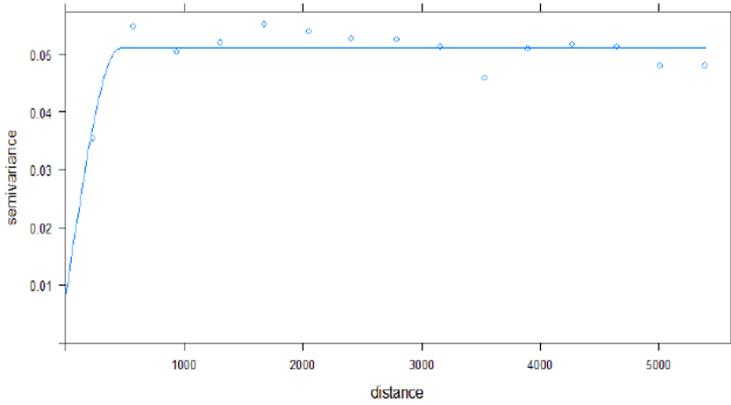
## Data Management Plan

Title of the thesis topic	Development of a spatially explicit Active Learning (AL) method for crop areas mapping from satellite image time series.
Date of the plan	11/11/2020
Brief description of the topic.	This topic seeks to develop an AL algorithm that will include the spatial domain precisely that will be used to perform mapping of crop areas using satellite images taken at different times
Name of researcher	Beatrice Anthony Kaijage
Sources of data	<ul style="list-style-type: none"> <li>• Nationaal Georegister - Ground truth data (Crop data) <a href="https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/dd8e0fb8-0f09-40ba-a884-7e23c0680ae2">https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/dd8e0fb8-0f09-40ba-a884-7e23c0680ae2</a></li> <li>• Sentinel 2A images- European Space Agency (ESA)- <a href="https://scihub.copernicus.eu/dhus/#/home">https://scihub.copernicus.eu/dhus/#/home</a></li> <li>• IGISMAP(Netherlands boundary shapefile) <a href="https://map.igismap.com/share-map/export-layer/Netherlandsshapefile30/f718499c1c8cef6730f9fd03c8125cab">https://map.igismap.com/share-map/export-layer/Netherlandsshapefile30/f718499c1c8cef6730f9fd03c8125cab</a></li> </ul>
Organizations that own the data to be used	Nationaal Georegister Netherlands (NGR) European Space Agency (ESA) IGISMAP Basisregistratie Parcelen (BRP) of the Netherlands Enterprise Agency
How the data is to be used	For implementing the proposed AL thesis topic.

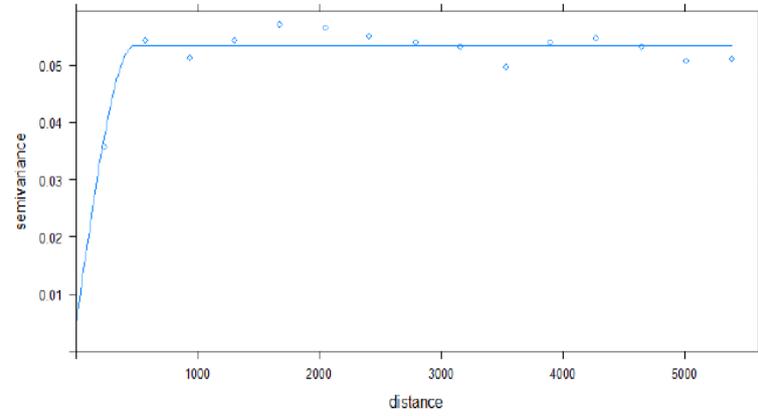
# APPENDIX B

## Variograms of the monthly NDVI images

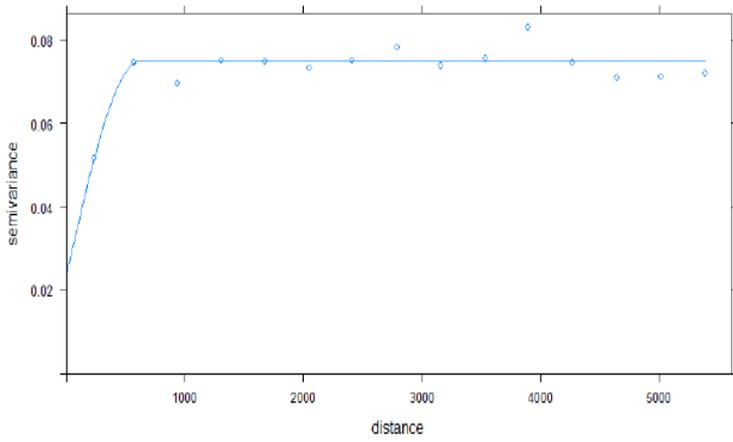
January



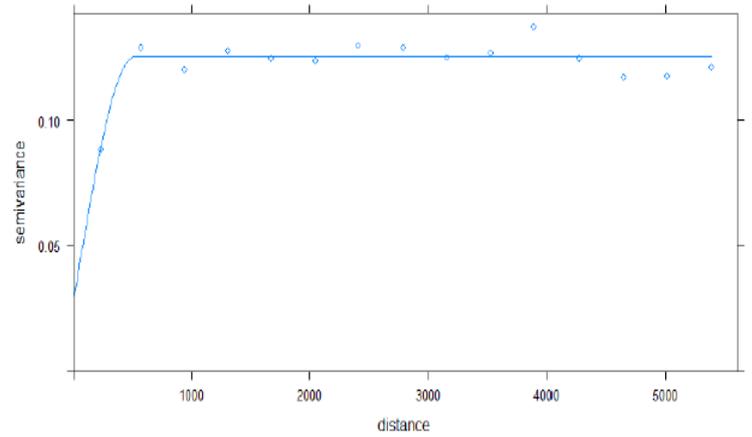
February



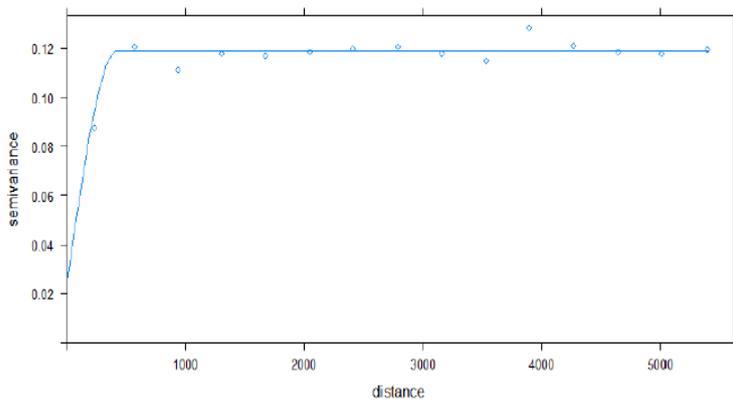
March



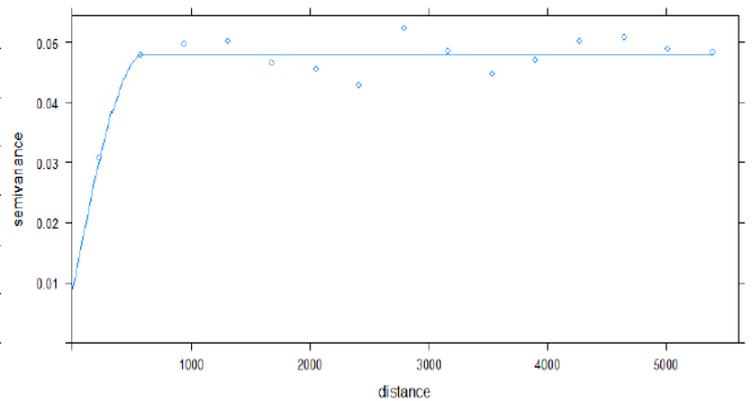
April



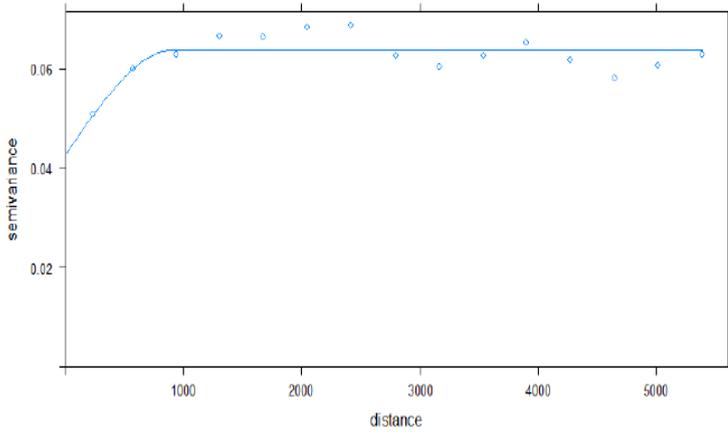
May



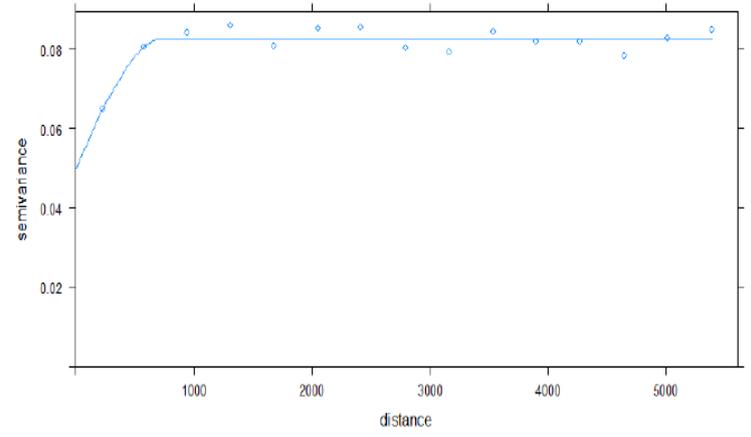
June



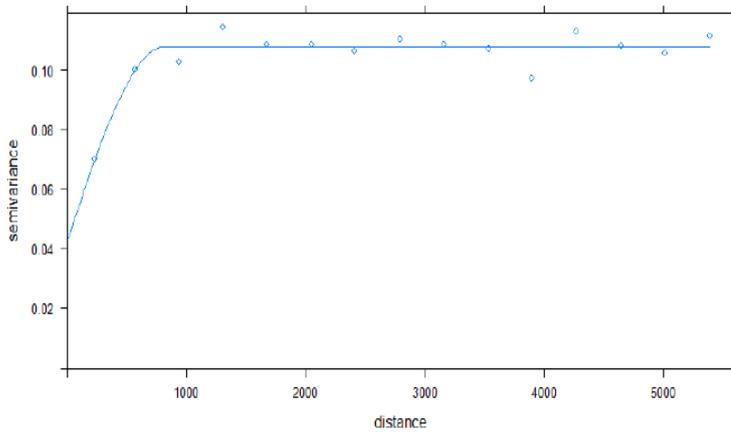
July



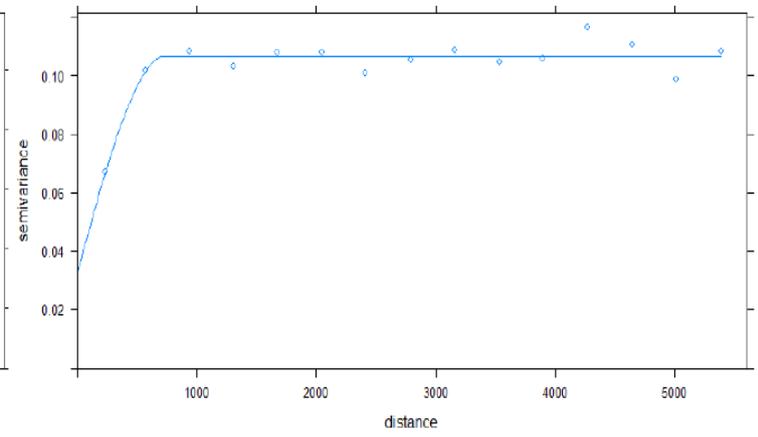
August



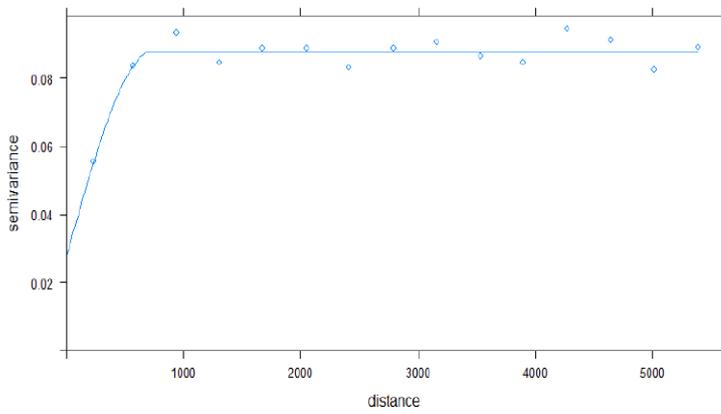
September



October



November



December

