Centre for Geo-Information

Thesis Report GIRS-2014-02

Agricultural land use patterns and malaria vector abundance

A case study of the spatial relationships between agricultural land use and mosquitoes of the Anopheles gambiae complex and the Anopheles funestus complex on Rusinga Island, Kenya

Annemieke Mulder

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Agricultural land use patterns and malaria vector abundance

A case study of the spatial relationships between agricultural land use and mosquitoes of the Anopheles gambiae complex and the Anopheles funestus complex on Rusinga Island, Kenya

Annemieke Mulder

Registration number 91 08 31 586 080

Supervisors:

Ron van Lammeren

Tobias Homan

Alexandra Hiscox

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

29-05-2014

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2014 - 02
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgements

I want to thank everybody who participated, helped me out, gave me motivation to continue and gave me useful advice. In the first place I want to thank my supervisors for their input: Ron van Lammeren, Tobias Homan and Alexandra Hiscox. Furthermore, I want to thank the people from icipe who assisted me during my fieldwork: Richard Mukabana, Ibrahim Kiche, Thomas, David Owaga, Collins Mweresa, Roselyne Sumba, Jackton Arija, Charles Wambua and Oscar Mbare. I also want to thank some of the teachers of the MGI chairgroup: Harm Bartholomeus and Aldo Bergsma. They helped me out when I was having difficulties with the program ArcGIS and helped me to solve those problems. I also want to thank Planet action (Digital Globe) for providing the Quickbird image and USGS for providing the Landsat image which were used during this research.

Last but not least I want to thank my family and boyfriend for their support during the whole process of starting up, gathering data, being sick, continuing with the analyses and finishing this study.

Abstract

The SolarMal project aims to reduce the malaria vector population of Rusinga Island, Kenya. Within this project there is information about the spatial and temporal component of those vectors, but analyses with those data where never performed. While those components are important determinants of vector abundance and spread. Therefore, the objective of this study was to study the spatial relationship between specific land use patterns and malaria vector abundance for the two crop growing seasons (March and October) separately in the SolarMal study area, Rusinga Island. The first step was to perform land use classifications for both seasons on basis of information about the different crop types occurring on the island. Furthermore data gathered in the field for two study areas in October 2013 and February/March 2014 about the different land use classes was used for this. The classifications were validated by means of an error matrix via which the total accuracy of the different classifications was calculated. The relational analysis by means of ordinary least squares (OLS) was performed in ArcGIS 10.1. The final products are the two agricultural land use classification maps, validation results and the coefficients of determination for the OLS analysis of the relationship between vector abundance and land use patterns.

The validation results for the land use classifications were < 30% in terms of total accuracy. Therefore, the separate crop type classes were merged, the classifications were repeated and the validation was performed again. From the classifications, it became clear that Rusinga Island is heterogeneous and the agricultural fields are mainly located along the lake shore. The hill tops are covered with shrubs and trees and slopes are mainly bare. The accuracy of the new classifications increased to 57% for the whole of Rusinga Island and to 55% and 44% for study areas 1 and 2 respectively for March and 35% for Rusinga Island as a whole and 36% and 43% for study areas 1 and 2 respectively for October. After the OLS analyses were performed, it became clear that there is no relationship between any land use pattern and vector abundance. Other studies indicated that there is a relationship between land use and larvae of the malaria vector, which can be studied in the future by use of the same models. Furthermore, a combination of land use patterns (percentages) near houses could possibly lead to finding a relationship between those land use type combinations and vector abundance.

Table of Contents

1. Introduction

Malaria is studied intensively in terms of entomological and health aspects, but the spatial-temporal component of vector abundance is often lacking (Appendix A). Preventive measures, such as insecticide treated bed nets, do not reduce the malaria vectors and the parasite has an emerging resistance to drugs and insecticides. Therefore, the health risk of malaria increases and other methods are needed to be able to reduce the risk of malaria. One of those methods is studied within the SolarMal project. The aim is to eliminate the malaria vector: mosquitoes of the *Anopheles gambiae s.l.* complex and the *Anopheles funestus* complex, on Rusinga Island (Appendix A – 6). In this project, data about the spatial-temporal component was gathered, but no analyses were done yet. This study focusses on those spatial-temporal component by using geo-information for studying the relationship between vector abundance and agricultural land use types for two different seasons. This is important, since different types of land use affect mosquito habitats differently and affect the abundance and distribution of those mosquitoes.

The vector population is related to space and time (Appendix A-3). Malaria is focussed around specific mosquito vector breeding habitats. In Africa the daily-movement of malaria vectors is typically between a few hundred metres and one km, it rarely exceeds 2-3 km (Carter et al. 2000). Furthermore, certain sites are hotspots of human malaria and clinical symptoms as well as asymptotic cases, usually households or clusters of households (Ernst et al. 2006). Spatial geographical knowledge permits vector control interventions to be targeted since spatial risk areas can be defined. This can increase the effectiveness of control measures (Wen et al. 2006; Moss et al. 2011). The use of geographical information systems (GIS) therefore can assist in targeted interventions against malaria. (Carter et al. 2000) However, the use of geo-information in the SolarMal project and thus within this study is different. It is only used to understand and study vector abundance and malaria transmission and not for adapting the placement of the intervention traps.

Moss et al. (2011) performed a landscape characterization by means of remote sensing (RS). They identified the environmental risk factors for malaria transmission and created a spatial risk map for the Southern Province, Zambia. This map could be used to guide the malaria control interventions. Their conclusion was that only 24% of the households would require malaria control interventions since there was clustering of observed malaria cases near certain land use types, such as third-order streams. This study thus gives more information on understanding vector abundance.

Furthermore, Myers et al. (2009) showed that malaria infection is significantly related to lower elevation areas and further away from administrative centres in the East Sepik province, Papua New Guinea ($P < 0.05$). Their aim was to evaluate geographic parameters, such as lower elevation and greater distance from administrative centre in a rural area in Papua New-Guinea. Another evaluation studied whether those geographic parameters were associated with malarial infection. They stated that knowledge of geography is important when studying insect-borne infectious disease, such as malaria.

Both studies of Moss et al. (2011) and Meyers et al. (2009) did not look at the vector abundance and the relationship between observed abundance and geographical factors such as land use. While there is also a relationship between agricultural land use and vector abundance (Diuk-Wasser et al. 2007). They found that the number of *An. gambiae* increased with irrigated rice cultivation. Young rice areas explained 86% of the inter-village variability abundance. Mutero et al. (2004) also found a relationship between rice cultivation and vector abundance. Furthermore, Beck et al. (1994) found a relationship between swamps, unmanaged pasture and vector abundance.

Rusinga Island has two rainy seasons of different lengths (3 months and 4 months) during which crops are grown by the farmers. There exists a correlation between the abundance of *An. gambiae s.s.* and rainfall in previous months (Mbogo et al. 2003). Since there appears to be an effect of seasonality on vector abundance and spread, it is important to study the difference between the land use and the relationship with vector abundance in those two seasons as well.

Thus, land use in relation with seasonality can be an important factor in explaining vector abundance. Therefore, the geographical spatial aspect possibly is important to get more information on vector abundance and malaria transmission on Rusinga Island.

1.1 Problem definition

The spatial relationship between land use types and vector abundance is important. Land use change affects mosquito habitat and in this way the abundance and distribution of those mosquitoes. (Van Wambeke et al. 2007) Certain habitats provide more favourable and more mosquito habitats than others as is proven by Beck et al. (1994). Swamps and unmanaged pasture are related to a higher vector abundance. The same is true for rice cultivation (Diuk-Wasser et al. 2007; Mutero et al. 2004).

Other studies showing the importance of including spatial relationships in malaria studies with regard to agricultural land use patterns are the studies of Van der Hoek et al. (2001), Boussalis et al. (2012) and Bukhari et al. (2011). For example, the study of Van der Hoek et al. (2001) proved that by means of a different management technique of rice fields the amount of mosquito habitats is reduced and therefore also the amount of vectors. Furthermore, Kebede et al. (2005) showed that the intensity of maize cultivation is associated with an increased risk for local inhabitants of getting malaria.

An explanation is that the variation of landscape characteristics affect the local climate differently. Different land use types influence microclimatic conditions including temperature, evapotranspiration and surface run off differently. Those can be all determinants for mosquito abundance and survivorship. Afrane et al. (2005) and Munga et al. (2006) found that open, treeless habitats experience warmer midday temperatures than forested habitats which leads to a shorter gonotrophic cycle of females of the *Anopheles gambiae* complex (Gonotrophic cycle = taking a blood meal and laying eggs, taking a blood male and laying eggs etc.). The same is found in Uganda. Cultivated fields had higher temperatures compared to natural wetlands and the amount of vectors increased with minimum temperatures (Lindblade et al. 2000). Thus, deforestation and cultivation create favourable conditions for the survival of *An. gambiae* larvae. This makes analysis of land use with regard to vector abundance essential for final malaria risk assessment and reducing malaria prevalence. (Patz and Olson 2006; Munga et al. 2006)

Especially the study of Mutero et al. (2004) showed that only reducing the vectors is not a solution for the reduction in malaria prevalence. As long as there are vectors present as well as people who are infected with the malaria parasite *Plasmodium falciparum,* malaria continues to exist. People need to take prophylaxis to reduce the amount of infections. Mutero et al. (2004) state that opportunities for reduction in vector abundance can be found in the local farming systems and that the success of such a project depends on the participation of the local farming community as well. The SolarMal project aims to reduce the malaria vector population of Rusinga Island, Kenya. The project emphasizes the need of participation of the local inhabitants in the research and in this way inform them about the health risk which malaria causes. Furthermore, the population is involved in decision making, ideas of implementation, assisting in data collection and instalment of traps. Thus, the participation of the local population is represented by the SolarMal project. However, the role of agriculture and the different types of land use within this sector with regard to the vector abundance are important aspects that are still lacking.

The studies carried out so far on malaria vector abundance were rarely based on spatial-temporal aspects by using geo-information. The SolarMal project does contain such GIS analysis including social-demographic data. However, the project is lacking the spatial-temporal relationship of vector abundance with agricultural land use patterns and seasonality (Moore and Carpenter 1999; Vanwambeke et al. 2007). Therefore, the relationship between agricultural land use patterns and the spatial-temporal aspect of vector abundance were analysed by use of geo-information within this study.

1.2 Objective and research questions

The objective of this study was to find out whether there is a spatial relationship between specific agricultural crops and malaria vector abundance for the two crop growing seasons separately in the SolarMal study area: Rusinga Island.

The hypothesis is that crops which need more irrigation in comparison with other crops are associated with different microclimatic conditions. Increased humidity leads to an increase in the amount of mosquito breeding sites (warm water pools) and therefore the amount of vectors in the direct surroundings of that area. Despite the high coverage of bed nets inside the houses for around a decade there is still malaria transmission. Therefore, the expectation is that increased vector densities lead to increased malaria transmission risk. (Ijumba et al. 2002; Klinkenberg et al. 2005; Norris 2004)

The first of two annual crop growing seasons starts in mid-March with sowing and ends in mid-August. The second crop growing season starts mid-August with sowing and the harvesting takes place in December/January. However the exact timing depends on which crop is used and on the timing of the rainy season of that year. (Table 1) When looking at the weather data of 2013 of those periods, it becomes clear that the first crop growing season has more rainfall peaks than the second season, especially in April 2013 (Figure 1). The relative humidity in this period is also higher when comparing it to the second season (above 60%, while in the second season there are also dips below 50%). The temperature for both seasons is comparable and fluctuates around the 25 degrees Celsius. (Kenya Meteorological Department 2013) The higher relative humidity and more rainfall peaks in the first season are an indication that the vector abundance in the first crop growing season probably is higher than in the second crop growing season. This is correlated with the abundance of the *An. gambiae s.l.* complex according to Minakawa et al. (2004) and Mbogo et al (2003).

Table 1: Crop calendar data per crop type used on Rusinga Island, Kenya, including information in which season the crops are planted (FAO 2013a).

Crop	Additional Information	Planting period - onset	Planting period - end	Sowing / Planting rate	Sowing/Planting rate unit
Bean common, dry	First season	15/03	31/03	40-100	kg/ha
Bean common, dry	Second season	15/08	31/08	40-100	kg/ha
Cowpea	Second Season	15/08	31/08	25-35	kg/ha
Tomato	Just before second season	01/08	15/08	1-1.5	kg/ha
Kale	First season	15/03	31/03	0.3	kg/ha
Kale	Second season	15/08	31/08	0.3	kg/ha
Maize	Second season	01/09	15/10	25	kg/ha
Sorghum	First season	01/04	15/04	7-10	kg/ha
Sorghum	Second season	01/09	15/09	7-10	kg/ha
Sweet potato	Just before second season	01/08	31/08	27,000	cuttings/ha
Pepper, sweet		15/04	15/06	5	kg/ha

Crop	Length of the cropping cycle	Harvesting period - onset	Harvesting period - end
Bean common, dry	70-90 days	01/06	30/06
Bean common, dry	70-90 days	01/11	01/12
Cowpea	75-100 days	01/11	20/12
Tomato	60-90 days	01/10	30/11
Kale	45-60 days	01/05	15/08
Kale	45-60 days	01/10	31/01
Maize	110-150 days	01/01	15/03
Sorghum	90-120 days	15/07	15/08
Sorghum	90-120 days	15/12	15/01
Sweet potato	90-120 days	15/11	15/12
Pepper, sweet	75 days	unknown	unknown

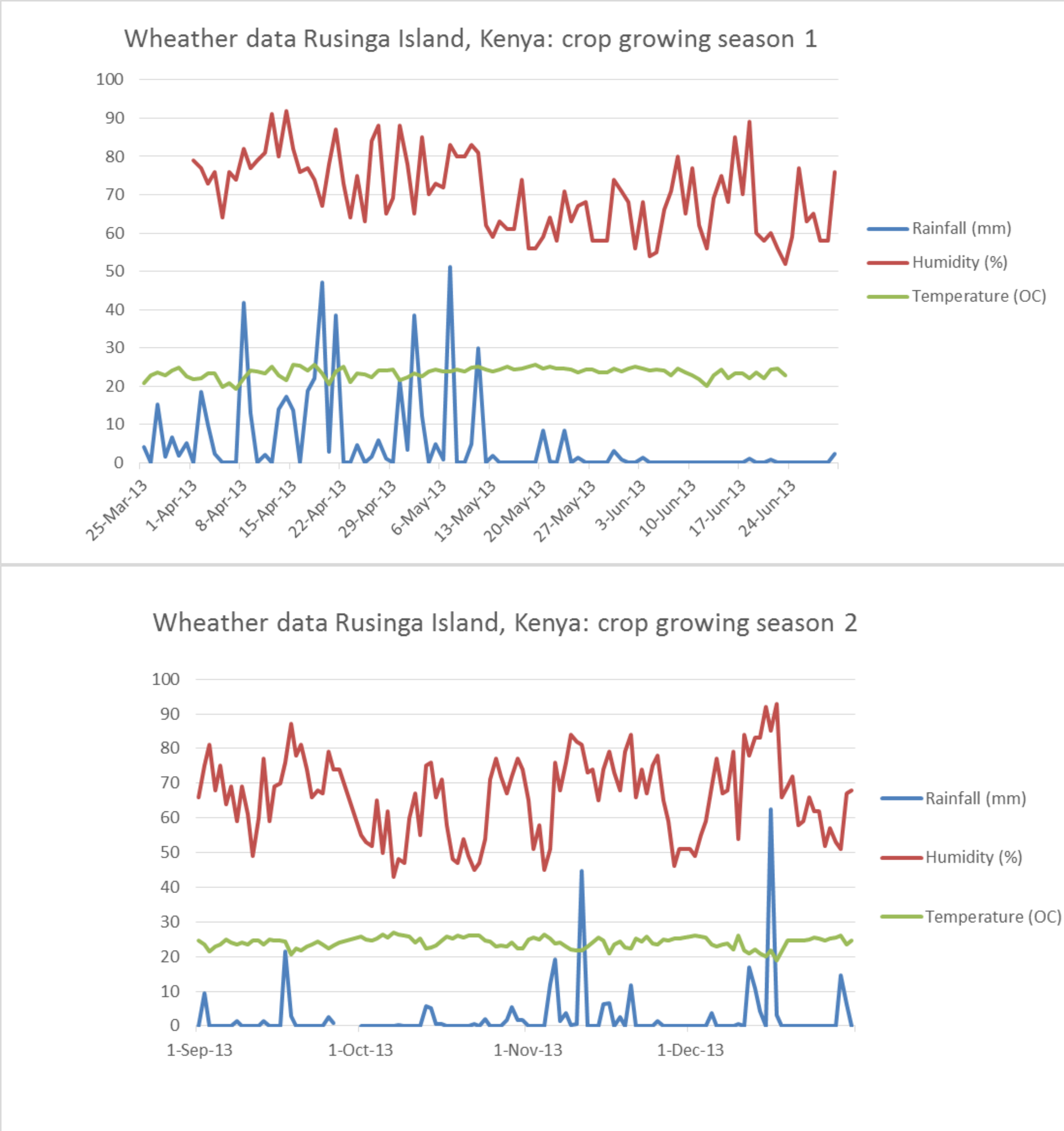


Figure 1: Weather data Rusinga Island, Kenya for the two crop growing seasons, including Rainfall in mm, Humidity as a % and Temperature in degrees Celsius. (Kenya Meteorological Department 2013)

According to the Food and Agriculture Organization of the United Nations (FAO 2013b) and Biovision (2012) the crop that has the highest water demand is maize, assuming that irrigation is necessary. So according to the hypothesis whenever maize is planted in an area, the abundance of vectors is highest in comparison to the other crops. Beans attract the least mosquitoes (Table 2). According to Kebebe et al. (2005) maize is associated with a higher vector abundance, since the maize pollen provide nutrition for larval Anopheles mosquitoes.

Furthermore, mosquitoes need places to hide from their predators. Dense vegetation and thus densely growing crops such as maize and sorghum provide those hiding places (Walton 2003). This can be another reason why mosquito density near especially maize is highest.

Table 2: Water requirements per growing season per crop type, based on data of the (FAO 2013b).

Crop	Water demand per growing season
Maize	500 - 800 mm
Tomatoes	400 - 600 mm
Beans	300 – 500 mm
Watermelon	400 – 600 mm
Kales	~225 mm*
Sorghum	450 – 650 mm
Sweet potatoes	500 – 700 mm

**The FAO did not have data on kales, so used information from Biovision (2012). However, this amount is based on growing kales in Western countries which are less dry than Kenya and the amount was given in cm/week. To get the amount of water needed in mm, the amount was multiplied by 10. The cropping period based on the crop calendar of the FAO was 60 days which is ~9 weeks, so the amount was multiplied by 9, which resulted in 225 mm water needed for the whole growing season*

Since much of the land cover close to homesteads of local families of Rusinga Island is devoted to agriculture and families are depending on them, the relationship between agriculture and vector abundance will be studied during this research by means of the following questions using two smaller study areas (Figure 2 & Chapter 2.1):

1. What agricultural land use patterns* based on crops are typical for Rusinga Island? **The term "land use patterns" is defined as spatial-temporarily delineated units of specific crop classes.*
2. What is the accuracy of the created land use classification maps?
3. Is there a relationship between specific agricultural crops and vector abundance for the two crop growing seasons separately?

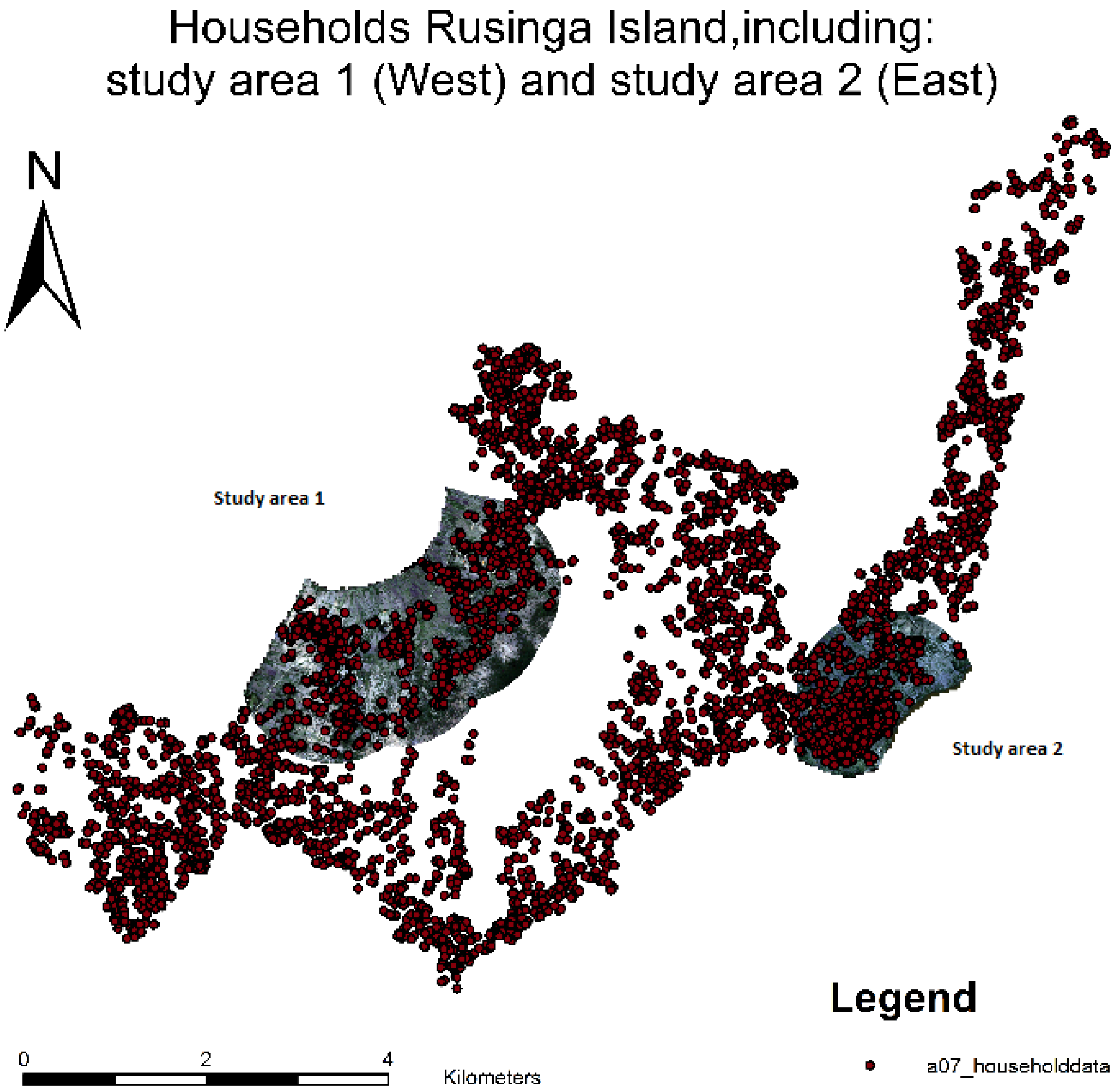


Figure 2: Distribution of households on Rusinga Island, Kenya, based on data gathered in the SolarMal project. Including study area 1 (west of Rusinga Island) and study area 2 (east of Rusinga Island).

1.3 Reading guide

In the methodology section of this thesis report, see Chapter 2, there is an elaboration on how the objectives were reached and how the expectations were studied in more detail. It contains more information about the study area (2.1), the data that is used for the analyses (2.2) and the method for the different types of analyses that were performed: data gathering, classification, validation and regression (2.3). Chapter 2.3 was divided in the seven phases of the flowchart as displayed in Figure 3. Chapter 3 contains the results of the different analysis according to the same set up as Chapter 2.3. Chapter 4 contains the conclusion, discussion and recommendation, after which the list of used literature is given together with the appendices which contain additional information according to the written text.

Furthermore, some references to a DVD are present in the text. On this DVD, the models built for the analyses of this study are included, together with PDF's of the literature used and additional data sources used.

2. Methodology

The steps taken to address the primary research objective of this thesis are given by the methodological framework (Figure 3).

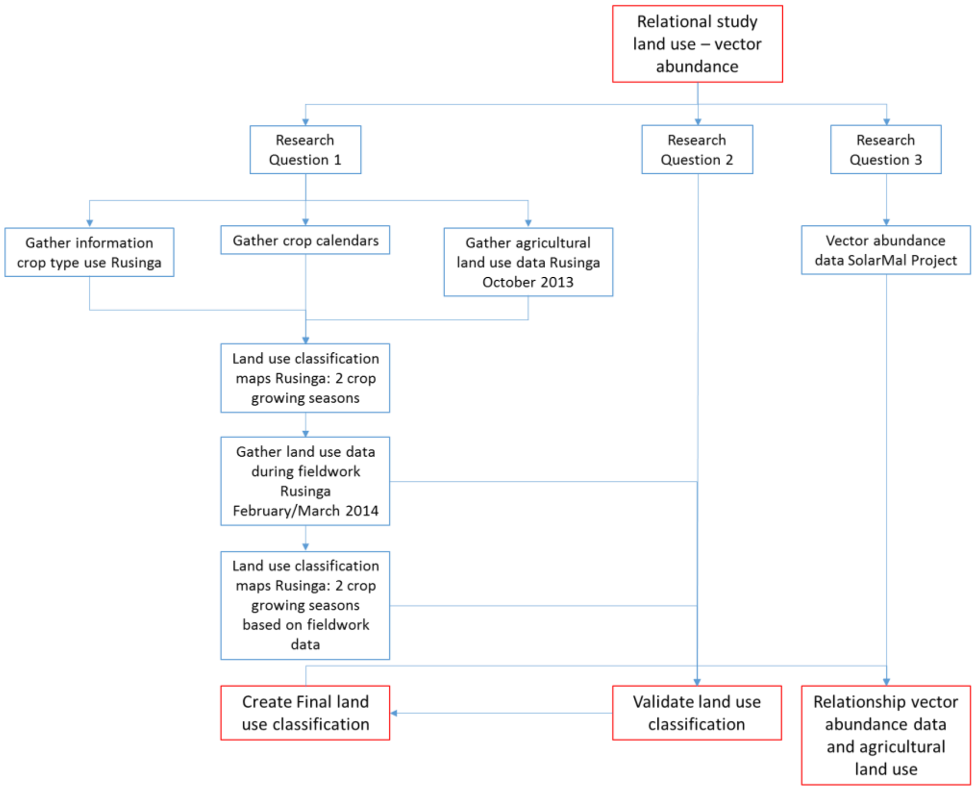


Figure 3: Flowchart global methodology: from research questions towards end products.

To get the final product of research question 1, collection of information on the agricultural land use patterns of the case study area, Rusinga Island, was needed. Furthermore, information was gathered about the different crop types that were used on the island. With this information, a land use classification was carried out for two crop growing seasons (March 2012 and October 2012) on the island and this map was validated by means of data gathered on Rusinga Island, Kenya. The validation was the result of research question 2 for which the accuracy of the classifications needed to be assessed. The newly created land use classification maps based on this validation were the results of research question 1. A validation was necessary since there was no land use map existing yet and accurate land use maps were needed for the study areas as defined in Chapter 2.1 (Hightower et al. 2000;Sipe and Dale 2003), which were interpolated for the whole island. Following the completion of land use classification the analysis for research question 3 of finding a relationship between vector abundance and agricultural land use was performed. This analysis contained a relational study between the agricultural land use patterns and the data on vector abundance on the island gathered during the SolarMal project. (Figure 3) Figure 4 shows the different activities done during this study divided in different phases.

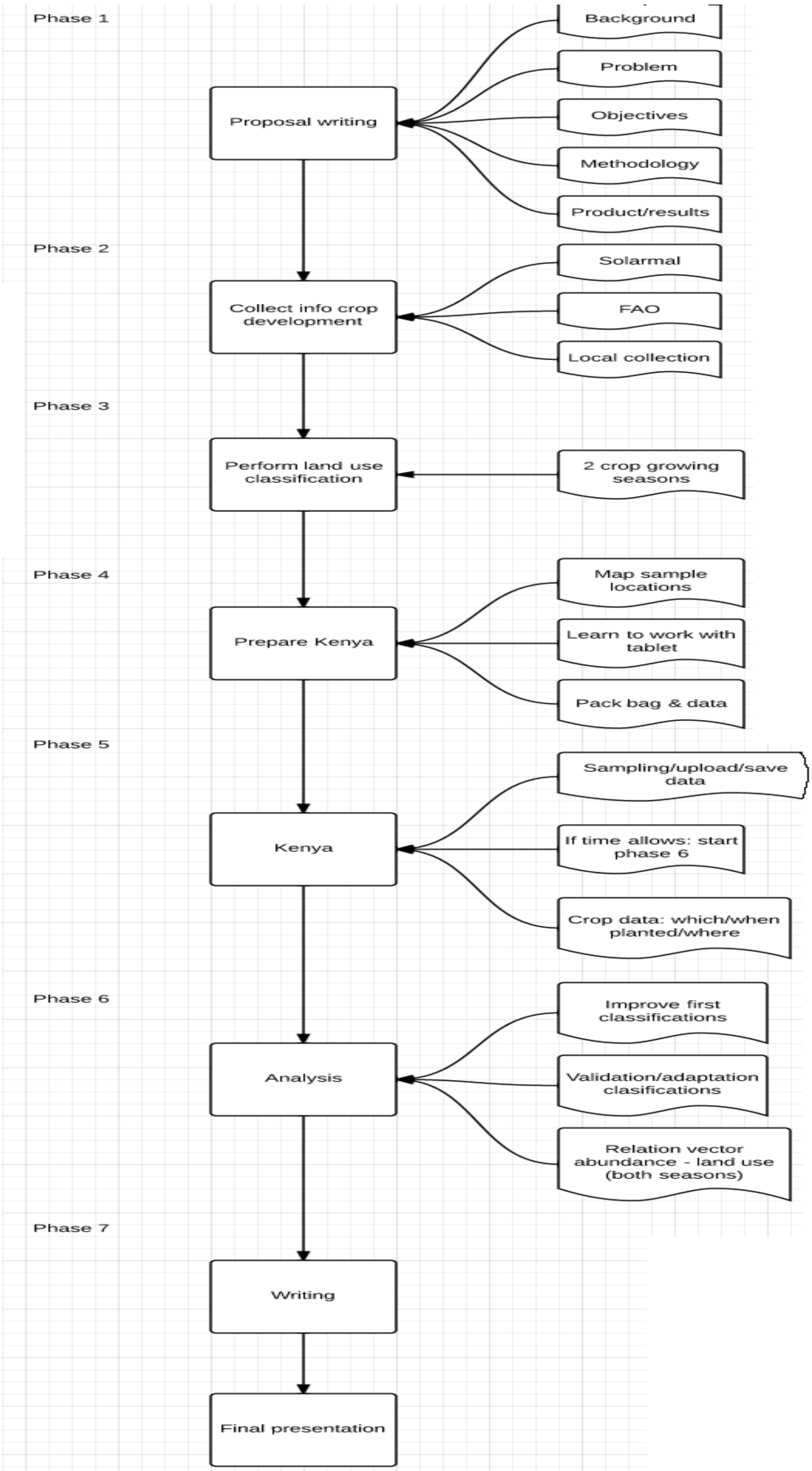


Figure 4: Flowchart methodology

2.1 Study area

Rusinga Island is located along the shores of Lake Victoria, Western Kenya in the Mbita division part of the South Nyanza district. This island was recently connected to the mainland through a causeway. Figure 5 shows the location of this island which is approximately 44 km^2 in size and lays at a mean elevation of 1125 m above sea level. The island population consists of 23,335 individuals and 4,062 households according to the census data gathered during the baseline phase of the SolarMal project (May 2012). Most households are found in the flat areas around the periphery of the island. The hill at the centre of the island is uninhabited, which is also displayed in Appendix B, Figure 38. The primary occupations of the local inhabitants are fishing and small-scale agriculture. Most people in the area are subsidence farmers with sorghum, maize, tomatoes and sukumawiki (kale) as the main crops. (Conelly 1994; Hiscox et al. 2012a)

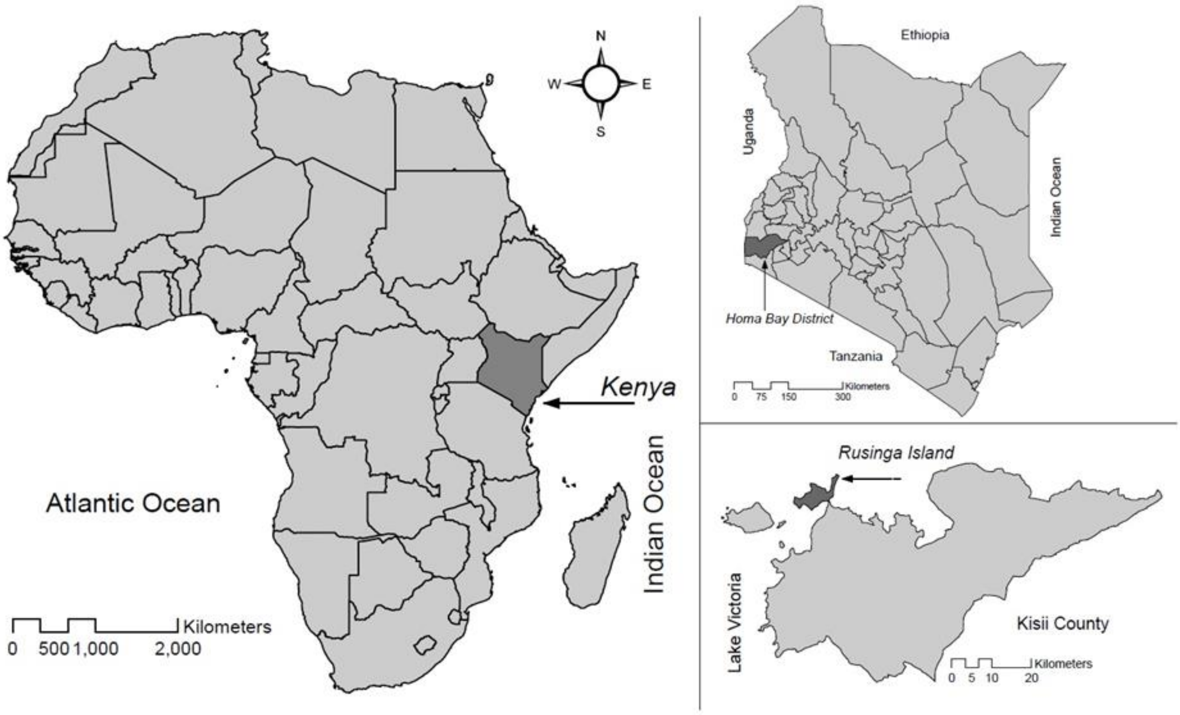


Figure 5: Rusinga Island, Kenya which is the study area of the SolarMal project. 1) Africa with its countries (Kenya), 2 Kenya with its counties (Homa Bay district), 3) Homa Bay county with Rusinga island (34 7' 5" E , 0 26'46" S / 34 14' 8" E, 0 20' 35" S).

A large part of the island is covered with rolling terrain and steep slopes (Figure 6). The average rainfall on the island lies between 700 and 1000 mm per year. There are traditionally two rainy seasons with a long rainy season from March to June and a short rainy season from October to December. During those rainy seasons, crops are planted and harvested (Table 1). Even during the long rainy season, precipitation is often insufficient to guarantee a good harvest. The average rainfall during that period is 535 mm in total for that season and the main crops used in that season are sorghum and maize. The land is divided in strips of approximately 50 m long and 20 m wide which are mostly located from the hillsides down to the lakeshore.

Figure 6: Topographical map of Rusinga Island, Kenya which is the study area of the SolarMal project. (34 7' 5" E , 0 26'46" S / 34 14' 8" E, 0 20' 35" S).

The agricultural fields of Rusinga Island are of particular interest according to the problem definition and objective mentioned in Chapter 1.2. To be able to study the relationship between agricultural crops and mosquito abundance and malaria prevalence, two study areas on Rusinga Island were defined. Those two study areas were defined in such a way that the first study area was located at a place where a high density of malaria vectors was found and where farmers were also present. The second study area was located at a place where the density of malaria vectors is low, but farmers and thus agricultural fields were present.

The vector abundance data and locations of farmers were already gathered by the SolarMal project and also mapped based on data gathered in April and June 2012. Figure 39, 40 & 41 in Appendix B, show those maps, including place indication of the two study areas. The map of Figure 7 which contains a hotspot analysis of the vector abundance data, shows that the hotspots (highest amount) of mosquito catches were found in the area which was defined as initial study area 1 and the coldspots (lowest amount) of mosquito catches were found in the area defined as initial study area 2. The method of this hotspot analysis is described below.

Using the WGS84 projection system the locations of the two study areas were as follows:

Study area 1:
X min: 34.147 dd
Y min: -0.418 dd
X max: 34.166 dd
Y max: -0.397 dd

Study area 2:
X min: 34.205
Y min: -0.418
X max: 34.217
Y max: -0.4

Those two study areas were used for doing the fieldwork on Rusinga Island. The two study areas were adapted by means of a hotspot analysis for the analyses that followed after the fieldwork (land use classification improvement, validation and regression) to improve the classification results (Chapter 2.3.4). This hotspot analysis determined where most of the vectors were (the hotspots) and whether those hotspots were significant or not. The two study areas described above were used to select the mosquito data for which the hotspot analysis was performed, since both areas contained agricultural fields, study area 1 contained a high vector abundance and study area 2 contained a low vector abundance.

Hotspots were defined as areas with a positive z-score > 1 and a p-value < 0.05 according to the Getis-Ord local statistic (Esri 2013a). Based on the results of this test, the first new study area 1, was defined in combination with the location of the initial study area 1. After the selection of those hotspots, the hotspots were buffered with a radius of 1 km, those buffers were merged and the "new" Study Area 1 was used for further analyses. The radius of 1 km was based on the estimation that mosquitoes rarely fly further than 1 km in search of blood according to a study of Beck et al. (1994) and Carter et al. (2000) The same was done for study area 2. However, instead of hotspots, so called coldspots were located in initial study area 2 according to the following conditions: a negative z-score < -1 and a p-value < 0.05 according to the Getis-Ord local statistic. Most coldspots were located in areas where there was no agriculture and outside initial study area 2. Therefore, the "new" study area 2 shifted to the right were coldspots still were found inside initial study area 2, but were not significant any more (0.05 > p-value < 0.1). The final result contained the hotspot analysis and the "new" study areas which can be found in Figure 8. In the end, the goal was to interpolate the analyses done for those two areas to the whole of Rusinga Island.

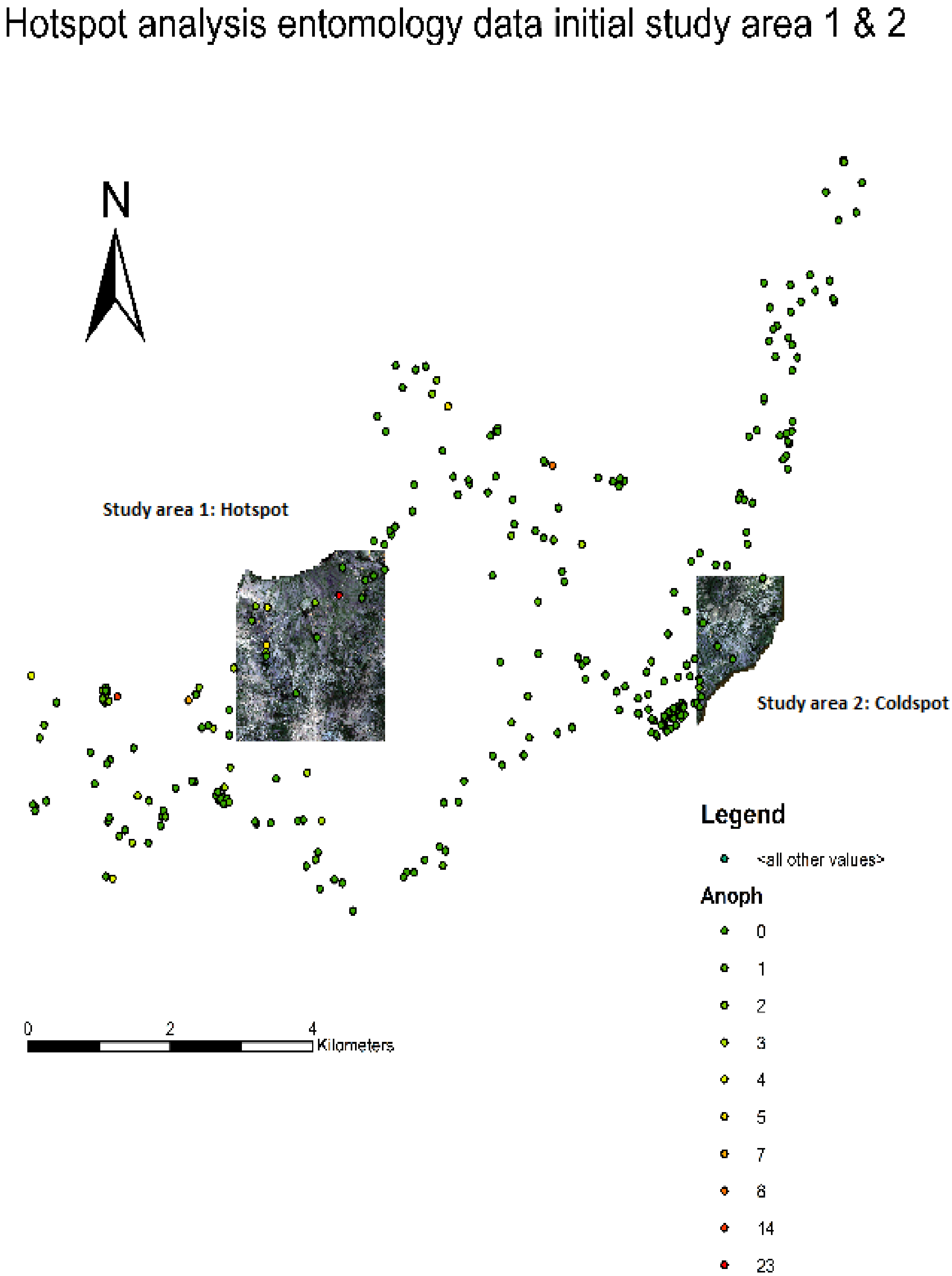


Figure 7: Entomology data SolarMal hotspot analysis indoor mosquito catches: most hotspots can be found in "initial" study area 1 and the coldspots in "initial" study area 2. (Based on Quikcbird image March)

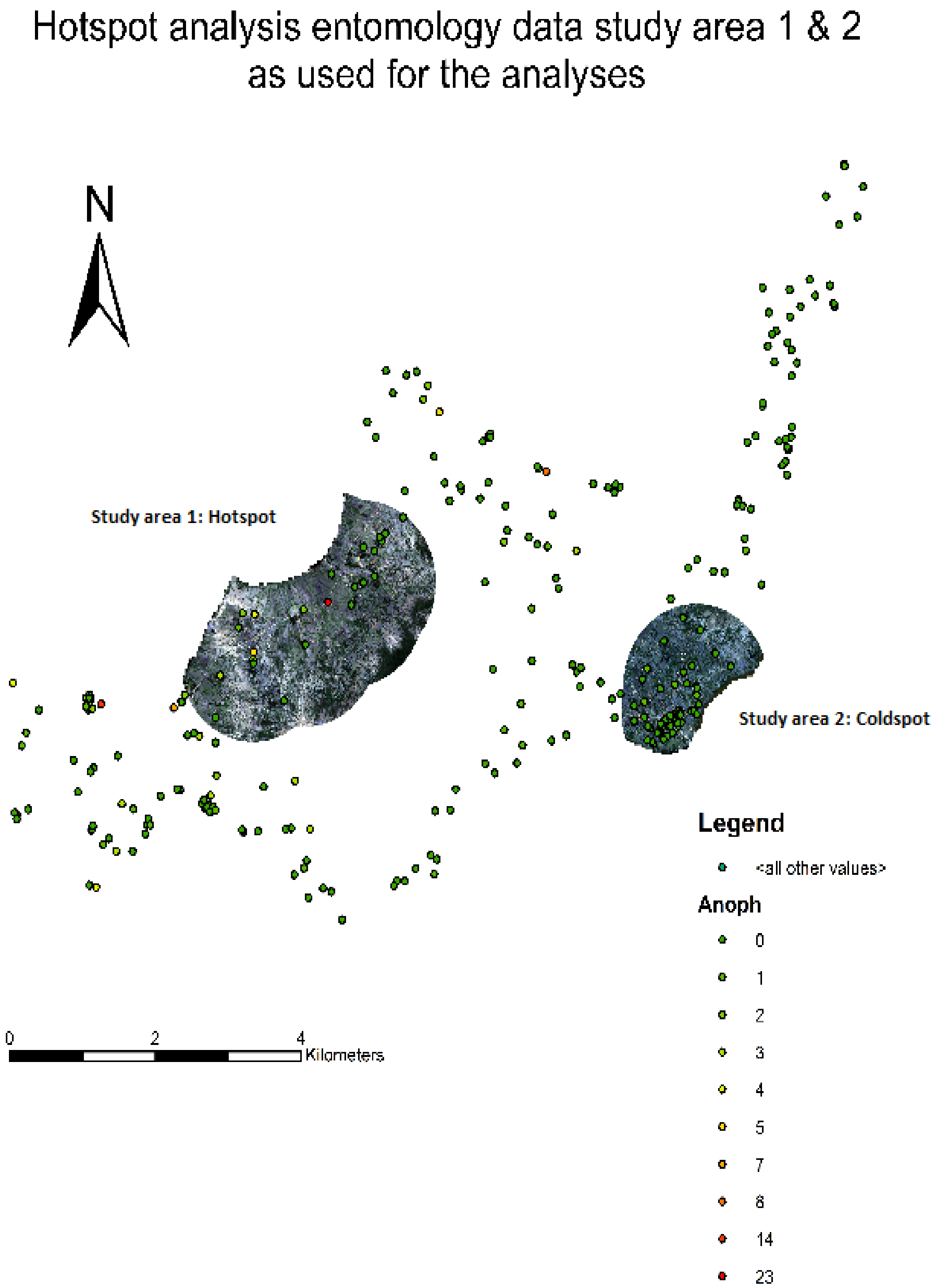


Figure 8: Entomology data SolarMal hotspot analysis indoor mosquito catches used for the creation of the two new study areas which are also displayed: most hotspots can be found in "new" study area 1 and the least in "new" study area 2. (Based on Quickbird image March)

2.2 Data

During this study the following datasets were used:

1. **Research question 1** (Figure 3)
   - Crop type information via Ibrahim Kiche
   - Crop calendars (planting period, sowing/planting rate, sowing/planting rate unit, preferred sowing/planting period, length of cropping cycle and harvesting period, Table 1, Chapter 1.2). (FAO 2013a)
   - Agrozone information of Rusinga Island = (lowlands – midland – humid zone) (Biovision 2012).
   - Data on locations (x- and y- coordinates) agricultural fields per crop type on Rusinga Island, Kenya, gathered from 15th - 18th of October 2013.
   - Quickbird satellite images of Rusinga Island (34 7' 5" E , 0 26'46" S / 34 14' 8" E, 0 20' 35" S) provided by Digital Globe. The standard of March 2012 was used, which already had radiometric and sensor correction. Ortho-rectification was not conducted yet, however an OR2a product was delivered to do this. Geometric correction was also not done, but the image had the basic spheroid WGS84. Both the panchromatic and the multispectral (4-band) images were used and both were 16-bit. The panchromatic image had a pixel resolution of 0.6 m. and one image band was panchromatic. The multispectral image had a pixel resolution of 2.4 m. and the image bands were blue, green, red and NIR. The file formats in use were TIF (for coordinates in meters) and TIL (for coordinates in degrees). (Digital globe 2013)
   - The same satellite image was needed, but than for October 2012. This image needed to be requested from Digital Globe by SolarMal. However, it was not possible to request a new image for this period. Therefore, Landsat ETM+ 7 imagery was used for Rusinga Island. This Landsat image contained a panchromatic band with 15 m spatial resolution and in the visible bands (blue, green, red, NIR and MIR) 30 m spatial resolution. Furthermore, there was a TIR band with 60 m spatial resolution. The satellite had 5% absolute radiometric calibration and the file format that was used is GEOTIFF. Although the Scan Line Corrector (OLC) failed, Rusinga Island was free from stripes, so the image from the first of October 2012 was used directly for classification (USGS 2014).
   - Dataset on land use patterns of October and March, gathered in February and March 2014 in the two study areas on Rusinga Island, Kenya. This gathering was based on predefined sampling locations.

2. **Research question 2** (Figure 3)
   - Quickbird satellite image, as described above.
   - Landsat satellite image, as described above.
   - Dataset on land use patterns gathered in February and March 2014 as described above

3. **Research question 3** (Figure 3)
    - Vector abundance data (SolarMal project). This dataset contains data on the amount of indoor and outdoor Anopheles catches including the exact GPS locations of those traps. This data was gathered using MMX traps in 80 randomly selected households every month from 2012 till 2015 on Rusinga Island, Kenya. For this study, the most recent data closest to March and closest to October were used for the analysis of all female *Anopheles gambiae* and *Anopheles funestus* together in that period, but also of the female *Anopheles gambiae* and *Anopheles funestus* mosquitoes separately. (Hiscox et al. 2012a)
    - Vector abundance data gathered from 07-09-2012 till 10-12-2013, still with the separation between *Anopheles gambiae*, *Anopheles funestus* and both vectors in total for the analysis. (Hiscox et al. 2012a)

2.3 Methods

To achieve the objectives and answer the research questions, tailor made methods were developed. Since the analyses was done in ArcGIS, all the models/toolsets to which is referred in the text are attached with a DVD. An overview of the toolsets and models used is given in Figure 9 a, b and c. Appendix D gives a description of the choices made within the particular models. In Figure 3 and 4 the flowcharts were given for the work that was completed during the study.

2.3.1 Research question 1

The first research question focussed on which agricultural land use patterns occur on Rusinga Island, Kenya. First, the obtained satellite images needed to be pre-processed. To be able to answer this question, data on the different crop types used on the island were collected. Furthermore, an unsupervised classification was done and fieldwork was performed to obtain more information about the land use patterns on Rusinga Island. Afterwards a supervised classification was carried out for March and October 2012 based on which the question was answered.

2.3.1.1 pre-processing

Figure 9a displays the toolset and the models that were used for the pre-processing of the satellite images. First, pre-processing of the Quickbird image of March was necessary. Then the study areas for this image were defined according to the hotspot analysis in which places with a lot of mosquitoes (hotspots) and places with no mosquitoes (coldspots) were defined. (Chapter 2.1)

2.3.1.2 Collect information on crop development

To identify the land use patterns on Rusinga Island, Kenya, information gathering on the different crop types used on the island is necessary. For this, the information of Ibrahim Kiche and the collected crop calendars were used (Chapter 2.2). Furthermore, the data about different crop types that was gathered by a fieldworker in October 2013 was used.

The data of October 2013 was collected by means of a Samsung tablet with Android OS (~10m accuracy). The field worker visited the two study areas (Chapter 2.1) and took one GPS point per field. The GPS point was taken in the middle of the agricultural fields and the fieldworker asked the owners of the fields which crops were growing there in March and which were growing there in October. This was written down per GPS point. The process was repeated until 4 fields were found for each crop in March and another 4 in October. Figure 10 displays how the sampling point was taken in the field and a sketch figure of the fieldwork was added in Appendix C. If the field had another shape as displayed in Figure 10 and it was possible to measure the length and width, the midpoint was determined based on the measurements (length/2 and width/2) and the ropes placed in a straight line from those midpoints onwards. If this was impossible as well, the midpoint of the field was estimated visually.

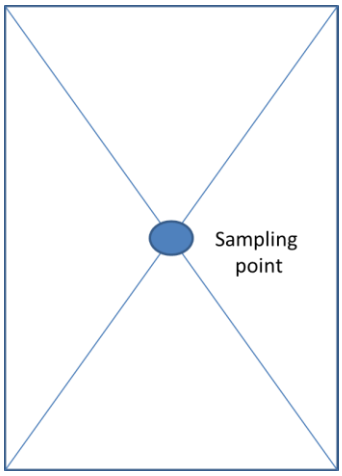


Figure 8: Sampling methods fields.

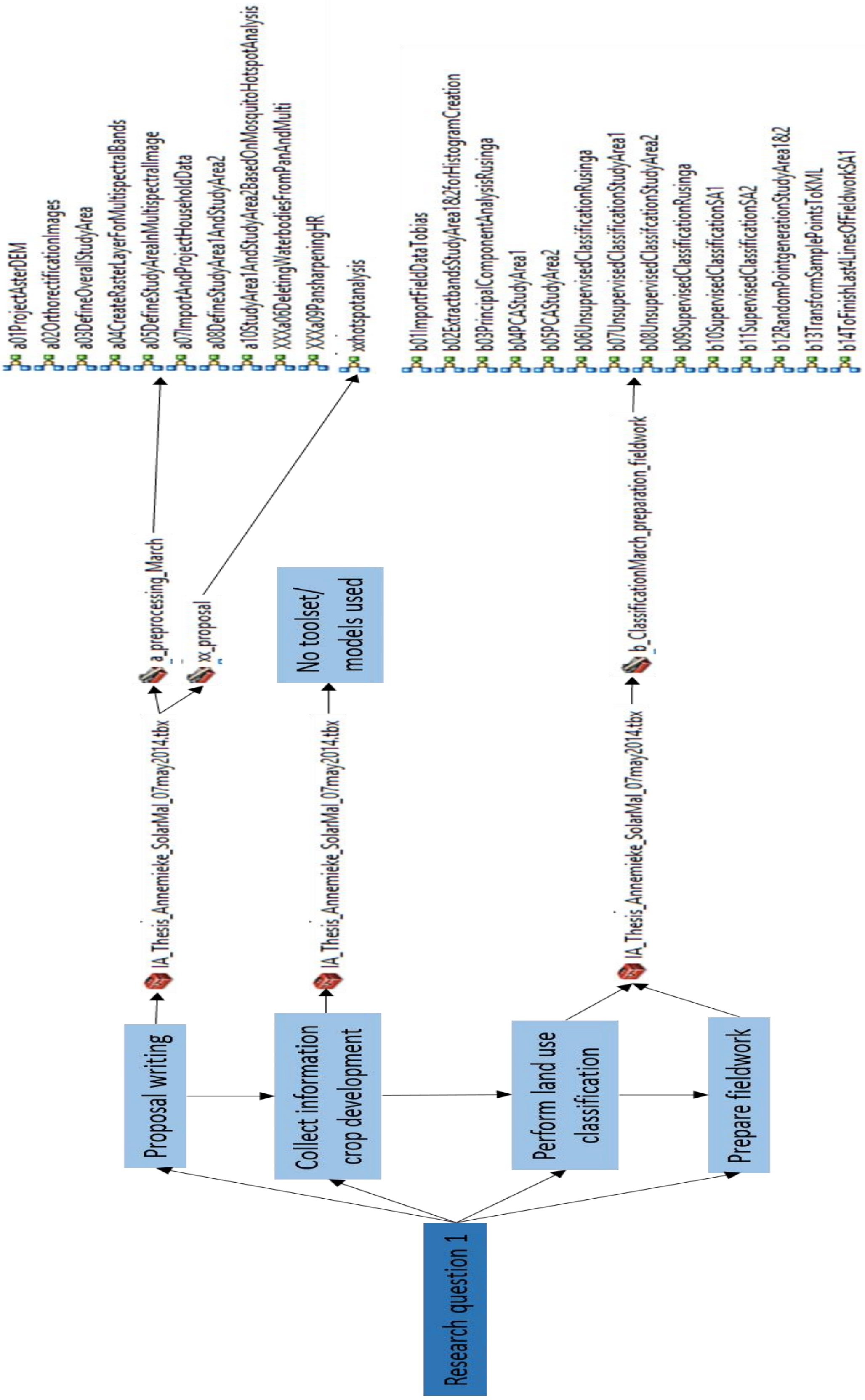


Figure 9a: Phase 1 – 4 including what was done per phase, which toolbox was used, which toolsets were created and which models were created within ArcGIS 10.1.

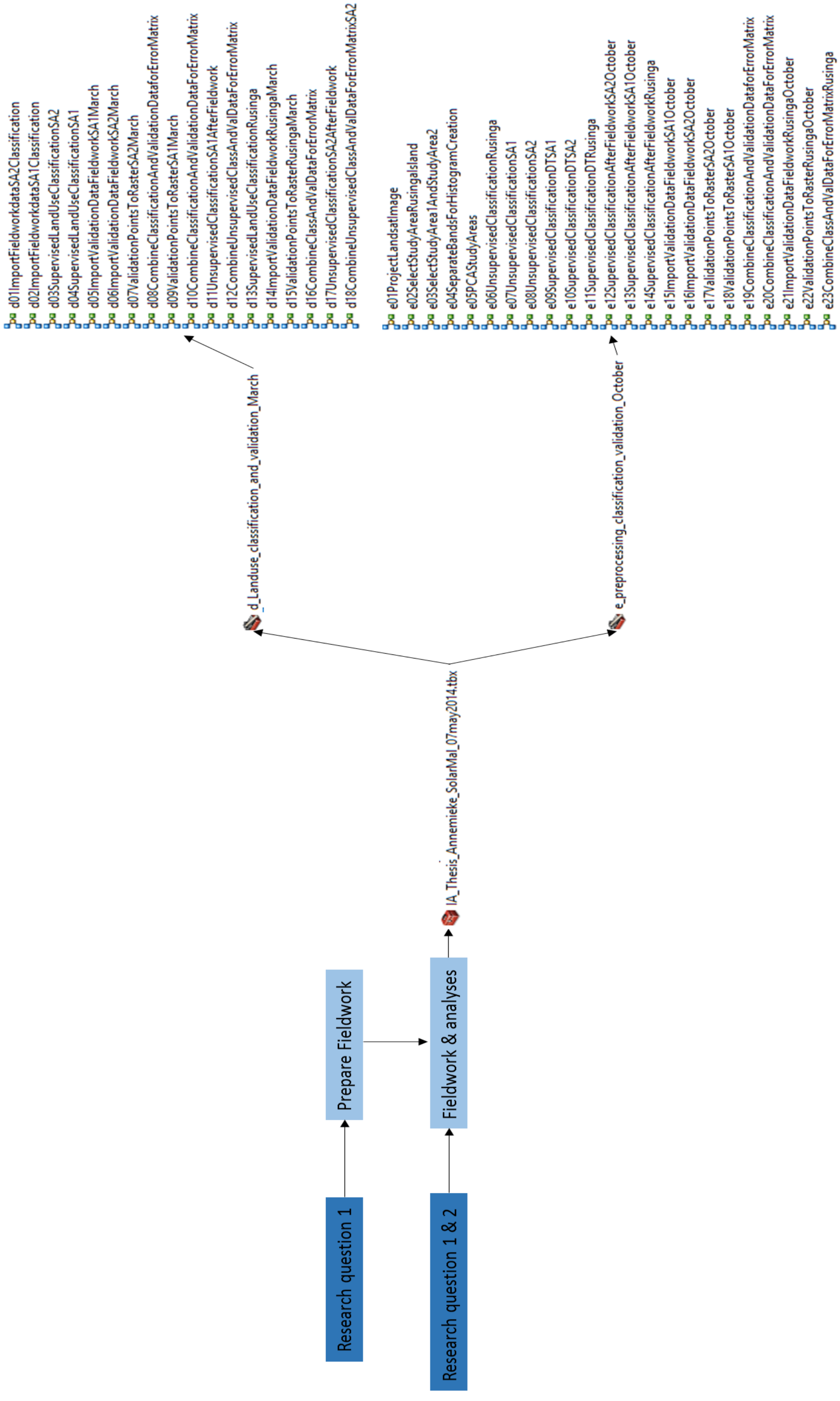


Figure 9b: Phase 4, 5 & 6 including what was done per phase, which toolbox was used, which toolsets were created and which models were created within ArcGIS 10.1.

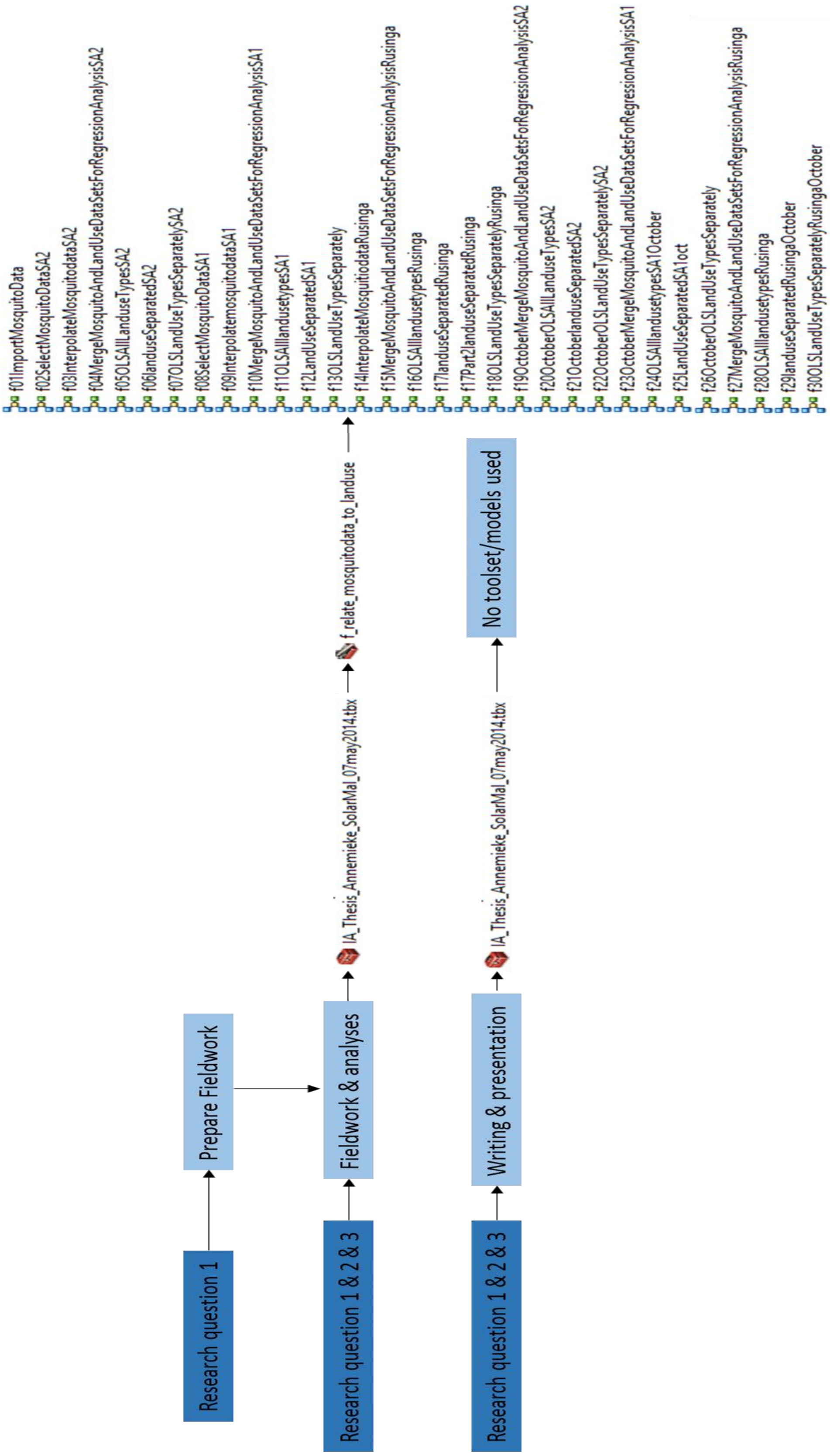


Figure 9c: Phase 4, 5 & 6 and 7 & 8 including what was done per phase, which toolbox was used, which toolsets were created and which models were created within ArcGIS 10.1.

2.3.1.3 Classification & preparation of fieldwork

After the pre-processing and data collection, the first unsupervised and supervised land use classifications were performed for the Quickbird image of March (Figure 9a). The analysis for October was done later, since it was not available yet. Both the unsupervised and the supervised classification were performed after the field data of October 2013 was imported. Then, the histograms per band were checked and a principal component analysis (PCA) was done. The PCA was performed to select the bands of the image for which the data was not correlated.

For the unsupervised classification, the Iso-cluster unsupervised classification tool of ArcGIS 10.1 was used. The tool calculated different clusters of neighbouring pixels which contained similar reflection values, produced a signature file and provided the classified image as output by means of the maximum likelihood classification function. The signature file contained a statistical description of the classes that were derived from the samples that were identified on the input raster. It consisted of two different parts: general information about all classes, such as input raster names, number of layers and number of classes and a part about signature statistics for each class: number of samples and the means and covariance matrices. Both the class mean vectors and covariance matrices were used by the Maximum Likelihood Classification tool. (Esri 2013b; Panda et al. 2009)

After the classification was done, the classification result needed to be smoothed by means of post-classification processing. This is necessary since the classification result can be very speckled containing a lot of small areas of different classes. This speckling makes the interpretation of a classification confusing and unclear. In order to make the classification more smooth, the majority filter, boundary clean and region group tool were used. The majority filter tool "*replaces cells in a raster based on the majority of their contiguous neighbouring cells*" (Esri 2013c; Fuller and Brown 1996). The boundary clean tool "*smooths the boundary between zones by expanding and shrinking it"* and the region group tool was used *"to remove small regions together with the set null tool which removed areas which are smaller than 10 pixels*" (Droppová 2011; Esri 2013d; Esri 2013e; Zhang et al. 2007). In the end, the nibble tool was used to dissolve the small areas. This tool "*replaces cells of a raster corresponding to a mask with the values of the nearest neighbours*" (Droppová 2011; Esri 2013f). (Figure 9a & Appendix D)

After the unsupervised classification, the supervised classification was carried out for Rusinga Island as a whole and the two study areas. This classification was based on the data that was gathered in October 2013. The first step of the supervised classification is not displayed in one of the models, but was performed via the spatial analyst image classification toolbar in ArcMap 10.1. The gathered samples were used to create training areas. A training area is an area with known properties, in this case which land use type occurs within that training area. The statistics within the training area are used to determine the decision boundaries for the classification of the whole study area. To create the training areas, the training sample drawing tool "polygon" was used from the image classification toolbar to create a polygon around the sampling point to select pixels of the same class containing the same reflection values which were defined as training samples. (Erdey-Heydorn 2008; Esri 2013g)

For the evaluation of the training samples, the histograms window, the scatterplot window and the statistics window were used to check the spectral characteristics of the training samples by means of the image classification toolbar (Dewan and Yamaguchi 2009; Esri 2013h). This means that was checked whether there is no overlap in reflection values between the different training samples. During the creation and evaluation of the samples, the training sample manager was used for saving, merging, deleting and clearing of the different classes (Dewan and Yamaguchi 2009; Esri 2013i).

The decisions of merging particular training samples were based on whether the histogram/scatterplot points overlap. If overlapping occurred, the training samples were merged into one class (Appendix D; Figure 42). (Esri 2013j; Erdey-Heydorn 2008). However, merging a particular class was also done if the results of the dendrogram showed that certain classes still had too much overlap. The dendrogram tool "*constructs a tree diagram (dendrogram) showing attribute distances between sequentially merged classes in a signature file*" (Erdey-Heydorn 2008; Esri 2013k). Those distances are also called between-class distances. In some cases the merging of the classes was adapted if the output of the dendrogram tool showed that the between-class distance was smaller than 0.9. This value was used, since any higher value results in only two defined classes during the classification and more classes were needed for the regression analysis.

If the results were acceptable, the maximum likelihood classification was performed. This classification method "*assumes that the cells in each sample are being normally distributed in the multidimensional space and is based on Bayes' theorem of decision making. It considers both covariance and variance of the class signatures when assigning each cell to one of the classes represented in the signature file. With the assumption that the distribution of a class sample was normal, a class was characterized by the mean vector and the covariance matrix. Given these two characteristics for each cell value, the statistical probability was computed for each class to determine the membership of the cells to the class*." (Esri 2013l; Pawlak 2002). In order to make the classification more smooth as was described above for the unsupervised classification, the majority filter, boundary clean and region group tool were used. The process of supervised classification is displayed in Figure 10. (Appendix D)

Those supervised classifications were used to create a map with the sampling locations based on the initial study areas and the order of those locations for the improvement of the supervised classification and the validation. Important is that all land use types are covered more or less equal in number of occurrences. Therefore, systematic random sampling was chosen, since this method made sure the sample plots were distributed evenly to all parts of the target area. It was spatially well-balanced. The difference with simple random sampling is that some land use types in that case may have many plots, while other land use types are not covered at all. Since there was a lot of variation in terms of land use on Rusinga Island, the chance that a particular land use type was not sampled with simple random sampling was too high (UPCC 2013). Systematic random sampling means that a random point was generated by means of a model (model b12; Figure 9a). In order to designate all the sampling points in the area, the points were placed from the first point with an equal distance between them. The outcome was a spatially well-balanced sampling points map that was used during the fieldwork in Kenya.

Li and Heap (2011) proved in their study that the sampling density had little impact on the performance of methods even over different spatial scales. Therefore, they think there is still a relationship between the number of sampling points and the performance of methods. They state that small scale phenomena, such as the land use patterns on Rusinga Island, need a high sampling density due to the relatively small study area. Large scale phenomena have a lower sampling density due to the larger study area. Thus, the size of the study area plays an important role in determining the sampling density. Since Rusinga Island and the two study areas defined for doing the sampling were relatively small (Chapter 2.1), a high sampling density was necessary. Kotrlik and Higgins (2001) noted that the type of data also determines the amount of sampling points.

In this case, categorical data (land use patterns) were studied for which the marginal error = 5%, α = 0.05 and a t-value = 1.96 was allowed. In order to determine the amount of variation in the area, a pilot study was performed on agricultural land use on Rusinga Island in October 2013. This data was used for a first classification via which the amount of variation in the area was determined. Both study areas were small (Study area 1 ≈ 5 km^2 and Study area 2 ≈ 2.5 km^2) and the amount of pixels per study area was large due to the high resolution of the Quickbird imagery. Furthermore, there was a lot of variation existing according to the first land use classification. Therefore, more than the maximum amount of 370 sampling points were needed to reflect the variation on the island properly according to the method of Kotrlik and Higgins (2001).

However, due to time limits it was not even possible to reach 370 sampling points per study area. It was estimated that a total amount of 50 points could be finished each week equating to 350 points over a seven week period. Therefore, based on the size and the time limit of two months sampling in Kenya the amount of sampling points and the distance between them was determined. For study area 1 the distance between the points was 150 m. which gave in total 210 sampling points and for study area 2 this was 75 m. which gave in total 270 sampling points. Since study area 1 was larger than study area 2, there was chosen to sample at a larger distance between the points in order to be able to cover the whole area within the time limits of the fieldwork. The points were placed by means of the drawing tools in ArcGIS and the points were exported as a layer file. Those layers were transformed into KML files to be able to upload them on the tablet which was used for the fieldwork.

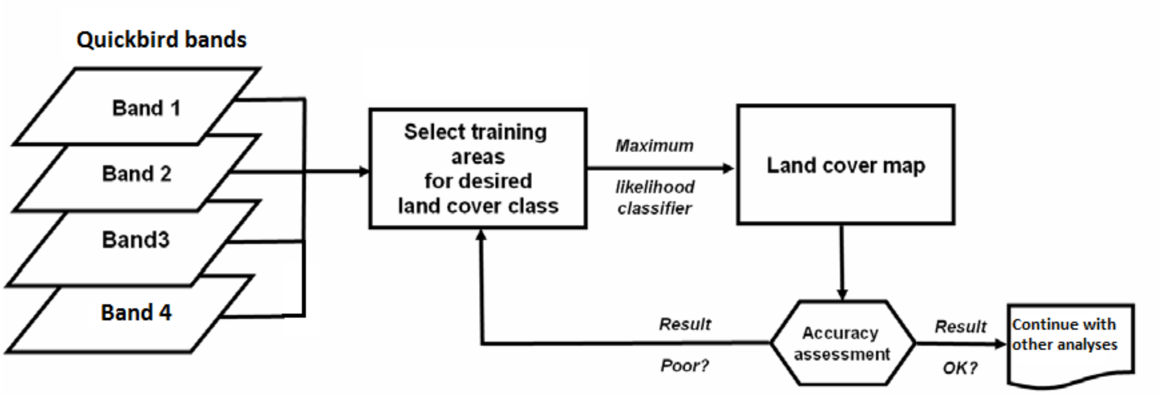


Figure 10: Scheme of supervised classification (Research question 1 and validation (Research question 2).

2.3.1.4 Fieldwork and Analysis: Improving the supervised classification

Since the first classification that was made for both March and October was only based on agricultural field data containing different crop types, the accuracy was not high. Furthermore, the amount of sampling points (24) per study area was too low to have a good enough accuracy. Therefore, the fieldwork on Rusinga Island was performed during February and March 2014 for the two study areas (Figure 9b & c). Per sampling point, the land use type for March and for October was recorded. In order to determine which part of the collected data was used for the improvement of the classification and which part of the dataset was used for the validation of the classification, random numbers were generated in the column behind the data. The data was sorted on basis of those random numbers (from low to high) and the first half was taken for the improvement of the classification. The other half was used for the validation of the dataset. Those datasets can be found on the DVD:

- *Fieldworkdata_Rusinga_Validation*
- *Fieldworkdata_SA1_validation2*
- *Fieldworkdata_SA2_validation2*
- *Fieldworkdata_SA1_classification*
- *Fieldworkdata_SA2_classification*

(For the classficiation of Rusinga island, both datasets of SA1 and SA2 are used)

Since the data exploration of the two study areas and of Rusinga as a whole was already done, a new data exploration of the same images used for the analyses was not necessary. Based on the imported data, more training samples per study area were created by means of the image classification toolbar. The histogram/scatterplot function on this toolbar helped to decide which classes were merged/were too different from each other to be merged before the dendrogram tool was used on the created signature files. Based on the outcome of this tool there was determined whether certain classes were merged. This was depending on the type of research and the amount of classes that should remain to have a valid study (Esri 2013m; Johnston et al. 2001).

For the classification done after the fieldwork, the threshold for merging was a between-class-distance of 0.9 in order to keep enough classes with a “large” enough between-class-distance. At least three classes should remain per study area in order to be able to continue with the data for the next analyses. (Chapter 2.3.3; Appendix D) This was done for the two new study areas and eventually the classification training samples were merged by using the training sample manager toolbar and the classification was also done for the whole of Rusinga Island.

2.3.2 Research question 2: Validation

Research question 2 focussed both on gathering groundtruth data in the field in February and March 2014 with which the supervised classification image was improved and on the validation.

The gathering of the groundtruth data and the improvement of the classification were already described in Chapter 2.3.1.4. After the improvement of the classification, the classification image was validated by means of the part of the data collected in February and March 2014 specified as validation data (Figure 9b; Appendix D). The points defined as validation points were adapted according to the merged classes in the classification and given the same "names" for each land use type. The data was interpolated by means of IDW (inverse distance weighted) and transformed from vector into raster.

IDW was chosen as interpolation method, since "*it estimates cell values by averaging the values of sample data points in the neighbourhood of each processing cell. If the point is closer to the centre of a cell being estimated, the more influence it has in the averaging process*." (Esri 2013n; Johnston et al. 2001) This method was important, because there were not a lot of mosquito data and the data were not well spread over the area: only near households, since sampling was done in- and outside the houses. Therefore, it was necessary that the points closer to the centre of the cell that is being estimated had more influence on the estimation of the value of that cell. Afterwards the error matrices were created (Appendix F).

By means of the created error matrices, the accuracy of the classifications were studied. The values of total accuracy were given in the range between 0 and 1 for which 0 means the accuracy does not deviate from a random classification and 1 means the classification is perfect. Also the kappa coefficient was calculated, which is a statistical measure of inter-raster agreement or inter-annotator agreement for qualitative (categorical) items. 0 means a low agreement, 1 a high agreement. The user and producer accuracy were calculated as well and given in percentage. If the accuracy was low, this meant the classification of that class was less good compared to other classes and compared to the validation dataset. The user's accuracy refers to the probability (%) that a pixel that is given a certain land use class is really that class. The producer's accuracy refers to the probability (%) that the land cover of a certain area on the ground is classified as such. Eventually the land use classification was adapted again in order to improve this accuracy if necessary. When this was finished, there was continued with the regression analyses. (GIS & Map Library 2013)

2.3.3 Research question 3: Relational analysis

In order to be able to answer the third research question, the vector abundance data was projected in ArcGIS as xy-data points. Furthermore, the land use classification maps were needed to find out the relationship between vector abundance and the agricultural land use patterns. To study this relationship, both datasets were merged by creating a raster dataset of the vector abundance dataset and extracting the values to the classification dataset. Then, the regression process started. Two methods of regression were used in ArcGIS. The first one is Ordinary Least Squares regression (OLS). This method was used in the first place, because it provides a global model for the variable that is tried to be understood or predicted. Only R^2 is used for analyses of the results. The coefficient of determination (R^2) shows how well the data points fit the developed statistical method. The values range from 0 to 1. A value of 0 means that the data does not fit the statistical model, a value of 1 means that there is an exact match between the data and the statistical model.

However, only one single regression equation is created and every land use type was separately investigated. Therefore, also Geographically weighted regression (GWR) was used. This method provides a local model by fitting a regression equation to every feature in the dataset. In this way, the relationship between the entomology data and each agricultural land use type can be investigated in total, including the non-stationarity of the regression coefficients over space. (Esri 2013o; Brunsdon 1998) The regression analyses was performed for every study area and for the whole of Rusinga Island investigating three different relationships: one containing all malaria vectors (both *Anopheles gambiae* females and *Anopheles funestus* females), one for *Anopheles gambiae* females only and one for *Anopheles funestus* females only. (Sounny-Slitine 2012)

3. Results

3.1 Research question 1

3.1.1 Information collection on crop development

To perform a supervised land use classification as a pilot study to define the amount of sampling points for the fieldwork, first information was gathered about the different crop types grown on the island. This was done via Ibrahim Kiche and crop calendars from the FAO (Table 1).

After those first data was studied in more detail, data on crop use was collected in October 2013. (Appendix C) The Excel file containing the data gathered in October 2013 can be found on the DVD belonging to this study and is named: *Data_Tobias_15_18_October_2013*. Based on those data the land use classes were determined (Appendix E).

The data of October 2013 was compared with the crop calendars of the (FAO 2013a) (Table 1) and some discrepancies were found as is given in Table 3.

Table 3: Comparison between data (FAO 2013a) and field data about the occurrence of particular crop types in a particular season on Rusinga Island, Kenya.

	FAO (2010)		Field data	
	1st season	2nd season	1st season	2nd season
Beans	x	x	x	x
Cow peas		x	x	
Tomatoes		x	x	x
Sukumawiki (Kale)	x	x	x	x
Maize		x		x
Sorghum	x	x	x	
Sweet potato		x	x	
Sweet pepper	x		x	x

Only for beans, kale and maize both datasets agreed in terms of their growing season. Sorghum and sweet potato grew only in one season according to the data gathered and could be grown in both seasons. Since only 24 points were sampled per study area it is possible that those crops are also growing in the other season and therefore the information from the (FAO 2013a) is assumed to be true.

The data gathered about cow peas, tomatoes and sweet pepper indicate that the crops are growing in another season than indicated by the (FAO 2013a). Unless the fact that only 24 points per study area were gathered and some crop types were missed in a particular season, the real-time gathered data give a better indication of the current situation and therefore were assumed to be true for the mentioned crop types. So, cow peas, beans and sorghum are the crop types that grow within a particular season and the other crop types grow in both seasons. This means that in terms of seasonality only three crop types deviate from each other.

3.1.2 Unsupervised land use classification

The first classification that was done in this study was the unsupervised classification to get a global idea of the variation of the land use on Rusinga island. The result of this for the whole of Rusinga is displayed in Figure 11 for March and Figure 14 for October. The results of this for the two study areas are displayed in Figure 12 & 13 for March and in Figure 15 and 16 for October.

As can be seen from Figure 11, 12 and 13, the spatial distribution of the land use was very heterogeneous on Rusinga Island. Especially, this becomes clear on the unsupervised classification results of the two study areas (Figures 12 & 13). At least every 50 meters a different type of land use occurred according to this land use classification. From Figure 14 it appeared that the variation is considerable less. However, the resolution of the image was lower. This means if the variation was occurring within the 30 m resolution this was not displayed within the classification. According to Figure 12, the variation occurred every 100 m at most. What was seen in the different classification results was that the amount of classes is lower than the specified number of classes in the model both for Rusinga Island (Figure 11 and 14), study area 1 (Figure 15) and study area 2 (Figure 13 and 16). An explanation for this is again the lower resolution of the Landsat image of October and/or that the specified number of classes in the model is the maximum number of clusters that can result from the clustering process and has the following conditions:

- "*The values of data and the initial cluster means are not evenly distributed. In certain ranges of cell values, the frequency of occurrences for these clusters may be next to none. Consequently, some of the originally predefined cluster means may not have a chance to absorb enough cell members.*
- *Clusters consisting of fewer cells than the specified Minimum class size value will be eliminated at the end of the iterations.*
- *Clusters merge with neighbouring clusters when the statistical values are similar after the clusters become stable. Some clusters may be so close to each other and have such similar statistics that keeping them apart would be an unnecessary division of the data*." (Esri 2013p)

In terms of visual inspection of the Figures, the classification of Rusinga Island (Figure 11 and 14) did show the larger patterns of land use on the island. The hill in the centre was clearly visible, just like the patterns of bare soil/rocks and even some agricultural fields were distinguished along the shoreline on the northern side of the island. The classification of study area 1 (March, Figure 12) showed a high variability and the only clear pattern that was seen is that from agricultural fields on the northern part of the study area. Furthermore, some part of the road is slightly visible. The classification of the same area, but than for October shows less variation (Figure 15). Only two classes were determined although the same dataset was used. Here, the image resolution plays a role again. The classification of study area 2 (March, Figure 13) also showed a high variability and some agricultural fields along the shoreline in the southern part of the study area were distinguished. However, those fields were not as clear as in study area 1. Some roads were distinguished as well. Figure 16 shows that the classification of October for the same area with the same dataset again has a low amount of variation due to the low resolution of the Landsat image. As a result, only 1 class was determined during the unsupervised classification.

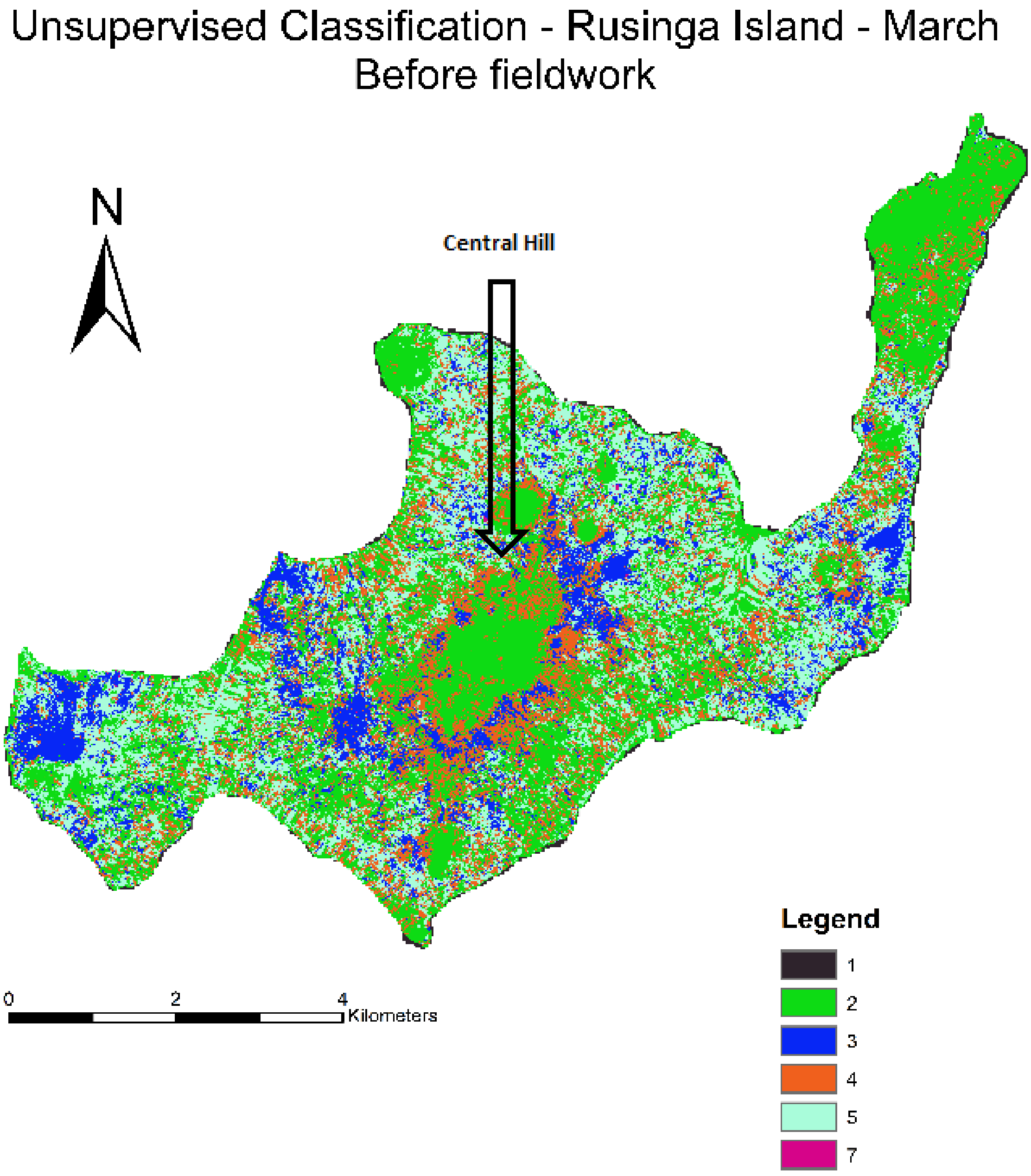


Figure 11: Unsupervised classification result of Rusinga Island – March, performed with 14 classes, only 7 are remaining (Since it is an unsupervised classification, the classes are not specified, instead each class has a number).

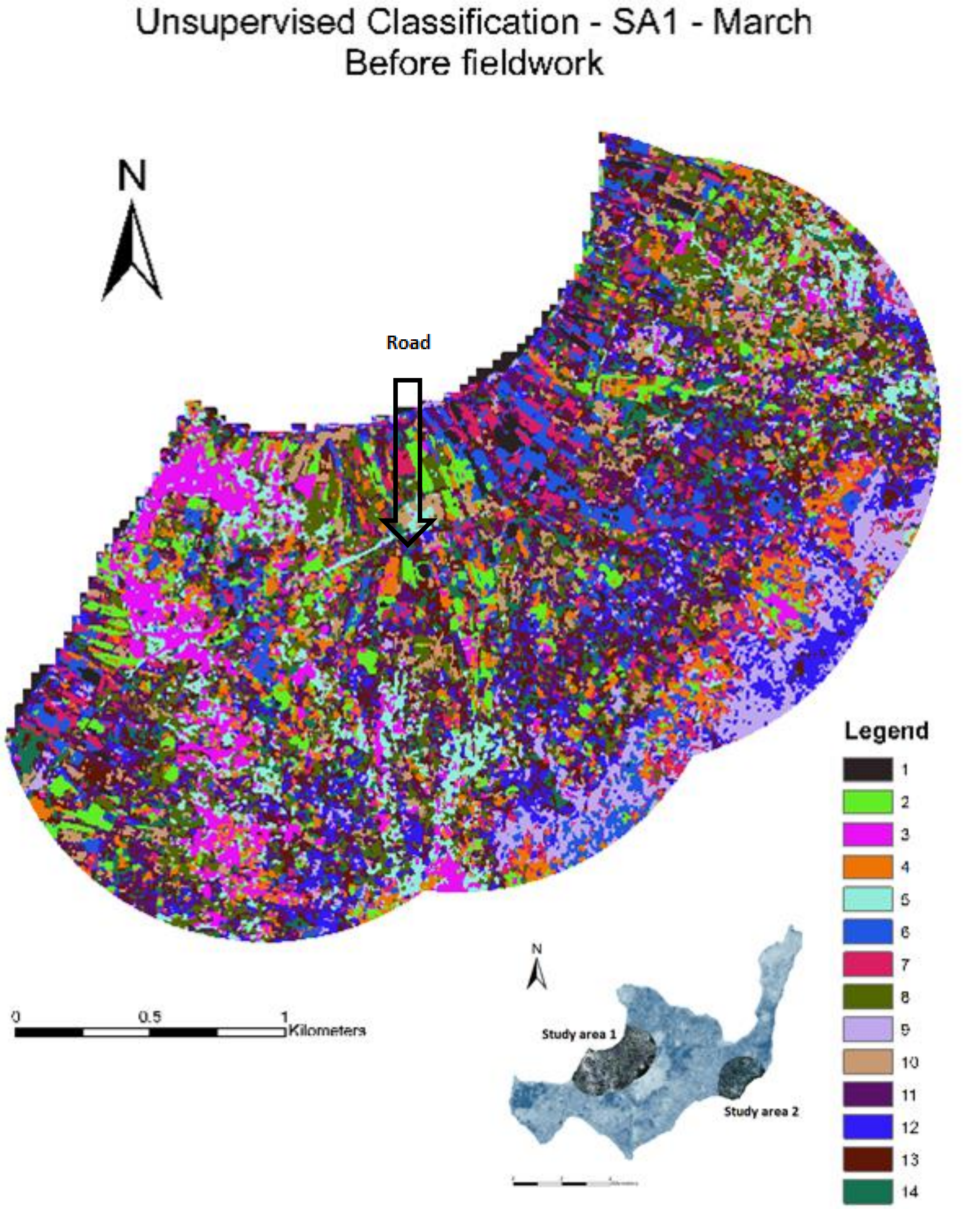


Figure 12: Unsupervised classification result study area 1, Rusinga Island, Kenya - March. (14 classes)

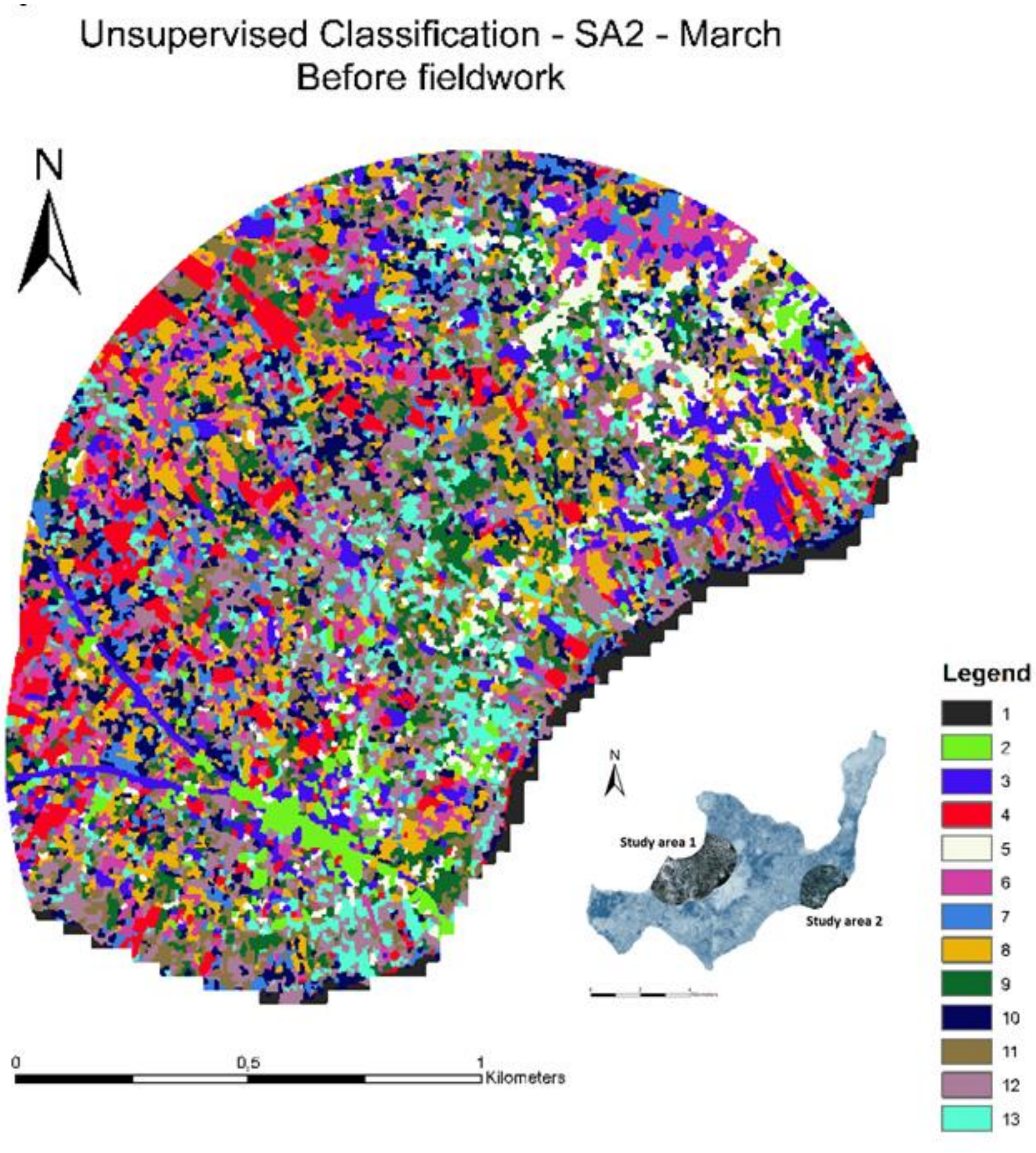


Figure 13: Unsupervised classification result study area 2, Rusinga Island, Kenya - March. (14 classes was the goal, only 13 remained after the analysis was done)

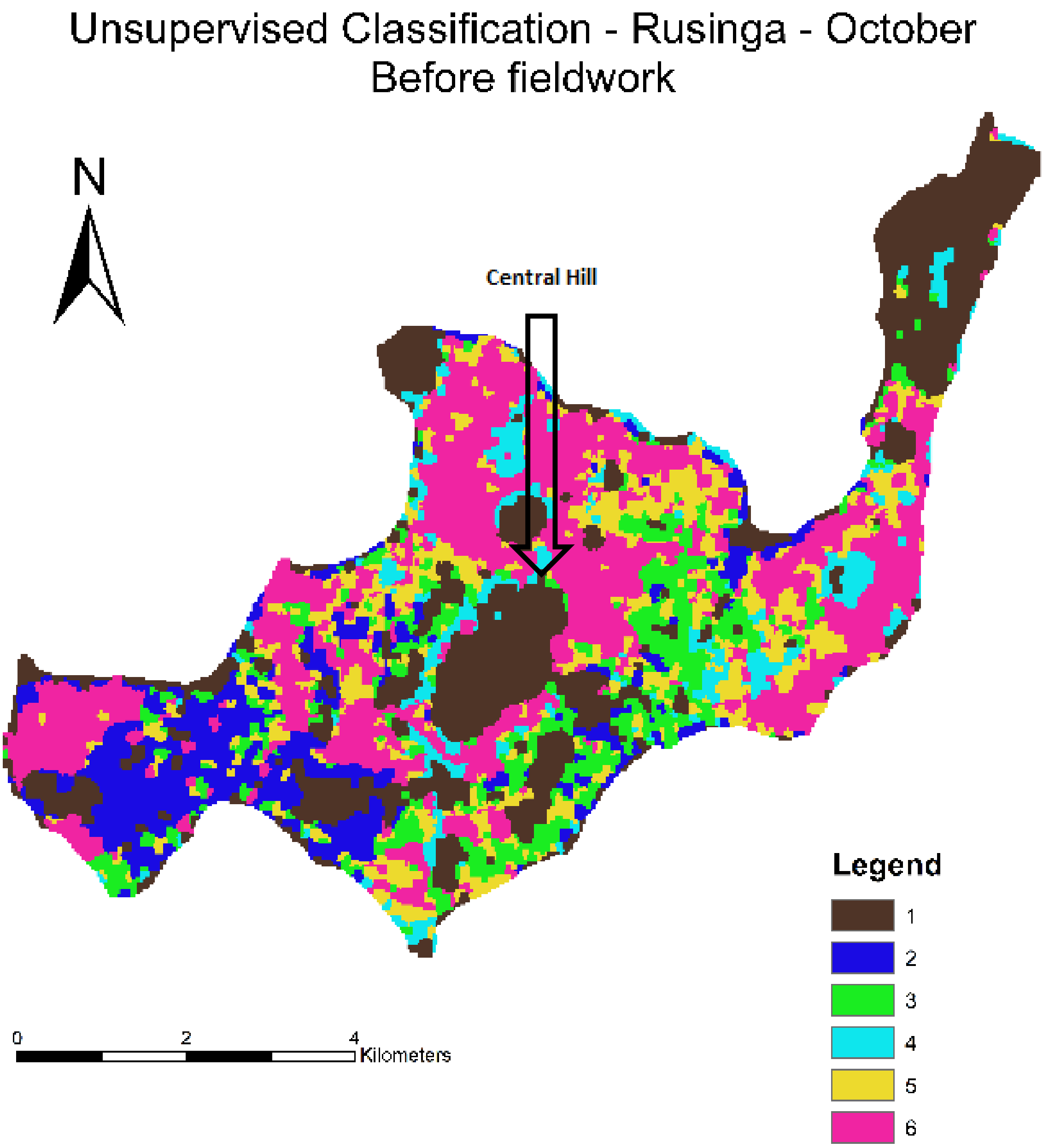


Figure 14: Unsupervised classification result of Rusinga Island – October, performed with 14 classes, only 6 are remaining (Since it is an unsupervised classification, the classes are not specified, instead each class has a number).

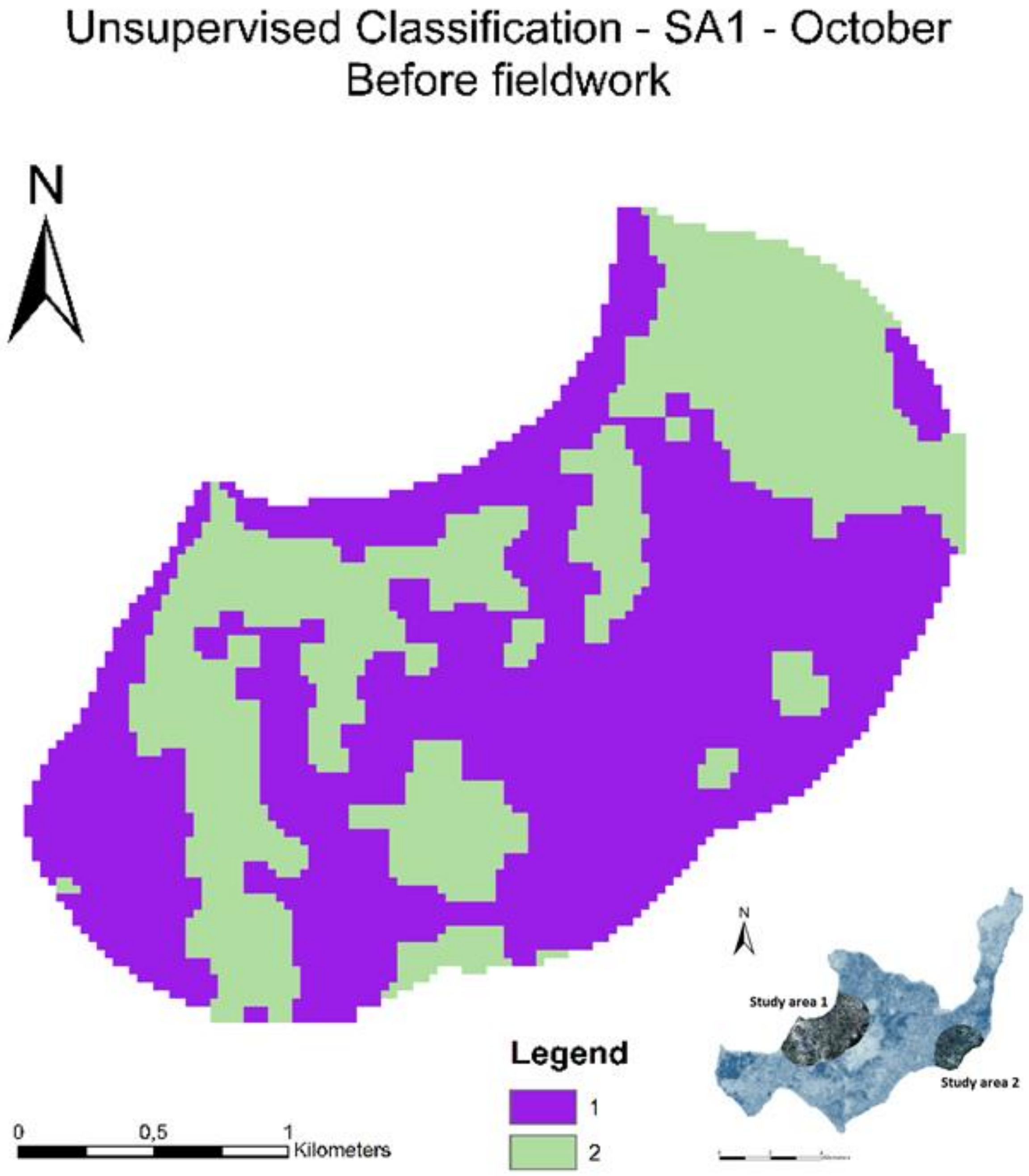


Figure 15: Unsupervised classification result study area 1, Rusinga Island, Kenya – October, performed with 14 classes, only 2 are remaining (Since it is an unsupervised classification, the classes are not specified, instead each class has a number)

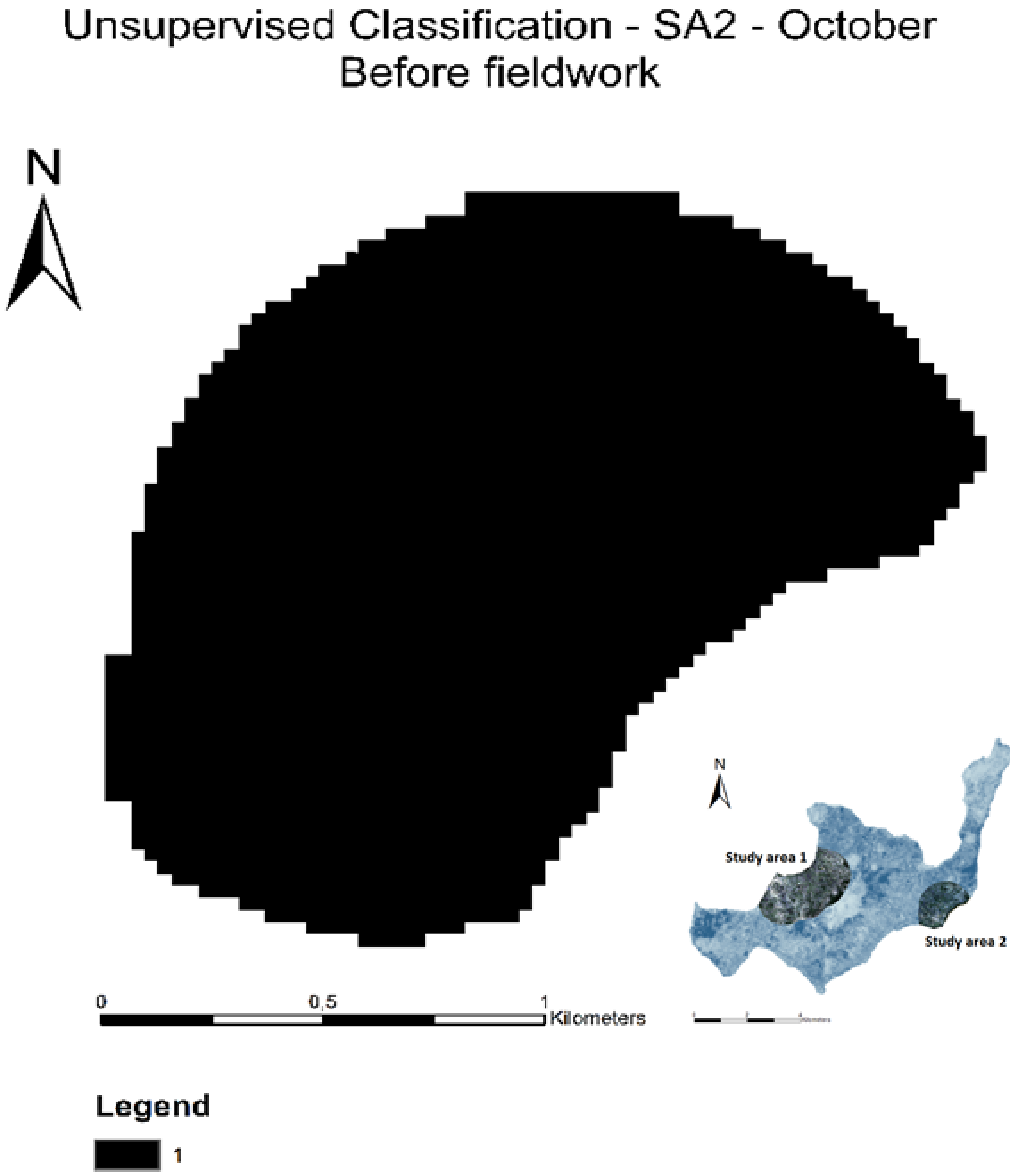


Figure 16: Unsupervised classification result study area 2, Rusinga Island, Kenya – October, performed with 14 classes, only 1 is remaining (Since it is an unsupervised classification, the classes are not specified, instead each class has a number)

From visual inspection of the images it became clear that for Rusinga Island (March, Figure 17) tomatoes and sorghum appear along the shoreline on the northern part of the island where the agricultural fields occurred during the unsupervised classification. However, for the image of October (Figure 20) only two classes were classified: beans and cow peas. Due to the low resolution of the image several classes felt in one pixel, which mixes the reflection values. For study area 1 the same is true for sorghum again (Figure 18). For the image of October (Figure 21), again less variation was displayed. The whole area was covered with cow peas, except for some small parts which were covered by beans or tomatoes. Most of the area was classified as cow peas, sukumawiki or sweet pepper. From the dendrogram, it also appeared that the between class distance was much smaller for this study area than for the other one and for Rusinga as a whole. The image of October for study area 2 (Figure 22) contained more variation. Although there were the same amount of classes, those classes were more spread over the area and the surface per class was larger. The result did not give any indication about a general pattern occurring as Figure 17 and Figure 18 did.

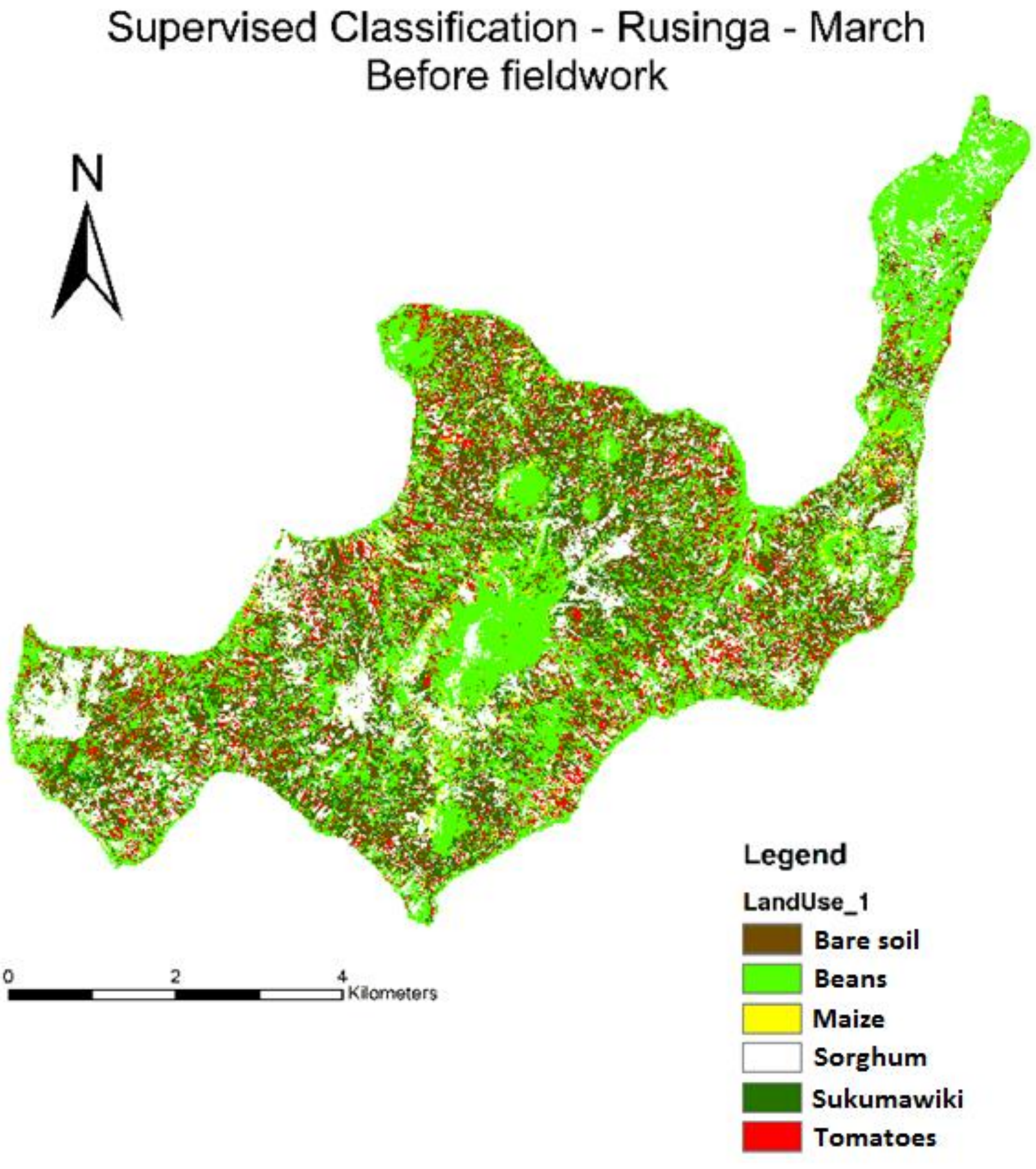


Figure 17: Supervised classification result of Rusinga Island – March.

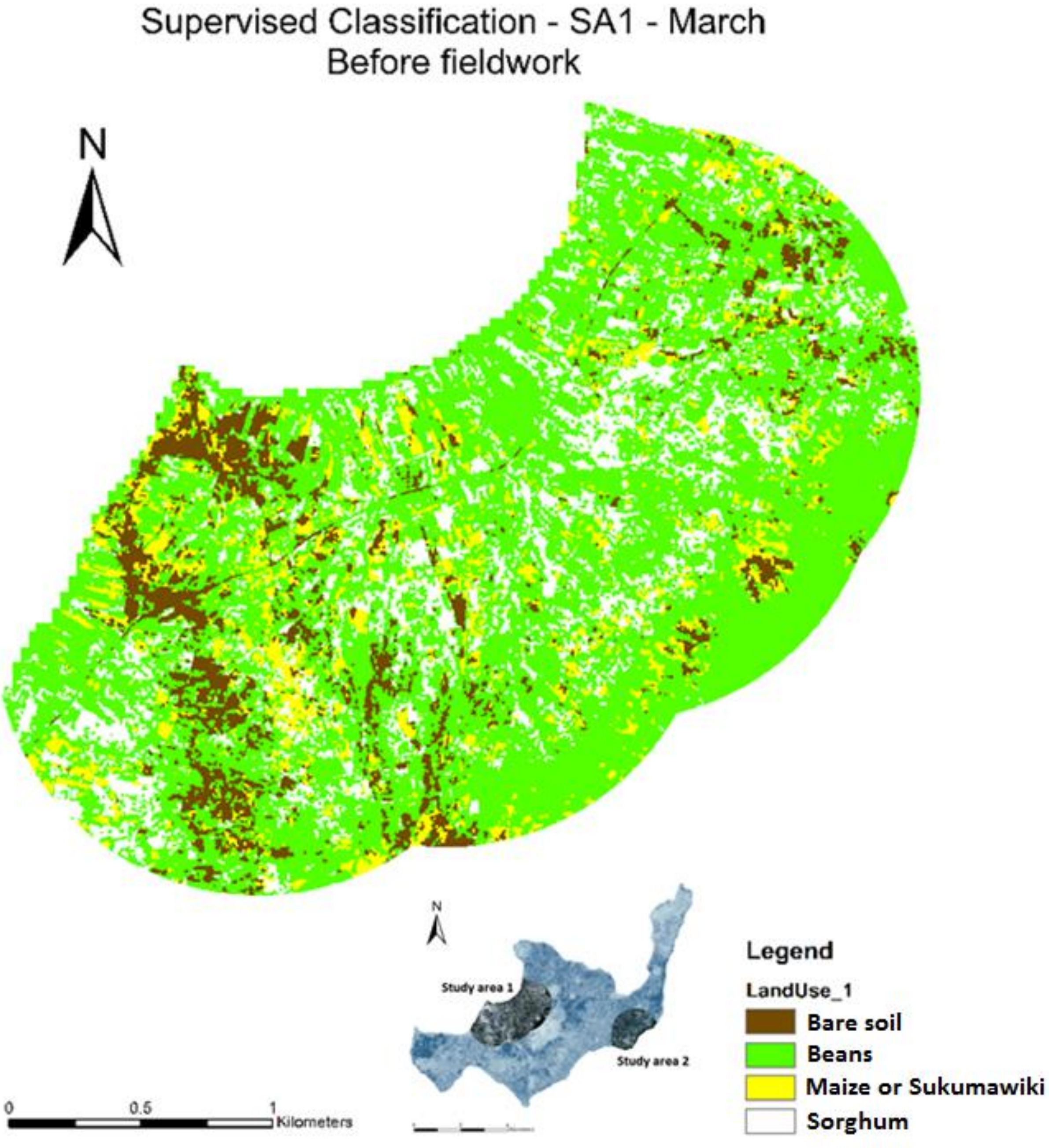


Figure 18: Supervised classification result study area 1, Rusinga Island, Kenya - March.

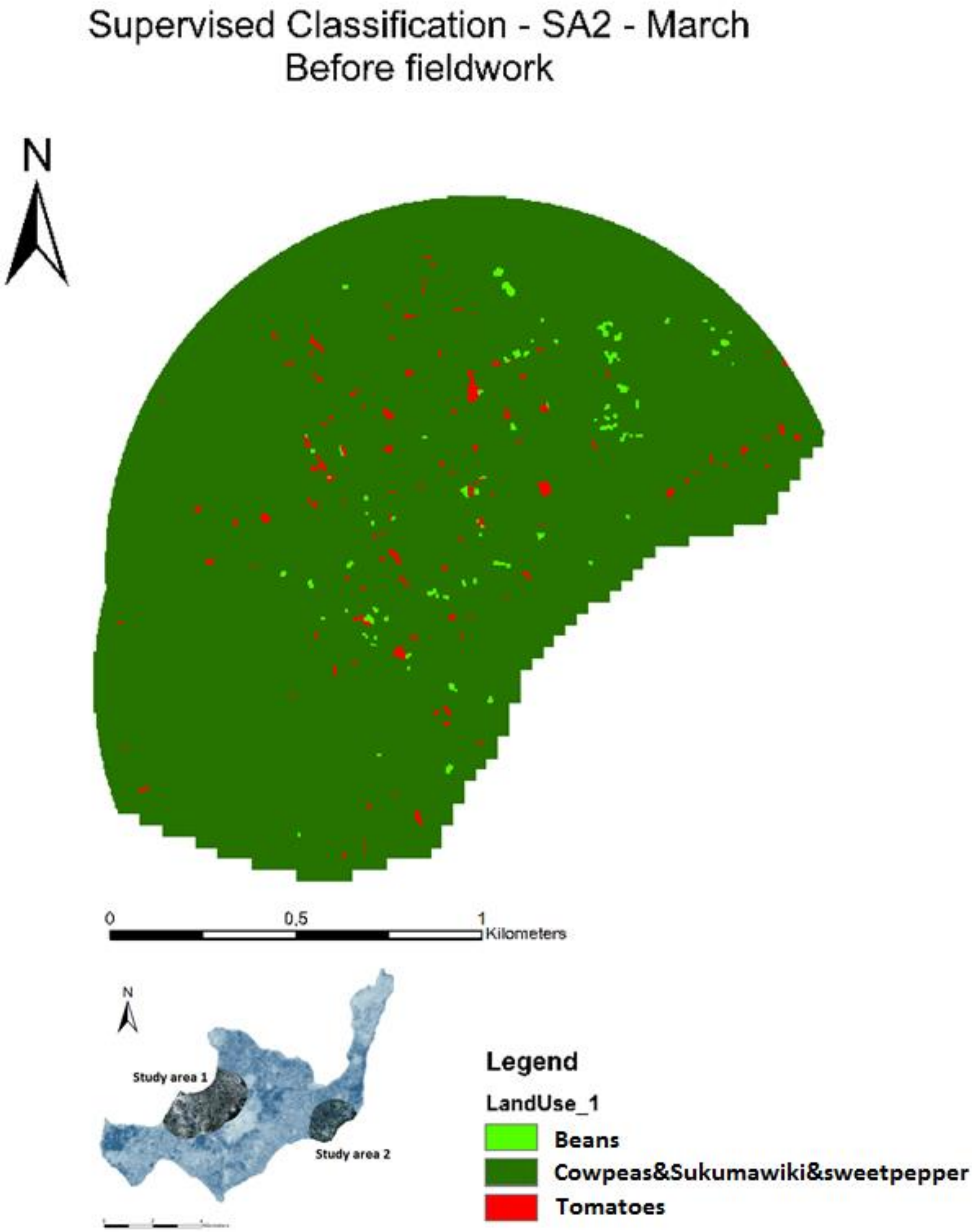


Figure 19: Supervised classification result study area 2, Rusinga Island, Kenya - March.

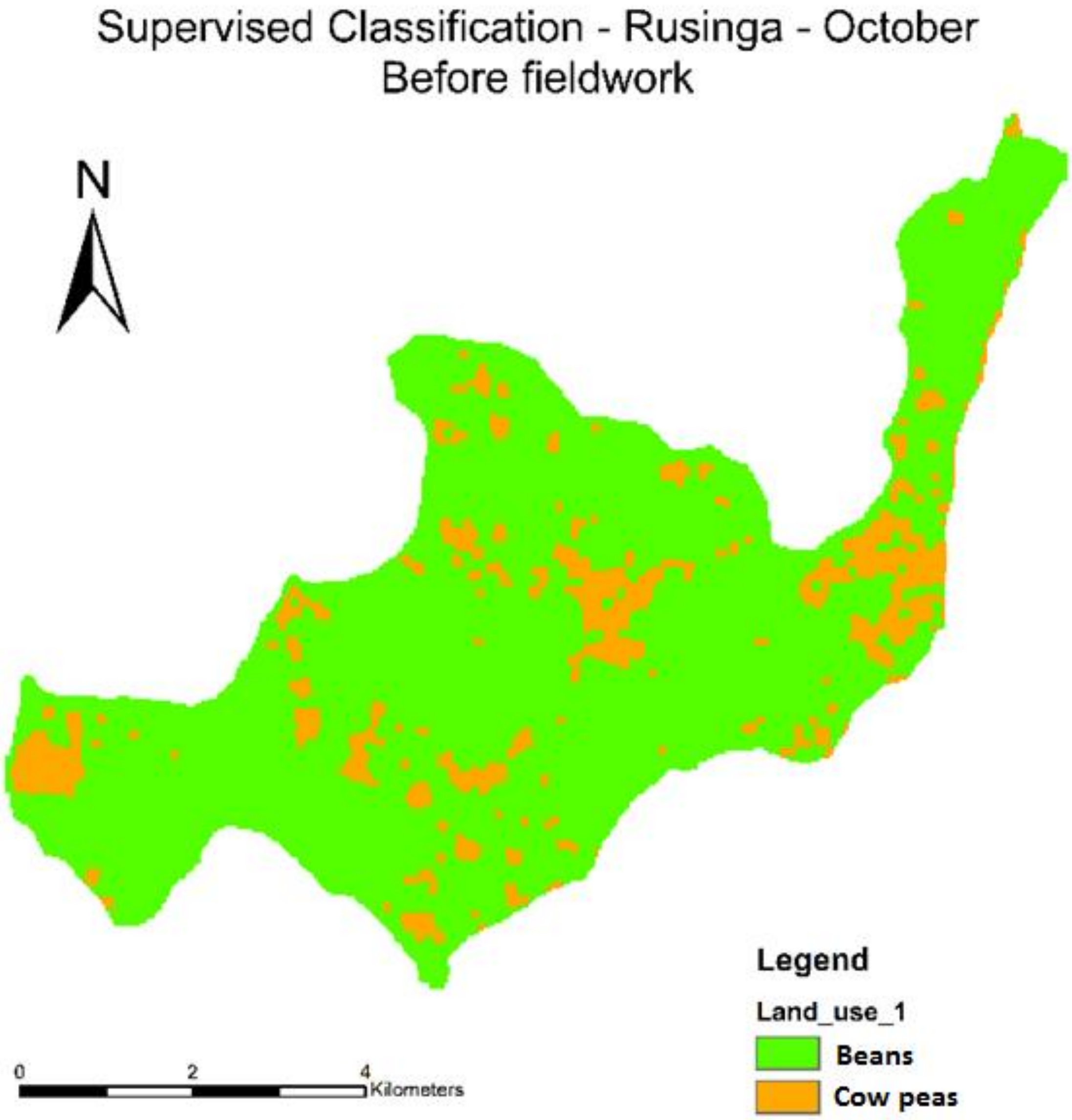


Figure 20: Supervised classification result of Rusinga Island – October.

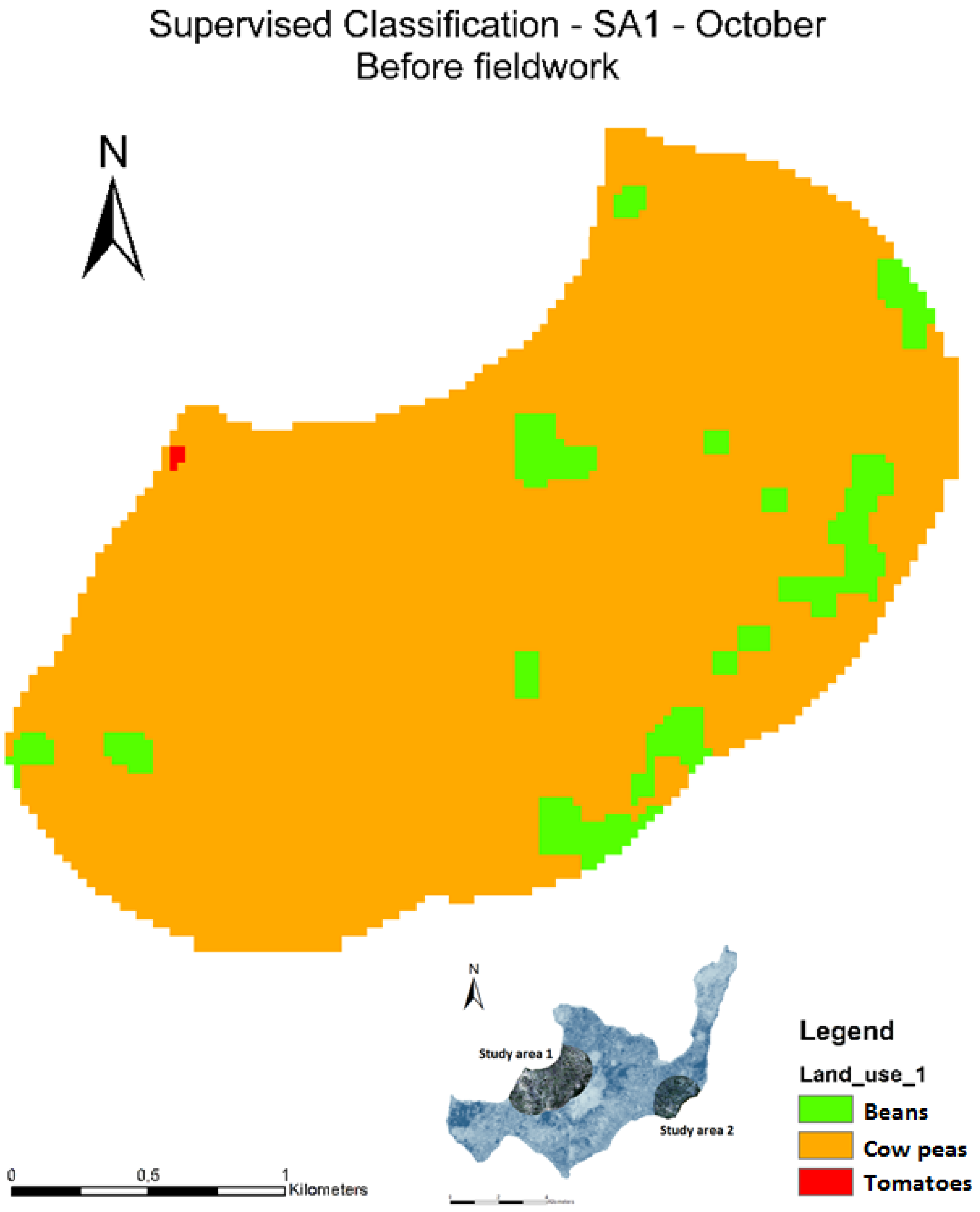


Figure 21: Supervised classification result study area 1, Rusinga Island, Kenya - October.

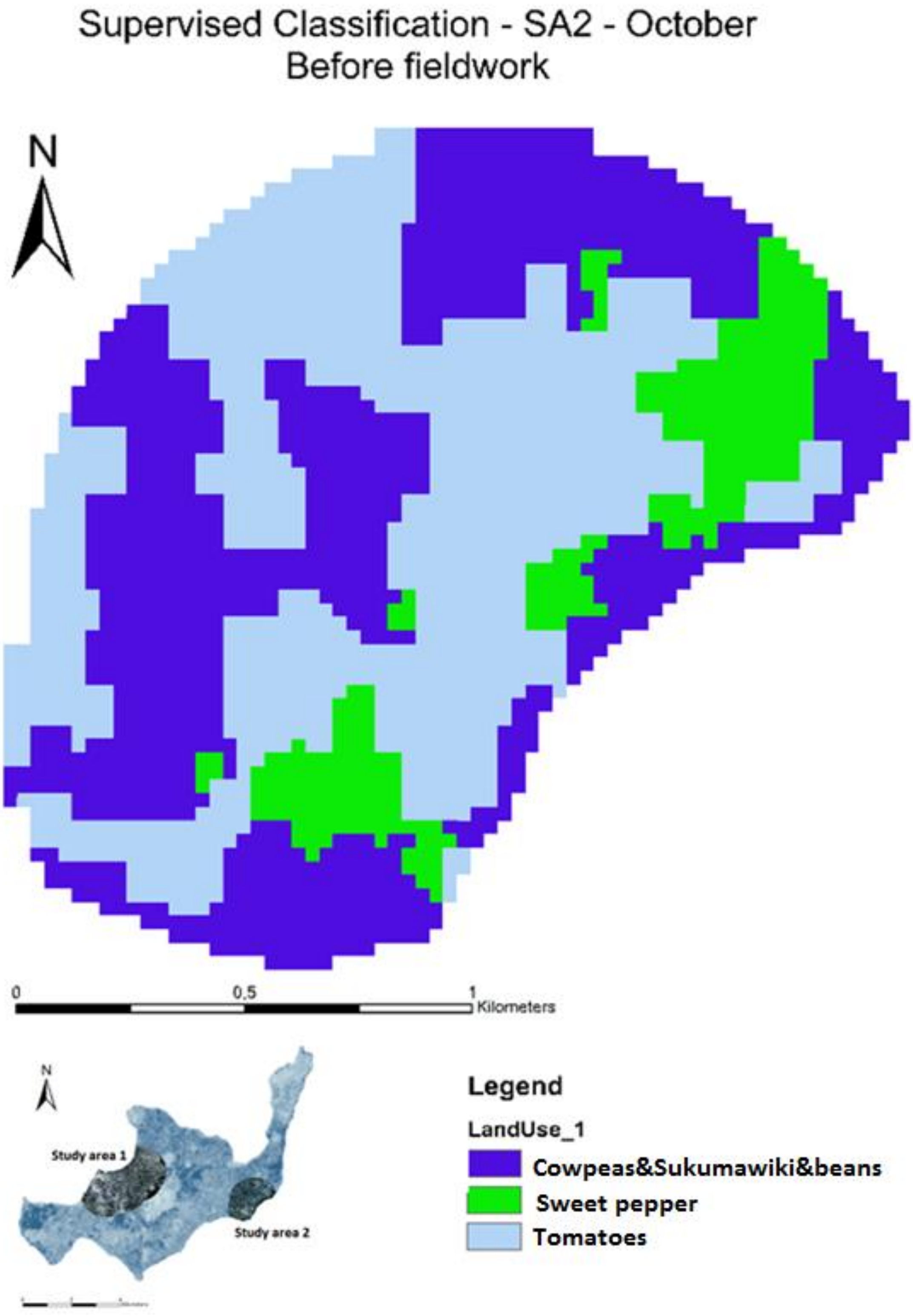


Figure 22: Supervised classification result study area 2, Rusinga Island, Kenya - October.

3.1.3 Preparation Fieldwork

The first supervised classifications were only based on the agricultural land use types which influences the accuracy and the completeness of the classification. For example in Figure 17, the central hill of Rusinga Island was covered with beans, while the main land use type there was shrubs and/or trees. To improve the land use classification, more classes were added and more agricultural fields were sampled. The fieldwork was done on Rusinga Island, as is described in Chapter 2.3.3 and 2.3.4. For the sampling, systematic random sampling was chosen as is explained in Chapter 2.3.3. The fieldwork maps that were produced can be found in Figure 23 and Figure 24.

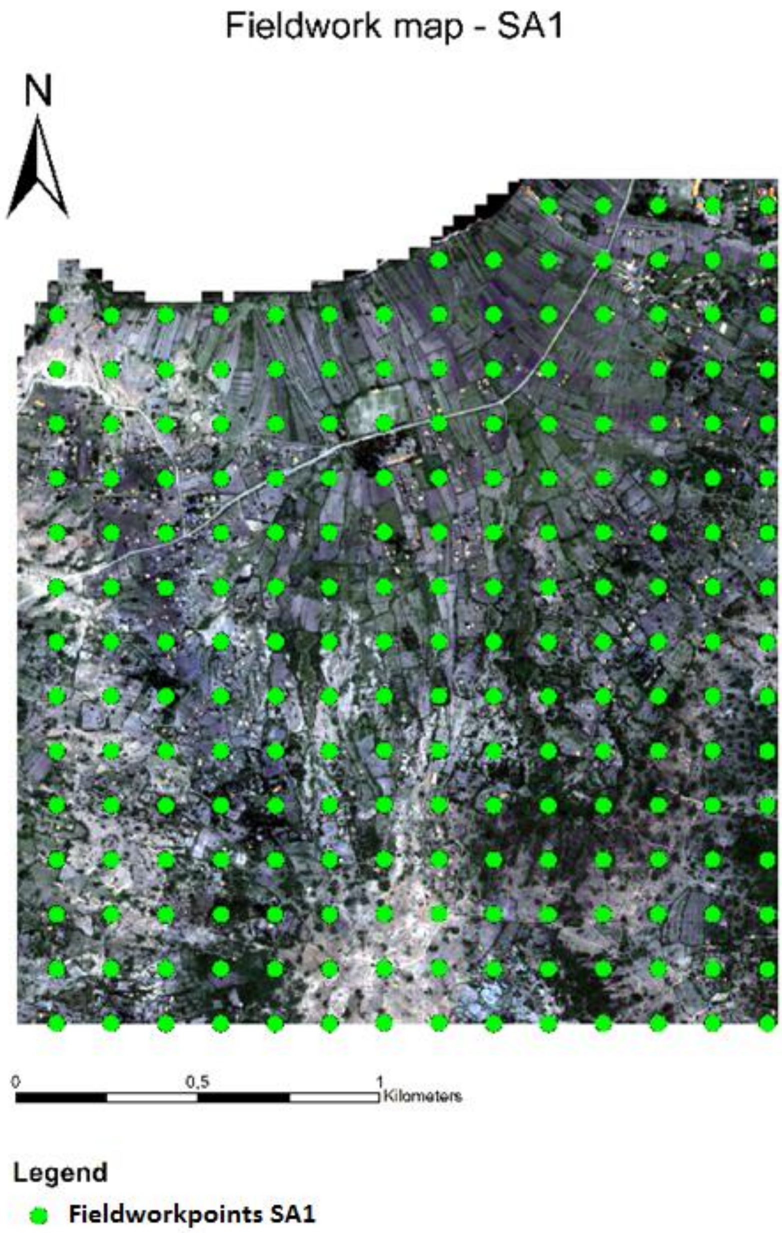


Figure 23: Fieldwork map study area 1, Rusinga Island, Kenya, containing the sampling locations.

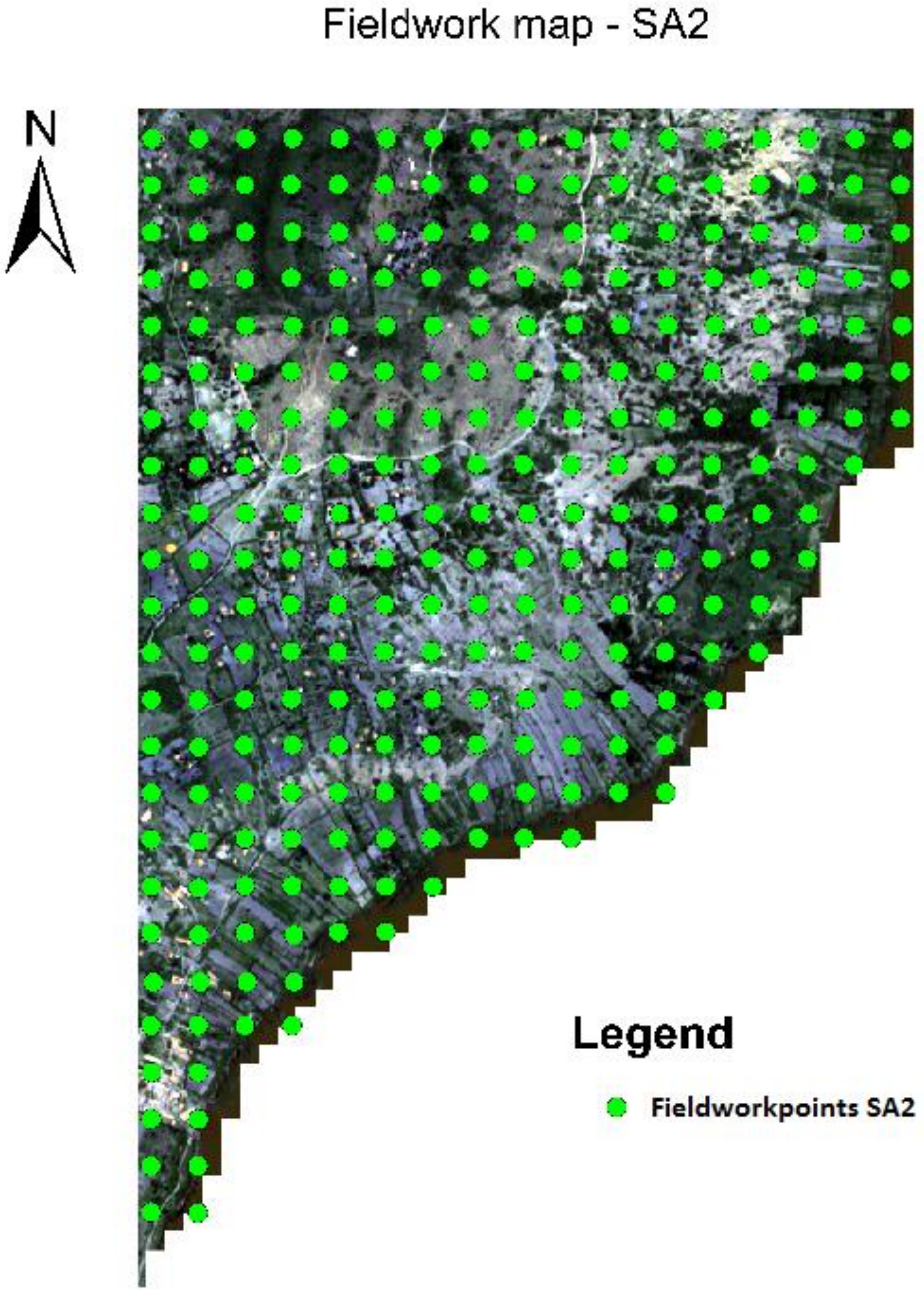


Figure 24: Fieldwork map study area 2, Rusinga Island, Kenya, containing the sampling locations.

3.1.4 Improving the supervised classification

The data gathered during the fieldwork were used for the improvement of the supervised classification. The result of the improved supervised classification for Rusinga Island can be found in Figure 25 and Figure 28, for study areas 1 & 2 in Figures 26 and 27 and Figures 29 and 30.

By means of visual inspection of the images it became clear that for Rusinga as a whole, the classification was accurate for the classes shrubs, water and rocks/bare soil. However, the classes building and agriculture were overrepresented in the classification. The classification of Rusinga Island in October was less good when comparing it to reality. The class containing bare soil, rocks as well as agriculture and pasture was underrepresented, especially the agricultural part along the lake shores was missing. The shrubs were overrepresented on the island. The class for trees seemed good. Trees were classified on top of the hill and along the lake shore. For study area 1 (Figure 26), rocks/bare soil was underrepresented due to classification of rocks/bare soil as agriculture and shrubs were overrepresented due to the same reason. The classification of study area 1 for October (Figure 29) had classified the class bare soil and agriculture and building/pasture quite well. However, the class containing rocks and shrubs/trees seemed to be underrepresented in this classification. Study area 2 (Figure 27) only had three classes in which agriculture also contained shrubs/trees and pasture when comparing it with the real-life situation. Therefore, agriculture was underrepresented, while building and rocks/bare soil were overrepresented. Figure 30 displays the classification of study area 2 for October. It seems that this classification is better in terms of global patterns within that area. However, there were more trees than there were in reality and less shrubs. The class agriculture was found along the lake shore. This was based only on knowledge on the area, however a validation dataset was created as well so there was looked at how accurate the land use classifications actually were.

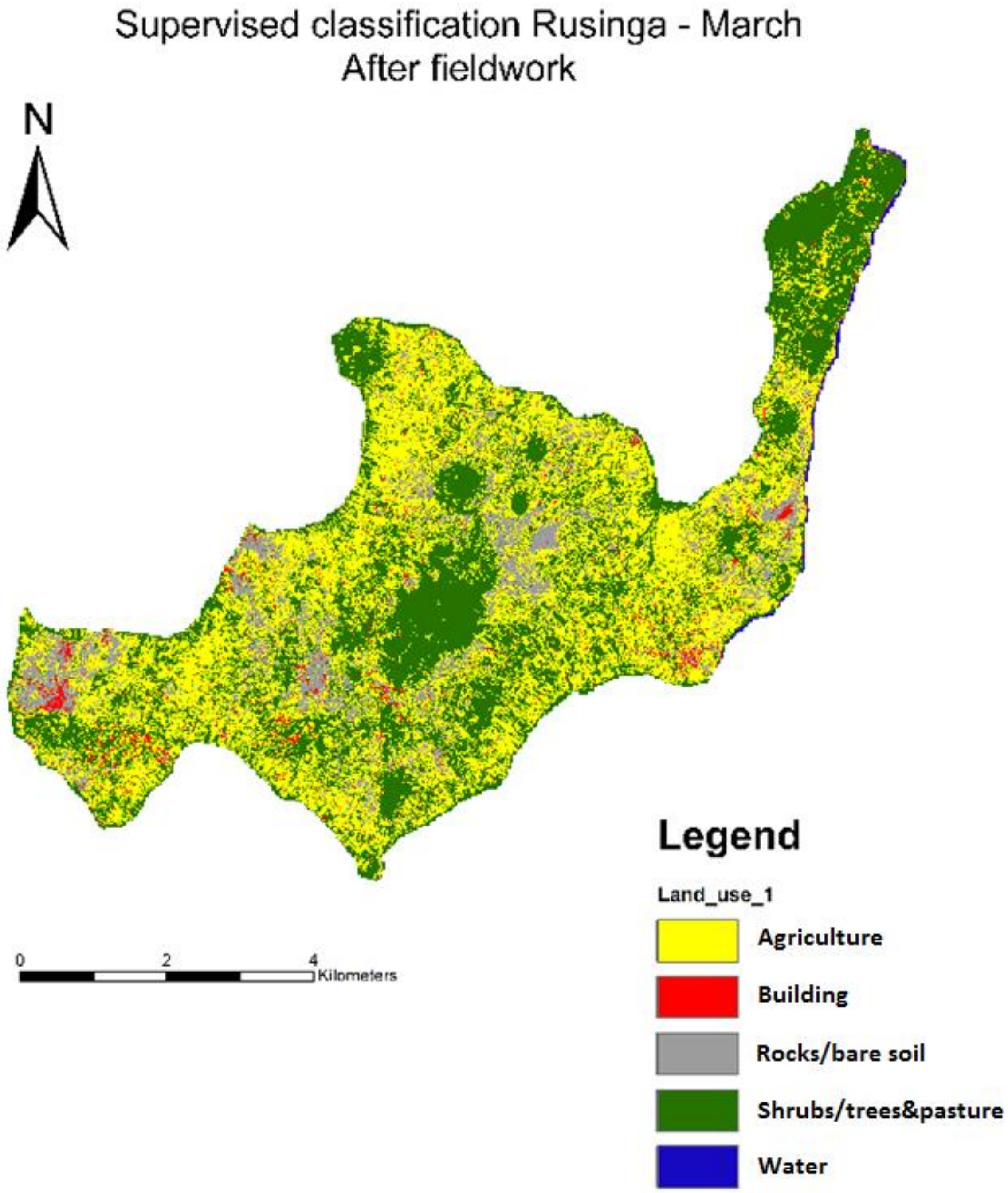


Figure 25: Supervised classification result of Rusinga Island – March – after fieldwork.

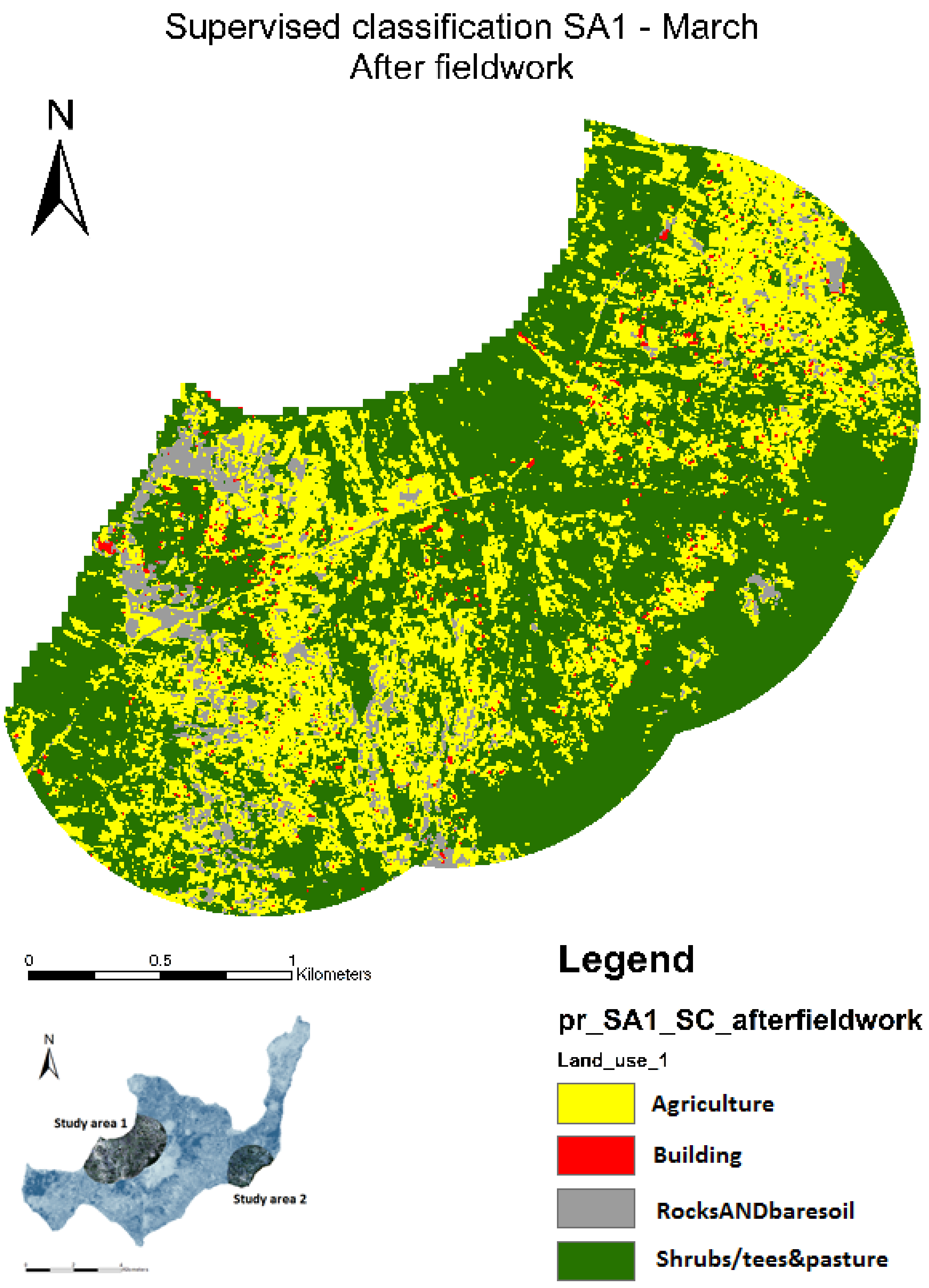


Figure 26: Supervised classification result of study area 1 (SA1) – March – after fieldwork.

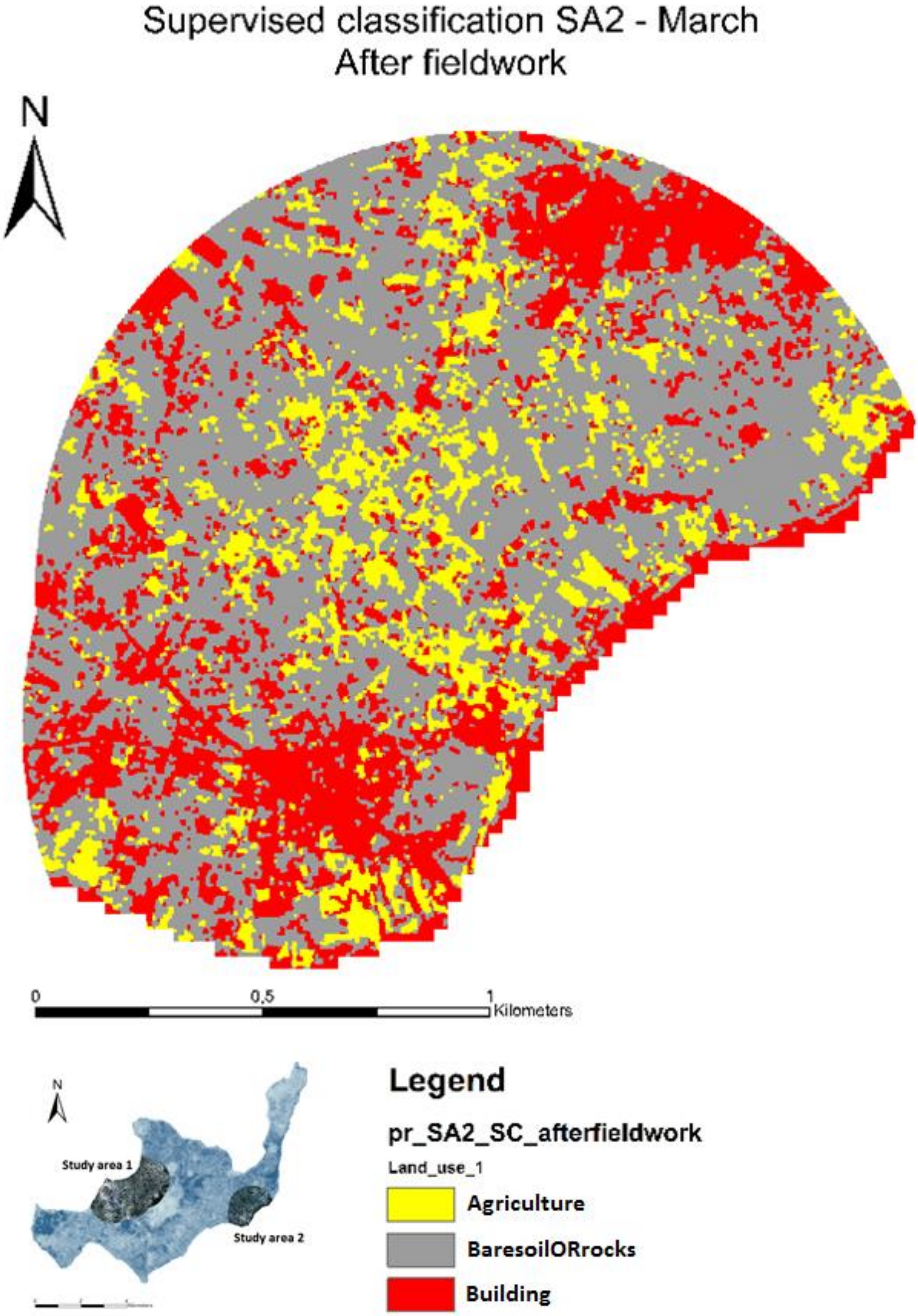


Figure 27: Supervised classification result of study area 2 (SA2) – March – after fieldwork.

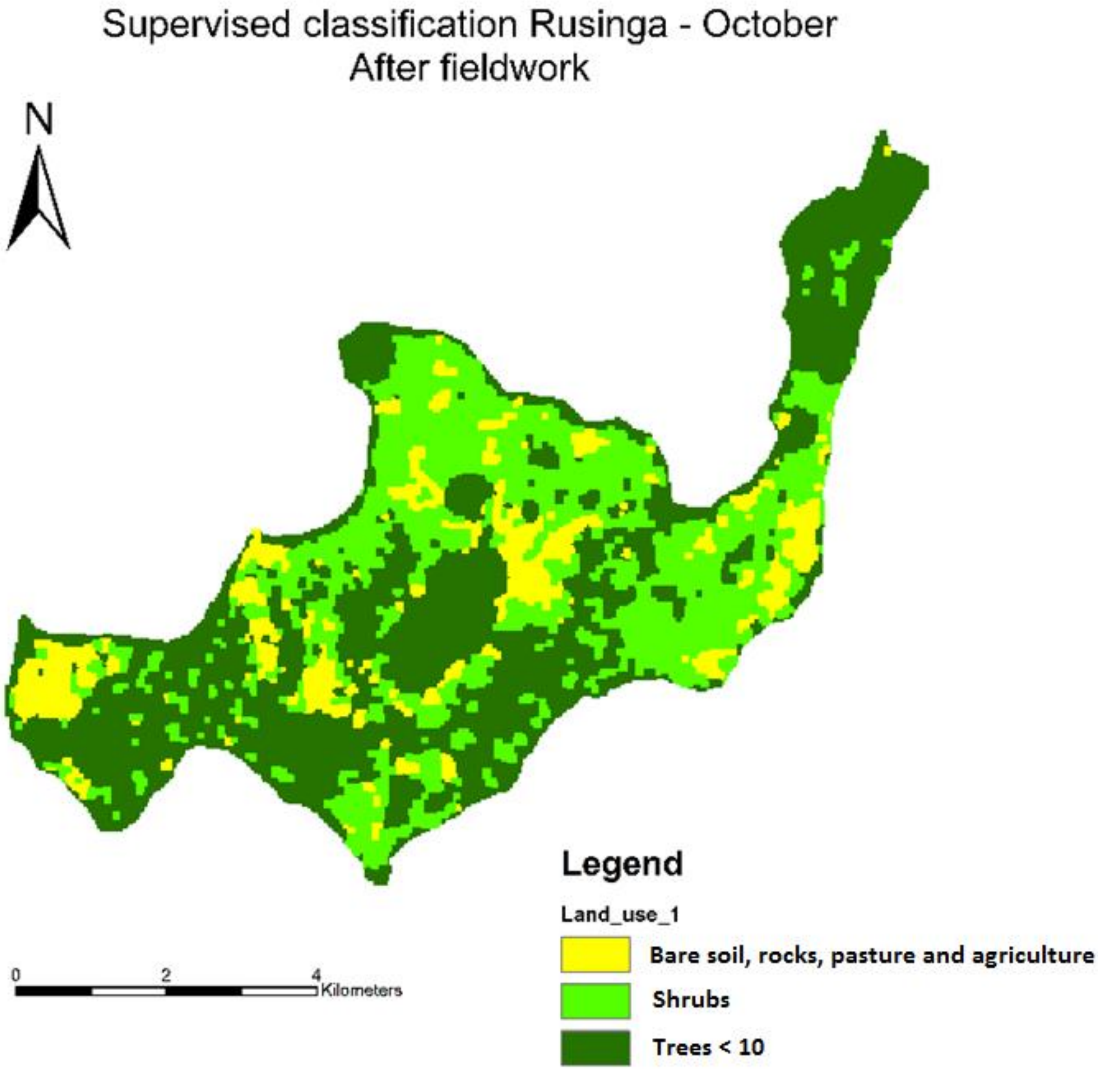


Figure 28: Supervised classification result of Rusinga Island – October – after fieldwork.

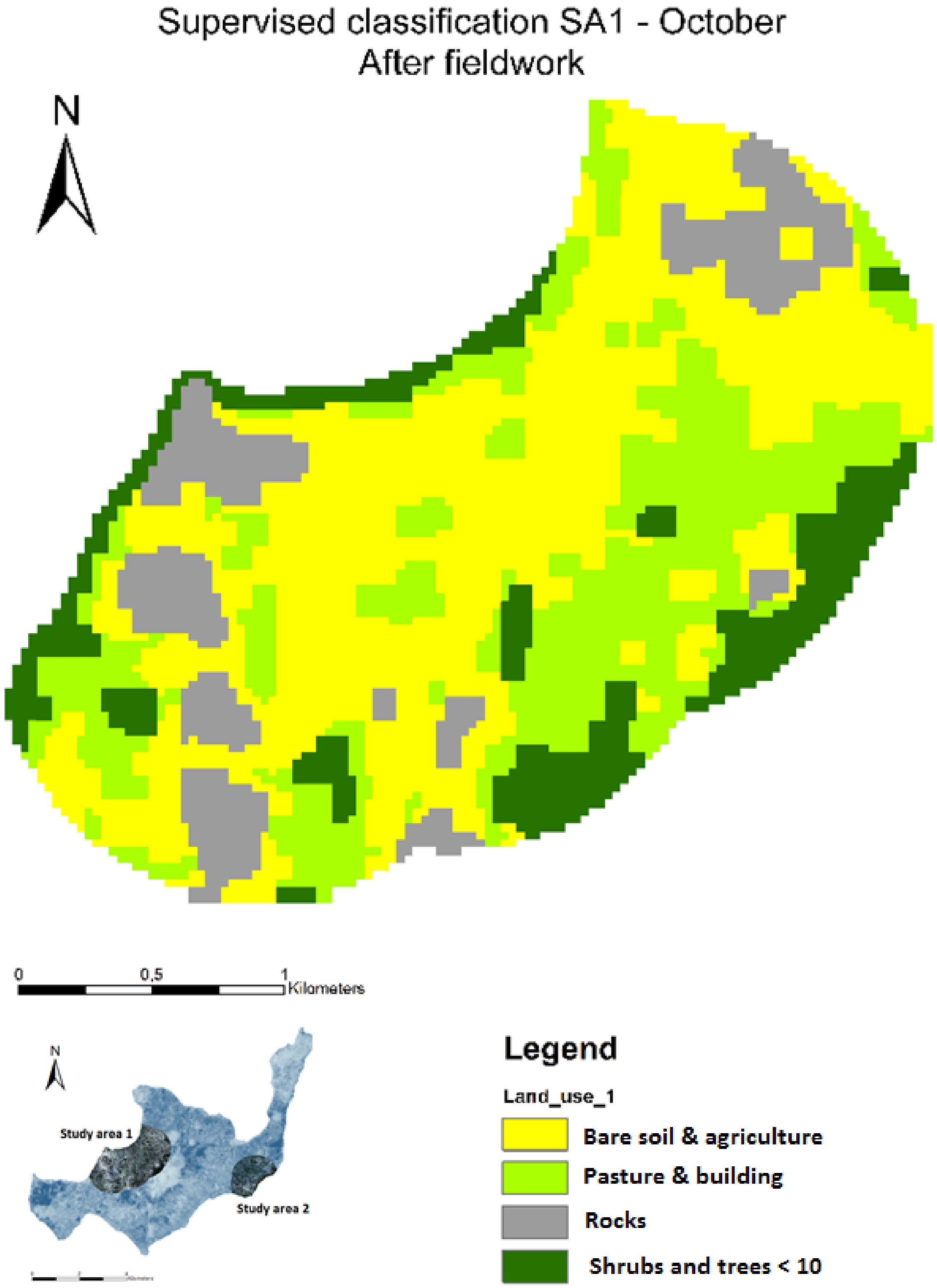


Figure 29: Supervised classification result of SA1 – October – after fieldwork.

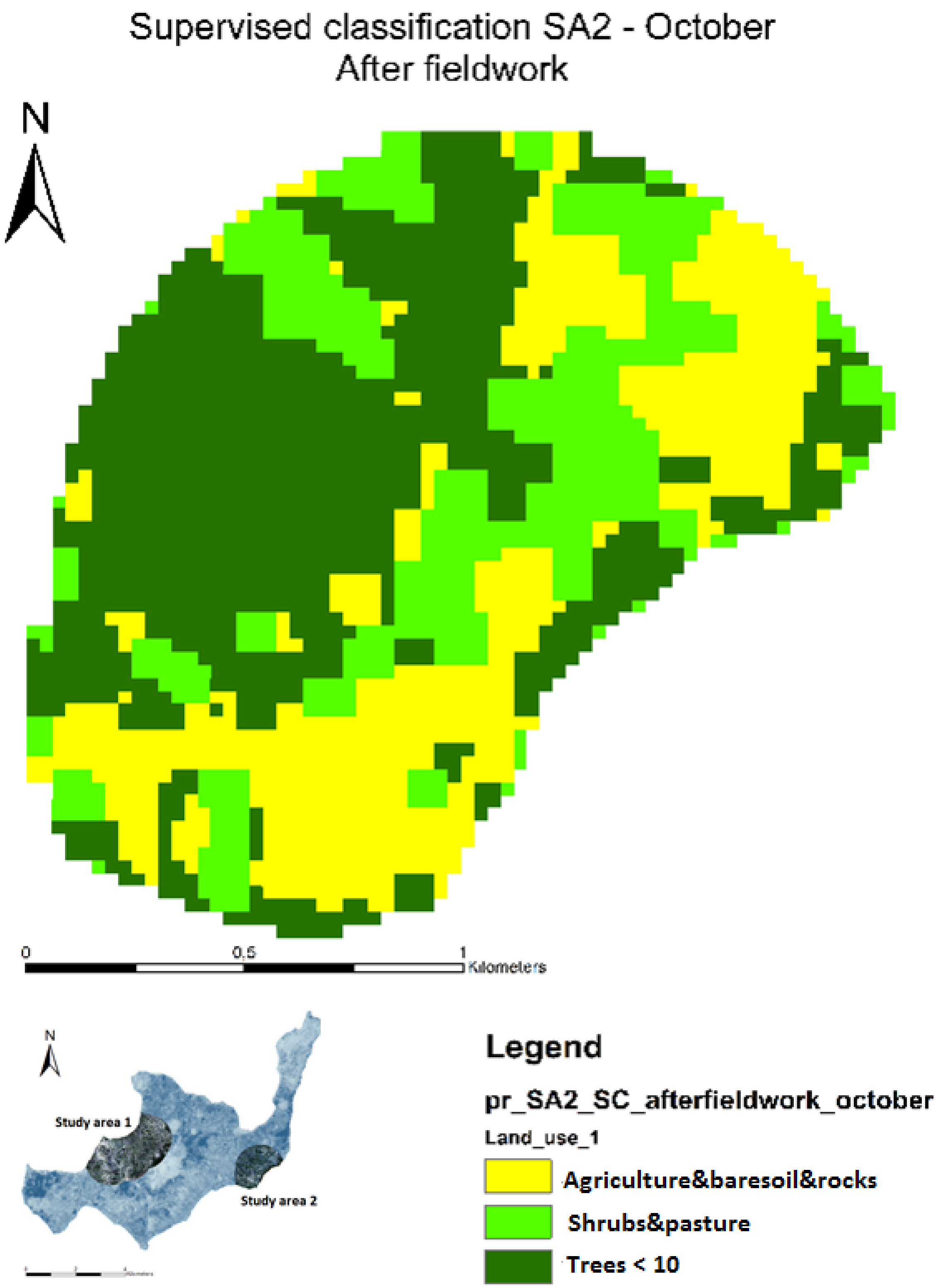


Figure 30: Supervised classification result of SA2 – October – after fieldwork.

3.2 Research question 2: Validation

The first objective of this study was to study the relationship between vector abundance and agricultural land use patterns. However, the overlap of the histograms of the different crop type classes was considerable and the between-classes values of the created dendrograms during the classification was exceeding the defined threshold of 0.9. Furthermore, the validation results showed an accuracy below 30%. Therefore, there was decided to merge those classes into one class: agriculture. The same is true for “rocks and bare soil” and “shrubs, trees and pasture”.

The classification process was repeated for the newly defined classes, which indeed increased the value of the between class distance, the total accuracy and the Kappa coefficient. The analyses were carried out both for the initial and for the new study areas. The initial study areas did match the study areas of the fieldwork, the new study areas did not match the fieldwork study areas, which can be found back in the results (Table 4a and b). The total accuracy of the initial study areas classification was higher than the total accuracy of the new study area since almost half of the validation data felt outside the ranges of the newly defined study areas. This means those validation data could not be used and together with the non-matching areas of the fieldwork and the analyses. This explains the lower values of total accuracy for the newly defined study areas.

0.00 values for the classes wet nature and for shrubs/trees as well for both old and new study area 2 were found in the table. Those classes were not classified during the classification as such, but were present in the validation dataset which leads to a value of 0.00. When using the formula to calculate the accuracy of those classes (Appendix F), there would always be a 0.00 value which caused the accuracy to be 0.00 as well. Furthermore, the classification of Rusinga Island had an accuracy of 56%, new study area 1 had an accuracy of 55%, while the new study area 2 only had a total accuracy of 22%. This means that the classification was not much better than a random classification. The accuracy for the images of October was for Rusinga Island 35%, for study area 1 36% and for study area 2 43%. So, study area 2 had a considerable better classification result than the image of March (43% compared to 22%). Since the classification accuracy of study area 2 was that low, an unsupervised classification for this study area was done in the hope that the accuracy improved by this. In this unsupervised classification, the classification data were used for defining which class is which after the unsupervised classification was finished. The result of this unsupervised classification can be found in Figure 31 and Table 4a and b contain the validation results. (Appendix F)

Table 4a: Error matrices outcomes March: Rusinga Island and both study areas including the total accuracy (range is between 0 and 1), kappa coefficient (range is between 0 and 1) and user's and producer's accuracy (%).

Rusinga island	User's accuracy	Producer's accuracy
Shrubs/trees&pasture	85,14	53,85
Water	50,00	100,00
Rocks/bare soil	68,97	46,51
Agriculture	38,79	75,00
Building	38,46	55,56

Total accuracy	0,57
Kapa coefficient	0,38

Study area 1 - old	User's accuracy	Producer's accuracy
shrubs/trees&pasture	70,00	80,77
rocksANDbaresoil	62,50	52,63
agriculture	45,45	47,62
building	83,33	35,71

Total accuracy	0,64
Kapa coefficient	0,45

Study area 1 - new	User's accuracy	Producer's accuracy
shrubs/trees&pasture	63,46	70,21
rocksANDbaresoil	21,05	36,36
agriculture	38,10	38,10
building	14,29	33,33

Total accuracy	0,55
Kapa coefficient	0,25

Study area 2 - old	User's accuracy	Producer's accuracy
Agriculture	45,65	53,85
Bare soil/rocks	48,57	60,71
Shrubs/trees	76,09	59,32
Water	100,00	100,00
Building	16,67	33,33
wet nature	0,00	0,00

Total accuracy	0,56
Kapa coefficient	0,36

Study area 2 - new	User's accuracy	Producer's accuracy
Agriculture	26,67	12,50
Bare soil/rocks	23,53	66,67
Building	14,29	100,00
Wet nature	0,00	0,00
Shrubs/trees	0,00	0,00

Total accuracy	0,22
Kapa coefficient	0,03

Table 4b: Error matrices outcomes October: Rusinga Island and both study areas including the total accuracy (range is between 0 and 1), kappa coefficient (range is between 0 and 1) and user’s and producer’s accuracy (%).

Rusinga island	User's accuracy	Producer's accuracy
Shrubs	27,36	37,66
Trees <10	18,46	52,17
bare soil, rocks, pasture	64,06	30,37

Total accuracy	0,35
Kapa coefficient	0,03

Study area 1 - new	User's accuracy	Producer's accuracy
shrubs/trees&pasture	75,00	10,34
rocksANDbaresoil	8,33	33,33
agriculture	43,18	73,08
building	32,00	29,63

Total accuracy	0.36
Kapa coefficient	0.13

Study area 2 - new	User's accuracy	Producer's accuracy
shrubs/trees&pasture	35,00	26,92
rocksANDbaresoil	13,79	57,14
agriculture	71,43	49,02

Total accuracy	0,43
Kapa coefficient	0,11

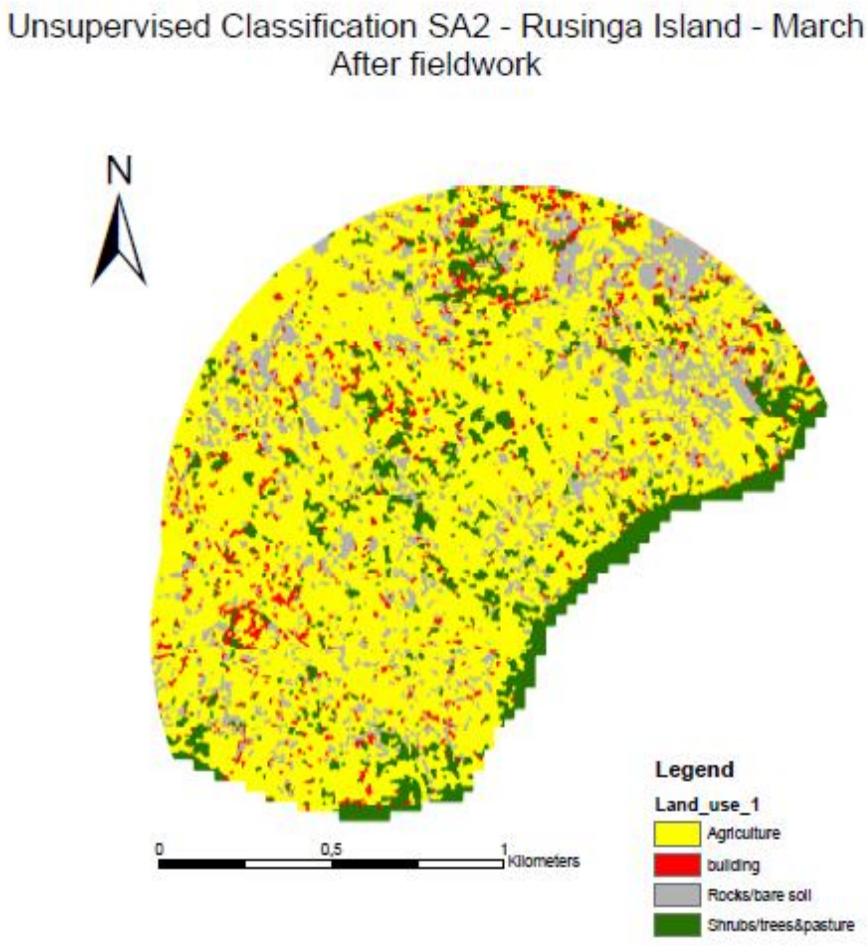


Figure 31: Unsupervised classification result of new study area 2, Rusinga Island, Kenya – March – after fieldwork.

Table 5: Error matrix outcome unsupervised classification "new" study area 2, including user's and producer's accuracy.

Study area 2 - new - Unsupervised	User's accuracy	Producer's accuracy
Shrubs/trees	26,67	25,00
Agriculture	23,53	71,88
Bare soil/rocks	14,29	33,33
Building	0,00	33,33
Wet nature	0,00	0,00
Total accuracy	0,44	
Kapa coefficient	0,17	

Table 5 makes clear that the unsupervised classification of the new study area 2 had a higher accuracy (44%) compared to the supervised classification of this area (22%). Since the outcome of the classification also influences the regression analysis, the unsupervised classified image of new study area 2 was used for further analysis.

3.3 Research question 3: relational analysis

To answer the third research question, the relationship between vector abundance and land use was studied by means of a regression analysis (OLS). As became clear during the land use classification and in the validation of the land use classification (Chapter 3.1 and 3.2), the agricultural crop type classes on their own were not different enough. Therefore, those classes were merged into one class: agriculture. The classification and the regression analyses were performed for the remaining land use classes including the class “Agriculture” to find out whether the low accuracy was the cause of finding no relationship between agricultural land use types and vector abundance. The results of the R^2 and R^2 adjusted values are given in Table 6 per analysis (*Anopheles gambiae* and *Anopheles funestus*, *Anopheles gambiae*, *Anopheles funestus*, each per study area and for Rusinga Island as a whole). An R^2 of 0 means the data does not fit the developed statistical model at all, a value of 1 means there is a perfect fit between the data and the model: the dependent variable is totally explained by the defined model.

Based on fieldwork experiences the definition of Rocks did not only contain rocks. This class also contained bare soil and the same is true for shrubs. Shrubs also contained trees and pasture. What was prominent was that there was no relationship at all between a particular land use type and malaria vector abundance, while this was expected according to the hypothesis (Chapter 1.2). The highest R^2 was found for the class Rocks in study area 2 for the combination of both malaria vectors, which had a value of 0.010529 for March. In October, the highest value was found for shrubs & pasture which was 0.048. This means that the goodness of fit of the model was 1% and even less for other land use types in March and 4.8% and less for other land use types in October. Those results did not say anything about a possible relationship and therefore it was decided not to continue with the geographically weighted regression. The geographically weighted regression had as goal to combine the classes into one model, but with those low R^2 values the regression results would not improve significantly.

What can be seen in Table 6 is that the class building in both analyses of study area 2 (March) had no results, since the data were multi-collinear. Furthermore, the analysis in study area 2 was not done for the malaria vector Anopheles, because there were not enough mosquitoes to do a viable analysis. The interpolation gave only values of 0.0 for the vector abundance and there was only 1 spot in study area 2 which contained a catch of 1 vector. This means that the regression analysis output certainly does not show any relationship, since every land use type contained the same amount of mosquitoes: 0.0. While in study area 1 there were enough mosquitoes of *Anopheles gambiae* present to perform the analysis.

Next to the values of the R^2 that are given, other statistics were calculated as well by the OLS tool in ArcGIS. Such as the Akaike's Information Criterion (AIC), which is a relative measure of the performance used to compare models, the smaller AIC indicates a superior model. In this case all AIC values were high, which means there is no superior model within the models determined. Furthermore, the joint F-statistic value was calculated which is used to assess the overall model significance. All models were non-significant according to those values. The joint F-statistics probability was calculated as well which is the probability that none of the explanatory variables have an effect on the dependent variable and this was indeed the case. The fourth value that was computed was the Wald statistics probability value, which is the probability that none of the explanatory variables have an effect on the dependent variables which was also the case for those models. Also the Jarque-Bera statistic was calculated which is used to determine whether the residuals deviate from a normal distribution. The probability values state that the residuals were also not normally distributed. The variance of the error term (residuals) was calculated as the last value and determined as sigma-squared. The variance appeared to be very large for all explanatory variables.

Table 6a: Outcomes OLS tool ArcGIS March: R^2 and R^2 adjusted per land use type for the three different datasets: both *Anopheles gambiae* and *Anopheles funestus*, *Anopheles gambiae* alone and *Anopheles funestus* only.

Rusinga Island - Both gambiae & funestus

Land use type	R2
Rocks	0,000148
Agriculture	0,002429
Water	0,005367
Shrubs	0,002131
Building	0,000018

Rusinga Island - gambiae

Land use type	R2
Rocks	0,000148
Agriculture	0,002426
Water	0,005365
Shrubs	0,002128
Building	0,000018

Rusinga Island - funestus

Land use type	R2
Rocks	0,000148
Agriculture	0,002429
Water	0,005367
Shrubs	0,002131
Building	0,000018

study area 1 - Both gambiae & funestus

Land use type	R2
Rocks	0,002527
Agriculture	0,001233
Shrubs	0,0029
Building	0,000126

study area 1 - gambiae

Land use type	R2
Rocks	0,002529
Agriculture	0,001234
Shrubs	0,002902
Building	0,000126

study area 1 - funestus

Land use type	R2
Rocks	0,002527
Agriculture	0,001233
Shrubs	0,0029
Building	0,000126

study area 2 - Both gambiae & funestus

Land use type	R2
Rocks	0,010529
Agriculture	0,000233
Water	0,001984
Shrubs	0,000567
Building	multicolinearity

study area 2 - funestus

Land use type	R2
Rocks	0,000567
Agriculture	0,001984
Water	0,000233
Shrubs	0,000112
Building	multicolinearity

Table 6b: Outcomes OLS tool ArcGIS October: R^2 and R^2 adjusted per land use type for the three different datasets: both *Anopheles gambiae* and *Anopheles funestus*, *Anopheles gambiae* alone and *Anopheles funestus* only.

Rusinga Island - Both gambiae & funestus

Land use type	R2
Trees < 10	0,003431
Shrubs	0,001985
bare soil, rocks, pasture and agriculture	0,000542

Rusinga Island - gambiae

Land use type	R2
Trees < 10	0,003427
Shrubs	0,001982
bare soil, rocks, pasture and agriculture	0,000541

Rusinga Island - funestus

Land use type	R2
Trees < 10	0,003431
Shrubs	0,001985
bare soil, rocks, pasture and agriculture	0,000541

study area 1 - Both gambiae & funestus

Land use type	R2
Rocks	0,000131
Agriculture & bare soil	0,005907
Shrubs	0,012353
Building & pasture	0,000049

study area 1 - gambiae

Land use type	R2
Rocks	0,000131
Agriculture & bare soil	0,005903
Shrubs	0,012352
Building & pasture	0,000049

study area 1 - funestus

Land use type	R2
Rocks	0,000132
Agriculture & bare soil	0,005907
Shrubs	0,012353
Building & pasture	0,000049

study area 2 - Both gambiae & funestus

Land use type	R2
Shrubs&pasture	0,047978
Trees<10	0,041917
Agriculture&baresoil&rocks	0,000262

study area 2 - funestus

Land use type	R2
Shrubs&pasture	0,047978
Trees<10	0,041917
Agriculture&baresoil&rocks	0,000262

3.4 Comparison of March and October

According to the third research question, a comparison between the two seasons in terms of land use and in terms of relationship with vector abundance data was necessary. It was difficult to compare the two seasons, since two different datasets with a difference in resolution and thus in accuracy were used. The land use was not displayed correctly (accuracy = 35% only) for the low resolution Landsat image (30m resolution), which made a comparison with March (2.4m resolution) not worthwhile. However, when looking at the gathered data without looking at the maps, it became clear that the land use in both areas was comparable both in terms of the proportion of land covered by those land use types and number of different land use types and this was not the aspect that causes the difference in low/high amount of mosquitoes in study area 2 and 1 respectively. Furthermore, the agricultural crop type classes were merged to increase the accuracy of the classified images and the original objective of assessing the relationship between vector abundance and agricultural crop types was not reached. However, the relationship between vector abundance and land use types more general was assessed.

The difference between the two seasons with regard to the relationship between vector abundance and land use types was not studied. Since there is no relationship found at all between any land use type and vector abundance on the island in any of the seasons, a comparison for the two seasons was not made.

4. Conclusion, discussion and recommendation

In this study, the objective was to find out whether there is a spatial relationship between specific agricultural crops and malaria vector abundance for the two crop growing seasons separately in the SolarMal study area: Rusinga Island. To reach this objective, three research questions were formulated and answered accordingly:

1. What agricultural land use patterns* based on crops are typical for Rusinga Island? **The term "land use patterns" is defined as spatial-temporarily delineated units of specific crop classes.*
2. What is the accuracy of the created land use classification maps?
3. Is there a relationship between specific agricultural crops and vector abundance for the two crop growing seasons separately?

4.1 Conclusions

The first research question focussed on the agricultural land use patterns that occurred on Rusinga Island. According to the data gathered in October 2013, the following land use types were defined: beans, cow peas, tomatoes, kales (sukumawiki), maize, sorghum, sweet potato, sweet pepper, bare soil, pasture, trees, water, shrubs, swamp and building. More information about the sowing and harvesting dates and in which crop growing season the different crops grow was given in Table 1, Chapter 1.2 . The data gathered in February/March 2014 can be found with the DVD: *fieldwork data study area 1* and *fieldwork data study area 2*. Those datasets were used to improve the first classification and to validate the data. Based on this knowledge, the data gathered in the field in October 2013, the data gathered in February/March 2014 and the first land use classification, the final land use classifications were performed.

From the fieldwork done in February and March 2014, it became clear that Rusinga Island was very heterogeneous in terms of land use patterns. Since the pixels within the satellite images agreed with 2.4 m and 30 m in reality and even variation within one meter did occur, those pixels should contain more than 1 class.

The final land use classifications that were created were based on the data gathered both in October 2013 and in February and March 2014. It became clear from those classifications and thus from the data gathered in the field that shrubs and trees were mainly growing on top of the hills of Rusinga island, water was only occurring in Lake Victoria (along the edges of Rusinga island) and rocks and bare soil were mainly located on slopes near the hill tops and along the shoreline. However, some bare soil/rock areas seemed to be classified as agriculture and this was also the reason why agriculture was overrepresented in the image of March (Figure 25). What can be seen is that most agriculture was found near the lake shore, although it also includes bare soil/rock areas. The class "building" was overrepresented as well. Along the lake shore there was one line of pixels determined as building, while it should be water. Since the image of October (Figure 28) only had three classes left from which one class was a mixture of all classes except trees and shrubs as was already expected and the class shrubs and the class trees were overrepresented. Those last classes probably include mixed pixels as well, which means they include pixels that belong to another class as well.

The second research question focussed on the validation of the classifications done. According to the validation done for the new study areas and Rusinga island as a whole, the classification of Rusinga island (March) had the highest accuracy of 57%, study area 1 had an accuracy of 55% and study area 2 one of 22%. After this validation the classification of study area 2 was adapted to an unsupervised classification with an accuracy of 44%. The validation results of October were the following: Rusinga island had an accuracy of 35%, study area 1 an accuracy of 36% and study area 2 an accuracy of 43%.

The third research question tried to reach the objective of this study which was finding out whether there is a relationship between specific agricultural crop types and vector abundance for the two crop growing seasons. The part of the objective to study the relationship between agricultural crop types and vector abundance was not reached, since the accuracy of the created maps was too low to work with. In this case, no relationship was found between a particular crop type and vector abundance which could be the answer to the last question. However, the analyses were performed for the merged class: agriculture and the other land use types. Still, no relationship was found for any type of land use with vector abundance.

Thus, the regression analysis did not reveal any relationship between vector abundance and land use which did not match the hypothesis as given in Chapter 1.2. One explanation is that the sampling of the land use classes was done differently when comparing it to the sampling method of the vector abundance. Land use was sampled according to the systematic random sampling principle and thus covered all the classes more or less equal. The difference with the vector abundance data is that those vector data are linked to the households on the island and there was sampled on basis of simple random sampling of 80 households per 6 weeks' time interval. Each household was sampled on two occasions. Every household had an equal chance of selection. However, due to the random sampling pattern, areas with a high housing density were oversampled. This means that there were not only few mosquitoes per location, but that the sampling points were also not spread equally over the island as well and only focussed on the populated areas of the island, where particular land use types such as shrubs are less common. (Hiscox et al. 2012b)

4.2 Discussion

4.2.1 Data

The first issue to discuss is the data that was used. The data gathered in the field was gathered at a different point in time then the satellite images were taken. The Quickbird image was taken on the 10th of March 2012 and the Landsat image was taken on the 1st of October 2012, while the data in the field was gathered between the 15th and 18th of October 2013 and during the whole month February and the first two weeks of March 2014. This mismatch caused a discrepancy between the data gathered and the images that were analysed with those data which leads to a lower accuracy of the classification done. It was possible to ask the farmers what they were growing on their fields two years ago, but this information was probably biased due to errors in the recall of information. Another problem faced was that especially in March, but also in October (Table 1, Chapter 1.2) crops were just sowed at the moment the satellite image was taken and not fully grown yet when the fieldwork was done. Which made the classification of the different crop types very hard. The satellite image only showed bare soil reflection values and agricultural fields were therefore classified as bare soil instead of agriculture. Engvall et al. (1977) and Mutua et al. (2010) used another method for specific crop type classification and was able to overcome this problem. There is elaborated on this method in section 4.3.1.

To improve the accuracy of the classifications done, the data could be gathered using a more accurate GPS or using a different sampling method. Especially for the Quickbird image this is necessary, since the resolution of that image is 2.4 m. while the GPS on the tablet could have a deviation of 10 m. and sometimes even 20 m. This means that there is discrepancy between the collected data and the Quickbird image. To reduce this error, the sampling can be done differently. What can be done to reduce the deviation of the tablet is approaching the sites from four directions to obtain a GPS fix of the sampling points. If the coordinates are all the same (GPS fix), the correct site is reached. The accuracy of the tablet is studied at the moment by a student and the results of his study can be used for the improvement of this study. One of the methods that can be used is called differential sampling (Cornelius et al. 1994). This method requires two (similar) GPS receivers and by use of those 2 receivers they cancel out their errors and give an accuracy of 2 – 15m.

The goal was in the end to compare the classification and regression results for Rusinga for two seasons: March and October. Therefore, another Quickbird image was necessary for the analysis of October 2012. Unfortunately, the company from which the image should be ordered stopped existing and it was not possible to acquire this image in time. This became clear after the fieldwork already was finished and some adaptations regarding the time schedule of finishing the study needed to be made. Another image was requested: a Landsat ETM+ image of the 1st of October, 2012 (section 3.2). This Landsat image has a much lower resolution (30 m) when comparing it to the Quickbird resolution (2.4 m), which makes the comparison between the two different seasons complicated due to the discrepancy in resolution.

The change to a Landsat image brought some troubles in ArcGIS: classifying the image was impossible, since the histogram function on the training sample manager toolbar gave an error: “out of memory”. However, the computer which was used for the analysis had a lot of memory left. It took a lot of time to solve it, but when one of the bands of the original image is displayed and the toolbar is opened first then everything works normal and the altered study areas could be uploaded and classified. Otherwise, the classification could have been done in Erdas Imagine or the data could be re-saved within ArcGIS or the mask tool in ArcGIS for the creation of the study areas could have worked out.

4.2.2 Methods

The first land use classifications of both study areas and of Rusinga Island as a whole showed that the island is very heterogeneous in terms of land use patterns. However, after performing the final classification, the different agricultural land use patterns were too similar and had to be merged into one class which is named agriculture. This became clear after using the dendrogram tool for which the between-class values for the separated agricultural land use patterns were smaller than the defined threshold of 0.9 and the accuracy was < 30%. After the merging, those values increased towards acceptable levels above 0.9 for the between-class values and between the 40% and 60% for the total accuracy. Since the pixels within the satellite images agreed with 2.4 m and 30 m in reality and even variation in land use patterns within one meter did occur on Rusinga Island, those pixels should contain more than 1 class. According to Forzieri et al. (2012) this causes mapping errors and thus a lower accuracy due to mixed spectral signatures.

Several studies proved that other land use types then specific agricultural crop types, such as unmanaged pasture, swamp and agriculture in general, can explain mosquito abundance as well (Afrane et al. 2005; Beck et al. 1994; Lindblade et al. 2000; Munga et al. 2006). Since only agriculture was defined as agricultural land use on the island, only something can be said about agriculture in general and about the four classified land use patterns occurring on the island, which are shrubs/trees and pasture, rocks/bare soil, buildings and water.

This means that the objective of this study is only partly reached, not the relationship between specific agricultural land use patterns and vector abundance was studied, but the relationship between agriculture/ other land use patterns and vector abundance was studied. This means the study did not focus on specific crop types, but on agriculture in general. However, the set-up of the study was not changed.

During the first supervised classifications done after the data gathering of October 2013 it became clear that some classes should be merged, since the distance between the classes according to the dendrogram was too small (< 0.9). Another issue that was already mentioned in section 4.1, was that some classes did not occur in one study area, while they did in the other. For example tomatoes did not occur in study area 1 according to the data, but did occur in study area 2 and bare soil did not occur in study area 2, but did in study area 1. Furthermore, study area 2 contained sweet pepper and cow peas, while study area 1 did not. Study area 1 contained maize and sorghum and study area 2 did not. Rusinga as a whole contained everything, except cow peas and sweet pepper. An explanation for this is that cow peas and sweet pepper were very similar to another class, Sukumawiki, in terms of how they appear in the image of March and were merged into that class. In Figures 17 and 18 this merged class is named Sukumawiki.

Agriculture appeared to exist especially close to the shoreline according to the classifications done before the fieldwork. Since there is no water on the island, the people depend on the water from the lake, which makes irrigation of crops in the middle of the island very difficult. When looking at the classification images after the fieldwork data is included, agriculture seems to be overrepresented in the image when comparing it to the real-life situation. The classification of shrubs/trees and pasture, water and rocks/bare soil seemed to be more reliable. The class building was overrepresented in both classifications done before and after the fieldwork. This means that more pixels were classified as building than there were in reality, which reduces the accuracy of the classification. An explanation for this is that the reflection of the roofs of the buildings is very high and can be compared to the reflection values of the waves breaking at the shoreline: when looking close to the classification of study area 2 the shoreline was classified as building. Some bare soil and rock areas were classified as building as well, probably the reflection of the light grey rock areas on the island is comparable to that of the light roofs of the buildings as well.

The classes shrubs and trees appeared to have similar reflection values according to the between-class values (< 0.9). Since trees were defined in the field as trees < 10, areas containing trees had a soil cover of shrubs. An explanation for this is that the trees often did not have a closed canopy which leads to similar reflection values of both classes. Rocks and bare soil both did not have vegetation cover and had a bright colour. This resulted in comparable reflection values as well.

The only clear pattern for the unsupervised classifications of study area 1 were the agricultural fields along the shoreline and the road. However, in the classification based on the land use pattern dataset of February and March 2014, the road could not be distinguished any more. The reason for this was that road was not defined as a separate class, it was defined as rocks/bare soil. This made the difference between road and its neighbouring land use patterns less clear, because the neighbouring land use types were often rocks/bare soil. The land use pattern rocks/bare soil seemed to be underrepresented and the land use pattern agriculture overrepresented. A reason for this is that the satellite image was taken in March, when crops were not fully grown yet. This means that a lot of bare soil was still visible and during the classification, the bare soil areas which were not agriculture in the image had the same reflection values as the training samples of agriculture and were therefore classified as agriculture.

Muthuri et al. (2005) showed that for example for maize it takes 90 days to be fully grown after sowing. This means for the second season for which the onset of maize on Rusinga Island was the first of September, that it was fully grown at the end of December. This can explain why agricultural fields often are classified as bare soil instead of agriculture.

Study area 2 basically showed the same patterns in land use as study area 1 in the first supervised classifications done before the fieldwork, but the classification after the fieldwork was slightly different. Only three classes were left, containing the classes agriculture, rocks/bare soil and buildings. Agriculture was merged with shrubs/trees and pasture since those classes showed too much overlap to perform a classification. Including this knowledge, the class agriculture seemed to be underrepresented, while building and rocks/bare soil seemed to be overrepresented based on knowledge of the area according to the same reasons as mentioned for study area 1. However, this time for rocks/bare soil and agriculture it seemed to be turned around: most agriculture is defined as rocks/bare soil instead of rocks/bare soil that was classified as agriculture.

According to the validation done for the new study areas and Rusinga island as a whole, the classification of Rusinga had the highest accuracy of 57%. Study area 1 had an accuracy of 55% and study area 2 one of 22% for March. For October this accuracy was respectively 35% for the whole of Rusinga Island and 36% and 43% for study area 1 and 2. So the over- and underestimation of the different classes is indeed the case. The classification result is not very accurate, since the accuracy values were lower than the lowest value found in literature: 64.5% (García-Mora et al. 2011). The accuracy of study area 2 (March) was very low compared to the other classifications and there was decided to do an unsupervised classification as well for which the classes were based on the classification dataset. The accuracy indeed became higher: 44%. This is not surprising since the different classes defined within the supervised classification were too similar. While the unsupervised classification defined classes in such a way that there is a large enough difference between the classes. Furthermore, an accuracy of 22% means that the difference with a random classification is small and therefore a higher accuracy result was expected as well.

One of the factors that caused this accuracy not to be higher was that the histograms of the different bands of the image were not normally distributed, but skewed to the right for every band due to atmospheric influences/background interference (Manolakis and Shaw 2002). A different factor were the classes which were determined during the fieldwork. In the field every crop type was determined as a specific agricultural crop type. However the sizes of the fields were very different. The fields near the houses were usually small (10 m length and width as maximum size), while the fields further away were larger (50 length and 20 m width). A different classification could be made when the small fields are defined as "small-scale agriculture" and the larger fields are defined as "large-scale agriculture" since the smaller fields have the greatest impact on mosquito house entry risk. The reason for this is that the crops grown there affect the humidity in and around the houses (Ijumba et al. 2002; Klinkenberg et al. 2005; Norris, 2004). Furthermore, the sampling points occurred in a heterogeneous environment for which several classes occurred within 1 m^2. Therefore, some classes were not similar to the class that was distinguished by the satellite due to mixed pixels. Other factors are already described in Chapter 4.2.1.

When the classification results of both March and October are compared to classification results of other studies by means of the results of the error matrices (Diuk-Wasser et al. 2007; Beck et al. 1994; Manandhar et al. 2009; Garcia-Mora et al. 2011) , it appears that the accuracy is low especially for the classifications of October. The studies that are already done have an accuracy of their classification between 60 and 90 %. When comparing the results of October (35%, 36% and 43% respectively) to those values, it appears that improvement of the classification still is necessary. The results for Rusinga and study area 2 for March are valid to be used for the regression analysis, since the values of 57% and 55% are just lower than the lowest accuracy found in literature. When looking at the user's accuracy which were calculated by means of the error matrices of the classified images of October for Rusinga and study area 1 and 2, it appears that the classes shrubs and trees for Rusinga contained the largest errors (user's accuracy was only 18% and 27% respectively). Furthermore, for study area 1 and 2, the class bare soil and rocks contained the largest errors, user's accuracy was only 8% for study area 1 and 14% for study area 2. This is explained by the discrepancy between date of satellite image and the date on which crops are fully grown (Chapter 5.2.1).

Since Ge et al. (2007) proved that the accuracy of land use classifications influence analyses that use those classifications as input afterwards and in their case on climatic simulations. Probably, this is also the case for the regression analysis. However, the exact effect is not studied and could include a whole new study.

The R^2 values were lower than expected, since a relationship between a particular land use type and vector abundance was expected. An explanation is that the fieldwork study areas were different from the study areas for which the analyses afterwards were done. The study areas were adapted, since the size of the first study areas was not based on something, while the new study areas now are based on the hotspot analysis and the distance a malaria vector can fly (Chapter 2.1).

Furthermore, the low R^2 values can be caused by the fact that the gathered mosquito abundance data of the whole year over the whole island in the SolarMal project are not representing correct numbers for the areas outside households, since those data are linked to the households. Next to this, the interpolation tool cannot cover the whole island with this dataset. This can lead to a lower R^2 than there actually is for particular land use types.

Another problem faced here, is that the household sampling was not accurate enough when comparing the data with the Quickbird image of March. This also has to do with the GPS of the tablets (Chapter 4.2.2) which is not accurate enough to provide data that matches the exact location of the sample points in the Quickbird image (Figure 32).

Figure 32: Pansharpened Rusinga Island, Kenya (Quickbird, March), displaying the discrepancy between the accuracy of the data gathering of the households (green point) and the household itself on the image.

R2 adjusted was not used. It takes into account the phenomenon of increasing the value automatically when extra explanatory variables are added to the model, but only single variables are studied due to the low R2 outcomes.

Since the relationship was that low, there was not continued with doing a geographically weighted regression (GWR) and also the autocorrelation of the residuals was not checked. The Global Moran's I tool, ArcGIS 10.1, measures the spatial autocorrelation based on feature locations and attribute values using the Global Moran's I statistic, which should be checked before continuing the analysis (Esri 2013q; Varga et al. 2013). So, if the regression results would be improved somehow, those analysis can still be done by another student for example.

4.2.3. Fieldwork

Also in terms of fieldwork some issues can be discussed. For a classification, a total accuracy of 100% would be great, however, it is impossible to reach this in practice. (Beck et al. 1994; Diuk-Wasser et al. 2007; Garcia-Mora et al. 2014; Manandhar et al. 2009) When comparing the classification results with other studies, the results were lower than the accuracy values found in literature (lowest was 64.5% from the study of Garcia-Mora et al. 2014). There were/are some options to obtain an even better result. In the first place both the sampling done by the fieldworker in October and the sampling done in March/April was facing some difficulties. Sometimes, a point could not be reached due to a (too) steep slope or too dense bushes and the land use type should be estimated from a distance. This can lead to a wrong land use class given to that point since there was not known were the point exactly was. Those points were not skipped in the analysis, because otherwise the idea of covering the whole study area equally would be lost. There were approximately 30 points per study area that could not be reached out of the 210 sampling points for study area 1 and 270 sampling points for study area 2.

Another issue was that areas contained two or three different types of land use even within 1 m^2. It was decided to look at which type of land use was present most and this was the land use type that was used for that sampling point. During the fieldwork, it appeared to be the case often. Together with the inaccuracy of the GPS this can lead to a discrepancy between the exact data point within the satellite image and the belonging land use type. A possibility is that this pixel is a mixed pixel within the image, which makes it difficult to classify not only in the field but also with regard to the land use pattern classification.

Furthermore, some of the points that fell in an agricultural field could not be reached, because of dense bushes and fences made by the farmers themselves and neighbours were asked what was growing there. Since it is possible that there were some translation and understanding problems with the farmers during the fieldwork, the data collected from those points could be different from reality. This is a known problem for gathering data about mixed cropping. There was also asked for those agricultural fields what was planted there in the other season (March or October). Those are all issues that influence the accuracy of the classification and indirect the results of the regression analysis.

An explanation for the non-presence of several crop types in the study area 1 and 2 is that there were only a limited amount of points sampled (24) and only four points per crop were located. Therefore, it could be possible that for one crop type, four fields of that crop were already sampled and the analysis was continued with another crop type.

4.3 Recommendation

4.3.1 Data

The first recommendation about the data is that it is wise to use the same satellite image for both seasons to be able to make a proper comparison between those seasons. Furthermore, the accuracy of the satellite images used needs to be the same as the accuracy of the GPS used for data gathering (Chapter 4.2.1). This is needed to get a more reliable classification result in terms of accuracy of the classified image. What also improves the accuracy is that the data that is gathered in the field should be matching the timing of the acquisition of the satellite images. In this case this means that the sampling and the satellite image should be gathered in the period for which the crops are fully grown.

Another option was that the different growth stages of the crop types in March 2014 were written down during the fieldwork. In this way, there could be known if and in what way the satellite images show the reflection values of the gathered crops. This was done in the study of (Mutua et al. 2010)

4.3.2 Methods

If the results of this study are going to be used, it is recommended to first improve the classification results as those influence all further analyses (Ge et al. 2006). This can be done by doing the fieldwork at the same time as the satellite image is taken and when the crops are fully grown as is stated in Chapter 4.2.1. The classes used also can be revised, since they are often a combination of several classes as was discovered in the field. When this is done, future sampling is more accurate. However, there are other methods of classification that can be used and which probably lead to a higher accuracy.

An example of this is doing the classification based on the vegetation cover. This is also done in the study by Beck et al. (1994) which had an accuracy of 90%. They used Landsat imagery as well as remotely sensed data to be able to identify the elements within their study area. So their classification was not based on data gathered in the field which solves the problem of discrepancies between data, the time of acquisition and the accuracy of the data. This leads to a higher accuracy of the classification. So on base of remote sensing classification techniques for example on base of the NDVI, the classification results will be improved.

In terms of the fieldwork done for the classification, the variation within the area was only determined by means of visual inspection. However, there are methods available that determine the variation within an area. In a follow up study, landscape metrics can be used to calculate the variation within the area in terms of land use. (McGarigal 2012)

4.3.3 Fieldwork

Agriculture appeared to exist especially close to the shoreline according to the classifications done before the fieldwork. To proof this an additional regression analysis could be done in ArcGIS which gives the relationship between the distance from the lake with the amount of land cover dedicated to agriculture.

4.3.4 Other options

In terms of the different satellite imagery used within this study, Landsat and Quickbird, the comparison is not impossible. However, if a proper comparison is made, the cell resolution of one of the images should be changed.

The regression analysis can be performed in a different way, since there are several methods for doing so which probably lead to an improved regression result. The first option is to do the same analysis, but than with larval data instead of adult female mosquitoes. Since larvae depend on water pools and certain land use patterns are linked with those, a relationship possibly is found. (Nmor et al. 2013) already studied other environmental factors to be linked with larvae (topographic variables derived from SRTM and ASTER DEMs) and did find relationships. The same study was done by a student for adult female mosquitoes and no relationship was found. For this land use classification this can also mean that due to the adult mosquito data there is no relationship found and when larvae data would have been used a relationship may have been found, since the spread of the mosquitoes is based on the location of the breeding sites (Figure 33). The studies which proved that there was a relationship between a particular land use pattern and malaria vectors also focussed on larvae and not on mosquito adults. (Beck et al. 1994; Van der Hoek et al. 2001; Kebede et al. 2005)

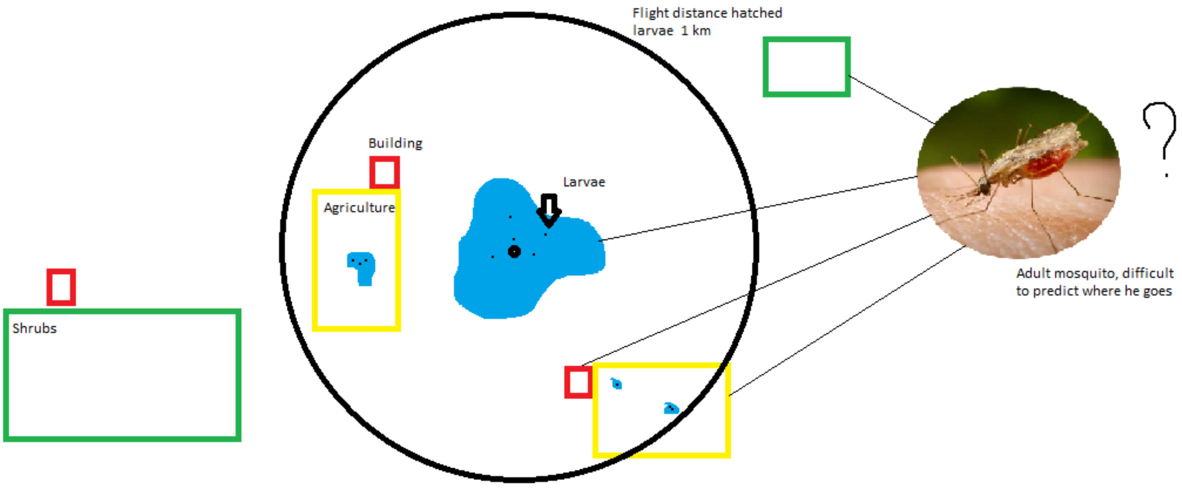


Figure 33: Comic with explanation why a relationship of vector larvae with land use is more easy to predict when comparing it to adult vectors (mosquitoes).

What also is a possibility, is to look specifically at the land use patterns near the houses as is done by Beck et al. (1994) as well. This can be done by creating a buffer around the houses, e.g. 1 km as that is the flight distance of mosquitoes. There can be calculated which percentages of different land use patterns the created buffer contains to get to know more about the proportion of particular land use patterns in the neighbourhood of houses. Then, statistical analysis can be done with the larvae and/or adult mosquito data to find out whether there is a relationship between vector abundance and a particular land use patterns.

Furthermore, the malaria vector abundance data could be used in a different way. Instead of using the exact numbers of malaria vectors, it can be changed into binomial data: are there mosquitoes at that particular location: yes (1) or no (0). Since there is difference in amount of mosquitoes, it probably can make a difference in what way the vector abundance data is used for the outcome of the regression.

If there is coming out a relationship with a particular patterns of land use after the changes being made to the analysis, it will be possible for the project to place their intervention traps more efficiently in a future malaria project: more traps in the mosquito attracting land use patterns. Another option is to reduce this land use pattern as much as possible, which is called land use management. However, if the land use is agriculture, this will be ethically, economically and probably politically impossible to do.

Another option for improving the classification results is to use the approach of Engvall et al. (1977). This approach does not use data gathered in the field, but used a time series of Landsat mean vectors for selecting agricultural fields for which ground truth data was available. This procedure was applied to individual Landsat pixels. The result was comparable to the ground truth data, which makes the method of Engvall et al. (1977) interesting for classifying the agricultural land use patterns on Rusinga Island with a high accuracy.

Literature

Afrane, Y.A., Lawson, B.W., Githeko, A.K., & Yan, G. (2005). Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of Anopheles gambiae (Diptera: Culicidae) in western Kenya highlands. *Journal of medical entomology, 42*, 974-980

Alonso, P., Lindsay, S., Armstrong, J., de Francisco, A., Shenton, F., Greenwood, B., Conteh, M., Cham, K., Hill, A., & David, P. (1991). The effect of insecticide-treated bed nets on mortality of Gambian children. *The Lancet, 337*, 1499-1502

Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., Meza, R.A., Roberts, D.R., Paris, J.F., & Spanner, M.A. (1994). Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American Journal of Tropical Medicine and Hygiene, 51*, 271-280

Biovision (2012). AEZs the FAO system. In http://www.infonet-biovision.org/default/ct/689/agrozones?search=agro

Blackwell, A., & Johnson, S.N. (2000). Electrophysiological investigation of larval water and potential oviposition chemo-attractants for Anopheles gambiae s.s. *Annals of Tropical Medicine and Parasitology, 94*, 389-398

Boussalis, C., Nelson, H.T., & Swaminathan, S. (2012). Towards comprehensive malaria planning: The effect of government capacity, health policy, and land use variables on malaria incidence in India. *Social Science & Medicine, 75*, 1213-1221

Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A, 30*, 1905-1927

Bukhari, T., Takken, W., Githeko, A.K., & Koenraadt, C.J. (2011). Efficacy of Aquatain, a monomolecular film, for the control of malaria vectors in rice paddies. *PLoS ONE, 6*, e21713

Carter, R., Mendis, K.N., & Roberts, D. (2000). Spatial targeting of interventions against malaria. *Bulletin of the World Health Organization, 78*, 1401-1411

Conelly, W.T. (1994). Population pressure, labor availability, and agricultural disintensification: The decline of farming on Rusinga Island, Kenya. *Human Ecology, 22*, 145-170

Cornelius, S., Sear, D., Carver, S., & Heywood, D. (1994). GPS, GIS and geomorphological field work. *Earth Surface Processes and Landforms, 19*, 777-787

de Souza, D., Kelly-Hope, L., Lawson, B., Wilson, M., & Boakye, D. (2010). Environmental factors associated with the distribution of Anopheles gambiae ss in Ghana; an important vector of lymphatic filariasis and malaria. *PLoS ONE, 5*, e9927

Dewan, A.M., & Yamaguchi, Y. (2009). Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka Metropolitan of Bangladesh during 1960–2005. *Environmental monitoring and assessment, 150*, 237-249

Diuk-Wasser, M.A., Toure, M.B., Dolo, G., Bagayoko, M., Sogoba, N., Sissoko, I., Traore, S.F., & Taylor, C.E. (2007). Effect of rice cultivation patterns on malaria vector abundance in rice-growing villages in Mali. *The American Journal of Tropical Medicine and Hygiene, 76*, 869

Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S.S., Yeung, S., Singhasivanon, P., Day, N.P.J., Lindegardh, N., Socheat, D., & White, N.J. (2009). Artemisinin Resistance in Plasmodium falciparum Malaria. *New England Journal of Medicine, 361*, 455-467

Droppová, V. (2011). The tools of automated generalization and building generalization in an ArcGIS environment. In, *Slovak Journal of Civil Engineering* (p. 1)

Engvall, J.L., Tubbs, J.D., & Holmes, Q.A. (1977). Pattern recognition of landsat data based upon temporal trend analysis. *Remote Sensing of Environment, 6*, 303-314

Erdey-Heydorn, M.D. (2008). An ArcGIS seabed characterization toolbox developed for investigating benthic habitats. *Marine Geodesy, 31*, 318-358

Ernst, K., Adoka, S., Kowuor, D., Wilson, M., & John, C. (2006). Malaria hotspot areas in a highland Kenya site are consistent in epidemic and non-epidemic years and are associated with ecological factors. *Malaria journal, 5*, 78

Esri (2013a). How Hot Spot Analysis (Getis-Ord Gi*) works. In. ArcGIS desktop help 10.1

Esri (2013b). How Create Signature works. In. ArcGIS desktop help 10.1

Esri (2013c). Majority Filter (Spatial Analyst) In. ArcGIS desktop help 10.1
Esri (2013d). Boundary Clean (Spatial Analyst). In. ArcGIS desktop help 10.1
Esri (2013e). Region Group (Spatial Analyst). In. ArcGIS desktop help 10.1
Esri (2013f). Nibble (Spatial Analyst). In. ArcGIS desktop help 10.1
Esri (2013g). Creating training samples. In. ArcGIS desktop help 10.1
Esri (2013h). Evaluating training samples. In. ArcGIS desktop help 10.1
Esri (2013i). Managing training samples. In. ArcGIS desktop help 10.1
Esri (2013j). Creating a signature file. In. ArcGIS desktop help 10.1
Esri (2013k). Dendrogram (Spatial Analyst). In. ArcGIS desktop help 10.1
Esri (2013l). How Maximum Likelihood Classification works. In. ArcGIS desktop help 10.1
Esri (2013m). Image classification using the ArcGIS Spatial Analyst extension. In. ArcGIS desktop help 10.1
Esri (2013n). Comparing interpolation methods. In. ArcGIS desktop help 10.1
Esri (2013o). Regression analysis basics. In. ArcGIS desktop help 10.1
Esri (2013p). How Iso Cluster works. In. ArcGIS desktop help 10.1
Esri (2013q). How Spatial Autocorrelation (Global Moran's I) works. In. ArcGIS desktop help 10.1
Evans, A.M. (1938). Mosquitoes of the Ethiopian Region. II.-Anophelini. Adults and Early Stages. *Mosquitoes of the Ethiopian Region. II.-Anophelini. Adults and Early Stages.*
Food and Agricultural Organization of the United Nations (FAO) (2013a). Crop calendar - An information tool for seed security.
In: http://www.fao.org/agriculture/seed/cropcalendar/welcome.do
Food and Agricultural Organization of the United Nations (FAO) (2013b). Crop Water Information. In: http://www.fao.org/nr/water/cropinfo.html
Forzieri, G., Moser, G., & Catani, F. (2012). Assessment of hyperspectral MIVIS sensor capability for heterogeneous landscape classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 175-184
Fuller, R., & Brown, N. (1996). A CORINE map of Great Britain by automated means. Techniques for automatic generalization of the Land Cover Map of Great Britain. *International Journal of Geographical Information Systems, 10*, 937-953
García-Mora, T.J., Mas, J.-F., & Hinkley, E.A. (2011). Land cover mapping applications with MODIS: a literature review. *International Journal of Digital Earth, 5*, 63-87
Ge, J., Qi, J., Lofgren, B.M., Moore, N., Torbick, N., & Olson, J.M. (2007). Impacts of land use/cover classification accuracy on regional climate simulations. *Journal of Geophysical Research: Atmospheres (1984–2012), 112*
Digital Globe (2013). Global Image. In http://www.digitalglobe.com/
Gouagna, L., Ferguson, H., Okech, B., Killeen, G., Kabiru, E., Beier, J., Githure, J., & Yan, G. (2004). Plasmodium falciparum malaria disease manifestations in humans and transmission to Anopheles gambiae: a field study in Western Kenya. *Parasitology, 128*, 235-243
Hightower, A.W., Ombok, M., Otieno, R., Odhiambo, R., & Hawley, W.A. (2000). A Geographical Information System Applied to a Malaria Field Study. *Geographical Targeting for Poverty Alleviation: Methodology and Applications*, 258
Hiscox, A., Maire, N., Smith, T., Mukabana, W.R., & Takken, W. (2012a). The SolarMal project: innovative mosquito trapping technology for malaria control. In (pp. 1-33): Wageningen UR
Hiscox, A., Mukabana, R., & Takken, W. (2012b). SolarMal six monthly progress report 1st july - 31st december 2012. In (pp. 1 -12)
Ijumba, J., Shenton, F., Clarke, S., Mosha, F., & Lindsay, S. (2002). Irrigated crop production is associated with less malaria than traditional agricultural practices in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 96*, 476-480
Johnston, K., Ver Hoef, J.M., Krivoruchko, K., & Lucas, N. (2001). *Using ArcGIS geostatistical analyst*. Esri Redlands
Kebede, A., McCANN, J.C., Kiszewski, A.E., & Ye-Ebiyo, Y. (2005). New evidence of the effects of agro-ecologic change on malaria transmission. *The American Journal of Tropical Medicine and Hygiene, 73*, 676-680
Kenyan Meteorological Department (2013). Weather data Rusinga Island, Kenya

Kleinschmidt, I., Schwabe, C., Shiva, M., Segura, J.L., Sima, V., Mabunda, S.J.A., & Coleman, M. (2009). Combining Indoor Residual Spraying and Insecticide-Treated Net Interventions. *The American Journal of Tropical Medicine and Hygiene, 81*, 519-524
Klinkenberg, E., McCall, P., Hastings, I.M., Wilson, M.D., Amerasinghe, F.P., & Donnelly, M.J. (2005). Malaria and irrigated crops, Accra, Ghana. *Emerging infectious diseases, 11*, 1290
Kotrlik, J.W.K.J.W., & Higgins, C.C.H.C.C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal, 19*, 43
Li, J., & Heap, A.D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics, 6*, 228-241
Lindblade, K.A., Walker, E.D., Onapa, A.W., Katungu, J., & Wilson, M.L. (2000). Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. *Tropical Medicine & International Health, 5*, 263-274
Lindsay, S.W., Jawara, M., Paine, K., Pinder, M., Walraven, G.E.L., & Emerson, P.M. (2003). Changes in house design reduce exposure to malaria mosquitoes. *Tropical Medicine & International Health, 8*, 512-517
Lindsay, S.W., Parson, L., & Thomas, C.J. (1998). Mapping the Ranges and Relative Abundance of the Two Principal African Malaria Vectors, Anopheles Gambiae Sensu Stricto and An. Arabiensis, Using Climate Data. *Proceedings: Biological Sciences, 265*, 847-854
Manandhar, R., Odeh, I., & Ancev, T. (2009). Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement. *Remote Sensing, 1*, 330-344
Manolakis, D., & Shaw, G. (2002). Detection algorithms for hyperspectral imaging applications. *Signal Processing Magazine, IEEE, 19*, 29-43
Maps & GIS Library (2013). Accuracy Assessment of an Image Classification in ArcMap. In http://www.youtube.com/watch?v=FaZGAUS_Nlo
Mbogo, C.M., Mwangangi, J.M., Nzovu, J., Gu, W., Yan, G., Gunter, J.T., Swalm, C., Keating, J., Regens, J.L., & Shililu, J.I. (2003). Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. *American Journal of Tropical Medicine and Hygiene, 68*, 734-742
McGarigal, K., SA Cushman and E. Ene (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. In. Amherst: University of Massachusetts
Meijerink, J., Braks, M.A.H., Brack, A.A., Adam, W., Dekker, T., Posthumus, M.A., Van Beek, T.A., & Van Loon, J.J.A. (2000). Identification of Olfactory Stimulants for Anopheles gambiae from Human Sweat Samples*, 26*, 1367-1382
Minakawa, N., Sonye, G., Mogi, M., & Yan, G. (2004). Habitat characteristics of Anopheles gambiae s.s. larvae in a Kenyan highland. *Medical and Veterinary Entomology, 18*, 301-305
Moore, D.A., & Carpenter, T.E. (1999). Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews, 21*, 143-161
Moss, W., Hamapumbu, H., Kobayashi, T., Shields, T., Kamanga, A., Clennon, J., Mharakurwa, S., Thuma, P., & Glass, G. (2011). Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community survey. *Malaria journal, 10*, 163
Mubyazi, G., Bygbjerg, I., Magnussen, P., Olsen, Ø., Byskov, J., Hansen, K., & Bloch, P. (2008a). Prospects, achievements, challenges and opportunities for scaling-up malaria chemoprevention in pregnancy in Tanzania: the perspective of national level officers*, 7*, 1-16
Mubyazi, G.M., Bygbjerg, I.C., Magnussen, P., Olsen, O., Byskov, J., Hansen, K.S., & Bloch, P. (2008b). Prospects, achievements, challenges and opportunities for scaling-up malaria chemoprevention in pregnancy in Tanzania: the perspective of national level officers. *Malar J, 7*, 135
Mukabana, W., Mweresa, C., Otieno, B., Omusula, P., Smallegange, R., Loon, J.A., & Takken, W. (2012). A Novel Synthetic Odorant Blend for Trapping of Malaria and Other African Mosquito Species*, 38*, 235-244

Munga, S., Minakawa, N., Zhou, G., Mushinzima, E., Barrack, O.-O.J., Githeko, A.K., & Yan, G. (2006). Association Between Land Cover and Habitat Productivity of Malaria Vectors in Western Kenyan Highlands. *The American Journal of Tropical Medicine and Hygiene, 74*, 69-75
Mutero, C., Kabutha, C., Kimani, V., Kabuage, L., Gitau, G., Ssennyonga, J., Githure, J., Muthami, L., Kaida, A., & Musyoka, L. (2004). A transdisciplinary perspective on the links between malaria and agroecosystems in Kenya. *Acta tropica, 89*, 171-186
Muthuri, C.W., Ong, C.K., Black, C.R., Ngumi, V.W., & Mati, B.M. (2005). Tree and crop productivity in Grevillea, Alnus and Paulownia-based agroforestry systems in semi-arid Kenya. *Forest Ecology and Management, 212*, 23-39
Mutua, F.N., Von Hagen, C., & Kuria, D. (2010). Cotton Mapping in Kenya: GPS-Based Data Collection–A Cost Comparison with High Resolution Satellite Imagery Mapping
Myers, W., Myers, A., Cox-Singh, J., Lau, H., Mokuai, B., & Malley, R. (2009). Micro-geographic risk factors for malarial infection. *Malaria journal, 8*, 27
Nchinda, T.C. (1998). Malaria: A Reemerging Disease in Africa. In, *Emerging Infectious diseases:Tracking trends and analyzing new and reemerging infectiuos disease issues around the world* (pp. 398 - 403). Geneva: World Health Organization
Nmor, J., Sunahara, T., Goto, K., Futami, K., Sonye, G., Akweywa, P., Dida, G., & Minakawa, N. (2013). Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. *Parasites & Vectors, 6*, 14
Norris, D. (2004). Mosquito-borne Diseases as a Consequence of Land Use Change. *EcoHealth, 1*, 19-24
Okumu, F.O., Killeen, G.F., Ogoma, S., Biswaro, L., Smallegange, R.C., Mbeyela, E., Titus, E., Munk, C., Ngonyani, H., Takken, W., Mshinda, H., Mukabana, W.R., & Moore, S.J. (2010). Development and Field Evaluation of a Synthetic Mosquito Lure That Is More Attractive than Humans. *PLoS ONE, 5*, e8951
Opiyo, P., Mukabana, W.R., Kiche, I., Mathenge, E., Killeen, G.F., & Fillinger, U. (2007). An exploratory study of community factors relevant for participatory malaria control on Rusinga Island, western Kenya. *Malaria journal, 6*, 48
Organisation, W.H. (1996). Investing in health research for development. In, *Report of the Ad Hoc Committee on Health Research Relating to Future Intervention Options*. Geneva: World Health Organisation
Organisation, W.H. (2012). World Malaria report 2012. In (pp. 1 - 2). Geneva: World Health Organisation
Panda, S.S., Hoogenboom, G., & Paz, J. (2009). Distinguishing blueberry bushes from mixed vegetation land use using high resolution satellite imagery and geospatial techniques. *Computers and electronics in agriculture, 67*, 51-58
Patz, J.A., & Olson, S.H. (2006). Malaria risk and temperature: Influences from global climate change and local land use practices. *Proceedings of the National Academy of Sciences, 103*, 5635-5636
Pawlak, Z. (2002). Rough sets, decision algorithms and Bayes' theorem. *European Journal of Operational Research, 136*, 181-189
Roberts, C.J.J., Jr. (2000). *Foundations of Parasitology*. (6th ed.). United States: McGraw-Hill Companies
Sauerwein, R.W., Roestenberg, M., & Moorthy, V.S. (2011). Experimental human challenge infections can accelerate clinical malaria vaccine development. *Nature Reviews Immunology, 11*, 57-64
Sipe, N.G., & Dale, P. (2003). Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia. *Malaria journal, 2*, 36
Sounny-Slitine, M.A. (2012). Geographic weighted regression on 911 phone calls. In. You Tube
UPCC (2013). Chapter 5.3: Sampling In Chapter 5: Cross-cutting Issues.
USGS (2013). Index of /srtm/version2_1/SWBD/SWBDeast.
In: http://dds.cr.usgs.gov/srtm/version2_1/SWBD/SWBDeast/
USGS (2014). EarthExplorer. In http://earthexplorer.usgs.gov/

Van der Hoek, W., Sakthivadivel, R., Renshaw, M., Silver, J.B., Birley, M.H., & Konradsen, F. (2001). *Alternate wet/dry irrigation in rice cultivation: a practical way to save water and control malaria and Japanese encephalitis?* : International Water Management Institute (IWMI).
Vanwambeke, S., Lambin, E., Eichhorn, M., Flasse, S., Harbach, R., Oskam, L., Somboon, P., Beers, S., Benthem, B.B., Walton, C., & Butlin, R. (2007). Impact of Land-use Change on Dengue and Malaria in Northern Thailand. *EcoHealth, 4*, 37-51
Varga, C., Pearl, D., McEwen, S., Sargeant, J., Pollari, F., & Guerin, M. (2013). Evaluating area-level spatial clustering of Salmonella Enteritidis infections and their socioeconomic determinants in the greater Toronto area, Ontario, Canada (2007 - 2009): a retrospective population-based ecological study. *BMC Public Health, 13*, 1078
Walton, W. (2003). *Managing mosquitoes in surface-flow constructed treatment wetlands*. UCANR Publications
Wen, T.-H., Lin, N.H., Lin, C.-H., King, C.-C., & Su, M.-D. (2006). Spatial mapping of temporal risk characteristics to improve environmental health risk identification: A case study of a dengue epidemic in Taiwan. *Science of The Total Environment, 367*, 631-640
Wikimedia (2013). Anopheles Range Map. In
Zhang, J., Pennington, D.D., & Liu, X. (2007). GBD-Explorer: extending open source java GIS for exploring ecoregion-based biodiversity data. *Ecological Informatics, 2*, 94-102

Appendix A – Background on Malaria

A-1 Malaria in Africa

In Sub-Saharan Africa, malaria has been a serious problem for a long time and is third in rank in deaths for the African population after pneumococcal acute respiratory infections and tuberculosis. Approximately 8.2 million cases of malaria are reported in Kenya every year out of a population of 30 million. Children younger than five years old and pregnant women are exposed to the highest risk. Every day, 72 children are dying due to a malaria infection. (Opiyo et al. 2007; World Health Organisation 1996, 2012) The factors that probably caused the rapid spread of malaria in Africa are: resistance of the malaria parasite to antimalarial drugs, such as chloroquine, the other quinolines and ACTs, and population movements of people who are not immune to malaria (Dondorp et al. 2009; Gouagna et al. 2004; Sauerwein et al. 2011). Other reasons for malaria still existing in Africa are that health care is not always accessible, education for the local inhabitants about malaria is not provided and local house constructions are not mosquito proof. (Lindsay et al. 2003; Mubyazi et al. 2008b)

A-2 Malaria and people

The type of malaria which is most common in Africa is caused by a parasite named *Plasmodium falciparum*, which causes the most deadly type of malaria. Malaria is transmitted to humans by means of a bite of a female mosquito during a blood meal to develop eggs. In Africa, mosquitoes of the *Anopheles gambiae s.l.* complex are responsible for most transmission. However, also *Anopheles funestus* and *Anopheles arabiensis* are common in Sub-Saharan Africa. (Lindsay et al. 1998; Organisation 2012) The parasites, in this stage called sporozoites, enter the human body after the bite and the sporozoites are released into the bloodstream from where they enter liver cells. After their multiplication, they cause the refraction of the liver and the parasite, merozoites in this stage, are released into the bloodstream where they invade the blood cells. When they are entering the blood cells they are asexual. However, some merozoites transform into gametocytes, which can be taken up by another susceptible female mosquito during a blood meal. (Sauerwein et al. 2011) This process is displayed and described in Figure 34.

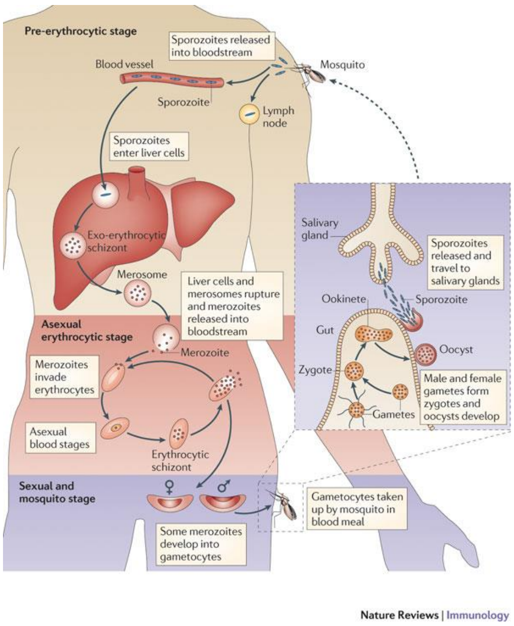


Figure 34: Process of sporozoites entering the human body and finally the infection of another vector (Sauerwein et al. 2011).

A-3 Malaria vectors in Africa

In Africa, the female mosquitoes that are most commonly infected with malaria are of the mosquito complex *Anopheles gambiae*. (Lindsay et al. 1998; Organisation 2012). This complex is widely distributed over Africa (Wikimedia 2013). Temperature has been identified as a key factor distinguishing the distribution of the complex. However, also rainfall and the relation between urban and rural areas are important factors. Furthermore, the temporal factor in terms of seasonality is of importance (Blackwell and Johnson 2000; Evans 1938; Roberts 2000; de Souza et al. 2010).

The mosquito has four life stages. First mosquitoes mate, than the first life stage is the egg, which should be directly laid on a water surface. In the second life stage the egg develops into a larva, which feeds on algae, bacteria and other micro-organisms in the water surface micro-layer and only dives below the water surface when disturbed. The water should have a warmer daytime temperature. After the larval stage, the mosquito transforms into a pupa. A pupa is comma shaped and it should come to the water surface to breathe. Both larvae and pupa can be found in small, sunlit (warm), temporary and turbid pools which are created by human or animal activity. After two or three days the pupa transforms into an adult mosquito. (Evans 1938; Meijerink et al. 2000; Minakawa et al. 2004)

The male adults survive on nectar, the female adults both on nectar and blood meals, which are needed for the development of eggs. (Evans 1938; Meijerink et al. 2000; Minakawa et al. 2004) Since female Anopheles mosquitoes are the disease vectors for malaria, they can become infected with the malaria parasite during this blood meal and transmit it to another person during the next blood meal, see Figure 32 (Sauerwein et al. 2011).

A-4 Symptoms & treatment

As a person is infected with the malaria type *Plasmodium falciparum* via a mosquito bite, the person usually experiences the symptoms 10 to 14 days later. This period is referred to as an incubation period. Those symptoms are observed when red blood cells lyse and release merozoites (Sauerwein et al. 2011). The symptoms of malaria include headache, joint pain, chills, vomiting and diarrhoea. Life-threatening symptoms are not common and quick response is only needed if the infected person suffers from acute malaria. For *P. falciparum*, the symptoms are more serious in nature, since small blood vessels of vital organs can get blocked, causing life-threatening organ failure. (Sauerwein et al. 2011)

Chloroquine was used for a long period in order to treat people with malaria. However, the parasite has become immune to chloroquine. Therefore, artemisinin combination therapies (ACTs) are now recommended by the World Health Organization (WHO). (Gouagna et al. 2004; Sauerwein et al. 2011) A problem that is recently occurring is that the parasite also becomes immune for those ACTs. This is shown in a research which is done in West Cambodia and North-Western Thailand (Dondorp et al. 2009). So malaria control research as well as curative research is vital.

A-5 Preventive measures

Since malaria is influencing the health condition of infected people and the malaria parasite is becoming resistant to all variants of drugs, infection should rather be prevented instead of treated afterwards and avoid becoming dependent on drugs (Gouagna et al. 2004; Sauerwein et al. 2011; Dondorp et al. 2009; (Organisation 1996, 2012)). The most common measure that is proven to be effective against bites of malaria vectors is using a long lasting insecticide-treated (LLIN) bed net (Alonso et al. 1991). However, also IRS and proper case management are also measures of the WHO. In some instances, the most vulnerable population: young children of less than 5 years old and pregnant women will receive intermittent preventive treatment. However, the policies for those treatments vary between countries. A combination of treatments leads to even better protection of people at risk. (Kleinschmidt et al. 2009; Mubyazi et al. 2008a)

A-6 The SolarMal project

Taking into account the preventive measures mentioned in A-5 and emerging resistance to drugs and insecticides, the scientific community need to find other ways to prevent and control malaria. Within the SolarMal project, the aim is to eliminate malaria from Rusinga Island, Western Kenya (Figure 35), by means of a novel malaria control measure in combination with the already existing national malaria control strategies of long-lasting nets and case management. The project will provide odour baited traps (OBTs) to all households to achieve mass trapping of malaria vectors.

In order to let the traps do their work, the traps should attract mosquitoes. Therefore, odour bait was developed and formulated in such a way that it mimics humans. Furthermore, a synthetic CO_2 replacement is used in the traps. An advantage of the odour is that it is reproducible, objective and has low development costs. (Mukabana et al. 2012; Okumu et al. 2010) This developed odour will be used in the newly developed mosquito trap, the Suna trap, which will be installed on Rusinga Island during the SolarMal interventions (Figure 36).

After a period of 2 years, SolarMal's hypothesis is that by lowering mosquito vector populations, the number of potentially infective bites a person receives will be reduced and this will eventually lead to malaria elimination on the island. To measure progress in the elimination of malaria, the demographic situation is constantly monitored in the field by gathering data on health and demography using a Health and Demographic Surveillance System (HDSS). This is vital to attribute effects to the intervention, geography and individuals. Also vector populations are continuously monitored, detecting seasonal and long-term changes in mosquito population density and species composition. Monitoring malaria is another important component of the project and parasite prevalence surveys are conducted continuously over the island, and RDT (Rapid Diagnostic Test) positive cases are treated with ACT or referred to local health services. (Hiscox et al. 2012a; Hiscox et al. 2012b)

For the SolarMal project, the nationwide adapted strategies of LLINs and case management are used together with mass trapping of the mosquito vectors. (Hiscox et al. 2012a; Hiscox et al. 2012b)

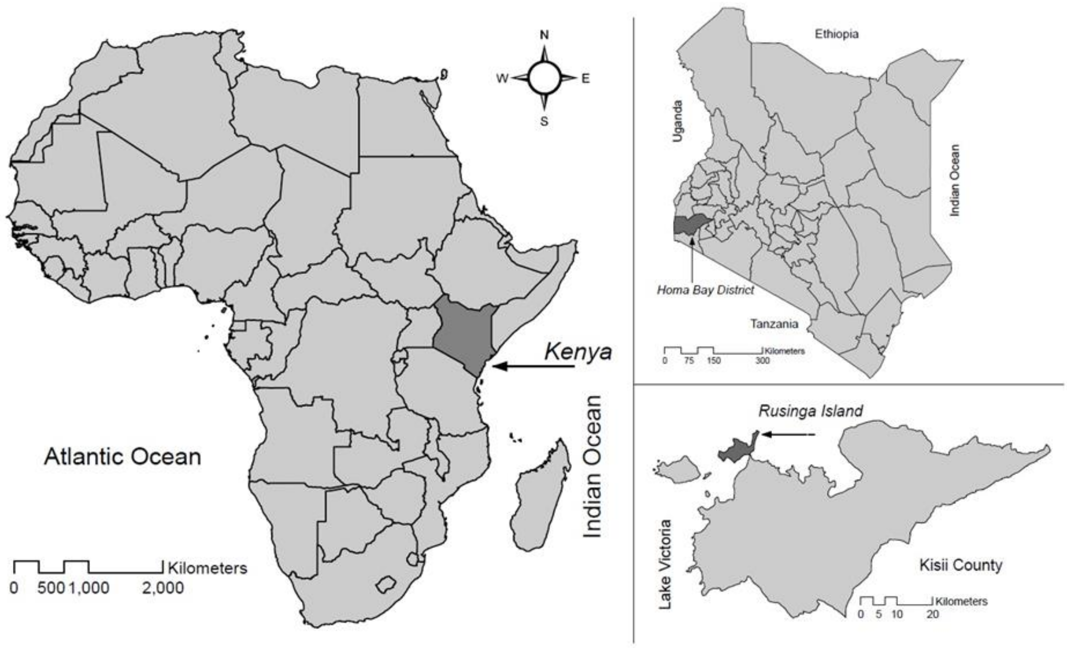


Figure 35: Rusinga Island, Kenya: the study area of the SolarMal project. 1) Africa with its countries (Kenya), 2 Kenya with its counties (Homa Bay district), 3) Homa bay county with Rusinga island (34 7' 5" E , 0 26'46" S / 34 14' 8" E, 0 20' 35" S).

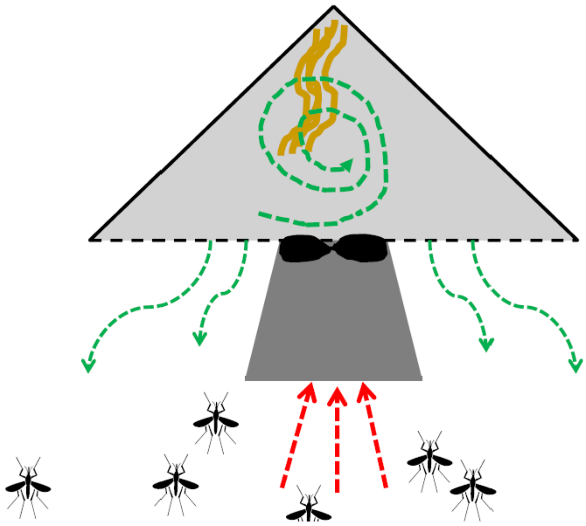


Figure 36: The Suna trap, an odour baited solar system (Hiscox et al. 2012a; Mukabana et al. 2012; Okumu et al. 2010).

The traps installed in- and outside the houses on Rusinga Island are solar-powered and the battery and wires are contained in a box. In addition to the OBT, residents receive a connection for charging their mobile phone. Furthermore, two LED light bulbs are installed in each household, thus providing a source of indoor lighting and reducing dependency on biofuels for lighting. All devices are powered by the solar panel which is installed on their roof. The solar panels and the traps are installed for free, since it is tried to let the SolarMal project be a holistic project that also develops other things for the people, like electricity and light. (Hiscox et al. 2012a; Hiscox et al. 2012b) For an overview, see Figure 37.

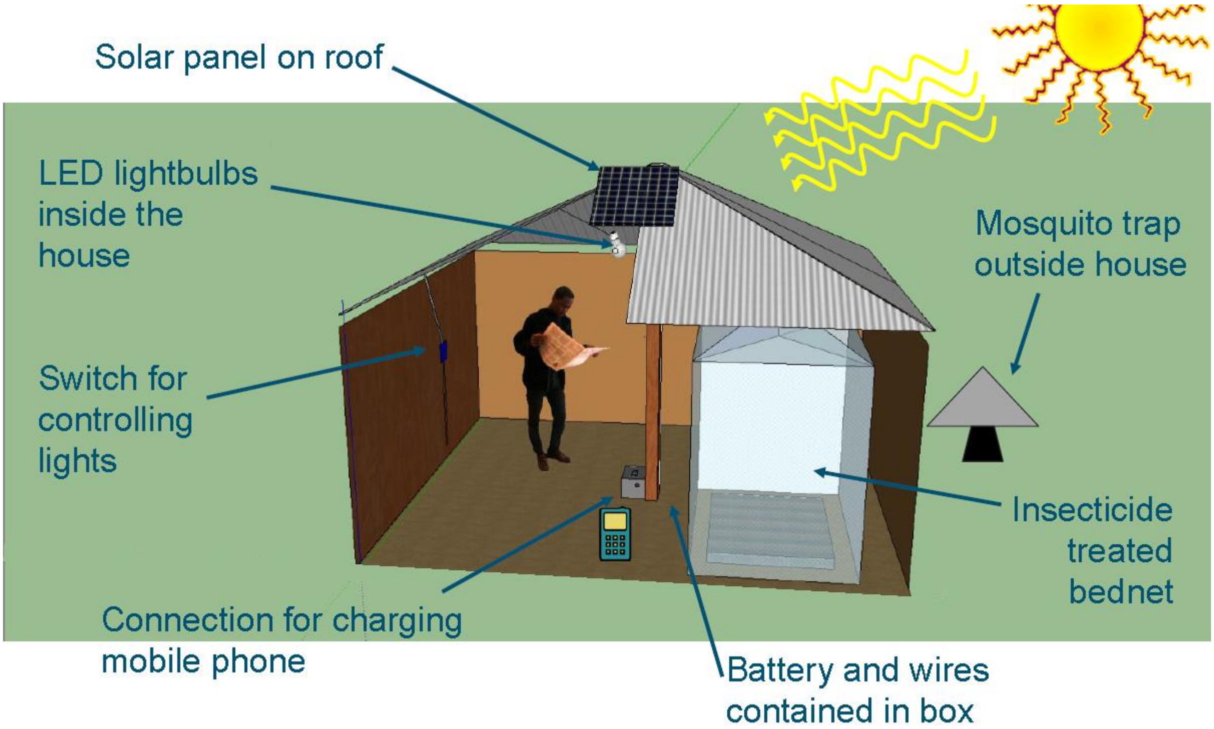


Figure 37: The different types of supplies provided during the SolarMal project (Hiscox et al. 2012a).

The SolarMal intervention done on the island will provide experimental outcomes, which can be divided into the health and demographic surveillance system (HDSS), parasitology, entomology and sociology. The roll out of the traps started in mid-2013 and should be finished within 24 months. A hierarchical study design is used to measure an effect of the intervention by randomly selected clusters of 50-51 households. Those households are linked to the neighbouring clusters that are similarly provided with a mosquito control tool within the same week. 9 of these clusters together form a metacluster and the island is divided into 9 metaclusters (a total of 81 clusters). At this time (July 2013), preliminary testing of the odour baited mosquito trapping system is already done and the roll-out of SolarMal interventions already started. It appears that the system met all requirements of the inhabitants and was efficient in providing enough power for all the electrical devices. (Hiscox et al. 2012b)

The SolarMal project as described above aims to let the local inhabitants participate in the research and in this way inform them about the health risk which malaria causes. Furthermore, the population is involved in decisions, ideas of implementation, helping in data collection and SMoT instalment. The communication is done by means of stakeholder groups, such as churches, women groups, beaches etc. and an official CAB (community advisory board) which represents the island. Since African people have their own methods for disease management, this public awareness is needed in order to increase awareness and understanding of the treatments being used. Furthermore, the SolarMal project is needed in order to get more background in epidemiology, socio-economic impact of the disease and the dynamics of the vector population in this region of Kenya. By means of good communication and placing this SolarMal project into a national context, the results can probably be translated into practical applications in other risk areas. (Nchinda 1998)

Appendix B – Maps

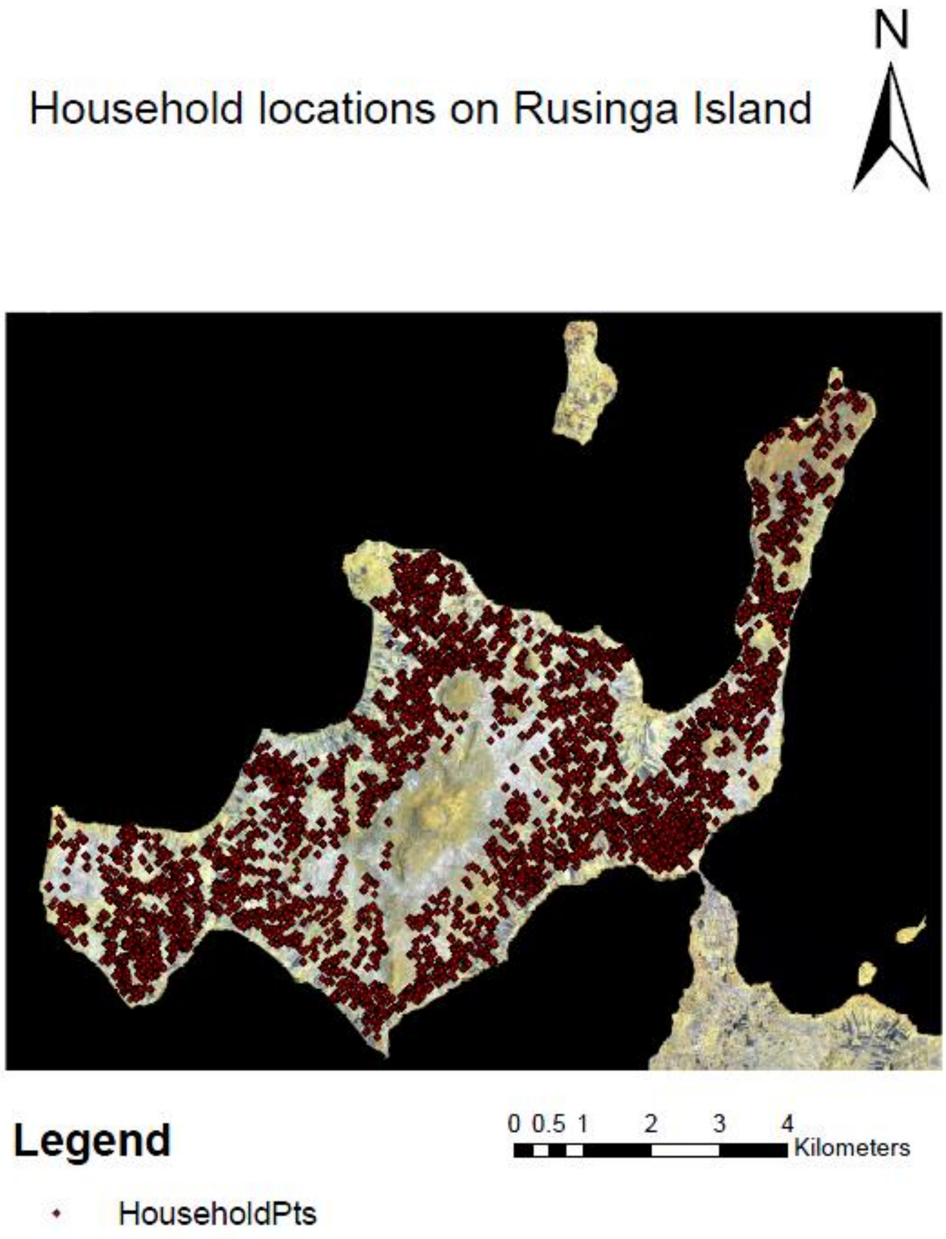


Figure 38: Distribution of households on Rusinga Island, Kenya, based on data gathered in the SolarMal project.

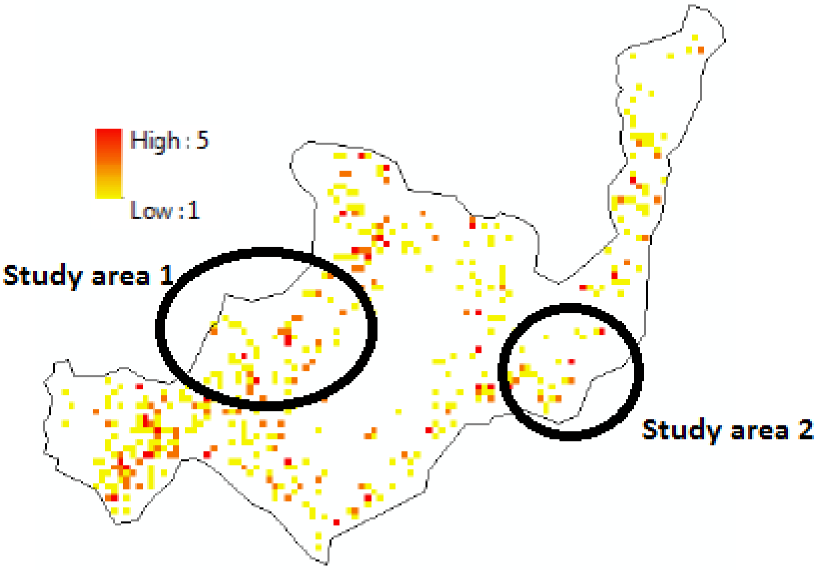


Figure 39: Distribution of farmers on Rusinga Island: in both areas farming practices are performed (Homan et al. Unpublished).

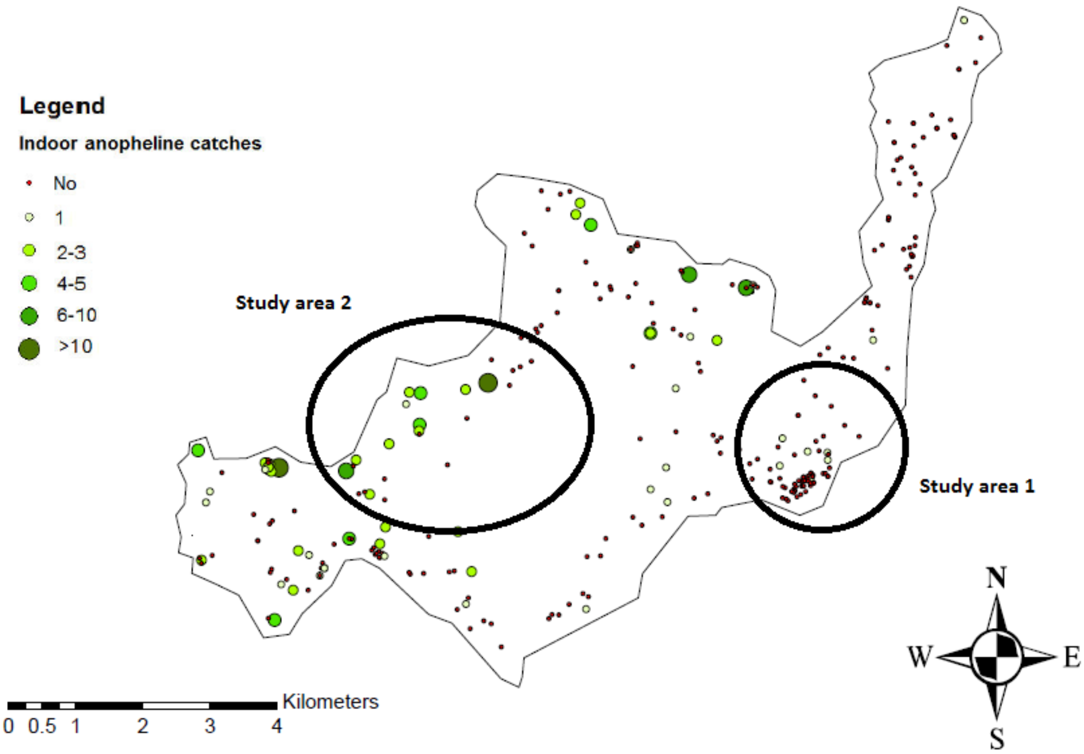


Figure 40: Amount of indoor Anopheline catches: a lot of catches in study area 1, (almost) no catches in study area 2 (Homan et al. Unpublished).

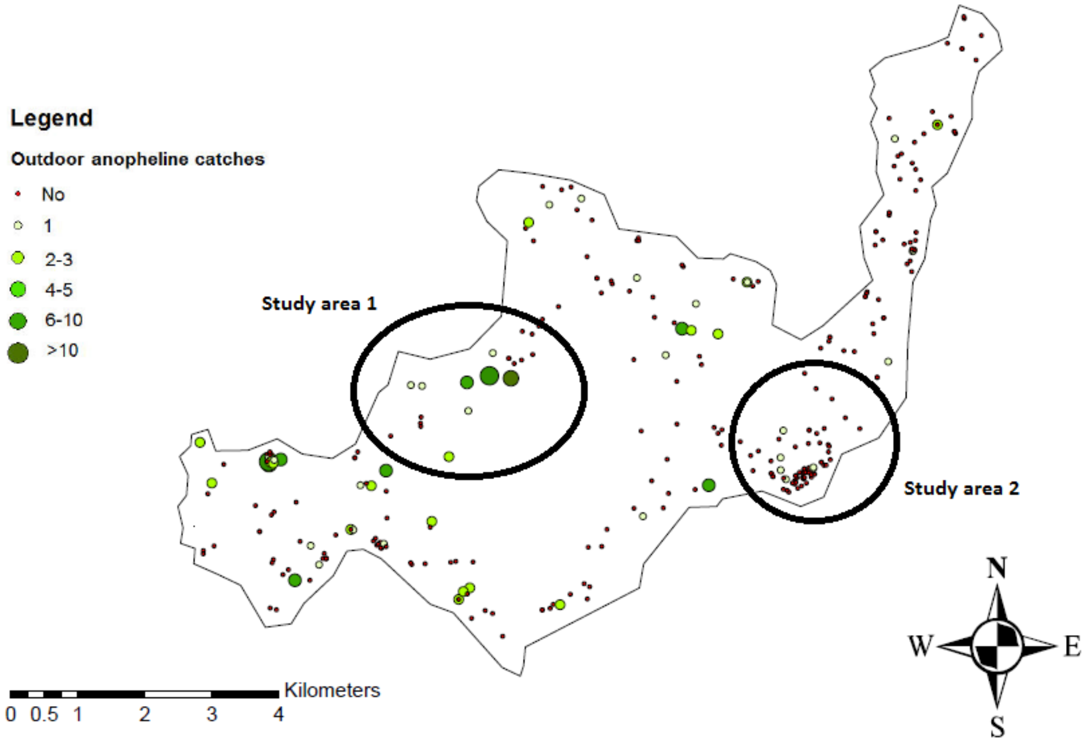
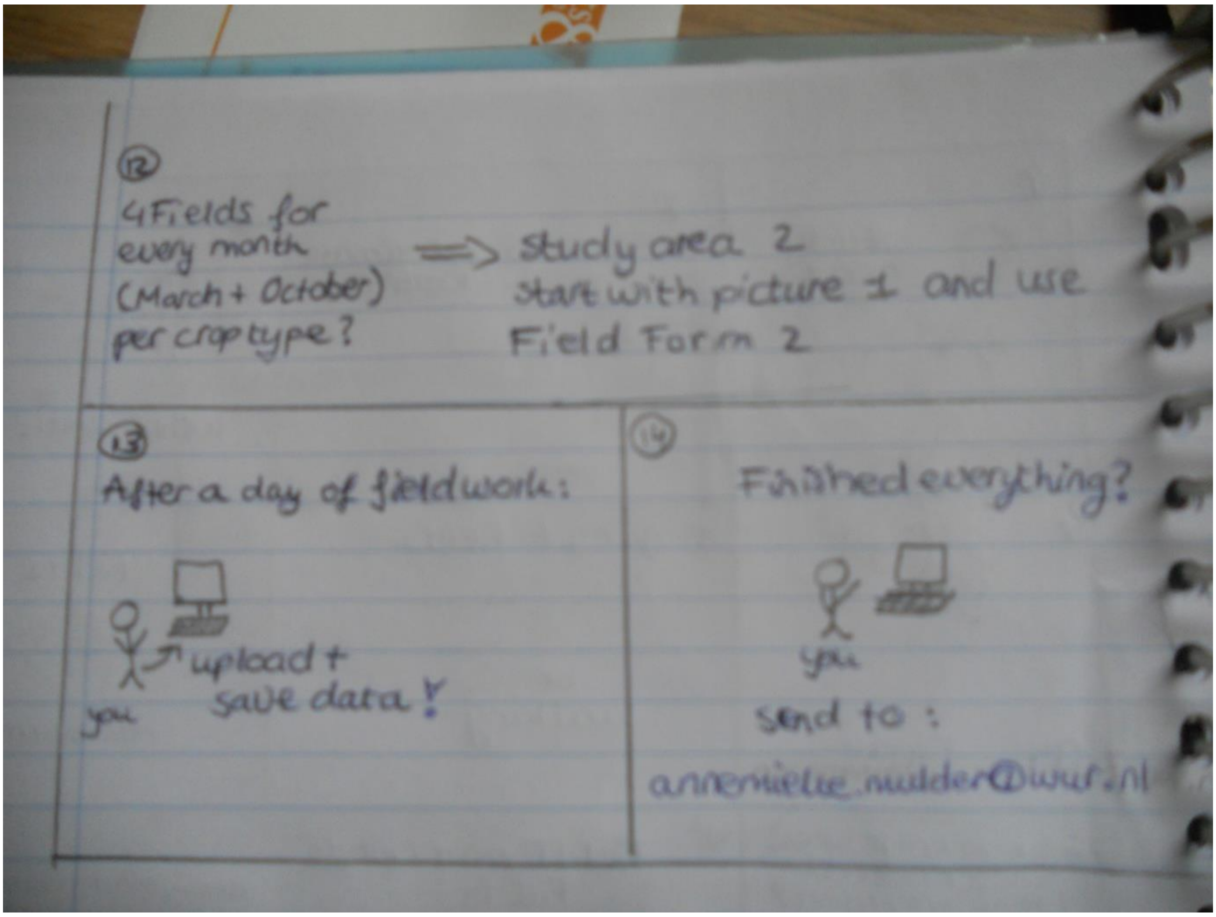


Figure 41: Amount of outdoor Anopheline catches on Rusinga Island: a lot of catches in study area 1, (almost) no catches in study area 2 (Homan et al. Unpublished).

Appendix C - Sketch image field work


⑫
4 Fields for
every month
(March + October)
per crop type?
⟹ Study area 2
Start with picture 1 and use
Field Form 2
⑬
After a day of fieldwork:
you
upload +
save data!
⑭
Finished everything?
you
send to:
annemieke.mulder@wur.nl

Appendix D - Toolbox description

The toolbox: IA_Thesis_Annemieke_SolarMal_"savingdate".tbx is used for doing the different pre-processing and analysis steps which are necessary for this particular study. This toolbox contains four different toolsets:

1. a_preprocessing_March
2. b_ClassificationMarch_preparation_fieldwork
3. d_Landuse_classification_and_validation_March
4. e_preprocessing_classification_validation_October
5. f_relate_mosquitodata_to_landuse
6. xx_proposal
7. zz_backup

Each of the above toolsets contains different models which are built to finally lead to viable results and answers to the proposed research questions of chapter 1.2. Underneath, the models will be discussed separately per toolset to explain certain choices which are made during the development of the models.

D-1 a_preprocessing_March

This toolset contains the following ten models, which were needed in order to pre-process the images. The choices made within those models will be explained here:

1. *a01ProjectAsterDEM*

   In this model, the given AsterDEM is being projected to the coordinate system which will be used for all the analyses done for this study: WGS_1984_UTM_Zone_36S with an output cell size of 30 as is induced from the raster information of the original DEM.

2. *a02OrthorectificationImages*

   The next step was to ortho-rectify the image which is done by means of the DEM which is created in the previous model. The raster was copied in order to preserve the previous adjustments. Both the panchromatic and the multispectral image were ortho-rectified by means of the provided DEM.

3. *a03DefineOverallStudyArea*

   Now that the images are in the correct coordinate system and are ortho-rectified, Rusinga Island should be selected from the images. In order to do so, the water bodies of Lake Victoria should be removed from the image to reduce the calculation time in future calculations. The SRTM water body with the code: e034s01f.shp is used for this purpose (USGS 2013). This file needs to be projected in the same coordinate system (WGS_1984_UTM_Zone_36S) first. After this, the polygon dataset needs to be transformed into a raster dataset. The new dataset gets the same cell size as the DEM of model a02. Furthermore, the default settings of the tool are being used. The water bodies are being set to null by use of the isnull tool. Those water bodies are being clipped to the same extent as the DEM dataset and by the tool set null, the waterbodies are being removed. In order to get rid of the small islands in the neighbourhood of Rusinga Island, the dataset is transformed into polygons again in order to be able to select only Rusinga Island. When this is done, Rusinga is copied to preserve this file and transformed into a raster file again since all analyses will be done with raster files.

4. *a04CreateRasterLayerForMultispectralBands*
   Now that the extend of Rusinga Island is created, the multispectral image should be adapted towards this shape. Before this can be done, the different bands (blue = band 1, green = band 2, red = band 3 and NIR = band 4) should be extracted from the image and copied to preserve the images of the different bands.

5. *a05DefineStudyAreaInMultispectralImage*
   To define the study area per band of the multispectral image, the raster calculator is being used. For the blue band, the following condition was used: Con("a03_Rusinga_final" == 1,"a04_blueband"). "a03_Rusinga_final" was the final extent of Rusinga created in model a03. The same condition is being used for the other bands, using the other raster layers created in model a04 of the correct bands.

   In order to get the correct reflection factor, the different values in the images of the different bands should be multiplied by a specific factor given in the metadata of the image. This is done by means of the raster calculator and for the blue band the condition was: "a05_Rusinga_Blueband" * 1.604120000000000e-02. The same condition is being used for the other bands, only with other factors:

Green band	=	1.438470000000000e-02
Red band	=	1.267350000000000e-02
NIR band	=	1.542420000000000e-02

   After those corrections, the bands could be combined into one image again by means of the tool ‘composite bands’.

6. *a07ImportAndProjectHouseholdData*
   The data gathered during the SolarMal project are gathered in an Excel file. In order to be able to use those date it should be imported in ArcGIS and projected in the correct coordinate system: WGS_1984_UTM_Zone_36S.

7. *a08DefineStudyArea1AndStudyArea2*
   In this study, two smaller study areas are defined: study area 1 on West Rusinga and study area 2 on East Rusinga. Since the multispectral image will be used for all the analyses, this image is clipped to the clip boxes of the two study areas, giving the two pre-processed images of the study areas.

8. *a10StudyArea1AndStudyArea2BasedOnMosquitoHotspotAnalysis*
   In this model, the hotspot/coldspot points where selected inside the old study areas. A buffer was created around those points of 1km (maximum flight distance of mosquitoes). Those buffers were dissolved and together formed the outside borders of the new study areas based on the hotspot anaylsis.

9. *XXXa06DeletingWaterbodiesFromPan*
   In this model, also the water bodies around the panchromatic image are deleted and Rusinga island is remaining together with the other islands. This was done when exploring both the panchromatic and the multispectral image.

10. *XXXa09PansharpeningHR*
    By the tool pan-sharpening, the resolution of the multispectral image is adapted to the resolution of the panchromatic image, which leads to a sharper image. The pan-sharpening type that is used is IHS. This type is chosen, because it uses Intensity, Hue, and Saturation colour space for data fusion. The other types were also investigated, but IHS resulted in the sharpest/detailed image. The pan-sharpened image was not used for the analyses phase, since the calculation time would become too long with such a high-resolution image.

D-2 b_ClassificationMarch_preparation_fieldwork

This toolset contains the following fourteen models, which focus on investigating the spread in land use on Rusinga Island and the preparation of the fieldwork maps. The choices made within those models will be explained here:

1. *b01ImportFieldDataTobias*
   In order to get a general idea of which types of agricultural land use there are on Rusinga Island, a fieldworker gathered some data in October 2013 via Tobias Homan. Those data are imported by means of this model and projected to the correct coordinate system: WGS_1984_UTM_Zone_36S.

2. *b02ExtractbandsStudyArea1&2forHistogramCreation*
   Exploring the data within the images before the classification is important, since the assumption in classification analysis is that the band data and the training sample data follow a normal distribution. Therefore, the different bands of the multispectral image created in model a05 are extracted from the image again. Per band, the distribution of the data is checked by using the histogram tool on the Training Sample Manager and therefore is not included in this model. The results are described in Chapter 3. (Esri 2013m)

3. *b03PrincipalComponentAnalysisRusinga*
   In this model, the bands created in model a05 are used as input for the principal component analysis (PCA) of those bands to remove correlation among the bands which is in fact still part of some pre-processing. The created text file is included in Chapter 3. Since the amount of information appeared to be limited in the NIR band, the NIR band was not included in the new multispectral image composite which was created after the PCA results had been evaluated.

4. *b04PCAStudyArea1*
   Model b03 contains the description of what is happening in this model as well. The only difference is that the focus here is on the first study area: Study Area 1, while the focus of the model b03 was on Rusinga Island as a total.

5. *b05PCAStudyArea2*
   Model b03 contains the description of what is happening in this model as well. The only difference is that the focus here is on the second study area: Study Area 2, while the focus of the model b03 was on Rusinga Island as a total.

6. *b06UnsupervisedClassificationRusinga*
   According to the description as given by (Esri 2013m), the unsupervised classification is performed for Rusinga Island as a whole in order to get a general idea of the distribution of the different land use types on the island. To perform the unsupervised classification, the “iso cluster unsupervised classification” tool is used. For this, the different bands needed to be extracted from the combined multispectral image which was created in model b03. The minimum class size for this classification is 30. This is based on the fact that the minimum class size should be 10x larger than the number of layers (3 bands in this case) (Esri 2013p).

   The number of classes is fixed on 14. This is based on the data which was gathered in the field in October 2013 by the fieldworker, study area had the highest variability and in total 14 classes. It contains both agricultural land use types and other types of land use to make the classification more complete. For the different land use types found on the island by the fieldworker, see Appendix E.

   In order to make the classification more smooth (post-classification processing), the majority filter, boundary clean and region group tool are being used. The region group tool is used to remove small regions together with the set null tool which removes areas which are smaller than 10 pixels. In the end, the nibble tool is used to dissolve the small areas.

7. *b07UnsupervisedClassificationStudyArea1*
   The same description counts for this model as is given for model b06. The only difference is that the unsupervised classification is only done for study area 1 instead of Rusinga Island as a whole.

8. *b08UnsupervisedClassifcationStudyArea2*
   The same description counts for this model as is given for model b06. The only difference is that the unsupervised classification is only done for study area 2 instead of Rusinga Island as a whole.

9. *b09SupervisedClassificationRusinga*
   For this classification, the first steps of the unsupervised classification done in model b02 and b03 can be used: the histogram creation and PCA. Than some steps follow which cannot be displayed in the model. For collecting the training samples the image classification toolbar is used. The polygon drawing is used for creating the training samples and the new class is shown in the training sample manager in which the name of the class can also be given. If this is finished, the training samples can be evaluated by means of the statistics tool and scatterplots tool (Figure 42).

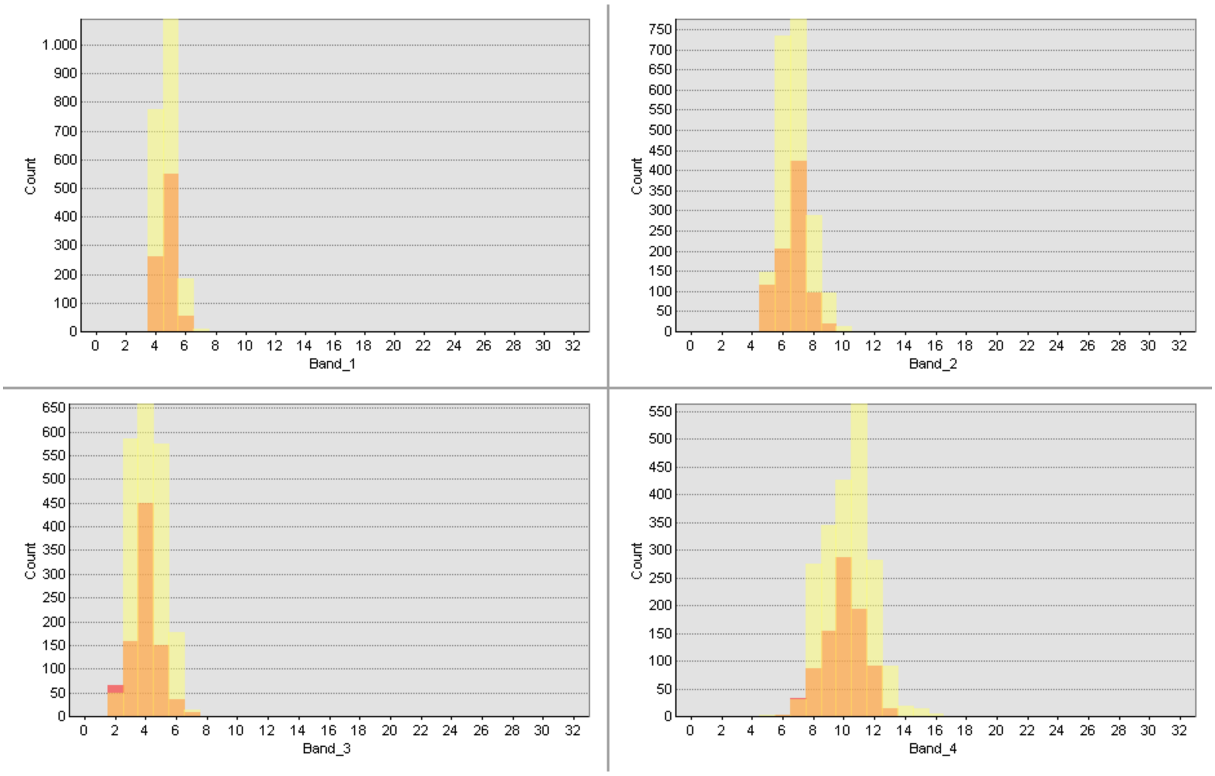


Figure 42: An example scatter plot of two training areas (both defined as pasture) which show overlap and were merged before the classification was done.

If there is not enough separation between the classes, there can be decided to adapt the classes and for example merge them. When the classification is good enough, the signature file is created. In order to investigate the signature file, the dendrogram tool is used. If the separation between the classes is too large, this can be adapted again, but it depends on the situation. In this case, all classes are preserved although they are very close to each other. Otherwise a too small amount of classes remains for this study.

After the creation of the signature file, the real classification can be done by means of the maximum likelihood classification tool. After this, the post-classification processing should be done. In order to make the classification more smooth, the majority filter, boundary clean and region group tool are being used. The region group tool is used to remove small regions together with the set null tool which removes areas which are smaller than 10 pixels. In the end, the nibble tool is used to dissolve the small areas.

10. *b10SupervisedClassificationSA1*
The same description counts for this model as is given for model b10. The only difference is that the supervised classification is only done for study area 1 instead of Rusinga Island as a whole.

11. *b11SupervisedClassificationSA2*
The same description counts for this model as is given for model b10. The only difference is that the supervised classification is only done for study area 2 instead of Rusinga Island as a whole.

12. *b12RandomPointgenerationStudyArea1&2*
    Since Rusinga Island is never mapped in terms of land use, both study areas should be covered in a way that every land use type on the island is covered more or less equally. Systematic random sampling is chosen for this, since it is spatially well balanced. Therefore, one random point per study area is needed to be generated within the extent of those study areas which is done in this model. In order to get all the sampling points in the area, the points should be placed from the first point with an equal distance between them. For study area 1 this is 150 m. which gives in total 210 sampling points and for study area 2 this is 75 m. which gives in total 270 sampling points. Since study area 1 was larger than study area 2, there was chosen to sample at a larger distance between the points in order to be able to cover the whole area within the time limits of the fieldwork (2 months). The points were placed by means of the drawing tools in ArcGIS and exporting the points as a layer file.

13. *b13TransformSamplePointsToKML*
    The layer files of the sampling points of both study areas were transformed into KML files in this model. This was necessary since Google Earth was used as the operating GPS system in the field and the files needed to be uploaded on the tablet into Google Earth.

14. *b14ToFinishLast4LinesOfFieldworkSA1*
    Unfortunately, the fieldwork could not be finished due to illness, therefore, someone else finished it. Only four lines of study area 1 needed to be done still, therefore those lines were exported in a map again, so that the other data points were not visible in the map anymore. This was necessary to have a better overview within the map of what still was needed to be done.

D-3 d_Landuse_classification_and_validation_March

This toolset contains the following eighteen models, which focus on the land use classification improvement after the fieldwork and the validation of those maps. The choices made within those models will be explained here:

1. *d01ImportFieldworkdataSA2Classification*
   The fieldwork data of study area 2 gathered in February and March 2014 is imported from excel to ArcGIS and projected to the correct coordinate system (WGS_1984_UTM_Zone_36S).

2. *d02ImportFieldworkdataSA1Classification*
   The same is done as in model d01, the only difference is that it is done for study area 1 instead of study area 2.

3. *d03SupervisedLandUseClassificationSA2*
   The same is done as in model b09, only with different input: fieldwork data of study area 2 of the fieldwork done in February and March 2014. The count is determined as <5 in this case, otherwise no classes are remaining and classes are needed to be able to find a relationship between mosquito vector abundance and agricultural land use types.

4. *d04SupervisedLandUseClassificationSA1*
   The same is done as for d04, with input of study area 1.

5. *d05ImportValidationDataFieldworkSA1March*
   To be able to validate the classification maps, validation data of study area 1 is necessary. This data is half the data of February and March 2014 which is uploaded from Excel to ArcGIS and projected in the correct coordinate system (WGS_1984_UTM_Zone_36S).

6. *d06ImportValidationDataFieldworkSA2March*
   The same is done as in model d05, only for the data of study area 2.

7. *d07ValidationPointsToRasterSA2March*
   To be able to compare the classification dataset with the validation data, both datasets should be in the same format: raster. Therefore, the validation points are transformed from vector to raster for study area 2 in this model.

8. *d08CombineClassificationAndValidationDataforErrorMatrix*
   Now that the datasets are in the same format, they are combined and a pivottable is created, which is the base of the error matrix (which is created in Excel) of study area 2.

9. *d09ValidationPointsToRasterSA1March*
   Same as model d07, only than for study area 1.

10. *d10CombineClassificationAndValidationDataForErrorMatrix*
    Same as model d08, only than for study area 1.

11. *d11UnsupervisedClassificationSA1AfterFieldwork*
    This model was created since first was thought that the accuracy of study area 1 was too low. Therefore, an unsupervised classification was done for this study area and indeed the accuracy increased. However, according to literature, the accuracy was not that bad, so there was decided to use the original classification result.

12. *d12CombineUnsupervisedClassAndValDataForErrorMatrix*
    The same as model d08, only than for the unsupervised classification of study area 2.

13. *d13SupervisedLandUseClassificationRusinga*
    The same as model d03, only then for Rusinga as a whole.

14. *d14ImportValidationDataFieldworkRusingaMarch*
    The same as model d01, only then for validation data of Rusinga as a whole.

15. *d15ValidationPointsToRasterRusingaMarch*
    The same as model d07, only then for Rusinga as a whole.

16. *d16CombineClassAndValDataForErrorMatrix*
    The same as model d08, only then for Rusinga as a whole.

17. *d17UnsupervisedClassificationSA2AfterFieldwork*
    The same as model d11, only this time for study area 2. Since the accuracy of the classification of this area was too low and it improved with this unsupervised classification, there was decided to use this classification result for further analyses.

18. *d18CombineUnsupervisedClassAndValDataForErrorMatrixSA2*
    The same as model d12, only now for study area 2.

D-4 e_preprocessing_classification_validation_October

This toolset contains the following 23 models, which focus on preprocessing, investigating the spread in land use on Rusinga Island and the improvement and validation of the classification maps after the fieldwork. The choices made within those models will be explained here:

1. *e01ProjectLandsatImage*
   Same as model a01, only for Landsat (October) instead of Quickbird (March) image.

2. *e02SelectStudyAreaRusingaIsland*
   Same as model a05, only for Landsat instead of Quickbird image.

3. *e03SelectStudyArea1AndStudyArea2*
   Same as model a10, only for Landsat instead of Quickbird image.

4. *e04SeparateBandsForHistogramCreation*
   Same as model b02, only for Landsat instead of Quickbird image.

5. *e05PCAStudyAreas*
   Same as model b03-b05, PCA's are combined now within 1 model, only now for Landsat instead of Quickbird image.

6. *e06UnsupervisedClassificationRusinga*
   Same as model b06, only for Landsat instead of Quickbird image.

7. *e07UnsupervisedClassificationSA1*
   Same as model b07, only for Landsat instead of Quickbird image.

8. *e08UnsupervisedClassificationSA2*
   Same as model b08, only for Landsat instead of Quickbird image.

9. *e09SupervisedClassificationDTSA1*
   Same as model b10, only for Landsat instead of Quickbird image.

10. *e10SupervisedClassificationDTSA2*
    Same as model b11, only for Landsat instead of Quickbird image.

11. *e11SupervisedClassificationDTRusinga*
    Same as model b09, only for Landsat instead of Quickbird image.

12. *e12SupervisedClassificationAfterFieldworkSA2October*
    Combines the models d01 and d03 in one model for Landsat instead of Quickbird.

13. *e13SupervisedClassificationAfterFieldworkSA1October*
    Combines the models d02 and d04 in one model for Landsat instead of Quickbird.
14. *e14SupervisedClassificationAfterFieldworkRusinga*
    The same as model d13, however, the analysis is done for Landsat instead of Quickbird.

15. *e15ImportValidationDataFieldworkSA1October*
    The same as model d05, however, the analysis is done for Landsat instead of Quickbird.

16. *e16ImportValidationDataFieldworkSA2October*
    The same as model d06, however, the analysis is done for Landsat instead of Quickbird.

17. *e17ValidationPointsToRasterSA2October*
    The same as model d07, however, the analysis is done for Landsat instead of Quickbird.

18. *e18ValidationPointsToRasterSA1October*
    The same as done in model d09, however, the analysis is done for Landsat instead of Quickbird.

19. *e19CombineClassificationAndValidationDataforErrorMatrix*
    The same as done in model d08, however, the analysis is done for Landsat instead of Quickbird.

20. *e20CombineClassificationAndValidationDataForErrorMatrix*
    The same as done in model d10, however, the analysis is done for Landsat instead of Quickbird.

21. *e21ImportValidationDataFieldworkRusingaOctober*
    Same as model d14, only the analysis is done for Landsat instead of Quickbird.

22. *e22ValidationPointsToRasterRusingaOctober*
    Same as model d15, only the analysis is done for Landsat instead of Quickbird.

23. *e23CombineClassAndValDataForErrorMatrixRusinga*
    Same as model d16, only the analysis is done for Landsat instead of Quickbird.

D-5 f_relate_mosquitodata_to_landuse

This toolset contains the following 30 models, which focus on relating the land use classification maps and thus the different land use types to the vector abundance data. The choices made within those models will be explained here:

1. *f01ImportMosquitoData*
   This model is made to import the data of the vector abundance on the island. This data is projected in the correct coordinate system (WGS_1984_UTM_Zone_36S).

2. *f02SelectMosquitoDataSA2*
   This model selects the mosquito data that is falling within study area 2 and the data is projected in the correct coordinate system (WGS_1984_UTM_Zone_36S).

3. *f03InterpolateMosquitodataSA2*
   This model is used to interpolate the mosquito data to the whole of study area 2. This is done by means of IDW (inverse distance weighting) for both *Anopheles gambiae* and *Anopheles funestus* as well as for both of them separately.

4. *f04MergeMosquitoAndLandUseDataSetsForRegressionAnalysisSA2*
   In this model, the mosquito vector abundance data that was interpolated in model f03 is transformed into a raster dataset and merged with the land use data set for the regression analysis that is coming.

5. *f05OLSAllLanduseTypesSA2*
   After the merging of model f04 is done, the regression analysis by means of the Ordinary Least Squares (OLS) tool is performed in this model for all land use types together.

6. *f06landuseSeparatedSA2*
   In this model, all land use types are separated to be able to perform the OLS per land use type to see if there is one land use type that explains the vector abundance best.

7. *f07OLSLandUseTypesSeparatelySA2*
   In this model the OLS is performed for every land use type separately.

8. *f08SelectMosquitoDataSA1*
9. *f09InterpolatemosquitodataSA1*
10. *f10MergeMosquitoAndLandUseDataSetsForRegressionAnalysisSA1*
11. *f11OLSAlllandusetypesSA1*
12. *f12LandUseSeparatedSA1*
13. *f13OLSLandUseTypesSeparately*
    Models f08 - f13 are exactly the same models as the models f02 - f07, however they are performed for a different study area: study area 1 and study area 2 respectively.
14. *f14InterpolateMosquitiodataRusinga*
15. *f15MergeMosquitoAndLandUseDataSetsForRegressionAnalysisRusinga*
16. *f16OLSAlllandusetypesRusinga*
17. *f17landuseSeparatedRusinga*
18. *f17Part2landuseSeparatedRusinga*

19. *f18OLSLandUseTypesSeparatelyRusinga*
    Models f14 – f18 are exactly the same as the models f03 – f07, however they are performed for Rusinga island as a whole instead of study area 2.

20. *f19OctoberMergeMosquitoAndLandUseDataSetsForRegressionAnalysisSA2*
21. *f20OctoberOLSAllLanduseTypesSA2*
22. *f21OctoberlanduseSeparatedSA2*
23. *f22OctoberOLSLandUseTypesSeparatelySA2*
24. *f23OctoberMergeMosquitoAndLandUseDataSetsForRegressionAnalysisSA1*
25. *f24OLSAlllandusetypesSA1October*
26. *f25LandUseSeparatedSA1oct*
27. *f26OctoberOLSLandUseTypesSeparately*
28. *f27MergeMosquitoAndLandUseDataSetsForRegressionAnalysisRusinga*
29. *f28OLSAlllandusetypesRusinga*
30. *f29landuseSeparatedRusingaOctober*
31. *f30OLSLandUseTypesSeparatelyRusingaOctober*

    Those models (model f19 – f30) are exactly the same models as the models f03 – f18, only the interpolation models are not done again since the same vector abundance data is used. Furthermore, the satellite image of October: the Landsat image is used in the analyses and not the Quickbird image.

D-6 xx_proposal

In this toolset, there is only one model included: xxhotspotanalysis. This model was created in order to investigate the distribution of the Anopheles mosquitoes over the island and in this way decide where the two study areas should be located. Since one area should be containing a lot of mosquitoes and the other a few. This is done by first uploading the data from the SolarMal project out of Excel and projecting those data to the correct coordinate system: WGS_1984_UTM_Zone_36S. After which the Hotspot Analysis Tool is being used. The distance method that is used is the Euclidean distance (default setting).

Appendix E - Land use types based on data fieldworker

Land use types Rusinga (March)

1. Tomatoes
2. Maize
3. Sorghum
4. Beans
5. Sukumawiki
6. Sweet potatoes
7. Cow Peas
8. Bare soil
9. Pasture – Grassland used for grazing or area with a mixture of grass and shrubs
10. Trees <10
11. Water
12. Shrubs - Short mature trees < 2m tall
13. Swamp - Presence of emergent aquatic plants
14. Building

Land use types Rusinga (October)

1. Sukumawiki
2. Cow peas
3. Beans
4. Tomatoes
5. Sweet pepper
6. Bare soil
7. Pasture – Grassland used for grazing or area with a mixture of grass and shrubs
8. Trees <10
9. Water
10. Shrubs - Short mature trees < 2m tall
11. Swamp - Presence of emergent aquatic plants
12. Building

Pictures different crops

1. Tomatoes (Pictures taken on Rusinga) – both study area 1 & 2

2. Maize (Google) – Only study area 1

3. Sorghum (Google) – Only study area 1

4. Beans (Pictures taken on Rusinga) – both study area 1 & 2

5. Sukumawiki (Pictures taken at Rusinga) – both study area 1 & 2

6. Sweet potatoes (Google) – Only study area 1

7. Cow Peas (Pictures taken at Rusinga Island) – both study area 1 & 2

8. Sweet pepper (Pictures taken at Rusinga) – only study area 2

Appendix F - Error matrix procedure

In order to obtain a value for the accuracy of the land use classification, the error matrix procedure is used. The error matrix itself is displayed in Figure 43. The calculations behind the procedure are explained in this appendix. (Map & GIS Library 2013)

Error matrix				
PR_SA1_SC_	Rocks/bare soil/building	Shrubs/trees&pasture	Agriculture	Ground truth
Rocks/bare soil/building				
Shrubs/trees&pasture				
Agriculture				
Total				

Figure 43: An example error matrix with which the total accuracy, producer's and user's accuracy could be calculated.

- The **total accuracy** of the classification shows how close the classification result is to reality. 0 means that there is no relation with reality at all, 1 means the land use classification is matching reality completely.
  To obtain the total accuracy of the classification: (sum of the blue cells)/red cell.*
- The **kappa coefficient** is a statistical measure of inter-rater agreement or inter-annotator agreement for qualitative (categorical) items, 0 means a low agreement, 1 a high agreement.
  To obtain the kappa coefficient: ((red cell * (sum of blue cells)) – ((total rocks vertical pink* total rocks horizontal green)+(total shrubs vertical pink*total shrubs horizontal green)+(total Agriculture vertical pink*total agriculture horizontal green)))/((red^2) - ((total rocks vertical pink* total rocks horizontal green)+(total shrubs vertical pink*total shrubs horizontal green)+(total Agriculture vertical pink*total agriculture horizontal green))).*
- **Ground truth** is used to compare pixels on the satellite image to what is there in reality in order to verify the contents of the pixel on the image. A value of 0 shows there is no similarity with reality, a value of 1 shows the classified pixels are similar to reality.
  To obtain the ground truth values: value land use type white cell/pink cell (total of that land use type).*
- **Commission** is the percentage of test pixels that are incorrectly classified as a particular class. To obtain the commission values: sum horizontal cells non-blue cells land use type/green cell land use type * 100.*
- **Omission** is the opposite of commission, also defined as percentage.
  To obtain the omission values: sum vertical cells non-blue cells land use type/ pink cell land use type * 100.*
- The **user's accuracy** refers to the probability (in percentage) that a pixel that is given a certain land use class is really that class.
  To obtain the user's accuracy: blue cell particular land use type/ green ground truth value that land use type * 100.*

- The **producer's accuracy** refers to the probability (in percentage) that the land cover of a certain area on the ground is classified as such:
  To obtain the producer's accuracy: blue cell particular land use type/ pink total value that land use type * 100.*

The total accuracy does not give details of the accuracy of individual classifications (per class), while they influence the total accuracy indirectly. In case of a low accuracy, the user and producer accuracy can be studied to find out which classes cause the low accuracy. Therefore also the user and producer accuracy is calculated per class:

*For more insight in the formulas, take a look at the excel tables of the different error matrices which can be found on the DVD belonging to this research. The three error matrices for the Quickbird image of March can be found under the following names:
- *error matrix March SA1*
- *error matrix March SA2 - merged agriculture*
- *error matrix March Rusinga*