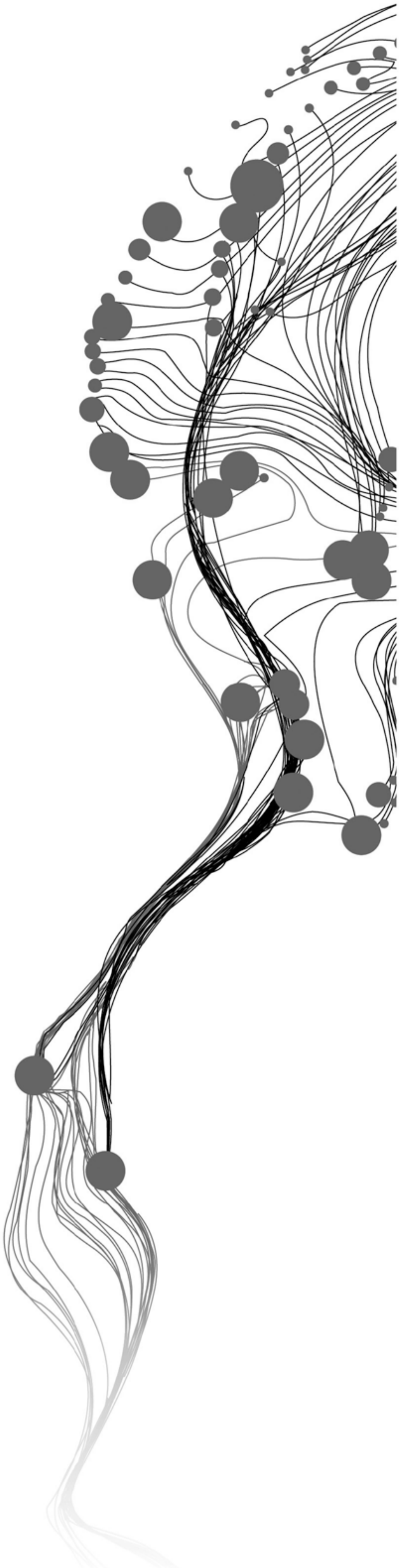
DELINEATION OF AGRICULTURAL FIELD BOUNDARIES USING RANDOM SETS

ALI GHOFRANI ESFAHANI
March, 2014

SUPERVISORS:
Dr. Ali A. Abkar
Prof. Dr. Ir. A. Stein



DELINEATION OF AGRICULTURAL FIELD BOUNDARIES USING RANDOM SETS

ALI GHOFRANI ESFAHANI
Enschede, The Netherlands, March, 2014

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

SUPERVISORS:
Dr. Ali A. Abkar
Prof. Dr. Ir. A. Stein

THESIS ASSESSMENT BOARD:
Prof. Dr. Ir. M.G. Vosselman (Chair)
Dr. M. Varshosaz (External Examiner)

ABSTRACT

Crop yield estimation of agricultural fields is important at the national and regional scale. Remote Sensing (RS) data are attractive for land cover identification, classification and estimation, especially in the agricultural regions. Modeling the agricultural fields as spatial objects can be helpful to reflect the extensional uncertainties and therefore to characterize inaccuracy in parcel size estimation.

In this research random sets are proposed for modeling agricultural parcels as extracted from an NDVI map of a Landsat 5 TM image, together with a quantification of their extensional uncertainty. The approach is applied to the agricultural fields in the Sharifabad region in Iran. These agricultural fields have both sharp and gradual transition boundaries. Thresholding approach using a random number from the normal distribution with mean and standard deviation values, of image segmentation is applied to generate random sets. The mentioned process is adopted for six parcels (agricultural field). It is noticeable that, the process of quantification of extensional uncertainty presents the parcels number 1 and 4 with larger extensional uncertainty than the other parcels. Geometric model used to delineate the agricultural field boundaries is handling non-rectangular shape (irregular shape boundaries) and this makes the approach applicable to various cases.

For validation of the results a Google Earth image is used as reference data. The overall accuracy of 91% shows that random sets are an effective tool for modeling the extensional uncertainty of the agricultural fields and for the delineation of the agricultural field boundaries, using its basic parameters and functions like the mean, covering function, level sets and variance.

Keywords:
Random sets theory, Remote Sensing, Agricultural fields, Uncertainty modeling, Crop yield estimation, Transition boundaries

ACKNOWLEDGEMENTS

I acknowledge and express my special gratitude to my supervisors Prof. Alfred Stein and Dr. Ali Abkar , for their scholarly advices, help and continuous encouragement have contributed significantly to the completion of this study. I want to thank again for providing me with constructive critics and comments which has immensely impacted the results of this thesis. A special thanks to Dr. Valentyn Tolpekin for his kindly comments and advices. Thank you so much!

I express my deep gratitude to Dr. Kourosh Khoshelham for his support.

Finally, I want to say special thanks for my family for their supports and affection during my studies. This work would not have happened without theirs advices, courage and support. I wish you a healthy and happy life into a bright future.

Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. MOTIVATION AND PROBLEM STATEMENT

According to Abkar and Sharifi (1995), "Crop yield estimation is very important in national and regional scale. Reliable and timely estimation of agricultural production has an important role in agricultural import or export policy, allocation policy for various resources and inputs, such as seed, fertilizer, pesticides, fuel, storage, etc.".

Remote Sensing (RS) data are attractive for land cover identification, classification and estimation, especially in the agricultural regions. Ferencz, et al. (2004), show that the agricultural application of satellite RS technology requires a quantitative processing of satellite remote sensing data with high accuracy and reliability. For yield prediction and for the estimation of crops it is necessary to achieve a very high accuracy and reliability (Ferencz, et al., 2004).

To identify the agricultural field boundaries with the neighbouring fields, we realize that fields may have different crop types but we do have spectral similarity in boundaries of these fields, that are as we can say "mixed boundary pixels" (Abkar, 1999). There is also a large amount of mixed boundary pixels, indicating uncertainty along the boundaries of agricultural fields. These mixed boundary pixels are to be included in the accuracy assessment. Thus proceeding, many pixels reflect extensional uncertainties and therefore, in turn lead to inaccuracy in parcel size estimation. It is noticeable that, Ji (1996) shows a field boundary as a set of locally connected pixels characterized by their abrupt spectral intensity variation on the image. Landsat images with a ground resolution of 30 by 30m have been commonly used in agricultural studies as for example in crop estimation procedures. In this way, each pixel has an important role in crop estimation. Hence the problem is to identify the boundaries between an agricultural field and its neighbouring fields. This is in particular a challenge if crops of neighbouring fields have a large spectral similarity.

Several methods have been proposed in the past to address this problem((Janssen, et al., 1995); (Ji, 1996); (Torre, et al., 2000); (Rydberg, et al., 2001); (Butenuth, et al., 2004); (Mueller, et al., 2004); (Ishida, et al., 2004); (Tiwari, et al., 2009); (Turker, et al., 2013)). Existing techniques to model the uncertainties of agricultural fields however have limitations, especially in delineation of field boundaries where the crop types of neighbouring fields are spectrally similar.

Performing a spatial analysis in the agricultural domain, geographical entities like fields,are extracted from images and are, represented with either crisp as or uncertain objects (Bandishoev, 2011). Real-world entities on RS images, however, may have gradual transition boundary due to either the scale of observation or thematic poor definition (Stein, et al., 2009). Uncertain spatial objects can be modeled by means of fuzzy set theory (Cheng, et al., 1999). Fuzzy set theory treats imprecise definitions of a concept by means of membership functions (Zhao, 2012). Since the assignment of the membership function is subjective in nature, a major obstacle of the fuzzy approach is to determinate the shape and the parameters of a membership function for a spatial object (Robinson, 2003). To handle the uncertainty in spatial objects, random sets theory have been identified as a generalization of other uncertainty theories ((Goodman, et al., 1997); (Mahler, 2007)). The membership function of a fuzzy set can be interpreted analogously to the probability of a random set covering a generic point (Goodman, 1982). As a stochastic method, random sets can be used for modeling spatial and spatial-temporal uncertainties of natural objects (Zhao, et al., 2010). The distribution function of a random set is called the covering function.

1.2. RESEARCH IDENTIFICATION

The overall objective of this research is to use random sets for modeling the extensional uncertainty of the agricultural fields in space.

1.2.1. Research objectives

The objectives of this research are:

1. To assess the suitability of random sets to delineate agricultural field boundaries.
2. To quantify the extensional uncertainty of the agricultural fields by random sets.
3. To identify the spectral variation within the agricultural fields.

1.2.2. Research questions

The research questions of this research are:

1. How to define a random set for an agricultural field?
2. How to implement covering functions for different fields with different crops?
3. How to deal with boundaries between agricultural fields that are represented as random sets?
4. What is the spectral variation within individual agricultural fields?
5. How to validate the results of random sets method in identifying agricultural field boundaries?

1.2.3. Innovation aimed at

The novelty of this research is on developing a random sets based method to identify the agricultural field boundaries.

1.3. THESIS OUTLINE

The remaining part of the thesis is organized in six chapters:

Chapter 2 demonstrates overall review regarding methods for delineation of field boundaries, uncertainty modeling and random set theory.

Chapter 3 presents an introduction of the study area, data which have been utilized and the softwares used to apply the proposed methodology.

Chapter 4 explains the proposed methodology of research approaches.

Chapter 5 presents the results of implementation of the proposed methods (Chapter 4) on the data, which have been described on chapter 3.
Chapter 6 discusses how the results analyzed and how they can be linked.
Chapter 7 presents the conclusion and recommendations of the research.

2. LITERATURE REVIEW

2.1. METHODS FOR DELINEATION OF FIELD BOUNDARIES

Janssen and Molenaar (1995) proposed an object based strategy in three stages for updating the field boundaries and the crop type of agricultural fields from a Landsat TM image data. They used information about the dynamics of object geometry to identify field boundaries. They reported good agreement between the resulting field geometry and the field geometry determined by a photo-interpreter through integration of remote sensing in a GIS environment. Ji (1996) used the dyadic wavelet transform to extract field boundaries directly from a Landsat TM imagery. The majority of the field boundaries were delineated but they reported that the use of a single date image fails to delineate some of the fallow field boundaries and boundaries between two cropped fields with similar spectral properties. Torre and Radeva (2000) addressed the problem of semiautomatic segmenting agricultural fields by region competition technique that integrates region growing and deformable models. They defined the initial positions manually, means initializing the process was time consuming. Rydberg and Borgefors (2001) presented a multispectral segmentation method for automated delineation of agricultural field boundaries in RS images. They integrated edge information from a gradient edge with a segmentation algorithm. The method which adopted was completely automatic and unsupervised, but need an operator to set some thresholds depending on the type of actual data. Butenuth, Straub, and Heipke (2004) presented the automatic extraction of field boundaries from aerial imagery using the segmentation of field areas based on the gradients in coarse scale of the imagery, carried out using watershed segmentation. GIS data were used to split the segments in detailed and improved preliminary results and finally, the derived field boundaries geometrically refunded using snake algorithm which extract only low-level features from images. Mueller, Segl and Kaufmann (2004) developed an object-oriented segmentation approach with special focus on shape analysis for the extraction of large, man-made objects, especially agricultural fields, in high-resolution panchromatic satellite imagery. The approach was combination of region- and edge-based techniques. They reported there is still lack of possibilities to control the segmentation process based on straight region boundaries of low contrast. Tiwari, Pande, Kumar, and Dadhwal (2009) presented the automatic extraction of field boundaries from high resolution satellite data using the potential of IRS P-6 LISS IV dataset for agriculture field boundary extraction. They classified the segmented regions to derive preliminary field boundaries. Finally, the derived field boundaries were geometrically refunded using snakes, but there are some problems (gaps, overlap) in snake refined boundaries. In most recent paper, Turker and Kok (2013) presented an approach for the automatic extraction of dynamic sub-boundaries within existing agricultural fields from remote sensing imagery using perceptual grouping. The results were promising and demonstrated the potential of the presented approach in the applications of agricultural field-based sub-boundary extraction and sub-field generation, but to fix the geometry of agricultural fields, they used trusty land parcels database that it is not in access most of the time.

There are some considerable results achieved by recent researches to model uncertain objects using random sets. Bandishoev (2011) worked on modeling the uncertainty of debris-covered glaciers, Zhao (2012) developed different techniques based on random sets to represent image objects with indeterminate boundaries, quantify their extensional uncertainties, and addressed uncertainty modeling in a spatial temporal change analysis. The methods were applied in the Poyang Lake area in China to classifying wetland vegetation and monitoring wetland inundation. The results demonstrated that the random set model enriches spatial and spatial-temporal modeling of phenomena which are uncertain in space and dynamic in time.

2.2. UNCERTAINTY MODELING OF SPATIAL OBJECTS

According to Stein, Hamm & Ye (2009) "Most of real-world entities on remote sensing images usually have gradual transition boundary due to either the scale of observation or thematic poor definition". It is noticeable that, agricultural fields also have gradual transition boundaries, especially where the crops of adjacent fields have spectral similarity. Quantification of extensional uncertainties of agricultural fields has an important role in delineation of field boundaries.

For modeling the uncertainty of agricultural fields as spatial objects, fuzzy set-based approach comes into mind. As Robinson (2003) reviewed various kinds of membership functions, a major obstacle of fuzzy set approach is determination of the membership function, because the assignment of the membership function is subjective in nature. For representing nature landscapes and uncertainty modeling using random sets, Bandishoev (2011) and Zhao (2012) worked on Debris-covered glaciers and wetland vegetation, respectively. As a recent research Zhou (2014) modeled the uncertainty of traffic island polygons using random sets method.

2.3. RANDOM SET THEORY

After the Matheron (1975) book, random sets theory has become an innate part of theory of probability and Molchanov (2005) and Nguyen (2006) gave advanced mathematical definitions on that. According to Zhao, Stein, & Chen (2010), "In conventional probability theory, random sets are generalizations of random variables. Unlike random variables, though, the probability laws of random sets are defined on elements which are sets rather than points". Stochastic geometry is the study of random patterns, whether of points, line segments, or objects(Kendall, 2003) and random sets theory acts as the core of stochastic geometry (Barndorff-Nielsen, et al., 1999) and relate particular closely by point process (Ortner et al. 2007, 2008; Renshaw et al. 2009). For the study of randomly varying population and randomly varying geometrical shapes, random sets theory is the base theory (Mahler, 2007).

Stoyan and Stoyan (1994) applied the random sets theory in particle statistics to study shape fluctuations of sand grains, whilst before that Serra (1980) and Diggle (1981) used the theory to model ore-sintering structure and incidence of heather respectively. In 2007 Serra also used random sets theory to model random spread process like tumor growth (Cressie, 1993) and fire spread (Vorob'ov, 1996). After using in time series analysis (Nuñez-Garcia and Wolkenhauer 2002), the first time of using random sets theory in analysis of the image was in 2003 by Gallego and Simo. Recent researches using the theory of random sets refers to Bandishoev (2011), Zhao (2012) and Zhou and Stein (2013). Bandishoev (2011) used the theory to modeling the uncertainty of debris-covered glaciers, Zhao (2012) worked on the uncertainty modeling in a spatial temporal change analysis to classifying wetland vegetation and monitoring wetland inundation and Zhou and Stein (2013) used random sets for modeling integrated uncertainties of traffic islands derived from airborne laser scanning points.

3. STUDY AREA, DATA AND SOFTWARE

This chapter presents the study area, data and software which have been utilized to apply the proposed methodology.

3.1. STUDY AREA

The area of interest for this study is located in Sharifabad region, as a town in the center of the Ghazvin province in the north-west of Iran. The geographical coordinates of the area are 36°10'44" N, 50°13'34" E. This area represents a typical agricultural region in Iran. The agricultural fields are large and also small with no specific shapes and mostly with different crop types. The elevation differences in Sharifabad region are very small. Figure 1 illustrates the study area in a snapshot of google earth. In this research a Landsat 5 TM satellite image with a Google Earth image from the study area, have been used to apply the proposed methodology.

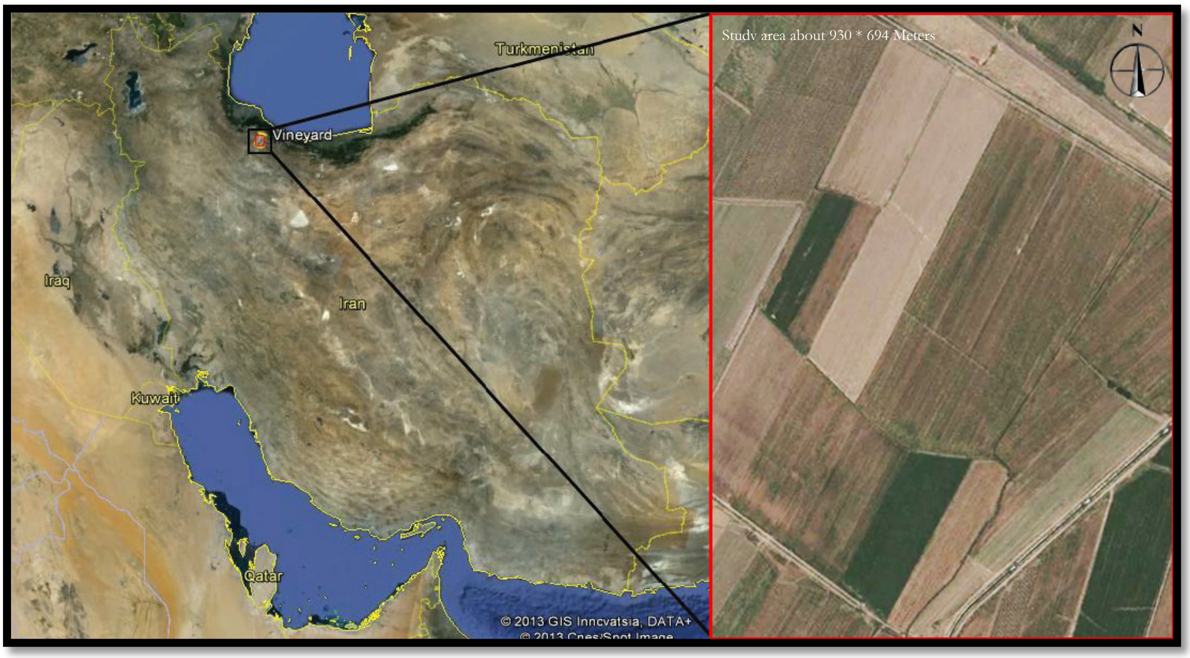


Figure 1. Study area (Sharifabad region) in a snapshot of Google Earth

3.2. DATA

3.2.1. Landsat 5 TM image

In this research a Landsat 5 Thematic Mapper (TM) image from U.S. Geological Survey (USGS: http://glovis.usgs.gov) has been utilized which acquired on 1th August 2011. The image consists of seven multispectral bands (of visible, near infrared and middle infrared) with 30m ground resolution for Bands 1 to 5 and 7. Figure 2 shows the Landsat 5 TM image and Table 1 presents the spectral bands of Landsat 5 TM image with related wavelengths.

Figure 2. Landsat 5 TM image from the study area, imagery date: 01/08/2011(Bands 5,4,3)

Table 1. Spectral bands of Landsat 5 TM image with related wavelengths

Bands of Landsat 5 TM	Description	Wavelength (μm)
Band 1	Blue	0.45 – 0.52
Band 2	Green	0.52 – 0.60
Band 3	Red	0.63 – 0.69
Band 4	Near Infrared	0.76 – 0.90
Band 5	Middle Infrared	1.55 – 1.75
Band 6	Thermal Infrared	10.40 – 12.50
Band 7	Middle Infrared	2.08 – 2.35

3.2.2. Google Earth image

In this research a georeferenced Google Earth image acquired on 18th of June 2011 has been used as reference data for validation of the results. This high resolution image has three bands (red, Green and blue). According to zmescience (2013) "Google Earth is a virtual globe, map and geographical information program that was originally called EarthViewer . It maps the Earth by the superimposition of images obtained from satellite imagery, aerial photography and geographic information system (GIS) 3D globe". Figure 3 illustrates the Google Earth image from the study area.

Figure 3. Google Earth image from the study area, imagery date: 18/06/2011

3.3. SOFTWARE

In this study the software packages of R, ENVI, ERDAS Imagine and ArcGIS have been used to apply the proposed methodology.

- **R Software**

R is a freeware statistical software package. The environment of the R software is for statistical computing and graphics. In this study the R is utilized for calculation of covering function, Random regions, core, median and support sets and also mean of the random sets.

- **ENVI**

ENVI software is a powerful remote sensing exploitation and image analysis software tool which includes links to popular GIS software programs. In this research it has been used for subsets extraction of the area of interest and calculating Normalized Difference Vegetation Index (NDVI) map from the Landsat 5 TM image.

- **ERDAS Imagine**

ERDAS Imagine is a remote sensing application with raster graphics editor abilities designed by ERDAS for geospatial applications. In this study it is utilized for georeferencing the Google Earth image and validation of the results.

- **ArcGIS**

ArcGIS is a GIS software and it is a product of Environmental Systems Research Institute (ESRI Ltd.) company. in this research ArcGIS has been used for on-screen digitizing on the reference image, editing and implement the Douglas–Peucker algorithm on the boundaries.

4. METHODS

Figure 4 illustrates the thesis workflow.

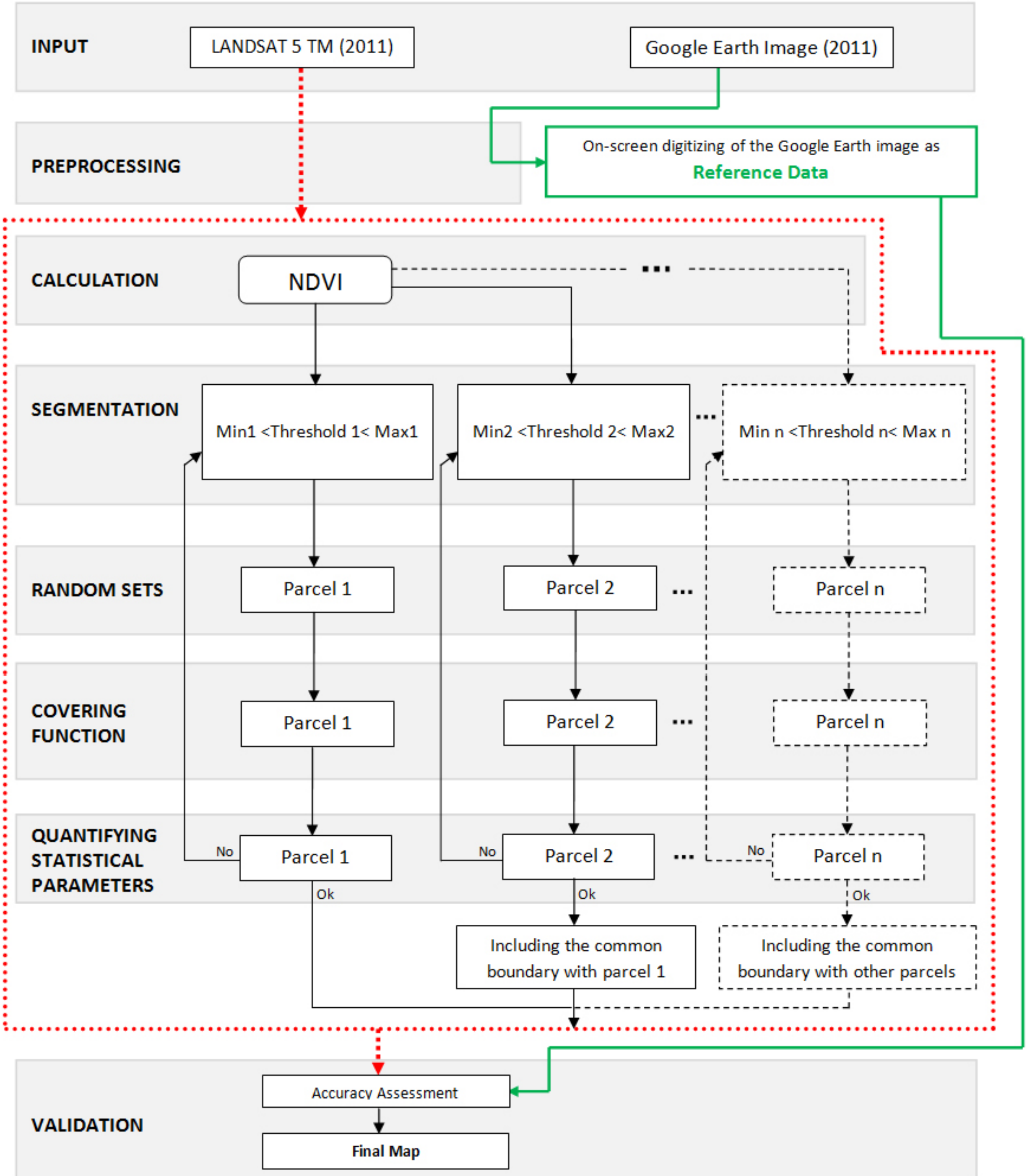


Figure 4. Thesis workflow

The flowchart of the methodology (Figure 4) shows the process of using random sets for modeling the extensional uncertainty of the agricultural fields and delineation of the field boundaries. In the first stage there are a Landsat 5 TM image and a Google Erath image both from 2011, as input data. After preprocessing of the data, there are also calculation of the NDVI map, segmentation using thresholding approach, random sets generation, covering function generation, quantification of the statistical parameters and validation stages. To apply the proposed methodology the NDVI map from the Landsat image will be generated. The NDVI is one of the most commonly used vegetation indices. Random sets will be generated based on the range of thresholding parameters (using normal distribution with mean and standard deviation values for each parcel) of segmentation on the NDVI. The covering function which is the distribution function of random sets will be produced using the segmentation results. Then, the core, median and support sets of random regions will be generated based on different level sets, the core set (Γ_c) describes the certain part of random set, the support set (Γ_s) describes the possible part of random set and the median set $(\Gamma_{0.5})$ is the 0.5- level set. The mean (Γ_m) of the random sets will be produced using the definition of Vorob'ev expectation, since it considers the set with a finite number of points like the set of pixels in image analysis. The mentioned process will be adopted for six parcels (agricultural field) and also it can be done for the other fields of the image (n parcels). For quantifying the statistical parameters of random regions, the values of set-theoretic variance and coefficient of variation will be calculated. After that the map of determined boundaries which also includes the common boundaries between fields will be simplified and the produced map enters into the validation stage. To validate the adopted method, a georeferenced Google Earth image from 18th of June 2011 will be used as reference data. After on-screen digitizing on the reference image, the accuracy assessment will be done for validation of the determined boundaries and the results will be compared.

This chapter presents the methods for applying the proposed methodology.

4.1. CALCULATION OF NDVI

Calculation of Normalized Difference Vegetation Index (NDVI) based on DN values, is as follows:

$$\mathrm{NDVI} = \frac{NIR - RED}{NIR + RED} \tag{4.1}$$

where RED and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared bands of the image, respectively.

4.2. IMAGE SEGMENTATION BASED ON THRESHOLDING APPROACH

Random set is generated by means of segmentation based on thresholding of NDVI image obtained from Landsat 5 TM data.

Thresholding of images is a segmentation method that can be applied to generate random sets by taking the value of image pixels into account. The idea of random set generation is that the extent of NDVI is sensitive to the different thresholds. So by slightly changing a threshold, a set of objects generates and construct focal elements of a random set (Bandishoev, 2011).

Segmentation based on thresholding applies a logical condition on pixel values:
$D \geq tval$
where D is the pixel value and $tval$ is a parameter called the threshold value. The result of this operation is either 0 (the result is false) or 1 (the result is true). Resulting image is a binary image (with values 0 or 1). The result of thresholding depends on the threshold value. For random sets we obtain segmentation results for a number of values (controlled by the variable *Nsteps*) of the threshold.

4.3. RANDOM SETS GENERATION

Suppose (Θ, σ_Θ) as a measure space and $(\Delta, \sigma_\Delta, Pr_\Delta)$ as a probability space, then $X(\omega)$ can be defined as a measurable function from the probability space Δ to the measurable space Θ. If we consider that, U is a set of subsets of Θ, i.e. $U \subseteq \mathcal{P}(\Theta)$, then a random variable from the sample space Δ to U , can be defined as a random set. It means that, in the space Θ the element u of the space U is a set instead of a point. Along with the above definitions, if we consider that, (U, σ_U) is a measurable space (σ_U a definition of σ-algebra on U), therefore a $(\sigma_\Delta - \sigma_U)$ -measurable mapping $\Gamma: \Delta \to U$ can be seen as a random set(Zhao, 2012).In Zhao (2012) definition, the distribution of random set Γ is as follows:

$$Pr_\Gamma(\mathcal{B}) = Pr_\Delta(\Gamma^{-1}(\mathcal{B})) = Pr_\Delta\{\, \omega \mid \Gamma(\omega) \in \mathcal{B}\} \qquad \forall\, \mathcal{B} \in \sigma_U \tag{4.2}$$

In this research random sets have been generated to apply the proposed methodology.

4.4. COVERING FUNCTION OF RANDOM SETS

On the space R^2 (Euclidean space), we consider an agricultural field as a random set. A random set Γ associates a function that takes values between 0 and 1. $Pr_\Gamma(x) : R^2 \to [0,1], x \in R^2$. This function is called the covering function of a random set. It is noticeable that the covering function can be interpreted as the probability of the element x in Euclidean space (R^2) being covered by the random set Γ (Zhao, et al., 2010). The space of an image has been denoted as $H \subset R^2$ where pixel x is the basic element, $x \in H$. The measure space that carries random spatial data models like random objects (agricultural fields) is described by U, where $U \subseteq \mathcal{P}(R^2)$. A random region (Rr) has been defined as a random set on U and can be introduced by the focal elements O_i with corresponding uncertainty assignments m_i, described as a collection of pairs$\{O_i, m_i\}$ (Zhao, et al., 2011). The focal elements are subsets of $R^2 : O_i \in \mathcal{P}(R^2)$. For example, n interpreters digitizing an agricultural field parcel from a Landsat image may obtain n different groups of indicator pixels, defined as $O_1, O_2, \dots, O_n$, like focal elements of the random region. The probability that pixel $x \in R^2$ is covered by the random region can be achieved by the covering function (Equation 4.3) as (Zhao, et al., 2011):

$$\hat{P}r_{\Gamma}(x) = \frac{1}{n}\sum_{i=1}^{n} H_{O_i}(x), \qquad x \in R^2; \quad O_i \in U \tag{4.3}$$

Where H_{O_i} is the indicator function:

$$H_{O_i}(x) = \begin{cases} 1, & x \in O_i \\ 0, & x \notin O_i \,. \end{cases}$$

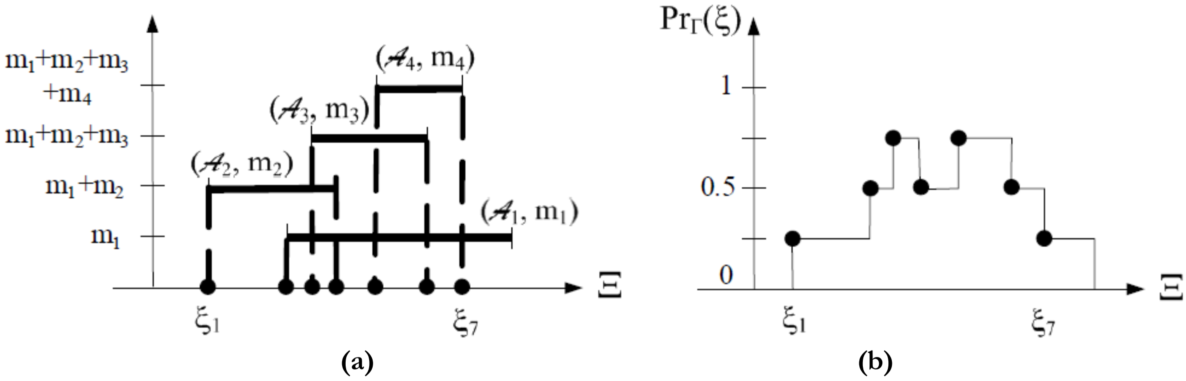


Figure 5. Four focal elements with equal uncertainty assignments construct a realization of a random set (a); Covering function of the random set estimated by focal elements (b) (Zhao, 2012).

4.5. CORE, MEDIAN AND SUPPORT SETS OF RANDOM REGIONS

The core, median and support sets of random regions have been obtained from the p-level set (Equation 4.4), as follows (Zhao, et al., 2011):

$$\Gamma_{\mathrm{p}} = \{x \in H, 0 \leq p \leq 1 : Pr_{\Gamma}(x) \geq p\} \tag{4.4}$$

The core set equals to:

$$\Gamma_{\mathrm{c}} = \{x \in H : Pr_{\Gamma}(x) = 1\} \tag{4.5}$$

The support set equals to:

$$\Gamma_{\mathrm{s}} = \{x \in H : Pr_{\Gamma}(x) > 0\} \tag{4.6}$$

And the median set as follows:

$$\Gamma_{median} = \{x \in H : Pr_{\Gamma}(x) \geq 0.5\} \tag{4.7}$$

4.6. MEAN OF RANDOM SETS

The mean (Γ_{m}) of the random set is a binary set to recognize the crisp extents of objects (agricultural fields) (Zhou, 2014). Among several expectations to present the mean of random set, Vorob'ev expectation has been determined by the covering function that is the average of the indicator functions (Stoyan, et al., 1994), and because it considers the set of pixels (with finite numbers) in image analysis. (Zhao, et al., 2011). The EA (mean area of the random set Γ) is determined as (Zhao, et al., 2011): $\mathrm{EA}(\Gamma) = \int_{\mathrm{R}^2} \mathrm{Pr}_{\Gamma}(\mathrm{x})\mathrm{dx}$ and the set Γ_{m} is as follows:

$$\Gamma_{\mathrm{m}} = \{x \in H, 0 \leq p_m \leq 1 : Pr_{\Gamma}(x) \geq p_m\} \tag{4.8}$$

where p_m is such that Γ_{m} has the area $\mathrm{EA}(\Gamma)$. It is significant that, when p_m value is equal to 0.5, mean and median are alike.

4.7. QUANTIFYING STATISTICAL PARAMETERS OF RANDOM REGIONS

It is noticeable that, for quantification of extensional uncertainty of the agricultural fields using random sets, the values of set-theoretic variance and coefficient of variation have an important role.

4.7.1. Set-theoretic variance

The common definition of variance in probability theory and statistics is that, variance measures how far a set of numbers is spread out. In this definition a small variance indicates that the data points are very close to the mean and from each other, while a high variance indicates that the data points are much spread out from the mean and from each other.

The corresponding set-theoretic variance of a random set is defined as (Zhao, 2012):

$$\Gamma_{\mathrm{var}}(x) = \mathrm{E}\,(H_{O_i}(x) - \mathrm{Pr}_{\Gamma}(\mathrm{x}))^2 \tag{4.9}$$

To quantify extensional uncertainty of the agricultural fields the level sets are utilized to present the spatial distribution of the different sizes of the random sets. It in significant that, pixels with $\Gamma_{\mathrm{var}}(x) = 0$ are expected to be in the certain part of the agricultural fields, whereas pixels with high values of Γ_{var} are expected to be in the boundary area. The sum of Γ_{var} , defined as SD, is as follows (Zhao, et al., 2011):

$$SD = \int_{\Gamma_{\mathrm{s}}} \Gamma_{\mathrm{var}}(x)\mathrm{dx}. \tag{4.10}$$

4.7.2. Coefficient of variation

The CV (coefficient of variation) is defined as (Zhao, et al., 2011):

$$CV = \frac{\int_{\Gamma_s} \sqrt{\Gamma_{var}\,(x)\mathrm{dx}}}{\int_{\Gamma_s} \hat{P}r_{\Gamma}(\mathrm{x})\mathrm{dx}} \tag{4.11}$$

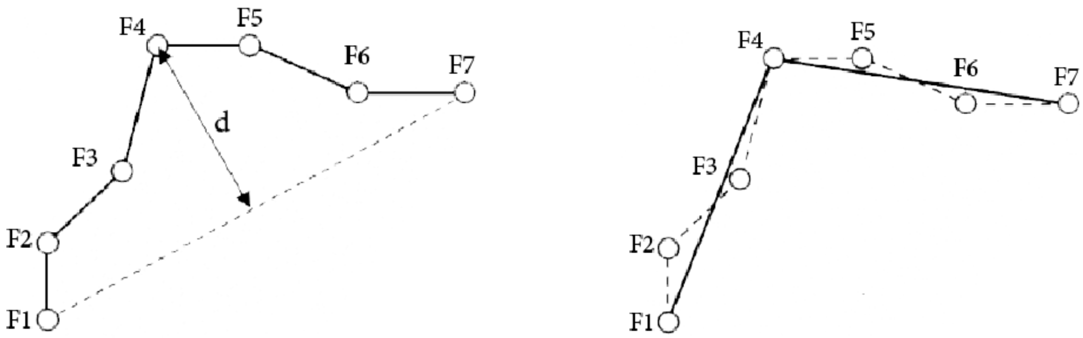
A high CV demonstrates a larger proportion of agricultural fields with a high Γ_{var} , then points to a large extensional uncertainty (Zhao, et al., 2011).

Since the CV utilized as a normalized and dimensionless measure, we can have comparison between different agricultural fields.

4.8. DOUGLAS-PEUCKER ALGORITHM

To simplify and better display the boundaries of the agricultural fields the Douglas–Peucker algorithm (Douglas, et al., 1973) has been used. According to Turker & Kok (2013) the algorithm is as follows: The first and the last points which are linked with a straight line segment are determined as the anchor (F_1) and the floater (F_7) points (Figure 6(a)). It is noticeable that, when we call a whole line as a simplified line, means all perpendicular distances from all intervening points falls above a user specified tolerance. If this situation is not met, we should define a new floating point which has the longest perpendicular offset from the straight line segment and repeat the mentioned process again. This process hangs on till the limitation of tolerance is met, then the simplification process is completed and should be stop (Turker, et al., 2013).

The algorithm is illustrated in Figure 6, in which the vertices F_1 and F_7 demonstrate the first and the last points on the line. In the first step, the longest perpendicular offset (d) to the straight line segment $F_1\,F_7$ is calculated from the vertex F_4 (Figure 6(a)). If d stays above the specified threshold value, the straight line segment $F_1\,F_7$ is separated into two line segments that are $F_1\,F_4$ and $F_4\,F_7$. The algorithm is recursively applied to these line segments. Thus, the line segment with the vertices F_1 , … , F_7 is reduced to two line segments that are $F_1\,F_4$ and $F_4\,F_7$ (Figure 6(b)) (Turker, et al., 2013).



(a) **(b)**

Figure 6. Line simplification using Douglas–Peucker algorithm.(a) before applying the line simplification and (b) after applying the line simplification algorithm (Turker, et al., 2013).

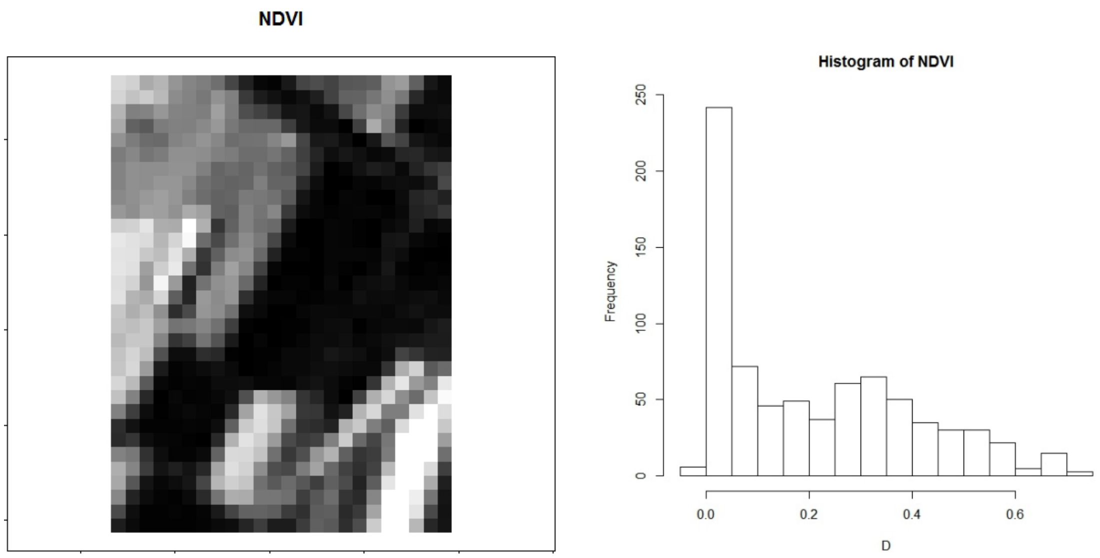
4.9. VALIDATION OF THE RESULTS

In this research to validate the proposed method in uncertainty modeling and delineation of agricultural field boundaries, the Google Earth image has been utilized as reference data. By on-screen digitizing of the reference image and doing accuracy assessment using 500 control points in ERDAS Imagine software, the process of validation of the results has been done.

5. RESULTS

5.1. NDVI (NORMALIZED DIFFERENCE VEGETATION INDEX) MAP

In this study the NDVI map was derived by calculating the NDVI using DN values (Equation 4.1). Where RED and NIR stand for the spectral reflectance measurements acquired in the red (band 3) and near-infrared (band 4) bands of Landsat 5 TM image, respectively. Figure 7 shows the NDVI map and histogram of NDVI.



(a) **(b)**

Figure 7. NDVI map from Landsat 5 TM image (a) and histogram of NDVI (b)

5.2. IMAGE SEGMENTATION

The thresholding approach to image segmentation is applied to generate random sets. In other words, random sets are generated by means of segmentation based on thresholding of NDVI image obtained from Landsat 5 TM data.

The threshold values may be chosen in various ways. It can be selected by inspection of images, as the human eye can estimate the correct values by using textural features of the image(Bandishoev, 2011).

The program applies a random number from the normal distribution with mean *(Mean)* and standard deviation (*SD)* values. Table 2 indicates the values of the mean, standard deviation and number of iterations for the six parcels.

Table 2. The values of the mean, SD (standard deviation) and number of iterations for the parcels 1-6

Parcel	*Mean*	*SD (Standard Deviation)*	*Number of Iterations*
1	0.4	0.05	60
2	0.04	0.005	60
3	0.13	0.01	60
4	0.38	0.05	60
5	0.036	0.002	60
6	0.34	0.01	60

The process of choosing the mean, standard deviation and number of iterations for each of the parcels is the same. It is almost like the process of getting some sample sets while there isn't any information about the crop type of each field (parcel). The first is inspection of the NDVI image, for instance: parcel 1. It is important to observe the pixel values of the NDVI image of parcel 1 as the human eye, also considering the average of the pixel values and textural features inside the parcel and around the conjectured boundary regions. Along with the above discussions, the calculated values for the mean and standard deviation of parcel 1 are 0.4 and 0.05, respectively. Intuitively, values closer to the mean should be more reliable and thus have more chances to be the preference of users. Therefore it has been adopted a normal distribution as our proposal distribution. In description of number of iterations in my research, it has been observed experimentally that, when the number of iterations equal to 60, we have stable covering functions in all six parcels. Figure 8 illustrates twelve segmented images generated from each of the six parcels.

Figure 8. Twelve segmented images from each of the parcels 1-6, generated with different random numbers

5.3. COVERING FUNCTION

Figure 9 presents the covering function of random sets for six parcels, in 2D and 3D models. From Equation 4.3, covering function is the distribution function of random sets which takes values between zero and one. In other words, the covering function can be interpreted as the probability of the element x (i.e. an image pixel) in R^2 being covered by the random set Γ.

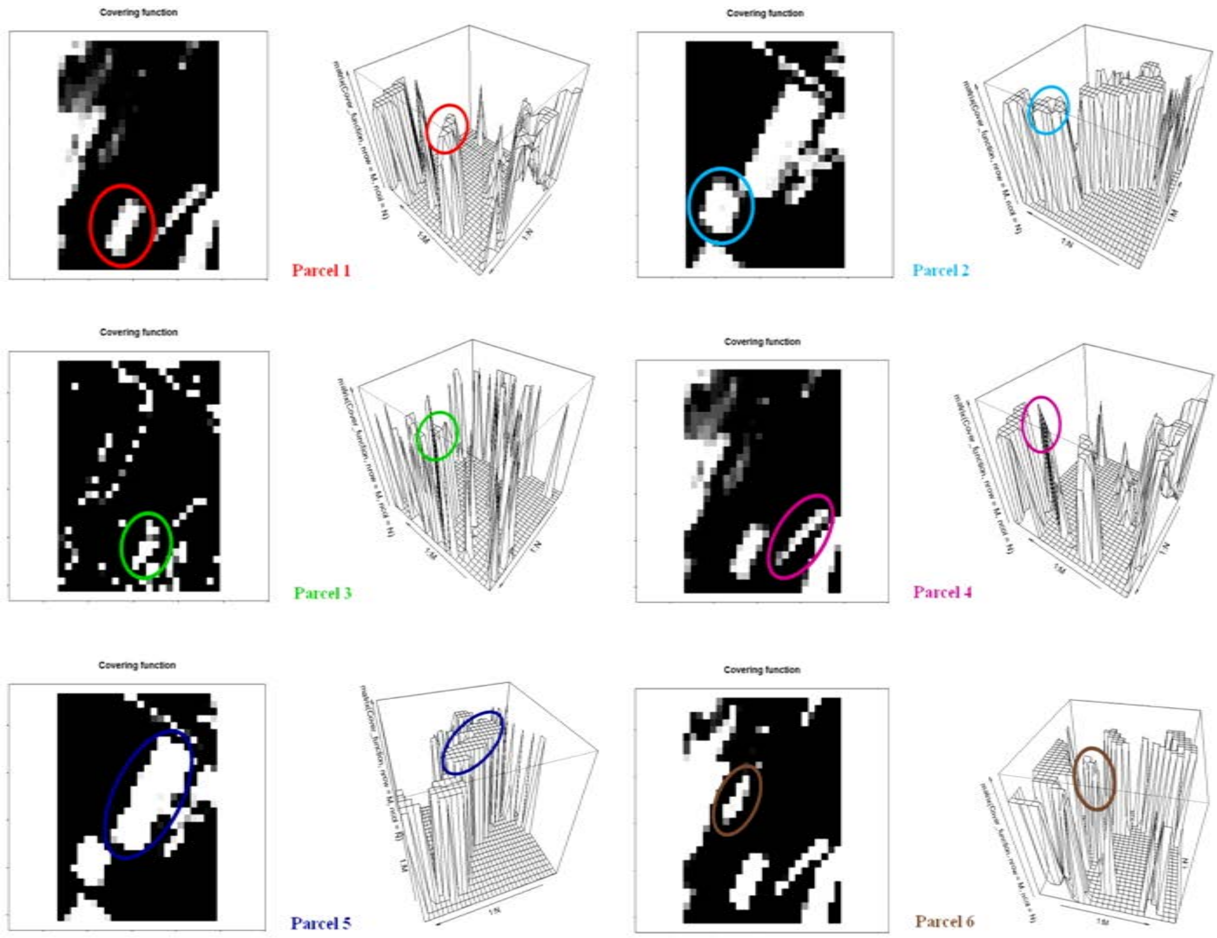


Figure 9. Covering function of random sets for parcels 1-6 in 2D and 3D models

5.4. CORE, MEDIAN AND SUPPORT SETS OF RANDOM REGIONS

According to Equations 4.5, 4.6 and 4.7, the core set (Γ_c) describes the certain part of random set (Γ), the support set (Γ_s) describes the possible part of random set (Γ) and the median set ($\Gamma_{0.5}$) is the 0.5- level set.

Figure 10 shows the core, median and support sets of parcels 1-6 and Figure 11 presents the contours of core, median and support sets for six parcels, derived from random regions.

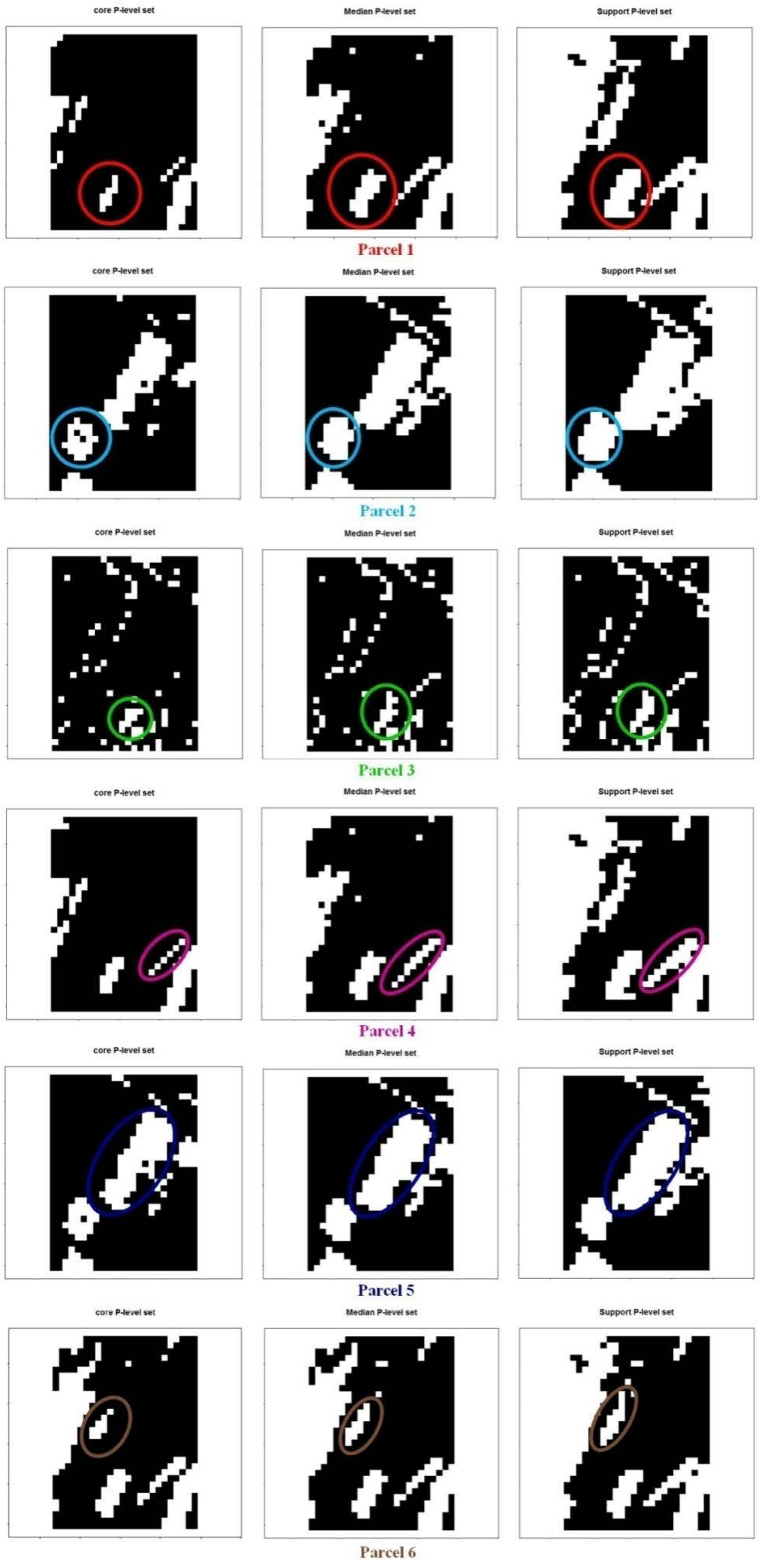


Figure 10. Core, median and support-level sets of parcels 1-6

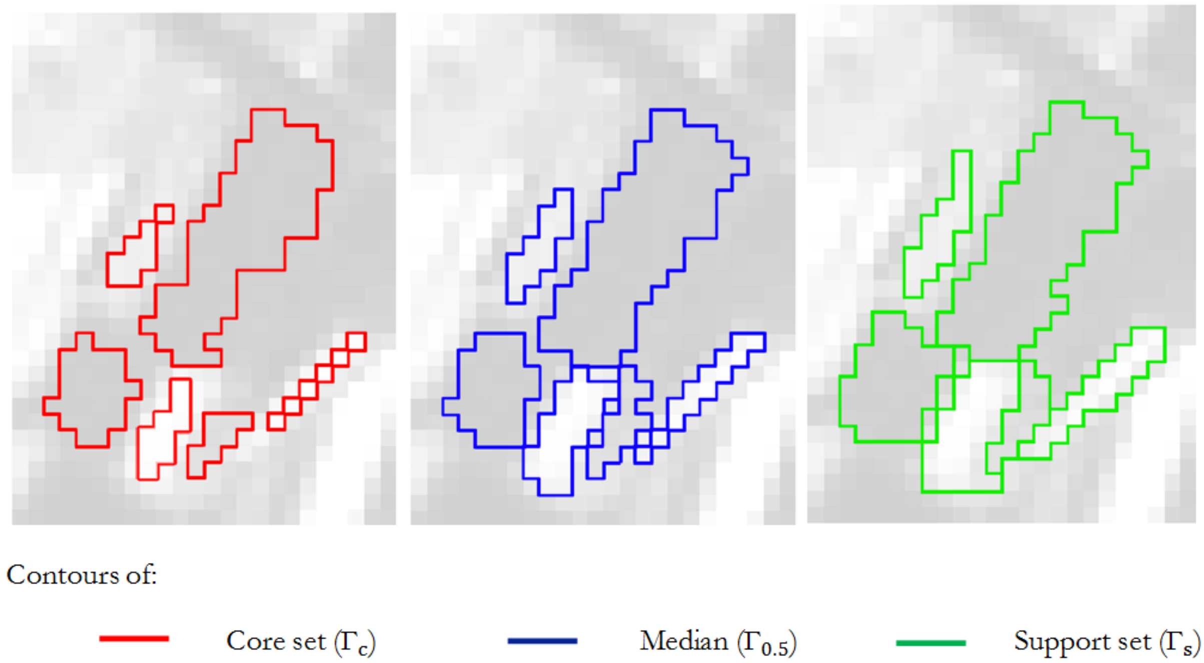


Figure 11. Contours of Γ_c, $\Gamma_{0.5}$, and Γ_s for parcels 1-6, derived from random regions

5.5. MEAN OF RANDOM SET

Figure 12 shows the Vorob'ev expectation for mean of the random sets, for six parcels.

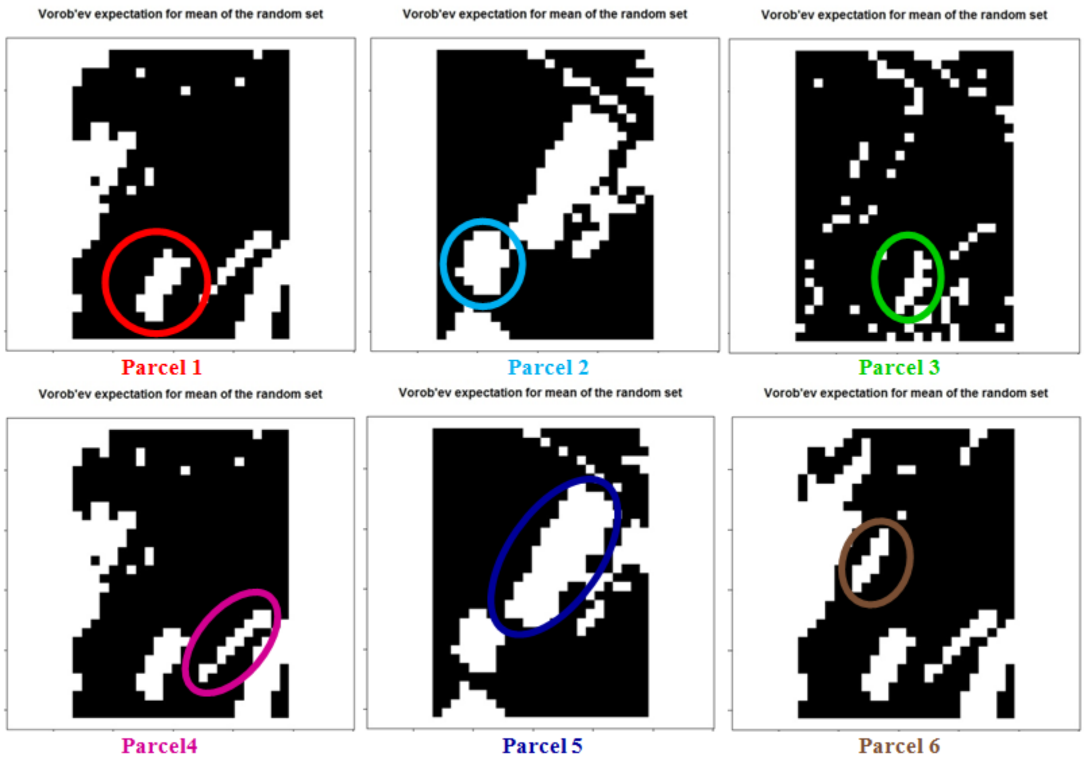


Figure 12. Vorob'ev expectation for mean of the random sets, parcels 1-6

Vorob'ev expectation (Equation 4.8) considers the set with a finite number of points like the set of pixels in image analysis. It is noticeable that in this part of research the mean and median for parcels 1, 2, 3&4 are identical.

5.6. QUANTIFICATION OF STATISTICAL PARAMETERS OF RENDOM REGIONS

For quantification of extensional uncertainty of the agricultural fields using random sets method, the values of set-theoretic variance and coefficient of variation have been presented.

5.6.1. Set-theoretic variance

The variance of a random variable has been defined as the expected value of the square of the difference between the random variable and the mean (Equation 4.9) (Zhao, et al., 2011).

From Equation 4.10, it in noticeable that, pixels with $\Gamma_{\text{var}}(x) = 0$ are expected to be in the certain part of the agricultural fields, whereas high values are expected to be in the boundary area. Figure 13, illustrates the set-theoretic variance of the random sets for six parcels and Table 3 indicates the values of sum of set-theoretic variance for these parcels.

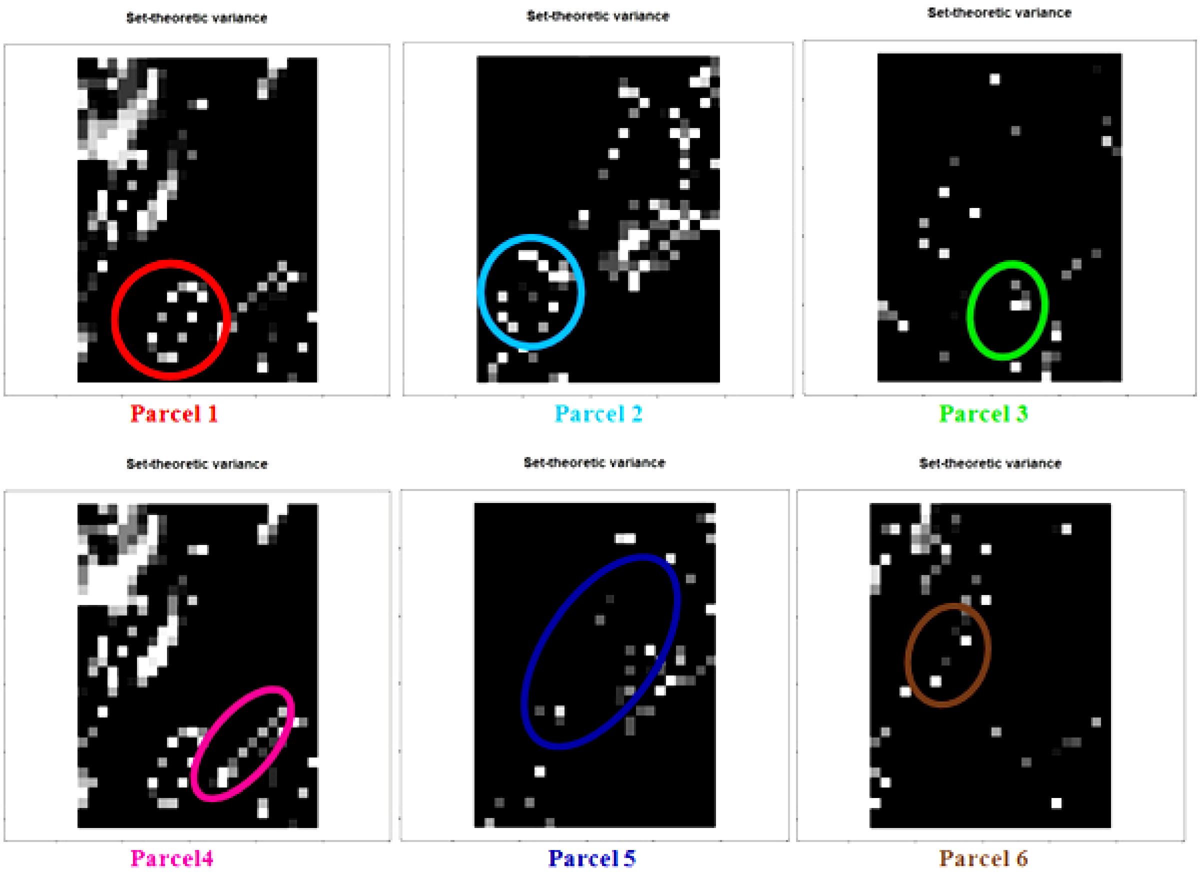


Figure 13. Set-theoretic variance of the random sets for parcels 1-6

5.6.2. Coefficient of variation

From Equation 4.11, since the CV utilized as a normalized and dimensionless measure, we can have comparison between different agricultural fields. A high CV demonstrates a larger proportion of agricultural fields with a high Γ_{var} , then points to a large extensional uncertainty (Zhao, et al., 2011).

Table 3. The values of SD (sum of set-theoretic variance) and CV (coefficient of variation) for the parcels 1-6

Parcel	*SD (Sum of set-theoretic variance)*	*CV (Coefficient of variation)*
1	24.43	0.44
2	15.61	0.19
3	4.88	0.15
4	26.87	0.42
5	5.15	0.08
6	7.37	0.09

Table 3 describes the quantification of extensional uncertainty of six parcels by numeric indicators, the SD (sum of set-theoretic variance) and CV (coefficient of variation). SD values are larger for parcel 1 than for parcel 3 and parcel 5 (Table 3), because the number of pixels with non-zero Γ_{var} values in parcel 1 is larger. As the CV takes the object area into account, the value of CV of parcel 1 is larger than that of parcel 3 which indicates that parcel 1 has a larger proportion of uncertain area. Although parcel 5 has a larger SD than parcel 3, its area is also larger than that of parcel 3, resulting in a lower CV.

Figure 14 present a schematic view of putting mean of the random sets on focused parcels. The results seem to be good enough to delineate the agricultural fields, at least in a glimpse.

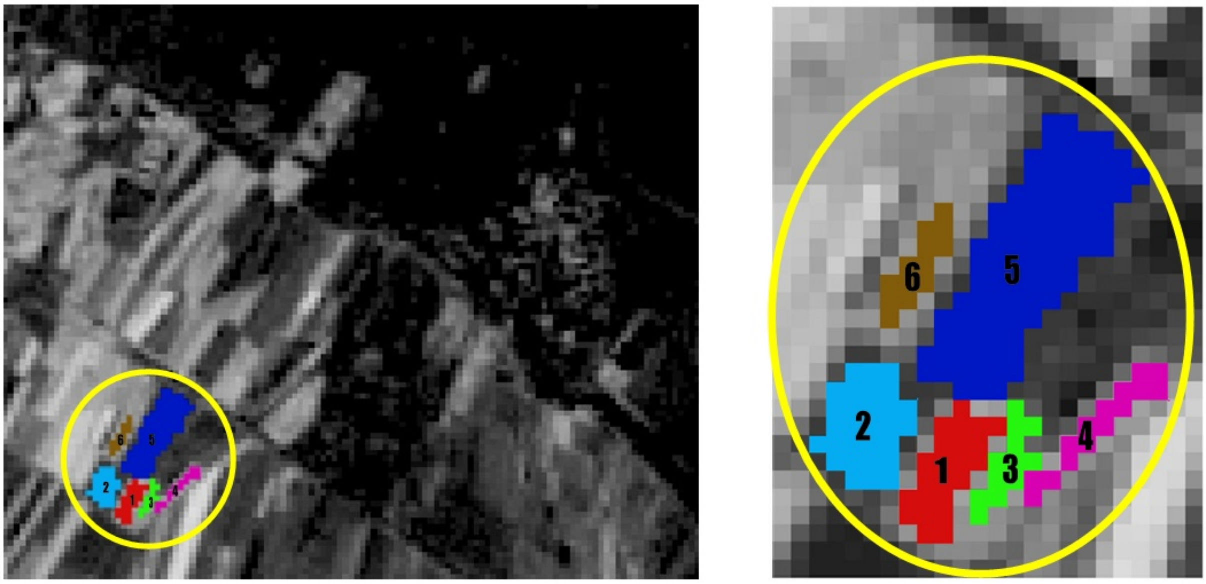


Figure 14. Schematic view of putting mean of the random sets on focused parcels (Parcels 1-6)

5.7. DELINEATION OF AGRICULTURAL FIELD BOUNDARIES

Figure15 illustrates the agricultural field boundaries (parcels 1-6) achieved from contours of the support set of each random region before and after resolving the overlap issue.

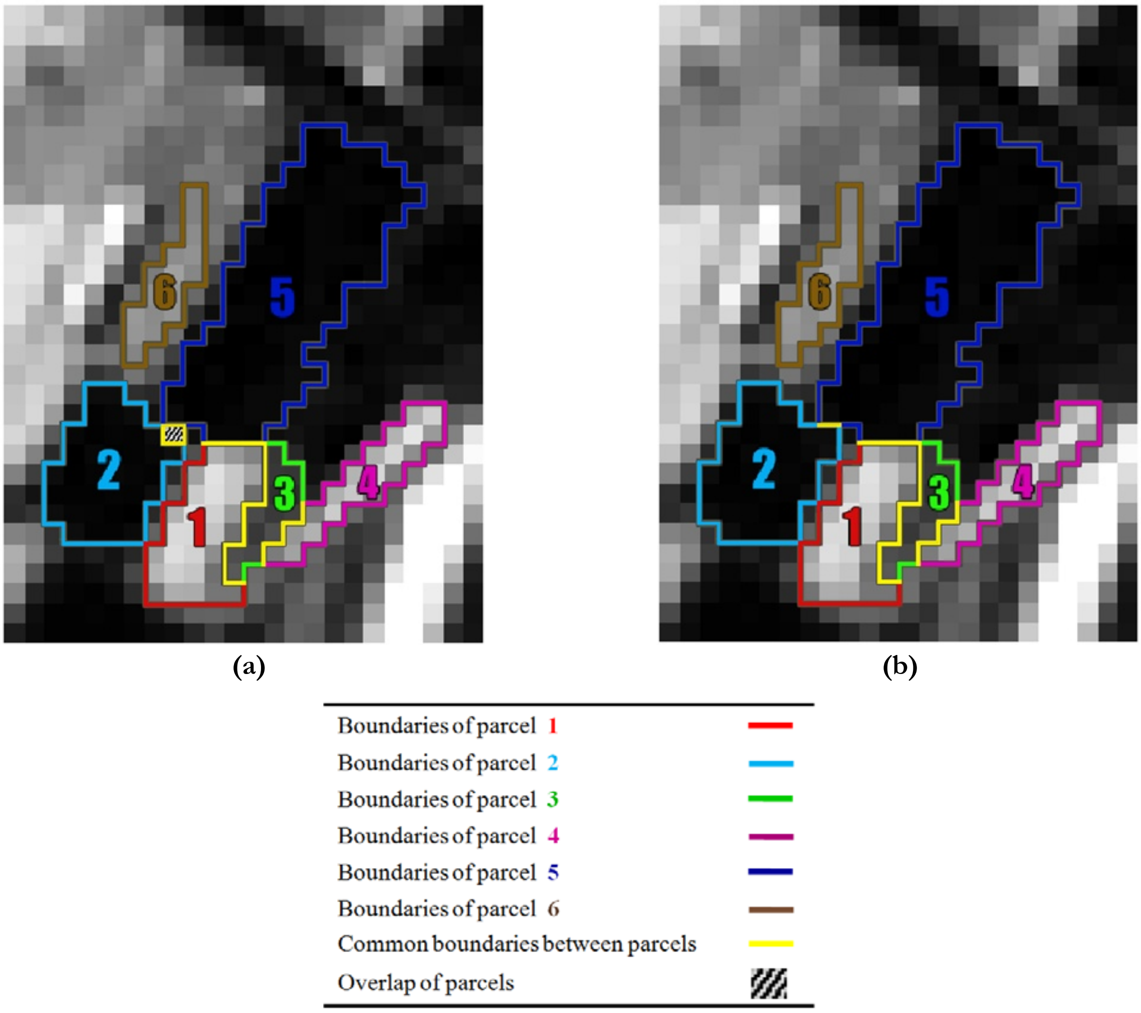


Figure 15. NDVI map from Landsat 5 TM image with the agricultural field boundaries (parcels 1-6) overlaid, (a) before and (b) after resolving the overlap issue of parcels

The contours of support set (Γ_S) which derived from random regions have been used to delineate the parcel boundaries. The support set describes the maximum area of a parcel, in other words describes the possible part of random set (Γ). It is noticeable by using the contours of support set we have some common boundaries between parcels and also some overlap of parcels. Obviously, in Figure 15(a) there are some common boundaries between parcels numbers 1&3, 3&4 and 1&5, also some overlap between parcels 2&5. If we consider on Figure 16 (describing the threshold values for parcels 1 to 6 in 60 iterations), there are two parcels at the bottom of the graph (parcels 2&5). These parcels show very similar values, means they are spectrally so similar to each other and this is the reason of overlap of these two parcels. It is significant that, there are also three parcels in the top of the graph (parcels 1, 4 & 6) show approximately similar values, whereas there is one parcel (parcel 3) with values in between the parcels on the top and bottom of the graph. The three parcels in the top of the graph (parcels 1, 4 & 6) don't share a border between each other, because they are non-adjacent parcels. It is important that, although parcel 3 is not similar to parcels 1&4 spectrally, but because it is adjacent with parcels 1&4, it shares some borders with these two parcels.

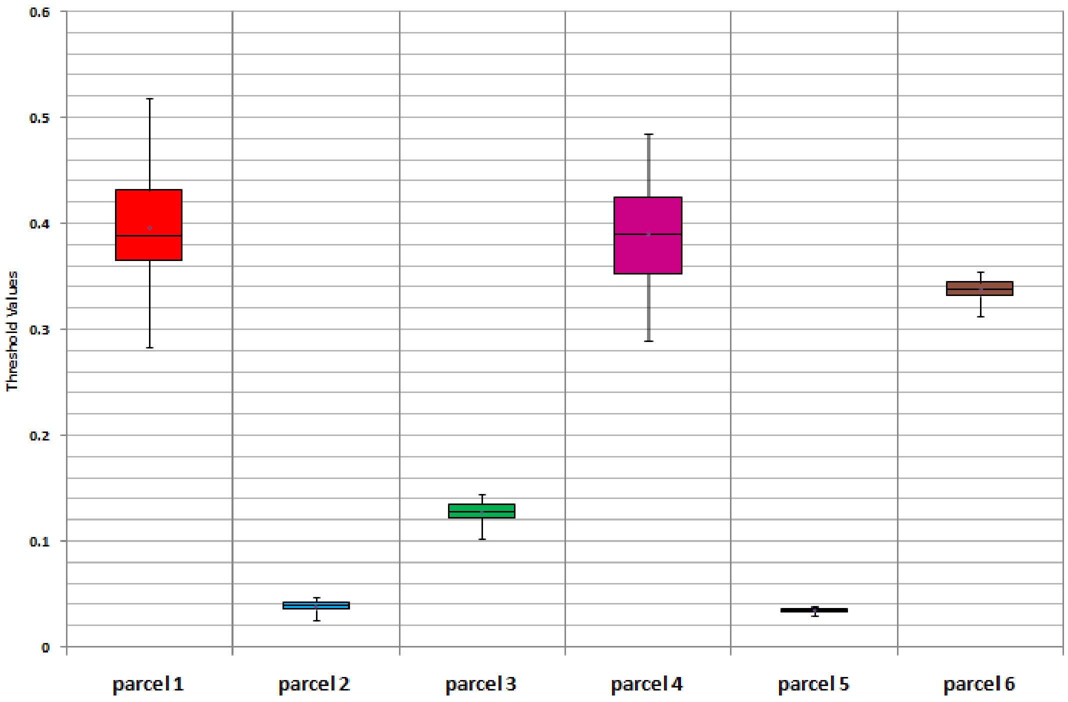


Figure 16. Threshold values for parcels1-6 in 60 iterations

5.7.1. Resolving the overlap issue

For resolving the overlap issue of parcels 2&5 (overlapped in pixel number 513), it is an interesting idea to use the covering function of random sets. An interpretation of the covering function is that it is the probability of a pixel x covered by the random region O. Since different fields (parcels) have different crops with different spectral reflection characteristics, so they have different covering function values at a focused pixel. If we consider on Figure 17, the covering function values for parcels 2 and 5 at the pixel number 513 is almost 0.51 and 0.08, respectively. It means the probability value that the overlapped area covered by the parcel 2, is the most. Figure 15(b) shows the agricultural field boundaries (parcels 1-6) after resolving the overlap issue. Obviously the overlapped area belongs to the parcel 2 and there is also a common boundary with the parcel 5. It is noticeable that, shared border was predictable from the graph of Figure 16.

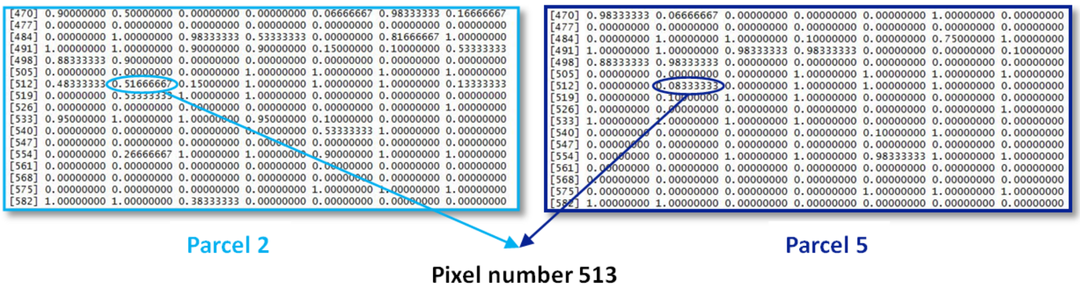


Figure 17. Covering function values of parcels 2&5 at the pixel number 513

5.7.2. Boundaries of the agricultural fields

If we consider on Figure 15(b), obviously the boundaries of the agricultural fields follow the pixel edges and in a whole view it leads to a blocky structure. To better display the boundaries we analyzed the pixel numbers of boundaries with the related covering function values for each pixel. Figure 18 illustrates the boundary between parcels1 and 3 with its related pixel numbers.

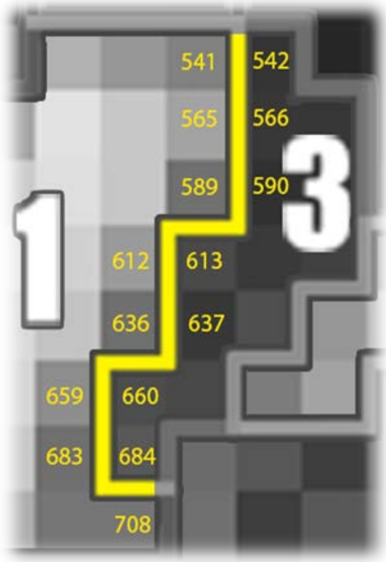


Figure 18. The boundary between parcels 1 and 3 with related pixels

The graph of Figure 19 describes the pixel numbers in horizontal axis and the values of the covering function in vertical axis of the graph. It is significant that, the covering function values of the related pixels of the boundary in the parcel 1 differ from the parcel 3. In other words from the trend of the graph it is obvious that, for displaying the boundaries it is not a must to follow the pixel edges and some methods or algorithms to simplify and better display the boundaries can be use. It is noticeable that, similar analyzes could be done for other boundaries.

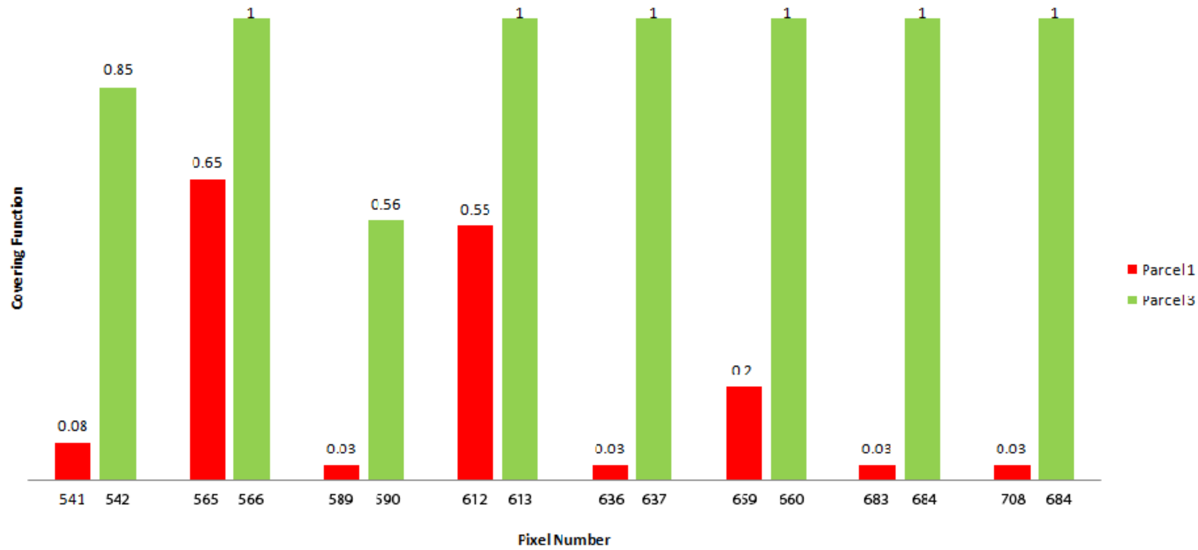
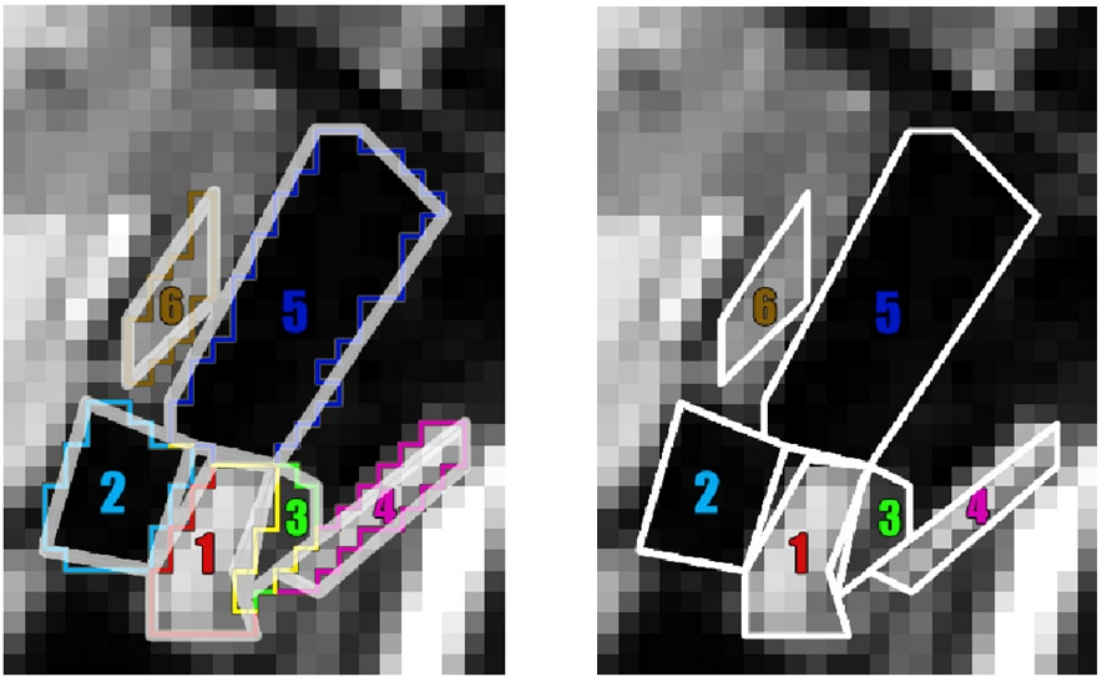


Figure 19. Covering function values for the pixels of the boundary between parcels 1&3

5.7.3. Implementation of the Douglas–Peucker algorithm

To implement the Douglas–Peucker algorithm(Douglas, et al., 1973), we used the ArcGIS software which uses Douglas–Peucker algorithm in its line and polygon simplification process. By choosing the tolerance of 38 meters on the map of the agricultural parcels (Figure 20), the good results have been achieved. Figure 20 illustrates the process, (a) before and (b) after applying the algorithm.



(a) **(b)**

Figure 20. Implementation of the Douglas–Peucker algorithm.(a) before and (b) after applying the algorithm

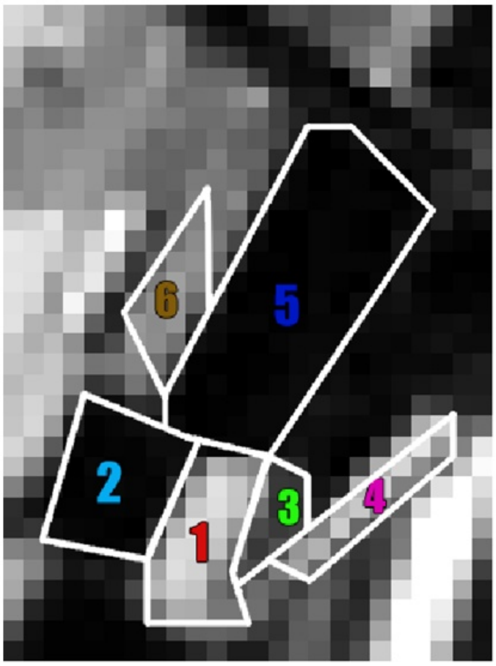


Figure 21. NDVI map with the agricultural field boundaries (parcels 1-6) overlaid, after resolving the small polygons issue

5.7.4. Resolving the issue of small polygons

In Figure 20(b), obviously there are some small polygons between the agricultural fields (i.e. between parcels 1, 2&5 and also between 1&3). There are different ways to resolve this issue like merging the parcels with neighbouring polygons that have the largest area or the longest shared border. In this part of research another approach has been utilized. The approach is as follows: for the triangle between parcels 1, 2 and 5, taking the shortest vertex and finding there the midpoint then combining that point with the sharp angle at the bottom. A similar approach has been used for the triangle between the parcels 1 and 3. At last, combining the bottom point of the parcel 6 with parcel 5 (using orthogonal projection). Figure 21 presents the outcome of this approach.

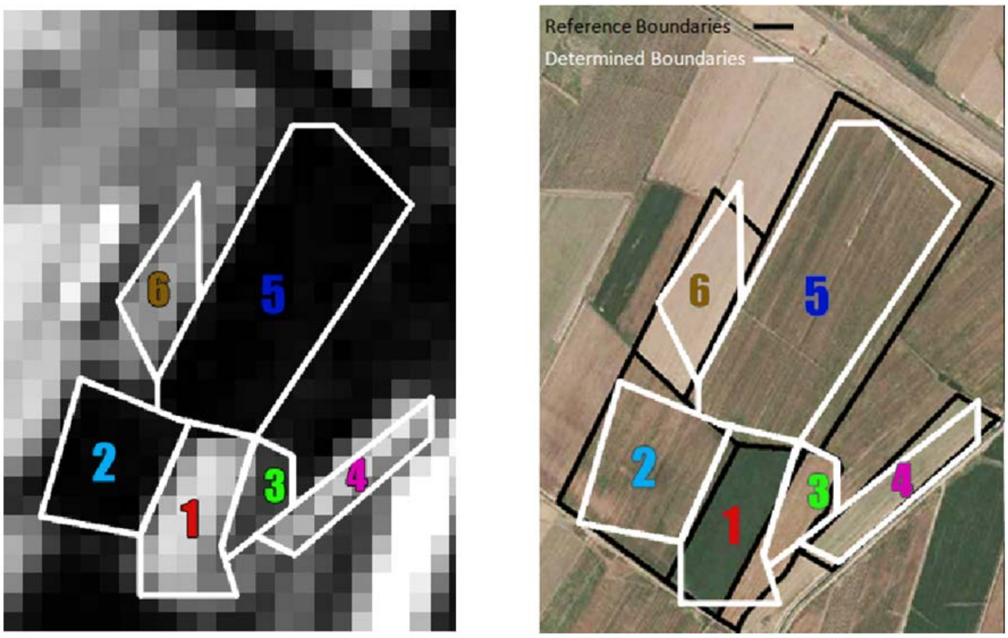
5.8. VALIDATION OF THE RESULTS

To validate the adopted method for delineation of the agricultural field boundaries, reference data are necessary. In this part of research, a georeferenced Google Earth image from 18th of June 2011 has been used as reference data, while the Landsat 5 TM image (base image) acquired on 1th August 2011.In other words the reference image taken 43 days before the Landsat image. Using visual interpretation, it seems that, along this period of time, there isn't any outstanding difference in the size of the objects or boundaries of the agricultural fields. For doing the accuracy assessment and finding the error matrix, first on-screen digitizing on the reference image was done using ArcGIS software. Next, the Landsat image was classified using the segments of obtained agricultural parcels as six sample sets for six classes of classification (using ERDAS Imagine software). The error matrix approach is most frequently used in accuracy assessment(Foody, 2002). The important accuracy assessment elements, such as overall classification accuracy (OCA), producer's accuracy (PA), user's accuracy (UA), and kappa coefficient (KC), can be derived from the error matrix. To generate the error matrix and doing accuracy assessment, 500 random points has been adopted. The analyst examined the random point plots and assigned a class value to each. By comparing these class values with the segments (parcels number) of reference map, the overall accuracy and the error matrix have been achieved. In this part of research the accuracy assessment has been done for validation of the determined boundaries before and after resolving the issue of small polygons and the results have been compared. Figure 22(b) and 23(b) illustrate the Google Earth image with the reference and the determined boundaries before and after resolving the polygons issue, respectively.

From the process of the accuracy assessment (Figure 24), an overall accuracy of 89.2% was achieved with a Kappa coefficient of 0.7930 for the determined boundaries before resolving the overlap issue, while an overall accuracy of 91% with a Kappa coefficient of 0.8360 was achieved for the determined boundaries after resolving the overlap issue. It is noticeable that, from the process, the increase of the value of the overall accuracy was predictable. To sum up, random sets method seems to be an appropriate method for delineation of the agricultural field boundaries. If we compare the study area in the Landsat image with the Google Earth image, obviously finding the exact boundaries of the agricultural fields from the Landsat image (30 by 30 meter) is not an easy task.

(a) **(b)**

Figure 22. (a) NDVI map from the Landsat 5 TM image with the determined boundaries before resolving the polygons issue, overlaid b) Google Earth image with the reference and the determined boundaries before resolving the polygons issue, overlaid (Black boundaries shows the reference boundaries and white boundaries shows the determined boundaries)



(a) **(b)**

Figure 23. (a) NDVI map from the Landsat 5 TM image with the determined boundaries after resolving the polygons issue, overlaid b) Google Earth image with the reference and the determined boundaries after resolving the polygons issue, overlaid (Black boundaries shows the reference boundaries and white boundaries shows the determined boundaries)

ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	313	356	309	---	---
Class 1	8	10	8	100.00%	80.00%
Class 2	38	26	26	68.42%	100.00%
Class 3	10	8	7	70.00%	87.50%
Class 4	15	18	14	93.33%	77.78%
Class 5	92	70	70	76.09%	100.00%
Class 6	24	12	12	50.00%	100.00%
Totals	500	500	446		

Overall Classification Accuracy = 89.20%

----- End of Accuracy Totals -----

KAPPA (K^) STATISTICS

Overall Kappa Statistics = 0.7930

Conditional Kappa for each Category.

Class Name	Kappa
Unclassified	0.6470
Class 1	0.7967
Class 2	1.0000
Class 3	0.8724
Class 4	0.7709
Class 5	1.0000
Class 6	1.0000

----- End of Kappa Statistics -----

(a)

ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	305	335	301	---	---
Class 1	15	21	13	86.67%	61.90%
Class 3	17	12	11	64.71%	91.67%
Class 2	30	24	24	80.00%	100.00%
Class 4	14	14	12	85.71%	85.71%
Class 5	105	87	87	82.86%	100.00%
Class 6	14	7	7	50.00%	100.00%
Totals	500	500	455		

Overall Classification Accuracy = 91.00%

----- End of Accuracy Totals -----

KAPPA (K^) STATISTICS

Overall Kappa Statistics = 0.8360

Conditional Kappa for each Category.

Class Name	Kappa
Unclassified	0.7398
Class 1	0.6073
Class 3	0.9137
Class 2	1.0000
Class 4	0.8530
Class 5	1.0000
Class 6	1.0000

----- End of Kappa Statistics -----

(b)

Figure 24. The overall accuracy and the kappa coefficient values,(a) before and (b) after resolving the polygons issue.

6. DISCUSSION

The topic of properly addressing agricultural field boundaries is addressed in this thesis. The applicability of the random sets method to delineate the agricultural field boundaries is discussed in detail in this chapter.

The first issue to consider is that, to identify the agricultural field boundaries with the neighbouring fields, we realize that fields may have different crop types but there is some spectral similarity in the boundaries of these fields. These boundaries cannot be extracted easily from satellite images (like Landsat images) by means of crisp-based classification, because these methods ignore uncertain areas or transition zones (Bandishoev, 2011). For modeling the uncertainty of the agricultural fields as spatial objects, fuzzy set-based approach comes into mind. As Robinson (2003) reviewed various kinds of membership functions, a major obstacle of fuzzy set approach is determination of the membership function, because the assignment of the membership function is subjective in nature. So in this study, for modeling the extensional uncertainty of the agricultural fields and delineation of the field boundaries, a random sets method has been proposed.

To apply the proposed methodology using random sets method, the NDVI map from the Landsat image was generated. The NDVI is one of the most commonly used vegetation indices. Random sets were generated based on the range of thresholding parameters of segmentation on the NDVI. The program of random sets generation, applied a random number from the normal distribution with mean *(Mean)* and standard deviation (*SD)* values (Table 2). The covering function which is the distribution function of random sets was produced using the segmentation results. The core, median and support sets of random regions were generated based on different level sets, the core set (Γ_{c}) describes the certain part of random set, the support set (Γ_{s}) describes the possible part of random set and the median set $(\Gamma_{0.5})$ is the 0.5-level set. The mean (Γ_{m}) of the random sets were produced using the definition of Vorob'ev expectation, since it considers the set with a finite number of points like the set of pixels in image analysis. The mentioned process was adopted for six parcels (agricultural field) and also it can be done for the other fields of the image (n parcels). For quantifying the statistical parameters of random regions and quantification of extensional uncertainty of the agricultural fields, the values of set-theoretic variance (SD) (Equation 4.10) and coefficient of variation (CV) (Equation 4.11) were calculated (Table 3). The variance of a random variable has been defined as the expected value of the square of the difference between the random variable and the mean (Zhao, et al., 2011). The coefficient of variation (CV) is used as a normalized and dimensionless measure which summarizes the dispersion of the distribution of a random set and allows making comparisons among different agricultural fields. Results (Table 3) presented the parcels number 1 and 4 with larger extensional uncertainty and so larger proportion of uncertain area, than the other parcels.

To delineate the agricultural field boundaries, contours of the support set of each random region were utilized. The support set describes the maximum area of a parcel, in other words describes the possible part of random set. It is significant that, by using the contours of support set, some common boundaries between parcels and also some overlap of parcels were visible (Figure 15). There were some common boundaries between parcels numbers 1&3, 3&4 and 1&5, also the overlap between parcels 2&5. For resolving the overlap issue of parcels 2&5 (overlapped in pixel number 513) (Figure 15), the covering function of random sets was used as an interesting idea. Different fields (parcels) have different crops with different spectral reflection characteristics, so they have different covering function values at a focused pixel (Figure 17). The probability value that the overlapped area was covered by the parcel 2, was the most (0.51). So the overlapped area belonged to the parcel 2. The boundaries of the agricultural fields followed the pixel edges, so to better display the boundaries (escape from the blocky structure) the pixel numbers of boundaries with the related covering function values for each pixel, were analyzed (Figure 19). It showed that some methods or algorithms to simplify and better display the boundaries can be use. The Douglas–Peucker algorithm (chapter 4.8) by the tolerance of 38 meters was utilized to simplify the boundaries and the good results were achieved. It is noticeable that, the issue of the small polygons which generated after the simplification process (Figure 20), was resolved by a general approach (Chapter 5.7.4).

To validate the adopted method for delineation of the agricultural field boundaries, a georeferenced Google Earth image from 18th of June 2011 was used as reference data, while the Landsat 5 TM image (base image) was acquired on 1th August 2011. Using visual interpretation, obviously along that period of time, there weren't any outstanding difference in the size of the objects or boundaries of the agricultural fields. For doing the accuracy assessment and finding the error matrix, on-screen digitizing on the reference image was done (Chapter 5.8). The accuracy assessment was done for validation of the determined boundaries before and after resolving the issue of small polygons and the results were compared.

The results of the accuracy analysis prove the efficiency of the random sets method in delineation of the agricultural field. If we compare the study area in the Landsat image with the Google Earth image, obviously finding the exact boundaries of the agricultural fields from the Landsat image (30 by 30 meter) is not an easy task. The overall accuracy of 91% with 500 control points is an outstanding result and shows the random set as an effective tool to delineate the agricultural field boundaries.

7. CONCLUSIONS AND RECOMMENDATIONS

The main objective of this research was to use random sets for modeling the extensional uncertainty of the agricultural fields and delineate the agricultural field boundaries.

7.1. CONCLUSIONS

This study presented the results of applying the random sets method as an effective tool for modelling the extensional uncertainty of the agricultural fields and delineation of their boundaries. As a starting point it was realized that it is not an easy task to determine the exact boundaries of the agricultural fields from the Landsat image because of its relatively poor resolution. The study is based on the outstanding property of the random sets method is identified. We conclude from this study that random sets are a relatively accurate tool for delineation of the agricultural field boundaries. It resulted into accurate crop acreage estimation and provided relevant parameters to distinguish between fields. It is further concluded that the geometric model used to delineate the agricultural field boundaries is efficient in handling non-rectangular shapes as provided by irregular shape boundaries. Hence, the algorithm can create irregularly shaped segments. This makes the approach generally applicable to a wide range of similar cases.

As far as we know, this study was the first attempt to delineate agricultural field boundaries using random sets, and it was also the first instance that issues of the common boundaries between random regions were faced, being in this case agricultural fields. Moreover, the study showed that random sets are suited to deal with the spectral overlap between regions.

In this research, agricultural field boundaries from a Landsat image have been delineated using the basic parameters of random sets, i.e. the mean, covering function, level sets and variance. By performing the accuracy assessment, outstanding results have been achieved with an overall accuracy of 91%.

This thesis as any research has its limitation and naturally, it can be followed and improved in many ways. Some of which are discussed in the next section.

7.2. RECOMMENDATIONS

In this research six parcels (agricultural fields) were utilized to address the proposed methodology. For further investigation it is recommended to increase the number of parcels (cover a larger area) to see the efficiency of random sets method for delineation of agricultural field boundaries when the number of parcels increases. Another recommendation belongs to the mean of the random sets, which is the generation of the mean of the random sets using other expectations (not only Vorob'ev expectation) to comparison and investigation on different expectations results.

LIST OF REFERENCES

Abkar, A. A. (1999). *Likelihood-Based Segmentation and Classification of remotely sensed images. A Bayesian Optimisation Approach for Combining RS and GIS.* Enschede, The Netherlands: University of Twente, ITC publication number 73.

Abkar, A. A., & Sharifi, M. A. (1995). Knowledge-based classification method for mapping of land cover using high resolution satellite data. *1st Conference on Space Technology and Developing Countries*, (pp. 1-19). Tehran.

Bandishoev, M. M. (2011). *The quality of glacier observation : the debris areas and their role in size estimation.* Enschede: University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC).

Barndorff-Nielsen, O., Kendall, W., & Lieshout, M. (1999). *Stochastic Geometry: Likelihood and Computation.* Chapman & Hall/CRC.

Butenuth, M., Straub, M. B., & Heipke, C. (2004). Automatic extraction of field boundaries from aerial imagery. *KDNet Symposium on Knowledge-Based Services for the Public Sector*, (pp. 3-4).

Cheng, T., & Molenaar, M. (1999). Objects with fuzzy spatial extent. *Photogrammetric Engineering and Remote Sensing , 65*, 797-802.

Cressie, N. (1993). *Statistics for spatial data.* Wiley-Interscience.

Diggle, P. (1981). Binary mosaics and the spatial pattern of heather. *Biometrics , 31*, 531–539.

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization , 10(2)*, 112-122.

Ferencz, C., Bognar, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timar, G., et al. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing , 25*, 4113-4149.

Foody, G. M. (2002). Status of land cover accuracy assessment. *Remote Sensing of Environment , 80*, 185-201.

Gallego, A., & Simo, A. (2003). Random closed set models: estimating and simulating binary images. *Image Analysis & Stereology , 22*, 133–145.

Goodman, I. (1982). Fuzzy sets as equivalence classes of random sets. *In R. Yager, editor, Fuzzy Sets and Possibility Theory: Recent Developments. Permagon Press.*

Goodman, I., Mahler, R., & Nguyen, H. (1997). Mathematics of data fusion. *Kluwer Academic Publishers , 37* (mathematical and statistical methods).

Ishida, T., Itagaki, S., Sasaki, Y., & Ando, H. (2004). Application of wavelet transform for extracting edges of paddy fields from remotely sensed images. *International Journal of Remote Sensing , 25*, 347-357.

Janssen, L. L., & Molenaar, M. (1995). Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on , 33*, 749-758.

Ji, C. (1996). Delineating agricultural field boundaries from TM imagery using dyadic wavelet transforms. *ISPRS journal of photogrammetry and remote sensing , 51*, 268-283.

Kendall, W. S. (2003, July). *madison*. Retrieved February 23, 2014, from University of warwick: http://web.warwick.ac.uk/statsdept/staff/WSK/talks/madison.pdf

Mahler, R. (2007). Statistical multisource-multitarget information fusion. *Artech House, Inc.*

Matheron, G. (1975). *Random sets and integral geometry.* New York: Wiley.

Molchanov, I. (2005). Random closed sets. In M. Bilodeau, F. Meyer, and M. Schmitt, editors. *Space, Structure, and Randomness , LNS183*, 135–149.

Mueller, R., Segl, K., & Kaufmann, H. (2004). . Edge-and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern recognition , 37*, 1619-1628.

Nguyen, H. (2006). *An Introduction to Random Sets.* Chapman & Hall/CRC.

Nuñez-Garcia, J., & Wolkenhauer, O. (2002). Random set system identification. *IEEE Transactions on Fuzzy Systems , 10(3)*, 287–296.

Ortner, M., Descombes, X., & Zerubia, J. (2008). A marked point process of rectangles and segments for automatic analysis of digital elevation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence , 30(1)*, 105–119.

Ortner, M., Descombes, X., & Zerubia, J. (2007). Building outline extraction from digital elevation models using marked point processes. *International Journal of Computer Vision , 72(2)*, 107–132.

Robinson, V. (2003). A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems. *Trans GIS , 7(1)*, 3-30.

Rydberg, A., & Borgefors, G. (2001). Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *Geoscience and Remote Sensing, IEEE Transactions on , 39*, 2514-2520.

Science, Z. (2013, May 9). *tag: google-earth.* Retrieved from ZME Science: http://www.zmescience.com/tag/google-earth/

Serra, J. (1980). The boolean model and random sets. *Computer Vision,Graphics and Image Processing , 12*, 99–126.

Serra, J. (2007). The random spread model. *In 8th International Symposium on Mathematical Morphology .*

Stein, A., Hamm, N., & Ye, Q. (2009). Handling uncertainties in image mining for remote sensing studies. *International Journal of Remote Sensing , 30(20)*, 5365-5382.

Stoyan, D., & Stoyan, H. (1994). *Fractals, random shapes and point fields.* Chichester: Wiley.

Tiwari, P., Pande, H., Kumar, M., & Dadhwal, V. K. (2009). Potential of IRS P-6 LISS IV for agriculture field boundary delineation. *Journal of Applied Remote Sensing , 3*, 033528-033528-9.

Torre, M., & Radeva, P. (2000). Agricultural-field extraction on aerial images by region competition algorithm. *15th International Conference on Pattern Recognition Proceedings.* (pp. 313-316). IEEE.

Turker, M., & Kok, E. H. (2013). Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS journal of photogrammetry and remote sensing , 79*, 106-121.

Vorob'ov. (1996). Random set models of fire spread. *Fire Technology , 32(2)*, 137–173.

Zhao, X. (2012). *Random sets to model uncertainty in remotely sensed objects.* Enschede: University of Twente, Faculty of Geo-Information Science and Earth Observation ITC. (ITC Dissertation 203).

Zhao, X., Stein, A., & Chen, X. (2010). Application of random sets to model uncertainties of natural entities extracted from remote sensing images. *Stochastic Environmental Research and Risk Assessment , 24(5)*, 713-723.

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*, 2548–2557.

Zhou, L. (2014). *Real uncertainty in traffic island polygons extracted from airborne laser point clouds.* Enschede, The Netherlands: University of Twente, Faculty of ITC.

Zhou, L., & Stein, A. (2013). A random set approach for modeling integrated uncertainties of traffic islands derived from airborne laser scanning points. *Photogrammetric Engineering & Remote Sensing , 79(9)*, 835-845.