**Clark University**
# Clark Digital Commons

International Development, Community and Environment (IDCE)

Master's Papers

5-2016

# Land Change Modeling: Summer 2015 Internship with Clark Labs

Xiaoyang Kong
*GISDE*, xikong@clarku.edu

Follow this and additional works at: https://commons.clarku.edu/idce_masters_papers

Part of the Earth Sciences Commons, and the Environmental Studies Commons

## Recommended Citation

# LAND CHANGE MODELING:

# SUMMER 2015 INTERNSHIP WITH CLARK LABS

**XIAOYANG KONG**

May 2016

Submitted to the faculty of Clark University, Worcester,

Massachusetts, in partial fulfillment of the requirements for

the degree of Master of Science in the department of International Development,

Community, and Environment

And accepted on the recommendation of

Yelena Ogneva-Himmelberger, Ph.D. Chief Instructor

# ABSTRACT

Land Change Modeling: Summer 2015 Internship with Clark Labs

Xiaoyang Kong

My internship with Clark Labs was completed during the summer of 2015, where I worked under the direction of Dr. Ronald Eastman. As a research assistant, the responsibilities of my internship was to process data for a project that intended to develop a web application for land change modeling. I learned a lot of knowledge about land change science and GIS applications from this internship. The experience I gained will definitely increase my competitiveness in my future career. I am fully satisfied with this internship and I would recommend this internship to anybody interested in GIS and land change science. The following chapters provide details about the organization, my responsibilities and an overall assessment of the internship.

_____

Yelena Ogneva-Himmelberger ,Ph.D.

Chief Instructor

# ACADEMIC HISTORY

**Name**: Xiaoyang Kong                                    **Date**: May 2016

**Place of Birth:** Rongchen, China                  **Date:** January 30<sup>th</sup> 1990

**Baccalaureate School:** Clark University        **Date:** Auguest, 2014

**Baccalaureate Subject:** B.A. Geography

# DEDICATION

For my parents, who have always encouraged me to pursue my academic goals.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

Geography has been my favorite subject since I was a middle school student in China, so I came to Clark University which has one of the best undergraduate programs in geography in the US to pursue my bachelor degree in geography. During my four years study period at Clark University, I gradually started to know about and get interested in geographic information science (GIS). Therefore, I decided to focus my studies on GIS and I was led to the Geographic Information Science for Development and Environment (GISDE) Master's program at Clark University. Graduation from this program requires the fulfillment of a GIS-related internship and I completed my internship at Clark Labs at Clark University.

My internship with Clark Labs took place in the summer of 2015 from June to August. While working in this research organization, I was able to utilize my GIS skills to do data processing for a project to develop a web application of land change. I also learned a lot about the professional world of GIS and land change science. This internship was undoubtedly a very useful learning experience for me and helped me to define my career path.

This internship paper is organized into several chapters. Chapter 2 describes the organization where I did my internship, including its mission, structure and strengths. Chapter 3 provides details on my internship responsibilities about the methodologies I used to complete my work. Chapter 4 describes the skills I learned from the internship and how I utilize the skills I learned from Clark University on the internship.

# CHAPTER 2: DESCRIPTION OF ORGANIZATION

Clark Labs, founded in 1987, is a non-profit organization dedicated to the study and evolution of geospatial technologies for efficient and responsible solution finding for environmental administration, sustainable resource development and impartial resource allocation.

Clark Labs provides professional services to meet a wide range of needs, from customized software development, analytical service, application research and training. Clark Labs' services leverage over twenty-five years of GIS and earth science expertise. Based in Clark University, Clark Labs brings unmatched skills to solve most important and demanding issues. Clark Labs has been actively developing geographic information system (GIS) software and its renowned product, IDRISI GIS and Image Processing software, is widely used by the professionals in a broad range of different industries not only in America, but also around the world in more than 180 countries. In 2015, Clark Labs released TerrSet Geospatial Monitoring and Modeling software which includes Image Processing tools of IDRISI along with additional constellation of software tools for monitoring and modeling the earth system.

Clark Labs is known for pioneering advancements in areas such as decision support, uncertainty management, classifier development, land change science, climate change monitoring and time series analysis. Clark Labs partners with a lot of organizations such as The Gordon and Betty Moore Foundation, Google.org, USDA, the United Nations, Conservation International and World Conservation Society and leverages its training base to develop and renew customized research tools which provide software solutions for the organizations in need and deploy the geospatial expertise to a variety of the world problems (Directions Magazine, 2015).

Earth system science is one of the most important research fields of Clark Labs. In order to provide tools for the research analysis, Clark Labs, with support from the Gordon and Betty Moore Foundation, developed the Earth Trends Modeler, a tool within TerrSet for monitoring the earth system based on time series analysis. Earth Trends Modeler is a very unique tool for Clark Labs to extract and analyze global climate trends and their impacts, examine the seasonal trends and recurrent climate patterns in space and time. No other software technology provides such a

coordinated suite of data mining tools needed by the earth system science community for climate change analysis and impact assessment (Clark Labs).

Clark Labs also focus on research of land change analysis and worked with Conservation International to develop a tool for modeling land change. In 2006, a revolutionary land planning and decision support tool, the Land Change Modeler, was released within the IDRISI software. Land Change Modeler allows user to model land cover transition potential and predict the quantity of change and where the change will happen. Land Change Molder is one of the most comprehensive tool for land change analysis and it provides a start-to-finish solution for land planning (Clark Lab). In 2007, Clark Labs developed the Land Change Modeler as an extension for ArcGIS which allow more users to have accesses to use this powerful tool for their land change analysis needs.

Recent projects of Clark Labs include the mapping of coastal marine aquaculture, land changing modeling and prediction, species modeling and mapping biodiversity offsets, detection of teleconnections of earth trends such as spatial and temporal analysis of climate cycles, carbon modeling of REDD (Reducing Emissions from Deforestation and Forest Degradation) programs, the detection of diseased trees, using hyper-spectral imagery, and the predictive modeling of invasive species using neutral networks (Clark Labs).

Clark Labs is headed by Professor Ronald Eastman who is the Director of Clark Labs and the prime developer and chief architect of IDRISI software. Other main staff members of Clark Labs include James Toledano, Exec Director of Clark Labs, Stefano Crema, Research Analyst, Tammy Woodard, Programmer, Hao Chen, Senior Researcher and Diane Sutter, Customer Service Manager. My position as the data analyst intern reported directly to Prof. Eastman. Since Clark Labs offers many internship opportunities to students of Clark University every year, about half

of the staff of Clark Labs are student interns and the joining and leaving of short-time student interns contribute to a variable staff composition of Clark Labs.

## CHAPTER 3: INTERNSHIP RESPONSIBILITIES

Clark Labs is currently working on a project to develop a cloud-based Land Change Modeler (LCM). This project is sponsored by ESRI and two final products will be created through this project. One of them is a paid version of the LCM based on ArcGIS online. Another one is a free web application of the LCM called Land Change Explorer (LCE) which is designed to provide an approach to let users to know about the basic functions of the LCM and attract their interests to buy the paid version of the LCM. All the data of the LCE are pre-stored. The main role of my internship was to do data processing for the LCE and the work can be divided into three steps: data pre-processing, calculating transition potentials and predicting land covers. Since the amount of data involved in the project was very large, I only completed a small part of the work during the summer and I continued to work on the project as a directed study with Prof. Ron Eastman in the following semester.

The first step of my work was data pre-processing. The LCE was designed to show land changes and their transition potentials in all the lower 48 states of the US from 2001 to 2011. Seven interns worked on the project in the summer, so each of us was assigned to work on several of the total 48 states. The states assigned to me included Colorado, Indiana, Maryland, South Carolina, Utah and Wisconsin. Table 1 shows the data we needed to download for each state. Other than some basic data processing work, like re-projection and transforming the format of the data, we spent most of time on developing variables for calculating transition potentials for each of the land cover transitions in each state during this step. Transition potential is used to measure the probability of land to go through a particular transition. It ranges from 0 to 1 and the

higher the value, the higher the probability the land to go through the transition. The variables used to calculate transition potential for a particular transition are driver variables that lead to the transition. For example, the transition potential for an area that changes from deciduous forest to high intensity developed area may be related to its distance to roads. The closer the area to roads, the higher the potential the area changes from deciduous forest to high intensity developed area. Then distance to roads would be a good driver variable for this transition. The variables used to calculate transition potentials were normally different for different areas, but ESRI thought that would be too complicated for users to use the LCM. Therefore, based on some previous researches, Prof. Eastman decided to use a same set of variables to calculate transition potentials for every state (see Table 2). All the variables were created based on the data we downloaded. It took us about 3 weeks to complete the data-preprocessing work for all the states.

The next step was to use the variables we created to calculate transition potentials for every state. One of the biggest advantages of the cloud-based LCM is that it utilizes multi-thread technique to increase the calculation speed. This is achieved by dividing the study area into separate parts and sending them to different computers to calculate transition potential for every part at the same time. Then the results for different parts are aggregated to get the transition potential for the whole study area. However, such technique cannot be realized based on traditional algorithms of LCM for calculating transition potential, like multi-layer perceptron (MLP) and logistic regression. Hence, Clark Labs developed a new algorithm called Weighted Normalized Likelihood (WNL) to calculate transition potentials for the cloud-based LCM.

Before we used WNL to model the transitions, we had done a test to compare the modeling accuracy between WNL and MLP. MLP is one of the most robust algorithms and it has the highest modeling accuracy among existing algorithms. For the test, we used WNL and MLP

5

to calculate transition potentials for three largest natural transitions (transitions from natural land covers to natural land covers) and three largest anthropogenic transitions (transitions from natural or developed land covers to developed land covers) from six states and then compared the result accuracy between the two algorithms. Based on our test, the average accuracy of WNL is very close to that of MLP and the accuracy of WNL for natural transitions is even a little bit higher than that of MLP which indicates that WNL is very effective and reliable.

After the test, we began to use the new algorithm to model transitions from 2001 to 2011 for every state. The variables used for the modeling are slightly different for different types of transition. Table 3 shows the input variables for different types of transition. At the very beginning, all the transitions are only divided into two types: natural transitions and anthropogenic transitions. However, we found that distance to cropland was an important variable for modeling transitions form other land cover types to cropland after we completed the modeling for several states, so we divided all transitions to cropland into one independent type and included Distance to Cropland in the input variables when we modeled this type of transition.  Another difference of the variables used for modeling different types of transitions was that we included County Subdivisions or Counties and Incorporated Places for anthropogenic transitions and not included it for natural transitions. This is because that sub-state level governments have powers for zoning in the state which could have a huge influence on anthropogenic land changes. The sub-state level governments that have powers for zoning could either be county subdivisions or counties and incorporated places, like cities, towns and villages, for different states. Therefore, we included county subdivisions in the input variables for states that county subdivisions have powers for zoning and included counties and incorporated places

in the input variables for states that counties and incorporated places have powers for zoning (see Table 4).

For each state, we only calculated transition potentials for transitions that were greater than 2 $km^2$ since calculating several small transitions is extremely time consuming and would not contribute to the effectiveness of our study.  Two kinds of output were created through this process for each transition. One of them was a transition potential image and Fig. 1 shows an example image of the transition potential from Woody Wetlands to Developed, Open Space in Florida which was one of the states assigned to me to work on during the summer. Based on Fig. 1, the areas with high transition potential values were close to developed areas along the coast line, so we could say that distance to developed area might be an important driver variable for this transition. Another output created through this process was a HTML file that contained summary statistics for the transition (see Fig. 2).  Skill, used to measure the accuracy of the transition potential, was a very important statistics in the file. In this case, the overall skill for the transition was 0.838 which was a very high value and indicated that the transition was modeled very well. Another important statistics was the importance weights for the input variables which were contained in the table at the bottom of the file (see Fig 2). According to the table, the importance weight for distance to high developed areas was 0.3994 which was the highest among the 9 variables. This indicated that distance to high developed areas was the most important driver variable that leaded to the transition in Florida. There were totally 105 transitions in Florida and I needed to model them one by one, so it was a very time consuming work. We only modeled transitions for 10 states during the summer and it was estimated that it would take several months to complete the work for the left states.

The last step would be predicting land covers in 2050 for all the states. By the end of the summer break, we had not started this step since we had to get the transition potentials first to predict the further land covers. The 2050 land covers will be predicted through Markov Chain analysis based on the land change data form 2001 to 2010. As long as the last two steps are finished, all the data will be stored into the LCE. Through the LCE, users can see the land changes from 2001 to 2011, see their transition potentials and predicted land covers in 2050 for all the lower 48 states.

## CHAPTER 4: INTERNSHIP ASSESSMENT

My internship with the Clark Labs was an invaluable learning experience that increased my knowledge of land change science and provided a great opportunity to apply what I learned from Clark University to real world issues. I was glad to work with several talented GIS scientists and developers which helped me to know about the current trends of GIS and what I need to learn to pursue the GIS work I am interested in in the future. The internship helped me to define the path and shape the goals of my future career.

I learned a lot during my summer at Clark labs, most of which was directly related to land change science. Land change has significant impacts on people and environment and research on land-use change has become an important element of global change research programs at national as well as international levels (Gutman). Therefore, the knowledge of land change science I learned through my internship was really valuable. I spent most of my time on modeling land cover transition potentials during the internship. The most important knowledge I learned was that input variables were a key element for modeling transition potentials. The results of modeling transition potentials were very sensitive to input variables. Even a small change in input variables might lead to a significant change in modeling accuracy and

importance weights of the input variables. For example, the variable of distance to cropland was not included in input variables, but we found that the modeling accuracy would increase a lot when we included it for transitions from other land covers to cropland after transitions for several states. Therefore, we decided to use the new variable to re-model all the transitions to cropland we had modeled before. Although Clark Labs had a lot of experience on land change analysis and did a lot of research for input variables of modeling transition potentials for this project before our modeling work, we still found some improvements that could be made on input variables during the modeling process. This indicated that land change science could not only depend on previous experience, exploration during the analysis process was also very important.

The skills I learned from Clark University helped me a lot to complete my summer internship. The majority GIS skills I used for my work were learned from the Advanced Vector GIS and Advanced Raster GIS courses. I regularly used tools in ArcGIS and TerrSet to do data processing for the project, such as buffer, project, and overlay. One of the most useful skills I learned from Clark University for the internship was how to use Macro Modeler in TerrSet. Macro Modeler is a graphical environment for building and executing multi-step models. Macro Modeler allows users to run more than one models at the same time and use the outputs of one model as the inputs for other models through iterations. There were a lot of repeated work I needed to do for the data. The use of Macro Modeler allowed me to do all the repeated work at once and I could do other works while the Macro Modeler was running. This saved a lot of time and significantly increased the work efficiency for me.

During the internship, the most useful skill I felt I need to learn in the future was programming skill. Developing the web application of LCM required a lot of programming work

and some GIS developers of Clark Labs were responsible for this work.  Some of the summer interns, including me, were interested in the programming work, however given our lack of programming experience, we were unable to contribute towards the team. Programming skills are not only useful for developing but also useful for data processing works. For example, the batch data processing achieved by the Macro Modeler I mentioned before could also be completed through programming and sometimes doing it through programming was even more convenient. In addition, I found some bugs in some of the new tools of TerrSet developed for the project when we used them. Since I was not involved in the developing work of the tools and I had not much programming knowledges, I could fix these problems by myself. I had to report the bugs to the developers and waited for them to fix the bugs so that I could continue my work. If I had had enough programming knowledges and could fix the bugs myself, it would increase the whole work efficiency a lot.

This internship was highly related to my studies since Clark Labs was a GIS-based research organization. Prof. Ran Eastman, the director of Clark Labs, was the instructor of advanced raster GIS class, so a lot of knowledges I learned from his class could be directly used for the internship. I would definitely recommend this internship to other GISDE students. The internship would provide a great opportunity to students to apply their skills of GIS and remote sensing. Clark Labs is a wonderful place to work and the staff are very friendly. Students who work there would have a chance to work with most talented GIS scientists and learned a lot from them.

## CHAPTER 5: CONCLUSION

My internship with Clark Labs was a great experience and I felt fortunate to work in a research organization with several talented GIS scientists. The LCE project was a great project

and would produce an extremely useful tool for people to analyze land change. I was honored to be involved in and made my own contributions to this project. During the internship, I learned a great deal about land change science and I got an opportunity to apply what I learned from Clark University in professional GIS world. I believe the knowledge and experience I gained from this internship will definitely increase my competitiveness in my future career. I am fully satisfied with this internship and I would recommend this internship to anybody interested in GIS and land change science.

# REFERENCES

*Clark Labs. 'Terrset Geospatial Monitoring And Modeling Software | Clark Labs'. N.p., 2015. Web. 29 Sept. 2015.*

*Directions Magazine. 'Clark Labs Receives Support From Google.Org To Develop On-Line Prototype Of Its Land Change Modeler Application To Be Run On Google Earth Engine'. N.p., 2015. Web. 29 Sept. 2015.*

Gutman, Garik. *Land Change Science: Observing, Monitoring and Undeerstanding Trajectories of Change on the Earth's Surface*. Vol. 1. Springer Science & Business Media, 2004.

**TABLES**

Table 1. Data needed to be downloaded for each state

| Data | Source |
|---|---|
| Land Cover | National Land Cover Database |
| National Elevation | ArcGIS Server |
| USA Surface Water | ArcGIS Server |
| Primary Road (TIGER/Line) | United States Census Bureau |
| Secondary Road (TIGER/Line) | United States Census Bureau |
| Local Road (TIGER/Line) | United States Census Bureau |
| Protected Areas | United States Geological Survey |
| County Subdivisions | United States Census Bureau |
| Counties and Incorporated Places | United States Census Bureau |

Table 2. Variables used to calculate transition potentials for each state

| Variables |
|---|
| Slope |
| Elevation |
| Distance to Water |
| Distance to Primary Roads |
| Distance to Secondary Roads |
| Distance to Local Roads |
| PAD (Protected Areas) |
| County Subdivisions |
| Counties and Incorporated Places |

Table 3.  Variables to be used in Modeling by Transition Type

| Natural Transition | Anthropogenic Transition | To Cropland |
|---|---|---|
| Slope | Slope | Slope |
| Elevation | Elevation | Elevation |
| Distance to Water | Distance to Water | Distance to Water |
| Distance to Primary Road | Distance to Primary Road | Distance to Primary Road |
| Distance to Secondary Road | Distance to Secondary Road | Distance to Secondary Road |
| Distance to Loacl Road | Distance to Loacl Road | Distance to Loacl Road |
| Distance to High intensity Developed | Distance to High intensity Developed | Distance to High intensity Developed |
| PAD (Protected Areas) | PAD (Protected Areas) | PAD (Protected Areas) |
|  | County Subdivisions or Counties and Incorporated Places | Distance to Cropland |

Table 4. List of States Modeled with County Subdivisions or Counties and Incorporated Places

| STATES WITH COUNTIES + INCORPORATED PLACES | STATES WITH COUNTY SUBDIVISIONS |
|---|---|
| Alabama | Arkansas |
| Arizona | Connecticut |
| California | Illinois |
| Colorado | Indiana |
| Delaware | Iowa |
| Florida | Kansas |
| Georgia | Louisiana |
| Idaho | Maine |
| Kentucky | Maryland |
| Montana | Massachusetts |
| Nevada | Michigan |
| New Mexico | Minnesota |
| Oklahoma | Mississippi |
| Oregon | Missouri |
| South Carolina | Nebraska |
| Texas | New Hampshire |
| Utah | New Jersey |
| Washington | New York |
| Wyoming | North Carolina |
|  | North Dakota |
|  | Ohio |
|  | Pennsylvania |
|  | Rhode Island |
|  | South Dakota |
|  | Tennessee |
|  | Vermont |
|  | Virginia |
|  | West Virginia |
|  | Wisconsin |

**FIGURES**
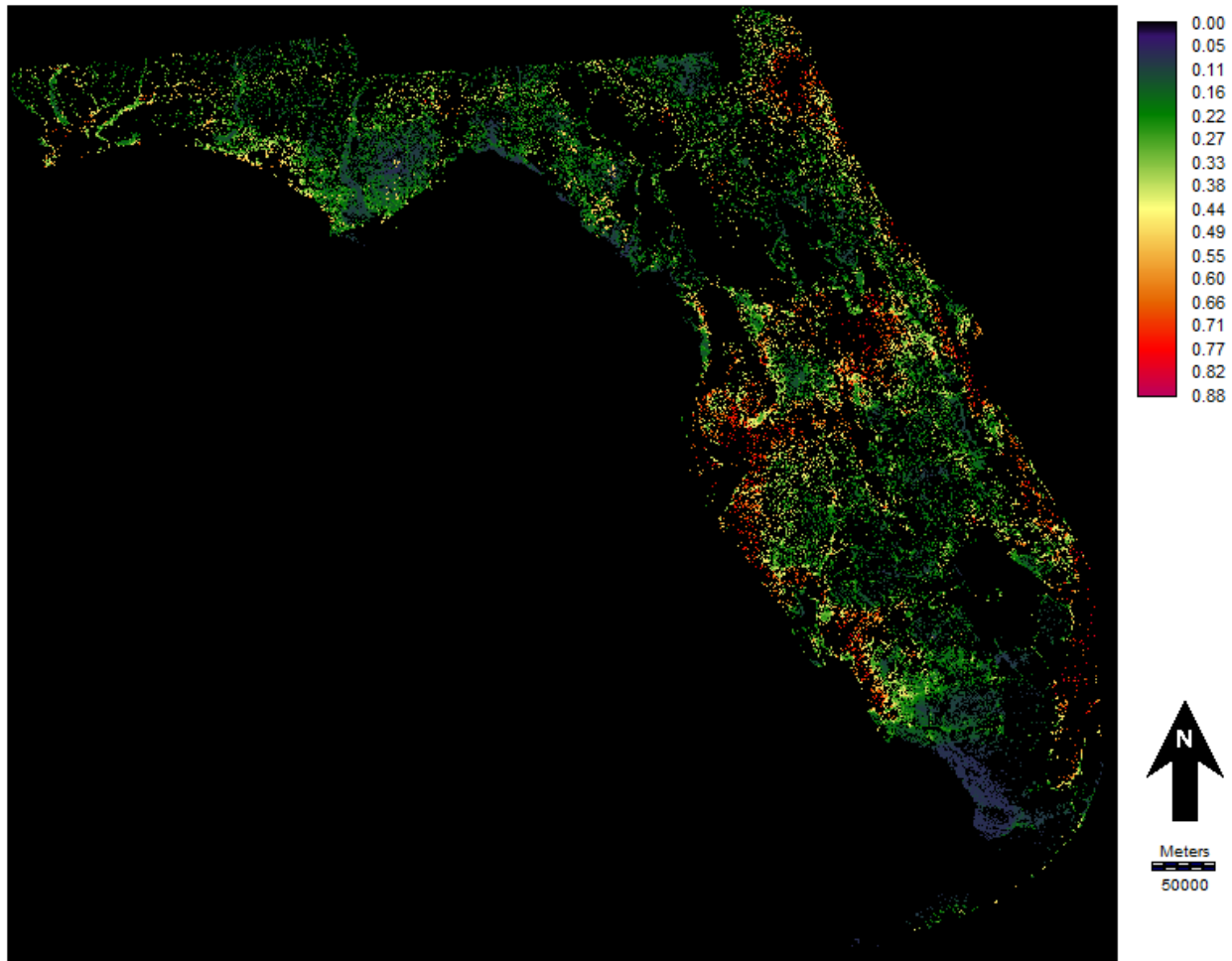


Fig. 1. Transition Potential from Woody Wetlands to Developed, Open Space in Florida

Summary statistics for FL_transition_potential_from_14_to_2

Accuracy for Change : 0.942580044269562

Accuracy for Persistence : 0.895158231258392

Total Accuracy : 0.918803930282593

Skill overall : 0.837607860565186

Skill for change : 0.885160088539124

Skill for persistence : 0.790316462516785

Total changed : 158406

Total persisted : 38979106

Training changed : 79531

Training persisted : 79554

Testing changed : 78875

Testing persisted : 79310

| variable | variable name | original weight | normalized weight | rank |
|---|---|---|---|---|
| 1 | S200_FL_Slope_sc_20150616_FINALMASK | 0.1117 | 0.0471 | 9 |
| 2 | S200_FL_Elev_sc_20150616_FINALMASK | 0.1147 | 0.0484 | 8 |
| 3 | S200_FL_DistWater_sc_20150617_FINALMASK | 0.2531 | 0.1068 | 6 |
| 4 | S200_FL_DistSecondary_sc_20150730_FINALMASK | 0.2684 | 0.1132 | 5 |
| 5 | S200_FL_DistPrimary_sc_20150730_FINALMASK | 0.1880 | 0.0793 | 7 |
| 6 | S200_FL_DistLocal_sc_20150730_FINALMASK | 0.3704 | 0.1562 | 2 |
| 7 | S200_FL_DistDev2001_THRESHFINAL | 0.3994 | 0.1685 | 1 |
| 8 | qual_FL_PADSTS_sc_20150617_FINALMASK | 0.3108 | 0.1311 | 4 |
| 9 | qual_FL_Counties_IncorporatedPlaces_FINALMASK | 0.3540 | 0.1493 | 3 |

Fig. 2. Summary Statistics for Transition Potential from Woody Wetlands to Developed, Open

Space in Florida