

MULTIPLE-POINT GEOSTATISTICS FOR IMAGE RE-CLASSIFICATION

CHUNHUI SHEN
February, 2016

SUPERVISORS:
Dr. V.A. Tolpekin
Prof.Dr.Ir. Alfred Stein



MULTIPLE-POINT GEOSTATISTICS FOR IMAGE RE-CLASSIFICATION

CHUNHUI SHEN

Enschede, the Netherlands, February, 2016

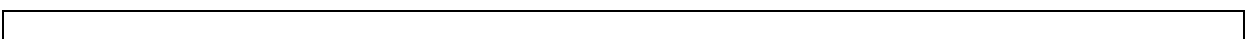
Thesis submitted to the Faculty of Geo-Information
Science and Earth Observation of the University of
Twente in partial fulfilment of the requirements for the
degree of Master of Science in Geo-information Science
and Earth Observation.
Specialization: Geoinformatics

SUPERVISOR:
Dr. V.A. Tolpekin
Prof.Dr.Ir. Alfred Stein

THESIS ASSESSMENT BOARD:
Prof.dr.ir. M.G. Vosselman (chair)
Dr.ir.L.J.Spreeuwers (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

A support vector machine (SVM) and multiple point geostatistics (MPS) combined classification method was developed and evaluated based on the MPS algorithm direct sampling (DS). DS is a MPS algorithm that can be used to simulate the spatial heterogeneity, meanwhile DS can stochastically rebuild the heterogeneous spatial structure of the training image. Training image is a representation of a spatial domain, it describes the spatial characteristics of a conceptual spatial structure, and it is essentially a database that stores the spatial patterns. SVM classify multi-spectral images based on defined spectral signature, MPS reproduces the physical reality in stochastic modelling by borrowing spatial patterns from the training image. As both SVM posterior probability and MPS class frequency represent classification certainty, SVM classification results and MPS simulations are combined by considering the SVM posterior probability and the MPS class frequency at the same time. MPS class frequency may provide high spatial classification certainty to the spatial domain which has low SVM spectral classification certainty. In this situation, the SVM classification accuracy could be improved.

ACKNOWLEDGEMENTS

I would like to take this chance to thank for all people who supported me during my MSc study period.

To my supervisors Dr. V.A. Tolpekin and Prof.Dr.Ir. Alfred Stein, I would like to express my sincere gratitude for your theoretical and technique support. It was my honour and pleasure to work with you.

To my family, thanks for your love and support.

TABLE OF CONTENTS

1.	Introduction.....	1
1.1.	Motivation and problem statement.....	1
1.2.	Research identification.....	3
1.2.1.	Research objectives.....	3
1.2.2.	Research questions.....	3
2.	Literature review.....	4
2.1.	Theoretical Background.....	4
2.2.	SNESIM post-processing methods.....	4
2.3.	Downscaling remotely imagery using ATPCK and MPS.....	5
2.4.	Pixel-based classification method works for very high resolution image of heterogeneous areas.....	5
2.5.	Modified form of direct sampling.....	5
2.6.	Pattern-based MPS algorithm based on the cross-correlation.....	6
3.	Concept and Methodology.....	7
3.1.	Direct Sampling.....	7
3.2.	Support vector machines.....	10
4.	Implementation.....	11
4.1.	Project set up.....	11
4.1.1.	Study area.....	11
4.1.2.	Data description.....	11
4.2.	Multiple point geostatistics application.....	11
4.2.1.	Selection of the experimental area.....	11
4.2.2.	Generation of SVM classification of the experimental area.....	12
4.2.3.	Generation of the training image.....	12
4.2.4.	Generation of the conditioning data.....	12
4.2.5.	MPS classification that uses DS algorithm.....	13
4.2.6.	Combination of the result of SVM classification and MPS classification.....	13
4.3.	Programing.....	14
5.	Results.....	15
5.1.	Effect of Training image selection.....	15
5.2.	Effect of conditioning data selection.....	18
5.2.1.	Select the conditioning data manually.....	18
5.2.2.	Select the conditioning data randomly.....	20
5.2.3.	Use pixels of high confidence as conditioning data.....	21
5.3.	Effect of apply the method to different experimental area.....	22
5.3.1.	Experimental area No.2.....	22
5.3.2.	Experimental area No.3.....	25
5.4.	Effect of using different MPS algorithm Parameters.....	27
5.4.1.	Maximum extension of the data events.....	28
5.4.2.	Maximum number of points in the data event.....	29
5.4.3.	Maximum fraction of scanned training image.....	29
5.4.4.	Type of distance variable and distance threshold.....	30
5.4.5.	Number of MPS simulations.....	30
6.	Discussion.....	32
6.1.	Interpretation of different inputs.....	32
6.2.	Evaluation of the method.....	33
7.	Conclusion and recommendation.....	35
7.1.	Conclusion.....	35

7.2. Recommendation	37
8. References.....	38

LIST OF FIGURES

Figure 1. Example of farm fields in Netherlands (Left) & Malis.....	2
Figure 2 WorldView-2 image of the Mali study site in growing season of May-November 2014	2
Figure 3. Assign conditioning data to simulation grid.....	8
Figure 4. How to define the data event of simulating node.....	8
Figure 5. a, b, c, d are the minimum and maximum value within the location vector.....	9
Figure 6. Compute the distance between the data event found in SG and data event.....	9
Figure 7. Assign the value of the simulating node in TI.....	10
Figure 8. WorldView-2 satellite image of training image area.....	15
Figure 9. Training set of training image area. It was collected in multispectral image.....	16
Figure 10. SVM classification result of training area.....	16
Figure 11. Manually drawn training image of the training area.....	17
Figure 12. Comparison of MPS classification of different training image.....	17
Figure 13. Worldview-2 satellite image of experimental area No.1.....	18
Figure 14. Example of manually selected Conditioning data,	18
Figure 15. Training set of training experimental area No.1.....	19
Figure 16. Comparison between SVM and MPS classification of the experimental area No.1.....	19
Figure 17. SVM and MPS combined classification of experimental area No.1	20
Figure 18. Two different MPS classification of experimental area No.1.....	21
Figure 19. Comparison between SVM classification result and MPS classification of experimental area No. 1.	22
Figure 20. SVM and MPS combined classification result of experimental area No.1	22
Figure 21. WorldView-2 satellite image of experimental area No. 2 in Mali.	23
Figure 22. Training set of experimental area No.2.....	23
Figure 23. SVM classification result of experimental area No. 2.....	24
Figure 24. MPS classification result of experimental area No. 2	24
Figure 25. SVM and MPS combined classification result of experimental area No. 2.....	25
Figure 26. Worldview-2 satellite image of experimental area No.3.....	25
Figure 27. Training set of experimental area No.3.....	26
Figure 28. SVM classification result of experimental area No. 3.....	26
Figure 29. MPS classification result of experimental area No. 3	27
Figure 30. SVM and MPS combined classification result of experimental area No. 3	27
Figure 31. MPS classification results with different number of simulations.....	31

LIST OF TABLES

Table 1 Classification results by using previous manually selected conditioning data (Figure.13).	19
Table 2 Experiment of different number of random conditioning data, OA and Kappa is the average value of 5 MPS classification results	21
Table 3. Experiments of different Maximum Extension applied to Experimental area No.1. OA and Kappa is the average value of 5 MPS classification results	28
Table 4. Experiment of different Maximum Extension applied to Experimental area No.2. OA and Kappa is the average value of 5 MPS classification results	28
Table 5 Experiment of different maximum number of points applied to Experimental area No.1. OA and Kappa is the average value of 5 MPS classification results.....	29
Table 6. Experiment of different maximum number of points applied to Experimental area No.2. OA and Kappa is the average value of 5 MPS classification results.....	29
Table 7. Experiment of different Maximum fraction of scanned training image. OA and Kappa is the average value of 5 MPS classification results.....	30
Table 8. Experiment of different number of MPS simulations. OA and Kappa is the average value of 5 MPS classification results	30

1. INTRODUCTION

1.1. Motivation and problem statement

With the growth of the world population, there is a rising food challenge: food production has to increase 60% by 2050 (FAO, 2009). To deal with impending food crises, the developing agriculture in Africa can be a viable solution. Agriculture in Africa remains highly undeveloped. Twenty five percent of the world's arable land is in Africa, but generates only 10 percent of the global agricultural output (Kariuki, 2011). In Sub-Saharan Africa, there are 183 million hectares currently used in agriculture, and some additional 451 million hectares of available arable land. This is more than the remaining arable land resource from the rest of the world. Aside from the development of Africa's agriculture, the other solution could be to improve the farming methods of the smallholders. Smallholder farmers is widely known to include small farmers who do not own or control the land they farm, most of them are dependent on family labour, and they produce on relatively small plots of land. They produce two thirds of the worlds food, but most of their farming methods are inefficient, their yield is low and they are not usually included in food production chains (de By & Zurita-Milla, 2015).

Spurring a Transformation for Agriculture through Remote Sensing (STARS) is a research project aimed at improving agricultural practices in Sub-Saharan Africa and South Asia by using modern remote sensing technology. In poor countries of Sub-Saharan Africa and South Asia, there are agriculture challenges such as unproductive soil, crop diseases, pests, flood and drought, which encumber farmers to produce consistently and sustainably. These challenges lead to difficulties in understanding the condition of the crop yield, the production level, the seasonal outlook, and marketing, which in turn lead to poor decision-making. For example, inaccurate assessment of food production leads to oversupply, causing food prices to plummet (STARS, 2015). Image analysis is applied to extract useful information from vast amount of data collected by using modern remote sensing technology, advices, suggestions and plans can be provided to farmers. It helps them to make better decisions on their agriculture practices, such as what type of cops to plant, when to plant, and where to plant them. This may bring better yields that increse the income and advance the livelihood of the farmers. With the information provided to higher level management departments, they can manage national food supply more effectively.

One important technical difficulty has prevented remote sensing technology to be used in many areas of STARS project. The technical challenge is how to deal with the identification of heterogeneity in spatial reality (STARS, 2015). Classification of the farm lands provides better understanding of farm practices and agricultural productions (Piiroinen et al., 2015). Classification always resulted in a degree of certainty which represents the similarity of an object with other samples from that class. However, there is a key issue of uncertainty which refers to what is 'known imperfectly', the uncertainty includes input variables, their locations and spatial supports, the model structure and parameters (Zhao et al., 2011). Unidentified spatial heterogeneity which is unrecognized complex spatial structure that distribute within a spatial domain leads to low classification certainty, it is not wise to make decisions based upon uncertain information. Geostatistics was applied to analysis and predict the spatial probability distribution of the unidentified spatial structure. However, traditional two-point spatial structure geostatistics is relying on the assumption of stationarity (Taylor & Eckley, 2013). It assumes that spatial statistical variation is similar over the entire spatial domain (Mariethoz & Caers, 2014). Very high resolution (VHR) satellite images are needed to observe

the heterogeneity from the sky. VHR satellite imagery provides sub-meter resolution, and is one of the highest quality images currently available from commercial remote sensing satellites. The heterogeneity is assessed by the complex texture in the images. Image texture constitutes patterns within an image region that act as visual property (Petrou & García Sevilla, 2006). In fact, because of lack of unified agriculture management and machinery power in poor countries, each farm field is managed by a single smallholder with manual labour. Their farm fields have irregular boundaries and farmers cannot remove objects like trees or big stones inside the farm fields. Farms have mixed levels of fertility causing differences in yield at the farm level.



Figure 1. Example of farm fields in Netherlands (Left) & Mali (Source: Google Earth)

Compared to the unified regular farm fields in developed countries, the lands belonging to poor smallholders are harder to visually identify (Figure 1). For example, the left image is an area of farm fields of Netherlands, it shows clear boundaries and homogeneous visual property. The right image consists two crop fields of sorghum and peanuts. In this image, the farm fields are with significant variation. The farm fields have significant variation including irregular boundaries, unremoved trees, the grows of same crop performance differently in one field, and other unidentified component. VHR satellite images are multi-spectral which can recognize various features not easily identifiable visually. However, visualization of some crops can be irregular and might generate totally different results using multi-spectral images (Figure 2). It is hardly possible to make the assumption of spatial stationarity in this context. Therefore, two-point spatial structure are expected to fail in description of the internal field variability.



Figure 2 WorldView-2 image of the Mali study site in growing season of May-November 2014(de By & Zurita-Milla, 2015)

1.2. Research identification

In order to solve the problem of heterogeneous farm lands classification, multiple-point geostatistics (MPS) is the main method that will be explored in this thesis. Other classification methods will be used to evaluate the classification result of MPS.

1.2.1. Research objectives

The main objective of this research is to implement and evaluate MPS in VHR satellite imagery for an heterogeneous area. Following are the sub-objectives:

- Implementation of SVM and MPS combined classification method for very high resolution satellite images.
- Classification of the VHR satellite image of different area during the growing season.
- Interpretation of results from classification methods.

1.2.2. Research questions

- Which heterogeneous area would be the study area?
- Which MPS algorithm is most suitable for this research?
- How to generate suitable training image?
- How to find patterns in training image?
- Depends on if the training image is classified or not, will MPS be applied to pixel class labels or pixel spectral values?
- What are the optimal parameter settings to obtain the best classification result?
- Does MPS provide better classification result of heterogeneous areas rather than support vector machine (SVM)?

2. LITERATURE REVIEW

This chapter covers the theoretical background and previous work relate to this study.

2.1. Theoretical Background

Traditional two-point statistics tools are variogram based statistical tools reproducing spatial structures by describing the dissimilarity of variables that observed at two separate spatial locations. The variogram essentially is a statistical device, it represents the spatial patterns in a mathematical form. It can be defined as the variance of the difference between field values across the fields. In traditional statistics, linear unbiased estimator like kriging is used for interpolation. However, the complexity of the patterns is severely limited. There are only two points are considered at the same time, it cannot capture mathematically the complexity of some spatial structures by describing merely correlations between two spatial locations. Two-point spatial structure is relying on the assumption of the spatial statistical variation is similar over the entire spatial domain. Similar experimental variograms could be produced by various strongly different types of reservoir heterogeneities. (Caers, 2004). Therefore, multiple-points geostatistics (MPS) are proposed for this situation. The first introduction to MPS was published in 1993 (Guardiano & Srivastava, 1993), which provided a basic conception for higher covariances statistics using training images. The training image is a representation of the physical reality, it acts as a database of spatial patterns. MPS uses multiple-point patterns from the training image to reproduce the physical reality in stochastic modeling. Correlations between multiple spatial locations are considered at a time. The training image replaces the variogram in MPS as a solution of spatial heterogeneity, it is much more intuitive since one can observe how the spatial heterogeneity could be visualized, prior to geostatistical simulation or estimation (Caers, 2004). MPS focus on modeling the spatial patterns of natural phenomena, which are examples of the heterogeneity that is expected to be present (Mariethoz & Lefebvre, 2014).

Several MPS algorithms are already proposed in literature, these algorithms can be roughly divided into two kinds of methods: pixel-based and pattern-based method. Among the pixel-based method, Normal Equation Simulation (NESIM) algorithm (Guardiano & Srivastava, 1993) is first algorithm based on MPS, however, it was not used widely because of the high CPU cost at that time. Then, Single Normal Equation Simulation (SNESIM) algorithm (Strebelle, 2002) is still widely used since it is sufficient, it provides better results than NESIM with reasonable computation speed, SNESIM uses a tree to store the spatial patterns. Based on SNESIM, Impala algorithm (Straubhaar et al., 2011), High order simulation (HOSIM) algorithm (Mustapha & Dimitrakopoulos, 2011) and Growthsim algorithm (Eskandari & Srinivasan, 2013) are proposed. Besides, direct sampling (Mariethoz et al., 2010) and simulated annealing (Peredo & Ortiz, 2011) are the two other methods that without the tree to store the spatial patterns. Stochastic Simulation with Patterns (SIMPAT) algorithm (Arpat & Caers, 2007), Filtersim (Zhang et al., 2006) algorithm, Patchwork Simulation (El Ouassini et al., 2008), and cross correlation based simulation (CCSIM) algorithm (Tahmasebi et al., 2012) are the main pattern-based methods. The pattern-based methods have better computation speed than pixel-based method.

2.2. SNESIM post-processing methods

Tang et al (2013) give an application of MPS involved classification method, the application applied MPS simulation for post-processing in order to provide better overall accuracy of a remote sensing classification. In this study, the study area was located in Serere district in Uganda. It represented by a sub scene of Landsat Thematic Mapper (TM) image that covers 9 km × 9 km area. Multiple Point Geostatistics (MPS) was used

to post process the result of maximum likelihood classification (MLC). The result of MLC was chosen as the training image, and processed by using MPS algorithm SNESIM. MPS post-processing methods are also compared with two different method for post-processing: traditional spatial filter and Markov random field method. After the comparison, MPS classification result has 89.65% overall accuracy (OA) which was beyond the other three methods. But compared to 89.46% OA of MRF, the improvement was 0.19%, it was very limited.

2.3. Downscaling remotely imagery using ATPCK and MPS

Tang et al (2015) presents a study using MPS to test if the spatial structure are remained during downscaling of the remote sensing image. In this study, area-to point cokriging (ATPCK) is used to deal with the problem where MPS cannot work alone at. The method was applied to two datasets. The study area was located in Serere district in Uganda with 9 km by 9 km sub scene. Other data set was based on QuickBird image. It was obtained from Kokilari Lagoon, Sri Lanka. The image covers a 1.12 km by 1.12 km sub scene. MPS using Pattern-based algorithm FILTERSIM was applied, to provide a set of simulation results. Multispectral image with a coarse spatial resolution which was used as the input band for ATPCK was taken as the training image for MPS simulation. The result of the study shows that MPS can keep the spatial structure. Thus, it is able to predict super-resolution image. The improved MPS simulation using ATPCK prediction for data integration can generally achieve a more accurate result than a simulation which does not account the change of support problem for the in the downscaling process. This study could be a good example of how to use MPS to process images of different resolution.

2.4. Pixel-based classification method works for very high resolution image of heterogeneous areas.

Piironen et al (2015) studied the highly heterogeneous crops of agriculture in sub-Saharan Arica area. The study area was 1 km × 3.5 km was located in the foot hills of Taita Hills, these high resolution images were from Aisa EAGLE pushbroom scanner and Nikon D3. They produce images with 64 bands and 0.6 m x 0.6 m resolution. The study focused on how accurately crop species in the study area can be classified using IS data with support vector machine (SVM). High spatial and spectral resolution images are classified using SVM. The conclusion of this paper is that very high spatial resolution image and pixel-based classification method works well in high heterogeneous areas. The OA for the classification was 90.8%. But there is a problem, those high resolution images could be very expensive and may not capable for large area. Thus, the smallholder farm owner in sub-Sahara cannot afford this data.

2.5. Modified form of direct sampling

Mariethoz et al (2013) proposed a modified form of direct sampling DS method. Original DS method is a pixel-based MPS method that was proposed in 2010 (Mariethoz et al., 2010). The algorithm looks for similar patterns in the training image. The most distinguishing feature when comparing with the previous methods like SNESIM is that the central node of the first found similar pattern in the training image is pasted in the simulation. Comparing to the original DS based on pasting on pixel at a time, this modified method provides the simulation which is done by pasting a bunch of nodes at a time. It is it is a pattern-based DS algorithm instead the original pixel based algorithm. It combines the core of the pixel based DS algorithm and the fast computation speed of other pattern-based MPS methods.

2.6. Pattern-based MPS algorithm based on the cross-correlation

Tahmasebi et al (2012) proposed a pattern-based MPS algorithm named cross-correlation simulation (CCSIM). Compared with the previous FILTERSIM method that depends on distance functions, CCSIM replaces the distance functions with the cross-correlation measures. Rather than process the whole pattern at the same time, CCSIM reproduce the spatial structure point by point. Therefore, it also provides significant improvement in the CPU performance, the computation time cost could be reduced by of 30%-80%, memory requirements is also reduced. A new approach that was invented to have better condition of training image, it is based on adaptive recursive template splitting. By comparing with FILTERSIM method in practical, CCSIM reproduces the connectivity and complexity of the spatial patterns better.

3. CONCEPT AND METHODOLOGY

This chapter describes the concept and the methodology that was used to fulfil the specified objectives of this research.

3.1. Direct Sampling

Direct sampling (DS) is an MPS algorithm that is statistically equivalent to previous MPS algorithms, e.g. SNESIM. It reproduces the spatial structure by borrowing spatial patterns from the training image. Furthermore, it allows MPS to deal with continuous variables and multivariate problems (Mariethoz et al., 2010). As DS can stochastically rebuild the heterogeneous spatial structure of the training image. This method can be used to simulate the geological heterogeneity. Therefore, DS algorithm is suitable for this research.

The input data of the DS algorithm includes:

- Location of nodes \mathbf{x} in a blank simulation grid (SG), SG is simply an empty spatial grid that will be filled with the spatial structure that is required to be reproduced by DS.
- All nodes \mathbf{y} of the training image (TI), training image is a representation of a spatial domain that describes the spatial characteristics of a conceptual spatial structure which is expected to be reproduced in the SG.
- A set of N conditioning data $Z(\mathbf{x}_i), i \in [1, \dots, N]$, conditioning data is assigned to the SG at the start of the simulation. DS search in the training image to find spatial patterns that match the spatial patterns which are formed by the conditioning data and simulated nodes.
- Maximum extension of the data event e , it describes the maximum extent of the spatial pattern.
- Maximum number of points in the data event n , it describes the maximum number of points in the spatial pattern.
- Threshold value t , t describes the acceptable difference in the comparison between the spatial patterns.
- Maximum fraction of the training image f . f describes the maximum percentage of the training image that could be simulated

The output of the algorithm is the simulated SG $Z(\mathbf{x})$.

The following lists the principle of the DS algorithm as introduced by the author (Mariethoz et al., 2010):

1. Assign the conditioning data to the SG. Each node in the conditioning data is assigned to its closest node in the SG. If the resolution or projection of the conditioning data is not the same with the designed SG, there could be more than one node in conditioning data should be assigned to the same node in the SG. In that case, the closest node in conditioning data will be used. Figure 6 shows an example of how the conditioning data is assigned to the SG.

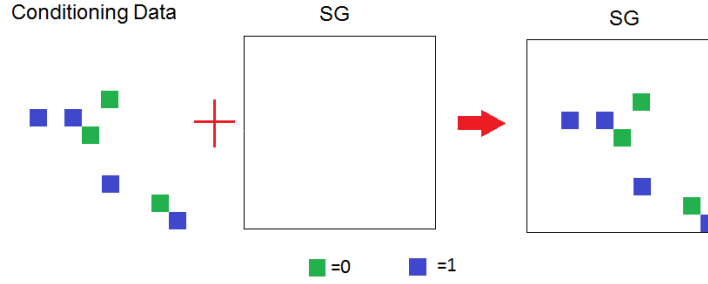


Figure 3. Assign conditioning data to simulation grid. In this example, the Conditioning data includes two types of nodes, green nodes has the variable value 0, and blue nodes has variable value 1.

2. A path is defined through the remaining nodes in the SG which so far no value was assigned. The path is a vector that contains the index of all the remaining nodes. This path can be defined differently, e. g. random, or unilateral (all nodes a visited with a regular order from one side of the grid). Random path is used in this research, the index is generated randomly. For each successive located node x in the path of SG:

- a. Obtain the data event of simulating node x , a data event is a representation of the spatial pattern in mathematical form. It is a set that includes the neighbours of the simulating node x . The neighbours are the closest nodes $\{x_1, x_2, x_3, \dots, x_n\}$ that are assigned or simulated in the SG. If no neighbour node is available, a node in TI is randomly took and its value is assigned to the simulating node in SG. In the example of figure 4, the maximum extension of the data events is 2 pixels. The found data event contains 3 neighbour nodes.

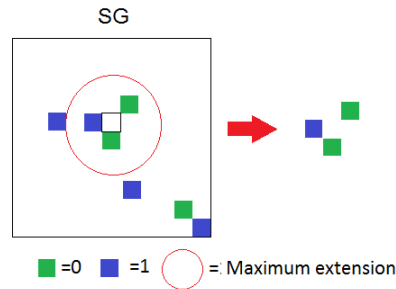


Figure 4. How to define the data event of simulating node. The small black box is the simulating node. The example data event of the simulating node includes one node with variable value 1 and two nodes with variable value 0.

- b. Define the collection of location vectors of the data event: $L = \{h_1, h_2, \dots, h_n\} = \{x_1 - x, x_2 - x, \dots, x_n - x\}$. These location vectors define the relative locations of all the neighbourhood nodes. In the example of figure 4, this vector is: $((1, 1), (-1, 0), (0, -1))$.
- c. Define the value vector $d_n\{Z(x + h_1), Z(x + h_2), \dots, Z(x + h_n)\}$ of the data event. It is a vector that contains the variable values of all the neighbour nodes. In the example of figure 4, this data event is: $(0, 1, 0)$.
- d. Define the search window in the TI. It is a collection of the locations of the nodes in TI that will be simulated. The minimum and maximum values of x and y of the location vectors defines the margin of the search window. In the example of figure 5, the yellow box represents an example of the search window.

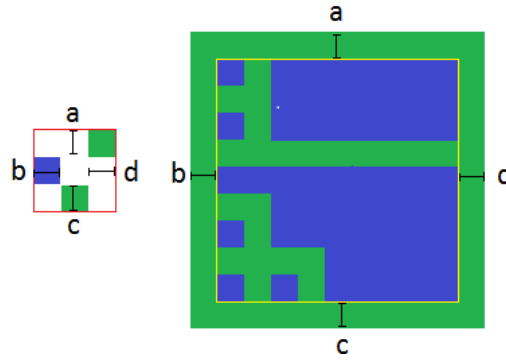


Figure 5. a, b, c, d are the minimum and maximum value within the location vector. The margin of the search window in training image is defined by a, b, c, d.

- e. Start from a random location and scan the whole search window lineally, for each node in the search window:
 - i. Find the nodes with the locations of the described data event in the training image. A random grid node could be selected in the search window.
 - ii. Calculate the distance $d\{d_n(x), d_n(y)\}$ between the data event x found in the SG and the data event y found in the TI. The distance of continuous or categorical variables is calculated in various ways. Because the class labels are used instead of continuous numbers in this research, the distance in this research is categorical distance. Here following equation to calculate the categorical distance:

$$d\{d_n(x), d_n(y)\} = \frac{1}{n} \sum_{i=1}^n a_i \in [0,1],$$

$$\text{where } a_i = \begin{cases} 0 & \text{if } Z(x_i) = Z(y_i) \\ 1 & \text{if } Z(x_i) \neq Z(y_i) \end{cases}$$

In the example of Figure 6, the distance between the data events is the distance between $Z(x) = (0, 1, 0)$ and $Z(y) = (1, 0, 0)$. Therefore its categorical distance is $2/3$.

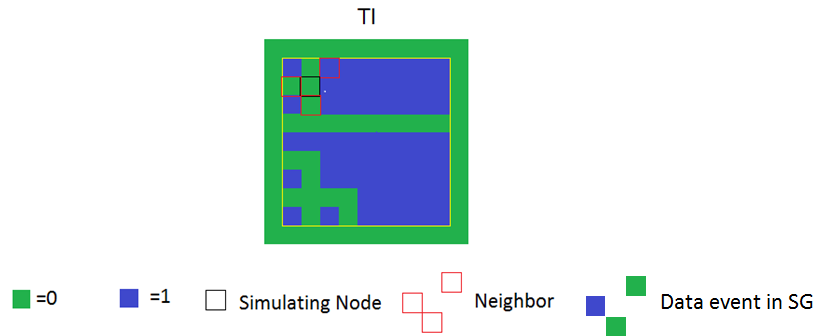


Figure 6. Compute the distance between the data event found in SG and data event of the simulating Node in TI. The data event $Z(x)$ found in SG has the variable value of $(0, 1, 0)$, the variable value of the data event $Z(y)$ of the simulating node in TI is $(1, 0, 0)$.

- iii. If the data event found in SG and the current data event in TI has the lowest distance so far, the value and the distance of the central node of the current data event is stored.

- iv. If the data event found in SG and the current data event in TI has the distance that is smaller than the threshold value t , the value of the node in search window will be sampled and assigned to SG. In the example of Figure 7, the data events found in TI and SG are the same, the value of the node in TI is sampled and assigned to SG.

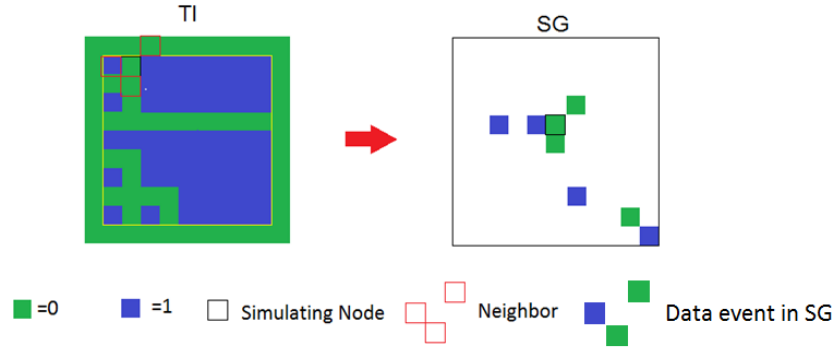


Figure 7. Assign the value of the simulating node in TI that with the matching data event to the simulating node in SG.

- v. If the simulations of finding the data event in training image processed a certain fraction of TI that is equal larger than maximum fraction of the training image f . The node in search window with the lowest distance will be accepted and its value will be assigned to SG.

3.2. Support vector machines

MPS reproduces the spatial structure, but it cannot classify the images with its own algorithm. Therefore, a classifier is needed in this situation. SVM with linear kernel is used in this research to classify the image before the MPS process. The SVM classification results and the MPS simulations are combined to produce the final classification result.

Support vector machines is a supervised learning models in remote sensing for machine learning (Steinwart & Christmann, 2008), which is widely used in classification and regression analysis. It provides a classification on high-dimensional feature space through kernel trick. In order to combine the MPS simulations with the SVM classification result, the posterior probability of SVM result is needed. It represents the degree of certainty of each class. However, Standard SVM does not produce the posterior probability. Posterior probability of the SVM result could be generate based on external method, e.g. Platt's method (Platt, 1999).

4. IMPLEMENTATION

This chapter presents the implementation of the SVM and MPS combined classification method that adopted to fulfil the specified objectives of this research.

4.1. Project set up

4.1.1. Study area

The study area of this research is in a village in Mali which is part of the STARS project. With the STARS project is aimed at improving agricultural practices in Sub-Saharan Africa and South Asia, there are plenty of data available in these areas. The study area contains many small farm fields of different crops. It is full of spatial heterogeneity: unclear boundaries, unknown crop types, trees, shadows, etc.

4.1.2. Data description

VHR satellite images of different seasons were selected for classification of the different type of crops. These images were obtained by Digital Globe's WorldView-2 sensor. The images were taken in May, July and October of 2014 in 2x2 meter resolution satellite image with 8 bands. Band 7/5/3 were used to display RGB bands while all bands were used for classification. There was also 0.5x0.5 meter resolution panchromatic image that available for comparison.

4.2. Multiple point geostatistics application

The quality of MPS simulations was decided based on the quality of training image and conditioning data. It was important to find suitable training image and generate reasonable conditioning data. Because of too much uncertainty within the study area, it was difficult to identify all the areas with different crop types. This MPS application only focus on identifying the maize fields from other type of crops, trees and bare soil. Therefore, tree, maize, bare soil and other crops are the four classes required to be classified. The application contains 6 steps:

- Selection of the experimental area.
- Generation of SVM classification of the experimental area.
- Generation of draw the training image.
- Generation of the conditioning data.
- Generation of the MPS simulations with obtained training image and conditioning data.
- Generation of the final classification by combine SVM classification estimation and MPS simulations.

4.2.1. Selection of the experimental area

By combining the satellite image with ground observed data, 10 maize fields were located. The combination of these maize fields are called experiment set. After viewing the experiment set, the common characteristics of these fields are identified: these maize fields were polygons with different shapes; they have trees which are growing inside the fields; some of the boundaries between maize fields, other crop fields and bare soil areas are not always regular; there are also buildings near the maize fields.

Buildings are out of the scope of this study, therefore the 3 maize fields near buildings were removed from the experiment set. There was also some unidentified area that could be shrubs or groves. In the satellite image that was observed in 7/5/2 RGB bands, these unidentified patterns has the similar spectral feature with trees. But they didn't have similar patterns with trees. The spatial patterns in training image and

simulation grid are expected to be similar. Therefore 3 maize fields near these shrubs or groves look like fields were also removed from the experiment set.

There were 4 maize fields remaining in the experiment set. They were spatially similar with one another. Visually, they have similar patterns of trees, fields and bare soil. In order to use the training image to generate MPS simulations, one maize field would be used to generate the training image (Figure 8), the other three were to be selected in the experimental area (Figure 12, Figure 19, Figure 23)..

4.2.2. Generation of SVM classification of the experimental area

Posterior probability of SVM classification is required in this study. For each pixel, the SVM posterior probability represent degree of certainty of each class by considering the similarity between the pixel and other pixels were classified to these classes. This posterior probability is used to combine with the MPS class frequency (chapter 4.2.6), and it is considered to be used to obtain pixels of high confidence which can be used as conditioning data (chapter 4.2.4.a). However, the standard SVM does not generate posterior probability. Platt's method was used to approximate the posterior probability in this research. Function "predict" (R Core Team, 2014a) of the R core package "statues" (R Core Team, 2014c) were used to implement this feature.

There have been studies that compared the performance between the different kernels of SVM. Generally, SVM with linear kernel almost never have higher classification accuracy than SVM with non-linear kernels, e. g. Gaussian kernel (Keerthi, 2003). However, linear SVM generally has better computation speed compare to non-linear SVM. In order to reduce the computation time, linear kernel SVM was used in this research.

Inputs of the SVM included the subset of satellite image of the experimental area, training set, test set and SVM regularization parameter C. Training set and test set consisted of the samples for training and test, and they were manually selected from the satellite image. Envi software was used to manually select the training set and test. Optimal value of C for different experimental areas was examined individually.

4.2.3. Generation of the training image

To generate the MPS simulations of the experimental area, a training image that is similar to the experimental area is required. MPS algorithms compare patterns in simulation grid with those in the training image in order to retrieve frequencies. Therefore, MPS algorithms are required to computing similarity between patterns. In order to generate a training image that contains the similar spatial patterns with the experimental area. SVM is used to classify the maize field that is selected in the previous step.

The SVM classification presents the spatial patterns of this field. However, the SVM classification is not the perfect, given that could be some obvious errors within. To avoid the SVM classification errors, the final training image was manually drawn on the basis of SVM results and visual interpretation of VHR satellite image.

4.2.4. Generation of the conditioning data

Because there was no available ground observed data, there was the need to generate conditioning data instead of observed conditioning data. There are few ways to generate the conditioning data:

- a. Select the conditioning data manually: select the conditioning data manually depends on what is

observed in the satellite image. It is similar with using the ground observed data, there could be mistakes when selecting the data.

- b. Select conditioning data randomly: select the pixels randomly from the SVM classification result: it is an automatic way. However, its quality relies on the accuracy of the SVM classification result. Correct and wrong classified pixels have the same possibility to be the conditioning data. It provides lots of uncertainty.
- c. Use pixels of high confidence as conditioning data: pixels of high confidence are the pixels that have high possibility to be classified correctly by SVM. For each pixel, posterior probability of the SVM result is used to judge if the pixel is confident or not. For each class, a percentage of pixels with the higher posterior probability are assumed to be classified correctly. For every MPS simulation, part of these pixels are selected randomly to be the conditioning data.

This conditioning data selection method avoids the manually work. It generated reasonable pixels that has great possibility to be classified by SVM correctly. Only a random part of these pixels are selected for each simulation also ensure the stochasticity of the MPS simulations.

4.2.5. MPS classification that uses DS algorithm

Direct sampling algorithm was used to generate MPS simulations of the experimental area. It generates the stochastic simulations. Therefore, MPS simulations with same training image, conditioning data and algorithms were not the same. MPS classification is not traditional classification, it is a kind of estimation of the classification. In order to generate one MPS classification result, different MPS simulations with same inputs were combined. These simulations were combined by using the class frequency. For each pixel, the class with the maximum class frequency was set to be the final result. For example, if a pixel is simulated to the same class 6 times in 10 MPS simulations, and no other class has higher class frequency then it, its value will be assigned to the MPS classification result.

4.2.6. Combination of the result of SVM classification and MPS classification

MPS generates the stochastic simulations with spatial information while SVM generates the image by using spectral information. In order to represent the classification probability of SVM classification result, posterior probability of the SVM result was calculated based on Platt's method. The posterior probability represents the classification certainty, and the classification certainty was treated as the classification probability. The class frequency of several MPS simulations is normalized to be the class probability of MPS classification. For example, if a pixel is classified to tree 5 times in 10 MPS simulations. The class probability of tree in this pixel is 50%.

In order to consider the spatial information and spectral information at the same time, the result of SVM classification and MPS simulation was combined by adding the SVM classification probability with the MPS classification probability. In each pixel, class probability represented the possibility of which class the pixel would be classified. When combining the two probabilities, there was no extra emphasis placed on either two sides. In each pixel, two probabilities were added together to get the combined class probability, and the pixel would be classified to the class that with the highest combined class probability. It is possible to set two class probabilities to have different fraction of the combined class probability.

4.3. Programing

The code is based on the implementation of the pixel based DS that was implemented by Mariethoz (Mariethoz, 2016). There was a plan to use the patch based DS code instead of the pixel based DS code. However, the patch based DS implementation has more than two thousands lines of code, and there was not enough time to learn it. The pixel based code was written in MATLAB (Mariethoz, 2016). This code is basic and slow, and it was not recommended to be used for benchmarking or comparison with other methods.

By considering flexible programming environment and familiar geostatistics packages, pixel based DS code was rewritten and was run in R environment (R Core Team, 2014b). R was purposely designed and implemented to make data analysis and statistics clearer and easier. R packages were used in this study includes: *rgdal* (Bivand et al., 2015), *MASS* (Venables & Ripley, 2002), *kernlab* (Karatzoglou et al., 2004), *rgl* (Adler & Murdoch, 2016), *raster* (Hijmans, 2015), *Rcpp* (Eddelbuettel & Francois, 2011). The MPS implementation of this study applied a loop over all the pixels in the simulation grid to obtain their data event. Within this loop, the MPS code looped all the pixels within in the entire search window to check if the data event found in training image was the same with the data event obtained in the simulation grid. The introduced loop within another loop is called nested loop. R is not a fast programing language. Loops have a bad rap in R, and they have a reputation for being slow. Nested loops computed even much slower. Therefore, the processing speed of the R version direct sampling code was slow, while the original MATLAB version takes around 50 seconds to process a MPS simulation of a 100×100 simulation grid with a 50×50 training image. The R version took around 6 minutes to finish it.

In order to improve the computation speed of this R version direct sampling code, the loop within the loop of the R code was replaced with a C++ function. R and C++ code was integrated by using the R package *Rcpp*. *Rcpp* allows C++ import and export many R data types and objects equivalently. This feature facilitates writing the code, and also makes the integration of other third party libraries easier (Eddelbuettel et al., 2016). The rewritten part of code loops all the pixels in the training image. Which has improved the computation speed a lot. With the nested loop replaced with C++ function, it can take around 0.0007 seconds to finish one loop of the part of the code that was rewritten to C++ with the previous simulation grid and training image. It is faster than the same part of MATLAB code which takes around 0.001 seconds to process the same loop. The entire process takes around 44 seconds which is also faster than the MATLAB version. Thus, it is expected to be faster after all the code has been rewritten in C++.

5. RESULTS

This chapter represents the MPS classification results with different training image, conditioning data and algorithm parameters.

Because of MPS simulations are stochastic, the MPS classification results with the same training image, conditioning data and MPS algorithm parameters are not the same. The classification overall accuracy (OA) and kappa are used to evaluate the classification result. In order to estimate the OA and kappa of the classification with the same training image, conditioning data and MPS algorithm parameters, average of OA and kappa of 5 classification results are used.

Here following the default input of the MPS classification experiments. It is the classification of experimental area No.1. It use manually drawn training image. Pixels of high confidence are used as conditioning data.

The default value of the algorithm parameters are:

- Maximum extension of the data events: 20 pixels
- Maximum number of points in the data event: 10 pixels
- Maximum fraction of scanned training image: 1
- Type of distance variable: categorical
- Number of MPS simulations: 10

5.1. Effect of Training image selection

In order to have a training image that has the similar spatial patterns with the simulating grid. SVM classification result of the selected area is used to be the training image. The Figure 8 shows the WorldView-2 satellite image of training image area.

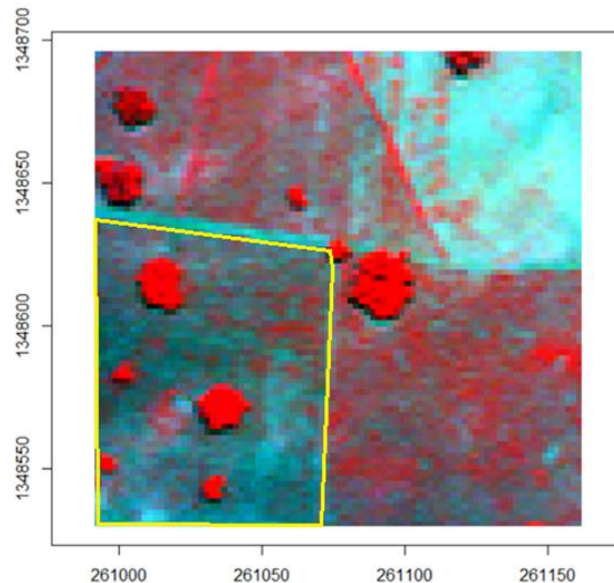


Figure 8. WorldView-2 satellite image of training image area. Band 7/5/2 was used to present the R/G/B bands. The known maize field are displayed with the yellow box.

In order to apply SVM classification to the training area, training set of the training area were selected in multispectral image using ENVI software.

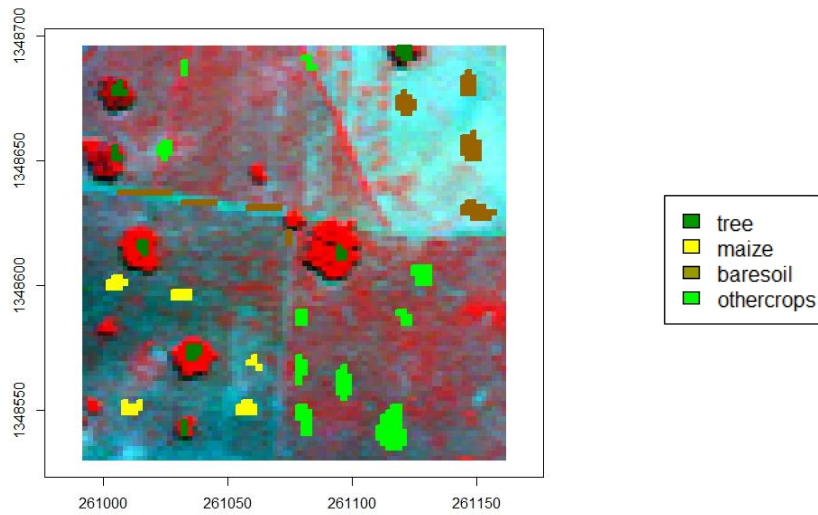


Figure 9. Training set of training image area. The training set was collected in multispectral image.

Since the Figure 10 represents some obvious errors within the SVM classification result, these errors will also be represented in the MPS simulations. For example: maize was not supposed to grow out of the maize field, but there are isolated pixels was classified as maize.

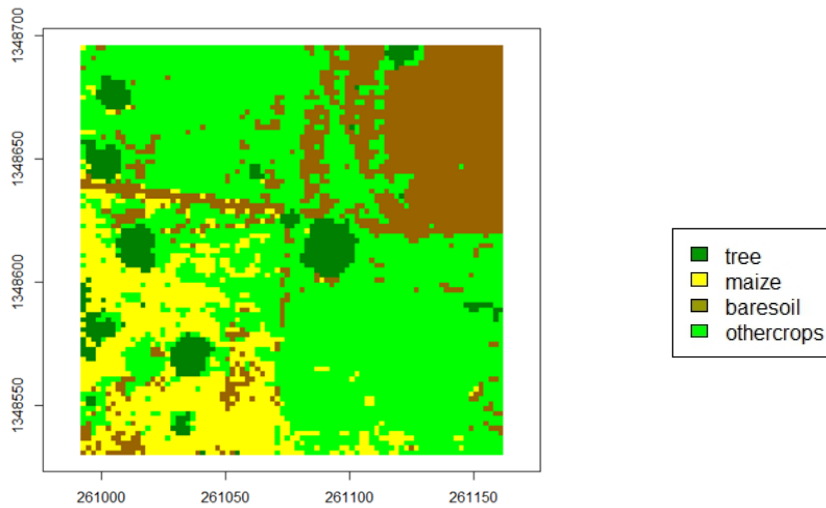


Figure 10. SVM classification result of training area (Figure. 8). This result was generated by using linear kernel. C parameter was optimized to be 15.

To avoid the SVM classification errors, the training image is manually drawn depends on the SVM result and VHR satellite image. In order to make significant difference with the SVM generated training image, the manually drawn training image is much smoother. Figure 11 shows the training image that was manually drawn in the Erdas Software.

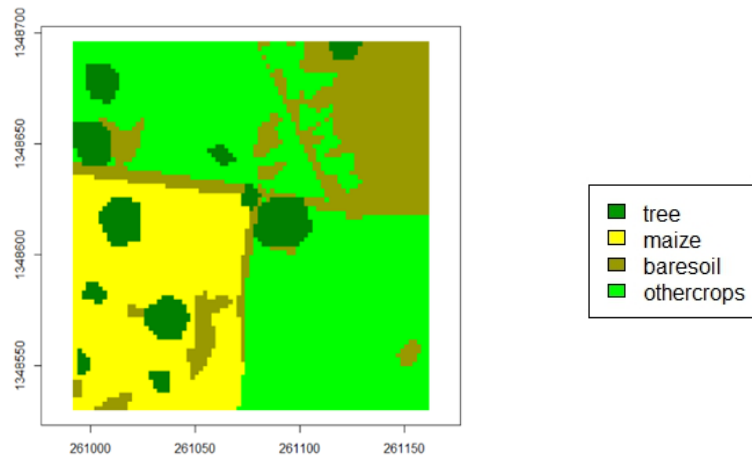


Figure 11. Manually drawn training image of the training area (Figure.8).

Figure 12 shows an comparison of the MPS simulations of these two different training image. The left image is the MPS simulation of the SVM generated training image, the right one is the MPS simulation of manually drawn training image. There are isolated maize pixels in the left image while there is none in the right image. Therefore, the MPS simulation of the SVM generated training image represents the errors of the training image.

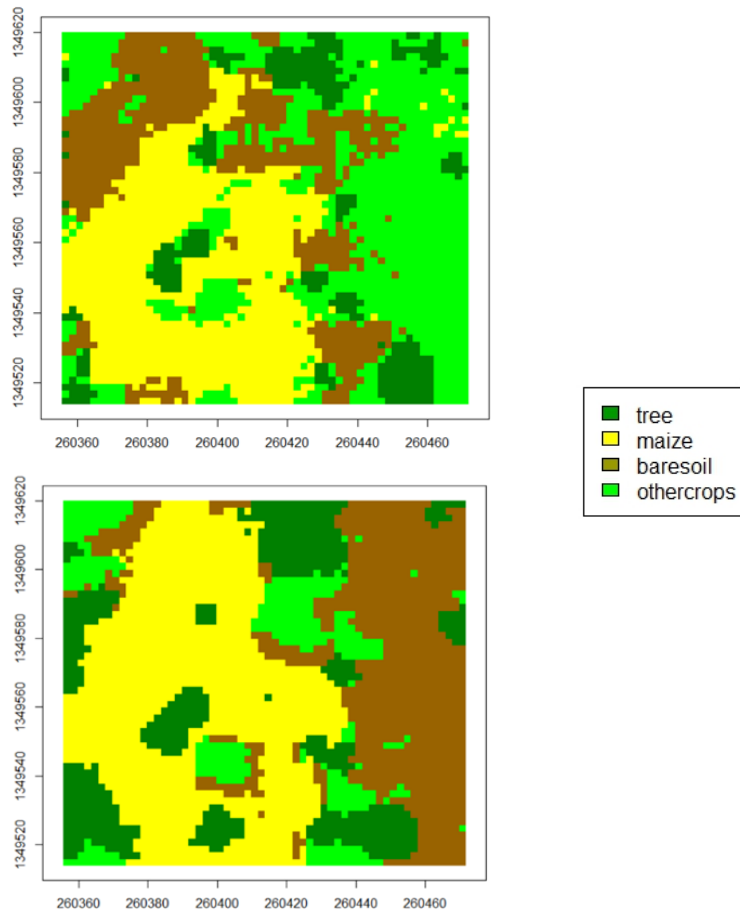


Figure 12. Comparison of MPS classification of different training image. The first image is the classification result by using the training image that was generated by SVM with linear kernel (Figure 9), the second image is the classification result by using the manually drawn training image (Figure 11).

5.2. Effect of conditioning data selection

Because there was not available ground observed data, conditioning data is required to be generated.

5.2.1. Select the conditioning data manually

An experiment was made to test if the conditioning data generate the satisfying classification result. The experimental area No.1 was used to make this experiment. Figure 13 shows the image of the experimental area No.1.

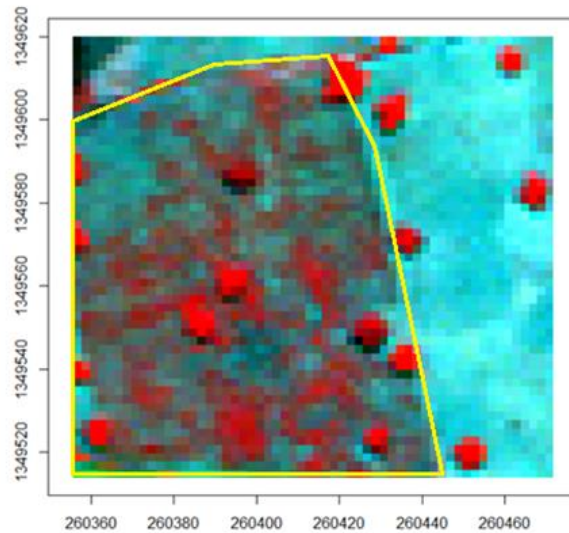


Figure 13. Worldview-2 satellite image of experimental area No.1. Band 7/5/2 was used to present the R/G/B bands. The known maize field are displayed with the yellow box.

Manually drawn conditioning points were expected to have high accuracy. In this experiment, there are 52 conditioning points was selected. The Figure 14 shows how the conditioning data is distributed among the experimental area No.1.

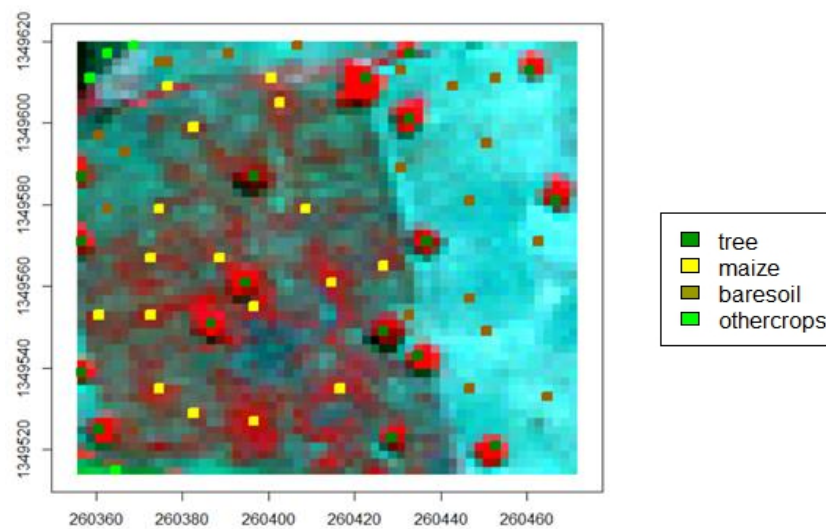


Figure 14. Example of manually selected Conditioning data, it is drawn on the Worldview-2 satellite image of experimental area No.1 (Figure.13).

In order to apply SVM classification to the experimental area No.1, training set of the training area were selected in multispectral image using ENVI software.

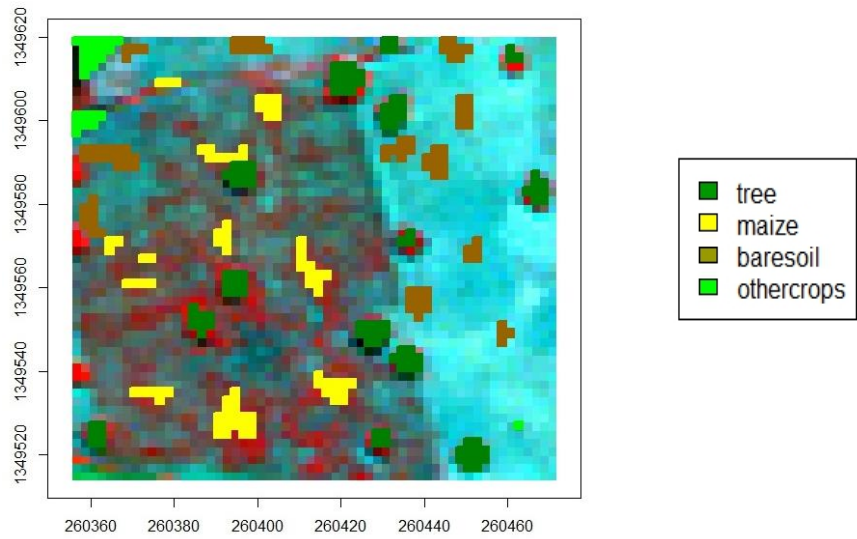


Figure 15. Training set of experimental area No.1 (Figure.13). The training set was collected in multispectral image.

After process the MPS application with this conditioning data, it generates unreliable result. Table 1 shows the OA and Kappa of 5 classification result by using this conditioning data.

OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
18.4%	0.013	94.3%	0.911
21.4%	0.106	94.0%	0.906
17.8%	0.086	93.4%	0.898
15.9%	0.058	93.4%	0.897
19.8%	0.062	94.1%	0.910

Table 1 Classification results by using previous manually selected conditioning data (Figure.13).

The result is not satisfied. By comparing the SVM classification result and MPS classification result visually, the MPS application with this manually selected conditioning data cannot reproduce the spatial structure correctly. Figure 16 shows the SVM classification result and one of those failed MPS classification result.

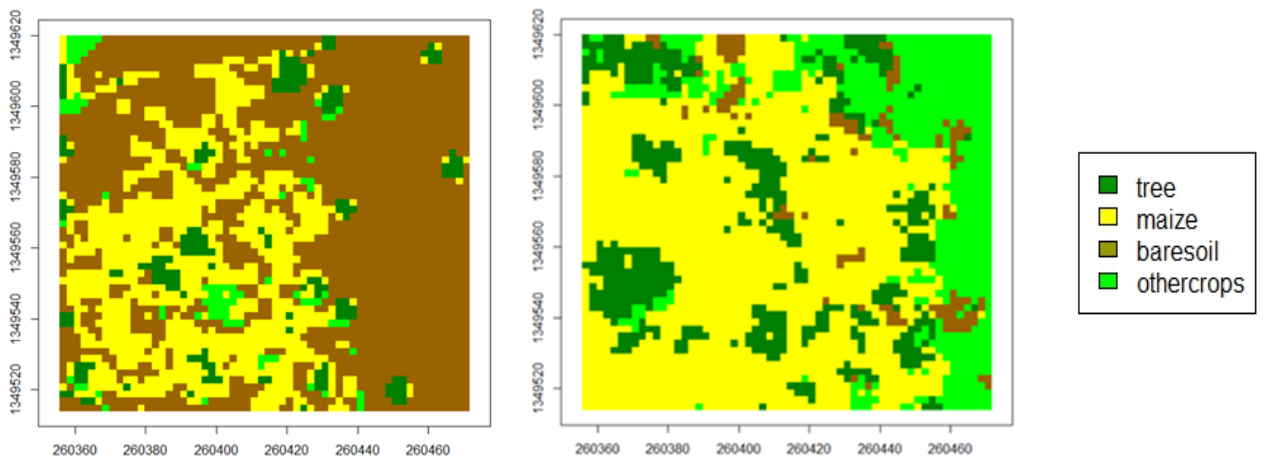


Figure 16. Comparison between SVM and MPS classification of the experimental area No.1 (Figure. 13). The left image is the SVM classification result. It was generated by using linear kernel. C

parameter was optimized to be 27. The second image is the MPS classification result by using the manually selected conditioning data (Figure 14).

Comparing with the SVM classification result that has 93.6% OA and 0.892 Kappa. After combine the MPS classification result with the SVM classification result, the final classification result is not worse than SVM result. Oppositely, it shows some unexpected improvements. By comparing the conditioning data with the classification test set, it shows part of them are repeated. As conditioning data will be represented in the final classification result, it means part of the test set are sure to be correct before test the classification accuracy.

By comparing the MPS classification result with the final classification result, it seems the classification estimation of the MPS classification result shows lots of uncertainty, it only modify the SVM classification result in limited pixels. Figure 17 presents the final classification result by combine the SVM classification and the MPS classification.

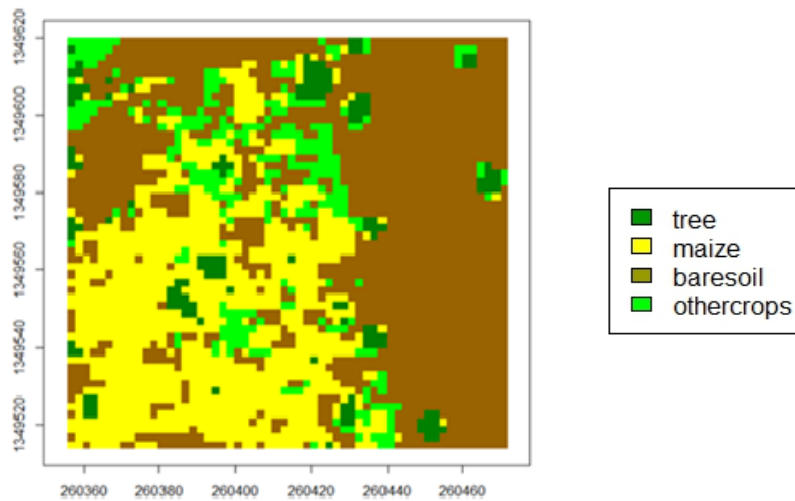


Figure 17. SVM (Figure. 14 left) and MPS (Figure. 14 right) combined classification of experimental area No.1 (Figure.12) with manually selected conditioning data (Figure 13)

5.2.2. Select the conditioning data randomly

Select the pixels randomly from the SVM classification result. Its quality relies on the accuracy of the SVM classification result. Correct and wrong classified pixels have the same possibility to be the conditioning data. The condition data of each simulation can be totally different. It is enlarge the stochasticity of MPS simulations. Figure 18 shows two MPS classification result with 100 random conditioning data. The classification results were bad, and they are totally different.

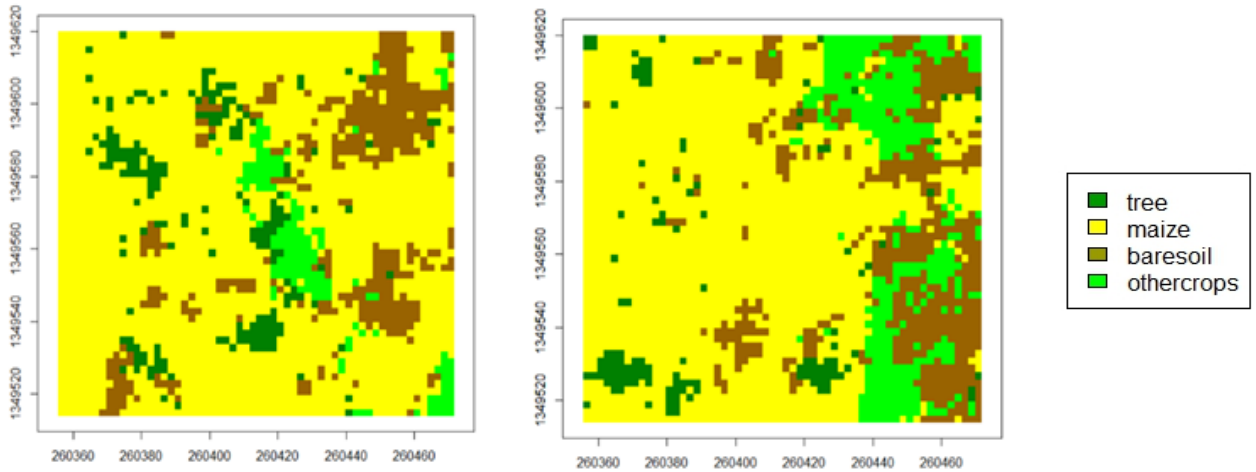


Figure 18. Two different MPS classification of experimental area No.1 (Figure. 13), they are generated by using same number of random conditioning points.

By testing the MPS application with different number of random conditioning data, here is a form of the OA and Kappa of different results. It shows the MPS results are becoming better with the increasing of the number of conditioning data.

Number of conditioning data	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
100	34.9%	0.134	93.0%	0.883
150	35.4%	0.136	93.0%	0.884
200	35.9%	0.153	93.2%	0.892
250	36.7%	0.159	93.3%	0.888
300	39.4%	0.162	93.9%	0.897

Table 2 Experiment of different number of random conditioning data, OA and Kappa is the average value of 5 MPS classification results

Comparing with the SVM classification result that has 93.6% OA and 0.892 Kappa, The final classification results with the random conditioning data is not better than the original SVM classification result.

5.2.3. Use pixels of high confidence as conditioning data.

Pixels of high confidence as conditioning data are the pixels that have high SVM classification certainty. After posterior probability of SVM result for each pixel is obtained. Pixels with high SVM posterior probability are defined as pixels of high confidence.

An experiment was made to check the performance of this conditioning data picking method. For each class, all the points with the most 5% high posterior probability are defined as pixels of high confidence. By viewing the posterior probability estimation of pixels of high confidence, most of them has above 90% the posterior probability. However, pixels of high confidence that belong to class bare soil only have around 65% classification proportion. It is because of the SVM doesn't classify this class as good as other three classes. In order to ensure the stochasticity of the MPS simulation, 20% of these pixels of high confidence are selected randomly to be the conditioning data in every single MPS simulation.

Figure 19 shows the comparison between SVM classification and MPS classification uses pixels of high confidence as conditioning data. In 5 MPS classifications, the average number of generated conditioning data was 323, the average OA of 5 results is 85.0%, and the Kappa is 0.773.

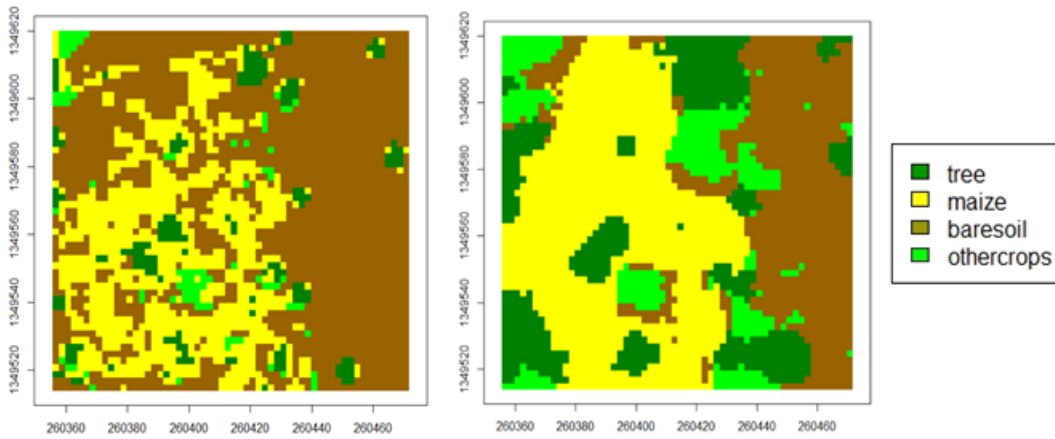


Figure 19. Comparison between SVM classification result and MPS classification of experimental area No. 1 (Figure.13). The left image is the SVM classification, the right image is the MPS classification uses pixels of high confidence as conditioning data.

After combining the SVM and MPS classification result, Figure 20 shows the final classification result. The OA of this result is 96.5%, the Kappa is 0. 0.942.

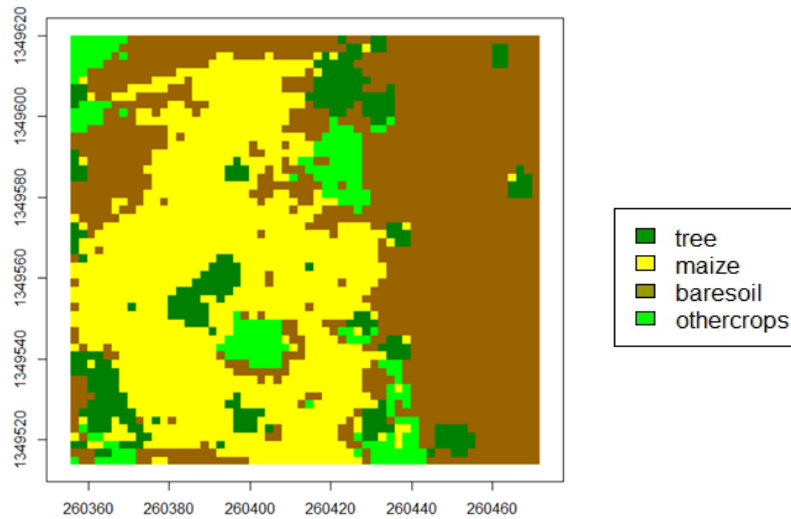


Figure 20. SVM (Figure. 17 left) and MPS (Figure. 19 right) combined classification result of experimental area No.1 (Figure. 13) by using the confident pixel conditioning data.

5.3. Effect of apply the method to different experimental area

In order to test the performance of the MPS application, it was applied to other to experimental areas. Here follows the details of the tests.

5.3.1. Experimental area No.2

Other maize fields were also used to test the performance of the MPS classification application. By comparing with the ground observed data. The known maize field are displayed with the yellow box. Figure 21 shows the sub set of the VHR satellite image of the experimental area No.2.

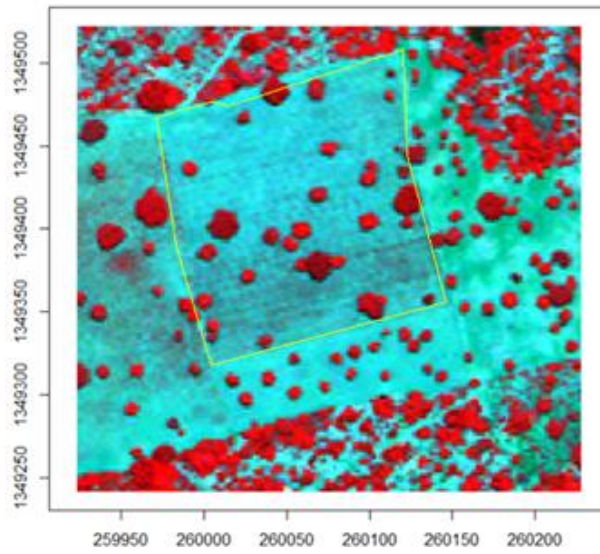


Figure 21. WorldView-2 satellite image of experimental area No. 2 in Mali. Band 7/5/2 was used to present the R/G/B bands. The known maize field are displayed with the yellow box.

In order to apply SVM classification to the experimental area No. 2, training set of the experimental area No. 2 were selected in multispectral image using ENVI software.

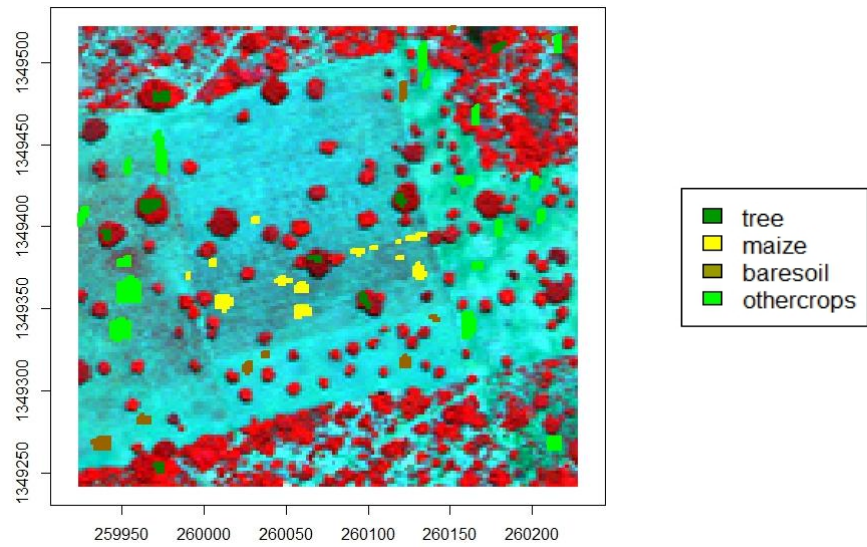


Figure 22. Training set of experimental area No.2 (Figure. 21). The training set was collected in multispectral image.

Figure 20 shows the SVM classification result. There is not significant class mixing that is observed in this result. The OA of the result is 88.3% and the Kappa is 0.877.

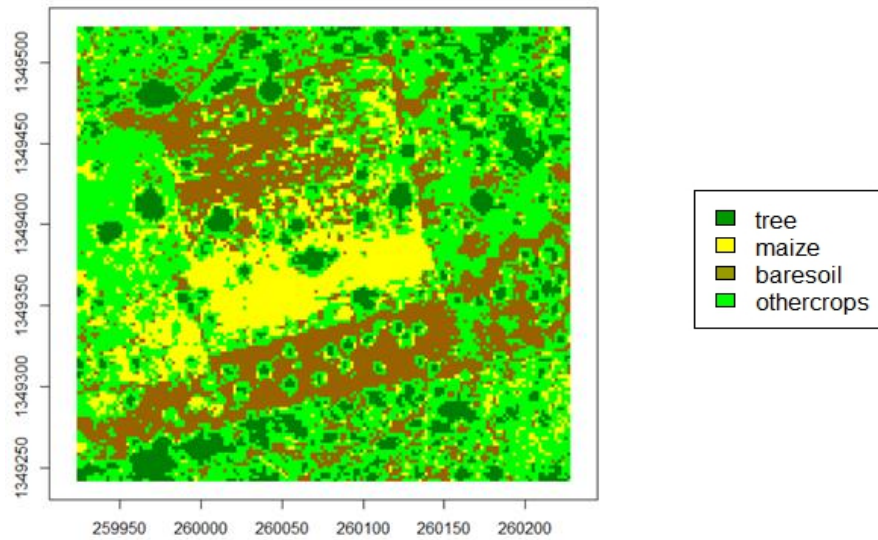


Figure 23. SVM classification result of experimental area No. 2 (Figure.21). It was generated by using linear kernel. C parameter was optimized to be 12.

Manually drawn training image is used to simulate the MPS classification result. Pixels of high confidence are used conditioning data. Figure 24 shows MPS classification result, it has OA 84.2% and Kappa 0.827.

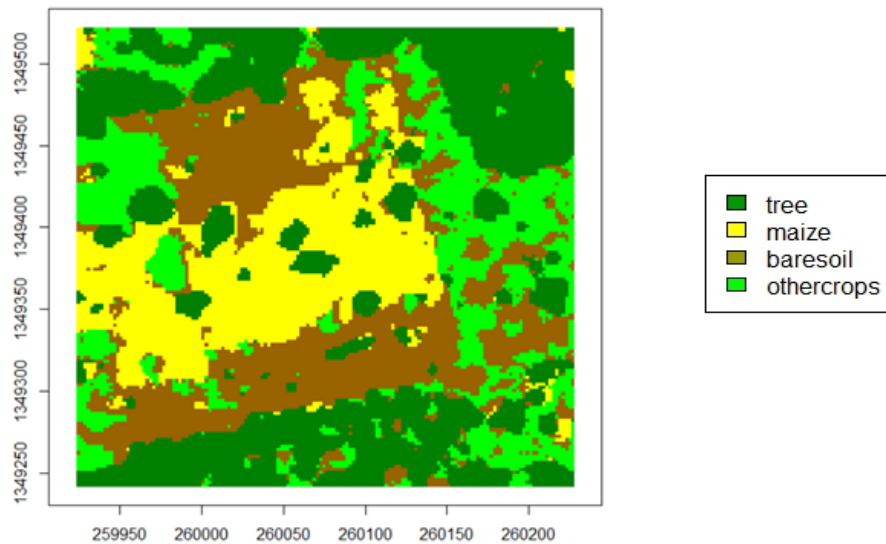


Figure 24. MPS classification result of experimental area No. 2 (Figure.19)

Because the manually drawn training image was really smooth, this MPS classification result looks much smoother than the original SVM classification. The complex physical reality is not represented in the classification.

Figure 25 shows the combined the MPS and SVM classification result. The OA of this classification result is 87.4% and the kappa is 0.851. It is lower than the SVM classification result. The reason of the bad classification was expected to be the divergence of the complexity between the experimental area No.2 and the training image.

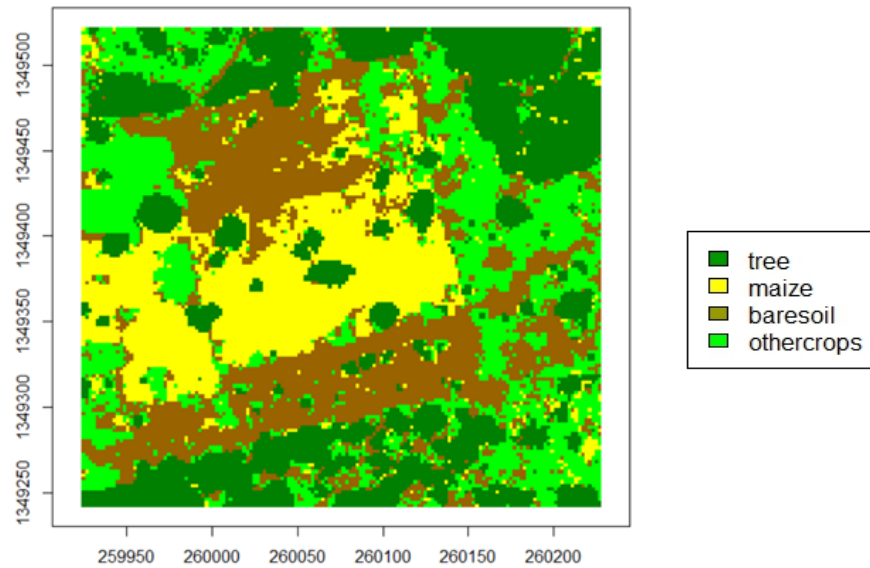


Figure 25. SVM (Figure. 23) and MPS (Figure.24) combined classification result of experimental area No. 2 (Figure 21).

5.3.2. Experimental area No.3

Here is an experiment with the experimental area No.3. It is also a maize field within the study area. By comparing with the ground observed data. The known maize field are displayed with the yellow box. Figure 26 shows the sub set of the VHR satellite image of the target area.

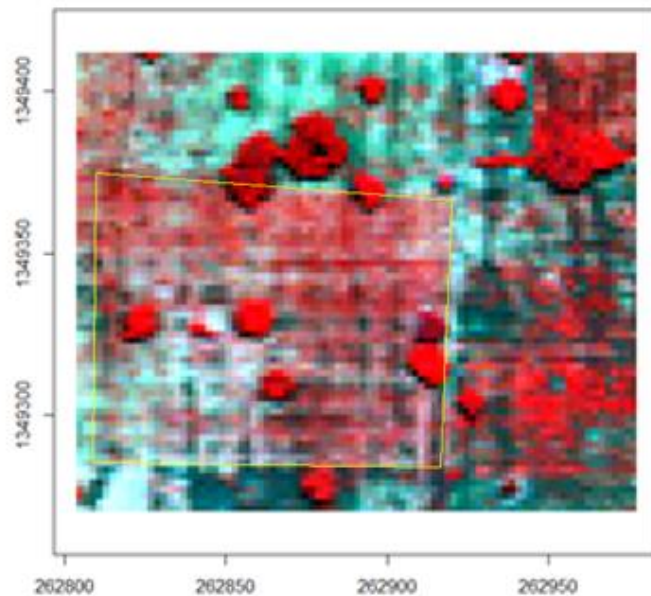


Figure 26. Worldview-2 satellite image of experimental area No.3. Band 7/5/2 was used to present the R/G/B bands. The known maize field are displayed with the yellow box.

In order to apply SVM classification to the experimental area No.3, training set of the experimental area No.3 were selected in multispectral image using ENVI software.

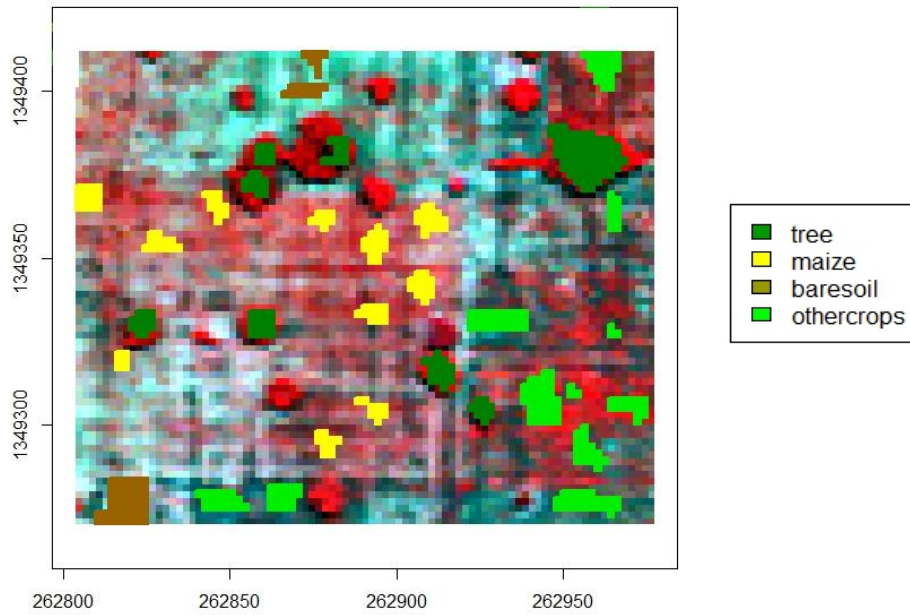


Figure 27. Training set of experimental area No.3. The training set was collected in multispectral image.

Figure 28 shows the SVM classification result. The significant mixing of the maize class and othercrops class is observed in this result. The OA of the result is 87.94% and the Kappa is 0.843.

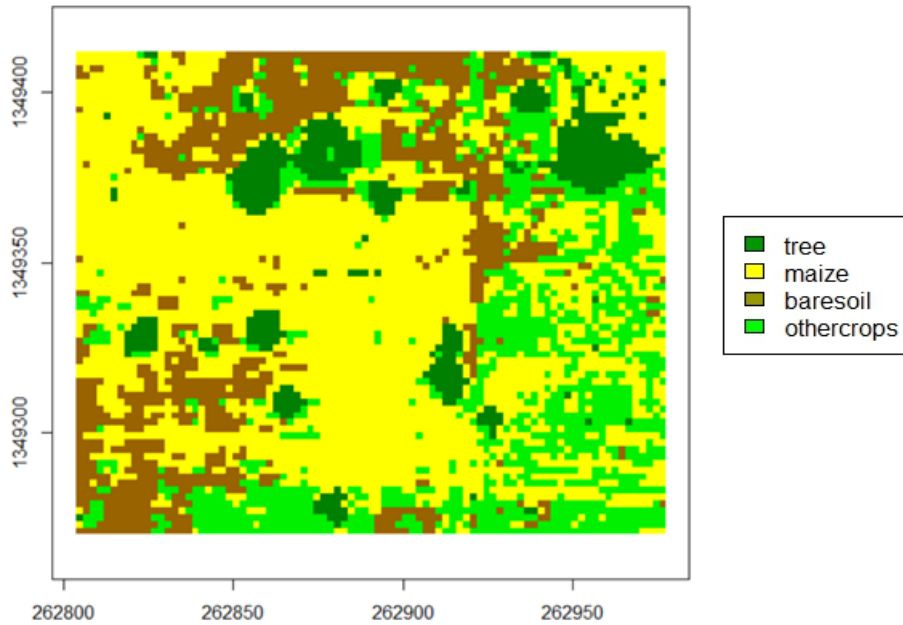


Figure 28. SVM classification result of experimental area No. 3 (Figure. 26). It was generated by using linear kernel. C parameter was optimized to be 21.

The manually drawn training image is used to simulate the MPS classification result. Pixels of high confidence in SVM result are used as conditioning data. Figure 25 shows the MPS classification result, it has OA 89.3% and Kappa 0.875.

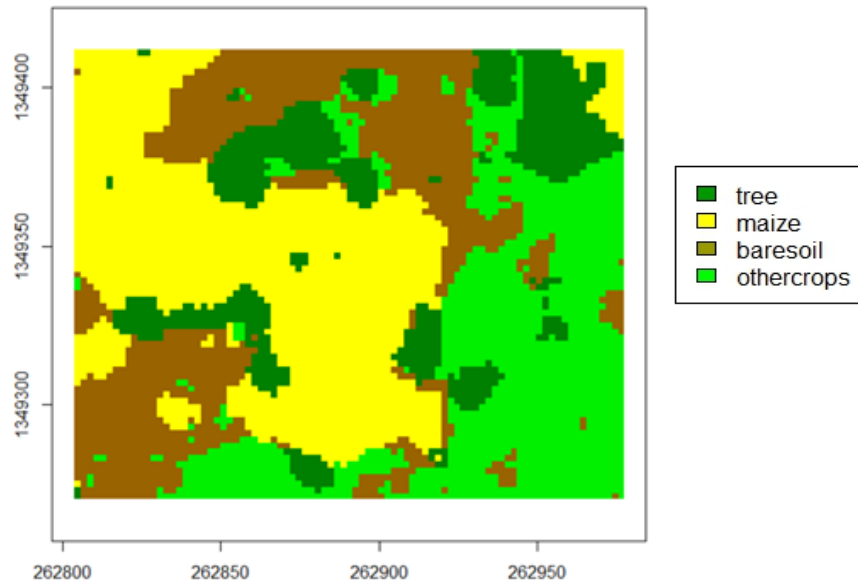


Figure 29. MPS classification result of experimental area No. 3 (Figure. 26).

Because the manually drawn training image was really smooth, this MPS classification result looks much smoother than the previous SVM classification result. The mixing of the problem is not here anymore.

Figure 26 shows the combined the MPS and SVM classification result. Because of the experimental area No.3 is more similar with the training image than the experimental area No.2. The classification result of this experiment is much better than the previous one. The OA of this classification result is 92.7% and the kappa is 0.901.

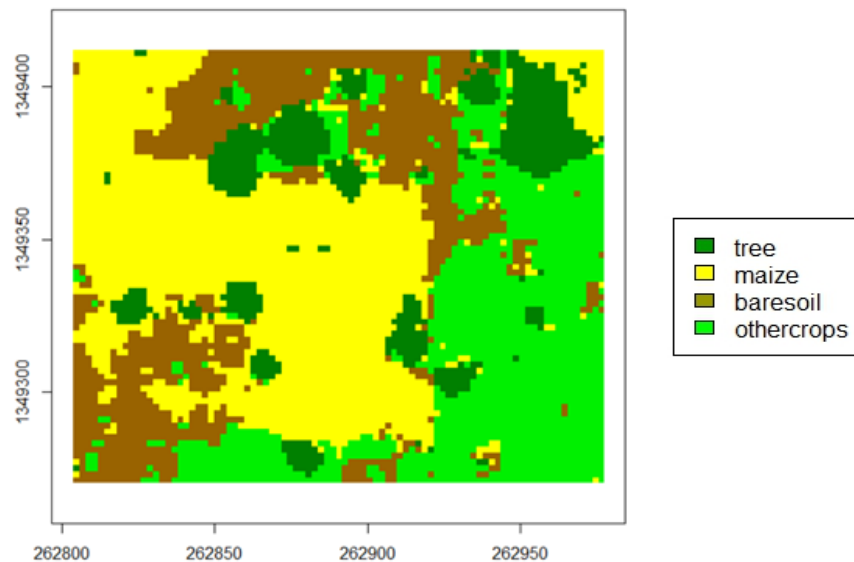


Figure 30. SVM (Figure. 28) and MPS (Figure. 29) combined classification result of experimental area No. 3 (Figure. 26).

5.4. Effect of using different MPS algorithm Parameters

Before process the Direct Sampling simulation, there are some algorithm parameters that have to be settled. These parameters includes maximum extension of the data events, maximum number of points in the data event, maximum fraction of scanned training image, distance threshold and type of distance variable.

In order to examine the optimal value of each parameter, many experiments were made to test how the testing parameter influences classification OA and Kappa.

This chapter presents the details of experiments of each parameter.

5.4.1. Maximum extension of the data events

Maximum extension of the data events describes the maximum extent of the spatial pattern. In order to capture enough points in the spatial pattern, the experiment starts with the value 8 pixels. Table 3 shows the result of experiments that applied to the experimental area No.1. These results shows the optimal maximum extension of this experimental area is around 18.

Maximum Extension (pixels)	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
8	84.5%	0.765	96.3%	0.938
10	86.5%	0.793	96.4%	0.941
12	85.5%	0.778	96.1%	0.935
14	84.7%	0.767	95.7%	0.928
16	81.5%	0.726	95.8%	0.930
18	88.3%	0.817	96.6%	0.943
20	85.0%	0.773	96.5%	0.942
22	83.1%	0.748	95.7%	0.928

Table 3. Experiments of different Maximum Extension applied to Experimental area No.1. OA and Kappa is the average value of 5 MPS classification results

Table 4 shows the result of experiments that applied to the experimental area No.2. These results shows the optimal maximum extension of this experimental area is around 14.

Maximum Extension (pixels)	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
8	82.7%	0.802	84.4%	0.824
10	84.2%	0.827	87.4%	0.851
12	83.9%	0.823	87.0%	0.845
14	84.7%	0.831	88.3%	0.859
16	84.5%	0.824	87.2%	0.846
18	84.3%	0.825	87.4%	0.867
20	84.2%	0.827	87.4%	0.851
22	80.9%	0.769	82.3%	0.803

Table 4. Experiment of different Maximum Extension applied to Experimental area No.2. OA and Kappa is the average value of 5 MPS classification results

The optimal maximum extension is different for different experimental area. It depends on the spatial structure of the experimental area. There was an assumption of the optimal maximum extension of the data event is depends on the size of the training image and simulation grid. However, this assumption is not approved in this research.

5.4.2. Maximum number of points in the data event:

Maximum number of points in the data event describes the maximum number of points in the spatial pattern. Table 5 shows the result of experiments that applied to the experimental area No.1. These results shows the optimal maximum number of points in the data event of this experimental area is around 14.

Maximum Number of Points	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
4	73.2%	0.614	95.3%	0.923
6	78.4%	0.684	95.9%	0.930
8	82.2%	0.736	96.3%	0.939
10	85.0%	0.773	96.5%	0.942
12	84.3%	0.760	96.5%	0.941
14	85.9%	0.781	96.7%	0.950
16	85.8%	0.779	96.1%	0.934
18	85.2%	0.700	94.7%	0.913

Table 5 Experiment of different maximum number of points applied to Experimental area No.1. OA and Kappa is the average value of 5 MPS classification results

Table 6 shows the result of experiments that applied to the experimental area No.2. These results shows the optimal maximum number of points in the data event of this experimental area is around 12.

Maximum Number of Points	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
4	79.1%	0.772	82.9%	0.825
6	81.3%	0.799	85.3%	0.830
8	82.7%	0.811	86.5%	0.842
10	84.2%	0.827	87.4%	0.851
12	84.5%	0.830	88.2%	0.863
14	82.9%	0.818	86.7%	0.850
16	81.4%	0.809	86.1%	0.834
18	80.2%	0.780	84.7%	0.813

Table 6. Experiment of different maximum number of points applied to Experimental area No.2. OA and Kappa is the average value of 5 MPS classification results

As same as the maximum extension of the data event, because the Experimental area No.1 and No. 2 has different spatial structure, the optimal maximum number of points in the data event is different for different experimental area. There was an assumption of the optimal maximum extension of the data event is depends on the size of the patterns in the training image and simulation grid. However, this assumption is not verified in this research.

5.4.3. Maximum fraction of scanned training image

When the application is looking for similar patterns from training image, maximum fraction of scanned training image determine how much percentage of the training image will be simulated. The value of maximum fraction is between 0 and 1, which represents the fraction between 0% and 100%. In order to avoid the algorithm slow down dramatically, the maximum fraction should be small. However, the result of

small value of maximum fraction was not satisfying. The experiments of this algorithm parameter starts from 0.5.

Maximum fraction	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
1	85.0%	0.773	96.5%	0.942
0.9	83.6%	0.755	96.4%	0.941
0.8	79.9%	0.706	96.1%	0.937
0.7	72.6%	0.620	95.5%	0.926
0.6	74.32%	0.639	95.0%	0.917
0.5	69.3%	0.576	94.4%	0.906

Table 7. Experiment of different Maximum fraction of scanned training image. OA and Kappa is the average value of 5 MPS classification results

Table 7 shows the OA and Kappa of classification result increases with the maximum fraction of the training image. It means the MPS simulations performances better when it simulates large fraction of the training image. For the purpose of avoiding the verbatim copy, the maximum fraction have to be smaller than the maximum possible value of 1. However, the verbatim copy is not considered to be a problem in this research, the optimal value of maximum fraction value is 1.

5.4.4. Type of distance variable and distance threshold

Distance threshold: most of the MPS algorithm set the pixel that have the pattern with the lowest distance to the simulation grid. In direct sampling algorithm, the final distance is compared with a threshold value. If the final distance is lower than the threshold value, the pixel is set to the simulation grid. However, because I'm looking for the "same patterns" instead of "similar patterns" within training image and simulation grid, my distance threshold value is 0.

Type of distance variable: because I'm using the class labels instead of continuous numbers, the type of distance variable is set to be categorical.

5.4.5. Number of MPS simulations

The number of MPS simulations describes how many MPS simulations are combined to get the final MPS classification result. Table 8 shows performance of MPS classification result with different number of MPS simulations, the optimal number of MPS simulations was tested to be around 20.

Number of simulations	OA MPS	Kappa MPS	OA SVM+MPS	Kappa SVM+MPS
5	80.8%	0.711	93.6%	0.896
10	85.0%	0.773	96.5%	0.942
15	86.9%	0.797	96.9%	0.949
20	88.6%	0.820	96.9%	0.948
25	87.7%	0.809	97.2%	0.953
30	85.5%	0.778	96.3%	0.939
35	82.4%	0.731	94.8%	0.921

Table 8. Experiment of different number of MPS simulations. OA and Kappa is the average value of 5 MPS classification results

In Figure 31, These three results shows the MPS classification result is getting smoother with the increase of the number of MPS simulations. Therefore, to avoid the too smooth MPS result, optimal value of the number of MPS simulations is still depends on the complexity of the classification area.



Figure 31. MPS classification results with different number of simulations, from left to right there are three different MPS classification results with 10, 20 and 30 simulations.

6. DISCUSSION

This chapter discusses the SVM and MPS combined classification method by interpolating the classification results.

6.1. Interpretation of different inputs

MPS simulations are stochastic simulations, hence same input may leads to different outputs. The input includes the training image, the conditioning data and algorithm parameters. The algorithm parameters includes maximum extension and maximum number of points of the data event, maximum fraction of scanned training image, distance threshold, and the number of MPS simulations. The following lists the observations and discussions of different inputs:

- Training image: experiments were conducted comparing the classification results generated by using SVM classification training image with manually drawn training image. In general, manually drawn training image performed better than the SVM classification image. SVM classification training image retains the classification errors, which are still present in the MPS simulations. However, manually drawn training image remains imperfect given its subjective dependence on the proponent visual observations.

The similarity between the training image and the target spatial domain is also important. As traditional two point geostatistics is relying on the similarity within the entire spatial domain, MPS is relying on the similarity between the training image and the target spatial domain. In this research, there was an assumption that “all maize fields are similar”. In reality, this assumption could be justified by visual analysis.

- Conditioning data: by considering the observations via the experiments with different conditioning data, the quality of the conditioning data was shown to have been effected by accuracy and quantity. Random generated conditioning data from the SVM classification also has low accuracy representing the errors of the SVM classification. Manually drawn conditioning data has the lowest number of points, leading to MPS simulations that reproduced totally different spatial structure from the SVM classification result. Using pixels of high confidence as conditioning data leads to larger number of points with high classification certainty than the other types of conditioning data, resulting into the highest classification accuracy.
- Maximum extension of the data event describes the maximum extent of the spatial pattern, it limits the distance between the simulating point and all points within the data event. The search window is defined by using the minimal and maximum values in the location vector of the data event in simulation grid. Therefore, the maximum extension of the data event also decides which part of the training image could not be simulated.

The optimal value of maximum extension of the data event varies for different training image. If the simulation grid is large, large maximum extension of the data event may generate huge data event, leading to tiny search window that cannot provide limited spatial patterns of the training image. The optimal value of this algorithm parameter also relies on the size of the spatial patterns in the training image and simulation grid. If the maximum extension is small and the spatial pattern is large, the data event cannot capture the characteristics of the spatial patterns.

- Maximum number of points of the data event: it describes the maximum number of points in the spatial pattern. The maximum number of points of the data event relies on the size of the spatial patterns. The recommended value of the algorithm parameter is between 10 for simple case and 60 for complex spatial structures (Mariethoz & Caers, 2014).
- Maximum fraction of scanned training image: it describes the maximum percentage of the training image that could be simulated. In standard direct sampling simulations, the results are not very sensitive with this parameter. Small maximum fraction reduces the computation time of the algorithm. The recommended value of this parameter commenced from 0.1 up to 0.5 (Mariethoz & Caers, 2014). However, in the experiments, the MPS classification resulted in better performance when simulating large fraction of the training image. The optimal value of Maximum fraction value is 1.
- Distance threshold: it describes the mismatching nodes in the data event. It is also a tool to avoid slow computation and verbatim copy of the training image. Because this MPS application is looking for the "same patterns" instead of "similar patterns" within training image and simulation grid, this MPS application is not avoiding the verbatim copy. Therefore, the threshold value is 0.
- Number of MPS simulations: it describes how many MPS simulations are combined to get the MPS classification result. By observing the MPS classification results with different number of simulations, MPS classification generated more smooth results while the number of MPS simulations increased. However, different spatial structures have distinct complexity: smoother does not mean better in every situation. In order to avoid the MPS classification to become too smooth, the optimal value of the number of MPS simulations should vary for different spatial structure with distinct spatial complexity.

6.2. Evaluation of the method

This SVM and MPS combined classification method was developed and evaluated based on the MPS algorithm DS. This method is a probability classifier which generates class probability by combining the posterior probability of the SVM classification and the Class frequency of the several MPS simulations. Both of the posterior probability and the class frequency represent the classification certainty. SVM classification posterior probability is generated by using the spectral information, MPS class frequency is depends on the spatial information. Thus, the final classification probability is a combination of spectral and spatial probability. Pixels with low classification certainty that represented by SVM classification posterior probability could have high classification certainty in MPS class frequency. Therefore, MPS class frequency is a tool that can be used to solve the problem of SVM classification uncertainty. It provides better classification rather than single SVM classification.

The quality of the final classification result is relying on both part of SVM and MPS. The accuracy of the SVM classification represents the posterior probability, randomly generated conditioning data and confident pixel conditioning data. This study was not focusing on optimize the SVM kernel and parameters in, but by optimizing the kernel and the SVM parameter, it is expected to provide better results. The accuracy of MPS classification is highly relying on the similarity between the training image and the classification area. Single training image is not expected to be suitable for many spatial structures. Various training images could be

generated or drawn for every different spatial structures individually. However, it requires a lot of efforts, and it is hard to determine if a training image is good or not. Currently, it is only possible to decide if the training image has the spatial patterns of the classification area by comparing them visually. This comparison is still relying on subjective dependence.

The code of the method was written for clarity rather than for computation efficiency. The computation speed of the current implementation is not satisfied. It takes more than half minute to process a small simulation grid of 100 pixels \times 100 pixels with a tiny training image of 50 pixels \times 50 pixels. To obtain the MPS classification frequency, many simulations are needed at a time. Therefore, one MPS classification with 10 MPS simulations requires more than 10 minutes. In reality, classification work could be applied to much larger area with thousands pixels \times thousands pixels. To process such a large area, the computation pressure could be thousands times more than the current experiments. It may take hours or days to be finished. The improvement of the computation efficiency is still required.

7. Conclusion and recommendation

7.1. Conclusion

The conclusion are made by answering the question of research objective:

1. Which heterogeneous area would be the study area?

Crop fields in Mali are used to be the study area in this research. There are known maize fields and many fields of other crops. These known maize fields are asumed to be simular with each other. In order to generate the MPS simulations of an area, a training image that is similar to the target area is required. Four known maize fields are selected, one of them are used to generate the training image, rest three are used to be the experimental area.

2. Which MPS algorithm is most suitable for this research?

Direct sampling algorithm is a MPS algorithm that can be used for the simulation of geological heterogeneity, meanwhile DS can stochastically rebuild the heterogeneous spatial structure of the training image. Therefore, Direct Sampling algorithm is the suitable MPS algorithm that be used in this research.

3. How to generate suitable training image?

In order to generate a training image that contains the similar spatial patterns with the experimental area. SVM classification result of the experimental area was used to be the training image. However, the SVM classification is not perfect, there are some obvious errors inside. The optimal training image is the manually drawn training image.

4. How to find patterns in training image?

For each simulating pattern in simulation grid, the information of the pixels within the simulating pattern is stored in two vectors. One describes the pixel locations, the other one describes the pixel values. The simulating pattern is compared with every training pattern in the training image that has the same location vector. By calculating the distance between the pixel value vectors of simulating pattern and all training patterns, most of the MPS algorithms get the training pattern that has the lowest distance to the simulating pattern. In direct sampling algorithm, after compare to each training pattern, the distance is compared with a threshold value. If the distance is lower than the threshold value, the training pattern is set to be the optimal pattern, and the loop ends here.

5. Depending on if the training image is classified or not, will MPS be applied to pixel class labels or pixel spectral values?

MPS can be used to reproduces the spatial structure, but it cannot classify the images with its own algorithm. Therefore, classified training image is required, MPS was applied to pixel class labels of the classified training image.

6. What are the optimal parameter settings to obtain the best classification result?

- Optimal maximum extension, maximum number of points of the data event and Number of MPS simulations is different for different input data, these three parameters depend upon the size and spatial patterns of the training image and simulation grid.
- Maximum fraction of scanned training image: Maximum fraction of scanned training image determine how much percentage of the training image will be simulated. The MPS simulations

performances better when it simulates large fraction of the training image. The optimal value of Maximum fraction value is 1.

- Distance threshold: It describes the mismatching nodes in the data event. This MPS application is looking for the "same patterns" instead of "similar patterns" within training image and simulation grid. Therefore, the threshold value is 0.

7. Does MPS provide better classification result of heterogeneous areas rather than SVM?
MPS classification result is not generating better result than original SVM classification. Combined class probability is generated by combining the posterior probability of the SVM classification and the class frequency of the several MPS simulations. Both of the posterior probability and the class frequency represent the classification certainty. Pixels with low classification certainty that represents by SVM classification posterior probability could have high classification certainty in MPS class frequency. In this sense, MPS could help SVM with the area that has the high spectral uncertainty. By fixing this uncertainty, the result could be better than the original SVM.

7.2. Recommendation

In order to increase the quality of the classified image, other classification methods could be used instead of linear SVM, for example: non-linear SVM, Markov Random Field.

The MPS could be applied to larger image, but this is limited by the computation speed of the algorithm and the quality of the training image. The improvement of the computation efficiency is required. Better method that could generate training images that contain the heterogeneity in spatial patterns of the entire large area.

Because MPS simulations are stochastic simulations, every MPS simulations with the same training image, conditioning data and MPS algorithm parameters produces different result. Random sets theory could be used to analyse the set of the simulations. Random sets method were based on the assumption that interpreted spatial structure with more uncertainties or vagueness will result in significant differences and randomness in different interpreting outputs. Therefore, the overall set of outputs can be considered as a random set with its variability can provide the degree of uncertainty (Zhao, 2012). This method could be applied to monitor the multi-temporal changes of spatial structures (Zhao et al., 2011).

The MPS method could be used to deal with multi temporal problems. For example: The temporal variation of a single crops could be predicted by using MPS.

Unseen or unclear part within the VHR image could be predicted by MPS. For example: After the image is classified to 4 classes that include tree, crop, bare soil and shadow. The value of shadow pixels could be removed, the value of these locations could be predicted by MPS.

8. REFERENCES

- Adler, D., & Murdoch, D. (2016). CRAN - Package rgl. Retrieved February 11, 2016, from <https://cran.r-project.org/web/packages/rgl/index.html>
- Arpat, G. B., & Caers, J. (2007). Conditional Simulation with Patterns. *Mathematical Geology*, 39(2), 177–203. <http://doi.org/10.1007/s11004-006-9075-3>
- Bivand, R., Keitt, T., & Rowlingson, B. (2015). Bindings for the Geospatial Data Abstraction Library. Retrieved from <https://cran.r-project.org/web/packages/rgdal/index.html>
- Caers, J. (2004). Multiple-point geostatistics : a quantitative vehicle for integrating geologic analogs into Stanford University , Stanford Center for Reservoir Forecasting. *AAPG Mem 80:383–394*, 383–394.
- de By, R. A., & Zurita-Milla, R. (2015). STARS project presentation. Retrieved April 20, 2015, from http://www.itc.nl/library/papers_2015/pres/deby_sta_ppt.pdf
- Eddelbuettel, D., & Francois, R. (2011). Rcpp: Seamless R and C++ Integration. Retrieved February 11, 2016, from <https://cran.r-project.org/web/packages/Rcpp/vignettes/Rcpp-introduction.pdf>
- Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Bates, D., & John, C. (2016). Package “Rcpp.” Retrieved February 11, 2016, from <https://cran.r-project.org/web/packages/Rcpp/Rcpp.pdf>
- El Ouassini, A., Saucier, A., Marcotte, D., & Favis, B. D. (2008). A patchwork approach to stochastic simulation: A route towards the analysis of morphology in multiphase systems. *Chaos, Solitons & Fractals*, 36(2), 418–436. <http://doi.org/10.1016/j.chaos.2006.06.100>
- Eskandari, K., & Srinivasan, S. (2013). Reservoir Modelling of Complex Geological Systems - A Multiple Point Perspective. *Society of Petroleum Engineers*, 49(08), 59–69. <http://doi.org/10.2118/139917-PA>
- FAO. (2009). News Article: 2050: A third more mouths to feed. Retrieved September 23, 2009, from <http://www.fao.org/news/story/en/item/35571/icode/>
- Guardiano, F. B., & Srivastava, R. M. Multivariate geostatistics: beyond bivariate moments, *Geostatistics Troia '92*. Vol. 1 133–144 (1993). Kluwer; Quantitative Geology and Geostatistics, 5.
- Hijmans, R. J. (2015). Geographic Data Analysis and Modeling. Retrieved from <https://cran.r-project.org/web/packages/raster/index.html>
- Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), 1–20. <http://doi.org/10.1016/j.csda.2009.09.023>
- Kariuki, J. G. (2011). The Future of Agriculture in Africa. Retrieved January 20, 2008, from <http://www.bu.edu/pardee/files/2011/11/15-PP.pdf?PDF=pardee-papers-15-africa>
- Keerthi, S. S. (2003). Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. Retrieved February 10, 2016, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.141.880&rep=rep1&type=pdf>
- Mariethoz, G. (2016). Additional resources - Multiple-point Geostatistics stochastic modeling with training images. Retrieved February 9, 2016, from <http://www.trainingimages.org/additional-resources.html>
- Mariethoz, G., & Caers, J. (2014). *Multiple-point Geostatistics: Stochastic Modeling with Training Images*. Wiley-Blackwell.

- Mariethoz, G., & Lefebvre, S. (2014). Bridges between multiple-point geostatistics and texture synthesis: Review and guidelines for future research. *Computers & Geosciences*, *66*, 66–80. <http://doi.org/10.1016/j.cageo.2014.01.001>
- Mariethoz, G., Renard, P., & Straubhaar, J. (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, *46*(11), 1–14. <http://doi.org/10.1029/2008WR007621>
- Mustapha, H., & Dimitrakopoulos, R. (2011). HOSIM: A high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. *Computers & Geosciences*, *37*(9), 1242–1253. <http://doi.org/10.1016/j.cageo.2010.09.007>
- Peredo, O., & Ortiz, J. M. (2011). Parallel implementation of simulated annealing to reproduce multiple-point statistics. *Computers & Geosciences*, *37*(8), 1110–1121. <http://doi.org/10.1016/j.cageo.2010.10.015>
- Petrou, M., & García Sevilla, P. (2006). Image Processing: Dealing with Texture. In *Image Processing: Dealing with Texture* (pp. 1–618). <http://doi.org/10.1002/047003534X>
- Piironen, R., Heiskanen, J., Möttöus, M., & Pellikka, P. (2015). Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 1–8. <http://doi.org/10.1016/j.jag.2015.02.005>
- Platt, J. C. (1999). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Retrieved February 8, 2016, from <https://www.cs.colorado.edu/~mozer/Teaching/syllabi/6622/papers/Platt1999.pdf>
- R Core Team. (2014a). R: Model Predictions. Retrieved February 12, 2016, from <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/predict.html>
- R Core Team. (2014b). R: The R Project for Statistical Computing. Retrieved February 12, 2016, from <https://www.r-project.org/>
- R Core Team. (2014c). R: The R Stats Package. Retrieved February 12, 2016, from <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
- Rezaee, H., Mariethoz, G., Koneshloo, M., & Asghari, O. (2013). Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. *Computers & Geosciences*, *54*, 293–308.
- STARS. (2015). STARS Fact sheet. Retrieved August 13, 2015, from <http://www.stars-project.org/en/about-us/>
- Steinwart, I., & Christmann, A. (2008). Support Vector Machines (pp. 7–13).
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., & Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, *43*(3), 305–328. <http://doi.org/10.1007/s11004-011-9328-7>
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point geostatistics. *Math. Geol.*, *34*(1), 1–22. <http://doi.org/10.1023/A:1014009426274>
- Tahmasebi, P., Hezarkhani, A., & Sahimi, M. (2012). Multiple-point geostatistical modeling based on the cross-correlation functions. *Computational Geosciences*, *16*(3), 779–797. <http://doi.org/10.1007/s10596-012-9287-1>

- Tang, Y., Atkinson, P. M., Wardrop, N. a., & Zhang, J. (2013). Multiple-point geostatistical simulation for post-processing a remotely sensed land cover classification. *Spatial Statistics*, 5(1), 69–84. <http://doi.org/10.1016/j.spasta.2013.04.005>
- Tang, Y., Atkinson, P. M., & Zhang, J. (2015). Downscaling remotely sensed imagery using area-to-point cokriging and multiple-point geostatistical simulation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 174–185. <http://doi.org/10.1016/j.isprsjprs.2014.12.016>
- Taylor, S. L., & Eckley, I. A. (2013). A test of stationarity for textured images. Retrieved from <http://www.maths.lancs.ac.uk/~nunes/papers/TOSpaper.pdf>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th ed. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4/>
- Zhang, T., Switzer, P., & Journel, A. (2006). Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38(1), 63–80. <http://doi.org/10.1007/s11004-005-9004-x>
- Zhao, X., Stein, A., Chen, X., & Zhang, X. (2011). Quantification of Extensional Uncertainty of Segmented Image Objects by Random Sets. *IEEE Transactions on Geoscience and Remote Sensing*, 49. <http://doi.org/10.1109/TGRS.2011.2109064>