Eli Baldwin

Master's Thesis

Agricultural Land Change in Zambia, 2000-2010-2015

Advisor: Professor Lyndon Estes

Clark University, Worcester, MA

Abstract:

Agriculture has been identified as the largest and most important Land Use/Land Cover (LULC) class in terms of LULC change, especially in terms of its gain from natural land covers. Agricultural land change can have important consequences for food security and environmental protection. There is a paucity of data and analysis of agricultural LULC change for Zambia. The literature shows many variables affect agricultural LULC change, including population, accessibility and environmental factors. This thesis analyzes the change in agricultural LULC, quantifies the strength of explanatory variables, and develops a model to predict the gain of agriculture in Zambia. There is a gain of agriculture between 2000 and 2010, and then a larger gain between 2010 and 2015. There are larger exchanges during the first time interval. Areas of shifting cultivation can be identified as areas with a higher proportion of toggle, meaning pixels that change to a different LULC class and then change back to the original class. Districts of Zambia with more agriculture tend to have medium gain intensities but low loss intensities. Districts of Zambia with less agriculture tend to have higher gain and loss intensities. Distance from existing agriculture is the most important variable in predicting gain of agriculture. Using distance from agriculture and holding the other variables constant decreases the calibration accuracy of the transition potential modeling part of the land change model by less than 1 percentage point. The predictions of gain of agriculture can get a considerable amount of the agriculture in the correct general area, but not the exact right area.

Table of Contents page #

Chapter 1: Introduction

Agriculture has been identified as the largest and most important Land Use/Land Cover (LULC) class in terms of LULC change, especially in terms of its gain from natural land covers (Lambin & Meyfroidt, 2011). Expansion of agriculture has been identified as the biggest contributor to deforestation in Eastern Africa (Gibbs et al., 2010) and Zambia (Vinya et al., 2011). Agricultural expansion can negatively affect savannah biomes as well (Searchinger et al., 2015). Agricultural land change can have important consequences for food security and environmental protection (Gibbs et al., 2010). Zambia has high levels of malnutrition and a rural population dependent on agriculture (World Development Indicators, 2020).

There is a paucity of data and analysis of agricultural LULC change for Zambia. I have thus far found only one study that analyzed LULC change in one region of Zambia (Petit, Scudder, & Lambin, 2001). The study was conducted in Lusitu, Southern Province, Zambia. The authors found that agriculture LULC was growing from 1986-1997, although that growth had slowed between 1992 and 1997. Agriculture gained mostly from forest and savannah. The authors found that agriculture pushed into natural LULC and then settlement expanded behind it, so some agricultural land was lost to settlement. Some agriculture also changed to grassland in the case of field abandonment, permanent or temporary. The study also found that Zambia's protected areas remained largely unchanged (Petit, Scudder, & Lambin, 2001). Most of the other studies I found in the literature from Africa and elsewhere found net growth of agriculture (Ali, Descheemaeker, Steenhuis, & Pandey, 2011; Ampofo, Sackey, & Ampadu, 2016; Calzada, Meave, Bon, & Figueroa, 2018; Gashaw, Bantider, & Mahari, 2014; Kindu, Schneider, Tekatay, & Knoke, 2015; Msofe, Sheng, & Lyimo, 2019). This growth in agriculture was most often found to come from forest, savannah/grassland, or bare lands. Some of these studies mentioned some loss of agriculture to settlement or grassland (Calzada et al., 2018; Kindu et al., 2015; Msofe et al., 2019). There were also some studies which mentioned that expansion of agriculture is happening on increasingly marginal land (Kindu et al., 2015; Stehfest, Doelman, & Mandryk, 2015). I have found two studies that found a net decrease in agriculture. One found that population change and poverty were influencing agricultural LULC change in Central Malawi from 1991-2015. (Munthali et al., 2019). The other study hypothesized that intensification of agriculture and fewer farmers were the causes (Kamwi, Chirwa, Manda, Graz, & Katsch, 2015).

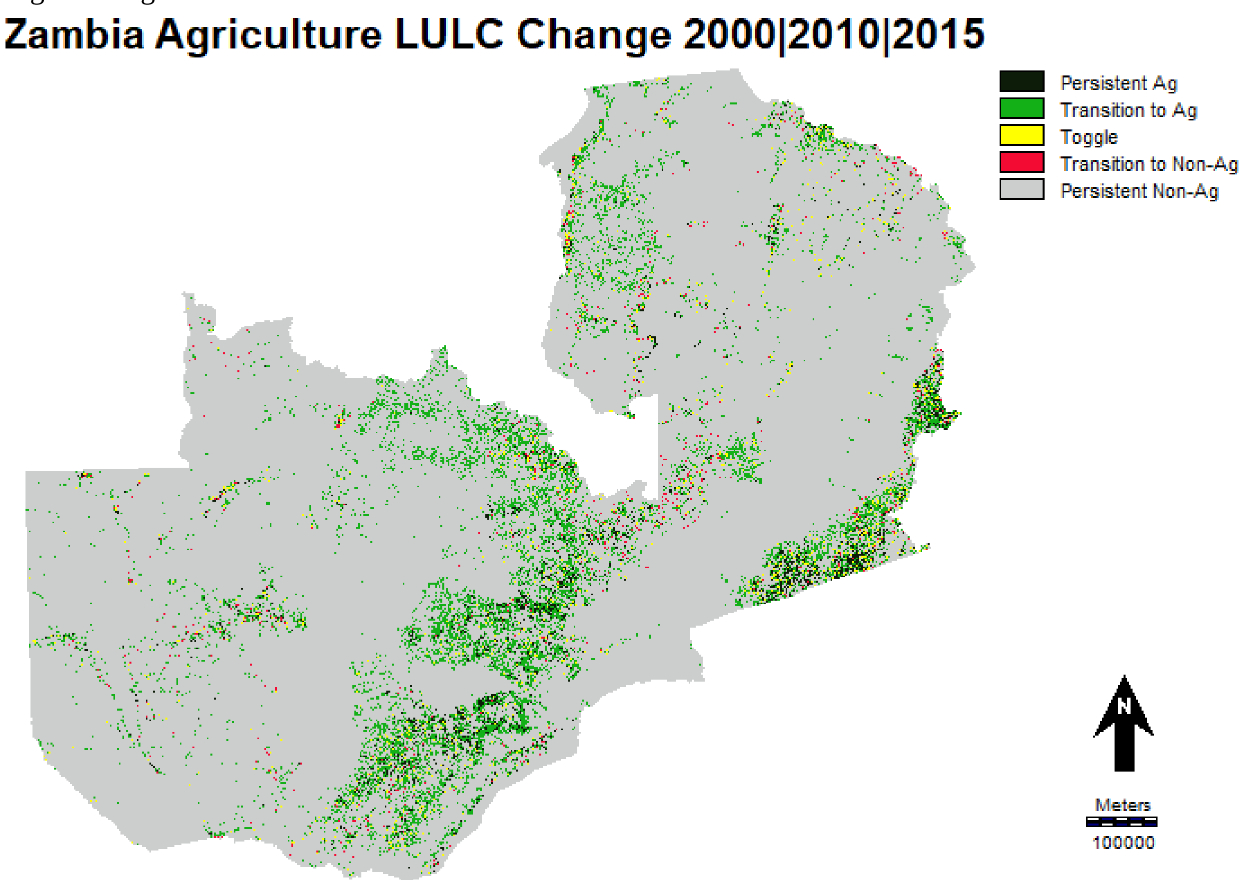
There are numerous LULC driver/predictor variables found in the literature. One of the most common seems to be population or population density change (Ali et al., 2011; Ampofo et al., 2016; Gashaw et al., 2014; Id et al., 2018; Kindu et al., 2015; Msofe et al., 2019; Munthali et al., 2019). Another group of commonly cited variables are accessibility. These include distance to existing agriculture, distance to roads, distance to settlements, and a variable called accessibility (Id et al., 2018; Kindu et al., 2015; Msofe et al., 2019; Stehfest et al., 2015). There are multiple raster datasets of accessibility that represent distance to nearest settlement of a certain size. Some of the literature also examined environmental factors such as rainfall, temperature, soil suitability and agro ecological zone and found that they effected the growth of agriculture (Calzada et al., 2018; Id et al., 2018; Kindu et al., 2015; Msofe et al., 2019; Stehfest et al., 2015; Yalew, Griensven, Mul, & Zaag, 2016). Elevation and slope were also found to be relevant variables (Id et al., 2018; Kindu et al., 2015; Msofe et al., 2019; Stehfest et al., 2015). Stehfast et al. (2015) recommend excluding pixels with slopes greater than 45 degrees when determining suitability for agriculture.

This project contributes to the agricultural LULC change literature by analyzing the entire country of Zambia at 30m resolution. This will be done in three stages. The first stage is analysis of binary agriculture and non-agriculture maps to determine the quantities of agriculture LULC change and their distribution in Zambia's districts. The second stage is the analysis of driver variables. The third stage is modelling the agricultural LULC change. All of the analyses will be carried out for three time points, meaning two time intervals, 2000-2010-2015.

Chapter 1a. Study Area:

The country of Zambia. Zambia is in a sub-tropical climate zone. Most of the vegetation cover in Zambia is woodlands, with Miombo being the largest type. The second largest vegetation cover category is grasslands (Vinya et al., 2011). The FAO estimates that there are 23,836,000 hectares of farmland in Zambia (2016).

Figure 1: Agriculture in Zambia.



According to Figure 1 agriculture in Zambia is concentrated in the central part of the country, and the southeast. The west and northeastern parts of the country have sparser agriculture. Toggle refers to the case when a pixel changed during the first time interval and then changed back to the original category in the second time interval. The northeastern part of Zambia is known to have more traditional shifting cultivation, which is called chitemene (Araki, 2007; Kapekele, 2006).

<u>Chapter 1b; Data</u>:

Name	Source	Date	resolution
Global Food Security-Support Analysis Data	NASA	2015	30m
Land Cover Project of the Climate Change Initiative	ESA	2016	20m
LULC 30m	Regional Centre for Mapping and Resource Development	2000, 2010	30m
GlobeLand30	National Geomatics Center of China	2000, 2010	30m
Elevation	USGS	2000	30m
Soil	FAO	2001	1km
Roads	Estes et al (2016)		vector
Protected areas	UNEP	2014	vector
Climate	WorldClim	1970-2000	1km
Population	NASA SEDAC	2000, 2010, 2015	250m
Accessibility	Global Environment Monitoring Unit - Joint Research Centre of the European Commission	2000	1km

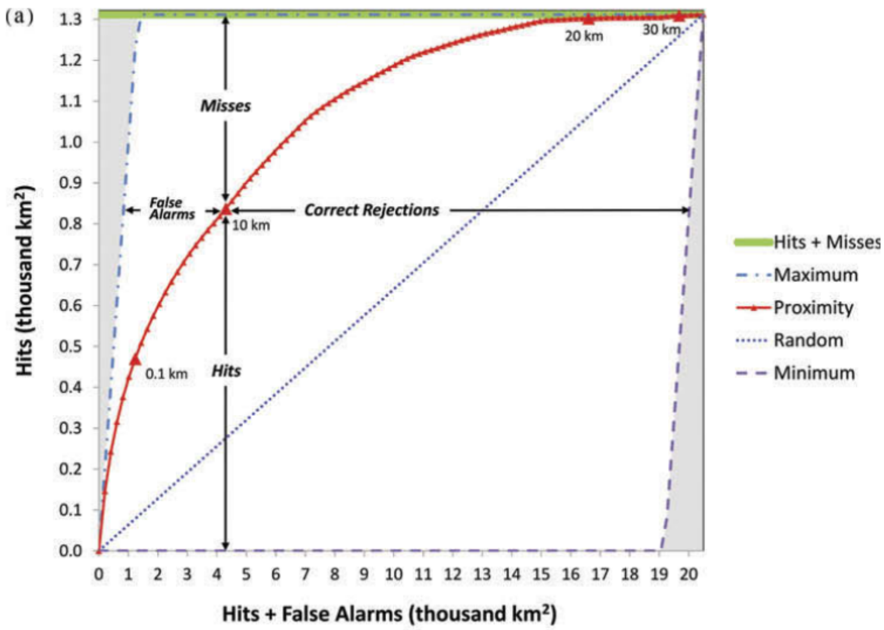
Priscilla Baltezar created maps of agriculture in Zambia for 2000, 2010 and 2015 using Bayesian fusion of the LULC datasets in the data table. This process was also bolstered by validation at approximately 800 points spread around Zambia using a stratified random sampling approach. This process of creating Bayesian updated land cover maps is described in (Baltezar et al., 2020). The terrain module in TerrSet was used to calculate slope from the elevation model. I calculated the difference and percent change in population for both time intervals using the image difference module in TerrSet. I used the climate data, which consists of monthly mean temperature, maximum temperature and precipitation in the crop climate suitability modeler in TerrSet. I calculated the suitabilities for maize, soy and cassava, the three main crops grown in Zambia (Zambia CountrySTAT, 2016), using the minimum suitability between temperature and precipitation. I then overlaid these three outputs using the mean operator to create one suitability map.

<u>Chapter 2, Methods</u>:

I calculated the cross tabulations for both time intervals, 2000-2010 and 2010-2015. Then I entered the cross-tabulation outputs in the PontiusMatrix42.xlsx (available at www.clarku.edu/~rpontius) which calculates gain and loss, and quantity and exchange for both. Quantity refers to the net gain or loss that a category experiences. Exchange refers to the changes back and forth below the net. Gain exchange and loss exchange will be equal for each category when there are only two categories (Pontius Jr. & Santacruz, 2014). I then calculated the average yearly gain quantity and exchange values because the two time intervals are different lengths of time. I created a scatterplot of the percentage of Zambia that is agriculture at each of the three years also adding points showing what percent of Zambia would be agriculture in 2010 and 2015 if only gross gain was taken into account. This figure is based on (Pontius & Malanson 2005). I calculated the three-way cross tabulation for all three time points. I calculated the persistence, incidence and toggle based on the results of the three-way crosstab. This analysis follows the methods of (Pontius Jr., Krithivasan, Sauls, Yan, & Sauls, 2017). I took the results from the three-way cross tabulation and aggregated the pixels to a coarser resolution in order to determine the areas of the country that had the most toggle relative to persistent and new agriculture (toggle/[toggle + persistent agriculture + new agriculture]). The hypothesis behind this is that there is more chitemene shifting cultivation in the northeast of Zambia (Araki, 2007; Kapekele, 2006), and so that area was expected to have more toggle. I used the zonal statistics tool to determine the amount of agriculture in each district of Zambia at each time point and the gain of agriculture during both time intervals. I then used these numbers to calculate the percent of each district that is agriculture, the gain and loss intensities for each district for both time intervals and the net gain and net percent change for each district from 2000-2015. The gain intensity is defined as the proportion of agriculture at the second time point that is new, i.e. was not agriculture at the first time point. The loss intensity is defined as the proportion of the agriculture at the first time point that transitions to non-agriculture at the second time point (Pontius Jr., 2019; Quan, Pontius Jr., & Song, 2019).

I created maps of agricultural LULC gain for both time intervals. I took a stratified random sample with the goal of getting 100,000 pixels from both categories of the Boolean map, which are 'gain of ag' and 'not gain of ag'. I did this because inputting the entirety of the data is too large for R studio. The sample for the first time interval had 100,000 pixels of 'not gain of agriculture' and 72, 364 pixels of 'gain of agriculture.' The sample for the second time interval had 100,000 pixels for both categories. This sample excluded national parks and pixels with slopes greater than 20% as these pixels are considered to be less suitable for agriculture (Estes, Searchinger, Tian, & Kehoe, 2016). Pixels that were persistent agriculture were also excluded. I then extracted the same sample of pixels from each of the driver variable maps. I then analyzed each driver variable using the Total Operating Characteristic (TOC), which is similar to the ROC, but contains strictly more information (Pontius Jr. & Si, 2014). The TOC measures how well a variable diagnosis a presence of something. In this case how well the variable maps diagnose presence of gain of agriculture. Below is an example TOC chart from Professor Pontius (Pontius Jr. & Si, 2014) (see figure 2). It shows how the TOC reveals the hits, misses, false alarms and correct rejections for each threshold. The correct quantity of presence can be found at the point where the Hits+Misses line meets the maximum line. The maximum line represents the perfectly correct diagnosis. The closer the curve is to the maximum line, the better the variable in question predicts the change in question. The TOC curve also produces an Area Under the Curve (AUC) identical to that of the ROC. The AUC is a summary statistic which ranges from 0 to 1 where 1 is perfectly correct and 0.5 is equivalent to that of an expected random line (Pontius Jr. & Si, 2014).

Figure 2: Explanatory example TOC graph from (Pontius & Si, 2014)

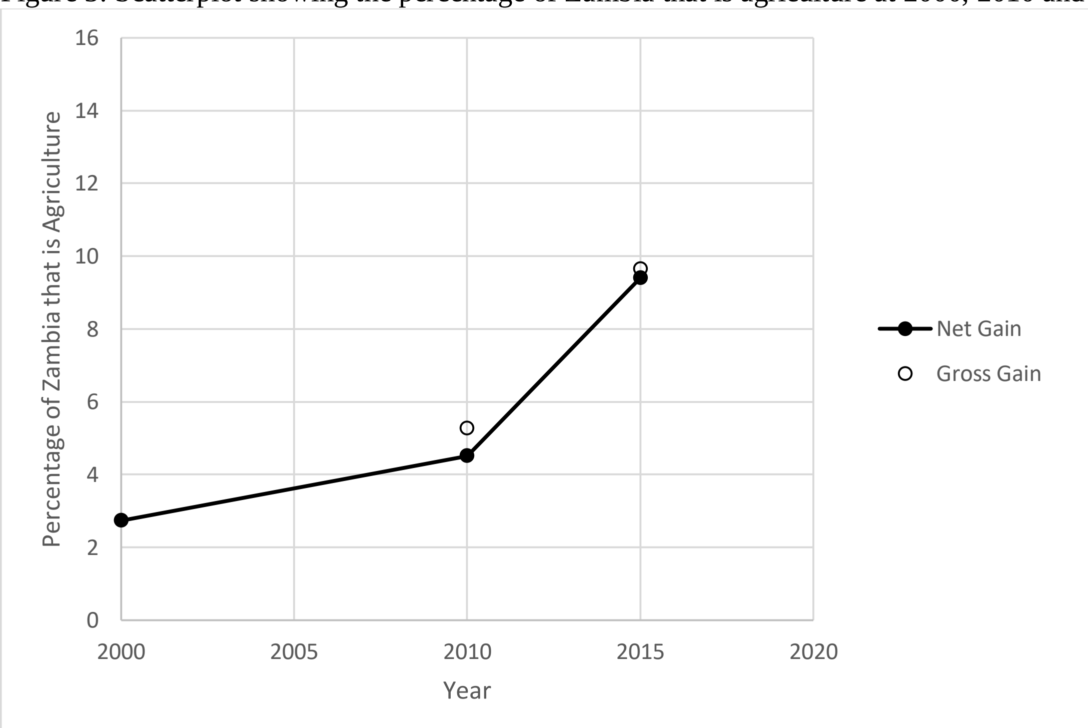


I used the Land Change Modeler (LCM) in TerrSet to model the gain of agriculture in 2015 using the maps of agriculture in 2000 and 2010. I modeled only the transition from not-agriculture to agriculture. I first did this with all the variables then with only the four variables that were shown to have strong relationships with gain of agriculture. I transformed the "distance from" variables using the natural log function and the soil index with the evidence likelihood transformation. I used the Multi-Layered Perceptron (MLP) to calculate the transition potentials. I also used the transition potential maps produced by the LCM as input variables to produce TOC graphs. I then extrapolated the land cover change from 2010 to 2015 using all variables and validated this against the actual land cover change from 2010 to 2015, creating maps of hits misses and false alarms for gain of agriculture and then a graph depicting quantities and exchanges for those categories as well as calculating the total percent correct. I then created a predicted land cover map for 2020 using the data from 2010 and 2015 as I thought it was important to use the most recent known agriculture data given that distance to existing agriculture is so important to the prediction. I then updated the scatterplot showing the percentage of Zambia that is agriculture to show predicted agriculture for 2015 and 2020. These are important to showing how gain quantity of agriculture is predicted since my land change model only models gain of agriculture.

Chapter 3, Results:

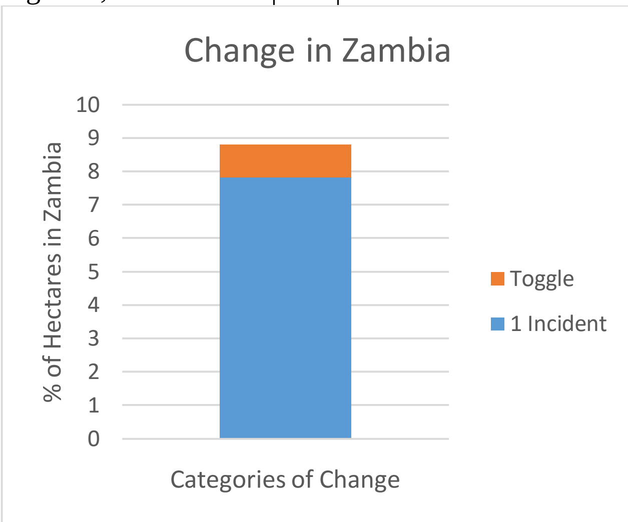
There was a net gain of agriculture from 2000-2010 (Figure 3). Then there was a bigger net gain of agriculture from 2010-2015. The exchange categories are larger for the first time interval than for the second (figure 3, also appendix 5). This is evidenced by the greater distance between the points representing gross and net gain in 2010 as compared to 2015.

Figure 3: Scatterplot showing the percentage of Zambia that is agriculture at 2000, 2010 and 2015.



The increased slope of the line between 2010 and 2015 reveals that the gain intensity is even greater in the second time interval. The percentage of Zambia that is agriculture in 2010 is ~4.5 and in 2015 is ~9.4, meaning there was a more than doubling of agriculture in that time interval. The average yearly graph shows exchange categories are larger in the first time interval on a yearly basis, but not as much larger as they appear on the interval basis. (appendix 5).

Figure 4, Zambia 2000|2010|2015



This graph (figure 4) shows that almost 9% of the land area of Zambia changed between 2000 and 2015. Approximately 1% of the pixels toggled, meaning changed and then changed back. A large amount of toggle can indicate error, but in this case may also indicate shifting cultivation.

Figure 5:

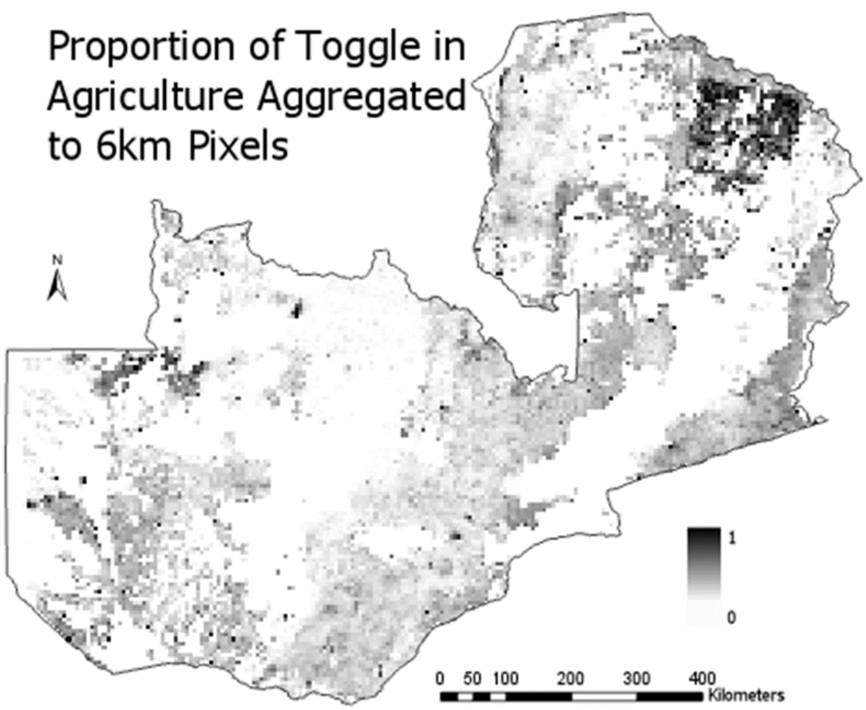
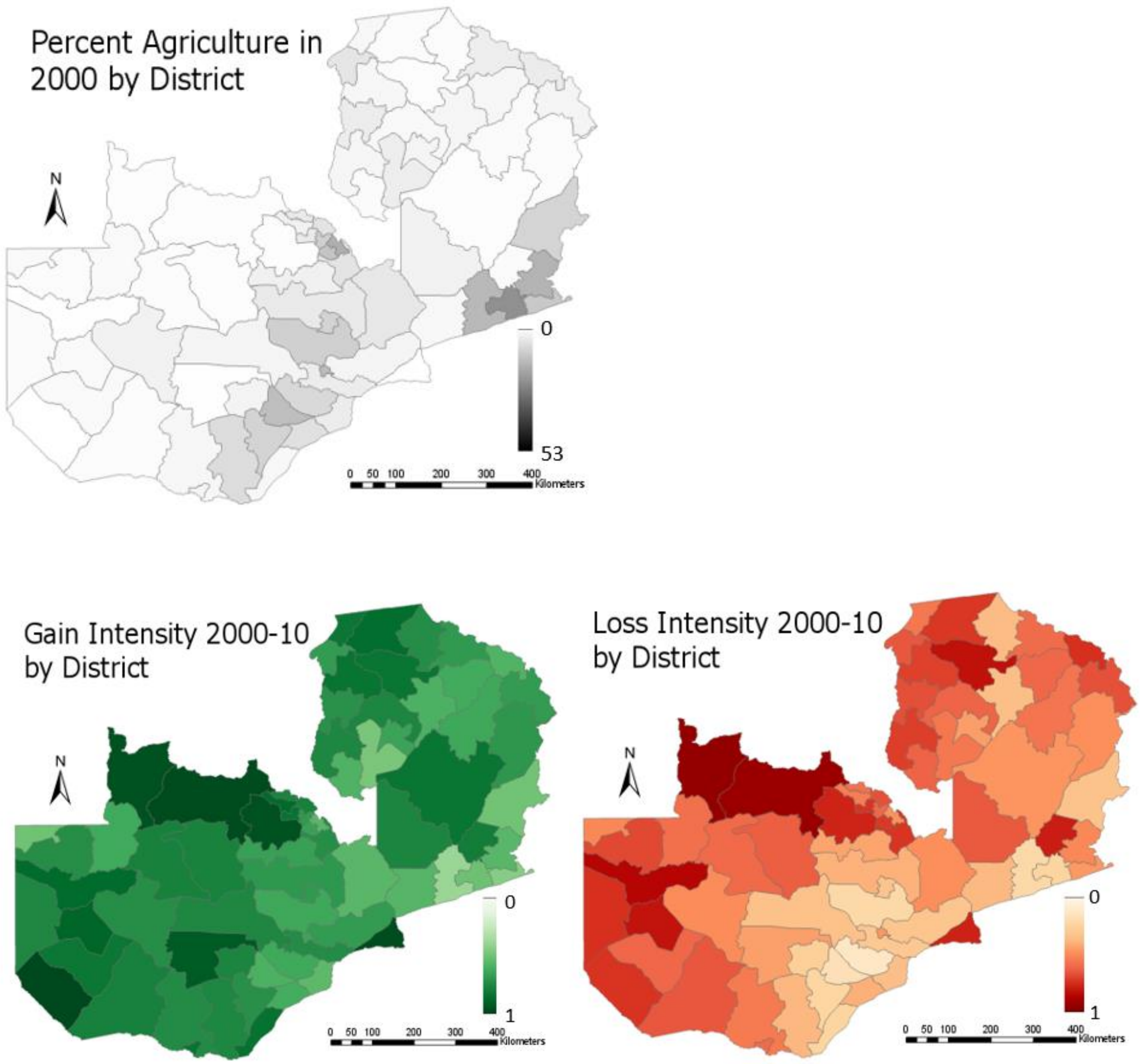
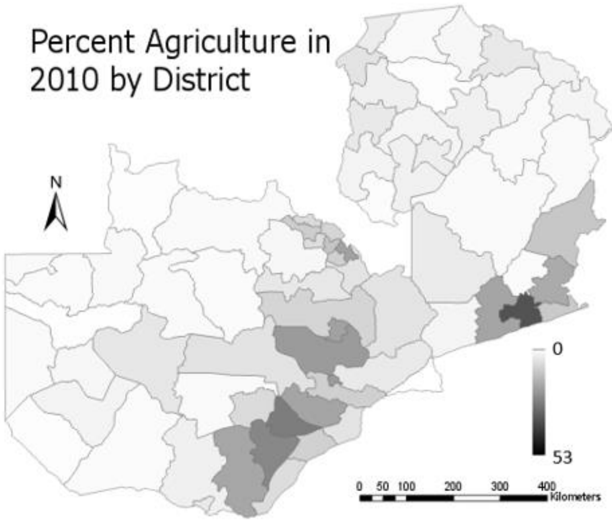
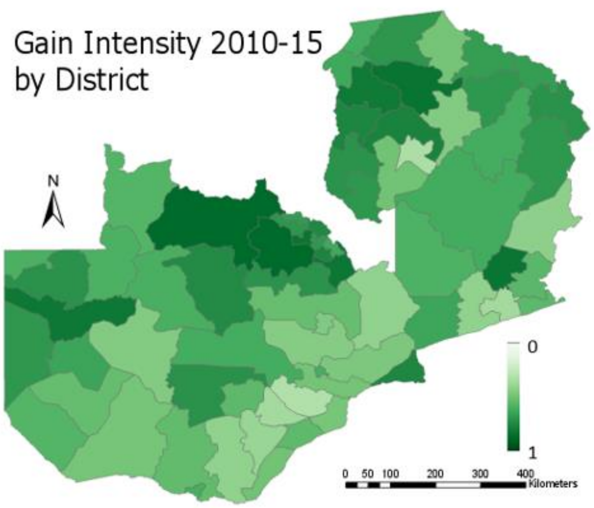


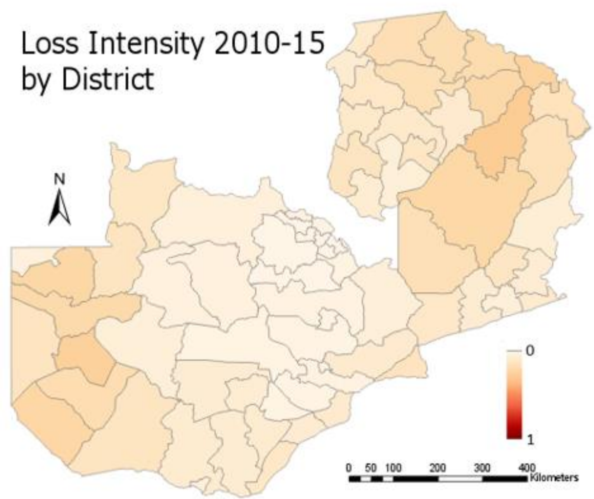
Figure 5 shows that there is toggle present in all areas of agriculture. However, there is a concentrated area with a higher proportion of toggle in northeastern Zambia. There are generally lower proportions of toggle in the areas of more intense agriculture in the central and southeastern parts of the country. There are some individual pixels with high proportions of toggle throughout the map.

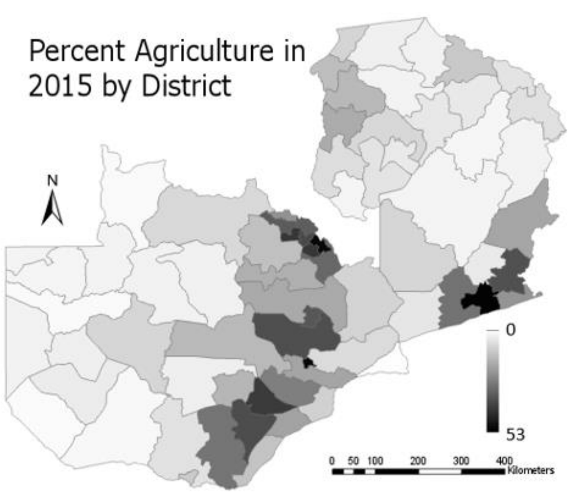
Figure 6: Breakdowns by District




Percent Agriculture in
2010 by District
N
0
53
0 50 100 200 300 400
Kilometers


Gain Intensity 2010-15
by District
N
0
1
0 50 100 200 300 400
Kilometers


Loss Intensity 2010-15
by District
N
0
1
0 50 100 200 300 400
Kilometers


Percent Agriculture in
2015 by District
N
0
53
0 50 100 200 300 400
Kilometers

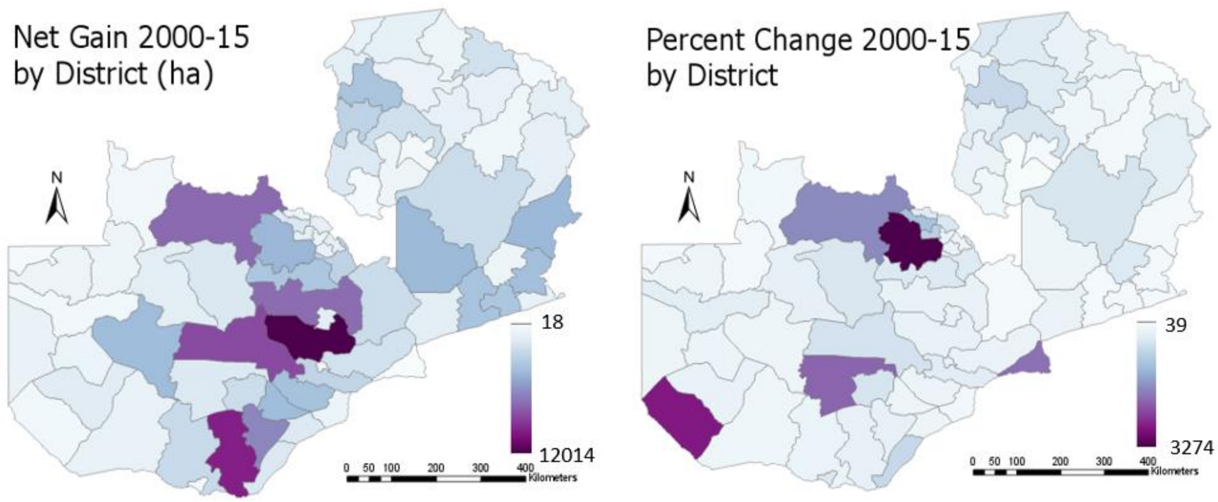
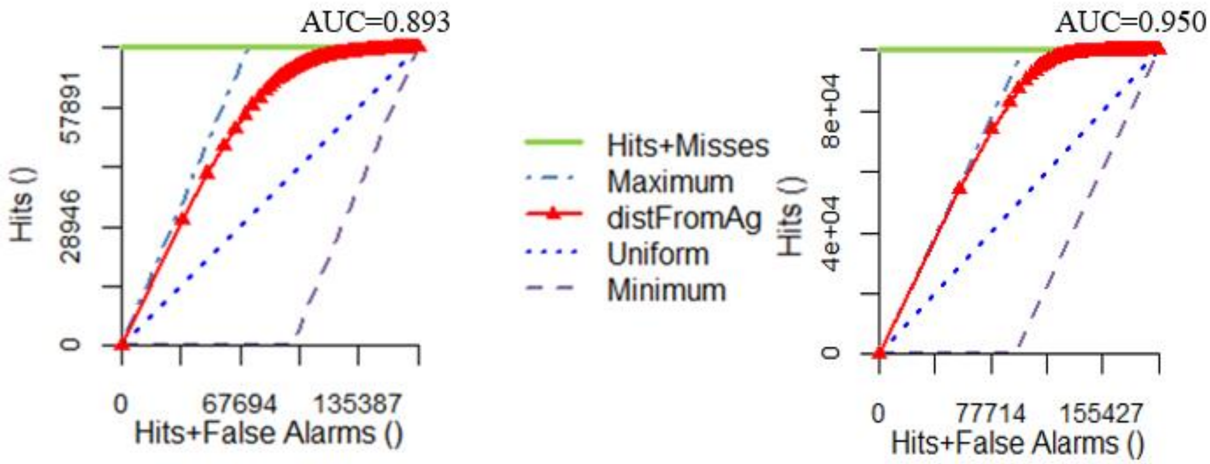


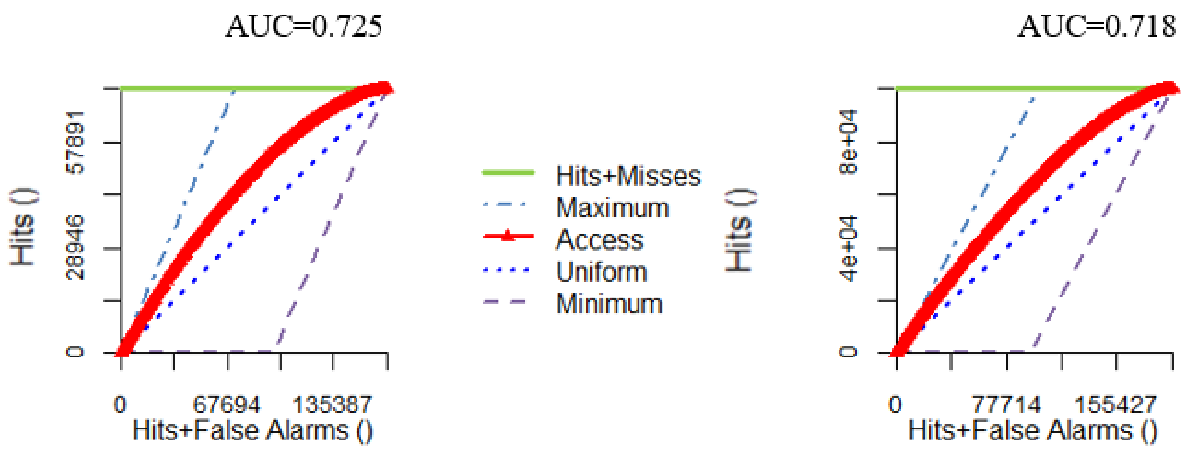
Figure 6 shows that the most intensely farmed districts in Zambia are in the central area and in the southeast. These areas have been the most intensely farmed districts since 2000, although their percent agriculture has increased. The first time interval is characterized by high gain and loss intensity for most districts. Some of the more intensely farmed districts have middling gain intensities, but particularly low loss intensities. The north-central districts have high gain and loss intensities. The second time interval is characterized by slightly lower gain intensities and much lower loss intensities. The northeastern and western districts have higher loss intensities than the other districts. The north-central districts continued to have high gain intensities. The largest net gains were in central and north-central districts. The districts with the highest percent agriculture in 2015 were mostly districts with the lowest loss intensities as opposed to the districts with the highest gain intensities.

Only four of the driver variables had TOC curves with an AUC greater than 0.53. The rest had AUCs within ± 0.055 of the expected random, 0.5. None of these curves varied much from the expected random line (see appendix 2). Those four driver variables were, in order of predictive ability, distance from existing agriculture, accessibility, distance from roads and population change (see figure 6).

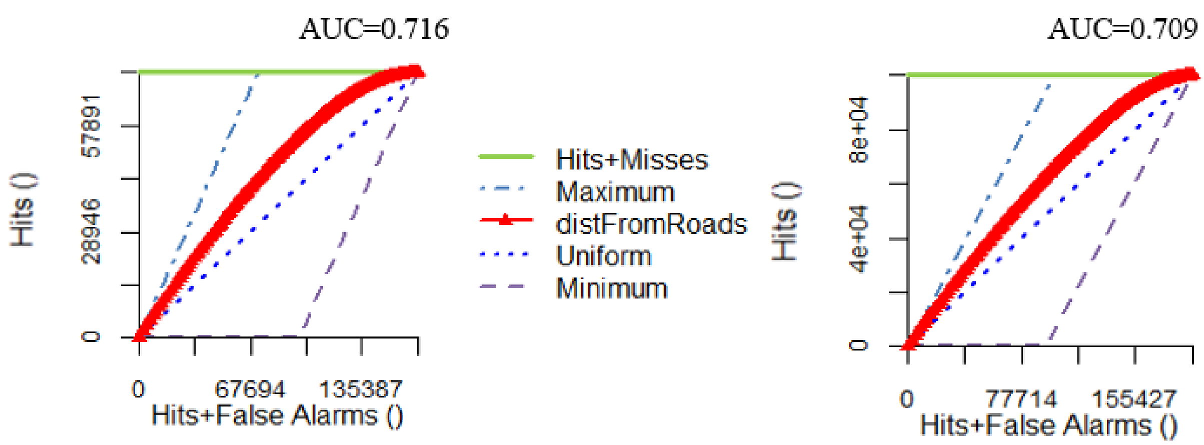
Figure 7. TOC graphs with AUCs First column is 2000-2010, second column is 2010-2015
Distance from Agriculture:



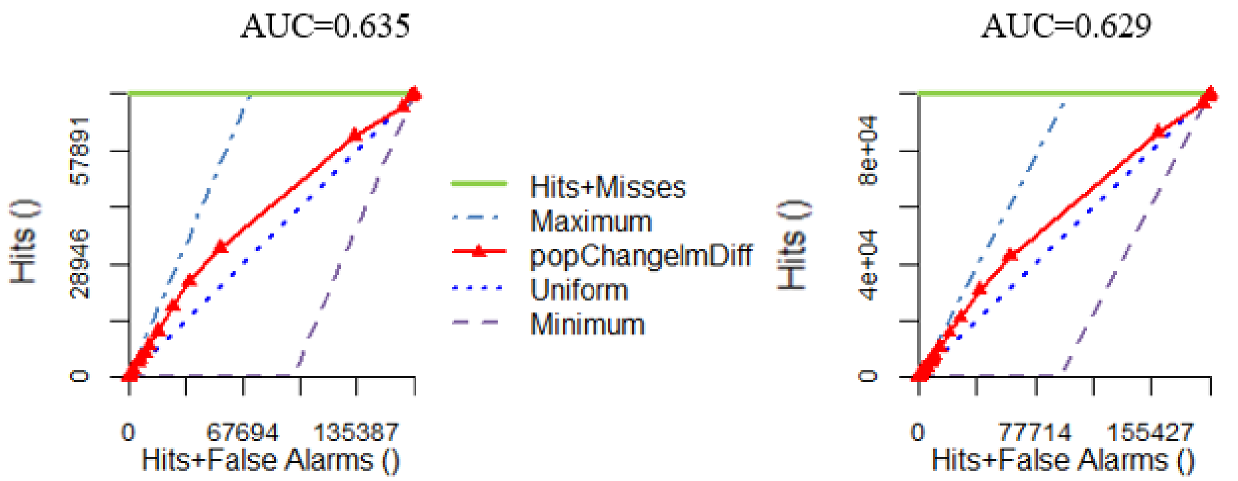
Accessibility (distance from population center >50,000):



Distance from Roads:



Population Change:



The distance from agriculture TOC curves show the strongest relationship with gain of agriculture. The AUC is near 0.9 for 2000-2010 and even higher, 0.95 for 2010 to 2015. That is closer to perfectly correct (1.0) than to random (0.5). Accessibility and distance from roads have smooth and similar curves with AUCs of around 0.715. The curve for population change is close to the maximum line for the highest values before moving towards the uniform line for middle and lower values. The other variables were elevation, slope, crop climate suitability, soil productivity index and percent population change. When I entered all the variables into the Land Change Modeler and created the transition potentials using the MLP, the report revealed that distance to existing agriculture was the most important variable, and that using only distance to agriculture would result in less than a 1 percentage point drop in calibration fit as compared to using all the variables (see appendix 3)

The MLP transition potentials model with four variables had a calibration fit of 82.20% and a skill of 0.6439. This was mostly attributable to distance from agriculture, the most important variable according to the report (see appendix 3). The MLP output with all variables as inputs had a calibration fit of 83.09% and a skill of 0.6618. This discrepancy is small enough that it is likely due to the uniqueness of every run of the MLP and not a substantial difference in the two models. Once again, distance from existing agriculture was the most important variable, and used alone had a calibration fit of the model less than one percentage point less as compared to the model that had all the variables (see appendix 3).

Figure 8: TOC graph for transition potentials from LCM with all 9 variables:

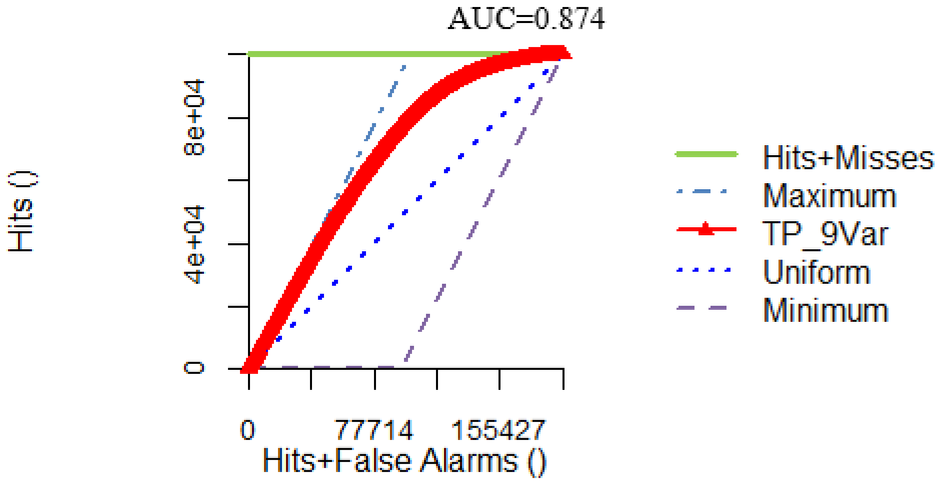
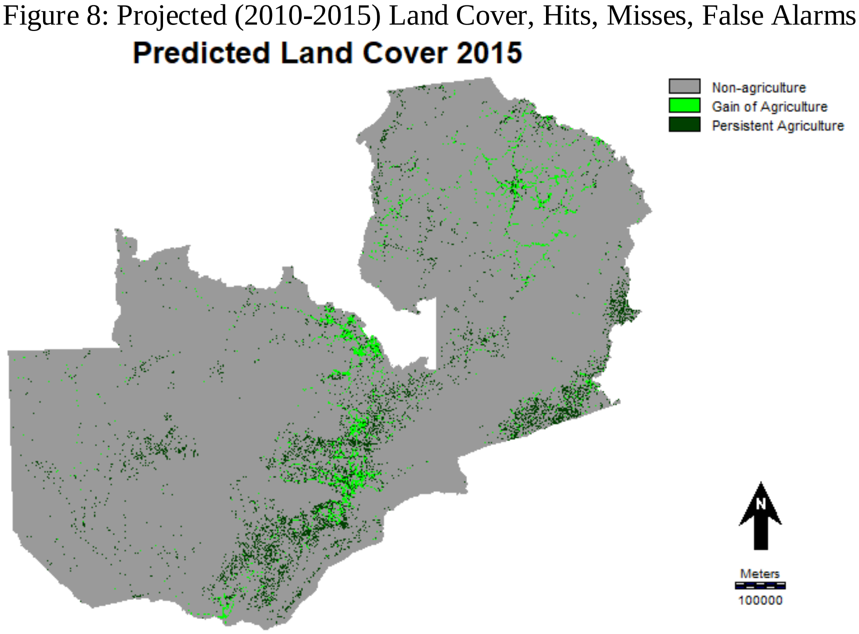


Figure 7 shows the TOC graph for transition potentials which has a smooth curve and an AUC of 0.874.

Figure 8: Projected (2010-2015) Land Cover, Hits, Misses, False Alarms



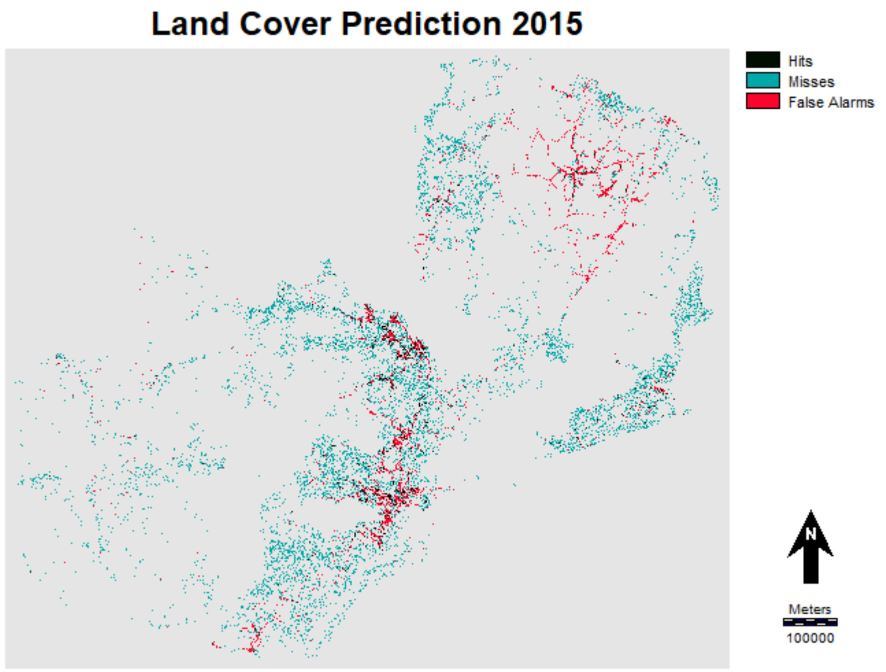


Figure 9: Validation of Prediction

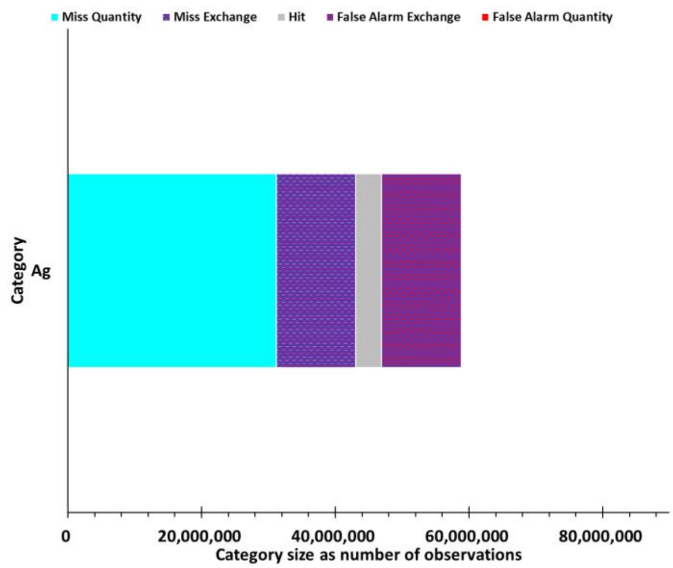


Figure 8 shows that the correct predictions of new agriculture (hits) are concentrated in the central area, which has the highest concentration of agriculture, and in the southeast and northeast sections of Zambia. The new agriculture that was not predicted (misses) are more spread out over the whole country, barring the large national parks. However, there are more misses in areas with more agriculture. The incorrect predictions of new agriculture (false alarms) are concentrated in the central and northeast sections of Zambia. Figure 9 shows that the Miss quantity is approximately as large as the exchanges and hits. The exchange categories are larger than the hit category. The overall accuracy of the prediction was 92.12% correct.

The run of the MLP for the transition potentials using 2010 and 2015 data yielded a calibration fit of 88.54% and a skill of 0.7707. Once again distance from agriculture was the most important variable, which used alone would have led to a decrease of less than 1 percentage point in the calibration fit of the model compared to a model that includes all variables (see appendix 3).

Figure 10: Projected (2015-2020) Land Cover

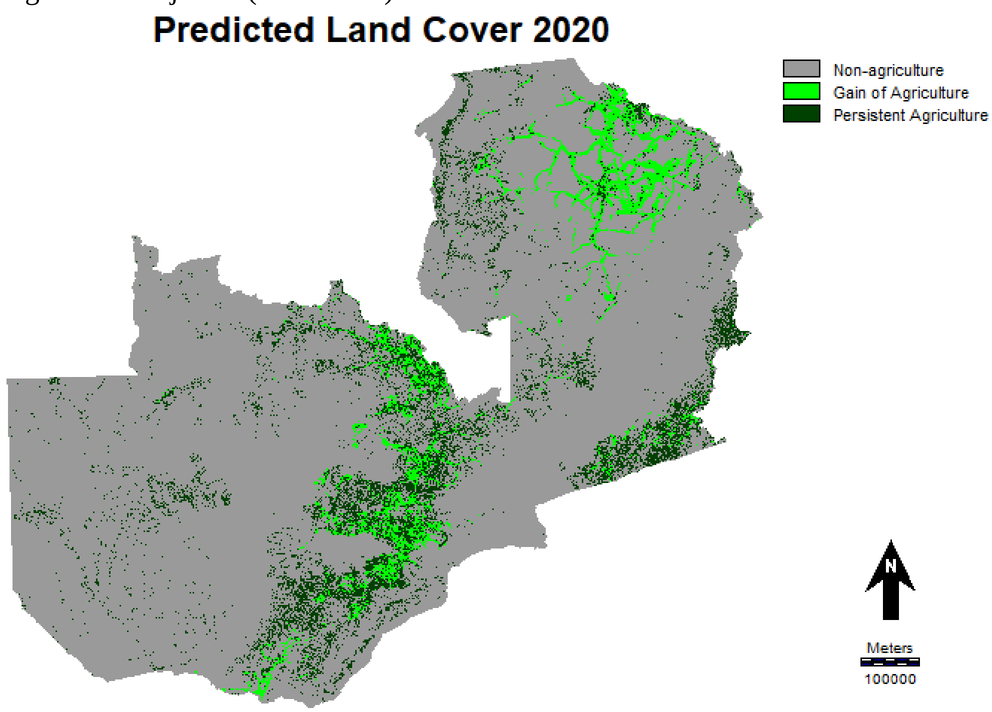


Figure 10 shows expansion of agriculture as expected. The predicted expansion is concentrated in the center, southeast and northeast of Zambia. The prediction in the northeast is similar to the over prediction for 2015.

Figure 11: Scatterplot showing percentage of Zambia that is ag, reference and predicted.

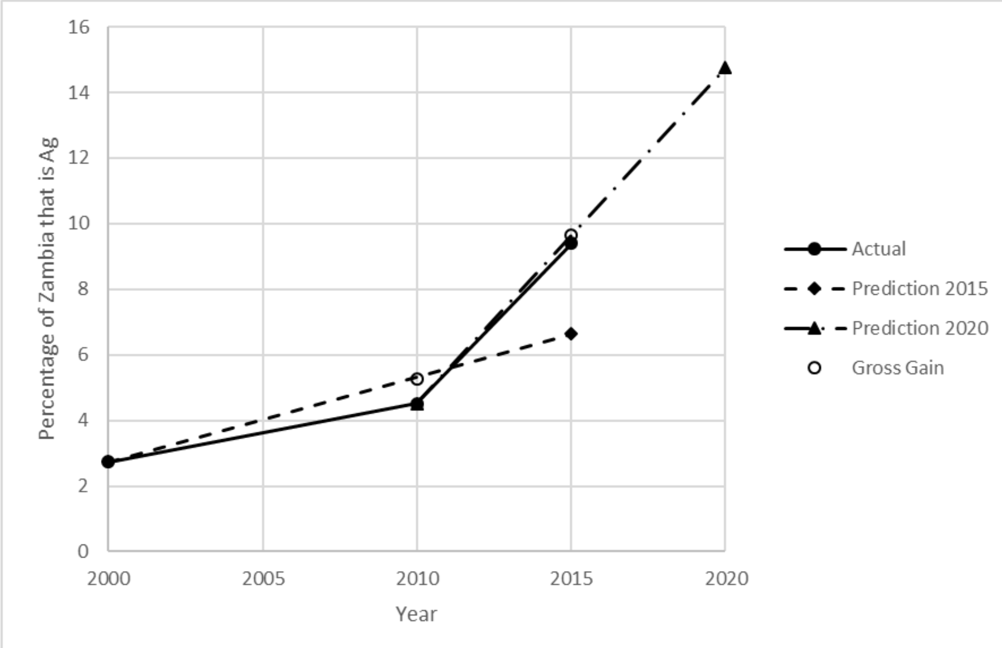


Figure 11 demonstrates how the prediction of quantity of agriculture for 2015 and 2020 are based on the gross gains of agriculture from the calibration time intervals, because only gain of agriculture was modeled. It also shows that the projected agriculture for 2015 covers over 3 percentage points less of Zambia than the reference agriculture.

<u>Chapter 4, Discussion</u>:

The fact that there are gains of agriculture found in both time intervals is in keeping with the trends found in my review of the literature in Chapter 1 with most studies showing expansion of agriculture. The amount of agriculture in Zambia more than tripled from 2000 to 2015, with most of the gain coming from 2010-2015. The larger gain in the second time interval is somewhat surprising because the duration of the second time interval is half the duration of the first time interval. The total gain of agriculture is likely smaller in reality. The larger exchange quantities in the first time interval could have two explanations. There might have been more shifting cultivation taking place in that time interval, and less in the more recent one. This is supported by the finding that the area in the northeast with the highest proportion of chitemene experiences a modest decrease in gain intensity but a large decrease in loss intensity as do other areas with lower concentrations of agriculture, while the central districts with the most agriculture and a smaller proportion of shifting cultivation experience modest declines in both gain and loss intensities. Higher gain and loss intensities with low net gain of agriculture are consistent with more shifting cultivation. This discrepancy could also be the result of error in the classifications. The gain in the second time interval seems too large. Data from FAOstat that is derived from MODIS imagery shows a much more modest gain for this time interval, although it is larger than the gain from 2001-2010 (see appendix 6). An accuracy assessment of the data based on approximately 600 validation points for each date reveals the following percent accuracies for the agriculture class:

Accuracy in terms of-	2000	2010	2015
Commission	87.69	76.70	56.03
Omission	52.29	65.29	87.80

This indicates that the classifications under predicted agriculture in 2000 and 2010, while over predicting agriculture in 2015, which could account for the apparent large gain of agriculture from 2010 to 2015. These accuracies are lower overall than those found by (Baltezar et al., 2020) which were for only agro ecological zone IIa. That zone includes most of the more concentrated areas of agriculture in the central and southeastern parts of the country, indicating that the classifications were more accurate for those areas. The toggle in the image does likely represent some error. However, there is a higher proportion of toggle in the northeast, which supports the idea that toggle primarily represents chitemene shifting cultivation. Most of the gain in agriculture in Zambia is coming in areas that already have a higher percentage of agriculture, where less agriculture is being converted to non-agriculture.

New agriculture is being created near existing agriculture. Given that the study area has large areas of non-agriculture that is not surprising, although the exclusion of national park land decreases the severity this issue. The other three variables that were found to be predictive are similar to distance from existing agriculture as they are focused on areas of human activity. However, in the models, these variables are dwarfed by distance from agriculture, which could indicate that they are showing the same phenomenon just less strongly. Elevation and slope were not found to be predictive; this is partly because high slopes were eliminated from the maps before the analysis, and some of the highest elevations in Zambia are in a national park (see appendix 1). Crop climate suitability and soil productivity index were not found to be predictive. This could indicate that farmers in Zambia have used up much of the most desirable land, and new farms are being made on less desirable land. According to the crop climate suitability model output, much of the land area of Zambia is highly suitable for the main crops produced in Zambia. The soil index only has six levels, and just one of those levels covers much of Zambia (see appendix 1). This could also indicate that there is a problem with the underlying data or the indices themselves.

The variable distance from roads is by far the most important in the LCMs predictions. This could be partly because Zambia is such a large study area most of which is not agriculture. The transition potentials produced by the MLP in the LCM have a lower AUC from the TOC graph than distance from roads alone does. This indicates that adding other variables and using more complex methods was not helpful in determining where new agriculture is being created. The MLP under predicted the gain of agriculture from 2010-2015 because there was a larger and more intense net gain of agriculture from 2010-2015 than there was from 2000-2010. The prediction placed new agriculture primarily in the central and northeast sections of Zambia. Many of these predictions in the central area were correct, many were wrong. The prediction over predicted the growth of agriculture in the northeast region. The prediction under predicted the gain of agriculture in the north-central area of Zambia in particular. The overall accuracy is 92%, but 91% of the pixels did not change at all from 2000 to 2015, which indicates that the high accuracy comes primarily from correctly predicting large amounts of persistence. The 2020 prediction shows more agriculture in the north-central region, reflecting the increases there between 2010 and 2015. This prediction also predicts more agriculture in the north-east region, which may again be an over prediction. There are many pixels near to agriculture in 2015 and the LCM has to decide which ones to allocate gain of agriculture to. The expansion is clearly along existing roads. The northeast region has particularly high CCSM values and one population center giving the area low accessibility values. These factors could be influencing the decision, although distance from roads is the most influential variable.

The land change model might be improved by using a naïve model that simply predicts gain of agriculture adjacent to existing agriculture or that simply predicts gain of agriculture near previous gain of agriculture. It is possible that the use of the variables in the model other than distance form agriculture is just confusing the issue. The quantity of agriculture in 2020 could have been predicted using the calibration time interval 2000-2015 which would result in a smaller prediction of gain. Our understanding of the patterns of agricultural LULC change in Zambia could be improved by an analysis of which other LULC categories agriculture is gaining from. Analyses of specific sections of Zambia, especially those with the most agriculture, could also be useful. When 2020 agriculture LULC data are available, the accuracy of the 2020 prediction could be assessed.

Chapter 5, Conclusions:

There is a net gain of agriculture between 2000 and 2010, and then a four times faster net gain between 2010 and 2015. The accuracy assessment indicates that the magnitude of gain is likely not that great in reality. The northeast of Zambia, which is known for having the most chitemene shifting cultivation, has the most 6km pixels with high proportions of toggle in agriculture. This indicates that toggle could be useful in identifying areas of shifting cultivation. Districts of Zambia with more agriculture tend to have medium gain intensities but low loss intensities, which is more consistent with settled agriculture. Districts of Zambia with less agriculture tend to have higher gain and loss intensities, which is more consistent with shifting agriculture. The central and southeastern districts in Zambia have the highest concentration of agriculture at all three dates and have the largest total gains of agriculture. Distance from existing agriculture is the dominantly important variable in predicting gain of agriculture. Other measurements of proximity to settled areas were less strongly related to gain of agriculture. Physical and environmental variables were not found to be related to gain of agriculture. Using only distance from agriculture gives a calibration fit of 1 percentage point less than the fit using all variables for the transition potential modeling part of the LCM. This reliance on distance from agriculture caused the model to miss a substantial area of new agriculture in north-central Zambia in the 2015 prediction. The model tended to predict gains of agriculture in the right general areas, but not necessarily the right exact location. This is evidenced by the clustering of hits, misses and false alarms in the central area of Zambia as well as the exchange categories being larger than the hits category. There was an under prediction of the amount of gain of agriculture because the rate of gain of agriculture increased form the calibration time interval to the validation time interval. The 2015 prediction over predicting expansion of agriculture in the northeast calls into question the reliability of the 2020 model’s prediction of expansion of agriculture in the northeast.

Acknowledgements:

Priscilla Baltezar has helped with this project since we were working together on a final project for GIS & Map Comparison, where I learned many of the techniques used in this thesis. Professors Lyndon Estes, Gil Pontius and John Rogan have provided guidance on this thesis and many other projects I have undertaken at Clark University.

Citations:

Ali, H., Descheemaeker, K., Steenhuis, T. S., & Pandey, S. (2011). Comparison of Landuse and Landcover Changes, Drivers and Impacts for a Moisture-Sufficient and Drought-Prone Region in the Ethiopian Highlands. *Expl Agric.*, *47*(1), 71–83. https://doi.org/10.1017/S0014479710000840

Ampofo, S., Sackey, I., & Ampadu, B. (2016). Ethiopian Journal of Environmental Studies & Management, *9*(4), 412–429.

Araki, S. (2007). Ten Years of Population Change and the Chitemene Slash-and-Burn System around the Mpika Area , Northern Zambia. *African Study Monographs*, *34*(Supplementary Issue), 75–89.

Baltezar, P., Crema, S., Baldwin, E., Ye, S., Crowley, M., Cardille, J., & Estes, L. (2020). *Support Vector Machine (SVM) Time Interval Mapping of Agriculture : A case study in Zambia Agroecological (AER) Zone IIa*. Clark University.

Calzada, L., Meave, J. A., Bon, C., & Figueroa, F. (2018). Lands at risk : Land use / land cover change in two contrasting tropical dry regions of Mexico. *Applied Geography*, *99*, 22–30. https://doi.org/10.1016/j.apgeog.2018.07.021

Estes, L., Searchinger, T., Tian, D., & Kehoe, L. (2016). Reconciling agriculture, carbon, and biodiversity in a savanna transformation frontier. *Philosophical Transactions of The Royal*

Society B Biological Sciences, (April 2017), 18. https://doi.org/10.1098/rstb.2015.0316

FAO (2016). Zambia, Country Profile. Retrieved from: http://www.fao.org/countryprofiles/index/en/?iso3=ZMB.

FAOSTAT (2020). Data, Land Cover. FAO. Retrieved from: http://www.fao.org/faostat/en/#data/LC

Gashaw, T., Bantider, A., & Mahari, A. (2014). Population dynamics and land use/land cover changes in Dera District, Ethiopia. *Global Journal of Biology, Agriculture & Health Sciences*, *3*(1), 137–140.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of the Sciences*, *107*(38), 1–6. https://doi.org/10.1073/pnas.0910275107

Id, J. S., Bunting, E., Zhu, L., Ryan, S. J., Id, H. V. H., Id, D. K., … Campo-besco, M. A. (2018). Using a coupled dynamic factor – random forest analysis (DFRFA) to reveal drivers of spatiotemporal heterogeneity in the semi-arid regions of southern Africa. *PLoS ONE*, *13*(12), 1–19.

Kamwi, J. M., Chirwa, P. W. C., Manda, S. O. M., Graz, P. F., & Katsch, C. (2015). Livelihoods, land use and land cover change in the Zambezi Region, Namibia. *Population and Environment*, *37*(2), 203–230.

Kapekele, E. M. (2006). *Determinants of slash and burn: The case of chitemene farming system in Zambia*. University of Pretoria.

Kindu, M., Schneider, T., Tekatay, D., & Knoke, T. (2015). Drivers of land use/land cover changes in Munessa-Shashemene landscape of the south-central highlands of Ethiopia. *Environmental Monitoring and Assessment*, *187*, 452–469. https://doi.org/10.1007/s10661-

015-4671-7

Lambin, E. F., & Meyfroidt, P. (2011). Global land use change , economic globalization , and the looming land scarcity. *Proceedings of the National Academy of the Sciences*, *108*(9), 3465–3472. https://doi.org/10.1073/pnas.1100480108

Msofe, N. K., Sheng, L., & Lyimo, J. (2019). Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain , Southeastern Tanzania. *Sustainability*, *11*, 505–32. https://doi.org/10.3390/su11020505

Munthali, M. G., Davis, N., Adeola, A. M., Botai, J. O., Kamwi, J. M., Chisale, H. L. W., & Orimoogunje, O. O. I. (2019). Local Perception of Drivers of Land-Use and Land-Cover Change Dynamics across Dedza District , Central Malawi Region. *Sustainability*, *11*, 832–857. https://doi.org/10.3390/su11030832

Petit, C., Scudder, T., & Lambin, E. (2001). Quantifying processes of land-cover change by remote sensing : Resettlement and rapid land- cover changes in south-eastern Zambia. *Internation Journal of Remote Sensing*, *22*(17), 3435–3456. https://doi.org/10.1080/01431160010006881

Pontius Jr., R. G. (2019). Component intensities to relate difference by category with difference overall. *Int J Appl Earth Obs Geoinformation*, *77*, 94–99. https://doi.org/10.1016/j.jag.2018.07.024

Pontius Jr., R. G., Krithivasan, R., Sauls, L., Yan, Y., & Sauls, L. (2017). Methods to summarize change among land categories across time intervals. *Journal of Land Use Science*, *12*(4), 218–230. https://doi.org/10.1080/1747423X.2017.1338768

Pontius Jr., R. G., & Santacruz, A. (2014). Quantity , exchange , and shift components of difference in a square contingency table. *Internation Journal of Remote Sensing*, *35*(21), 7543–7554. https://doi.org/10.1080/2150704X.2014.969814

Pontius Jr., R. G., & Si, K. (2014). The total operating characteristic to measure diagnostic ability for multiple thresholds. *International Journal of Geographical Information Science*, *28*(3), 570–583. https://doi.org/10.1080/13658816.2013.862623

Pontius Jr., R. G. & Malanson, J. (2005). Comparison of the Structure and Accuracy of Two Land Change Models. *International Journal of Geographical Information Science*, 19(2), 243-265.

Quan, B., Pontius Jr., R. G., & Song, H. (2019). Intensity Analysis to communicate land change during three time intervals in two regions of Quanzhou City , China. *GIScience & Remote Sensing*, 1–16. https://doi.org/10.1080/15481603.2019.1658420

Searchinger, T. D., Estes, L., Thornton, P. K., Beringer, T., Notenbaert, A., Rubenstein, D., … Herrero, M. (2015). Converting Africa's Wet Savannahs to Cropland. *Nature Climate Change*, *5*(April). https://doi.org/10.1038/NCLIMATE2584

Stehfest, E., Doelman, J., & Mandryk, M. (2015). *Assessment of global land availability: land supply for agriculture* (FOODSECURE No. 7).

Vinya, R., Syampungani, S., Kasumu, E.C., Monde, C. & Kasubika, R. (2011). Preliminary Study on the Drivers of Deforestation and Potential for REDD+ in Zambia. A consultancy report prepared for Forestry Department and FAO under the national UN-REDD+ Programme Ministry of Lands & Natural Resources. Lusaka, Zambia

World Development Indicators (2020). The World Bank Group. Retrieved from: https://datacatalog.worldbank.org/dataset/world-development-indicators

Yalew, S. G., Griensven, A. Van, Mul, M. L., & Zaag, P. Van Der. (2016). Land suitability analysis for agriculture in the Abbay basin using remote sensing , GIS and AHP techniques. *Modeling Earth Systems and Environment*, *2*(2), 1–14. https://doi.org/10.1007/s40808-016-

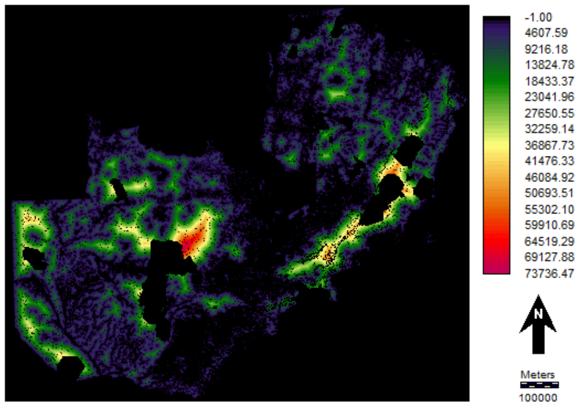
0167-x

Zambia CountrySTAT (2016). Central Statistical Office and Ministry of Agriculture. Retrieved from: http://zambia.countrystat.org/home/en/

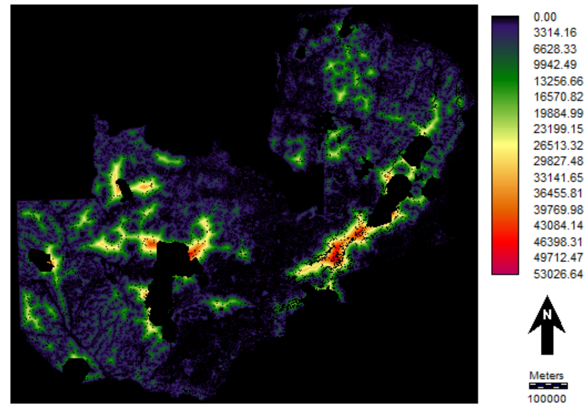
Appendices:

1 Driver Variables:

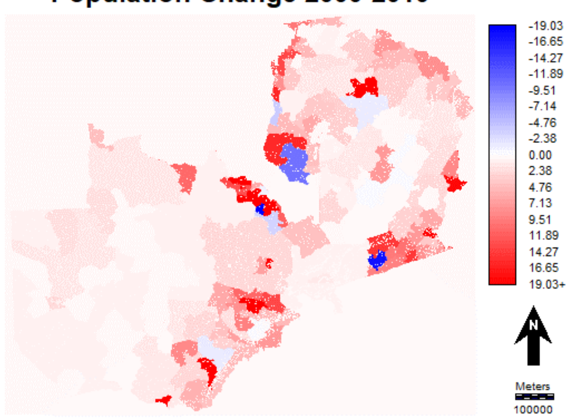
Distance From Agriculture in 2000



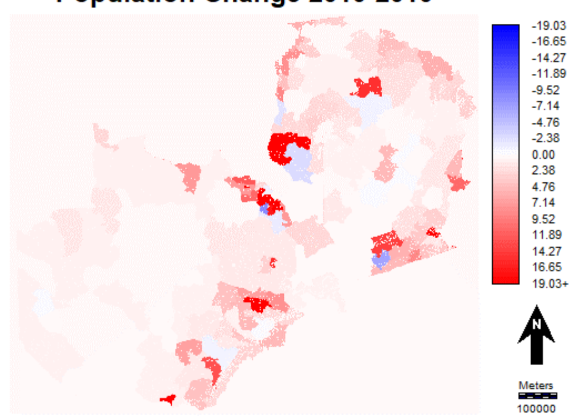
Distance From Agriculture 2010



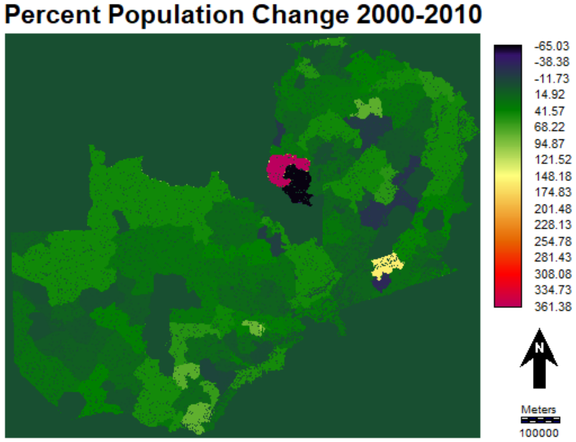
Population Change 2000-2010



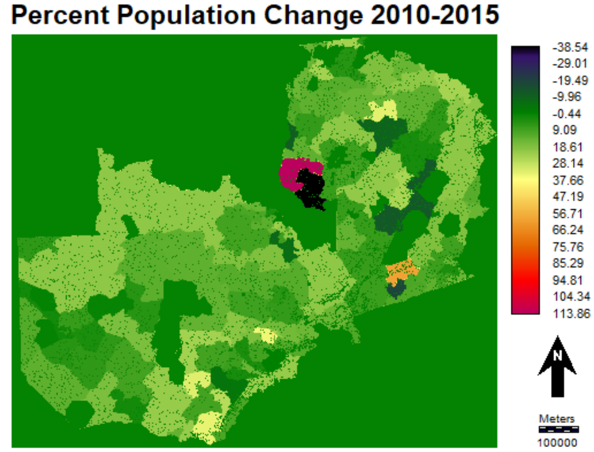
Population Change 2010-2015



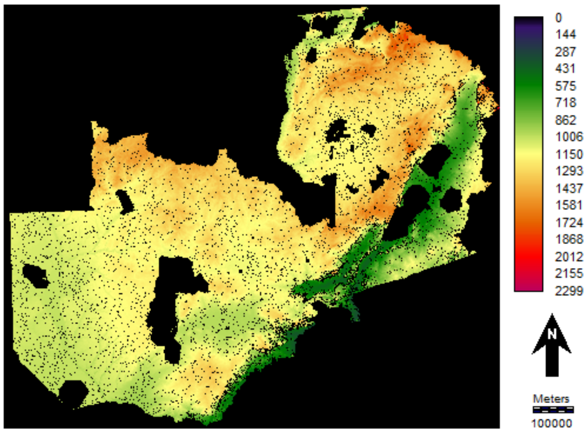
Percent Population Change 2000-2010



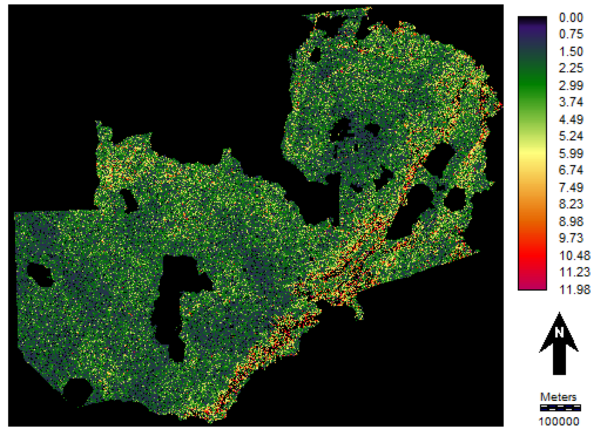
Percent Population Change 2010-2015



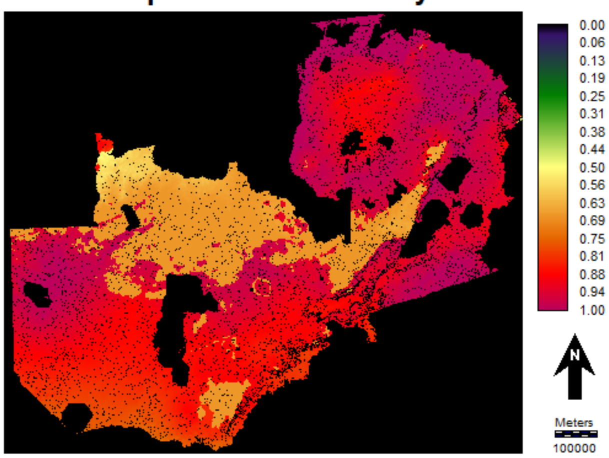
Elevation



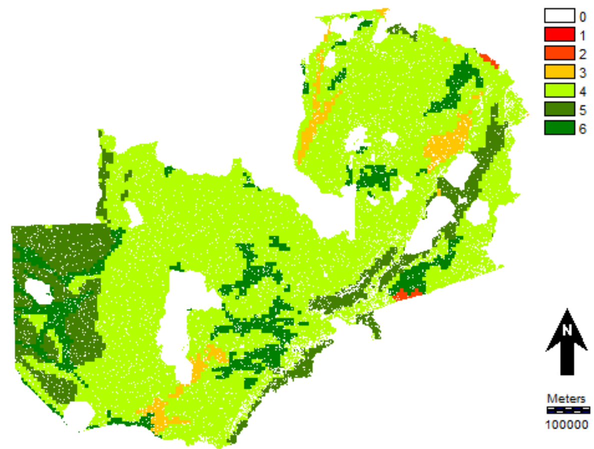
Slope in Degrees



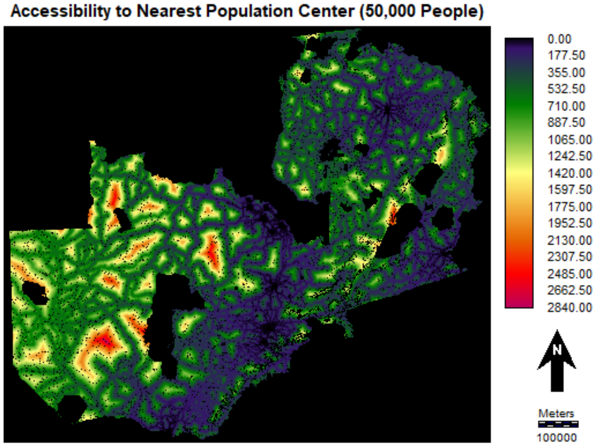
Crop Climate Suitability



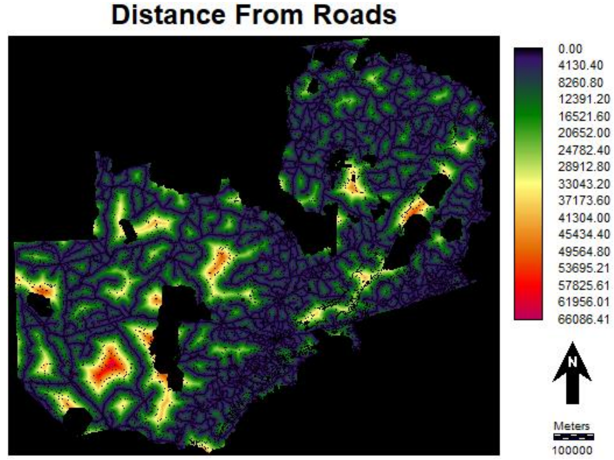
FAO Soil Productivity Index



Accessibility to Nearest Population Center (50,000 People)



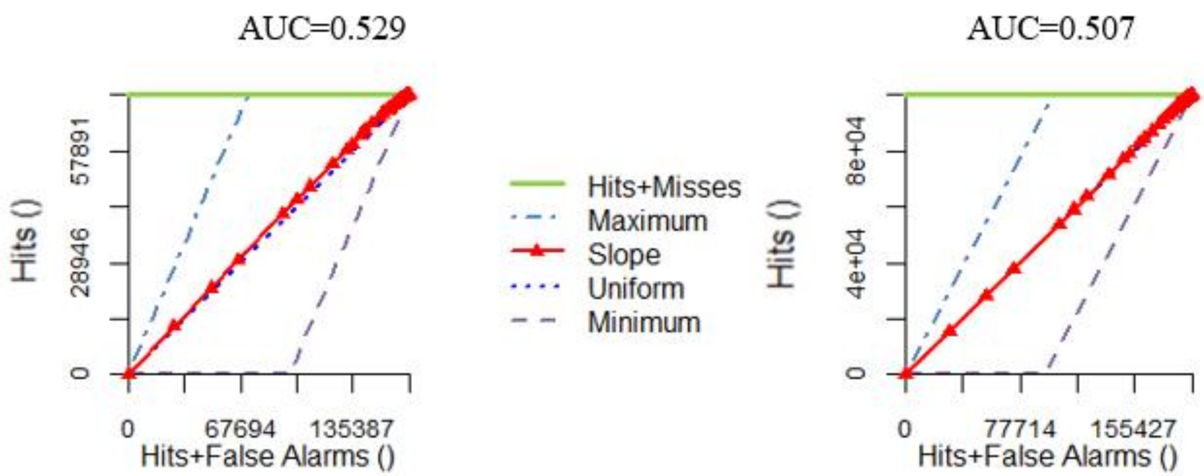
Distance From Roads



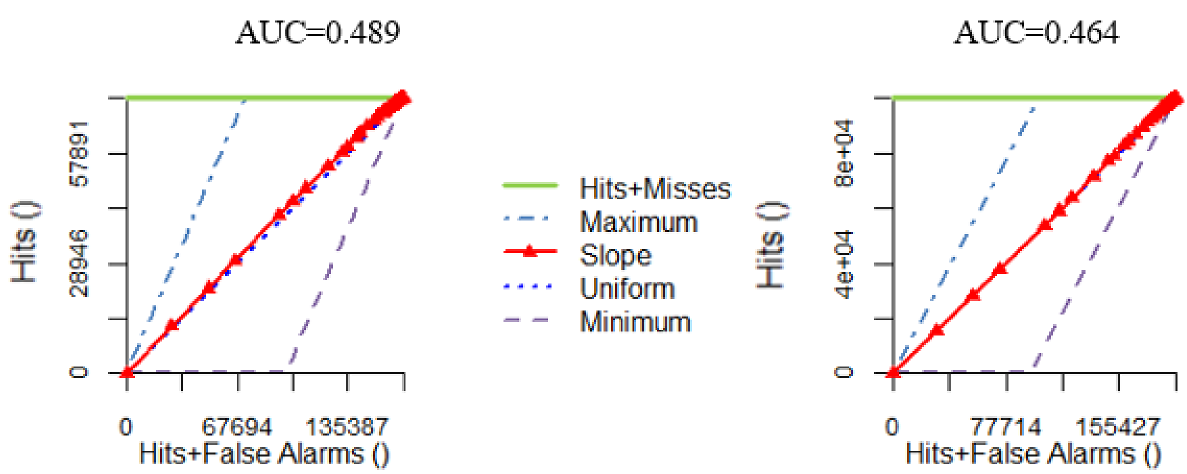
2. TOC curves not in the main section:

2000-2010 is in the left column, 2010-2015 in the right

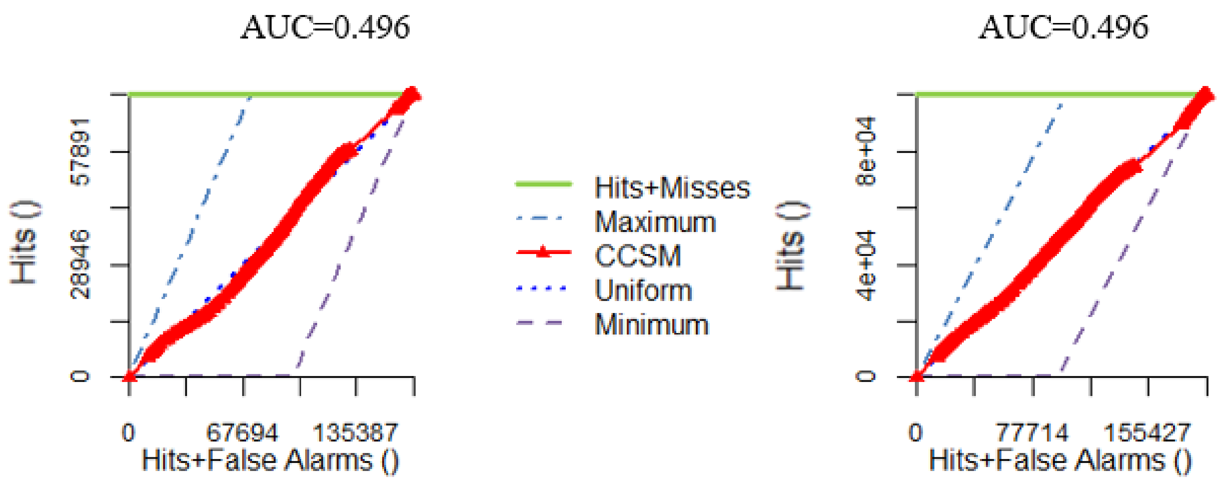
Slope:



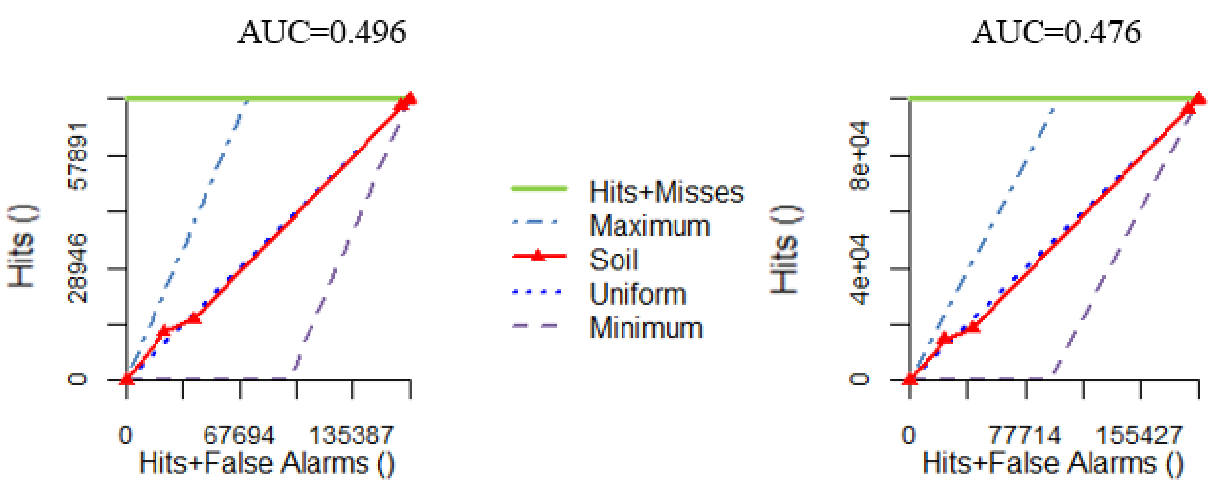
Elevation:



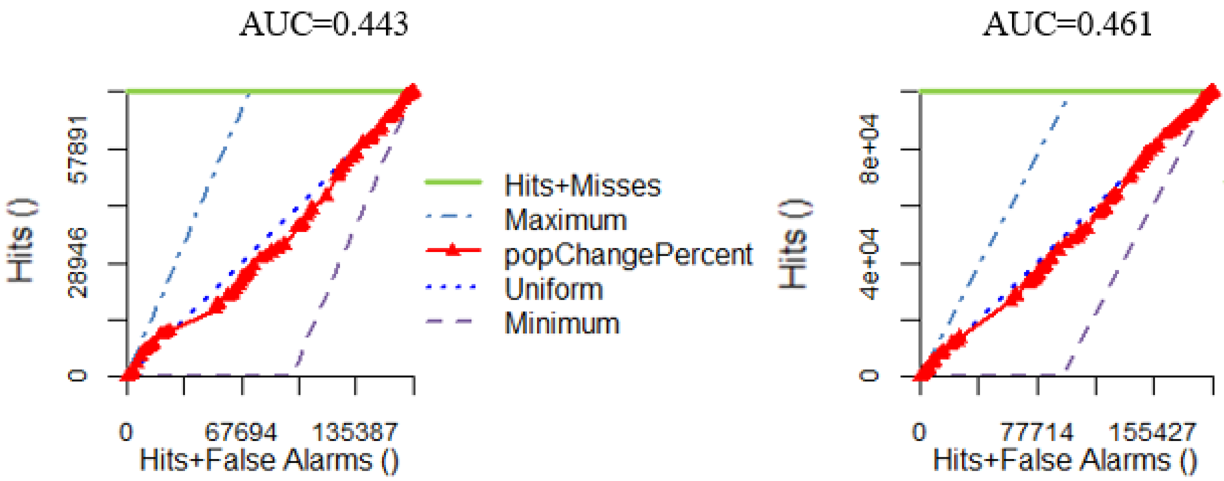
Crop Climate Suitability Modeler Index:



Soil Index:



Population (percent) change:



3. LCM MLP reports:

2000 and 2010 using all variables:

1) Input Files

Independent variable 1	acc_50k_processed2_ln
Independent variable 2	soil_prod_processed2_EL
Independent variable 3	ZAM_distFromAg_2000_processed_ln
Independent variable 4	pop_00_10_lmDiff_processed2
Independent variable 5	pop_00_10_pChange_processed2
Independent variable 6	ccsm_3crop_min_avgSuit_processed2
Independent variable 7	DEM_processed2
Independent variable 8	slope_processed2
Independent variable 9	dist_from_roads_processed_ln
Training site file	ZAM_ag_00_10_Train_non-a_to_agric

2) Parameters and Performance

Input layer neurons	9
Hidden layer neurons	7
Output layer neurons	2
Requested samples per class	10000
Final learning rate	0.0001
Momentum factor	0.5
Sigmoid constant	1
Acceptable RMS	0.01
Iterations	10000
Training RMS	0.3553
Testing RMS	0.3504
Accuracy rate	83.09%
Skill measure	0.6618

1) Forcing a Single Independent Variable to be Constant

Model	Accuracy (%)	Skill measure	Influence order
With all variables	83.09	0.6618	N/A
Var. 1 constant	82.98	0.6596	2
Var. 2 constant	83.12	0.6624	9 (least influential)
Var. 3 constant	64.31	0.2863	1 (most influential)
Var. 4 constant	83.12	0.6624	8
Var. 5 constant	83.05	0.6610	6
Var. 6 constant	83.12	0.6624	7
Var. 7 constant	83.01	0.6602	5
Var. 8 constant	83.01	0.6602	4
Var. 9 constant	83.00	0.6600	3

3) Backwards Stepwise Constant Forcing

Model	Variables included	Accuracy (%)	Skill measure
With all variables	All variables	83.09	0.6618
Step 1: var.[2] constant	[1,3,4,5,6,7,8,9]	83.12	0.6624
Step 2: var.[2,5] constant	[1,3,4,6,7,8,9]	83.22	0.6644
Step 3: var.[2,5,4] constant	[1,3,6,7,8,9]	83.19	0.6638
Step 4: var.[2,5,4,7] constant	[1,3,6,8,9]	83.18	0.6636
Step 5: var.[2,5,4,7,6] constant	[1,3,8,9]	83.21	0.6642
Step 6: var.[2,5,4,7,6,1] constant	[3,8,9]	83.10	0.6620
Step 7: var.[2,5,4,7,6,1,9] constant	[3,8]	82.96	0.6592
Step 8: var.[2,5,4,7,6,1,9,8] constant	[3]	82.74	0.6548

2000 and 2010 using only 4 variables:

1) Input Files

Independent variable 1	acc_50k_processed2_ln
Independent variable 2	ZAM_distFromAg_2000_processed_ln
Independent variable 3	pop_00_10_lmDiff_processed2
Independent variable 4	dist_from_roads_processed_ln
Training site file	ZAM_ag_00_10_Train_non-a_to_agric

2) Parameters and Performance

Input layer neurons	4
Hidden layer neurons	3
Output layer neurons	2
Requested samples per class	10000
Final learning rate	0.0001
Momentum factor	0.5
Sigmoid constant	1
Acceptable RMS	0.01
Iterations	10000
Training RMS	0.3591
Testing RMS	0.3584
Accuracy rate	82.20%
Skill measure	0.6439

1) Forcing a Single Independent Variable to be Constant

Model	Accuracy (%)	Skill measure	Influence order
With all variables	82.20	0.6439	N/A
Var. 1 constant	81.92	0.6383	2
Var. 2 constant	60.59	0.2118	1 (most influential)
Var. 3 constant	82.18	0.6435	4 (least influential)
Var. 4 constant	82.01	0.6401	3

3) Backwards Stepwise Constant Forcing

Model	Variables included	Accuracy (%)	Skill measure
With all variables	All variables	82.20	0.6439
Step 1: var.[3] constant	[1,2,4]	82.18	0.6435
Step 2: var.[3,4] constant	[1,2]	82.05	0.6409
Step 3: var.[3,4,1] constant	[2]	81.57	0.6315

2010 and 2015 using all variables.

1) Input Files

Independent variable 1	acc_50k_processed2_ln
Independent variable 2	ccsm_3crop_min_avgSuit_processed3
Independent variable 3	dist_from_roads_processed_ln
Independent variable 4	ZAM_distFromAg_2010_processed_ln
Independent variable 5	DEM_processed2
Independent variable 6	pop_10_15_lmDiff_processed2
Independent variable 7	pop_10_15_pChange_processed2
Independent variable 8	slope_processed2
Independent variable 9	soil_prod_processed2
Training site file	zam_ag_10_15_Train_Non-a_to_Agric

2) Parameters and Performance

Input layer neurons	9
Hidden layer neurons	4
Output layer neurons	2
Requested samples per class	10000
Final learning rate	0.0001
Momentum factor	0.5
Sigmoid constant	1
Acceptable RMS	0.01
Iterations	10000
Training RMS	0.2896
Testing RMS	0.2886
Accuracy rate	88.54%
Skill measure	0.7707

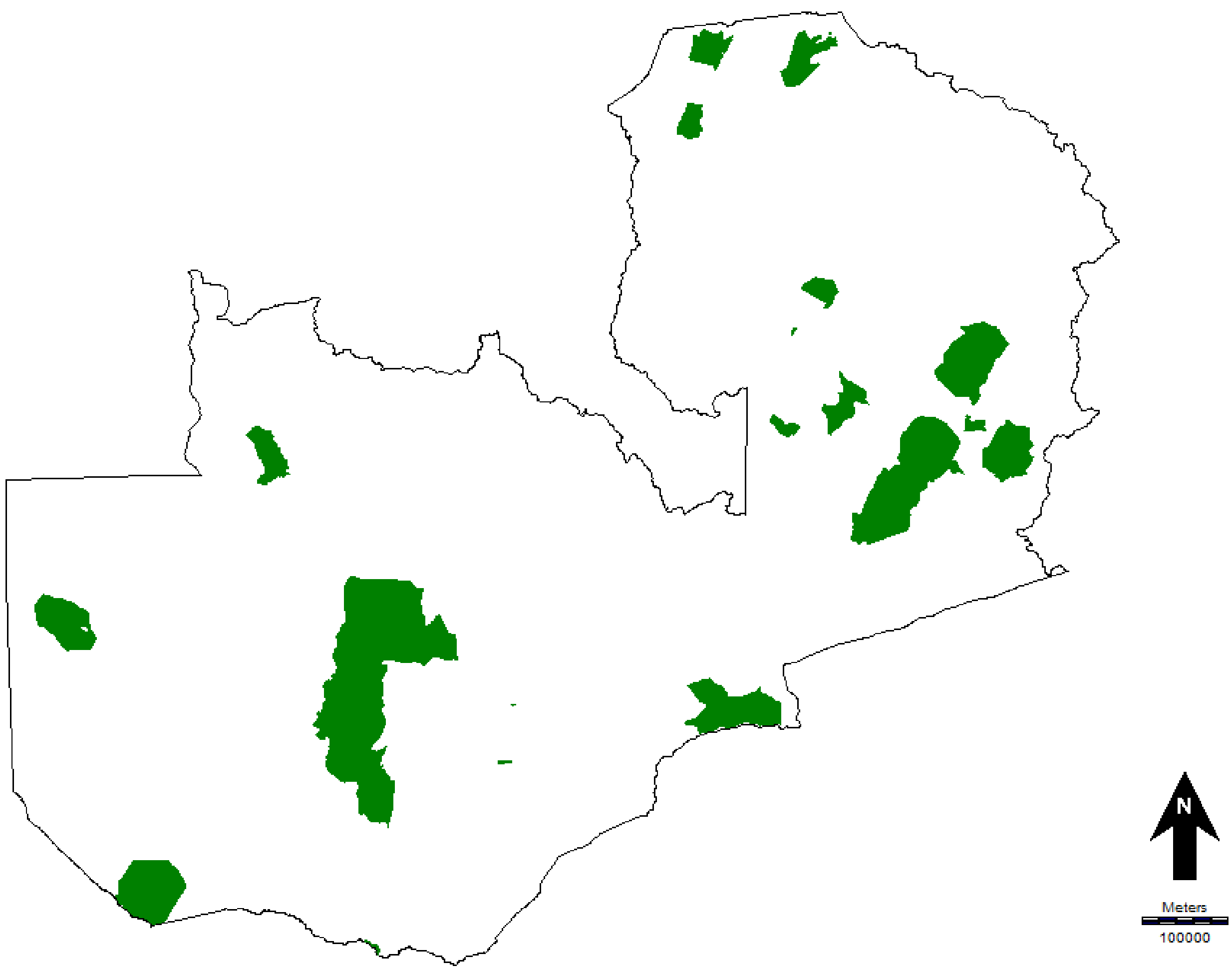
1) Forcing a Single Independent Variable to be Constant

Model	Accuracy (%)	Skill measure	Influence order
With all variables	88.54	0.7707	N/A
Var. 1 constant	88.41	0.7683	2
Var. 2 constant	88.70	0.7739	9 (least influential)
Var. 3 constant	88.62	0.7723	8
Var. 4 constant	52.43	0.0487	1 (most influential)
Var. 5 constant	88.59	0.7717	7
Var. 6 constant	88.43	0.7687	3
Var. 7 constant	88.56	0.7711	5
Var. 8 constant	88.58	0.7715	6
Var. 9 constant	88.45	0.7691	4

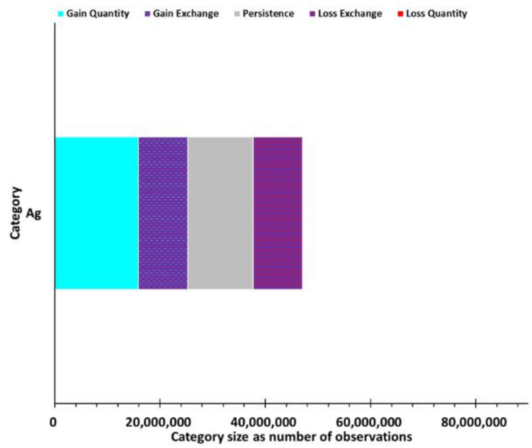
3) Backwards Stepwise Constant Forcing

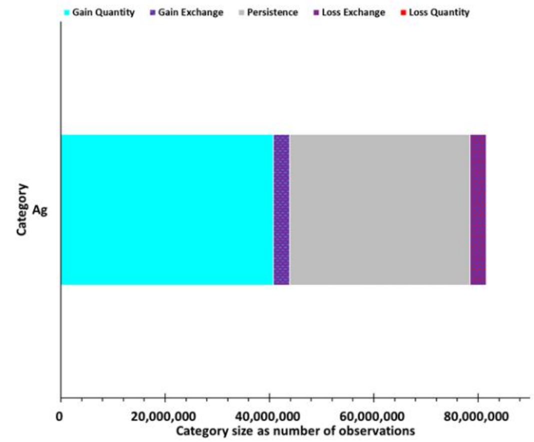
Model	Variables included	Accuracy (%)	Skill measure
With all variables	All variables	88.54	0.7707
Step 1: var.[2] constant	[1,3,4,5,6,7,8,9]	88.70	0.7739
Step 2: var.[2,7] constant	[1,3,4,5,6,8,9]	88.70	0.7739
Step 3: var.[2,7,8] constant	[1,3,4,5,6,9]	88.71	0.7741
Step 4: var.[2,7,8,5] constant	[1,3,4,6,9]	88.69	0.7737
Step 5: var.[2,7,8,5,3] constant	[1,4,6,9]	88.62	0.7723
Step 6: var.[2,7,8,5,3,6] constant	[1,4,9]	88.57	0.7713
Step 7: var.[2,7,8,5,3,6,9] constant	[1,4]	88.44	0.7689
Step 8: var.[2,7,8,5,3,6,9,1] constant	[4]	88.51	0.7701

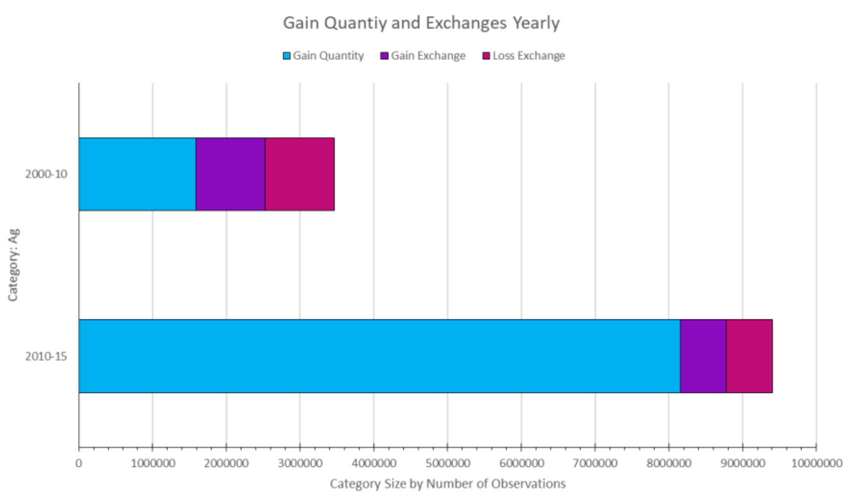
4. National Protected Areas of Zambia



5. Gain Quantity, exchange and persistence. The first time interval is on the left, the second on the right. Second row is average yearly gain quantity and exchanges:







The amount of agriculture in 2000 is the union of the persistence and the loss exchange (and the loss quantity, which in this case is 0). The amount of agriculture in 2010 is the union of the persistence, gain exchange and gain quantity. The Non-agriculture category is not shown here, because the only different information it contains is the large amount of non-agriculture persistence.

6. Area of Agriculture in Zambia from 2001-2015 from FAO stat based on MODIS data.