

# **Integrating Semi-Supervised Learning and Expert System for wetland vegetation classification using Sentinel-2 data**

NASIR FARSAD LAYEGH

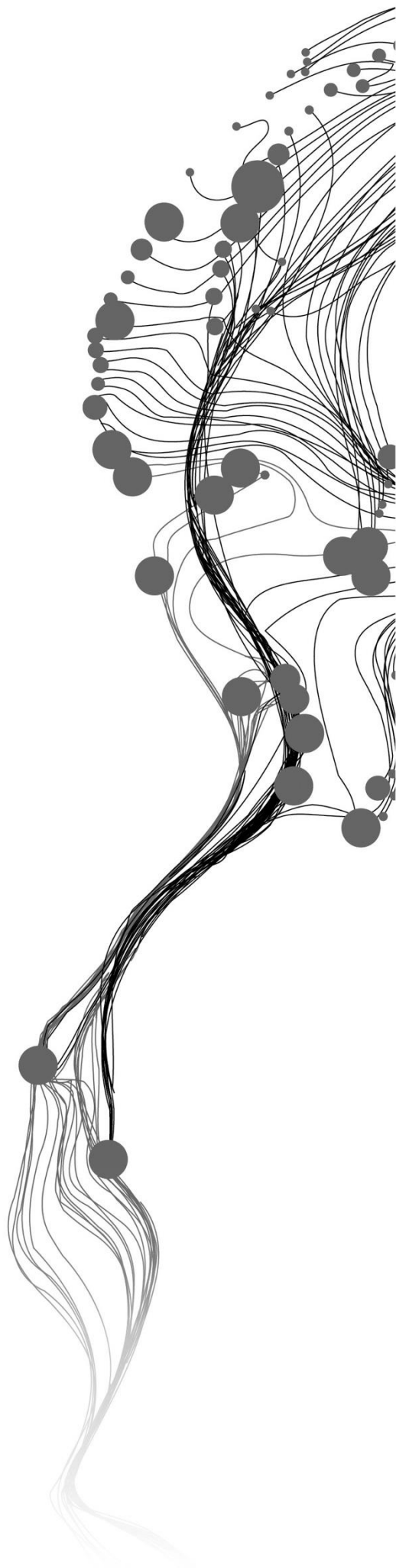
June, 2017

SUPERVISORS:

Dr. R. Darvishzadeh

Prof. Dr. A.K. Skidmore





# **Integrating semi-supervised learning and expert system for wetland vegetation classification using Sentinel 2 data**

**NASIR FARSAD LAYEGH**

Enschede, The Netherlands, June, 2017

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: Geo-information Science and Earth Observation for Environmental Modelling and Management

## **SUPERVISORS:**

Dr. R. Darvishzadeh

Prof. Dr. A.K. Skidmore

## **THESIS ASSESSMENT BOARD:**

Prof.Dr. A.D. Nelson (Chair)

Dr. C. Persello (External Examiner, University of Twente)

---

#### 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

In complex classification tasks, such as vegetation classification, the high similarity between different vegetation classes and the low similarity within a vegetation class can confuse the classification algorithm in assigning correct class labels to unlabeled samples. Therefore, the main objective of this study is to design and implement a new classification methodology by integrating expert system and semi-supervised learning (SSLES) for classification of wetland vegetation cover using Sentinel-2 data. Furthermore, the potential of different Sentinel-2 spectral band combinations for the classification of vegetation types are investigated.

The vegetation classes of Schiermonnikoog island, the Netherlands, were identified utilizing Sentinel-2 data of 17 July 2016, RapidEye data of 18th July 2015 as well as an existing vegetation map belonging to 2010 and ancillary field data of the study site (the latter were considered as the expert knowledge).

The proposed approach consists of three main steps: object-based image analysis (OBIA), SSLES and post-classification. Using OBIA the satellite image is segmented to generate image objects and extract the image features. Using SSLES training samples are increased by labeling some of the most certain unlabeled samples and finally, in post-classification, classification is performed on the training sets.

To evaluate the performance of SSLES, its obtained classification accuracy was compared to the one obtained from a standard supervised classifier (i.e. Random Forest), in terms of overall accuracy. To validate the results of both methods, a test set extracted from reference data, i.e. vegetation map, was used. At the end, McNemar's significance test was performed to compare the obtained results of both methods.

The obtained results revealed that using our proposed method, the classification performance of a standard supervised method can improve significantly. Moreover, further analysis of the obtained results from our proposed method showed that the accuracy increased without hurting the per-class classification accuracies of the vegetation classes. We observed that the combinations of all Red-edge spectral bands of Sentinel-2 yielded the highest classification accuracy for the wetlands vegetation cover classification with the overall accuracy of 83.6%.

In a future study, the potential applicability of the trained model in the proposed approach, for different wetlands or different biomes such as forest and croplands utilizing different remote sensing data could be examined. Furthermore, to increase the efficiency of SSLES, different strategies for similarity measurement and graph-construction should be investigated.

**Keywords:** Semi-supervised learning, Expert system, Object-based image analysis, Sentinel-2, wetland, vegetation cover

---

## ACKNOWLEDGEMENT

First, I would like to thank for the opportunity and financial support of European Union under Erasmus plus scholarship that made it possible for me to experience life and study at Lund University and ITC and University of Twente.

I also want to thank the staff of GIS center and Geo Centrum, Lund University for making my first year a great experience, with their motivation and dedication to work. I would like to thank to Prof. Dr. Petter Pilesjö for a warm welcome in Lund, greatly appreciated help and advice in practical challenges. I would like to thank ITC as well for organizing GEM program.

I give my gratitude to Dr. Roshanak Darvishzadeh for her help and supports during my stay in ITC. A special thanks to Prof. Dr. Andrew Skidmore for the discussions and ideas in this thesis and it was a valuable experience for me to learn from him. I also thank Nina Amiri for helping me in developing the ideas of this thesis.

And at the end, I want to thank the Karate communities in Lund and Enschede for keeping me strong and balanced.

I want to dedicate this work to my family and friends for their support and encouragement.

---

# Table of Contents

---

<b>1. Introduction .....</b>	<b>1</b>
1.1. Background .....	1
1.2. Overview of classification methods.....	1
1.3. Overview of image analysis approaches for classification.....	2
1.4. Challenges with existing classification approaches.....	2
1.5. Semi-supervised learning (SSL) methods.....	3
1.6. Problem statement and justification .....	4
1.7. Aim and Research objectives.....	5
1.8. Research questions.....	5
1.9. Hypotheses .....	6
<b>2. Study area and materials.....</b>	<b>7</b>
2.1. Study area.....	7
2.2. Materials.....	8
2.3. Reference data and sampling .....	9
<b>3. Methods .....</b>	<b>11</b>
3.1. Object-based image analysis .....	12
3.1.1. Image segmentation .....	12
3.1.2. Image segmentation evaluation .....	13
3.1.3. Feature extraction.....	13
3.1.4. Feature selection .....	14
3.2. Semi-Supervised Learning and Expert System (SSLES) .....	17
3.2.1. Graph-based Semi-supervised learning.....	17
3.2.2. Expert system .....	19
3.2.3. SSLES algorithm .....	20
3.3. Post-classification and evaluation .....	23
3.3.1. Post-classification .....	23
3.3.2. Classification evaluation .....	24
<b>4. Results.....</b>	<b>26</b>
4.1. Data pre-processing/data preparation .....	26

---

4.1.1. Image segmentation.....	26
4.1.2. Feature extraction .....	29
4.2.3. Feature selection.....	30
4.2. Expert system rules and <i>a priori</i> probabilities.....	32
4.3. Model selection (Parameter optimization) .....	35
4.3.1. SSL parameters.....	35
4.3.2. Classification parameters .....	36
4.4. Model performance evaluation .....	37
4.5. Result of McNemar significance test .....	38
4.6. SSLES performance.....	39
<b>5. Discussion.....</b>	<b>42</b>
5.1. Assessment of SSLES .....	42
5.2. Object-based image analysis .....	42
5.3. Expert system design.....	43
5.4. Performance and drawbacks of SSLES .....	44
5.4.1. Classification accuracy.....	44
5.4.2. SSLES drawbacks.....	45
5.5. The scalability of SSLES .....	47
<b>6. Conclusion and recommendations .....</b>	<b>49</b>
6.1. Conclusion .....	49
6.2. Recommendations .....	49
<b>7. List of references.....</b>	<b>50</b>



## LIST OF FIGURES

---

Figure 2. 1. Location of Schiermonnikoog national park .....	7
Figure 3. 1. Procedure of methodology in this study.....	11
Figure 3. 2. Main architecture of OBIA in this study .....	16
Figure 3. 3. Illustration of label propagation procedure in this study .....	19
Figure 3. 4. SSLES algorithm flowchart .....	22
Figure 3. 5. Thematic representation of decision boundary changes .....	23
Figure 4. 1. Results of segmentation evaluation process.....	27
Figure 4. 2. Final segmentation result .....	28
Figure 4. 3. Reference samples among the image objects .....	29
Figure 4. 4. Feature selection results for band groups.....	32
Figure 4. 5. Expert rules weights for ancillary data. ....	33
Figure 4. 6. Analysis of SSL parameters.....	35
Figure 4. 7. Analysis of ntree parameter of RF classifier.....	36
Figure 4. 8. Classified map of the study area.....	38
Figure 4. 9. Producer's accuracy matrices.....	40
Figure 4. 10. Detailed comparison of two classification approaches. ....	41
Figure 5. 1. Example of irregular graph in 1-NN case .....	46
Figure 5. 2. User's accuracy map derived using SSLES.....	46
Figure 5. 3. Mean of features distance between 10 vegetation classes. ....	47

## LIST OF TABLES

---

Table 2. 1. Vegetation classes in Schiermonnikoog Island .....	10
Table 4. 1. Number of image objects for three different subsets .....	28
Table 4. 2. Extracted features and their description .....	29
Table 4. 3. Number of extracted features for different groups of band combinations .....	30
Table 4. 4. A priori probabilities estimation of the vegetation classes .....	33
Table 4. 5. Expert rules based on the distance of classes to streams and residential areas .....	34
Table 4. 6. Adjacency matrix of vegetation types .....	35
Table 4. 7. Classification results for six groups of band combinations .....	37
Table 4. 8. Statistically significance tests result .....	38
Table 5. 1. McNemar test for two cases of using all and 50% of training samples .....	45

# 1. INTRODUCTION

## 1.1. Background

Accurate assessment of vegetation condition is essential for ecosystem management and preservation of biological diversity (Keith et al., 2011). In other words, retrieving the current state of vegetation cover in an ecosystem can help to initiate vegetation protection and restoration programs efficiently (Egbert et al., 2002). Therefore, it is necessary to have accurate and up-to-date information about the status of the vegetation cover in the ecosystem through regular monitoring.

Among terrestrial ecosystems, wetlands play a key role in providing essential ecological services such as wildlife habitat, groundwater recharge, flood control sediment filtration and pollutant removal (Rundquist et al., 2001). Hence, monitoring the vegetation changes in a continuous and semi-automatic manner is vital for their management to improve the effectiveness of management strategies (Lindenmayer & Likens, 2009). Traditional field inventory monitoring methods to assess the condition and distribution of wetland vegetation cover are expensive (Xie et al., 2008). A viable approach is the use of satellite remote sensing data, which provides the advantages of large area coverage, ongoing data collection, improved spatial resolution and potentially cost-effective for wetland monitoring and mapping purposes (Immitzer et al., 2016; Mui et al., 2015).

Using remote sensing images, which contain reflectance values of the observed surfaces as a function of wavelength by a sensor, allows discriminating the different species based on their spectral behavior (Landgrebe, 2002). The advent of remote sensing technologies with a new generation of high-resolution multi-spectral images, such as Sentinel-2, which covers a wide spectral range (400 nm - 2400 nm) and includes red edge region, has provided new opportunities to monitor and model the vegetation surface at different spatial and temporal scales and may allow discrimination of different vegetation types. A common approach to distinguish different vegetation types using remote sensing data is using image classification (Richards, 2013).

## 1.2. Overview of classification methods

Image classification, in remote sensing, is defined as the process of generating spatially explicit generalizations that represent individual classes (Franklin, 2001), i.e. vegetation classes in the context of this study, such as high shrub. There are two main categories of classification algorithms, unsupervised and supervised algorithms (Richards, 2013). In the unsupervised methods, unlabeled samples, i.e. data samples that their actual class on the ground is unclear, is used to describe the hidden structure of the dataset because detailed knowledge such as ground truth data is not available. These methods can be used to group the data into clusters, based on their observable features (Karakos et al., 2005; Patrick & Costello, 1970). Nonparametric Bayesian classification (John & Langley, 1995), fuzzy c-means (Ahmed et al., 2002) and ISODATA (Iterative Self-Organizing Data Analysis Technique) (Dunn, 1973; Repaka et al., 2004)

are three major unsupervised classification algorithms. In the supervised algorithms, true class labels of training data are clear and they are being exploited to train the classifier and find the mapping function (Anuta, 1977). Some of the common supervised classification algorithms are: Random Forest (Breiman et al., 1984; Breiman, 1999; Breiman, 2001), Linear Discriminant Analysis (Cochran, 1964; Klecka, 1980), Maximum Likelihood methods (Cam, 1990), Nearest neighborhood algorithm (Cover & Hart, 1967), Neural networks (Bishop, 1996) and Support Vector Machine (Cristianini & Shawe-Taylor, 2000).

### **1.3. Overview of image analysis approaches for classification**

In the literature, there are two general image analysis approaches, utilizing supervised algorithms, for the classification of vegetation types from remote sensing data, i.e., Pixel-based Image Analysis (PBIA) and Object-based Image Analysis (OBIA). PBIA techniques are mainly statistically oriented which consider the spectral properties of every pixel within the area of interest (Weih & Riggan, 2010) without taking into account the spatial context of the pixels. The main problems when using these techniques are, shadowed or noisy pixels (Dalponte et al., 2014), low classification accuracy due to the so-called “salt-and-pepper” problem (Ouyang et al., 2011), as well as the spectral variability (Peña-Barragán et al., 2011). An alternative approach to be able to overcome these problems is to use object-based techniques (Blaschke & Strobl, 2001). The main goal of the OBIA is segmenting the image and constructing a hierarchical network of homogeneous objects (Devadas et al., 2012) by integrating the spectral properties and spatial or contextual information of pixels within an object in the classification process (Blaschke, 2010). Assigning all the pixels in a defined object as a specific class can solve the mentioned common problems of the PBIA. Furthermore, previous studies have demonstrated that object-based classification approaches typically out-performs pixel-based approaches in terms of classification accuracy (Fu et al., 2017; Mui et al., 2015; Weih & Riggan, 2010).

### **1.4. Challenges with existing classification approaches**

Although using a supervised algorithm exploiting OBIA for classification can improve the results in terms of classification accuracy, there are still some challenges in classification that need to be considered. One of these challenges lies in defining the appropriate classification algorithm (Blaschke, 2010). An appropriate classification algorithm should have the following characteristics:

- high generalization ability and classification accuracies with respect to other algorithms (Chapelle et al., 2010);
- convexity of the cost function which always allows one to reach the optimal solution (Álvarez-Meza et al., 2016);
- effectiveness in addressing ill-posed problems (i.e. low number of high dimensional labeled sample) (Wang et al., 2016a; Wang et al., 2016b)

- ability to use an unbalanced number of a training set, which is common in vegetation classification that includes a small proportion of labeled samples belong to the less frequent classes (Mellor et al., 2015)
- ability to use a lower number of training samples to train the classifier, as collecting a sufficient number of samples requires time and expensive fieldworks (Mellor et al., 2015).

Otherwise, often it becomes difficult for the traditional classifiers to offer a satisfactory performance. A relevant advanced solution to the aforementioned classification challenges is the introduction of semi-supervised learning (SSL) techniques (Board & Pitt, 1989). The semi-supervised learning is a type of supervised learning technique that combines a few labeled samples with many unlabeled samples to perform classification (Chapelle et al., 2010). The main idea of this technique is to exploit the structural information of unlabeled samples, as the available and inexpensive source of information, in the feature space to improve the decision boundaries and find a more accurate classification rule than using only labeled samples (Dalponte et al., 2015a; Persello & Bruzzone, 2014).

### 1.5. Semi-supervised learning (SSL) methods

In the past few years, several paradigms of SSL have been proposed for classification of remote sensing data which can be divided into 4 major categories: (1) Generative mixture models, (2) Low-density separation algorithms (3) Self-learning approaches, and (4) graph-based methods.

1) *Generative mixture models*: This method is based on the estimation of the joint probability  $P(x, y | \theta)$  assuming a particular model for the data (e.g., Gaussian mixture model), where  $\theta$  is the parameter vector of the model that should be estimated from the observations. The parameter  $\theta$  can be estimated by joint exploitation of both labeled and unlabeled samples. Then the classification can be performed based on Bayes' rule. Expectation-Maximization (EM) algorithm is the instance of a method to estimate the parameter  $\theta$ . For remote sensing image classification, Shahshahani & Landgrebe (1994) proved the positive effect of the unlabeled samples for classification in the context of a Gaussian Maximum-Likelihood classifier. Tadjudin & Landgrebe (2000) used unlabeled samples to update the parameters of Maximum Likelihood classifier by means of the EM algorithm. Jackson & Landgrebe (2001) showed that information contained in semi-labeled (i.e. labeled by classifier) samples of two Gaussian distributions in terms of the Fishers Information Matrix can improve the classification accuracy.

2) *Low density separation algorithms*: The aim of these algorithms is to push the decision boundary away from the unlabeled data. The most common approach is using maximum margin algorithms such as support vector machines (SVMs). The method of maximizing the margin for unlabeled as well as labeled data is called the Transductive SVM (TSVM) (Chapelle et al., 2010). The main idea of this method is that the decision boundary has to pass in low density regions, and this is obtained by adding an additional regularization term on the unlabeled data to the standard SVM optimization problem. In the field of

remote sensing, Bruzzone et al. (2006) proposed a TSVM method by utilizing unlabeled samples to address ill-posed problems and Dalponte et al. (2015) implemented a Semi-supervised SVM (S3SVM) approach for individual tree crown classification.

3) *Self-learning*: This method is known as one the earliest idea about using the unlabeled data in the classification. “Self-training”, “self-labeling” and “decision-directed learning” are the other terms that refer to this approach. This method uses a supervised algorithm repeatedly. It starts by training the classifier using labeled data only. At each step, a part of the unlabeled data is being classified according to the current decision function and the most confident unlabeled data and their corresponding class labels are being added to the training set. Then the classifier is being trained using its own prediction data, as additional labeled data (Agrawala, 1970; Chapelle et al., 2010; Fralick, 1967; Scudder, 1965). As this method uses its own prediction so there can be reinforcement classification error problem. This problem can be avoided by choosing an appropriate model and “unlearning” those unlabeled points if the prediction confidence drops below a threshold (Maulik & Chakraborty, 2011). For remote sensing data classification, Maulik & Chakraborty (2011) and Dopido et al. (2013) applied a self-learning method for the classification of multi-spectral and hyper-spectral satellite images, respectively.

4) *Graph-based methods*: This method defines a graph, where the labeled and unlabeled samples are the nodes and the connecting edge between nodes reflects the similarity between them. Each labeled sample spreads its label information to its connected unlabeled neighbor samples until a global stable condition is achieved. For remote sensing data classification, Camps-Valls et al. (2007) proposed a graph-based SSL for classification of hyper-spectral data by introducing composite kernel framework to extract contextual information from data. Gu & Feng (2012) implemented a graph-based approach for hyper-spectral data classification in which they generated the graph weights by solving a L1 optimization problem. Ma et al. (2015) designed a spectral-spatial regularized graph-based SSL for hyper-spectral data classification by introducing a new method for similarity measurement and a spatial regularizer.

## 1.6. Problem statement and justification

Among the SSL methods, graph-based approaches have recently received significant attention due to their ability to provide a relatively high classification accuracy while retaining computational simplicity (Kim & Choi, 2014; Ma et al., 2015; Zhu et al., 2003). Although the graph-based algorithms can improve the classification performance by the use of the distribution of the unlabeled samples, they suffer from some limitations (Bruzzone et al., 2006; Kim & Crawford, 2010). One of these limitations occurs in complex classification tasks such as vegetation types where two samples from same vegetation class may show different characteristics (i.e. low similarity) and two samples from two different vegetation classes show identical characteristics (i.e. high similarity). This similarity problem can confuse the graph-based algorithm in assigning correct class labels to unlabeled samples. In this case, unlabeled samples



can be detrimental for the classification as they may degrade the accuracy by misguiding the classifier (Chapelle et al., 2010).

The contribution of “Expert Knowledge” in the labeling process of the graph-based algorithm may help to solve the aforementioned similarity problem. In this context, the expert knowledge is considered as experience and existing knowledge of the expert in the specific domains of study, technical practices and prior information on the study area (Booker & McNamara, 2004; Hayes-Roth, Waterman, & Lenat, 1983). As such to tackle the similarity problem, an Expert System (ES) integrated with an SSL algorithm may increase the certainty of labeling of the samples. The ES, should have the ability to classify the unlabeled samples, selected by SSL, and assign the most probable class label to them. ESs are artificial intelligence based systems that are able to exploit the domain experts’ knowledge as well as the experience and use that for decision making (Hayes-Roth et al., 1983). The proposed ES, in this study, should have the ability to classify the unlabeled samples, selected by SSL, and assign the most probable class label to them.

### **1.7. Aim and Research objectives**



Motivated by above insights, in this study, a classification methodology is proposed for the classification of wetland vegetation cover using Sentinel-2 data by integrating Semi-Supervised Learning and Expert System (SSLES). The main idea of the proposed approach is to construct a graph based on image features, derived from OBIA, and use ES in the labeling process of SSL to assign the most probable class labels to the selected unlabeled samples.

In this study, two objectives are defined to address the proposed approach using Sentinel-2 data, as follows:

-  To investigate the performance of SSLES for wetland vegetation classification by comparing its obtained accuracy with the one obtained from a standard supervised classification, for the same condition.
-  To find the most informative spectral group of band combinations of Sentinel-2 data which improves the vegetation classification in terms of overall classification accuracy.

### **1.8. Research questions**

Based on the objectives above, following research questions can be derived:

-  Is there a statistically significant difference in the classification accuracy for vegetation classification between the SSLES and the standard supervised classifiers?
-  Which group of band combination of Sentinel-2 data can provide the highest accuracy (in terms of overall classification accuracy) for the wetland vegetation classification?

## 1.9. Hypotheses

Hypothesis 1:

- ✚ There is a statistically significant difference between the classification accuracy of the SSLES and the standard supervised method, e.g. RF, according to McNemar's significance test.

Hypothesis 2:

- ✚ Red edge band combinations of Sentinel-2 data can provide the highest accuracy (in terms of overall classification accuracy) for wetland vegetation cover discrimination.



## 2. STUDY AREA AND MATERIALS

### 2.1. Study area

The study area is the Schiermonnikoog island which is one of the barrier islands surrounded by the North Sea from the north and Wadden Sea from the south, located between 53° 27' 20" N - 53° 30' 40" N latitude and 06° 06' 35" E – 06° 20' 56" E longitude (Figure 2. 1). The island originally belongs to the Dutch province of Friesland. The main part of the island is natural landscapes including dunes, beaches and polders. The vegetation cover in the south and south-east shore of the island has adapted to the regular inundation sea water and has formed the salt marsh (Schmidt & Skidmore, 2003).

There are around 120 different vegetation species growing in the salt marsh and among them, the 10 most dominant species were considered in this study for classification, namely: High matted grass, Low matted grass, Agriculture, Forest, Green beach, Tussock grass, High shrub, herbs, Low salix shrub, Low hippophae shrub. The natural vegetation cover has a large spatial and temporal variability, due to the dynamic influences of tide, wind, and grazing (Vrieling et al., 2017).



Figure 2. 1. Location of Schiermonnikoog national park. (Source: <http://www.np-schiermonnikoog.nl/>)

## 2.2. Materials

### 2.2.1 Sentinel-2 data

Two different satellite data were used in this study. The first satellite imagery used in this study is the standard Sentinel-2 Level-1C product, which is in UTM/WGS84 projection and its per-pixel radiometric measurements are provided in Top of Atmosphere (TOA) reflectance (ESA, 2015). The Sentinel-2 image of the study area is acquired on 17 July 2016 belonging to the relative orbit of R008 and was downloaded from ESA Sentinel-2 Pre-operation Hub (<https://scihub.copernicus.eu/>). The atmospheric correction of the image was performed using Sen2Cor software (ESA, 2015) and the top of canopy reflectance (TOC) was calculated for further analysis.

Sentinel-2 offers a multispectral sensor in 13 bands from 443 to 2190 nm with three different geometric resolutions: 60m, 20 m and 10m. The 10 m resolution spectral bands are: Blue (B, 490 nm), Green (G, 560 nm), Red (R, 665 nm) and Near Infrared (NIR, 842 nm). Four red edge/NIR bands with central wavelength at 705 nm, 740 nm, 783 nm and 865 nm, respectively, short wave infrared-1 (SWIR1, 1610 nm) and short wave infrared-2 (SWIR2, 2190 nm) bands have 20 m resolution. The other bands (which were not utilized in our study), are coastal (C, 443 nm), water vapor (WV, 1375 nm) and cirrus (CI, 1376) and have a spatial resolution of 60 m (Novelli et al., 2016).

In order to achieve higher classification accuracy and also assess the capability of Sentinel-2 data in classifying the vegetation types, we combined the spectral bands into different groups (combinations) to be assessed. Based on the literature, the most important regions of the spectrum to study vegetation cover are: Red-edge, Shortwave infrared and Red-infrared regions (Mui et al., 2015; Tigges et al., 2013; Delegido et al., 2011; Darvishzadeh et al., 2009; Gilmore et al., 2008; Tucker, 1979). Consequently, six groups of band combinations of the Sentinel-2 spectral data are considered in this study to classify the wetland vegetation cover, as follows:

- (1) All spectral bands
- (2) Red and Infrared bands
- (3) All Red-edge bands
- (4) All shortwave infrared bands
- (5) Red, Infrared and Red-edge bands
- (6) Red-edge and shortwave infrared bands

### 2.2.2. RapidEye data

In this study, a RapidEye image is acquired for Schiermonnikoog island on 18th July 2015. This multi-spectral imagery covers five spectral bands of blue (0.44–0.51  $\mu\text{m}$ ), green (0.52–0.59  $\mu\text{m}$ ), red (0.63–0.685  $\mu\text{m}$ ), red-edge (0.69–0.73  $\mu\text{m}$ ), and near-infrared (0.76–0.85  $\mu\text{m}$ ) with a spatial resolution of 5 m. The pre-processed data was obtained at level 3A, which means that the radiometric, geometric corrections and geo-referencing were applied. The image covers 25 km  $\times$  25 km with the orthorectified pixel size of 5 m  $\times$  5 m. As the weather condition was clear during the image acquisition, no further atmospheric correction was applied.

In this study, the RapidEye data was used for image segmentation of object-based image analysis, as this data has finer spatial resolution compared to Sentinel-2 data.

### 2.3. Reference data and sampling

The reference data used in this study, include field observations of dominant vegetation species for 30 vegetation plots collected in July 2015 and a vegetation map belonging to 2010 (Pranger & Tolman, 2012), which was obtained from experts' visual interpretation of 1:10.000 aerial photographs combined with extensive field inventory. These reference data are used in this study to collect training and test sets and also as expert knowledge. As this study is utilizing an object-based approach for classification, the required features (attributes) for classification need to be calculated at the object level. In this case, the sampling unit for training and validation should be an object (polygon) too (Grenier et al., 2008; Schöpfer & Lang, 2006; Tiede et al., 2006). To select the training and test samples, stratified random sampling strategy is conducted, where each vegetation class is used as a stratum. To choose representative samples, the formula developed by Cochran (1977) is used, as follows:

$$n_0 = \frac{Z^2 pq}{\varepsilon^2} \quad (2.1)$$

where  $n_0$  is the sample size,  $Z^2$  is the abscissa of the normal curve that cuts off an area  $\alpha$  at the tails ( $1 - \alpha$  equals the desired confidence level, e.g., 95%),  $\varepsilon$  is the desired level of precision,  $p$  is the estimated proportion of a stratum that is present in the study area, and  $q$  is  $1-p$ . The value for  $Z^2$  is found in statistical tables which contain the area under the normal curve. The  $p$  value is computed based on area proportion for each vegetation class with 95% of confidence level and  $\pm 5\%$  precision, the resulting sample size becomes 650 for 10 vegetation classes that were extracted from the reference data. A number of 434 more samples were collected from the reference data as independent test samples (2/3 of the number of the training samples) and were used for validation of the final classification results. Table 2.1 reports the name of vegetation classes as well as their area, area proportion and the number of samples per stratum (vegetation class).

INTEGRATING SEMI-SUPERVISED LEARNING AND EXPERT SYSTEM FOR WETLAND VEGETATION  
CLASSIFICATION USING SENTINEL-2 DATA

---

*Table 2. 1. Vegetation classes in Schiermonnikoog Island, based on coverage area and number of samples*

<i>Class name</i>	<i>Area (m<sup>2</sup>)</i>	<i>Area proportion(%)</i>	<i>Number of training samples</i>	<i>Number of test samples</i>
<b><i>High matted grass</i></b>	8286023.966	28.47	160	107
<b><i>Low matted grass</i></b>	6886990.92	23.66	142	95
<b><i>Agriculture</i></b>	2940049.402	10.1	71	47
<b><i>Forest</i></b>	2353857.364	8.09	58	39
<b><i>Green Beach</i></b>	2326664.757	7.99	58	39
<b><i>Tussock grass</i></b>	1780060.747	6.12	45	30
<b><i>High Shrub</i></b>	1774066.818	6.1	45	30
<b><i>herbs</i></b>	1376571.176	4.73	35	23
<b><i>Low Salix Shrub</i></b>	952389.7279	3.27	25	17
<b><i>Low Hippophae Shrub</i></b>	425417.5454	1.46	11	7
<b><i>Sum</i></b>			650	434

### 3. METHODS

The main architecture of the proposed classification approach in this study contains three main parts: Object-based image analysis (OBIA), Semi-Supervised Learning & Expert System (SSLES) and post-classification. Using OBIA, the satellite image will be segmented to generate image objects and related image features. Using SSLES, training samples will be increased by labeling some of the most certain unlabeled samples. In the post-classification step, final classification will be performed on the training sets. The procedure of methodology of this study is shown in Figure 3. 1.

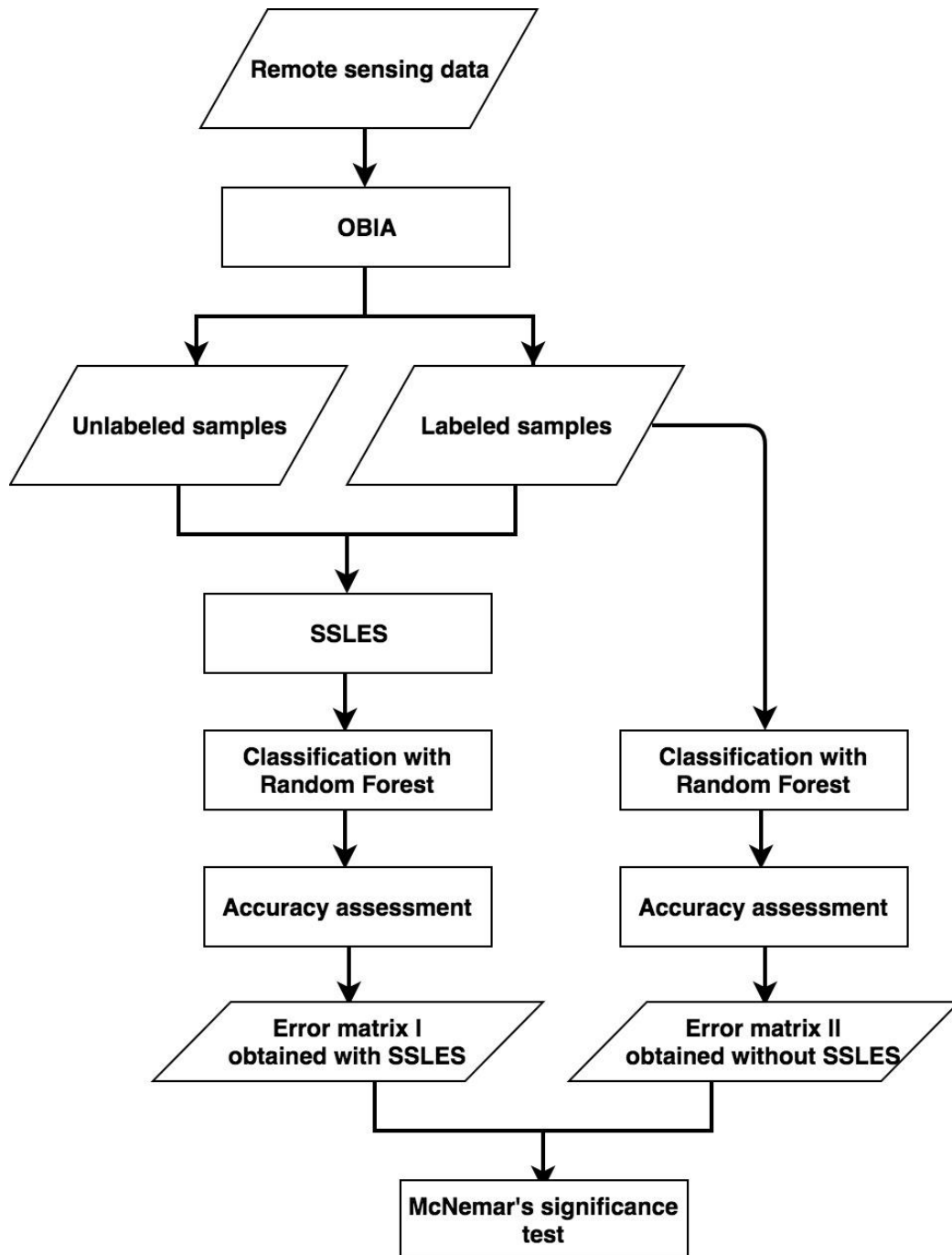


Figure 3. 1. Procedure of methodology in this study

### 3.1. Object-based image analysis

#### 3.1.1. Image segmentation

Object-based image analysis mainly involves segmentation and feature extraction steps. Segmentation is the process by which an image is partitioned into a set of spatially contiguous image objects composed of a group of pixels with homogeneity or semantic significance (Blaschke, 2010). After organizing groups of adjacent pixels into image objects (segments), they will be treated as a minimum classification unit. The generated image objects are spectrally more homogeneous within individual regions than between them and their neighbors, i.e. low within object spectral variation and high between object spectral variation. Ideally, they have distinct boundaries, and they are compact and representative (Yu et al., 2006).

The image segmentation method of Mean Shift (MS) is employed in this study to generate image objects. MS method is a non-parametric iterative clustering algorithm that does not require prior knowledge about the number and the shape of the image objects (segments). The method is based on a density mode searching and clustering techniques. It defines an empirical probability density function and estimates the modes of the densest regions in the space by finding the local maxima of the probability density function that is estimated by the kernel density estimation method (Comaniciu & Meer, 2002). When finding the locations of modes in the image, the object (segment) associated with that mode is delineated based on the local structure in the feature space.

In this study, the MS segmentation was performed following the Comaniciu & Meer (2002) work and same terminology was used. Remote-sensing imagery is typically represented as a spatial-range joint feature space. The spatial domain denotes the coordinates and locations for different pixels, and the range domain represents the spectral signals for different channels (Huang & Zhang, 2008). The modified multivariate kernel of Comaniciu & Meer (2002) is defined as the product of these two radially symmetric kernels as:

$$K_{h_s h_r}(x) = \frac{C}{h_s^2 h_r^d} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \quad (3.1)$$

where  $C$  is the normalization parameter,  $k(x)$  is kernel profile,  $x^s$  is the spatial part;  $x^r$  is the range part of a feature vector;  $h_s$  and  $h_r$  are the kernel bandwidths for spatial and range domains, respectively. The dimensionality of the joint domain is  $d=2+p$  (two for spatial domain and  $p$  for the spectral domain). Two issues need to be considered before applying MS to remote sensing data. First, MS algorithm cannot be applied on high dimensional data (Comaniciu & Meer, 2002; Georgescu et al., 2003) so the density should be analyzed along lower dimensionality if the dimensionality of data is high<sup>1</sup>. Second, since the MS feature-space analysis is task-dependent, the kernel bandwidths parameters should be tuned based on the defined task's requirements.

---

<sup>1</sup> The reason is the empty-space phenomenon by which most of the mass in a high-dimensional space is concentrated in a small region of the space (Comaniciu & Meer, 2002).

### 3.1.2. Image segmentation evaluation

Several criteria have been developed by researchers for quantitative evaluation of segmentation results (Clinton, 2016; Möller et al., 2007; Radoux & Defourny, 2008; Zhang, 1996). In this study, the method proposed by Clinton (2016) and Möller et al. (2007) is adapted which evaluates the segmentation quality by measuring both topological and geometric similarity between segmented objects and reference objects (polygons). The method relies on the ratio of intersected area of the segment and reference object area. At first, the overlapping area between segments and reference objects needs to be extracted. Two important metrics can be defined:

$$Oversegmentation_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i)} \quad (3.2)$$

$$Undersegmentation_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(y_j)} \quad (3.3)$$

Where  $x_i$  is the  $i^{th}$  reference objects,  $y_i$  is the  $i^{th}$  image segments and  $area(x_i \cap y_j)$  represents the intersection area of reference objects and segments. Assuming that the over-/under-segmentation define a 2D space, the result of a perfect segmentation (i.e. over-/under-segmentation = 0) is a point at the origin of this space. So the Euclidean norm of a vector with coordinates of over-/under-segmentation from the origin can represent a measure of the quality of a segmentation (Levine & Nazif, 1985). The distance D can be defined as the square root of the sum of squares of over-segmentation and under-segmentation.

The resulted index D is interpreted as the “*Closeness*” in the space defined above to an ideal segmentation, in the context of reference objects. D value lies in the range of [0,1] that zero represents the condition of a perfect segmentation where the segments match the reference objects entirely (Clinton, 2016).

### 3.1.3. Feature extraction

Based on the image segmentation result, features can be defined and extracted. Extraction of features depends on the specific task of classification (e.g. land cover, vegetation cover, tree, etc.) and also the availability of data layers. For mapping wetland vegetation, three categories of features which were recognized important in previous studies (Pham et al., 2016; Tigges et al., 2013; Mathieu et al., 2007; Yu et al., 2006; Haralick et al., 1973) were considered, as follows:

- (a) a set of spectral features consisting mean, standard deviation, median, minimum and maximum values of pixels within an image segment;
- (b) geometrical features representing the area and perimeter of an image segment;
- (c) a set of textural features including GLCM (Gray-Level Co-occurrence Matrix) and GLDV (Gray-Level Difference Vector). GLCM indicates how often different combinations of gray levels of two pixels at a fixed relative position occur in an

image; a different co-occurrence matrix exists for each spatial relationship. GLDV is the sum of the diagonals of the GLCM and counts the occurrence of references to the neighbor pixels' absolute differences (Haralick et al., 1973).

After extracting the features, in the band combinations selection step, the features were selected for each group of band combinations. Except for the geometrical features that are common features between the groups of band combinations, spectral and textural features depend on the existence of spectral bands of Sentinel-2 in the groups.

#### 3.1.4. Feature selection

To minimize the redundancy and inter-correlation among the extracted features and to increase the relevance to the target, it is necessary to select a subset of features (Tang et al., 2014) using a feature selection algorithm, prior to the classification process. For this purpose, Sequential Forward Feature Selection (SFFS) algorithm is conducted, in this study. The main advantage of SFFS algorithm is its simplicity and speed (Marcano-Cedeno et al., 2010).

Assume that  $F = \{f_1, \dots, f_N\}$  denotes the original feature set, consists of  $N$  features and  $X_k = \{x_1, \dots, x_k\}$  represents the subset of features containing  $k \leq N$  features. SFFS algorithm, as a bottom-up search procedure, aims to select the feature subset  $X_k$  out of  $F$  in a way that the performance of the underlying recognition system, i.e. classifier in this study, remains unchanged or even improves while  $K$  declines (Schenk et al., 2009). The algorithm initiates with an empty feature set  $X_k$  and gradually adds features that can minimize the cost function. A cost function  $J(X_k)$  is defined in terms of MisClassification Error (MCE), where  $J(X_i) > J(X_j)$  represents better performance of the feature set  $X_j$  that has lower MCE. At each iteration, the feature to be included in the feature set, is selected among the remaining available features in  $F$ , which have not been added to  $X_k$ . So, the generated subset  $X_k$  should have better performance in terms of the defined cost function, compared with the addition of any other feature. The algorithm continues to add features until the subset  $X_k$  meets the stopping criteria (De Silva & Leong, 2015; Aha & Bankert, 1996).

Basically, the stopping criterion of SFFS algorithm is “No further improvement in the MCE”, which means the algorithm will stop if adding a new feature is not improving the MCE. But in the case of using this criterion, there is the probability of under-fitting problem and the algorithm may not include some important features, because it may stop before reaching to the minimum MCE of the whole feature set (Guyon & Elisseeff, 2003). Moreover, in the case of including more features, there can be a possibility of over-fitting. Considering the over-fitting problem,  $k$ -fold cross-validation approach is used to calculate the MCE (De Silva & Leong, 2015). This validation approach partitions the data into  $k$  disjoint subsets of equal size. In the  $i^{\text{th}}$  fold of cross-validation procedure, the  $i^{\text{th}}$  subset is used to compute the MCE of the model trained on the remaining  $(k-1)$  subsets. The final MCE of the model trained on the entire dataset is then computed by averaging the MCEs obtained over all  $k$  folds. Additionally, to be able to control and overcome the under-fitting problem, the defined stopping criteria were considered as:



- (1) Adding a new feature changes the MCE by more than 10%
- (2) No statistically significant improvement in MCE values in 5 features ahead.

These criteria can help to find global minimum in MCE for whole dataset (i.e. where the feature sub-set generates the highest classification accuracy) by skipping small changes in accuracy and ignoring local variations in MCE (Guyon & Elisseeff, 2003). Algorithm 1 represents the SFFS algorithm utilized in this study.

---

**Algorithm 1: SFFS**

---

**Inputs:**

- The original feature set  $F$

**Outputs:**

- The subset of features  $X_k$

**Steps:**

For each  $f_i$  in  $F$ :

*Step 1:* Add  $f_i$  to  $X_k$ , temporarily.

*Step 2:* Compute  $MCE_i$  with k-fold cross-validation

*Step 3:* Find the best feature ( $f_b$ ) so that  $f_b = \arg \min_{f_i \in F} J(X_k)$

*Step 3:* Add the feature to  $X_k$

Remove the feature from  $F$

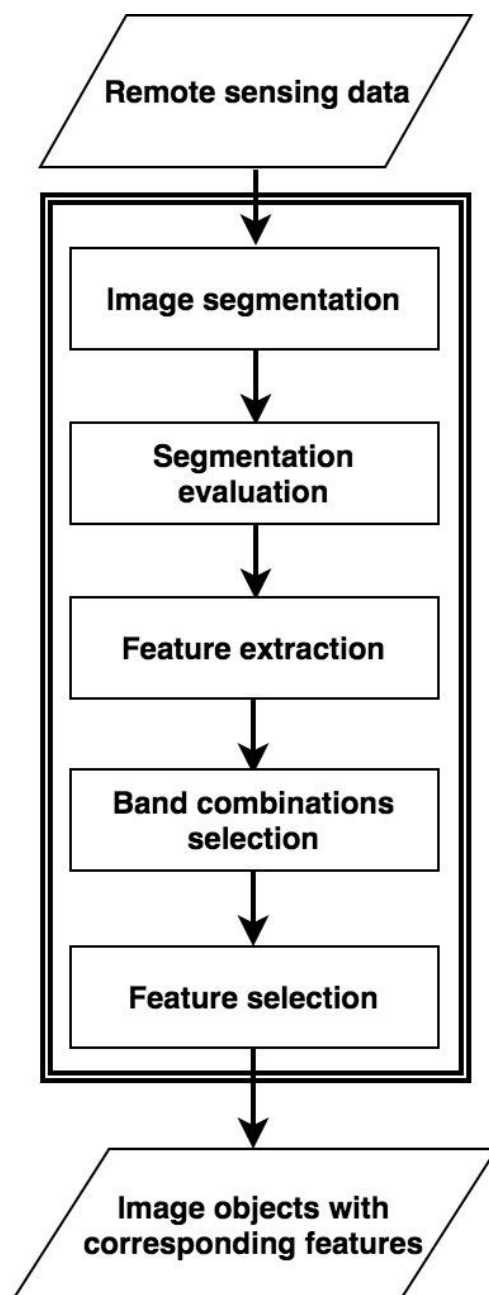
Set  $k = k + 1$ ;

*Step 4:* Evaluate the  $X_k$ . If it meets stopping criteria, then terminate.

Otherwise, go to step1.

---

By applying the feature selection algorithm, final outputs of OBIA part were being produced which are image objects and corresponding features. These objects will be treated as input data for SSLES part of this study. Figure 3. 2 is demonstrating architecture of OBIA.



*Figure 3. 2. Main architecture of OBIA in this study*

## 3.2. Semi-Supervised Learning and Expert System (SSLES)

### 3.2.1. Graph-based Semi-supervised learning

The graph-based semi-supervised learning method is used in this study to increase the training samples. Generally, graph-based SSL methods involve two main steps. The first step is graph construction, in which a graph is generated by connecting the labeled and unlabeled data and the second step is the label inference process.

#### 3.2.1.1. Graph construction

Graph-based semi-supervised learning methods rely upon the construction of a graph representation that connects similar samples. Suppose  $X_L = \{x_1, x_2, \dots, x_l\}$  are the labeled samples and  $X_U = \{x_{l+1}, x_{l+2}, \dots, x_{l+u}\}$  are the unlabeled samples and there are  $m$  classes denoted as  $C = \{c_1, c_2, \dots, c_m\}$ . Let the vector of class labels be  $Y = (Y_L, Y_U)^T = \{y_1, y_2, \dots, y_l, y_{l+1}, \dots, y_{l+u}\}$ , where  $Y_L$  and  $Y_U$  are composed of  $l$  and  $u$  labels of the labeled and unlabeled samples, respectively.

Graph-based semi-supervised learning aims to construct a graph  $G = (V, E)$  connecting similar samples, where  $V$  consists of  $N = L + U$  samples and edges  $E$  reflects the similarity between samples (Ma et al., 2016). The similarities are represented typically by a symmetric weight matrix  $W \in \mathbb{R}^{N \times N}$ . A cell  $W(i, j)$  corresponds to the similarity between the  $x_i$  and  $x_j$  samples.

Graph construction takes place in three steps (Jebara et al., 2009):

(1) *Similarity calculation*: Initially it is required to calculate the similarity between samples. In this study, the similarity measure between two samples is based on image features obtained from OBIA and is calculated by Euclidean distance as follows:

$$\|x_i - x_j\|^2 = \sqrt{(fe_{1i} - fe_{1j})^2 + (fe_{2i} - fe_{2j})^2 + \dots + (fe_{p_i} - fe_{p_j})^2} \quad (3.4)$$

where  $fe_{p_i}$  denotes the  $p^{\text{th}}$  image feature of the image object  $i$ . By computing similarity among all the samples a fully connected matrix can be generated as  $G_0 \in \mathbb{R}^{N \times N}$ .

(2) *Graph Sparsification*: This step removes the edges from  $G_0$  if it is supposed to not be a link between two samples. At first step, the edges connecting two labeled samples belonging to two different classes and the edge between unlabeled samples need to be removed, because there is no edge connection between them. After edge refining, the main sparsification can be applied to the  $G_0$  graph. The approach for sparsification, conducted in this study, is the k-Nearest Neighbors (kNN) method. For each labeled sample, kNN method chooses  $k$  closest samples (i.e. with the highest similarity values) to connect.

(3) *Graph re-weighting*: Using the sparsified graph and corresponding similarity values in  $G_0$ , a weighting scheme is applied to compute the weight of each edge. The Gaussian kernel is used in this study to generate the weight matrix denoted as  $W^w$  as follows:

$$W_{i,j}^W = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) & \text{if } x_j \in NB_K^W(x_i) \\ 0 & \text{Otherwise} \end{cases} \quad i, j \in \{1, 2, \dots, l + u\} \quad (3.5)$$

where  $\sigma$  is the Gaussian kernel bandwidth,  $\|x_i - x_j\|$  is the similarity measure between samples, and  $NB_K^W(x_i)$  is a set of  $K$  nearest neighbors of sample  $x_i$  that have the highest similarity values.

Finally, after generating the graph that connects labeled and some unlabeled samples, the class labels of unlabeled samples can be inferred using label propagation (Chapelle et al., 2010; Szummer & Jaakkola, 2001).

### 3.2.1.2. Label propagation

The main idea of label propagation is to propagate the class labels from labeled samples to the unlabeled samples, by minimizing a specific energy function (Zhu et al., 2003), until all the samples have labels. The energy function should consider two criteria (Wang et al., 2014):

- (1) minimizing the loss function, meaning that the predicted class labels of the labeled samples should be identical to the existing ones
- (2) minimizing the smoothness function, meaning that two neighbor (similar) samples are most likely to have the same class label.

In this study, the energy function proposed by Rohban & Rabiee (2012) is used, which is defined as follows:

$$\begin{aligned} \min_f \sum_{i \in \{1, 2, \dots, l\}} (f_i - y_i)^2 + \frac{1}{2} \sum_{i, j \in \{1, 2, \dots, l+u\}} W_{i,j}^W (f_i - f_j)^2 \\ = (f_l - y_l)^T (f_l - y_l) + \frac{1}{2} f^T \Delta f \end{aligned} \quad (3.6)$$

where  $f = (f_l, f_u)^T$  consists of  $f_l$  and  $f_u$  which are the predicted class labels of the labeled and unlabeled samples, respectively.  $\Delta$  is called graph Laplacian matrix obtained by  $\Delta = \mathbf{D} - \mathbf{W}$ , where  $\mathbf{D}$  is the diagonal degree matrix given by  $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ .

The class labels will propagate from the labeled samples to the unlabeled samples based on the probability values of links between samples and the link with the highest probability value will determine the class label of the unlabeled samples. This probability value is the probability that a node (samples) in the graph belongs to a specific vegetation class. The procedure stops whenever all the unlabeled samples get a specific class label.  $\mathbf{P} \in \mathbb{R}^{(l+u) \times (l+u)}$  is the probability matrix in label propagation defined as  $\mathbf{P} = \mathbf{D}^{-1} \mathbf{W}^W$ .

To better depict the idea of label propagation, a graph is constructed in Figure 3. 3 representing a set of labeled (i.e. colored circles) as well as unlabeled (i.e. white circles) vegetation samples<sup>2</sup>. In this graph, there is no connection between two labeled samples with different labels (e.g. “forest” and “herbs”), nor between two unlabeled samples. Using the abovementioned probability equation, the probability value is computed for each link between a labeled sample and the connected unlabeled sample. As an example, in the middle of the graph, it can be seen that the four labeled samples “herbs”, “herbs”, “forest” and “high shrub” are feeding into the unlabeled sample in the center. And this unlabeled sample will get the label of “forest” because of the larger probability, which is propagated from the “forest” labeled sample.

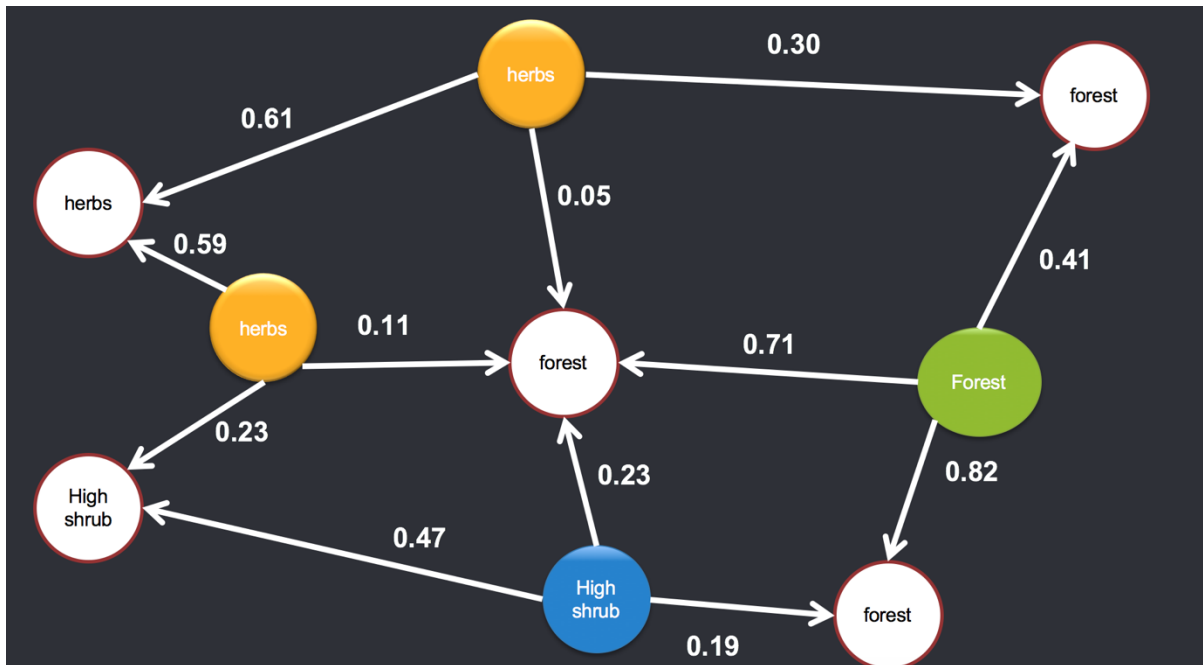


Figure 3. 3. Illustration of label propagation procedure in this study. Colored circles represent four different labeled samples and white circles represent unlabeled samples. Values indicate the probability of links. Propagation direction is shown by the direction of the arrows, i.e. arrow is always from a labeled sample to an unlabeled sample..

The final output of SSLES will be a set of labeled samples. These labeled samples will be merged with the samples in the original training set to generate a new extended training set.

### 3.2.2. Expert system

The expert system (ES) approach used in this study is described in details by (Skidmore, 1989) and the same terminology is used here. The ES in this study is developed to answer the research question of “What vegetation type is probable to occur in a given image object?”. Using a set of image features and some ancillary data the ES can infer the most probable vegetation type that may occur in an image object.

<sup>2</sup> As it was discussed before, in the context of this study samples are image objects with related image features.

Assuming that there exist  $m$  wetland vegetation classes denoted as  $C_i = \{c_1, c_2, \dots, c_m\}$  occurring in the study area and let  $E_b$  be an item of evidence (for  $b=1, \dots, k$  items of evidence such as image features) known in an image object. A hypothesis ( $H_a$ ) can be defined that class  $C_i$  occurs in an image object. A rule can be defined for this condition as  $E_b \rightarrow H_a$ .

Bayes' theory is used in the ES to compute the probability of the rule that the hypothesis ( $H_a$ ) occurs in an image object given a piece of evidence ( $E_b$ ), i.e.,

$$P(H_a|E_b) = \frac{P(E_b|H_a)P(H_a)}{P(E_b)} \quad (3.7)$$

where,  $P(E_b|H_a)$  is the *a priori* conditional probability that there exists a piece of evidence  $E_b$  (e.g., a mean slope of less than  $0.1^\circ$ ) given a hypothesis  $H_a$  (e.g., High shrub vegetation type) that class  $C_i$  occurs in a specific image object.  $P(H_a)$  is the probability for the hypothesis ( $H_a$ ) that class  $C_i$  occurs in an object and is estimated by the experienced vegetation scientist from the expected extent of each vegetation unit in the area. On iterating with the  $b = \{2, 3, \dots, k\}$  items of evidence,  $P(H_a|E_b; b=1)$  (i.e., the *a posteriori* probability of  $H_a$ , given  $E_b$ , for  $b=1$ ) replaces  $P(H_a)$  in Equation 6.  $P(E_b)$  is the "classical marginal probability," and is the probability of the evidence alone or the probability that any object has an item of evidence  $\{E_b\}$ .

The ES developed in a forward chaining manner which is data (evidence) driven process. This process starts with a set of evidence and uses rules to extract more data (evidence), if available, and works forward to the final conclusion (Hayes-Roth, 1985; Kendal & Green, 2007). It needs to be mentioned that the evidence  $\{E_b\}$  should be independent; otherwise,  $P(E_b)$  would become larger or smaller and perhaps cause the *a posteriori* probabilities to be incorrect.

Following methodology was used to define the probability rules to be able to generate the knowledge-base of ES:

- Generating a histogram population of each feature layer, e.g. slope.
- Dividing each histogram to 10 equal percentiles, representing frequency of occurrence each vegetation type at each percentile of feature layer
- Normalizing frequency values by fitting a normal distribution

### 3.2.3. SSLES algorithm

Due to the similarity problem (i.e. low within class similarity and high between class similarity), the developed ES is integrated with SSL to increase the certainty of the labeling process of the SSL. Algorithm 2 displays the steps of SSLES. Inputs of this algorithm are image objects with corresponding features extracted in OBIA. In this algorithm, steps 1-4 are related to SSL that generates a graph, based on image features, and propagates the class labels to the potential

unlabeled samples. In step 5 developed ES performs an independent class label prediction on the selected unlabeled samples of the SSL. Then if both ES and SSL approaches agreed on a same class label, the sample was added to the training samples; otherwise, it was returned to be classified by the post-classifier. The flowchart of SSLES algorithm is represented in Figure 3. 4.

---

**Algorithm 2: SSLES**

---

**Inputs:**

- A set of labeled object objects  $O_L$
- A set of unlabeled object objects  $O_U$
- Extracted features  $f_i$  for each object

**Outputs:**

- The classification result for all the image objects

**Steps:**

For every  $o_i \in O_L, i=1:l$

1. Measure the similarity to all unlabeled samples using Eq. 3.4
  2. Sort the similarity values and choose the first K unlabeled samples with the highest similarity values
  3. Construction of graph: Construct the graph  $W^w$  with the KNN method using Eq. 3.5 and calculate the probability matrix P
  4. Prediction of the labels of the unlabeled objects: predict the labels exploiting Label propagation using Eq. 3.6
  5. Prediction of the labels of the unlabeled objects: predict the labels exploiting Expert system
  6. If both approaches have predicted same class label  
Then add the object to the labeled object set  
Unless move back to unlabeled object set
  7. Post-classification: Train RF with labeled samples and predict the labels of test samples
-

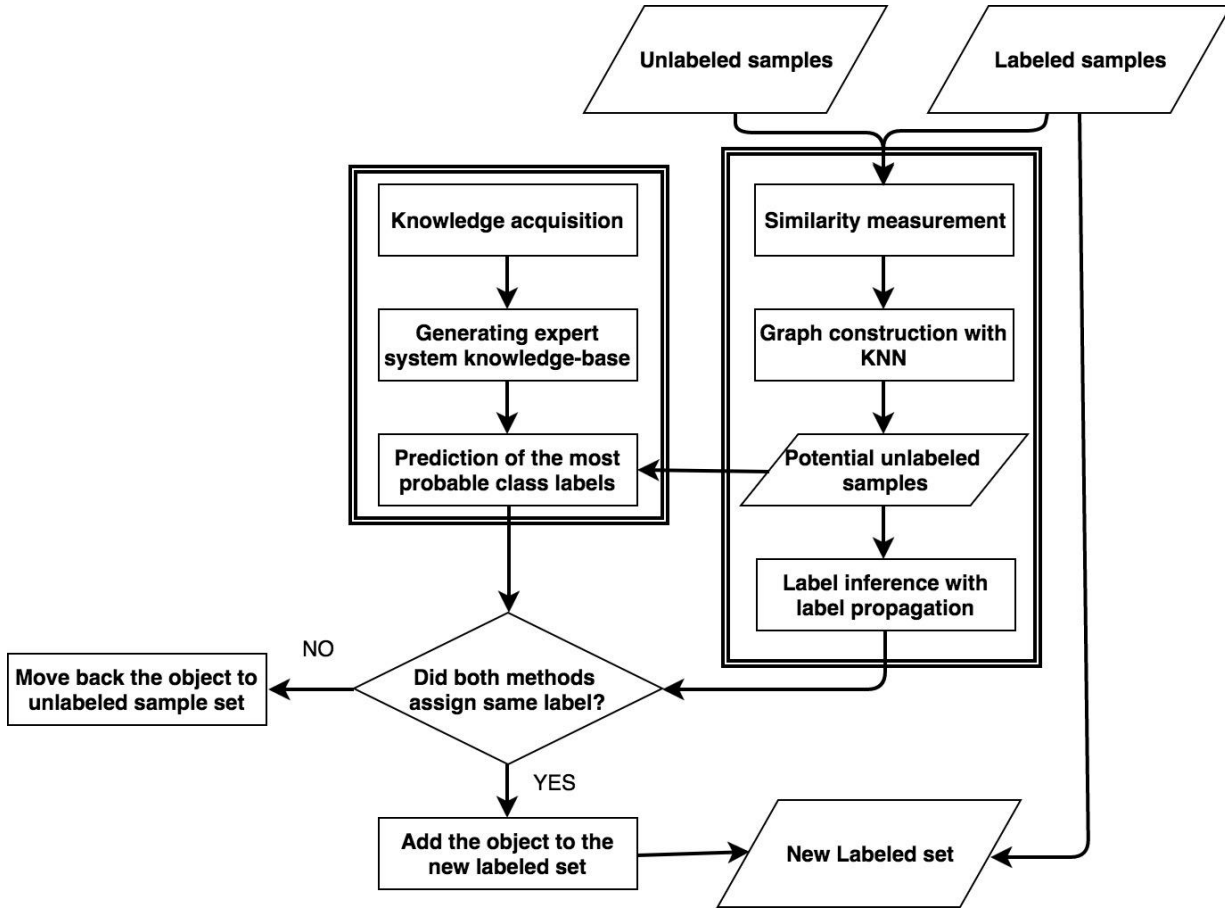


Figure 3. 4. SSLES algorithm flowchart

An important point to be mentioned about the proposed methodology is the question that “Where to put the decision boundary between two classes if they are similar?”. To be able to answer this question first it is required to discuss the effect of using unlabeled samples in the context of SSL (Chapelle et al., 2010). As an example in Figure 3. 5(a), assuming that two classes are coherent groups (e.g. Gaussian), using unlabeled samples with labeled samples shifts the decision boundary that is made using labeled samples only. This shift of decision boundary will be toward the actual decision boundary only if the unlabeled samples get the correct class label in SSL procedure, which may not happen because of similarity problem. Otherwise, obviously utilizing unlabeled samples will deteriorate the classifier. The use of the expert system to filter out the uncertain unlabeled samples in the proposed approach applies a second shift to the decision boundary which may move it as close as possible to the actual decision boundary (Figure 3. 5(b)).



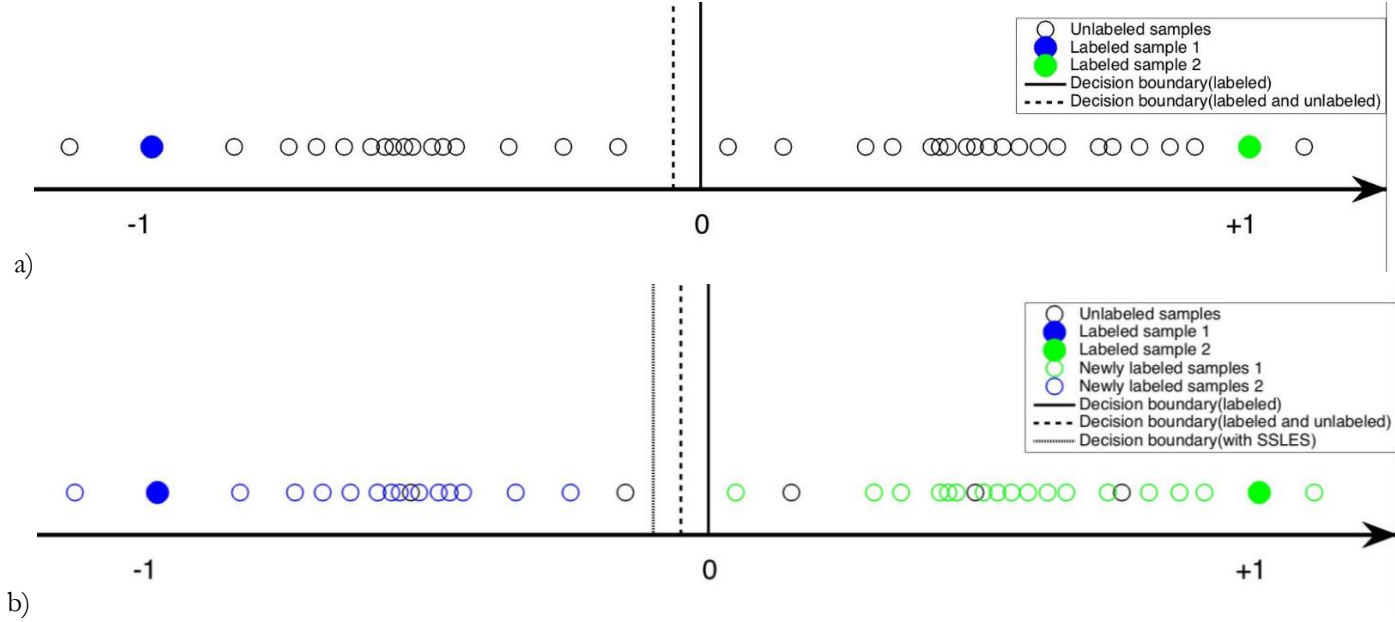


Figure 3. 5. Thematic representation of effect of using unlabeled samples on the decision boundary when a) using unlabeled samples and b) using unlabeled samples obtained from SSLES

### 3.3. Post-classification and evaluation

#### 3.3.1. Post-classification

The post-classification step is performed using the Random Forest (RF), which is a non-parametric ‘ensemble learning’ algorithm (Breiman, 2001; Brieman, 1999). RF uses a bootstrap aggregation technique (bagging) to generate sub-sets of training data with which to build an ensemble of decision trees (base classifiers). The bagging process involves resampling the original training set with replacement, resulting in a greater diversity of decision trees, thereby improving classifier stability and accuracy (Breiman 1999). Each decision tree provides a classification result for the samples not chosen and the output class is obtained as the majority vote of the outputs of a large number individual trees. Moreover, in constructing trees, as some training data instances may be used more than once or not at all, correlation between trees is reduced, and as a result, RF is more robust to variations in input data and less sensitive to mislabeled training data or over-fitting (Pal, 2005).

An RF classifier consists of  $n$  trees ( $ntree$ ) to be grown and each tree is generated from  $m$  variables ( $mtry$ ). In the context of this study, inputs to train the RF classifier are image objects with related image features. Each tree in RF classifier takes a subset of image objects from the training set and then a subset of related image features of those objects and trains a classifier. The final class prediction of the RF classifier is based on majority vote of  $n$  trees (Pal, 2005). Thus,  $ntree$  and  $mtry$  are two user-defined parameters required to generate a random forest classifier and as different values of these two parameters have the effect on the final results of the classifier, they should be tuned to get the optimum results. Regarding the  $mtry$  parameter of RF classifier, several studies have proved that the default value of this parameter (i.e. square root of a total number of the

input features) can provide satisfactory results (Immitzer et al., 2016; Immitzer et al., 2012; Liaw & Wiener, 2002). Consequently, only the *ntree* parameter needs to be tuned, in this study.

In this study, RF classifier is used in two different cases. In the first case, the classifier is trained using the original training set and in the second case, the training set generated from SSLES is used to train the classifier. While using two different training set to train the classifier, the results of the both case are evaluated using the same test set.

### 3.3.2. Classification evaluation

#### 2.3.2.1. Overall accuracy

For the validation of the classification results, accuracy assessment based on error matrix is conducted (Congalton & Green, 2008). The error matrix measures an overall accuracy for the classification, as well as a producer and user's accuracy for each feature class. Overall Accuracy (OA), as the descriptive static derived from error matrix, and kappa coefficient, as another discrete multivariate static, are used as the performance metrics to compare the classification results. In addition, producer's accuracy and user's accuracy are used to compute the individual classes' accuracies (Congalton, 1991). In this study, the final evaluation of the classification results for two cases of classification with SSLES and without SSLES will be performed using the same 434 test samples extracted from reference data.

#### 2.3.2.2. McNemar's test

In this study the same test set is used to assess the accuracy of two classification methods (i.e. with SSLES and without SSLES) and the comparison of two classification results is conducted using McNemar's test. As stated by Foody (2004), for two dependent samples, the statistical significance of the difference between two proportions may be evaluated using McNemar's test. It is a non-parametric test that is based upon a 2x2 confusion matrix that represents the count of correctly an incorrectly classified samples (Table 3. 1) (McNemar, 1947). The null hypothesis states that both classification models generate the same accuracy.

*Table 3. 1. Confusion matrix of McNemar's test*

	Test1-Correct	Test1-Incorrect	Sum
Test2-Correct	N <sub>11</sub>	N <sub>12</sub>	N <sub>11</sub> +N <sub>12</sub>
Test2-Incorrect	N <sub>21</sub>	N <sub>22</sub>	N <sub>21</sub> +N <sub>22</sub>
Sum	N <sub>11</sub> +N <sub>21</sub>	N <sub>12</sub> +N <sub>22</sub>	N <sub>0</sub>

The test is based on a chi-square statistic, computed as follows:

$$\chi^2 = \frac{(N_{12} - N_{21})^2}{(N_{12} + N_{21})} \quad (3.8)$$

Where  $N_{12}$  is the number of samples predicted correctly by the second classifier and incorrectly by the first classifier and  $N_{21}$  represent samples correctly classified by the first classifier and incorrectly by second classifier.  $\chi^2$  follows a Chi-squared distribution with one degree of freedom. The statistical significance can be obtained with the derived  $\chi^2$  compared against tabulated chi-square values. For this, using a chi-squared distribution, P-value is computed for the test. P-value is defined as the probability of finding chi-squared value more than the computed one when the null hypothesis of the test is true. A small value for P is evidence to reject the null hypothesis, while larger values of P means the null hypothesis should not be rejected.

## 4. RESULTS

In this chapter the results of the model implementation and performance comparison will be reported in following order: Data pre-processing/data preparation will report the results of the OBIA. Then the results of deriving rules from sources of knowledge as well as generating the knowledge-base of expert system will be discussed. Next, in the model selection section, the optimal parameter set will be described and the final experimental results of the proposed approach will be presented. Finally, the last section will present the results of McNemar's significance test on the obtained results.

### 4.1. Data pre-processing/data preparation

Before using Sentinel-2 data for further analysis, first, its geometric accuracy was checked. For this purpose, it was co-registered to the RapidEye image, which was already geo-referenced and also had a higher spatial resolution. The co-registration was done using a second-degree polynomial transformation and nearest neighbor resampling with an RMS error of 0.47. To account for the Sentinel-2 spatial characteristics, 10-meter bands were resampled to 20 m and 60-meter bands were excluded from further processing. After geometric correction, the Sentinel-2 image became ready for further analysis in OBIA, which was designed in four steps: image segmentation, segmentation evaluation, feature extraction and feature selection.

#### 4.1.1. Image segmentation

Image segmentation was applied on RapidEye image to obtain homogeneous objects in the images. As this multi-spectral image had a higher spatial resolution compared to the Sentinel-2 image, it could result in a higher spatial accuracy of image objects (Laurent et al., 2014). RapidEye image's five spectral bands (i.e. RGB, red edge, and near IR) in combination with the derived NDVI map, were used as input for the segmentation.

The mean shift segmentation algorithm was implemented in MATLAB\_R2015b to segment the satellite image. This algorithm has two important parameters that need to be optimized to be able to get the optimum result from the segmentation process. To find out the parameter set, the values of  $h_s$  (spatial radius) and  $h_r$  (range radius) were changed empirically within the range of [0, 10] and segmentation results were evaluated using the training polygons from reference map. D (closeness to ideal segmentation) as the segmentation evaluation metric was computed based on the over- segmentation and under-segmentation rates. The D values of segmentation results, with respect to changes in  $h_s$  and  $h_r$ , are reported in Figure 4. 1.

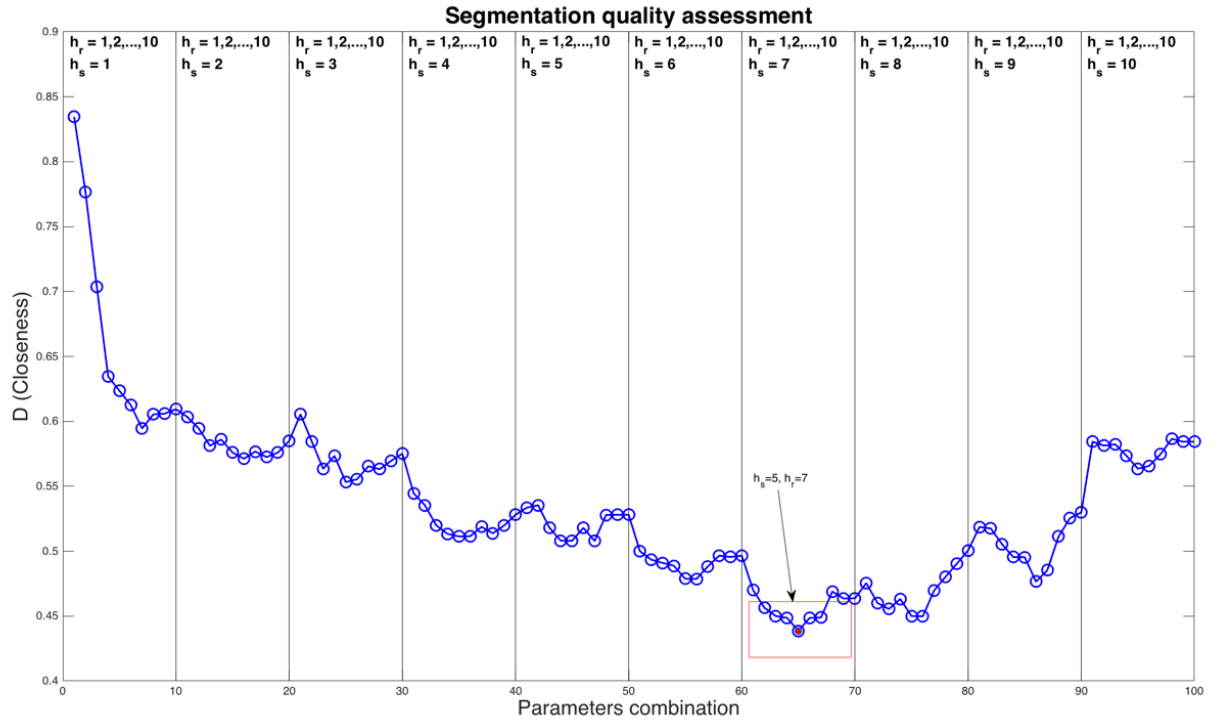
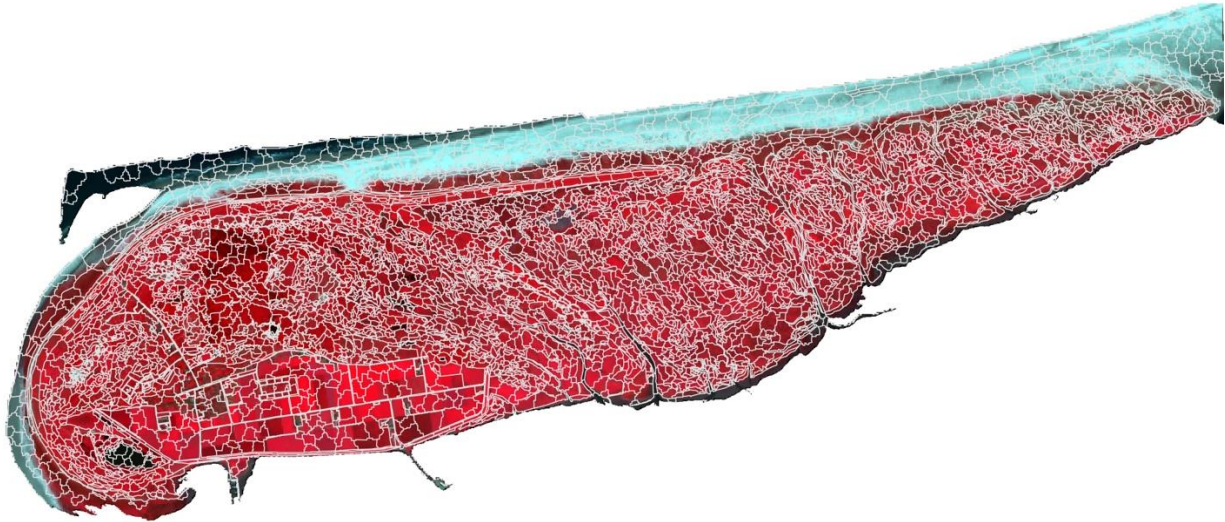


Figure 4. 1. Results of segmentation evaluation process

As shown in the figure above, when the  $h_s$  is small, as the result of obvious over-segmentation, the segmentation yields in higher values of  $D$ , while increasing in  $h_r$  shows improves the  $D$  value. With the rise of  $h_s$ , over-segmentation phenomenon falls off and  $D$  value falls too. However, the minimum value of  $D$  occurs when  $h_s = 7$ . As  $h_s$  continues to increase,  $D$  value starts to enlarge as the result of over-segmentation. The parameter  $h_r$  improves the  $D$  value as it increases, but just after a certain value (i.e. 5 or 6), due to under-segmentation  $D$  value starts to deteriorate. The best segmentation result (i.e. least  $D$  value) happened at  $h_s = 5$  and  $h_r = 7$ , so the optimum segmentation result was obtained using these two parameter values. Figure 4. 2 illustrates the final result of segmentation on the false color composite map of the study area.



*Figure 4. 2. Final segmentation result with  $hs=5$  and  $hr=7$ , on false color composite of Sentinel-2*

A total number of 5230 image objects were delineated with a mean size of 30 pixels. Then the image objects were divided into three data sets: labeled (training) objects<sup>3</sup>, unlabeled objects, and testing objects, where labeled and test objects were generated using the reference vegetation map, and all the remainder objects in the images were set as unlabeled. To generate the labeled objects, the segmented map layer was overlaid with the reference samples' layer and any image object that contained the centroid of a reference sample (polygon) and had more than 50% overlap with it, was treated as a training object. The same approach was applied to generate the test objects. The numbers of the generated objects for three datasets are listed in Table 4. 1. Further in Figure 4. 3 a part of the collected labeled samples in the study area is presented.

*Table 4. 1. Number of image objects for three different subsets*

Labeled(training) Objects	Test Objects	Unlabeled Objects	Total number of Objects
650	434	4146	5230

---

<sup>3</sup> An object (image object) in the context of this study, as the result of segmentation, represents a vegetation patch. If the (vegetation) class label of a sample is known, then that object is a labeled sample otherwise, that object is unlabeled.





Figure 4. 3. Reference samples among the image objects

#### 4.1.2. Feature extraction

There existed 13 information layers to extract features including DEM, slope, NDVI and 10 spectral bands of Sentinel-2 (excluding three 60m bands). Table 4. 2 summarizes the list of extracted features with related descriptions.

Table 4. 2. Extracted features and their description. B1-10: Sentinel-2 spectral band number 1 to 10

INFORMATION LAYER	FEATURE NAME	FEATURE DESCRIPTION	TOTAL NUMBER OF FEATURES
B1-10, DEM, SLOPE, NDVI	Mean	Mean of pixel values inside the object	16
B1-10, DEM, SLOPE, NDVI	Median	Mode of pixel values inside the object	16
B1-10, DEM, SLOPE, NDVI	Min	Minimum of pixel values inside the object	16
B1-10, DEM, SLOPE, NDVI	Max	Maximum of pixel values inside the object	16
B1-10, DEM, SLOPE, NDVI	STD	Standard deviation of pixel values inside the object	16
B1-10	GLCM_H	GLCM homogeneity from 10 spectral bands of Sentinel	10
B1-10	GLCM_C	GLCM contrast from 10 spectral bands of Sentinel	10

# INTEGRATING SEMI-SUPERVISED LEARNING AND EXPERT SYSTEM FOR WETLAND VEGETATION CLASSIFICATION USING SENTINEL-2 DATA

<b>B1-10</b>	GLCM_D	GLCM dissimilarity from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLCM_E	GLCM entropy from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLCM_SD	GLCM standard deviation from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLCM_M	GLCM mean from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLCM_A	GLCM angular 2nd moment from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLCM_Cor	GLCM correlation from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLDV_M	GLDV mean from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLDV_E	GLDV entropy from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLDV_C	GLDV contrast from 10 spectral bands of Sentinel	10
<b>B1-10</b>	GLDV_A	GLDV angular second moment from 10 spectral bands of Sentinel	10
-	Area	Area of an object	1
-	Perimeter	Perimeter of an object	1

A total of 187 spectral, geometrical and textural features were extracted from the information layers. The feature extraction procedure was carried out for all six groups of band combinations and the number of features ranged from 51 to 187. Each group consisted of 17 basic features, including two geometrical features and 15 spectral features for DEM, slope and NDVI. The rest of the features were band related, i.e. in the case of including any spectral band in a group, it would add 12 textural features and 5 more spectral features. As an instance, group 1 consisted of 17 basic features and as it included 10 bands of Sentinel-2  $10 \times (5+2)$  more features were included in the group, which became 187 in total. Table 4. 3 represents the detailed number of extracted features for different band combinations.

*Table 4. 3. Number of extracted features for different groups of band combinations*

	<b>Spectral features</b>	<b>Textural features</b>	<b>Geometrical features</b>	<b>Total number of features</b>
<b>Group 1</b>	65	2	120	187
<b>Group 2</b>	25	2	24	51
<b>Group 3</b>	45	2	72	119
<b>Group 4</b>	35	2	48	85
<b>Group 5</b>	25	2	24	51
<b>Group 6</b>	45	2	72	119

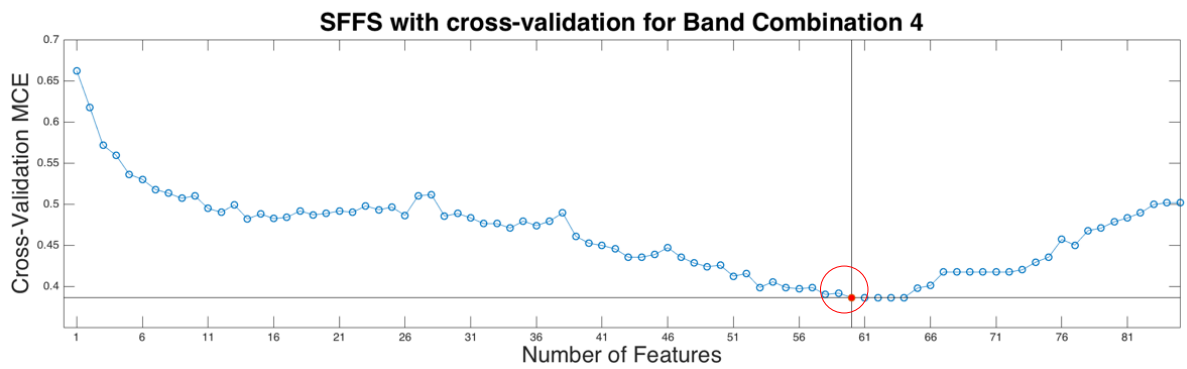
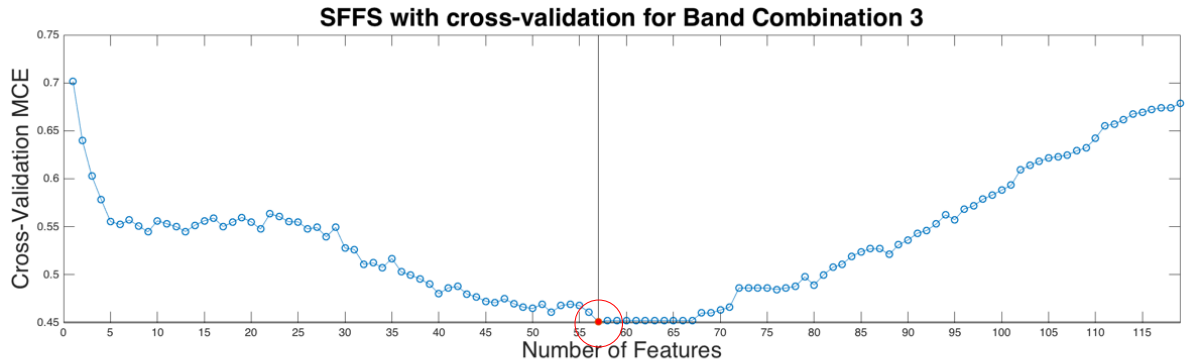
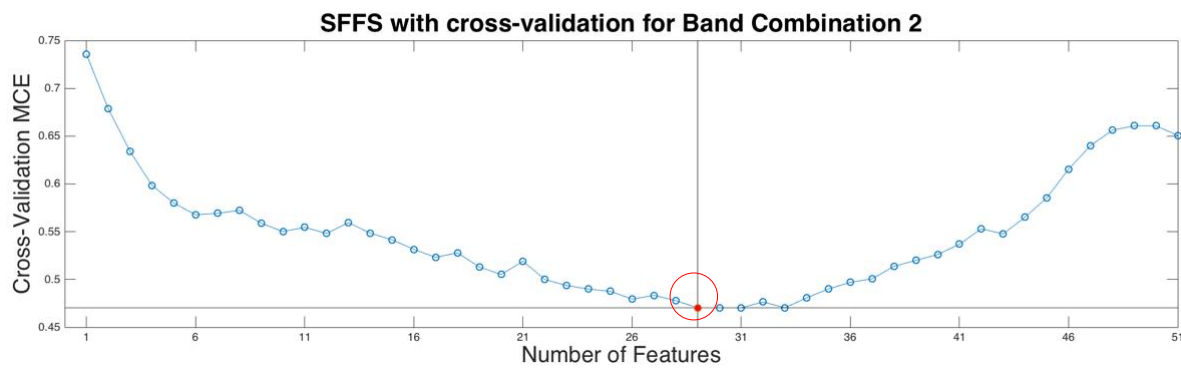
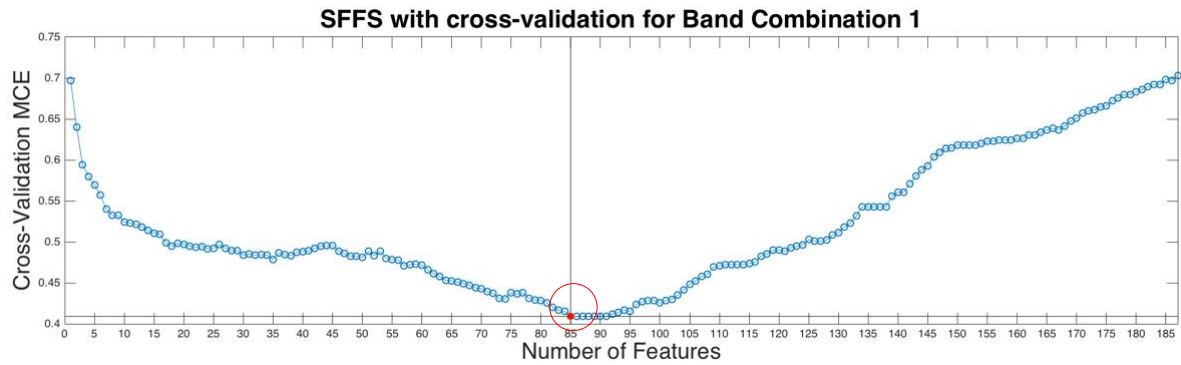
## 4.2.3. Feature selection

Sequential forward feature selection (SFFS) algorithm was implemented in MATLAB\_R2015b and was applied on six datasets for the band groups to find the optimal feature subsets. In this study, 10-fold cross-validation approach was used in each iteration of the SFFS algorithm to validate the results and compute the misclassification error (MCE). Figure 4. 4 shows the changes of 10-fold cross-validation MCE by adding more feature from the feature set that results in the



# INTEGRATING SEMI-SUPERVISED LEARNING AND EXPERT SYSTEM FOR WETLAND VEGETATION CLASSIFICATION USING SENTINEL-2 DATA

lowest MCE, at each iteration of the algorithm. The optimum number of features is highlighted with a red circle.



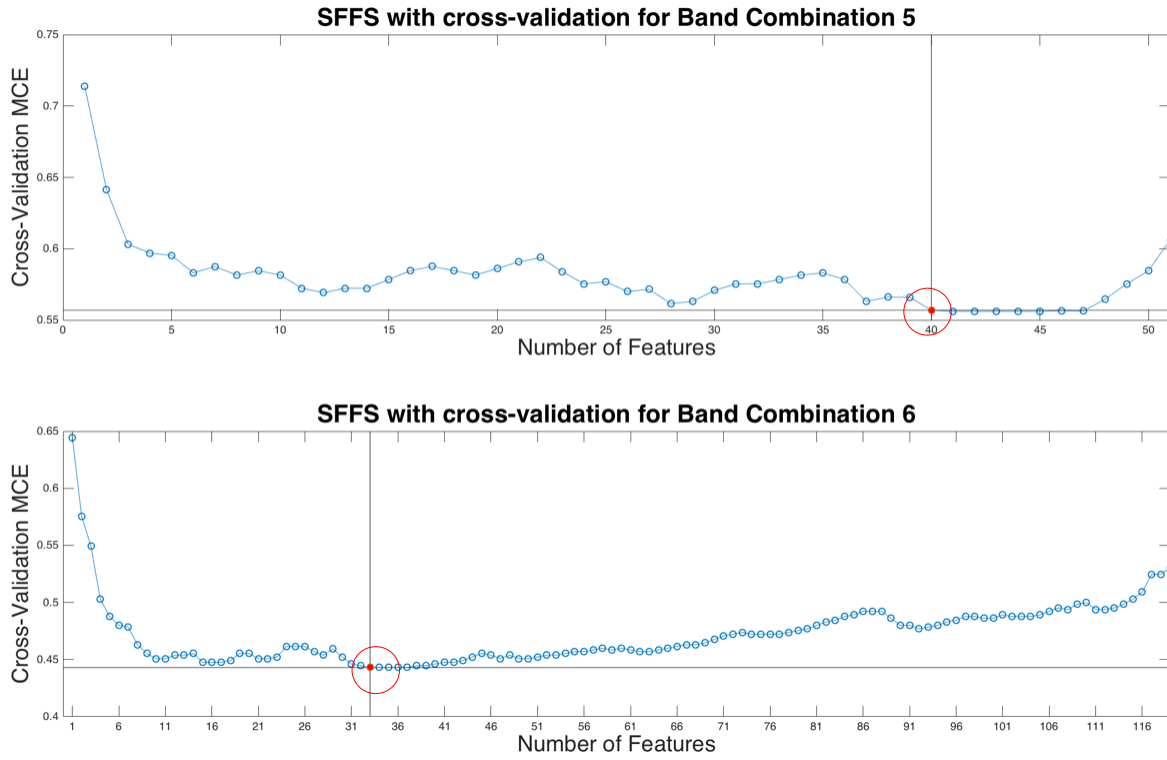


Figure 4. 4. Feature selection results for band groups. Optimal numbers of features are highlighted with a red circle.

## 4.2. Expert system rules and *a priori* probabilities

Three available sources of knowledge were used to generate the knowledge-base of ES, as follow:

1. A reference vegetation map of the study area, generated in 2010 (Pranger & Tolman, 2012), as this map is generated with expert visual interpretations of aerial photographs and extensive fieldworks, it contains some level of experts' knowledge.
2. Ancillary data, including field records of dominant vegetation types for 30 vegetation plots, NDVI data, slope and DEM maps
3. Published resources by scientists and experts about vegetation types in wetlands and Schiermonnikoog wetland and ecology of the study area (Rundquist et al., 2001; Schmidt et al., 2004; Bird, 2008; Vrieling et al., 2017).

The training samples with required features extracted from ancillary data are statistically analyzed to define the probability of occurrence of a vegetation type within an image object.

The result is a graph for each feature layer that displays the probability values of occurrence of an evidence (i.e. a set of features) ( $E_b$ ) given a hypothesis (i.e. specific vegetation type) ( $H_a$ ), or in other word  $P(E_b|H_a)$ . *A priori* probabilities for the vegetation classes and the initial conditional probabilities for all the feature layers are estimated using the reference vegetation map (Table 4.4).

# INTEGRATING SEMI-SUPERVISED LEARNING AND EXPERT SYSTEM FOR WETLAND VEGETATION CLASSIFICATION USING SENTINEL-2 DATA

Table 4. 4. A priori probabilities estimation of the vegetation classes. The probability of occurrence of each vegetation type in the study area is presented. HMG: High Matted Grass, LMG: Low Matted Grass, Ag: Agriculture, Fr: Forest, GB: Green Beach, TG: Tussock Grass, Hr: Herbs, LSS: Low Salix Shrub, LHS: Low Hippophae shrub

Class name	HMG	LMG	Ag	Fr	GB	TG	HS	Hr	LSS	LHS
Probability	0.28	0.24	0.1	0.08	0.08	0.06	0.06	0.05	0.03	0.01

Figure 4. 5 shows six sets of the derived expert rule weights. These rules belong to mean and standard deviation values of the feature layers.

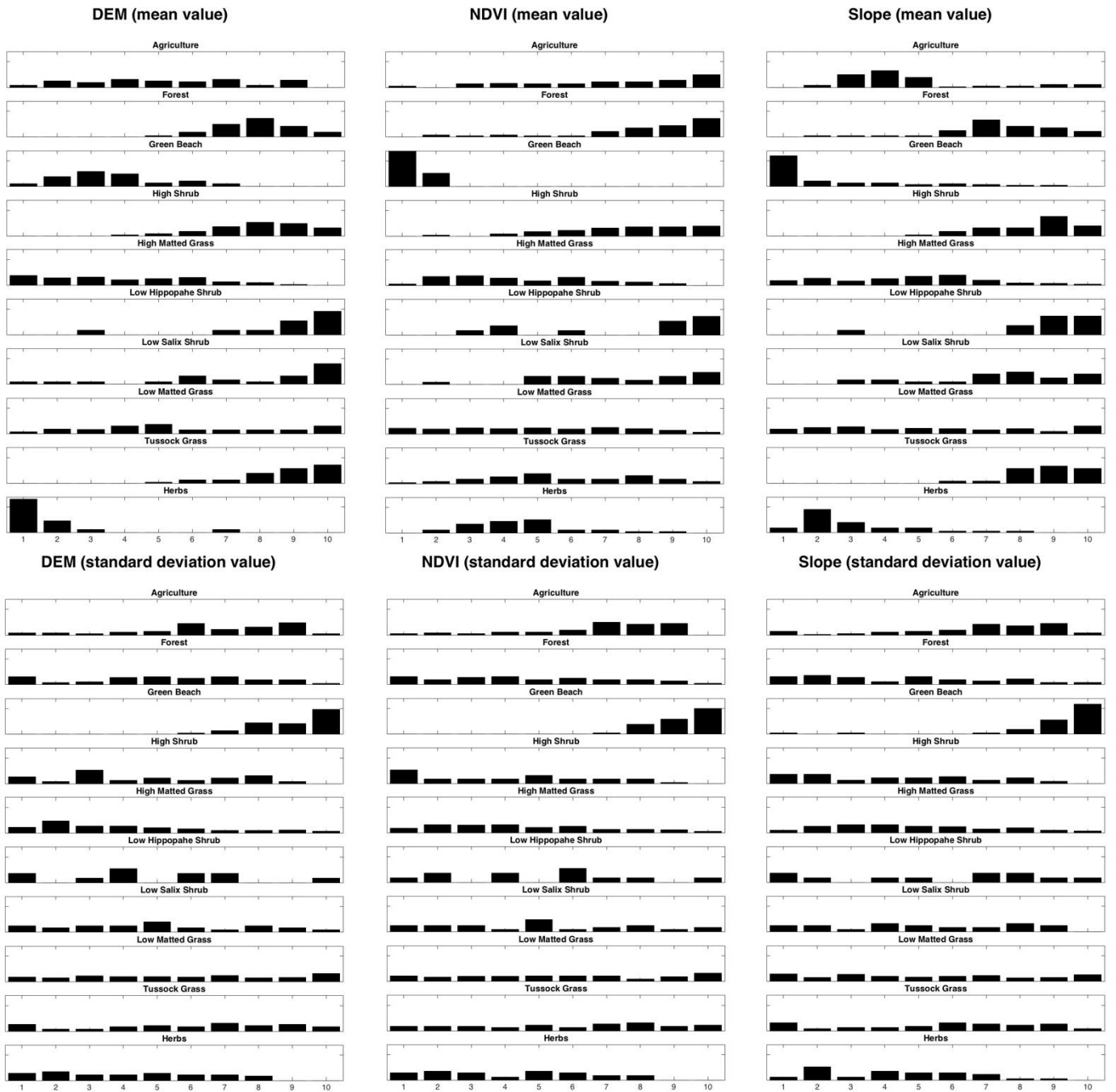


Figure 4. 5. Expert rules weights for ancillary data. Y axis shows the initial conditional probability value and X axis shows the 10 quantiles

Using the knowledge resource, three more facts extracted to generate a new set of rules. These three facts are as follows:

- (1) Dependence of some vegetation classes to water resources (streams). For instance, in Schiermonnikoog island, herbs patches are mostly found adjacent to streams in south-eastern parts of the island.
- (2) The presence of specific vegetation classes around the residential areas. For instance, agriculture classes only exist adjacent to the residential area and also as the result of human activities and ecological reasons, some classes such as high matted grass and herbs are rare to appear close to the residential areas.
- (3) A forestry program in the island to protect sand drifting and to protect the residential areas. In 1990 there was a forestry program in the island to protect sand drifting and to protect the residential areas, therefore, forest class appears mostly adjacent to the village (Beukeboom, 1976).

Using the method described before, the measured distances between classes, based on above-mentioned facts and rules, are divided into three quantiles and then the probability of occurrence of vegetation species at each distance quantile is computed (Table 4. 5).

*Table 4. 5. Expert rules based on the distance of classes to streams and residential areas*

<b><i>Vegetation classes</i></b>	<b><i>Distance to streams</i></b>			<b><i>Distance to residential area</i></b>		
	quantile 1	quantile 2	quantile 3	quantile 1	quantile 2	quantile 3
<i>Agriculture</i>	0.11	0.3	0.58	0.9	0.05	0
<i>Forest</i>	0.06	0.21	0.72	0.41	0.32	0.25
<i>High matted grass</i>	0.4	0.45	0.1	0	0.13	0.85
<i>Low matted grass</i>	0.21	0.31	0.47	0.23	0.4	0.31
<i>Tussock grass</i>	0.09	0.21	0.7	0.27	0.52	0.18
<i>High shrub</i>	0.19	0.22	0.56	0.12	0.45	0.42
<i>Low hippopabe shrub</i>	0	0.24	0.71	0	0	0.96
<i>Low salix shrub</i>	0.22	0.1	0.6	0.21	0.76	0
<i>Green beach</i>	0	0.51	0.49	0	0	0.99
<i>Herbs</i>	0.61	0.38	0	0	0	0.99

Finally, a contextual check is performed to ensure that the predicted class for an image object is ecologically possible to appear adjacent to other vegetation classes. A matrix (Table 4. 6) is set up that provides contextual information about the species that can naturally occur adjacent to other species. Adjacent objects in this study are the objects that share the same border. The predicted class for each object will be checked with this matrix. According to Table 4. 6, if the predicted class for an object can naturally occur alongside the adjacent objects, then the object can get the predicted class label as the final label, otherwise, that object will be moved back to unlabeled samples.

Table 4. 6. Matrix representing vegetation types that can occur adjacent to other vegetation types based on expert ecological knowledge. HMG: High Matted Grass, LMG: Low Matted Grass, Ag: Agriculture, Fr: Forest, GB: Green Beach, TG: Tussock Grass, Hr: Herbs, LSS: Low Salix Shrub, LHS: Low Hippopahe shrub

	Ag	Fr	GB	GS	HMG	LHS	LSS	LMG	TG	Hr
Ag	×	×						×	×	
Fr	×	×						×	×	
GB			×		×			×		
HS				×			×	×		
HMG			×		×					×
LHS						×		×	×	
LSS				×			×	×		
LMG	×	×	×	×		×	×	×		
TG	×	×				×			×	
Hr					×					×

### 4.3. Model selection (Parameter optimization)

#### 4.3.1. SSL parameters

For graph construction in SSL, there were two parameters that must be tuned for optimum graph construction. The first parameter was RBF kernel width, which was used to generate the similarity matrix, and was tuned in the range  $\sigma = \{0.1, 0.2, \dots, 2\}$ . The other parameter to be tuned was k nearest neighbors of KNN algorithm and to find the optimal value for this parameter, it was tuned in the range of  $k = \{1, 2, \dots, 20\}$ . First, the  $\sigma$  value was changed while k was constant (Figure 4. 6(a)) and then the k value was changed while  $\sigma$  was constant (Figure 4. 6(b)).

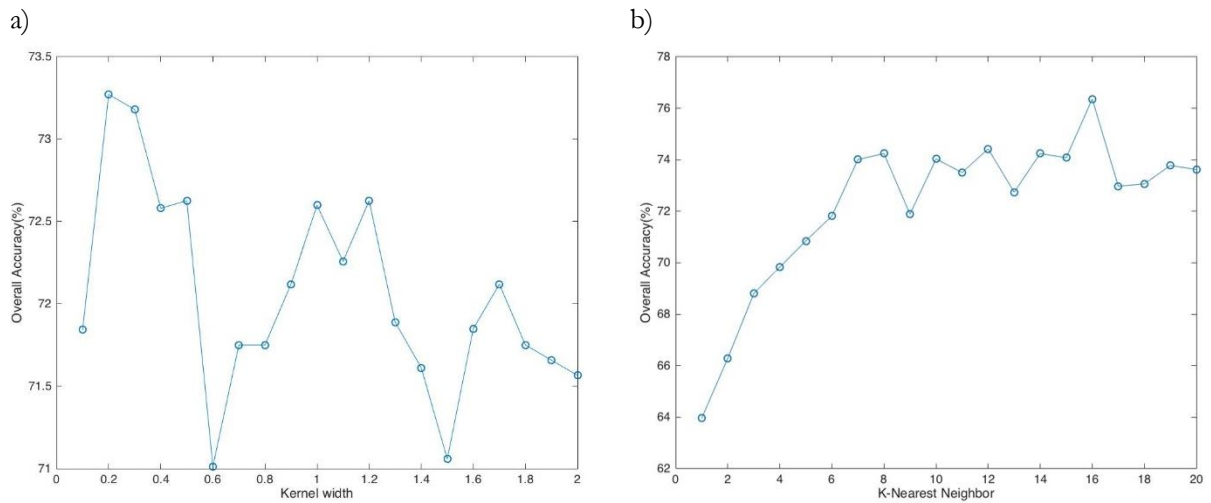


Figure 4. 6. Analysis of SSL parameters; a) kernel width vs OA, b) number of k neighbor vs OA

According to Figure 4. 6(a), the highest accuracy can be achieved when the kernel width is 0.2. In

the weight formula (i.e.  $W(i, j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$ ) lower value of kernel width can result in higher value for weight matrix and this can result in assigning higher weights to false neighbor objects and when this parameter has high values, weight matrix will include small weights (close to zero). For the parameter  $k = 16$  is chosen as the optimum value as this value resulted in highest OA.

#### 4.3.2. Classification parameters

Regarding the *mtry* parameter of RF classifier, several studies have proved that the default value of this parameter (i.e. square root of a total number of the input features) can provide satisfactory results (Immitzer et al., 2016; Immitzer et al., 2012; Liaw & Wiener, 2002). To tune the *ntree* parameter, its value was changed in the range of {10, 800} and *ntree* value yielding highest OA was chosen as the optimal value of this parameter (Figure 4. 7).

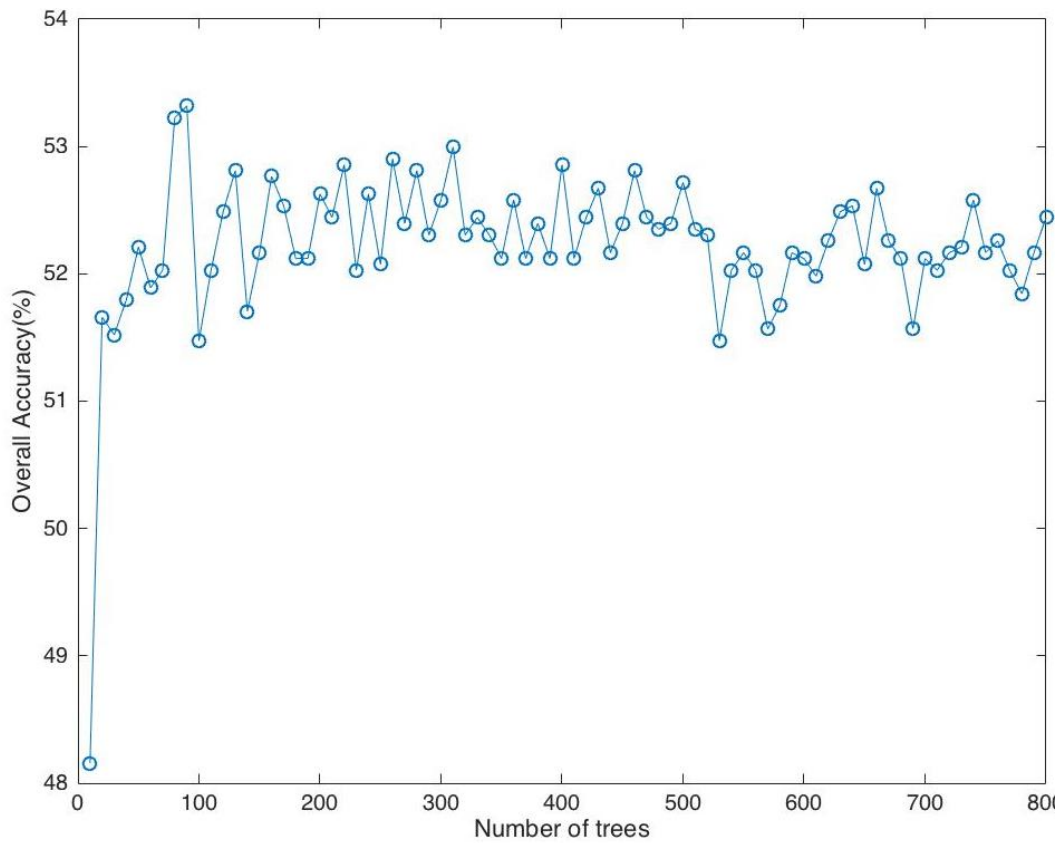


Figure 4. 7. Analysis of *ntree* parameter of RF classifier

According to Figure 4. 7, maximum accuracy was achieved when *ntree*=80. So this value was chosen to perform the final classification. After finding the optimum values for the three parameters of the model, the model can be applied to the data for further evaluation and comparison.

#### 4.4. Model performance evaluation

In two sets of experiments, i.e. with SSLES and without SSLES, accuracy procedures of both procedures for six groups of band combinations was evaluated in terms of overall accuracy and Cohen's kappa coefficient using the same test set. Table 4. 7 reports the classification scores, where overall accuracy and kappa coefficient statistics are displayed. Two conclusions can be obtained from Table 4. 7. First, it is remarkable that SSLES can produce better results compared to the use of RF classifier solely. This fact applies for all the six groups of band combinations. Second, the highest accuracy is observed for the 4th group of band combination which is Red-Edge bands only.

Table 4. 7. Classification results for six groups of band combinations

DATASET	WITHOUT SSLES		WITH SSLES	
	<i>OA(%)</i>	<i>kappa</i>	<i>OA(%)</i>	<i>kappa</i>
<b>GROUP 1</b>	59.4	0.51	81.1	0.67
<b>GROUP 2</b>	52.1	0.41	73.5	0.57
<b>GROUP 3</b>	54.4	0.45	74.6	0.59
<b>GROUP 4</b>	61.3	0.53	83.6	0.70
<b>GROUP 5</b>	44.9	0.32	67.8	0.49
<b>GROUP 6</b>	59.9	0.52	79.9	0.67

However, classification with the group of band combinations 1 and 6 could yield higher OA, too. As the differences in the OA value of these three results were small, it was necessary to assess if they have statistically significant differences. The lowest classification accuracy in the table above belongs to the Shortwave Infrared spectral bands with 44.9% of OA, however using SSLES could increase its OA to 67.8%. Figure 4. 8 illustrates the final classified map of the study area with the group 4 dataset using the SSLES in classification.



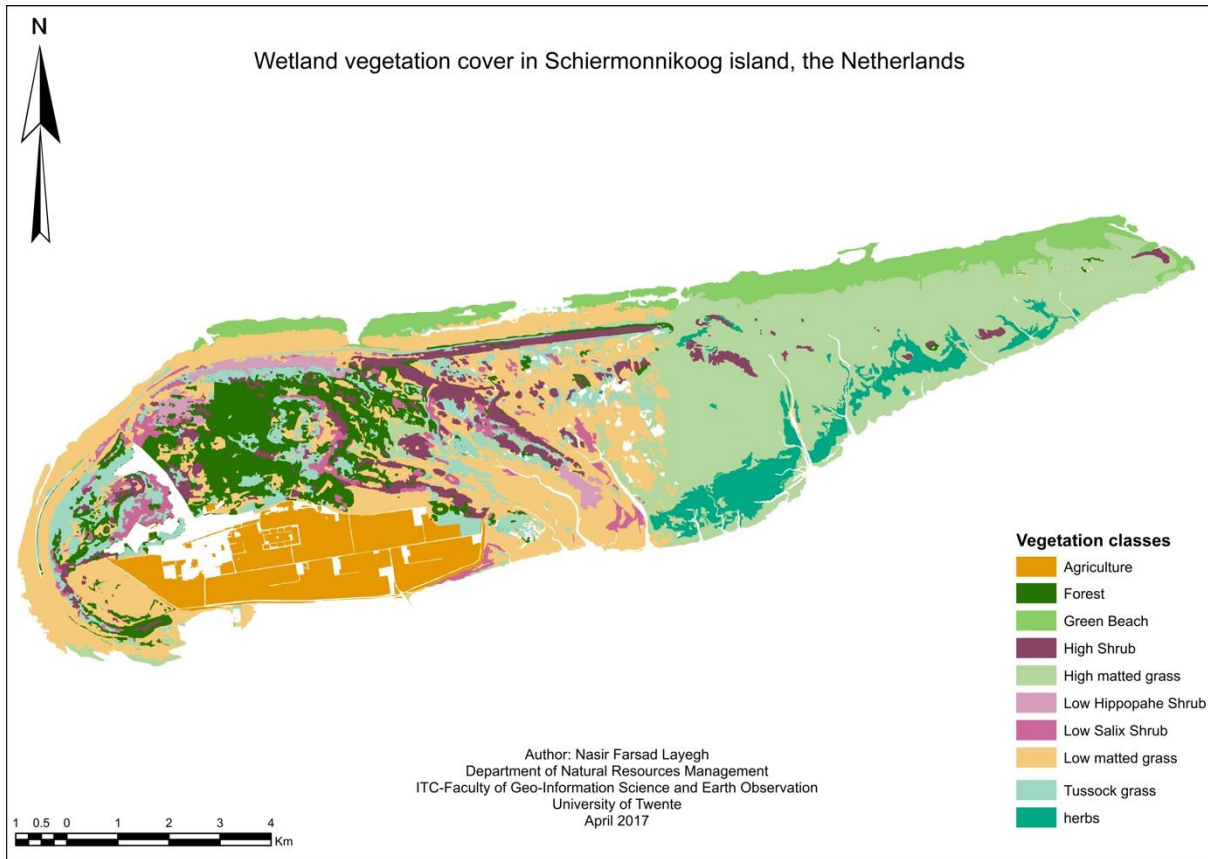


Figure 4. 8. Classified map of the study area with Red-edge spectral bands of Sentinel-2 data using SSLES. The obtained overall accuracy is 83.6%.

#### 4.5. Result of McNemar significance test

To examine the significance of the obtained results, McNemar test at 95% confidence level and 1 degree of freedom is conducted. First, the test is applied on the three groups of band combinations (i.e. groups 1, 4 and 6) as their OA values are close to each other to find out the most informative band combination. The second test is applied on the obtained results from two classification approaches for group 4, with the highest accuracy, to find if using SSLES, can significantly increase the accuracy or not. For each test first, the related confusion matrix was generated (Table 3.1) and then the chi-squared value was calculated (Formula 3.8). Finally, using a chi-squared distribution, P-value was computed for each test. Table 4. 8 represents the results of the three significance tests conducted in this study.

Table 4. 8. Statistically significance tests result on band combinations 1-4, 6-4 and with and without SSLES for band combination4, b.c.: Group of band combination

##### Test 1

Models	Trained classifier with b.c. 1	Trained classifier with b.c. 4
H0	The trained classification model with b.c. 1 and 4 has equal accuracy for predicting true class label	
Result	P_value = 0.5186	chi-squared = 0.4167
	The test result shows no statistically significance difference.	



<b>Conclusion</b>	Accept the H0.	Both models have same classification accuracy.
<b>Test 2</b>		
<b>Models</b>	Trained classifier with b.c. 6	Trained classifier with b.c. 4
<b>H0</b>	The trained classification model with b.c. 6 and 4 has equal accuracy for predicting true class label	
<b>Result</b>	P_value = 1.0	chi-squared = 0.0
	The test result shows no statistically significance difference.	
<b>Conclusion</b>	Accept the H0.	Both models have same classification accuracy.
<b>Test 3</b>		
<b>Models</b>	Classification with SSLES	Classification without SSLES
<b>H0</b>	The trained classification model with SSLES and without SSLES has equal accuracy for predicting true class label.	
<b>Result</b>	P_value = less than 0.0001	chi-squared = 33.8273
	The test result shows statistically significance difference.	
<b>Conclusion</b>	Reject the H0.	Models have different accuracies.

#### 4.6. SSLES performance

To have a better overview of the obtained results and performance of the proposed classification model, the confusion matrices of both classification methods are presented in Figure 4. 9, to report the partial accuracies of every vegetation class. In these two matrices, each row (class)'s cell value is divided by the number of ground truth samples belonging to that class (i.e. producer's accuracy) and results are presented as a map.

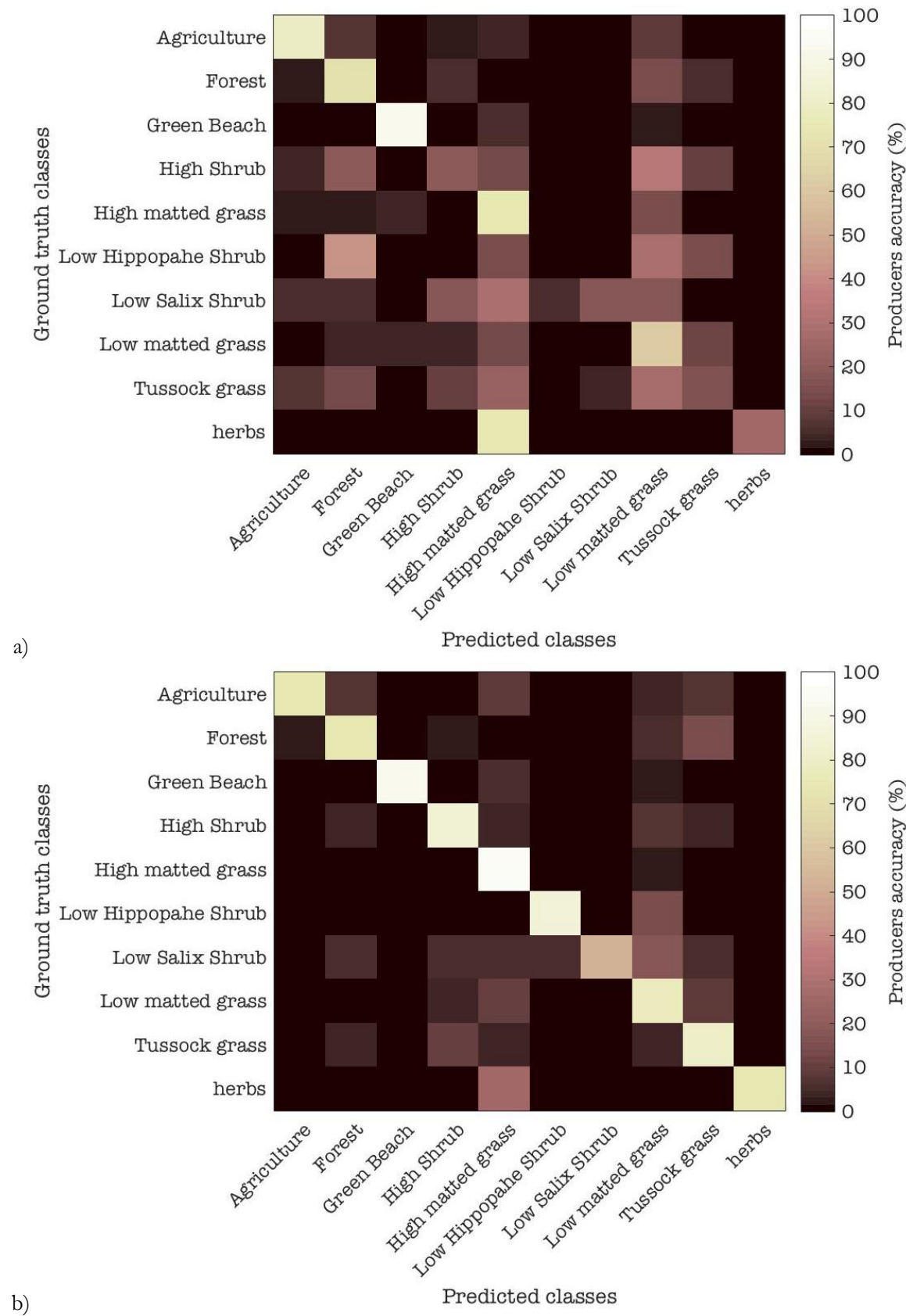


Figure 4. 9. Producer's accuracy matrices for: a) model without SSLES, b) model with SSLES.

In Figure 4. 9, the ideal condition of classification would be, to have the white color in main diagonal and the black color at the other cells, which means 100% of correct classification of the classes and 0% misclassification.

To get a better insight into the SSLES's performance and its improvement over simple RF classification a new map is generated. This map is the result of subtracting the confusion matrices of SSLES and simple RF methods. In the resulting matrix that was presented as a map, if the value of a cell had an increase, it is labeled “**positive change**”, and if the value had a decrease, it is labeled as “**negative change**”, otherwise the cell gets “**no change**” label (Figure 4. 10).

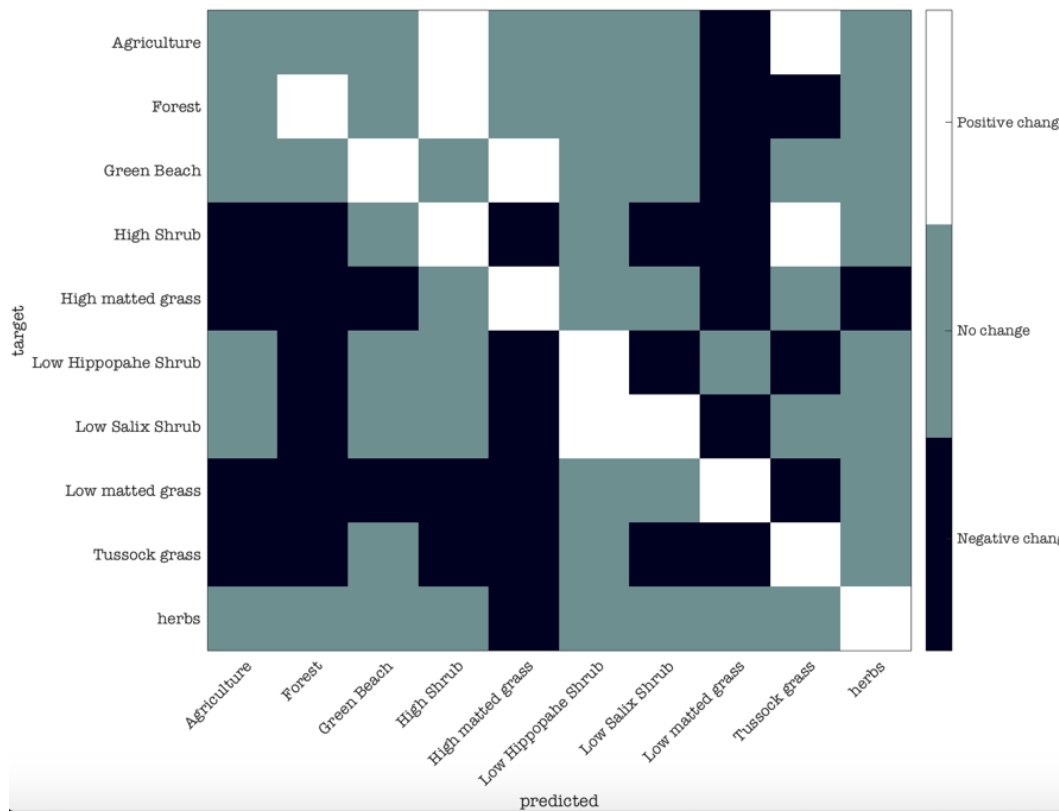


Figure 4. 10. Detailed comparison of two classification approaches, in confusion matrix level.

As it can be observed from Figure 4. 10, all vegetation classes have improvement in their classification accuracy except for “Agriculture” that had no change. In an ideal situation, it is expected that all the positive changes happen in the main diagonal elements, and there is no positive change in the off-diagonal elements, which means not having more misclassified vegetation type. However, in six off-diagonal cells, there is a positive change.

## 5. DISCUSSION

### 5.1. Assessment of SSLES

In this study, SSL and ES algorithms were integrated into a single algorithm for wetland vegetation cover mapping. The results of this study showed that using SSLES in classification approach could improve the classification performance of a standard supervised method significantly. Furthermore, according to Table 4. 7, the highest classification accuracy was achieved using the group of Red-edge bands combination.

According to the results of significance tests, band combinations of group 1 (that includes all Sentinel-2 spectral bands), group 6 (that includes red-edge and shortwave infrared bands) and group 4 (that includes all red-edge bands) did not show statistically significance difference. Consequently, based on the parsimony principle, band combination group 4 could be chosen as the most informative band combination of Sentinel-2 data that results in higher OA value for vegetation classification, as this group contains the least number of spectral bands. This result confirms the findings of Immitzer et al. (2016) in which the high values of the Sentinel-2 red-edge bands for vegetation classification are highlighted.

Based on the result of McNemar's test (Test 3 of Table 4. 8), classification with SSLES generated significantly better results compared to the one "without SSLES". Based on the derived results in Figure 4. 9(a), "Green Beach", "Low matted grass" and "High matted grass", "Agriculture" and "Forest" classes have been classified more accurately than other classes. "Low hippopae shrub" is the less accurate classes with less than 10% of user's accuracy. It can be seen that a considerable number of "Herbs" are classified as "High matted grass" and obviously the classifier had confused the classes "Low matted grass" and "High matted grass" with other classes, because most of the classes have some samples classified as these two classes. However, comparing figure 4.8(a) and figure 4.8(b) reveals the positive effect of SSLES on partial accuracies for all classes. For instance, "Low hippopae shrub" class's accuracy could increase to more than 80%. The lowest accuracy belongs to "Low salix shrub" which is 58%. Additionally, the result of a further analysis in figure 4.9 implies that SSLES can increase the OA without hurting the per-class classification accuracies.

In the following sections, the performance of OBIA, expert system and SSLES, as well as the proposed methodology's limitations and scalability, will be discussed.

### 5.2. Object-based image analysis

In this study, utilizing object-based image analysis could help to avoid some of the problems in pixel-based classification such as noisy pixels, "salt and pepper" effect and spectral mixing (Devadas et al., 2012). However, some points need to be discussed regarding the object-based approach that can affect the obtained classification results:

- 1.1- Segmentation:** (a) Segmentation quality has a direct effect on the classification accuracy. If segmentation has low quality (e.g. having dominant over- or under-segmentation), the delineation of two different adjacent classes may not happen properly in the image, hence the extracted features of an object will be the mix of features of two different classes (Liu & Xia, 2010; Song et al., 2005). To reduce the negative effect of segmentation, some factors such as the level of details in image analysis and geometric characteristics of classes should be taken into consideration (Hamedianfar & Shafri, 2016). However, in the present study, the segmentation results were compared to the reference map polygons to be able to choose the optimum segmented map. Nevertheless, in the case of having uncertainty in the reference vegetation map and consequently in the reference samples there may exist some level of uncertainty in the result of segmentation as well. (b) The other point about segmentation is the spatial resolution of used data. Using finer resolution data combined with 5m RapidEye image could improve the quality of segmentation to some extent.
- 1.2- Features extraction:** As mentioned before, the features that were used in this study were based on literature and their contribution to (wetland) vegetation classification had been proved as well. To our expertise, the extracted number of features were enough to get an acceptable result, furthermore adding more features could have resulted in correlation between features. However, the contribution of more features needs further investigations which was beyond the scope of this study.
- 1.3- Feature selection:** Even though stopping criteria was defined to stop the algorithm when the optimum number of features was reached, the algorithm was set to continue running for the entire feature set to find out the effect of using the whole feature set for classification. Based on the results of SFFS algorithm, three important conclusions can be made. First, according to Figure 4. 4, using whole features degrades the classification accuracy while using a feature selection algorithm and using a subset of feature for classification improves the accuracy. It can be because of the existence of noisy, correlated or unrelated features in the feature set that do not contribute sufficiently to vegetation cover classification. Second, among the selected features, the majority were textural features which show the importance of these features on vegetation classification. Third, it can be seen that using the defined stopping criteria, the algorithm could successfully find the best feature sub-set resulting in the least MCE.

### 5.3. Expert system design

The designed expert system could utilize contextual information and prior knowledge about the study area in terms of an available vegetation map and published scientific resources to discriminate the wetland vegetation types efficiently. The defined role of the expert system could allow the SSL to increase the certainty of labeling potential objects and avoid assigning false class

labels by refining less certain or uncertain class labels. The expert system handles this uncertainty through the use of probability rules that aim to link environmental parameters and the location of vegetation types, where it can be most likely a vegetation type may occur (Skidmore, 1989). For example, herbs are more likely to occur at low elevations and steeper slopes, so that *a priori* information was included in the set of rules (Figure 4. 5). In addition, the expert system has the capability of incorporating other useful parameters to generate a relationship between the data layer and the dependent variable being modeled (i.e. a specific vegetation type). Additional dependent variables can also be generated, if the necessary and required data is available, and the relationships between the environmental variables and the dependent variable such as soil type or soil moisture index are known.

The expert system has two main disadvantages that need to be discussed as well because they may negatively affect its performance. The first disadvantage is the wrong prediction of the expert system that may assign wrong class labels to the image objects due to insufficient rules (Murray et al., 2009). For instance, a specific vegetation type was not known to occur within an image object with specific features. The possible solution to this problem might be the inclusion of additional data layers and establishing the relationship between the dependent variables and new data layers. The second disadvantage of the expert system can be the conflict of knowledge from different sources, which can cause inconsistencies in the existing knowledge, in the knowledge-base, on the factors affecting the decision of expert system. To overcome this problem, the “correctness” of expert system’s prediction needs to be evaluated using the criteria that the sources of knowledge and experts define.

## **5.4. Performance and drawbacks of SSLES**

### **5.4.1. Classification accuracy**

The experimental results show that using SSLES in classification approach can improve the classification performance of a standard supervised method significantly. There can be two reasons behind the performance of SSLES. The first reason lies in constructing the graph in SSL, based on image features rather than conventional methods of using spatial neighborhood and spectral similarity. This could help to obtain a better estimation of the underlying class label of the potential unlabeled image objects. The second reason is probably due to the role of ES in labeling process of SSL, which increased the certainty of labeling by refining samples with uncertain labels assigned. By increasing reliability of SSL’s output samples, the number of training samples increased and RF classifier could learn more from training samples and finally, could provide a better class prediction of unlabeled samples.

However, it was assumed that the reference training samples used in this study were representative and sufficient to be able to train a standard supervised classifier, to have a benchmark to evaluate and compare the performance of the proposed SSLES. Nevertheless, using such high number of training samples in an SSL approach can increase the chance of generalization problem and over-fitting of the classifier. To test the robustness of the proposed SSLES against the mentioned problems a new experiment was conducted on data belonging to

group 4, but with 50% of reference training samples (i.e. 325 reference training samples for 10 vegetation classes) and the result was evaluated by the same test set used before in this study. The OA of this experiment became 80% and compared to the case of using all the training samples no statistically significant difference was observed (Table 5. 1).

*Table 5. 1. McNemar test for two cases of using 100% and 50% of reference training samples. b.c.: Group of band combination*

<b>Models</b>	Trained classifier with b.c. 4 With 100% of reference training samples	Trained classifier with b.c. 4 With 50% of reference training samples
<b>H0</b>	The trained classification model with both datasets has equal accuracy for predicting true class label	
<b>Result</b>	P_value = 0.7272 The test result shows no statistically significance difference.	chi-squared = 0.1216
<b>Conclusion</b>	Accept the H0.	Both models have same classification accuracy.

#### 5.4.2. SSLES drawbacks

In the present study, the obtained results could have been negatively affected by various factors. The first factor to mention is the uncertainties associated with the input data (used for object-based image analysis) and reference data. The issues caused by these uncertainties were already discussed in section 1 of this chapter. The second factor is the limitation of SSLES itself. For instance, there can be some level of limitations in the design and implementation of the expert system as well as SSL. In terms of expert system, assigning the quantitative values for the *a priori* probabilities, had also some uncertainties, since it was based on the reference data. In generating knowledge-base of the expert system, using knowledge and experience of experts which could have resulted stronger knowledge-base, was not possible, due to time constraints.

Regarding SSL, two points need to be discussed about the graph construction. The first point is the effect of parameter kernel width. Lower values for the kernel bandwidth can affect the quality of the constructed graph by removing the edge between two samples as the result of the low weight value and in worse condition, this can result in the singularity of the transpose of weight matrix and failure in the model run. Consequently, the range of 0.1-2 was used to find the optimum value of the kernel parameter. The second point is the use of K-NN method for graph construction. A recent study by (Jebara et al. 2009) has shown that KNN could result in irregular graph construction while theoretically it is expected to generate a regular graph. KNN often generates graphs that each node (i.e. image object in this study) is connected to more than K neighbors. For instance, Figure 5. 1 illustrates a 1-NN case where 5 nodes are connected, while central nodes are connected to 4 nodes. In this case, the algorithm may end up assigning same class labels to 4 nodes, as the class label of central node (Subramanya & Talukdar, 2014).

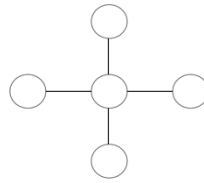


Figure 5. 1. Example of irregular graph in 1-NN case

This problem is more pronounced in the present study because of high similarity problem of vegetation classes. In this case, an image object (representing a specific vegetation type) can be in K-nearest neighbor of more than one labeled image objects (that represent different vegetation types) and can result in confusion of SSL algorithm in assigning the correct class label for that image object.

A closer look into the classification accuracies of each class in terms of user's accuracy, reveals that highest misclassification happens for "Low salix shrub" with 24%. Other classes such as "low hippopahe shrub", "Tussock grass" and "High shrub" had an accuracy of less than 60%. Mainly the classifier had classified these classes wrongly as "low matted grass", "High matted grass" and "Tussock grass" (Figure 5. 2).

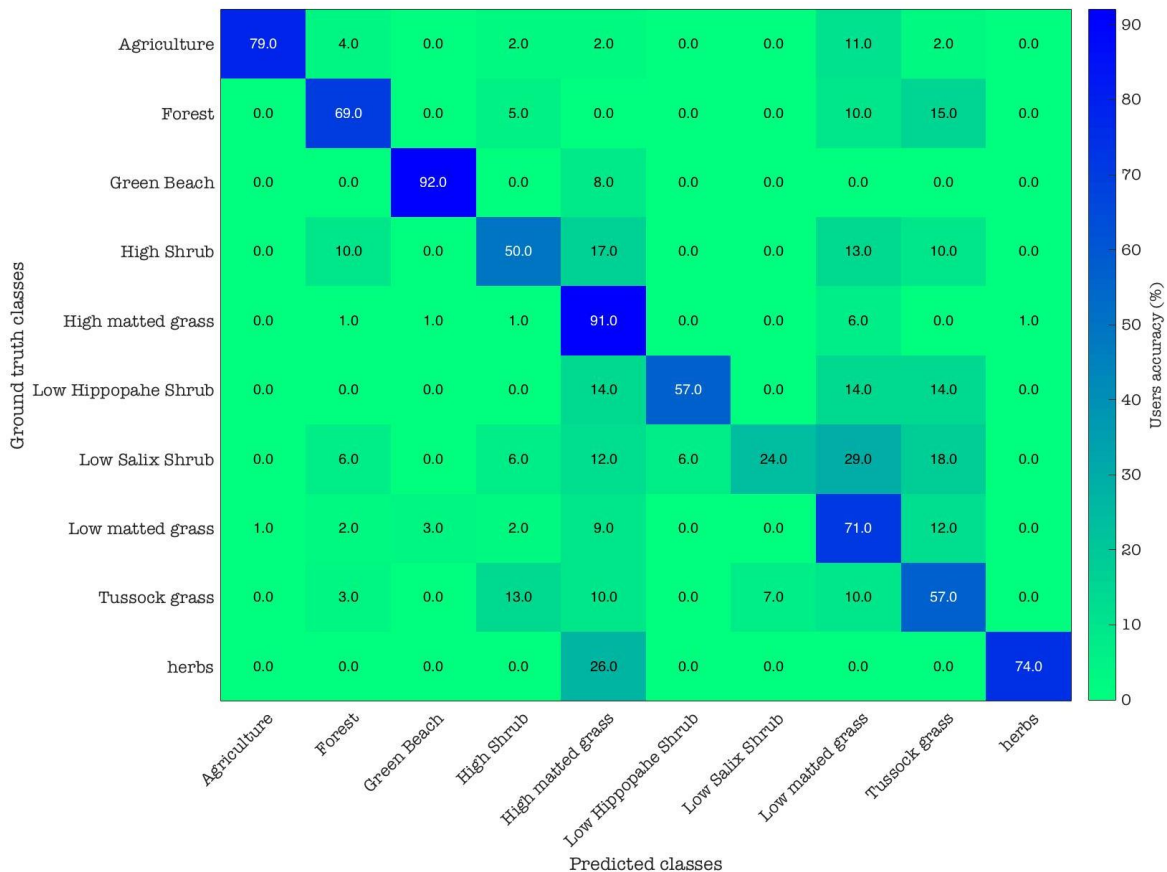


Figure 5. 2. User's accuracy map derived using SSLES



According to Figure 5. 3, feature distance between misclassified classes and “High matted grass” is very small, consequently, as the result of a problem in graph construction, they are being misclassified. Even though the feature distance between the misclassified classes and “Low matted grass” is high enough not to be confused by SSLES, another reason of this misclassification can be mixed features of different classes as the result of poor segmentation results.

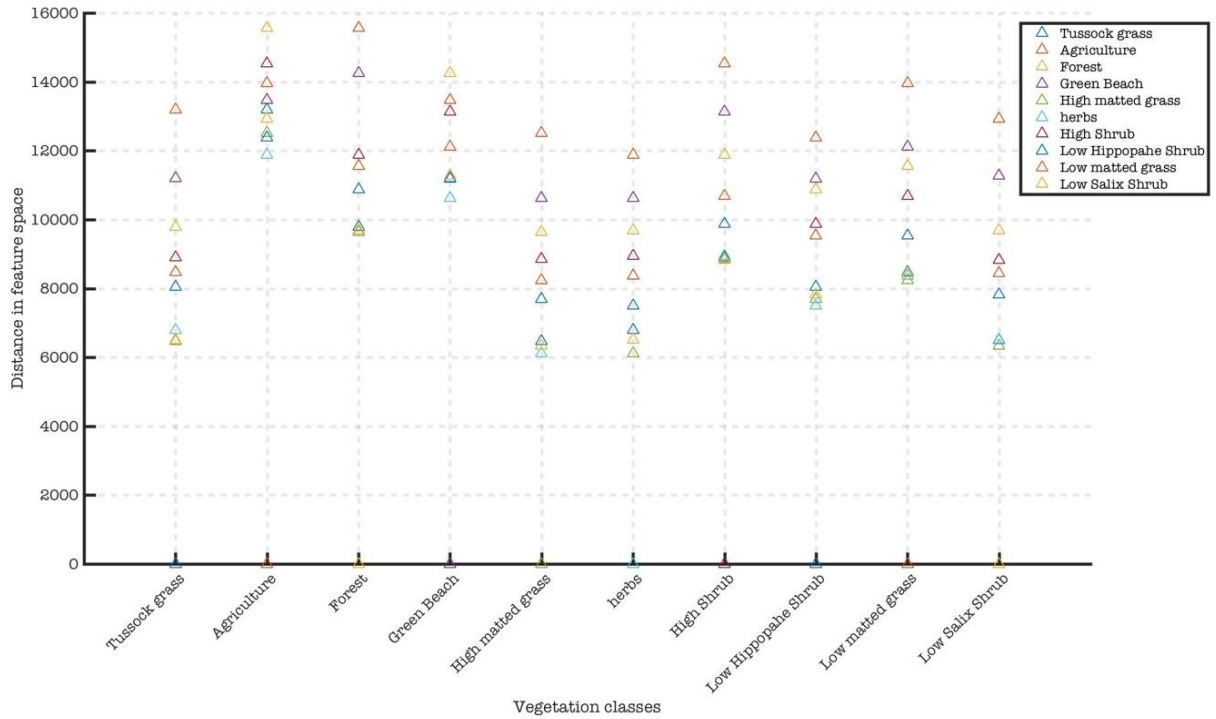


Figure 5. 3. Mean of features distance between 10 vegetation classes. The distance was computed based on Euclidean distance.

### 5.5. The scalability of SSLES

The basic data required to feed SSLES are image objects and related features. As long as the vegetation types (or any other objects on the Earth’s surface) are distinguishable in the used remote sensing data and required features can be extracted, SSLES can work effectively. However, using different remotely sensed data can result in different segmentation quality and also a different number of image features which could affect the final results. Depending on the size of objects on the Earth’s surface, as the resolution of the image used in segmentation is higher, more objects can be identified in the image, furthermore, the border between classes can be drawn more precisely.

The most important part of SSLES is the expert system. Consequently, to be able to use the proposed approach for a wetland vegetation cover classification in a different location, some modification may be required based on the ecology of that study area. In this study, because of time constraints, it was not possible to establish a direct contact with experienced scientists, experts and ecologists who could have knowledge about the vegetation cover in the study area.

Consequently, the assumption of this study is not to exploit an available vegetation map to generate the knowledge-base, rather the assumption is to use any available source of knowledge and prior knowledge. The sources of knowledge can be either human expert or other resources such as scientific researches, published works, etc. made by human experts. But in the case of using the proposed approach for a different application (e.g. to classify different objects on Earth surface) a new knowledge acquisition and rule generation process needs to be designed for expert system based on the new application and new Earth objects to be classified.

## 6. CONCLUSION AND RECOMMENDATIONS

### 6.1. Conclusion

In this study, a new approach for semi-supervised classification of wetland vegetation cover is proposed. The algorithm constructs a graph based on extracted spectral, textural and geometrical image features from OBIA. It uses Euclidean distance to compute the (feature) similarity between image objects where K nearest neighbors are being selected for labeling. This exploits the probable fact that two similar objects may have the same class label. Using expert system to supervise labeling process of semi-supervised learning algorithm was the key point in this study that could increase the certainty of labeling process. The capability of OBIA, semi-supervised learning and expert system for classification, particularly in the field of remote sensing image classification, have already been investigated in the literature. The novel contribution of this study is to integrate these three, in a single algorithm, for vegetation cover classification. Results of this study prove the effectiveness of the proposed algorithm for the challenging problem of vegetation cover classification where some vegetation classes show similar characteristics. The effectiveness of the proposed algorithm is assessed using the recent Sentinel-2 multi-spectral image and the proposed algorithm's performance was compared with a standard supervised classification method (i.e. Random Forest).

Experimental results indicated that the proposed SSLES model performed well. The capability of Sentinel-2 image's bands in vegetation classification was assessed using six important and common band combinations and results proved that band combinations of all Red-edge bands yielded the highest overall accuracy for wetland vegetation cover classification.

McNemar's significance test was also conducted to further validate the superiority of proposed SSLES. According to McNemar significance test, there was no significance difference between the three of band combination groups, i.e. (1) all sentinel-2 red-edge bands, (2) all sentinel-2 bands and (3) sentinel-2 shortwave infrared and red-edge bands.

### 6.2. Recommendations

The potential applicability of the trained model of the proposed approach on different wetlands, as well as different biomes such as forest and croplands with different remote sensing data could be analyzed in a future study. There is an imminent need to investigate different strategies further in order to reliably and in a most informative way select unlabeled samples. Using alternative approaches for graph construction instead of KNN as well as using alternative similarity measure methods such as JM distance should be of high priority for future works. Finally, big data or large-scale technologies may be investigated, in order to apply the presented approach for classification of big data in remote sensing.

## 7. LIST OF REFERENCES

- Agrawala, A. K. (1970). Learning with a probabilistic teacher. In *IEEE Transactions on Information Theory* (p. 16:373–379).
- Aha, D. W., & Bankert, R. L. (1996). A Comparative Evaluation of Sequential Feature Selection Algorithms (pp. 199–206). [http://doi.org/10.1007/978-1-4612-2404-4\\_19](http://doi.org/10.1007/978-1-4612-2404-4_19)
- Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., & Moriarty, T. (2002). A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3), 193–199. <http://doi.org/10.1109/42.996338>
- Álvarez-Meza, A. M., Molina-Giraldo, S., & Castellanos-Dominguez, G. (2016). Background modeling using Object-based Selective Updating and Correntropy adaptation. *Image and Vision Computing*, 45, 22–36. <http://doi.org/10.1016/j.imavis.2015.11.006>
- Anuta, P. E. (1977). COMPUTER-ASSISTED ANALYSIS TECHNIQUES FOR REMOTE SENSING DATA INTERPRETATION. *GEOPHYSICS*, 42(3), 468–481. <http://doi.org/10.1190/1.1440719>
- Beukeboom, T. J. (1976). *The Hydrology of the Frisian Islands*. Amsterdam: Rodopi Bv Editions.
- Bird, E. (2008). *Coastal Geomorphology. An Introduction*. *Igarss 2014*. <http://doi.org/10.1007/s13398-014-0173-7.2>
- Bishop, C. M. (1996). *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. <http://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Blaschke, T., & Strobl, J. (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift Für Geoinformationssysteme*, 14(6), 12–17.
- Board, R., & Pitt, L. (1989). Semi-supervised learning. *Machine Learning*, 4(1), 41–65. <http://doi.org/10.1007/BF00114803>
- Booker, J. M., & McNamara, L. A. (2004). Solving black box computation problems using expert knowledge theory and methods. *Reliability Engineering & System Safety*, 85(1–3), 331–340. <http://doi.org/10.1016/j.ress.2004.03.021>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <http://doi.org/10.1023/A:1010933404324>
- Breiman, L. (1999). *Random forests—random features*. Berkeley.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. CA: Belmont:Wadsworth.
- Bruzzzone, L., Chi, M., & Marconcini, M. (2006). A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11), 3363–3373. <http://doi.org/10.1109/TGRS.2006.877950>
- Cam, L. Le. (1990). Maximum Likelihood: An Introduction. *International Statistical Review / Revue Internationale de Statistique*, 58(2), 153. <http://doi.org/10.2307/1403464>
- Camps-Valls, G., Bandos Marsheva, T. V., & Zhou, D. (2007). Semi-Supervised Graph-Based Hyperspectral Image Classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(10), 3044–3054. <http://doi.org/10.1109/TGRS.2007.895416>
- Chapelle, O., Schölkopf, B., & Zien, A. (2010). *Semi-Supervised Learning*. England-London: The MIT Press.
- Clinton, N. (2016). An accuracy assessment measure for object based image segmentation AN ACCURACY ASSESSMENT MEASURE FOR OBJECT BASED IMAGE, (October), 1189–1194.
- Cochran, W. G. (1964). On the Performance of the Linear Discriminant Function. *Technometrics*, 6(2), 179–190.
- Cochran, W. G. (1977). *Sampling techniques*. New York: John Wiley and Sons.
- Comaniciu, D., & Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE*

- Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.  
<http://doi.org/10.1109/34.1000236>
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [http://doi.org/10.1016/0034-4257\(91\)90048-B](http://doi.org/10.1016/0034-4257(91)90048-B)
- Congalton, R. G., & Green, K. (2008). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Second Edi). Boca Raton, FL, USA: CRC Press.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <http://doi.org/10.1109/TIT.1967.1053964>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and other kernel-based learning methods*. New York, NY, USA: Cambridge University Press.
- Dalponte, M., Ene, L. T., Marconcini, M., Gobakken, T., & Næsset, E. (2015). Semi-supervised SVM for individual tree crown species classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110, 77–87. <http://doi.org/10.1016/j.isprsjprs.2015.10.010>
- Dalponte, M., Ørka, H. O., Ene, L. T., Gobakken, T., & Næsset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sensing of Environment*, 140, 306–317. <http://doi.org/10.1016/j.rse.2013.09.006>
- Darvishzadeh, R., Atzberger, C., Skidmore, A. K., & Abkar, A. A. (2009). Leaf Area Index derivation from hyperspectral vegetation indices and the red edge position. *International Journal of Remote Sensing*, 30(23), 6199–6218. <http://doi.org/10.1080/01431160902842342>
- De Silva, A. M., & Leong, P. H. W. (2015). Grammar-Based Feature Generation for Time-Series Prediction, 105. <http://doi.org/10.1007/978-981-287-411-5>
- Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, 11(12), 7063–7081. <http://doi.org/10.3390/s110707063>
- Devadas, R., Denham, R. J., Pringle, M., Centre, R. S., & Percinct, E. (2012). Support Vector Machine Classification of Object-Based Data for Crop Mapping , Using Multi-Temporal Landsat Imagery, XXXIX(September), 185–190. <http://doi.org/10.5194/isprsarchives-XXXIX-B7-185-2012>
- Dopido, I., Li, J., Marpu, P. R., Plaza, A., Bioucas Dias, J. M., & Benediktsson, J. A. (2013). Semisupervised Self-Learning for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7), 4032–4044. <http://doi.org/10.1109/TGRS.2012.2228275>
- Egbert, S. L., Park, S., Price, K. P., Lee, R.-Y., Wu, J., & Duane Nellis, M. (2002). Using conservation reserve program maps derived from satellite imagery to characterize landscape structure. *Computers and Electronics in Agriculture*, 37(1–3), 141–156. [http://doi.org/10.1016/S0168-1699\(02\)00114-X](http://doi.org/10.1016/S0168-1699(02)00114-X)
- ESA. (2015). SENTINEL-2 User Handbook, (1), 64. <http://doi.org/GMES-S1OP-EOPG-TN-13-0001>
- Foody, G. M. (2004). Thematic Map Comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5), 627–633. <http://doi.org/10.14358/PERS.70.5.627>
- Fralick, S. C. (1967). Learning to recognize patterns without a teacher. In *IEEE Transactions on Information Theory* (p. 13:57–64).
- Franklin, S. E. (2001). *Remote Sensing for Sustainable Forest Management*. Boca Raton, FL, USA: CRC Press.
- Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., ... Jin, X. (2017). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological Indicators*, 73, 105–117. <http://doi.org/10.1016/j.ecolind.2016.09.029>
- Georgescu, Shimshoni, & Meer. (2003). Mean shift based clustering in high dimensions: a texture classification example. In *Proceedings Ninth IEEE International Conference on Computer Vision* (pp. 456–463 vol.1). IEEE. <http://doi.org/10.1109/ICCV.2003.1238382>
- Gilmore, M. S., Wilson, E. H., Barrett, N., Civco, D. L., Prisloe, S., Hurd, J. D., & Chadwick, C. (2008).

- Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh. *Remote Sensing of Environment*, 112(11), 4048–4060.  
<http://doi.org/10.1016/j.rse.2008.05.020>
- Grenier, M., Labrecque, S., Benoit, M., & Allard, M. (2008). Accuracy assessment method for wetland object-based classification. *GEOBLA 2008 - Pixels, Objects, Intelligence*, 6.
- Gu, Y., & Feng, K. (2012). L1-graph semisupervised learning for hyperspectral image classification. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1401–1404). IEEE.  
<http://doi.org/10.1109/IGARSS.2012.6351274>
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, (3), 1157–1182.
- Hamedianfar, A., & Shafri, H. Z. M. (2016). Integrated approach using data mining-based decision tree and object-based image analysis for high-resolution urban mapping of WorldView-2 satellite sensor data. *Journal of Applied Remote Sensing*, 10(2), 25001. <http://doi.org/10.1117/1.JRS.10.025001>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 610–621.  
<http://doi.org/10.1109/TSMC.1973.4309314>
- Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM*, 28(9), 921–932.  
<http://doi.org/10.1145/4284.4286>
- Hayes-Roth, F., Waterman, D., & Lenat, D. (1983). *Building Expert Systems*. MA, USA: Addison-Wesley, Reading.
- Huang, X., & Zhang, L. (2008). An Adaptive Mean-Shift Analysis Approach for Object Extraction and Classification From Urban Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12), 4173–4185. <http://doi.org/10.1109/TGRS.2008.2002577>
- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, 4(12), 2661–2693.  
<http://doi.org/10.3390/rs4092661>
- Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, 8(3), 166.  
<http://doi.org/10.3390/rs8030166>
- Jackson, Q., & Landgrebe, D. A. (2001). An adaptive classifier design for high-dimensional data analysis with a limited training data set. *IEEE Transactions on Geoscience and Remote Sensing*, 39(12), 2664–2679.  
<http://doi.org/10.1109/36.975001>
- Jebara, T., Wang, J., & Chang, S.-F. (2009). Graph construction and b -matching for semi-supervised learning. *International Conference on Machine Learning (ICML)*, 1–8.  
<http://doi.org/10.1145/1553374.1553432>
- John, G. H. , & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In *UAI'95 Proceedings of the Eleventh conference on Uncertainty in artificial intelligence* (pp. 338–345). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Karakos, D., Khudanpur, S., Eisner, J., & Priebe, C. E. (2005). Unsupervised classification via decision trees: An information-theoretic perspective. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, V(1). <http://doi.org/10.1109/ICASSP.2005.1416495>
- Keith, D. A., Martin, T. G., McDonald-Madden, E., & Walters, C. (2011). Uncertainty and adaptive management for biodiversity conservation. *Biological Conservation*, 144(4), 1175–1178.  
<http://doi.org/10.1016/j.biocon.2010.11.022>
- Kendal, S., & Creen, M. (2007). *An Introduction to Knowledge Engineering* (p. 290). London: Springer London. <http://doi.org/10.1007/978-1-84628-667-4>
- Kim, K.-H., & Choi, S. (2014). Label propagation through minimax paths for scalable semi-supervised

- learning. *Pattern Recognition Letters*, 45, 17–25. <http://doi.org/10.1016/j.patrec.2014.02.020>
- Kim, W., & Crawford, M. M. (2010). Adaptive Classification for Hyperspectral Image Data Using Manifold Regularization Kernel Machines. *IEEE Transactions on Geoscience and Remote Sensing*. <http://doi.org/10.1109/TGRS.2010.2076287>
- Klecka, W. R. (1980). *Discriminant analysis* (No. 19). University of Cincinnati: Sage publications.
- Landgrebe, D. (2002). Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, 19(1), 17–28. <http://doi.org/10.1109/79.974718>
- Laurent, V. C. E., Schaepman, M. E., Verhoef, W., Weyermann, J., & Chavez, R. O. (2014). Bayesian object-based estimation of LAI and chlorophyll from a simulated Sentinel-2 top-of-atmosphere radiance image. *Remote Sensing of Environment*, 140, 318–329. <http://doi.org/10.1016/j.rse.2013.09.005>
- Levine, M. D., & Nazif, A. M. (1985). Dynamic Measurement of Computer Generated Image Segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2), 155–164. <http://doi.org/10.1109/TPAMI.1985.4767640>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, (2), 18–22.
- Lindenmayer, D. B., & Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution*, 24(9), 482–486. <http://doi.org/10.1016/j.tree.2009.03.005>
- Liu, D., & Xia, F. (2010). Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, 1(4), 187–194. <http://doi.org/10.1080/01431161003743173>
- Ma, L., Ma, A., Ju, C., & Li, X. (2016). Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification. *Pattern Recognition Letters*, 83, 133–142. <http://doi.org/10.1016/j.patrec.2016.01.022>
- Marcano-Cedeno, A., Quintanilla-Dominguez, J., Cortina-Januchs, M. G., & Andina, D. (2010). Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society* (pp. 2845–2850). IEEE. <http://doi.org/10.1109/IECON.2010.5675075>
- Mathieu, R., Aryal, J., & Chong, A. k. (2007). Object-Based Classification of Ikonos Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas. *Sensors*, 7(11), 2860–2880. <http://doi.org/10.3390/s7112860>
- Maulik, U., & Chakraborty, D. (2011). A self-trained ensemble with semisupervised SVM: An application to pixel classification of remote sensing imagery. *Pattern Recognition*, 44(3), 615–623. <http://doi.org/10.1016/j.patcog.2010.09.021>
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153–157. <http://doi.org/10.1007/BF02295996>
- Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 155–168. <http://doi.org/10.1016/j.isprsjprs.2015.03.014>
- Möller, M., Lymburner, L., & Volk, M. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 9(3), 311–321. <http://doi.org/10.1016/j.jag.2006.10.002>
- Mui, A., He, Y., & Weng, Q. (2015). An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 30–46. <http://doi.org/10.1016/j.isprsjprs.2015.08.005>
- Murray, J. V., Goldizen, A. W., O’Leary, R. A., McAlpine, C. A., Possingham, H. P., & Choy, S. L. (2009). How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *Journal of*

- Applied Ecology*, 46(4), 842–851. <http://doi.org/10.1111/j.1365-2664.2009.01671.x>
- Novelli, A., Aguilar, M. A., Nemmaoui, A., Aguilar, F. J., & Tarantino, E. (2016). Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *International Journal of Applied Earth Observation and Geoinformation*, 52, 403–411. <http://doi.org/10.1016/j.jag.2016.07.011>
- Ouyang, Z.-T., Zhang, M.-Q., Xie, X., Shen, Q., Guo, H.-Q., & Zhao, B. (2011). A comparison of pixel-based and object-oriented approaches to VHR imagery for mapping saltmarsh plants. *Ecological Informatics*, 6(2), 136–146. <http://doi.org/10.1016/j.ecoinf.2011.01.002>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <http://doi.org/10.1080/01431160412331269698>
- Patrick, E., & Costello, J. (1970). On unsupervised estimation algorithms. *IEEE Transactions on Information Theory*, 16(5), 556–569. <http://doi.org/10.1109/TIT.1970.1054534>
- Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6), 1301–1316. <http://doi.org/10.1016/j.rse.2011.01.009>
- Persello, C., & Bruzzone, L. (2014). Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11), 6937–6956. <http://doi.org/10.1109/TGRS.2014.2305805>
- Pham, L. T. H., Brabyn, L., & Ashraf, S. (2016). Combining QuickBird, LiDAR, and GIS topography indices to identify a single native tree species in a complex landscape using an object-based classification approach. *International Journal of Applied Earth Observation and Geoinformation*, 50, 187–197. <http://doi.org/10.1016/j.jag.2016.03.015>
- Pranger, D. P., & Tolman, M. E. (2012). *Toelichting Bij De Vegetatiekartering Schiermonnikoog Op Basis Van False Colour-luchtfoto's 1: 10.000 [Explanation to the Vegetation Mapping Schiermonnikoog 2010 on the Basis of False Colour Aerial Photographs 1:10.000, in Dutch]*. Delft, The Netherlands.
- Radoux, J., & Defourny, P. (2008). Quality assessment of segmentation results devoted to object-based classification. In *Object-Based Image Analysis* (pp. 257–271). Berlin, Heidelberg: Springer Berlin Heidelberg. [http://doi.org/10.1007/978-3-540-77058-9\\_14](http://doi.org/10.1007/978-3-540-77058-9_14)
- Repaka, S. R., Truax, D. D., State, M., Kolstad, E., & Manager, C. (2004). Comparing spectral and object based approaches for classification and transportation feature extraction from high resolution multispectral imagery. *ASPRS Annual Conference*, (May). <http://doi.org/10.1.1.84.3803>
- Richards, J. A. (2013). *Remote Sensing Digital Image Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg. <http://doi.org/10.1007/978-3-642-30062-2>
- Rohban, M. H., & Rabiee, H. R. (2012). Supervised neighborhood graph construction for semi-supervised classification. *Pattern Recognition*, 45(4), 1363–1372. <http://doi.org/10.1016/j.patcog.2011.09.001>
- Rundquist, D. C., Narumalani, S., & Narayanan, R. M. (2001). A review of wetlands remote sensing and defining new considerations. *Remote Sensing Reviews*, 20, 207–226.
- Schenk, J., Kaiser, M., & Rigoll, G. (2009). Selecting Features in On-Line Handwritten Whiteboard Note Recognition: SFS or SFFS? In *2009 10th International Conference on Document Analysis and Recognition* (pp. 1251–1254). IEEE. <http://doi.org/10.1109/ICDAR.2009.130>
- Schmidt, K. S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85(1), 92–108. [http://doi.org/10.1016/S0034-4257\(02\)00196-7](http://doi.org/10.1016/S0034-4257(02)00196-7)
- Schmidt, K. S., Skidmore, a. K., Kloosterman, E. H., van Oosten, H., Kumar, L., & Janssen, J. a. M. (2004). Mapping Coastal Vegetation Using an Expert System and Hyperspectral Imagery. *Photogrammetric Engineering & Remote Sensing*, 70(6), 703–715. <http://doi.org/10.14358/PERS.70.6.703>
- Schöpfer, E., & Lang, S. (2006). Object fate analysis – a virtual overlay method for the categorization of object transition and object-based accuracy assessment. *1st International Conference on Object-Based Image*



*Analysis.*

- Scudder, H. J. (1965). Probability of error of some adaptive pattern-recognition machines. In *IEEE Transactions on Information Theory* (p. 11;363–371).
- Shahshahani, B. M., & Landgrebe, D. A. (1994). The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5), 1087–1095. <http://doi.org/10.1109/36.312897>
- Skidmore, A. (1989). An expert system classifies eucalypt forest types using thematic mapper data and a digital terrain model. *Photogrammetric Engineering and Remote Sensing*, 55(10), 1449–1464.
- Song, M., Civco, D. L., & Hurd, J. D. (2005). A competitive pixel-object approach for land cover classification. *International Journal of Remote Sensing*, 26(22), 4981–4997. <http://doi.org/10.1080/01431160500213912>
- Subramanya, A., & Talukdar, P. P. (2014). *Graph-Based Semi-Supervised Learning*. Morgan & Claypool Publishers.
- Szumner, M., & Jaakkola, T. (2001). Partially labeled classification with Markov random walks. *Advances in Neural Information Processing Systems*, 945–952. <http://doi.org/10.1.1.25.1955>
- Tadjudin, S., & Landgrebe, D. (2000). Robust parameter estimation for mixture model. *IEEE Trans. Geosci. Remote Sens.*, 38(1), 439–445.
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature Selection for Classification: A Review. *Data Classification: Algorithms and Applications*, 37–64. <http://doi.org/10.1.1.409.5195>
- Tiede, D., Langa, S., & Hoffmann, C. (2006). Supervised and forest type-specific multi-scale segmentation for a one-level-representation of single trees. *1st International Conference on Object-Based Image Analysis*, 36(September 2002), 4.
- Tigges, J., Lakes, T., & Hostert, P. (2013). Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of Environment*, 136, 66–75. <http://doi.org/10.1016/j.rse.2013.05.001>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [http://doi.org/10.1016/0034-4257\(79\)90013-0](http://doi.org/10.1016/0034-4257(79)90013-0)
- Vrieling, A., Skidmore, A. K., Wang, T., Meroni, M., Ens, B. J., Oosterbeek, K., ... Paganini, M. (2017). Spatially detailed retrievals of spring phenology from single-season high-resolution image time series. *International Journal of Applied Earth Observation and Geoinformation*, 59, 19–30. <http://doi.org/10.1016/j.jag.2017.02.021>
- Wang, L., Hao, S., Wang, Q., & Wang, Y. (2014). Semi-supervised classification for hyperspectral imagery based on spatial-spectral Label Propagation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97, 123–137. <http://doi.org/10.1016/j.isprsjprs.2014.08.016>
- Wang, S., Lu, J., Gu, X., Du, H., & Yang, J. (2016). Semi-supervised linear discriminant analysis for dimension reduction and classification. *Pattern Recognition*, 57, 179–189. <http://doi.org/10.1016/j.patcog.2016.02.019>
- Wang, X., Gao, Y., & Cheng, Y. (2016). A non-negative sparse semi-supervised dimensionality reduction algorithm for hyperspectral data. *Neurocomputing*, 188, 275–283. <http://doi.org/10.1016/j.neucom.2014.12.127>
- Weih, R. C., & Riggan, N. D. (2010). Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII, 1–6.
- Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1(1), 9–23. <http://doi.org/10.1093/jpe/rtn005>
- Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric*

*Engineering and Remote Sensing*, 72(7), 799–811. <http://doi.org/10.14358/PERS.72.7.799>

Zhang, Y. J. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8), 1335–1346. [http://doi.org/10.1016/0031-3203\(95\)00169-7](http://doi.org/10.1016/0031-3203(95)00169-7)

Zhu, X., Ghahramani, Z., & Lafferty, J. (2003). Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *Proceedings of the 20th International Conference on Machine Learning* (pp. 912–919).