International Development, Community and
Environment (IDCE)

8-2016

Cloud-based Land Change Modeler Application: Summer Internship with Clark Labs

Lei Rong
lrong@clarku.edu

Cloud-based Land Change Modeler Application:

Summer Internship with Clark Labs

Lei Rong

June 2015

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, Academic Advisor

ABSTRACT

Cloud-based Land Change Modeler Application: Summer Internship with Clark Labs, Worcester, MA

Lei Rong

This report provides a detailed illustration of my internship experience with Clark Labs in Worcester, MA, during the summer of 2015. The internship is about two Land Change Modeler projects which conducted by Esri. The first one is a cloud-based Land Change Model on ArcGIS online for professional change analysts. Another one is a free cloud-based App for land change analysts and the public to explore land change of 48 states in United State. This work mainly relied on TerrSet skills and Land Change Modeler skills. As an GIS analyst, my responsibilities were data extraction, preparation, and modeling test. In June 6th, I started my internship at Clark Labs, and am continuing work till now. I think I gained a lot GIS and non-GIS experience during this time period. In an academic perspective, I got a full understanding of Land Change Modeler and TerrSet during this time period. Socially, I made a connection with students and staffs in Clark Labs. Overall, it is a great experience to me!

_____

Yelena Ogneva-Himmelberger ,Ph.D.
Chief Instructor

ACADEMIC HISTORY

Name: Lei Rong   **Date**: June 2015

Place of Birth: Changsha, Hunan   **Date:** October 24th 1991

Baccalaureate School: Beijing Normal University

Date: May, 2014

Baccalaureate Subject: B.Sc., Environment Science

Occupation and Academic Connection since date of baccalaureate degree:

Clark University

DEDICATION

I dedicate this to my family, and all those who I look up to.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF MAPS

LIST OF TABLES

Chapter 1: Introduction

In 2015 spring semester, I decided to pursue an internship to fulfill my Clark University GISDE program. I chose internship track for several reasons. First, for my career plan, I would like to apply a GIS related job when I graduated in Clark University. Experience is critical for applying GIS related job. Therefore, I think a real world GIS experience will help my study. At last, I want my summer vacation to be more meaningful. After I applied for 15 to 20 internships and jobs, I got a GIS analyst position which worked on a Land Change Model project in Clark Labs. The study and skills I learned at Clark University provided me a solid academic background beforehand. Besides that, I got this internship because I did a final project in advance raster course in 2015 spring semester which was strongly related to this internship project. In other words, the study at Clark University provided me this opportunity.

In June 6th, I started my internship at Clark Labs, and am continuing work till now. I think I gained a lot GIS and non-GIS experience during this time period. In the following sections, I would like to talk in detail about my internship experience. I structured this paper into four parts, description of organization, description of my internship and project; and internship assessment; and conclusion. Overall, it was a great experience, and I learned a lot from it.

Chapter 2: Organization description

Organization: Clark Labs

Clark Labs started in 1987 as a non profit organization providing software and research projects for GIS and image data analysis. TerrSet is the software that is provided by Clark Labs. Over 25 years of geospatial software development, currently, TerrSet becomes a powerful environment analysis software. It has variables tools and models for vector and raster based analysis. In general, TerrSet has over 300 analysis tools and mainly focuses in eight features which are IDRISI GIS Analysis, IDRISI Image Processing, LCM, GeOSIRIS, ESM,ETM, and CCAM.

The IDRISI GIS tool set aims to provide wide analytical tools for raster and vector data. IDRISI Image Processing system provides image restoration, enhancement, transformation and classification tools for the user. Land Change Modeler (LCM) is a model for land plan decision making and prediction. Habitat and Biodiversity Modeler (HBM) aims to provide a model for species habitat assessment and biodiversity research. GeOSIRIS aims to deal with forest deforestation, carbon emission and agricultural planning. Ecosystem Services Modeler (ESM) is a spatial decision support application that helps with sustainable development. Earth Trends Modeler(ETM) is a model that analyzes raster data in a time series to gain the observation of change over time. And Climate Change Adaptation Modeler(CCAM) models the global climate change by detecting sea level rise and other observations.

Clark Labs has a strong connection with Graduate School of Geography at Clark University. Meanwhile, Clark Labs has staff and visiting scholars from the geography and environment fields. Therefore, Clark Labs also works on a number of analytical or application research projects in which the TerrSet software is used. Recent projects include mapping marine aquaculture using Landsat imagery in Southeast Asia which is funded by the Gordon and Betty Moore Foundation and Land Change Modeler supported by Esri ArcGIS online.

Besides TerrSet software and related projects, Clark Labs also provides professional trainings and services for public users. Services include tech support, analytical service, software customization, etc. Training includes training classes and workshops for Terrset users. Clark Labs can offer specific training classes or workshops for specific topics or research groups. For instance, in 2010, Clark Labs conduct a national-level REDD planning training class for the Indonesian Government staff with the support of Conservation International. And in 2013, they conducted a conservation professional workshop for Peru.

At present, Clark Labs have about 10 regular staffs including manager, programmer, analyst, geographers and environmental scientists. Professor Ron Eastman is the director of Clark Labs. Ron Eastman is also a professor in Clark University Geography department. His main research interests are geographic information systems, earth system informatics, conservation GIS, remote sensing, and cartography. He got his PHD degree in 1982 from Boston University, and served as Director of the Graduate School of Geography in Clark University from 1998-2002.

My office is located at the left corner of second floor. There are five other Master students that work with different projects in the same office. The atmosphere of Clark Labs is relaxed and comfortable. We can chat and communicate with each others during the time. Clark Labs has a special culture tradition which is, in the summer time, three to four staffs or students cook a lunch for everyone in Clark Labs each week. In this summer, we provided and shared lunch every Thursday. It is a great opportunity for me to chat with Ron Eastman and other staffs at Clark Labs. Also, it was an opportunity to share your food and culture to other people. In the later of July, Siqi, Shuhong, and myself cooked a traditional south China meal for the whole Clark Labs. It was a fun experience, and I was so happy to share food from my hometown.

Chapter 3: Project description

Land Change Model for United States

This project is associated with Esri. In February 2015, Esri conducted Clark Labs to work on a cloud-based Land Change Model on ArcGIS online for professional change analysts. And later on, J. Ronald Eastman, the director of Clark Labs decided to operate an additional free cloud-based App for land change analysts and the public to explore land change of 48 states in United State. The difference between the professional version and the free explorer is professional LCM allows users to specify their own data for analysis, and the free explorer only allows users to view results that are provide by Clark Labs. In this summer, a group of over eight students, including Masters and Phd students, started to

provide the data and analysis for these two projects. In the meantime, several programmers and scientists joined together to build the web application.

Our mission in the summer was basically data extraction and processing. We needed to download and extract two land cover images and eight explanation variables. And test three types of model in order to get a better model for cloud-based app. In the follow sections, I would like to explain land cover image and each variables in detail.

Data Collection

Data for this project includes

1) National Land Cover Database (NLCD) land cover data for 2001 and 2011

2) National Elevation Data

3) CENSUS TIGER/Line (primary, secondary and local roads)

4) U.S. Census counties & county subdivisions;

5) USGS GAP PAD Protected Areas Database

6) USA Surface Water

7) Buffered & non-buffered state shapefile (US Albers)

Data Preparation

All raster and vector data were projected to Albers Conical Equal Area coordinate system(ALBERSUS83). This projection system has less absorption in area. In this project, we primarily use 2001 and 2011 NLCD land cover images to conduct our research. NLCD stands for national land cover database, which was provided by the Multi-Resolution Land Characteristics consortium. There are 30 meter spatial resolution land cover image based

on Landsat satellite data, classified by decision-tree classification, and applied the same 16

class land cover classification scheme across the whole United States. The 16 land cover

types are Open Water; Ice; Developed, Open Space; Developed, Low Intensity;

Developed, Medium Intensity; Developed, High Intensity; Barren Land; Deciduous Forest;

Evergreen Forest; Mixed Forest; Shrub/Scrub; Grassland/Herbaceous; Pasture/Hay;

Cultivated Crops; and Emergent Herbaceous Wetlands. Our mission for dealing with land

cover data is to clip each state, convert it from tiff file to rst file, and project to

ARBERUS82 projection system.

Eight variables are extracted and used in our model to explain the change. The

eight variables are: elevation, distance to primary roads, distance to secondary roads

distance to local roads, protected area, distance to surface water, slope, and distance to high

developed area.

2001 and 2011 Land Cover Image

In ArcMap,we clipped 2001 and 2011 land cover image by target buffered state.

In TerrSet, we used GDALIDRISI to convert geo-tiff image to rst type image. We used

SHAPEIDR and RASTERVECTOR tools to import and transfer non-buffered state

boundary to *.rst file in TerrSet. For states that have coastline, we did an additional

process that should be done in order for masking out ocean area. Figure 1 and figure 2

shows the land cover image of Kansas in 2001 and 2010.

Variables:

      In this project, as I talked before, we have eight explanatory variables. Roads data are from TIGER LINE database, which is created by United States Census Bureau. Protected area is created by USGS National Gap Analysis Program. The four types of road are vector files, while the surface water and elevation are raster layers. For variable extraction, the first step was to download all vector and raster files from certain websites. In the second, we needed to import these files to Terrset; project to Albers Conical Equal Area coordinate system (ALBERSUS83); covert to rst file; and crop to state boundary by OVERLAY operator. Distance to high development area and crop are variables that were extracted from the 2001 land cover image. Distance operator was applied to create distance images for 2001 high development area, primary roads, secondary roads, local roads, cropland, and surface water. Figure 3 and figure 6 show distance to primary road and water. The slope variable was created based on elevation image by DISTANCE module in TerrSet. Protected area has been rescaled to a range of 1 to 5. 1 present to non-protection area, and 5 present to completely protected area. (Figure 4) County subdivision data is created by simple import in raster. (Figure 5)

Model

      In the test period, we used several states to test all three land change modules. In order to avoid bias on size of state, we randomly selected eight states from all size levels. For me, I chose Kansas as my test state.

LCM

Land Change Module (LCM) in IDRISI is for land change analysis and modeling. The first task for this module is to do land change analysis from 2001 to 2011. In our project, we modeled only transition potentials over 2 square kilometer. Therefore, we created a change map that ignored the transition lower than 2 square kilometer and exported the text info into excel.

In NCLD land change map, in general, there are 16 categories which I talked about before. In our analysis, we divided these 16 land change categories into two general class, natural and developed, shown at table *. Developed categories only include four development categories. In the original stage, we considered cropland and pasture as developed, but later on, we decided to switch those to natural categories due to a county subdivision problem. For transition between natural and development, we decided to use all variables including county subdivision, and for transition between natural to natural, we decided to use all variables excluding county subdivision.

MLP

MLP is based on equal sized samples of historical change and persistence. It is a great method to create change potential images in TerrSet. The only problem is it takes a very long time. For my test state, Kansas, it took about half hour for each transition potential image creation. It is a very time consuming work, and obviously this method doesn't suit for a cloud-based app.

WNL

WNL stands for weighted normalized likelihood which calculate probability of change based on each variable. the calculation equation is as follows:

$$NL = P(x|C)/[p(x|C)+P(x|\sim C)]$$

NL produces a range from 0 to 1. If the value close to 1 means that this area is likely to change. And if a value is close to 0 it means that this place is likely not to change. And a value that is close to 0.5 means that the variables that included in the model cannot explain the change in this area. Therefore, we aggregate all NLs by using weighted linear combination method.

NLA

NLA stands for normalized likelihood average, which calculates the variables' probability in the same way as WNL. The only different is it considers the problem with protected area.

Compare MLA, WNL, and NLA

Overall, all methods can produce credible results. In terms of the similarity of three methods, we conducted a linear regression of these three methods. According to the linear regression result, we can see that WNL and NLA methods have very high correlation, and MLP shows much different than WNL and NLA. This might be caused by the different aggregation types. With the testing of all three methods, we finally decided to use NLA as our research arithmetic.

Result

For the project goal was for our team to mapping all 48 states in United State. Up till end of August, we finished just four to five states. And our plan is to finish all states in

United States by February, 2016. And we will continues to model all 48 states. In terms of result, I would like to use Kansas as an example to explain the output of WNL model. Kansas is a state that is located in the Midwestern United States. Topeka is the capital of Kansas, and major cities include Wichita, Kansas City, and Overland Park. According to 2014 national population data, the total population in Kansas is 2,904,021. In the meantime, Kansas is a one of the most productive agricultural states. It produces large amounts of wheat, corn, sorghum, and soybeans. According to our land cover image, in 2011, about 80% of land in Kansas is cropland. Based on our LCM result, we can see that there are over 100 transitions occurred during the 2001 to 2011 time period. There are 66 transitions are over 2 square kilometer. Figure 7 and Figure 8 show transition potential image from Development, Open Space to Development, Medium and from Deciduous Forest to Development, Medium. According to figures, red pixels presents a high transition potential, and blue pixels show a high resistance of change.

Problem and Solution

The data processes are not as right as rain. We also faced several problems in this project. For example, we got problems with our road map. As I talked before, this project originally used 2010 TIGER LINE road database for extract primary, secondary, and local roads. TIGER stands for Topologically Integrated Geographic Encoding and Referencing, which is a GIS database that was created by United States Census Bureau. When we tried to extract each state's primary road from this 2010 database, we found that there are several isolated line fragments presented which obviously are not parts of primary road. They are

10

classification error made by this database. But, they greatly affect our outputs. Due to this reason, we tried to fix this problem in several ways. For example, we tried to delete these line fragments manually. In the result, we finally found out that this problem is fixed in TIGER LINE 2014 data. Therefore, we switch our primary and secondary road data from 2010 to 2014. We also faced a problem in the WNL modeling section. In the original plan, we decided to use all eight variables in predicting all transition potential images. But, later on, we found out that if we include counties subdivision variable in natural to natural transition, we will get a weird result which is shown in figure *. According to this figure, we can see that weird squares appear in the image. These squares are not right. Due to this problem, we finally exclude counties subdivision in explaining natural to natural transition. Overall, our team faced problems and obstacles. But, we finally got over those problems.

Chapter 4: Internship Assessment

My internship at Clark Labs was an incredible learning experience that greatly increased my skills in Terrset and Land Change Modeler. I used TerrSet everyday when I work. The experience and skill I gained will greatly help for my future career and increase my competition in GIS job market. I also gained experience in how to fix and deal with problems during research. As I mentioned before, during the internship period, we always faced problems that we did not expect. It was a great opportunity that I can work together with Ron Eastman and project manager on fixing problems. In the meantime, my project manager shared a lots tips and personal experiences to me.

The culture of Clark Labs is to have an international lunch together each week. It was a great opportunity for me to talk with Ron Eastman, staffs, and colleagues in Clark Labs. Also, I have been talked a lot to my colleagues, with we have similar GIS background, we have a lot of common topics. Overall, I think I learn a lot in this internship, and I successfully complete this internship.

Chapter 5: Conclusion

Overall, I think it was a great experience in my life, and I did a good job. In the meantime, I get the opportunity to continuing work on this project for now. It is very nice to work on a huge project from start to finish. In an academic perspective, I got a full understanding of Land Change Modeler and TerrSet during this time period. I think TerrSet is a key software we should learn in our graduate study at Clark University. Therefore, the experience at Clark Labs consolidated my knowledge in TerrSet. Socially, I made a connection with students and staffs in Clark Labs. They shared their experiences and understanding in GIS and other aspects. It was very amazing. Therefore, I strongly recommend this internship for students who like and want to make a career in remote sensing filed.
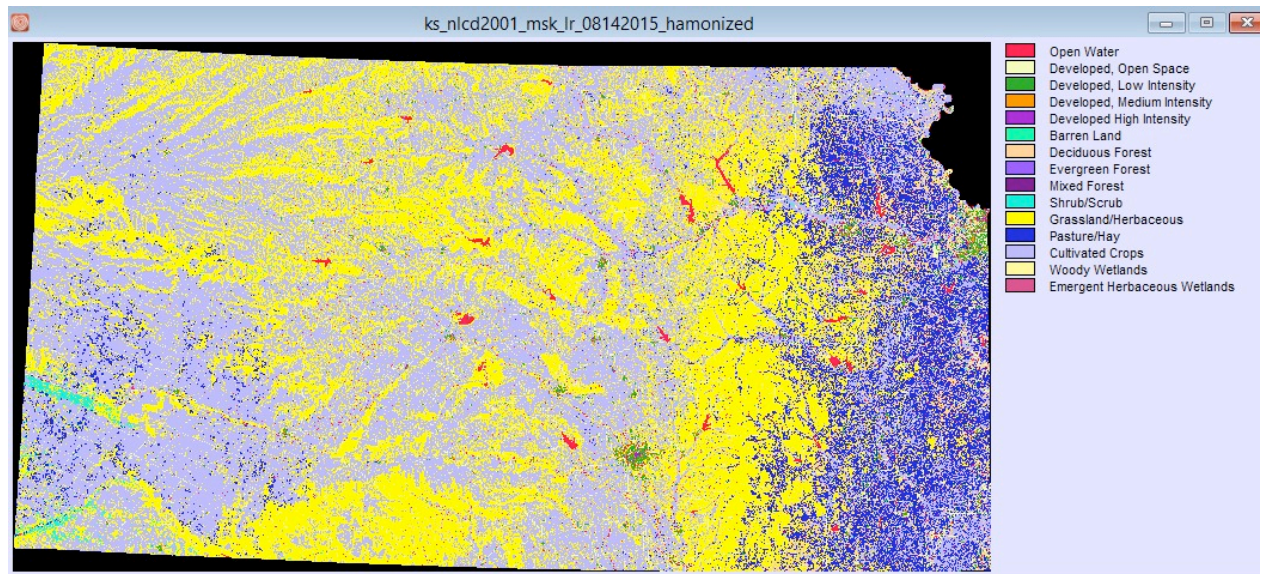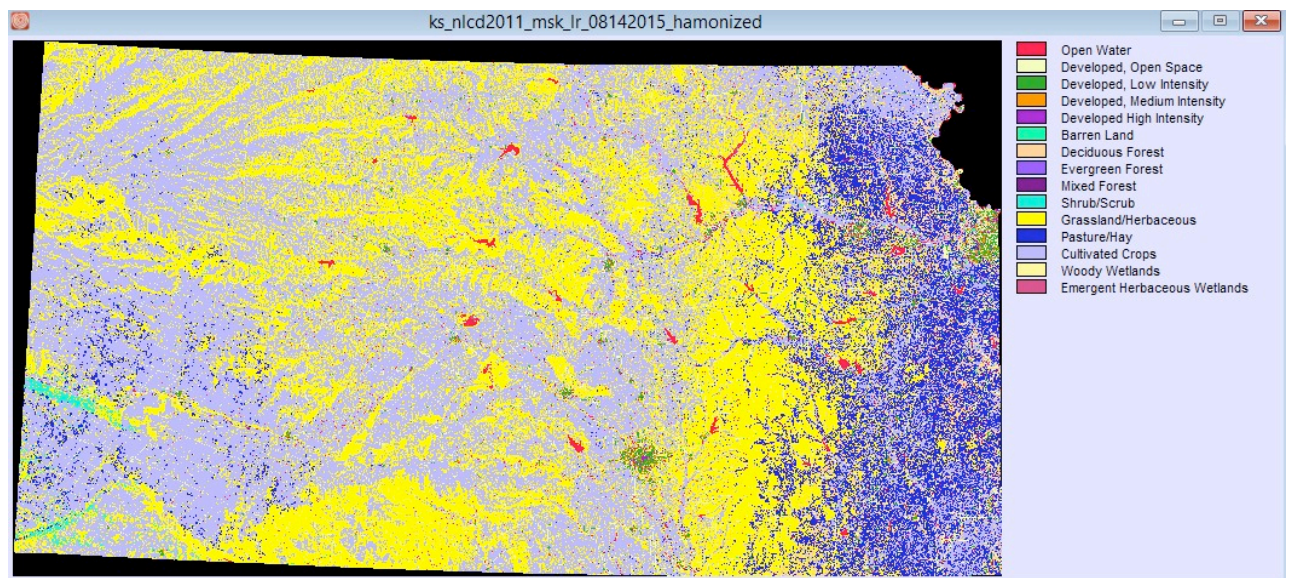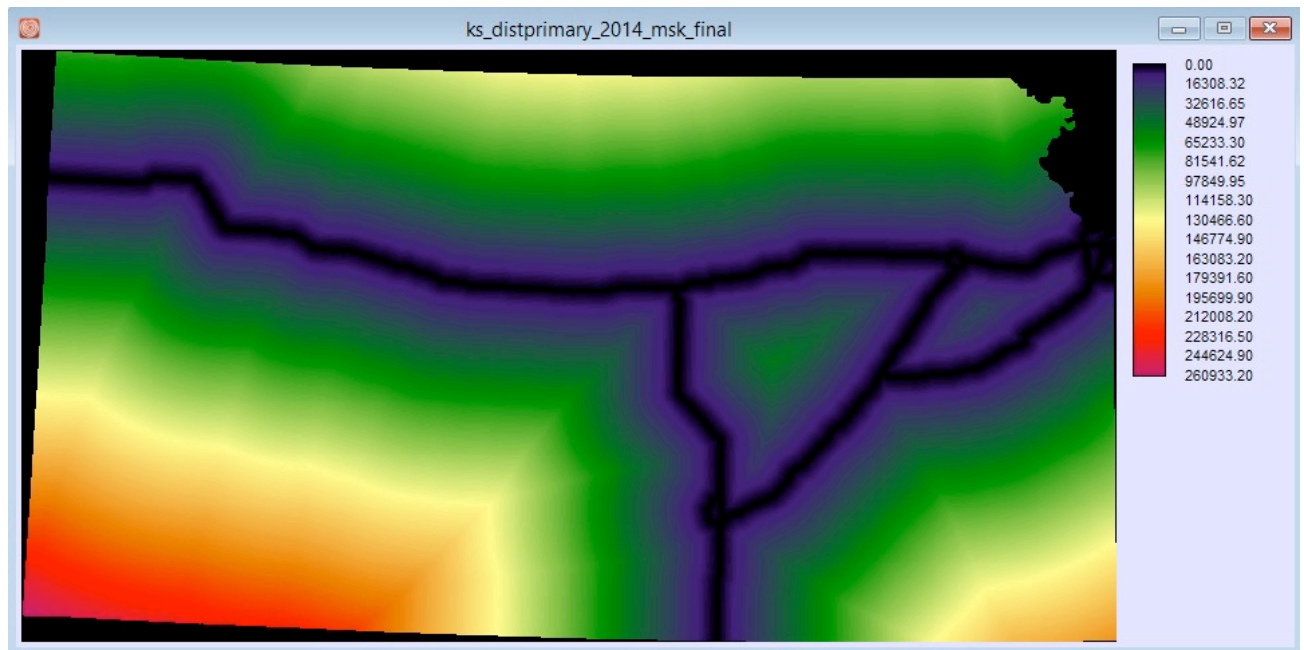
Figures:

Figure 1:
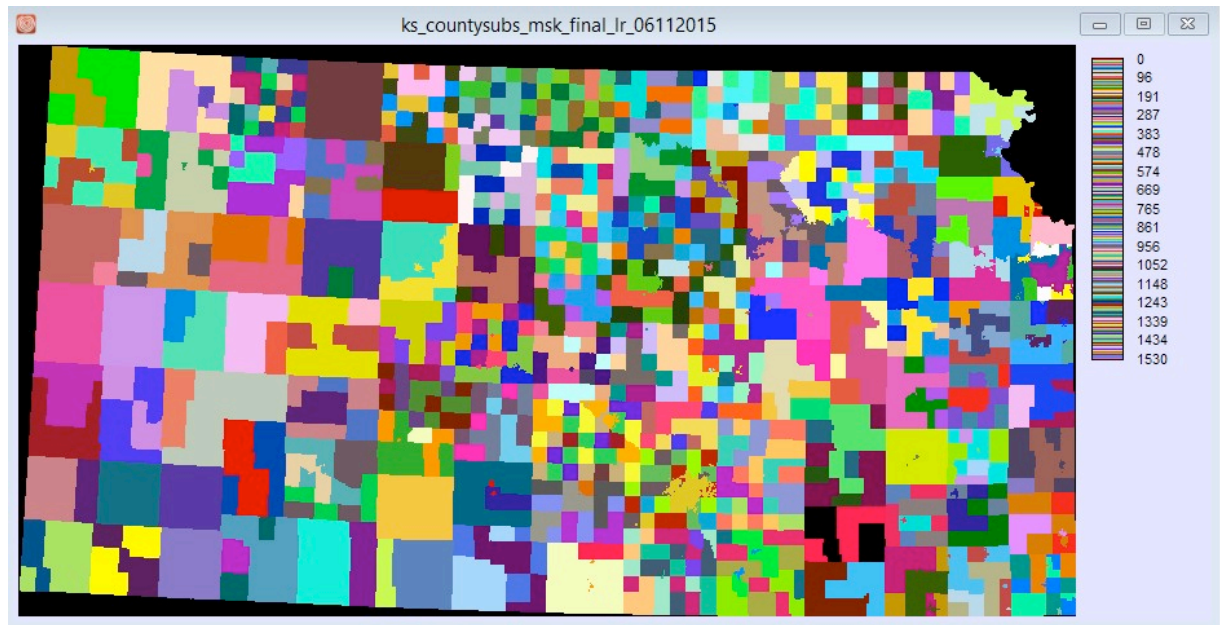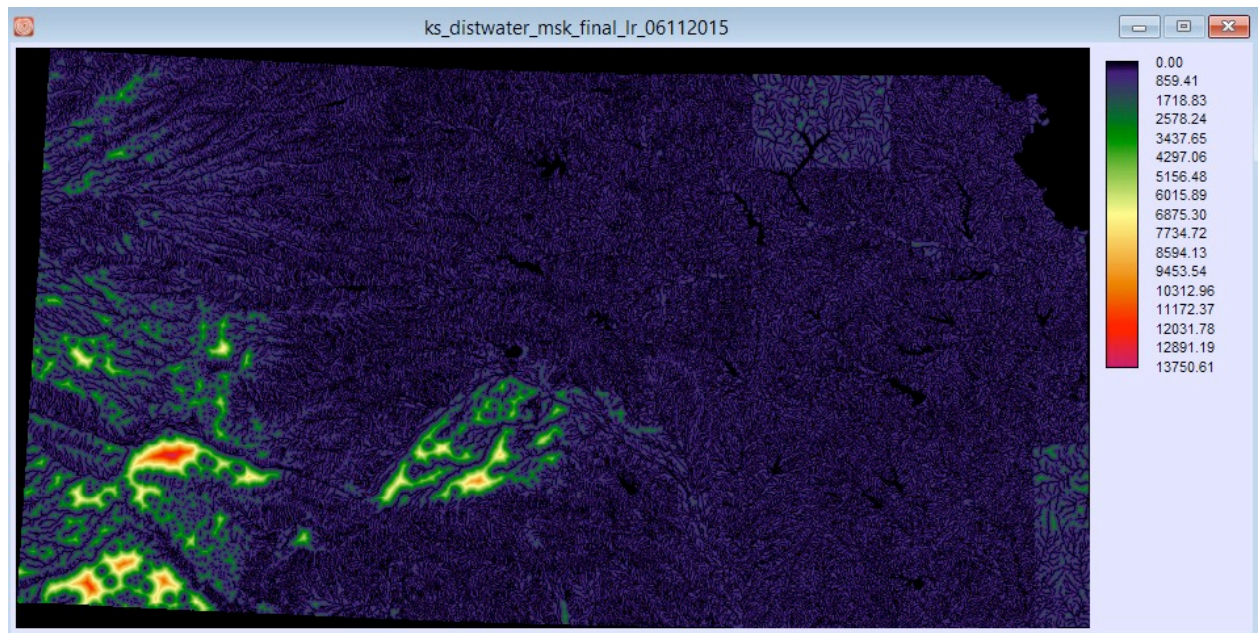

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:
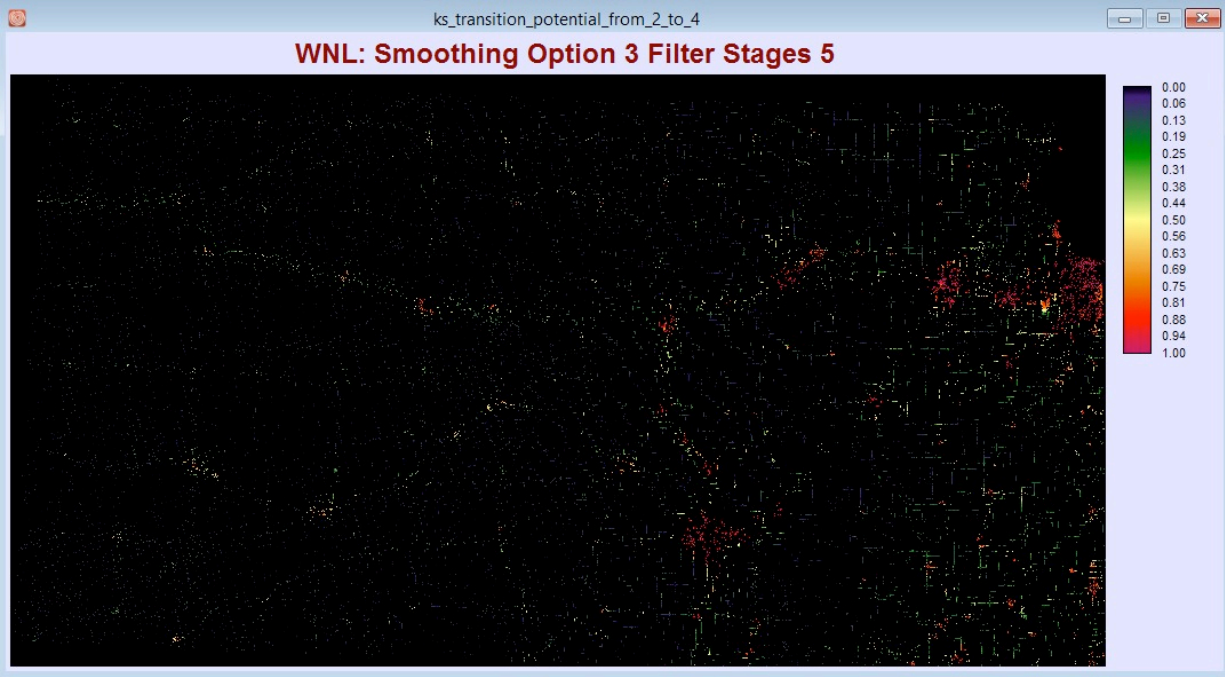

Figure 8:
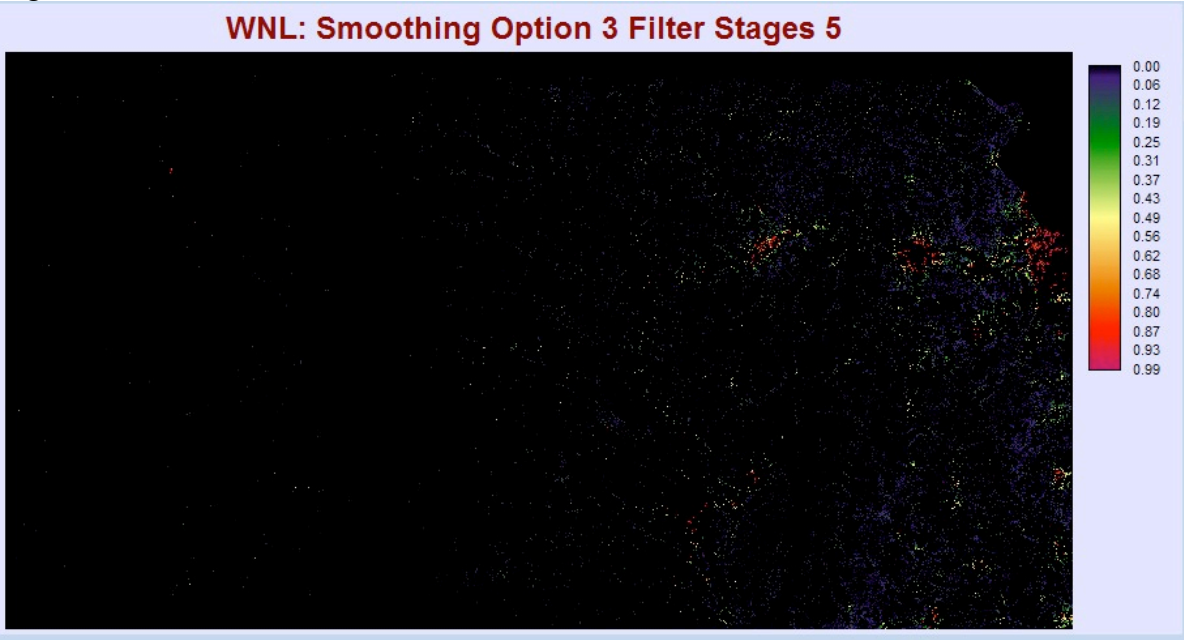
Table:

Time schedule:

Time period	Accomplish
8st, June to 25th, June	Data extracted and preprocessing within state: VT, MT, NM, NJ, GA, SD, AL, DE, KS, NH
26th, June to 14th, July	LCM, MLA, WNL, model test for Kansas
15th, July to 24th, July	Revising and cleaning primary and secondary road data error for all previous states
25th, July to 14th, August	Updating road data from 2010 data to 2014 data; revising distance to 2010 high development area for all previous states
15th, August, to now	New WNL model for all states Finished state: KS, SD, NJ