QUASI-EMPIRICAL AND SPATIO-TEMPORAL VULNERABILITY MODELING OF ENVIRONMENTAL RISKS POSED TO A WATERSHED

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Papia Faustina Rozario

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Environmental and Conservation Sciences

March 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

QUASI-EMPIRICAL AND SPATIO-TEMPORAL VULNERABILITY MODELING OF ENVIRONMENTAL RISKS POSED TO A WATERSHED

By

Papia Faustina Rozario

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Peter G. Oduor
Chair

Dr. Stephanie S. Day

Dr. G. Padmanabhan

Dr. Saleem Shaik

Approved:

3/31/2017 — Date

Dr. Craig Stockwell — Department Chair

ABSTRACT

Water quality assessment is crucial in investigating impairment within agricultural watersheds. Seasonal and spatial variations on land can directly affect the adjoining riverine systems. Studies have revealed that agricultural activities are often major contributors to altering water quality of surface waters. A common means of addressing this issue is through the establishment and monitoring the health of riparian vegetation buffers along those areas of stream channels that would be most susceptible to the threat. Remote sensing and Geographic Information Systems (GIS) offer a means by which impaired areas can be identified, so that subsequent action toward the establishment of riparian zones can be taken. Modeling the size and rate of land use and land cover (LULC) change is an effective method of projecting localized impairment.

This study presents an integrated model utilizing Analytical Hierarchical Process (AHP), Markov Chain Monte Carlo (MCMC) simulations, and geospatial analyses to address areas of impairment within the Pipestem Creek watershed, a part of the Missouri Watershed James Sub-region of North Dakota, USA. The rate and direction of LULC change was analyzed through this model and its impact on the ambient water and soil quality was studied. Tasseled Cap Greenness Index (TCGI) was used to determine the loss of forested land within the watershed from 1976 to 2015. Research results validated temporal and spatial relations of LULC dynamics to nutrient concentrations especially those that would be noted at the mouth of the watershed. It was found that the levels of Total Dissolved Solids (TDS) were much higher for the years 2014 to 2016 with a discernible increased localized alkalizing effect within the watershed. Fallow areas were seen to produce significant amounts of sediment loads from the sub-watershed. LULC distribution from 2007 to 2015 show that it is possible to project future land use change patterns. About 89.90%

likelihood of increment in agricultural land leading to a 77.44% likelihood of decrement in forested land in the area was noted for years 2007 to 2015. TCGI generated higher values for years 1976 to 2000 and it gradually reduced for 2000 to 2011 indicating loss of forested land.

ACKNOWLEDGEMENTS

With gratitude and appreciation, I acknowledge my major advisor, Dr. Peter G. Oduor, for his consistent support, guidance, challenges, patience, encouragement, advice, direction and constructive criticisms during my years at North Dakota State University. Taking his classes sparked my interest in this research and our numerous discussions on science and life have had important impacts on my world views that extend well beyond academia. A big thank you to my graduate committee members, Dr. Stephanie S. Day, Dr. G. Padmanabhan and Dr. Saleem Shaik, for their willingness to serve on my committee, providing suggestions and guidance.

I would also like to thank the federal agencies who funded this research including the United States Department of Agriculture and North Dakota Forest Service. I would like to thank the following entities for travel grants: NDSU Graduate School and Interdisciplinary Studies, Environmental and Conservation Sciences Program, the College of Science & Mathematics at NDSU, North Dakota View Scholarship and the Association of American Geographers.

The faculty and staff of the Department of Geosciences have all been very friendly and wonderful during my period of study here. Special thanks to Dr. Craig Stockwell, Director of Environmental and Conservation Science Program. Thanks to all the administrative staff (former and present) of the Department of Geosciences: Suzy Schmoll, Rita Slator, Phyllis Murray, Nancy Suttle, Carla "CJ" Marsh-Becker, and Carole Huber for their support during my studies.

I am deeply indebted to the faculty members of the Department of Geography in Loreto College specially Mrs. Sharmila Ray Kumam, Mrs. Soma Ganguly, Dr. Sushma Sahai, Mrs. Sabiha Sethwala, Mrs. Sumita Ghosh and Mrs. Ricci Chatterjee for building my knowledge and capacity of Geography.

A huge thank you to Dr. Buddhika Madurapperuma for serving as part of my field crew and Deidra Lies for taking the role of my lab supervisor. To my colleagues Dr. Buddhika Madurapperuma, Deidra Lies, Dr. Anthony Wamono and Mohammad Anar, thank you for all your support during my studies. I am deeply indebted to Deidra Lies and Kate Warner for being such incredible friends.

Last, but certainly not the least, I extend my sincerest gratitude to my family. Words can never express my gratitude for your love, support and understanding. I am especially thankful to my husband, Rahul and my sister, Margaret.

DEDICATION

To Ma

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

CHAPTER 1. INTRODUCTION

1.1. Natural vegetation of North Dakota

North Dakota's natural vegetation can be divided into three categories namely the forests, grasslands and the wetlands (Kotchman, 2010). An area saturated by surface water or groundwater for a long period of time such that it supports vegetation adaptive to that region is called a wetland (Wikum et al., 1974). These wetlands can be permanent or temporary in terms of holding water. North Dakota has a large area under wetlands. Areas with a higher water density, in central North Dakota, are called the prairie potholes region (Seelig, 2009; Wikum et al., 1974). The grasslands, also called the native prairies, are found extensively and comprises of mainly three different types - tall grass, mixed grass and short grass (Burns et al., 1990). Tall grass is concentrated along the Red River valley, mixed grass is found in the western part of North Dakota and short grass is found in the elevated areas of Missouri region of North Dakota (Madurapperuma, 2013). The forests are mainly of riparian origin and occur along the rivers in narrow strips (Rozario et al., 2016). Aspen and oak forests along with Juniper forests are found widespread (Kotchman, 2010).

1.2. North Dakota soil

North Dakota has a general soil type known as Williams named after Williams County, North Dakota. It is a loamy soil, which has a high content of organic matter. The soil is also naturally fertile supporting cultivation of various crops (Seelig, 2009). Soil use in North Dakota is generally dominated by field crop and native grass production. In addition, soils are used for waste disposal, construction material, building sites, recreational sites and wildlife habitat (Sweeney, 2009). Because of its unique properties, a soil exhibits various levels of suitability for different uses (Foth, 1978).

1.3. North Dakota watersheds

North Dakota has 52 watersheds spread over its 53 counties (EPA Web Portal). According to the Watershed Boundary Dataset, a surface water drainage is determined using topography and hydrographic principles and not on the county or the political boundaries (North Dakota State Water Commission, 2011). Any drainage area of no particular size is called a Hydrologic Unit (HU) as part of the hydrologic unit hierarchy and all units are given a code called the Hydrologic Unit Code or HUC (Laitta et al., 2004). A digit is assigned after each HUC, which signifies the level or subset (North Dakota State Water Commission, 2011). Watershed models are used to fill in the gaps in our understanding of watershed hydrology, the effects of human influence on the landscape, and identify nutrient sources (Ma et al., 2000). Physical watershed models require data from multiple sources, including meteorology, hydrology, water quality, permitted effluent discharges, and even satellite imagery of soils, land use and topography (Moriasi et al., 2007). Model calibration and validation are critical steps in model application, producing a representation of the watershed being assessed and predicting water quality based on the criteria of the study (Moriasi et al., 2007).

1.4. Background of land use and land cover change assessment

The natural environment of North Dakota has been modified to create built environments. This is reflected as a change in land use. According to the National Agricultural Statistics Service (NASS, 2012), 90% of the forested land in North Dakota has been converted into farmland. Land cover transition at such a large scale over a period creates environmental issues such as water pollution, soil degradation and erosion and this transition needs to be monitored and analyzed (Matson et al. 1997). Land cover generally refers to the biophysical material on the Earth's surface such as forest and urban areas while land use refers to the human use of the land

at a particular point in time and examples of this will include wheat farms, and wild life parks (Van Lier et al., 2011). Deforestation, agriculture, expanding farmlands and urban centers are a few of the ways in which man is changing the world's landscape (Foley et al., 2005). Although these activities vary from one place to the other, their impact on the earth's surface is usually the same and combined; these activities paint a picture of man's contribution in degrading the environment (Alemayehu et al., 2009). Conversion of forestlands to agricultural lands has been identified as the major form of land cover modification (Matson et al. 1997). The quest to develop better means of using natural resources and at the same time understand their impact on the environment has, over the years led to the development and improvement of maps and other methods of landscape analysis. Our ever-increasing use of the Earth's resources have led to both short and long term effects on the environment, and for decades remote sensing has played a major role in the understanding of the consequences of man's actions (Makhamreh et al., 2011). Change detection (monitoring changes in pixel value between images of a given location acquired at different times) using remote sensing has been considered of great importance in the monitoring of the Earth's well-being (Van Oort P.A.J., 2007). Change detection analyses are used to monitor the dynamic nature of biophysical and anthropogenic features on the Earth's surface. As aforementioned, it is important that such changes be monitored so that their contribution to global environmental change can be fully understood (Morawits et al., 2006). Change detection analysis is performed using multi-date imagery (Singh, 1989). Single date imagery show the land uses and land covers for a particular point in time but multi-date imagery show the land use and the land cover of a particular place at different points in time, (t_1, t_2….t_n) (Jensen, 2005). Land use like commercial, residential, transportation, utilities, cadastral, and land cover mapping have been especially improved over the years by the use of multi-date

imagery, which have been used in cases of progressive or gradual environmental changes such as erosion or reforestation for which more than one image may be necessary (Le Hegarat-Mascle and Seltz, 2004; Jensen, 2005). Of the many different change detection techniques that exist, two main categories can be identified. One category involves techniques that first detect change and then assign classes to the detected change (e.g., principal component analysis and image differencing) (Singh, 1989). A second category of techniques first assigns classes and then detects the changes between the different classes (Singh, 1989). An example of this second category of techniques is the post classification method of change detection (Van Oort P.A.J, 2007). Change detection analysis takes into consideration image characteristics such as spatial (and look angle), radiometric, temporal and spectral resolutions (Madurapperuma et al., 2015). For the most part, the type of land use or land cover to be studied and the level of detail needed in the study, determines the type of sensor to be used like, Landsat 5 -5 band image , Landsat TM -7 band image , SPOT, or Landsat Enhanced Thematic mapper (ETM) among others (Jensen, 2005). Visual change detection analysis (comparing the difference between two or more image visually, without any band analysis) is a basic form of change detection, cannot be grouped under any of the above categories, and has been successfully used by the National Wetlands Inventory (Wilkie and Finn, 1996). Unfortunately, visual change detection is time consuming and tedious, thus making automated (software driven) change detection analysis a more widely used procedure (Wilkie and Finn, 1996). Our ability to monitor successional changes in the environment has been made more practical since the launch of Earth resource sensing satellites. Change detection analysis, may be enhanced through anniversary date synchronization (Lillesand and Kiefer 2000). Using anniversary date images (images from same month but different year) ensures that the sun's angle of incidence is the same on both days of image

capture (Lillesand et al., 2004). In the case of post-classification change detection analysis, it is necessary that both images have high classification accuracy and accuracy assessment determines how well the classified image corresponds with what actually exists on the Earth surface (Jensen, 2005). A Confusion Matrix is way of assessing accuracy of a classification model that contains information about actual and predicted classifications done by a classification system (Provost and Kohavi, 1998). The term confusion relates to the true positives, true negatives, false positives and false negatives that are associated with this matrix. True positive is the proportion of positive cases that were correctly identified (Provost and Kohavi, 1998). For example, if the Confusion Matrix is used to predict change in forested areas to non-forested areas from 1976 to 1991, the cases in which the model predicts yes and there is actual change, then it is a true positive. Similarly, the case in which the matrix predicts no and there is no actual change then it is said to be a true negative. True negative is defined as the proportion of negative cases that were classified correctly (Provost et al., 1998). False positive is the proportion of negative cases that were incorrectly classified as positive whereas false negative is the proportion of positive cases that were incorrectly classified as negative (Provost et al., 1998). Accurate spatial registration, similar acquisition sensors same spatial, spectral and radiometric resolutions, of the images are all necessary for an effective change detection analysis to be performed (Jensen, 2005). In most cases, the above factors depend on the feature under study like lake levels, tidal stage (affected mostly by the moon), wind and soil moisture, might also be important in change detection analysis (Lillesand et al., 2004).

1.5. Geographic Information Systems (GIS) and Remote Sensing (RS)

A geographic information system (GIS) lets us visualize, question, analyze, and interpret data to understand relationships, patterns, and trends (Burrough, 1986). Remote Sensing (RS)

and Geographic Information Systems (GIS) help us in understanding how land use and land cover change over time and space, to understand the relationships between anthropogenic and natural phenomena and to make proper land management decisions (Kamusoko and Aniya 2006). The output of a remote sensing system is usually an image representing the scene being observed. Geographic Information Systems and Remote Sensing provide ideal platforms for monitoring and analyzing the spatio-temporal change in land use and land cover (Rozario et al., 2016). Studying the spatial pattern of land use gives us an idea of the underlying human activities at work in that particular region. Georegistered multi temporal remote sensing data helps determine the change associated to land use and Land cover. Collaborating field data with remotely sensed data can give us an inexpensive and faster way of assessing such change. *In situ* soil samples and water samples would reflect the increase or decrease in the nutrients present. This would enable us determine whether the water or the soil of that region is contaminated and to what extent.

1.6. Title justification

Quasi – empirical methods are a way of gaining knowledge by means of direct and indirect observation (Patton, 1990). Collecting field data and using it in research is a quasi – empirical approach used in this research. A Bayesian Markov Chain Monte Carlo model is an example of such a model. This research proposes a vulnerability model, that is, an index of the ability to produce clean water (I_{APCW}) which generated a composite score of the watershed highlighting areas of impairment within the watershed. The I_{APCW} model was able to predict areas within the watershed that were vulnerable to nutrient and sediment loading. The hybrid watershed model developed in this research is a spatial-temporal model since the preliminary

data was collected across time as well as space and spatial dependence of variables were analyzed.

1.7. Description of study area

The Missouri River sub-basin is approximately 635,500 acres, spreading over four counties of North Dakota namely Foster, Kidder, Stutsman, and Wells in the Missouri Region – James River Sub-Region of North Dakota. James River, Maple River, Pipestem Creek, Beaver Creek and Spring Creek are located in this sub basin. Of the 635,500 acres, Stutsman County contains 65%, Wells 22%, Foster 8%, and Kidder 5 % (NRCS, 2007). There are approximately 450 farms in the sub-basin. These farmlands produce commodities ranging from soybeans, wheat, barley, corn, canola, sunflowers, and field peas to beef cattle, swine, poultry, and bees. This region is mostly converted to agricultural land from forested land (NRCS, 2007). Thin buffers of riparian forests exist along the rivers. Figure 1.1. shows the map of the study area-Pipestem Stem creek along with James River. Pipestem Creek starts from the Pipestem Dam downstream to its confluence with the James River, which is about 5.6 miles. The mean annual precipitation is from 13 to 22 inches and mean elevation ranging from 1,280 to 2,560 feet. The type of soil found at this location is Williams–Bowbells loams which is a well-drained, non-saline clay loam with calcium carbonate of about 20%. Figure 1.2. shows a part of the watershed near Pingree, North Dakota, which was one of the sampling locations. Figure 1.3. shows a view of the watershed from Sykeston, North Dakota. Human activities like construction of impervious surfaces affect watersheds. Urban areas are appropriate examples of such surfaces. The amount of runoff is thus increased to foster surface erosion. This could be prevented if a buffer of forests is present in the areas in question. Most forest cover are of riparian origin and are found scattered along riverbanks. Table 1.1. shows the Land cover percentage within the

Missouri Watershed James subregion where the percentage area allotted for cultivated crops is much higher that the forest covers. Table 1.2. shows the Agricultural Land Suitability within the Missouri Watershed James subregion in North Dakota indicating a high suitability for grassland.

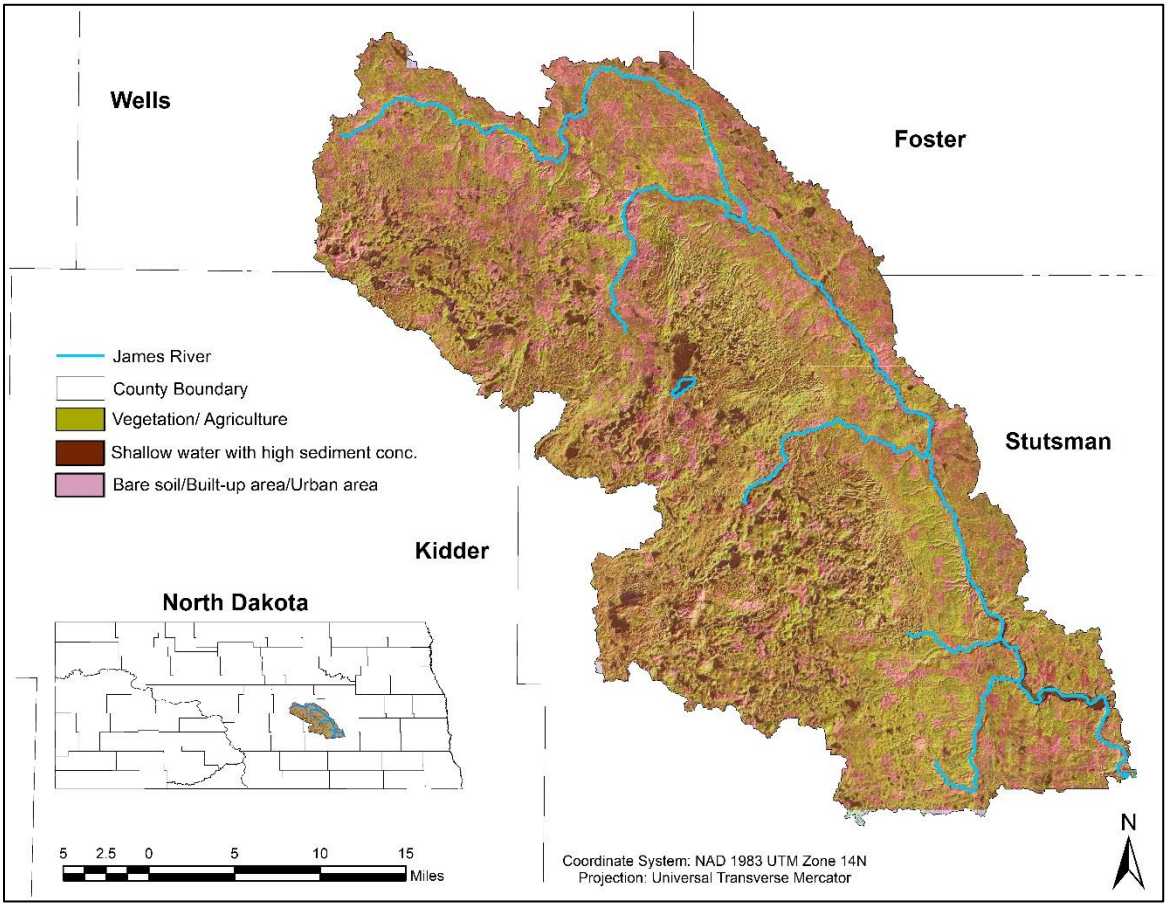

Figure 1.1. Map of study area depicting Pipestem Creek watershed and James River in North Dakota. Data for the boundaries of the watershed, county boundary and river were downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011. Vegetation, hillshade, shallow water and bare soil cover were derived from Landsat images, downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: January 26, 2011.

Figure 1.2. An image of the Pipestem Creek watershed from Pingree in North Dakota, which was one of the sampling locations. Picture was taken on July 18, 2011.

Figure 1.3. An image of the Pipestem Creek watershed from Sykeston in North Dakota, which was one of the sampling locations. Evidence of eutrophication can be seen here. Picture was taken on July 18, 2011.

Table 1.1. Land cover percentage within the Missouri Watershed James subregion. Data was acquired from the NRCS database 2007 (https://www.nrcs.usda.gov). Accessed: March 28, 2011.

Land cover	**Area (hectares)**	**Percentage**
Developed open space	28758.24	0.148883
Developed Low intensity	106.65	0.000552
Developed Medium Intensity	924.57	0.004787
Barren Land	53494.38	0.276943
Deciduous Forest	257179.5	1.331431
Evergreen Forest	30946500	0.016021
Mixed Forest	514.08	0.002661
Shrub land	503344.26	2.605837
Grassland/Herbaceous	5776106.58	29.90318
Cultivated Crops	6242172.75	32.31603
Small Grains	4432623.48	22.94791
Fallow Land	1424838.87	7.376459
Urban Grasses	13927.68	0.072104
Forested wetlands	21897.45	0.113364
Emergent wetlands	557043.48	2.883841
Total	50259431.97	100

Table 1.2. Agricultural Land Suitability within the Missouri Watershed James subregion. Data was acquired from the NRCS database 2007 (https://www.nrcs.usda.gov). Accessed: March 28, 2011.

Land cover	Area (hectares)	Percentage
Shrub land	566090.37	45.6379
Grassland/Herbaceous	649650.6	52.37448
Forested wetlands	24654.24	1.987612
Total	1240395.21	100

1.8. Dissertation layout

The dissertation consists of five chapters including an introduction, three chapters, which stem from published papers and the last chapter focusing on general conclusions and future work. Chapter 1 is a general introduction on the natural vegetation, soils and watersheds of North Dakota. It also highlights a literature review of the various methods of assessing land use and land cover change and the role of Remote Sensing and GIS in it. The details of the study area – Pipestem Creek watershed in North Dakota is included in this chapter, along with the conclusions drawn from this study. Chapter 2 discusses about developing a spatial model using Analytical Hierarchical Process (AHP). Several input parameters such as forest cover, riparian buffers, agricultural land, soil erodibility factor, road density and housing density were processed as independent datasets and then weights were assigned for each dataset. These datasets were modeled on a GIS platform to derive a single dataset showing areas of impairment within the watershed. An Index called the index of the ability to produce clean water (I_{APCW}) was generated for the watershed. This model was then compared to a distributed model, AnnAGNPS, and

results showed that fallow areas produced significant amount of sediment loads from the sub-watershed. The same locations generated a low $I_{APCW.}$ This chapter also explains the water quality and soil chemistry results acquired from the impaired areas within this watershed. Therefore, this model ascertains the increase in sediment load within the watershed indicating a change in LULC. Chapter 3 includes a comparison of the land use change detection statistics with a stochastic model such as the Markov Chain Monte Carlo Model. Changes from forested land to non-forested land was quantified for the years 2007 to 2015. The results showed concomitancy between the two methods. The Markov Chain Monte Carlo Model however had higher accuracy and was able to project future change. The combination of the spatial model discussed in chapter 2 and the stochastic model discussed in chapter 3 proved to be a good fit to assess LULC change and determine areas of impairment within a watershed. Chapter 4 discusses the implementation of a remote sensing technique - Tasseled Cap Transformation, and a geostatistical technique - Exponential Kriging interpolation to assess the intensity and direction of LULC change within the watershed. This model also studies the spatial structure within the watershed. The final chapter concludes the dissertation and proposes future work that can be done as an extension on this research.

1.9. Conclusion

This research addresses two key questions: (1) What amount of change has occurred in land use and land cover within the particular agricultural watershed? (2) How much change will take place in the future and in what direction? In relation to the stated questions, our four central objectives were met: (1) to create map overlays that will help quantify forest cover, open water areas and developed areas, (2) to develop indicators that will help prioritize areas that could be designated as areas for conservation within this watershed, (3) to analyze the direction, rate, and

spatial pattern of land use change within the watershed using a transition model, e.g. Markov Chain Monte Carlo (MCMC) Simulation, and, (4) to use a validation and prediction model such as Kriging to evaluate the change. This research project contributes to the growing field of work on land use change and its implications on riverine systems by introducing a new technique of incorporating Markov chain models in land use change maps. This will broaden our understanding of anthropogenic activities affecting land use change. GIS, RS and land use pattern metrics were coupled to quantitatively characterize the LULC change within an agricultural watershed. Changes quantified using remote sensing technologies provide observations which may show critical adverse and undesirable environmental impacts, hence requiring crucial sustainable land management policies and practices to avoid the endangering of the environment and sustainable development. This research presents geostatistical and deterministic methods to model uncertainties in image-derived estimates of the areal extent of developing land use policies. These uncertainties have a spatial attribute concerning the size of land use and proportion of land use. The spatial model using AHP utilized fewer parameters compared to other distributed models and was successful in replicating similar results. The Markov-CA model performance in predicted LULC distribution reveals the potential and the merit of using this approach for projecting land use change in similar agricultural areas. The simulated potential distribution of the LULC classes indicated that the changes the landscape has experienced in the recent past are likely to continue.

1.10. References

Alemayehu, F., Taha, N., Nyssen, J., Girma, A., Zenebe, A., Behailu, M., Deckers, S. and Poesen, J., 2009. The impacts of watershed management on land use and land cover

dynamics in Eastern Tigray (Ethiopia). *Resources, Conservation and Recycling*, 53(4), pp. 192-198.

Burrough, P.A., 1986. Principles of geographical information systems for land resources assessment. *Oxford University Press*. p. 54.

Burns, Russell M.; and Honkala, Barbara H., tech. coords. 1990. Silvics of North America: 1. Conifers; 2. Hardwoods. *Agric. Handb*. 654. Washington, DC: U.S. Department of Agriculture, Forest Service. Vol. 2. p. 877.

EPA portal .https://cfpub.epa.gov/surf/state.cfm?statepostal=ND. Assessed 2/7/17.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K. and Helkowski, J.H., 2005. Global consequences of land use. *Science*, *309*(5734), pp. 570-574.

Foth, H.D., 1978. Fundamentals of soil science. *Soil Science*, 125(4), p. 272.

Hegarat-Mascle, L., & S, S., 2004. Automatic change detection by evidential fusion of change indices. *Remote Sensing of Environment*, 91, pp. 390-404

Jensen, J.R., 2005. Introductory Digital Image Processing: a Remote Sensing Perspective (third ed) *Prentice Hall. Upper Saddle River*, NJ. pp. 505–512.

Johnson, W.C., Burgess, R.L. and Keammerer, W.R., 1976. Forest overstory vegetation and environment on the Missouri River floodplain in North Dakota. *Ecological Monographs*, 46(1), pp. 59-84.

Kamusoko, C. and Aniya, M., 2007. Land use/cover change and landscape fragmentation analysis in the Bindura District, Zimbabwe. *Land degradation & development*, 18(2), pp. 221-233.

Kotchman, L.A. 2010. North Dakota Statewide assessment of forest resources and Forest resource strategy. Resource Bulletin, *North Dakota Forest Service*, p. 85.

Laitta, M.T., Legleiter, K.J. and Hanson, K.M., 2004. The national watershed boundary dataset. *Hydro Line: GIS for Water Resources Summer*, pp. 1-7.

Lillesand, T., Kiefer, R., 2000. Remote Sensing and Image Interpretation 3rd Ed. *John Wiley & Sons, Inc.*

Lillesand, T., Kiefer, R., & Chipman, J. R.., 2004. Remote Sensing and Image Interpretation 5th Ed. *John Wiley & Sons, Inc.*

Ma, L., Ascough, J.C., Ahuja, L.R., Shaffer, M.J., Hanson, J.D. and Rojas, K.W., 2000. Root zone water quality model sensitivity analysis using Monte Carlo simulation. *Transactions of the ASAE*, *43*(4), pp.883-896.

Madurapperuma, B.D., 2013. From Bray-Curtis ordination to Markov chain Monte Carlo simulation: Assessing anthropogenically-induced and/or climatically-induced changes in arboreal ecosystems. *Doctoral dissertation, North Dakota State University*.

Madurapperuma, B., Rozario, P., Oduor, P. and Kotchman, L., 2015. Land use and Land cover Change Detection in Pipestem Creek Watershed, North Dakota. *International Journal of Geomatics and Geosciences*, 5, pp. 416-426.

Makhamreh, Z., 2011. Using remote sensing approach and surface landscape conditions for optimization of watershed management in Mediterranean regions. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(5), pp. 213-220.

Matson, P.A., Parton, W.J., Power, A.G. and Swift, M.J. 1997. Agricultural intensification and ecosystem properties. *Science* 277, pp. 504–509.

Morawitz, D.F., Blewett, T.M., Cohen, A. and Alberti, M., 2006. Using NDVI to assess vegetative land cover change in central Puget Sound. *Environmental monitoring and assessment*, 114(1-3), pp. 85-106.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), pp. 885-900.

North Dakota State Water Commission. 2011. Mitigation Plan. http://www.swc.nd.gov/4dlink9/4dcgi/GetContentPDF/PB-1956/Mitigation Plan.pdf. Assessed: August 1, 2015.

NRCS, 2007. SSURGO Data Layers. http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm. Assessed: June 18, 2016.

Patton, M.Q., 1990. Qualitative evaluation and research methods. *SAGE Publications, inc.*

Provost, F.J., Fawcett, T. and Kohavi, R., 1998, July. The case against accuracy estimation for comparing induction algorithms. *ICML*, 98, pp. 445-453.

Provost, F. and Kohavi, R., 1998. Guest editors' introduction: On applied research in machine learning. *Machine learning*, *30*(2), pp.127-132.

Rozario, P.F., Oduor, P., Kotchman, L. and Kangas, M., 2016. Quantifying Spatiotemporal Change in Land use and Land Cover and Assessing Water Quality: A Case Study of Missouri Watershed James Sub-Region, North Dakota. *Journal of Geographic Information System*, 8(06), pp. 663-682.

Seelig, B., 2009. Salinity and sodicity in North Dakota soils. *NDSU Extension Bulletins* 57.

Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, 10(6), pp. 989-1003.

Sweeney, M.D., Richardson, J.L. and Patterson, D.D., 2009. Understanding North Dakota's Soil Resource. *Farm research* Vol.41, No.1. pp. 3–8.

Van Lier, O.R., Luther, J.E., Leckie, D.G. and Bowers, W.W., 2011. Development of large-area land cover and forest change indicators using multi-sensor Landsat imagery: Application to the Humber River Basin, Canada. *International Journal of Applied Earth Observation and Geoinformation*, 13(5), pp. 819-829.

Van Oort, P.A.J., 2007. Interpreting the change detection error matrix. *Remote Sensing of Environment*, 108(1), pp. 1-8.

Wikum, D.A. and Wali, M.K., 1974. Analysis of a North Dakota gallery forest: vegetation in relation to topographic and soil gradients. *Ecological Monographs*, 44(4), pp. 441-464.

Wilkie, S., David & Finn, J. T., 1996. Remote Sensing Imagery for Natural Resources Monitoring: A Guide for First - Time Users. *Columbia University Press*.

CHAPTER 2. QUANTIFYING SPATIOTEMPORAL CHANGE IN LAND USE AND LAND COVER AND ASSESSING WATER QUALITY: A CASE STUDY OF MISSOURI WATERSHED JAMES SUB-REGION, NORTH DAKOTA, USA[1]

2.1. Abstract

Spatial causal effects on water quality are essential in identification of vulnerable watersheds. Modeling land use variables is an effective method of projecting localized impairment. This study presents an integrated index, designed to gauge the ability of an 8-digit Hydrologic Unit Code watershed in its ability to produce clean water. This index, I_{APCW}, can be successfully applied on a geospatial platform. In this study we utilized I_{APCW} to address forest cover dynamics of an impaired watershed, that is, Missouri Watershed James Sub-region in North Dakota. Specific parametric functions were analysed and combined within a customized GIS interface to provide a multi-faceted structured technique to derive I_{APCW}. These included ambient forest cover, housing density, agricultural land, soil erodibility and road density; it can be lucidly ascertained that where a prevailing forest cover undergoes conversion processes, the secondary effect may spur an exponential increase in water treatment costs. These parameters when projected statistically validated temporal and spatial relations of land use/land cover dynamics to nutrient concentrations especially those that would be noted at the mouth of the watershed. In this study, we found that the levels of Total Dissolved Solids (TDS) were much higher for the years 2014 to 2016 with a discernible increased alkalizing effect within the

[1] The material in this chapter was co-authored by Papia F.Rozario, Dr.Peter Oduor, Larry Kotchman and Michael Kangas (Published in Journal of Geographic Information System, 8, 663-682. https://doi.org/10.4236/jgis.2016.86053). Papia F. Rozario had primary responsibility for conducting this research including collecting field samples. Papia F. Rozario was the primary developer of the conclusions that are advanced here. Papia F. Rozario also drafted and revised all versions of this chapter. Dr.Peter Oduor served as proofreader and checked the math in the statistical analysis conducted by Papia F. Rozario.

watershed. When I_{APCW} was compared to Annualized Agricultural Nonpoint Source (AnnAGNPS), the spatial distribution generated by the AnnAGNPS study showed that fallow areas produced significant amounts of sediment loads from the sub-watershed. These same locations in this study generated a low I_{APCW}.

2.2. Introduction

Water and sediment supply, and their management, are critical to many hydraulic project operations. Trend analysis of water quality data is an essential environmental diagnosis of a stream allowing evaluation of how the water body has responded to change in land use over a period of time (Sadoddin et al., 2012). They directly impact sustainable use of reservoirs, water quality, and riparian habitat (Youg G. Lai, 2005). However, we are limited by the tools and methodology available to understand the future impacts on a larger scale. Water and sediment supply has been measured only at limited locations and over a limited time period. Thus we need a predictive model that provides a viable alternative. Sekar and Randhir (2007) developed prioritization maps to characterize conjunctive water harvesting potential that is based on benefits and costs. The results of their study demonstrate that a spatially variable harvesting strategy can be used to minimize runoff loss and to augment water supplies.

Changes in the composition of soil take place due to change in Land Use Land Cover (LULC). LULC is an integrated term that includes both categories of LULC and analysis of changes is of prime importance to understand many social, economic and environmental problems (Pelorosso et al., 2008). Land use (LU) and Land cover (LC) are the two fundamental components describing the terrestrial environment in relation to both natural and anthropogenic processes (Jansen and di Gregorio, 2002; Bender et al., 2005; Mendoza et al., 2010, 2011). Environmental modifications worldwide are mostly caused due to LULC changes, thus it

emerges as a key research question (Xiao et al., 2006). Quantifying landscape patterns enable us to identify and evaluate temporal changes and enable the study of the effects of pattern on ecological processes (Turner 1989). Jensen (2005) in his investigation of urban landscape perceived land use as a way by which human beings utilize land while land cover exists as a natural environmental system. Remote sensing and Geographic Information Science (GIS) techniques have been effectively utilized to identify and quantify periodic change in the landscape and its consequent environmental impacts (Madurapperuma et al., 2015). Land cover is an important parameter for monitoring agricultural, hydrological and watershed modeling which constitute necessary tools for development, planning and management of natural resources in a particular region (Madurapperuma et al., 2015). Past research has shown that increase in agricultural land use has direct consequence on sedimentation, nutrients, and pesticides in streams (Osborne and Wiley, 1988; Soranno et al., 1996). Land use change detection is therefore a critical requirement for the assessment of potential environmental impacts and developing effective land management and planning strategies (Leh et al., 2011). Surface water bodies are the potential recipients of the contaminations contained in surface runoff from their catchments (Chin, 2006). This makes surface water quality monitoring an important parameter. There are limited resources available for conservation that can be allocated to the erosion susceptible areas. These areas can be highlighted through mapping, monitoring and prioritizing (Fei et al., 2010). Erosion risk mapping of the area can enable the decision makers to prioritize the susceptible areas for conservation measures in accordance with their level of vulnerability (Iqbal et al., 2014). According to USDA Forest Service, protecting and managing forests in source watersheds is an essential part of future strategies for providing clean safe drinking water thus an Index of the Ability to Produce Clean water (I_{APCW}) can be generated through GIS overlay

Analysis to prioritize impaired watersheds. Spatial Multi Criteria Decision Making (MCDM) has also become one of the most useful methods for land use and environmental planning, as well as water and agricultural management (Davidson et al., 1994; Ahamed et al., 2000; Joerin et al., 2001; Ceballos-Silva and López-Blanco, 2003; Sicat et al., 2005; Chen et al., 2007). As a result, the request for GIS models and tools supporting collaborative decisions has increased over the last decade (Kollias and Kalivas, 1998; Karnatak et al., 2007; Reshmidevi et al., 2009; Chen et al., 2009). GIS-based MCDM involves a set of geographically defined basic units (e.g. polygons in vectors, or cells in rasters), and a set of evaluation criteria represented as map layers or attributes. Based on a particular ranking schema, it ultimately informs a spatially complex decision process by deriving a utility of these spatial entities through overlaying the criterion maps according to the attribute values and decision maker's preferences using a set of weights. Therefore, besides criteria selection, criteria weight severely impacts the results of the MCDM (Chen Y. et al, 2010). Nutrients in a water body such as nitrogen and phosphorus are considered to be pollutants when these nutrient concentrations become excessive, causing some organisms to proliferate at the expense of others (Davis and Masten, 2004). The situation is significantly multiplied by eutrophication, which is caused by excessive algae growth in a water body from surrounding agricultural watersheds due to the excessive presence of the necessary growth nutrients and ambient conditions that promote algal blooms. This enhanced plant growth reduces the dissolved oxygen levels when the plants decompose, potentially hindering the survival of fish and other aquatic life that depend on pristine conditions (Ritter and Shirmohammadi, 2001). These physical and chemical changes may interfere with the recreational and aesthetic uses of the water body, while both taste and odor problems may make the water less desirable for water supply and human consumption (Ritter and Shirmohammadi, 2001). Thus, it is essential to

estimate and qualify nutrient contaminations within the watershed. The objectives of this current study are to assess and analyze the LULC changes and to prepare a risk map through weighted overlay of influencing factors such as vegetation, rainfall, LULC, soil data and water quality data. Figure 2.1. shows the sampling sites within the Pipestem Creek watershed. In the process, we also identified the potential areas showing levels of vulnerability to change in soil and water quality.

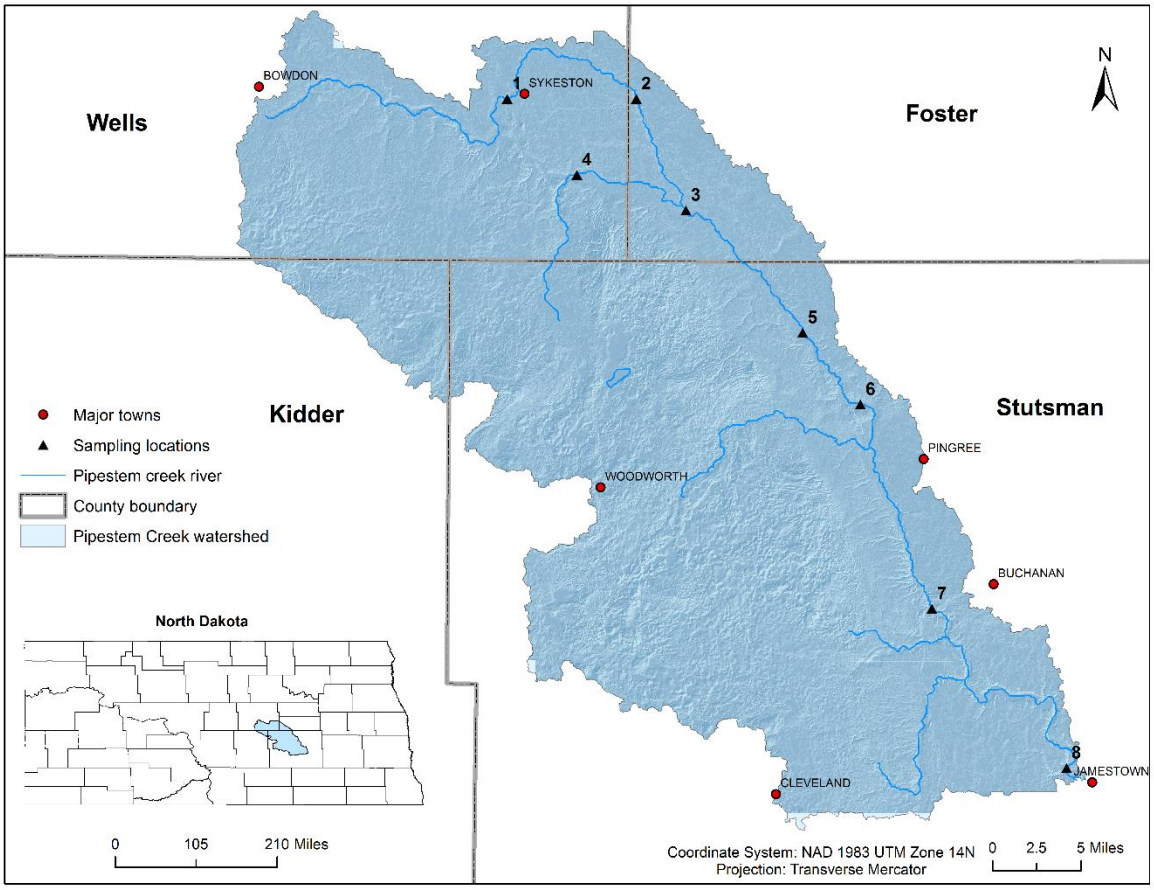

Figure 2.1. Pipestem Creek in North Dakota, USA depicting sampling locations around the watershed. Data for the boundaries of the watershed, county boundary, towns and river were downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

2.3. Materials and methods

2.3.1. Data processing and GIS analysis

Historic data of Pipestem Creek was used to perform a spatial analysis and identify localised areas of impairment within the watershed. Forest Cover including riparian forests and agricultural land use data was acquired from United States Department of Agriculture National Agricultural Statistics Service (NASS). Soil erodibility dataset was acquired from United States Department of Agriculture National Resources Conservation Service (NRCS). Road network data and year 2000 Housing Density data was acquired from the North Dakota GIS Hub. The land use and land cover was classified using the Anderson classification system (Anderson, 1976). NLCD data, a raster dataset was imported to ArcMap®9.3, a GIS software, where only the study area was clipped. Each attribute dataset was processed individually to produce a raster grid. To summarize forest cover, the "Tabulate Areas" function was used in ArcMap®9.3, to calculate the acreage of forested land for the watershed. The percent of the watershed that is forested was calculated by dividing the acreage of forested land by the total watershed land acreage (Barnes et al., 2009). The results were saved to the attribute field of this shapefile, which was then converted to a 30 m raster dataset (Figure 2.3.). The percent forest was reclassified into the four categories. Category break points were entered as half integers between the intervals. For example, a value of 24.5 was the break point for percent forest land scored as low or moderate. The results were saved in the corresponding attribute field. Agricultural land was summarized using grid values from the NLCD 2001 dataset of North Dakota. The same method was replicated to tabulate the areas under agricultural land. The percent agricultural land was reclassified into the four categories summarized in Table 2.1. Category break points were entered as half integers between the intervals. For example, 30 was the threshold for percent

agricultural land scored as low. The results were saved in the attribute field of this shapefile which was then converted to a 30 m raster dataset (Figure 2.4.). For riparian forest cover, the acreage of riparian forested land was divided by the total acreage of riparian buffer in the watershed. The percent riparian forest cover was reclassified into the four categories summarized in Table 2.1. Category break points were entered as half integers between the intervals. A value of 29 was the break point for percent riparian forest scored as low. The results were saved in the attribute field of this shapefile which was then converted to a 30 m raster dataset (Figure 2.2.). The North Dakota national roads dataset was split into east and west portions using the county boundary shapefiles. The "Repair geometry" tool for the east and west roads dataset was used to repair self-intersecting lines. The "multipart to single part" tool on each dataset was used to join multipart lines. Each road shapefile was converted to a coverage arc. A "Simplify Line" tool was run on each layer to remove excessive vertices. The simplification tolerance was set to 10m. A "Line Density" function was applied on each of the resulting coverage. Parameters were set as cell size to be 30m, search radius to be 565 m (to equal a search area of approximately 1 km^2) and the units were set as square kilometre. Finally, the East and West line density raster was merged into one raster dataset (Figure 2.5.). The results were sorted into four quartiles, and reclassified with values 1-4 as shown in table 2.1. To summarize the soil erodibility map, first, the soil dataset was clipped to the watershed boundary. Then this dataset was converted to a raster format using *kffact* field as the grid value. The *kffact* field is the *k* factor in the soil which contains the erodibility values in the dataset. The grid (Figure 2.6.) was then reclassified as shown in table 2.1. The raw housing data file was reclassified into 15 classes to reduce the file size. The classification used was based on 2000 U.S. Census Bureau block (SFI) data developed by the Natural Recourse Ecology Lab (Barnes et

al., 2009). Each of the 15 classes was assigned a range of housing density, for example, class 5 ranged from 32 to 49. To summarize the housing density data, the raw 2000 housing density data was clipped to the watershed area and resampled from a 100 m grid to a 30 m grid (Figure 2.7.). The raw grid values in units per hectare were converted to acres/unit using the relation (Barnes et al., 2009) in equation 2.1.

((units/ha)/1,000) * 1 ha/2.47 acres = units/acre (invert) = acres/unit, (2.1.)

Table 2.1. (a) Percent forest cover reclassification; (b) Percent agricultural land reclassification, (c) Percent riparian forest cover reclassification; (d) Road density reclassification; (e) Soil erodibility reclassification; (f) Housing density reclassification.

(a)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Percent forest land (F)	0 – 24	25 – 49	50 – 75	>75

(b)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Percent agricultural land (A)	>30	21 – 30	10 – 20	<10

Table 2.1. (a) Percent forest cover reclassification; (b) Percent agricultural land reclassification; (c) Percent riparian forest cover reclassification; (d) Road density reclassification; (e) Soil erodibility; (f) Reclassification and Housing density reclassification (continued).

(c)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Percent riparian forest cover (R)	0 – 29	30 – 50	51 – 70	>70

(d)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Road density (D, quartiles)	75 – 100th percentile	50 – 74th percentile	25 – 49th percentile	0 – 24th percentile

(e)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Soil erodibility (S, k factor)	>0.34	0.28 – 0.34	0.2 – 0.28	0 – 0.2

Table 2.1. (a) Percent forest cover reclassification; (b) Percent agricultural land reclassification; (c) Percent riparian forest cover reclassification; (d) Road density reclassification; (e) Soil erodibility; (f) Reclassification and housing density reclassification (continued).

(f)

Attribute	**Rating for 30-meter grid cell**			
	Low (1 point)	**Moderate (2 points)**	**High (3 points)**	**Very High (4 points)**
Housing density (H, acres per housing unit in 2000)	< 0.6 acre/unit	0.6 – 5.0 acres/unit	5.0 – 20.0 acres/unit	> 20.0 acres/unit

This generated a new reclassified dataset with 15 classes. The 15 value classes were categorised into four housing density classes: rural, exurban, suburban, and urban where rural ranged from 1 to 8 and assigned a reclassified value of 4; exurban was 9 to 10 assigned a value of 3; suburban was 11 to 12 which was assigned a value of 2 and for the urban class, ranging from 13 to 15 assigned a value of 1. So weightage was assigned based on this as shown in table 2.1. Using raster calculator *add* function for the six raster, resulted in a grid output with values ranging from 6 to 24 for each 30 m grid cell. Using ArcMap® 9.3 Spatial Analyst, a weighted overlay was performed using the data that included forested land, riparian cover, agricultural land suitability, soil erodibility, road density and housing density data. Equation 2.2 shows the underlying concept in a weighted overlay process.

$$I_{APCW} = R + F + S + A + D + H \qquad (2.2.)$$

where, F = forest land (percent); A = agricultural land (percent); R = riparian forest cover (percent); D = road density (quartiles); S = soil erodibility (k factor); H = housing density (acres per housing unit in 2011).

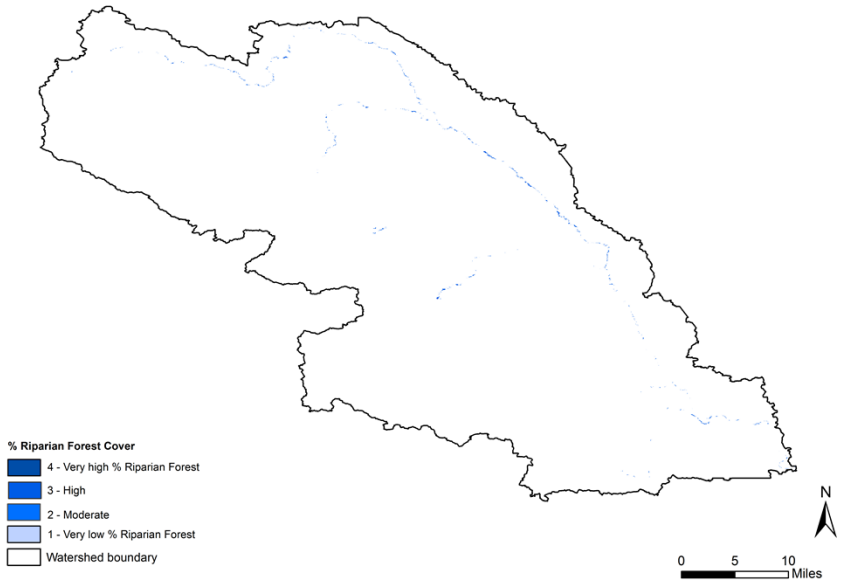

Figure 2.2. GIS weighted scoring showing percentage change in riparian forests within the Pipestem Creek watershed. Riparian forest cover was derived from NASS datasets. Data for the watershed boundary was downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

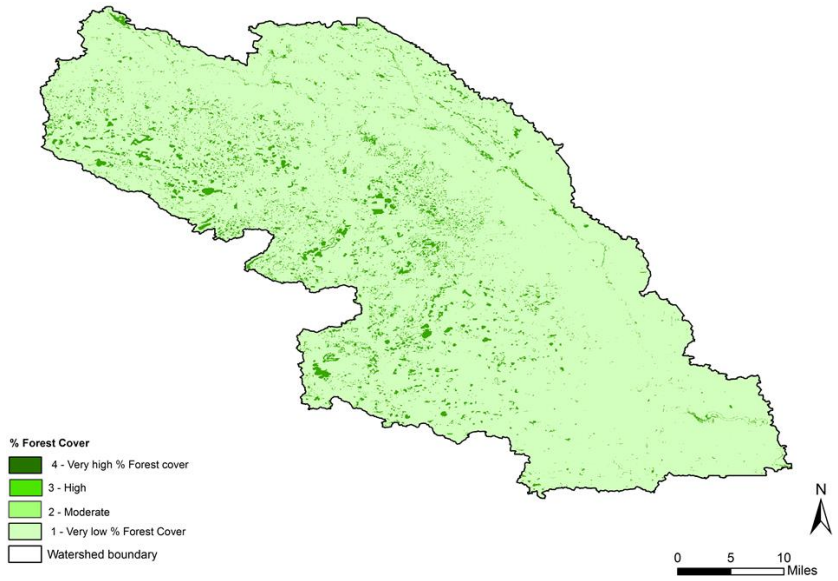

Figure 2.3. GIS weighted scoring showing percentage change in other forests within the Pipestem Creek watershed. Forest cover was derived from NASS datasets. Data for the watershed boundary was downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

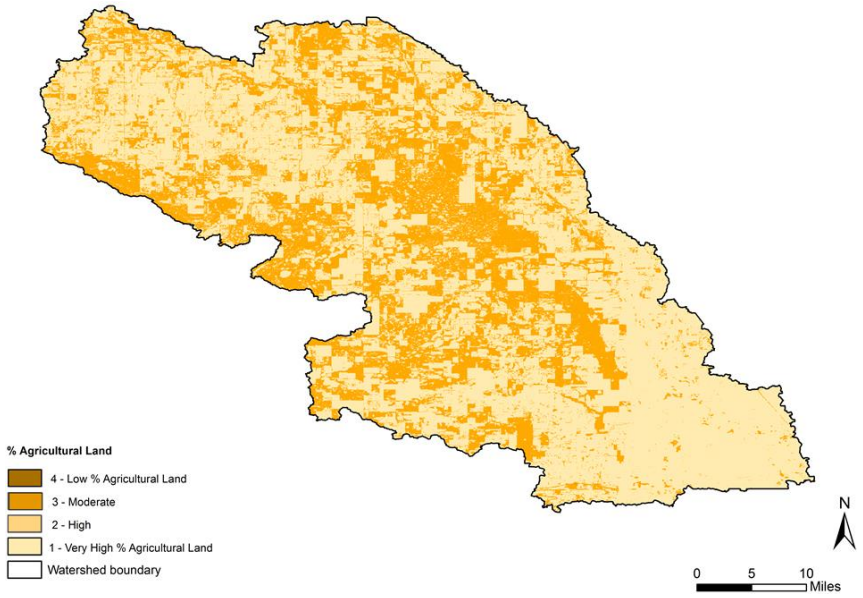

Figure 2.4. GIS weighted scoring showing percentage of agricultural land within the Pipestem Creek watershed. Percentage of Agricultural land was derived from NLCD. Data for the watershed boundary was downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

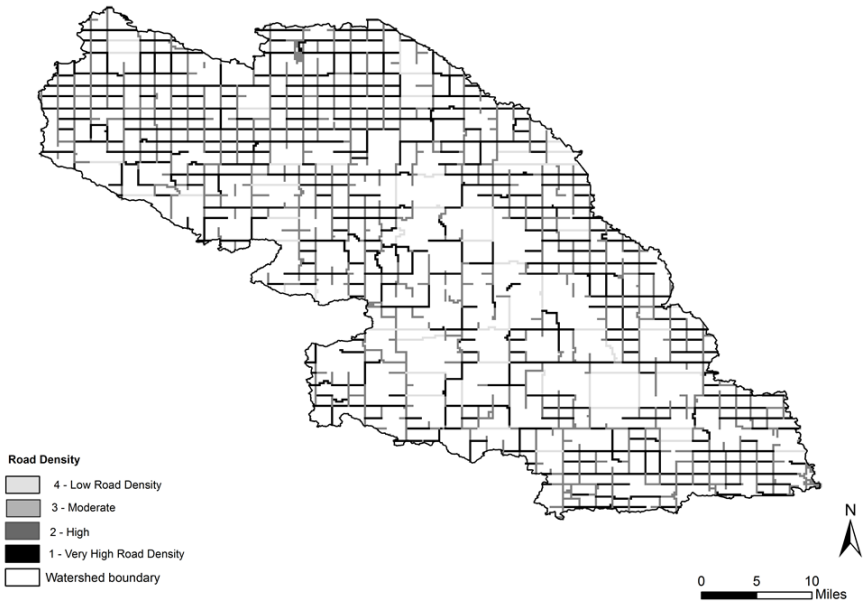

Figure 2.5. GIS weighted scoring showing percentage of road density within the Pipestem Creek watershed. Road Density was derived from datasets acquired from the North Dakota GIS Hub Data Portal. Data for the watershed boundary was also downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

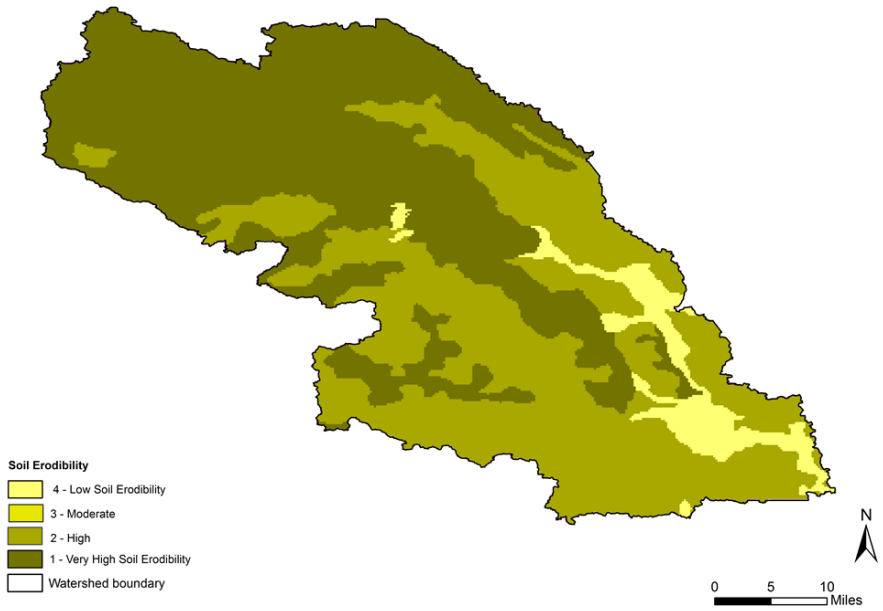

Figure 2.6. GIS weighted scoring showing soil erodibility within the Pipestem Creek watershed. Soil erodibility data was derived from NRCS datasets. Data for the watershed boundary was downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

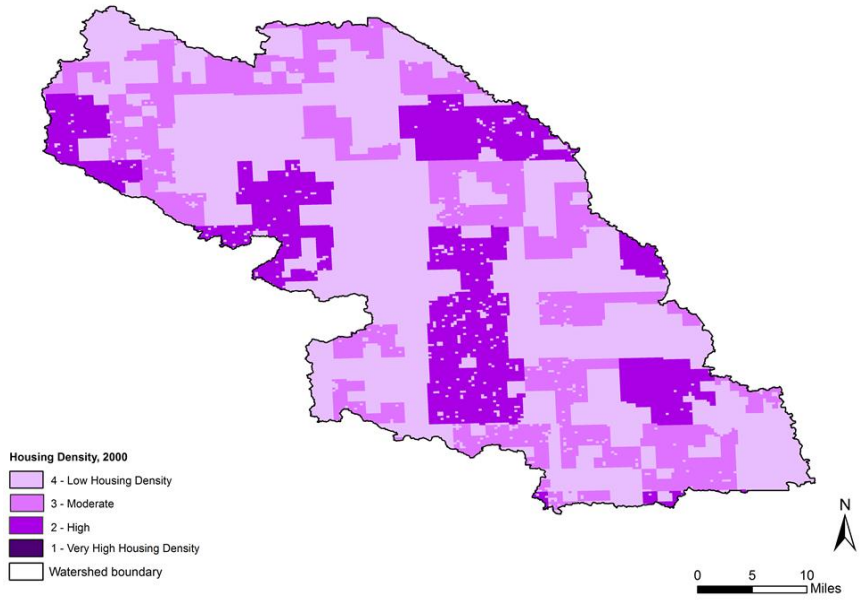

Figure 2.7. GIS weighted scoring showing percentage of housing density within the Pipestem Creek watershed. Housing Density was derived from datasets acquired from the North Dakota GIS Hub Data Portal. Data for the watershed boundary was also downloaded from the North Dakota GIS Hub Data Portal (http://www.nd.gov/gis/). Accessed: January 26, 2011.

I_{APCW} is the Index of the Ability to Produce Clean Water. All the variables were classified into four classes: low, moderate, high and very high. This was done to maintain an equal influence. The weighted overlay tool generated a final raster dataset, which was a mean composite score by watershed. Figure 2.8 shows the map of the study area with the I_{APCW}. The I_{APCW} was originally used in a study conducted by Barnes et al., in 2009 where the index was generated for a watershed at a regional level.

2.3.2. Analysis of soil data

Soil samples were collected from the top soil layer (0-6) inch to capture leachable ions in 2011 within Pipestem Creek watershed. The sampling sites were selected such that they incorporated the entire study area. Areas that depicted a low composite score in the map of the I_{APCW} (Figure 2.8) were kept in consideration while choosing the sampling points. Soil testing was done to analyze 12 different elements, which included Sodium, Potassium, Calcium, Magnesium, Copper, Chlorine, Nitrate, Phosphate, Zinc, Iron, Manganese and sulphur. Nitrate Electrode Method was used to read the Nitrate concentration in the soil where 20g of the soil sample was added to 50ml of an extracting solution. The suspension was stirred with a magnetic stirrer and then read on a pH/ion meter. Chlorides were also estimated using a pH/ion meter. The Olsen Test was used to detect Phosphorus levels in the soil where the colorimeter is used to produce an intensity and standard curve to determine phosphorus concentration in the soil. Estimates of available Potassium (K) in the soil were done by Atomic adsorption/emission Spectrometer which gives a standard curve for K by emission. The result was multiplied by 10 to give ppm in a soil (mg K/kg). Estimates of exchangeable calcium and magnesium were also acquired using Atomic adsorption/emission Spectrometer by adsorption and Sodium estimates by emission. Inductively coupled plasma (ICP) was used to measure sulphate levels in solution, as

well as organic and inorganic S. The advantage of ICP is in its low standard errors. Micronutrients such as Zinc, Iron, Manganese and Copper were estimated using an Atomic absorption spectrophotometer.

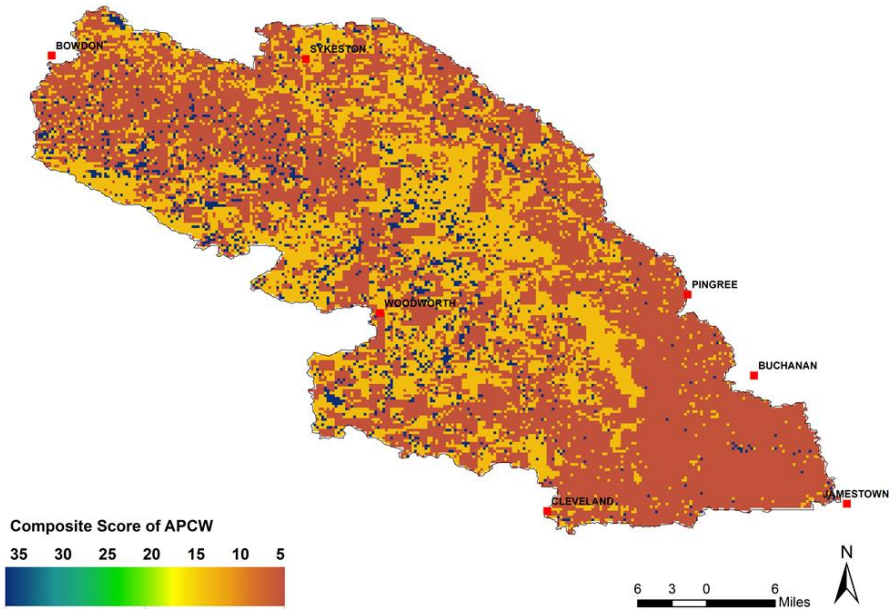

Figure 2.8. Index of the ability to produce clean water (I_{APCW}) within the study area. The composite score shows areas within the watershed having higher or lower ability to produce clean water.

2.3.3. Analysis of water quality data

Water samples were collected from 8 different sites along the Pipestem Creek-James subregion of the Missouri river in 2011 and 2016. Again, areas that depicted a low composite score in the map of the I_{APCW} (Figure 2.8.) were kept in consideration while choosing the sampling points. These were independently tested in an EPA certified laboratory (Fargo Cass Public Health Environmental Laboratory in North Dakota) using standard EPA methods, sound Colorimetric and Ion Chromatography (IC) principles. The EPA 300.0 method (Pfaff, 1993) was used to analyse Nitrate-Nitrite, Sulphate, Chloride and fluoride amounts in the water. All samples that contained particles larger than 0.45 microns and reagent solutions that contain

particles larger than 0.20 microns were filtered prior to any IC analysis. This method involves introducing 2-3 ml of the water sample to an Ion Chromatograph where the anions of interest are separated, measured, using a system comprised of a guard column, analytical column, suppressor device, and conductivity detector. A 1 mL of concentrated eluent (7.3 100X) to l00 mL of each standard and sample was added for presence of negative peaks near the fluoride peak can usually be eliminated by the addition of the equivalent. The EPA 200.7 method (Martin et al., 1992) was used to quantify amounts of Calcium, Iron, Manganese, Magnesium, Sodium, Potassium and total hardness. This method involves multi-elemental determinations by an Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) using sequential or simultaneous instruments. The instruments measure characteristic atomic-line emission spectra by optical spectrometry. SM2320 B Titration Method was used for carbonates, bicarbonates, hydroxides and total alkalinity where a pH meter was used (APHA, 1915). This method involves hydrolysis of solutes that react with the addition of standard acid whereas alkalinity would depend on the pH used. The amount of Total Dissolved Solids (TDS) was analysed using the SM2540C method where a well-mixed sample is filtered through a standard glass fiber filter, and the filtrate is evaporated until dry in a weighed dish to constant weight at 180°C. The increase in dish weight represents the total dissolved solids (Brown, 1979).

2.4. Results and discussion

GIS Weighted Overlay scoring (Figure 2.8) on a 30-meter grid generated an index that ranked (high to low) the watershed based on its ability to produce clean water. Results generated were primarily driven by affiliated land use. Areas that are darker with a high composite score from 25 to 30 represent areas within the watershed having a higher ability to produce clean water which corresponds to high riparian cover, high forest cover, low soil erodibility, low agricultural

land, low road density and low housing density. Similarly, areas that are lighter in colour, with a low composite score from 5 to 15, are areas that have a very low ability of producing clean water corresponding to low riparian cover, low forest cover, high soil erodibility, high agricultural land, high road density and high housing density.

Table 2.2. *In situ* water sample data from the sampling locations collected within the Pipestem Creek watershed in 2011 showing the sediment load.

Site number	**Total Dissolved Solids (TDS) mg/l**	**Total Hardness (TH) mg/l**	**Total Alkalinity (TA) mg/l**
1	852	474	245
2	628	336	163
3	1020	493	292
4	408	186	176
5	383	206	104
6	345	264	62
7	637	381	46
8	532	310	87

This indicates a direct geographic connection between forests, water and people. Forest cover (Figure 2.3) on the southern part of the watershed is almost negligible as most of the agricultural fields lie there. This region shows a very low I_{APCW}. The entire watershed has a very low riparian buffer (Figure 2.2) to protect the streams from the adjacent land use. Soil erosion ability within the watershed is shown in figure 2.6. where the darker areas depict very high soil erosion probability whereas the lighter areas show very low soil erosion probability. This

indicates a direct geographic connection between forests, water and people. Forest cover (Figure 2.3) on the southern part of the watershed is almost negligible as most of the agricultural fields lie there. Figure 2.4 shows the percentage of agricultural land. Darker areas signify higher percentage of fallow land whereas lighter areas signify less agricultural practice. Figures. 2.5 and 2.7 show road density and housing density information consecutively. These variables do not depict vivid information due to the fact that the study area is predominantly an agricultural watershed with very few residential land and proper metalled roads.

Table 2.3. *In situ* water sample data from the sampling locations collected within the Pipestem Creek watershed in 2011 showing anion content.

Site number	**Nitrate-Nitrite as N (mg/l)**	**Phosphate (mg/l)**	**Chloride (mg/l)**	**Sulphate (mg/l)**	**Bicarbonate (mg/l)**
1	0.2	1.14	15.9	411	245
2	0.2	0.9	5.63	317	163
3	0.2	2.15	12.8	503	292
4	0.1	2.2	8.39	533	176
5	0.2	0.41	13.5	169	104
6	0.1	1.8	3.55	96	62
7	0.2	0.91	11.2	284	46
8	0.2	0.98	10.9	208	87

Table 2.4. *In situ* water sample data from the sampling locations collected within the Pipestem Creek watershed in 2011 showing cation content.

Site number	Calcium (mg/l)	Iron (mg/l)	Magnesium (mg/l)	Manganese (mg/l)	Potassium (mg/l)	Sodium (mg/l)
1	77.6	0.02	67.9	0.042	17.7	73.8
2	61.1	0.228	44.5	0.03	10.4	59.8
3	86.9	0.055	67	0.087	12.4	121
4	41.1	0.316	20.3	0.135	8.13	50.8
5	37.8	0.02	27.2	0.351	8.07	33.8
6	52.9	0.02	31.9	0.09	15	12.5
7	73	0.035	48.3	0.095	11.8	60.4
8	62.1	0.02	37.7	0.425	12.2	53.4

The tables display the data from the water samples collected. The presence of Total Dissolved Solids (TDS) in large amounts ranging from 852 mg/l to 1020 mg/l can be seen (Table 2.2). Higher levels of Sulphate (SO_4) were seen ranging from 96mg/l to 533 mg/l (Table 2.3). Recommended limits of Sulphate in water for water used as a domestic water supply are below 250 mg. Although such high levels are not toxic to humans directly, they are an indicator of non-point source pollution within a watershed (Xia et al.,2016). Bicarbonates were present in the water ranging from 87mg/l to 250mg/l in most sites (Table 2.3). Thus, the total hardness of water was very high ranging from 150 to 500mg/l, in turn raising the alkalinity of water. Nitrate, Phosphate and Chloride levels were not significant to be studied. The soil samples contained soluble salts such as sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and magnesium (Mg^{2+}) along

with anions chloride (Cl^-), sulphate (SO_4^{2-}), nitrate (NO_3^-), bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}). Out of these, Calcium was found in very high levels ranging from 2,000 ppm to 2,900 ppm followed by magnesium and sodium. Since the watershed is enclosed within agricultural lands, fertilizer residues washed out from the adjoining fields could attribute to these higher levels of nutrients. Agricultural land use within watersheds have been linked to increased nutrient concentrations in river waters via wastewater, fertilizer use, cultivation of N fixing crops, and atmospheric deposition (Xia et al., 2016). The cation content (Table 2.4) in water showed presence of calcium ranging from 37 to 87mg/l, Sodium ranging from 53 to 122mg/l and Magnesium ranging from 20 to 68 mg/l which is not significantly high but they can accumulate leading to a high Sodium Absorption Ratio in the water.

A comparison of the water quality data was done within this study to verify and quantify the water quality data (Table 2.3 and 2.4). Excessive nutrient loading causes eutrophication of lakes and streams (Chislock et al., 2013). Increased nutrient and algae concentration can lead to water quality problems when these concentrations reduce water clarity, harm wildlife and reduce recreational uses (Carpenter et al., 1998). Decaying algae decreases dissolved oxygen concentration making the streams and lakes unable to support fish and other aquatic life (Carpenter et al., 1998). A higher concentration of Total Dissolved solids (TDS) was noted along with Total Hardness (TH) and Total Alkalinity (TA) (Table 2.2). Field data from 2011 and USGS field data from 2013 to 2016 from Pingree in North Dakota were compared which produced extremely high levels of TDS ranging from 1130mg/L to 1340mg/L along with Total Hardness ranging from 618mg/L to 1130mg/L (Table 2.5 and 2.6). The EPA Secondary Regulations advise a maximum contamination level (MCL) of 500mg/L for TDS (Regulations, 2012). Generally, high levels of TDS are caused by the presence of potassium, chlorides and

sodium which is evident from the data collected within the Pipestem Creek. Sulphate salts are a major contaminant in natural waters. Results from the field data have shown that sulphate content in the watershed is very high. Problems caused by sulphates are most often related to their ability to form strong acids which changes the pH (DeZuane, 1997). Sulphate ions also are involved in complexing precipitation reactions which affect solubility of metals and other substances (DeZuane, 1997). High sodium concentrations found in water samples indicate a high pH, lack of oxygen inadequate nutrients in the water. This region has noticeable excessive algal bloom. High levels of bicarbonate are found which has increased the Sodium Adsorption Ratio (SAR) Index of the water. The SAR is a ratio that measures the relative concentration of sodium to calcium and Magnesium. This Index is indicative of the alkalizing effect within the watershed (Clark et al., 2006). The FAO Document Repository states that excessive sodium in water promotes soil dispersion and structural breakdown, which can result in water infiltration problem due to soil dispersion and plugging and sealing of the surface pores similar to water with low alkalinity (Ayers et al., 1985). Intensive agricultural activities affect ecological and environmental quality and affect water quality (Hosono et al., 2007). One of the most significant impacts is from increased non-point source pollution loading, which has caused serious water pollution problems in recent decades (Ouyang et al., 2016). Soil samples generated similar results showing that the soil was excessively alkaline with presence of large amounts of soluble salts. These may be the effects of land use, which include irrigation patterns such as using excess salts on agricultural fields or use of excessive fertilizers. This watershed can be termed as impaired owing to the fact that it is a source of Non-point source pollution resulting from the extensive agricultural fields (Ouyang et al., 2016). The surface geology of this region is composed of glacial till which is mostly clay, naturally occurring aluminium silicate (NRCS,

2005). Clay imparts plasticity and is relatively impermeable to water making this region aptly facilitate surface runoff, letting nutrients to flow into the stream.

Table 2.5. Comparison of year-wise water quality data for the Pipestem Creek outlet at Pingree in North Dakota. Data was acquired from USGS Data portal (https://waterdata.usgs.gov). Accessed on August 15, 2016.

Year	Nitrate (mg/l)	Phosphate (mg/l)	Sulphate (mg/l)
2013	0.03	0.121	573
2014	0.016	0.172	802
2015	0.33	0.07	408
2016	0.12	0.09	612

Table 2.6. Comparison of year-wise water quality data for the Pipestem Creek outlet at Pingree in North Dakota. Data was acquired from USGS Data portal (https://waterdata.usgs.gov). Accessed on August 15, 2016.

Year	Bicarbonate (mg/l)	Calcium (mg/l)	Sodium (mg/l)	Potassium (mg/l)	Manganese (mg/l)	Magnesium (mg/l)	TDS (mg/l)	TH (mg/l)
2013	337	125	115	13.9	138	71.4	1130	689
2014	826	232	206	16.8	270	97.1	1870	1130
2015	619	146	122	9.61	575	69.9	1110	678
2016	543	167	143	11.4	256	121	1340	817

2.5. Comparison to a distributed model

The Annualized Agricultural Non-Point Source Model (AnnAGNPS) is used to evaluate non-point source pollution in impaired watersheds. It is currently used by many USA agencies to investigate non-point source pollution problems. The distributed parameter feature of the model allows spatial simulation (Pease et al., 2010). Pease *et al.* (2010) used the AnnAGNPS model to determine the nutrient status of the Pipestem Creek in North Dakota, USA. It was also used to predict the total runoff. The effectiveness of the model relies on the fact that it could only be effectively applied for a large agricultural watershed. Runoff predicted by the AnnAGNPS model for the Pipestem creek watershed between 2004 and 2006 showed a coefficient value of 3.17. This relatively low coefficient value indicated that the AnnAGNPS model predicted runoff from the watershed satisfactorily (Pease et al., 2010). To further validate the present study, the AnnAGNPS model, which is a distributed model, was compared to our spatial model in predicting impaired areas within the watershed. A Boolean 'And' operation was executed using 3D Analyst in ArcMap 10.4.1. The mathematical operation generated an output value of one, if both the input values were found true. The average annual sediment load data from the AnnAGNPS model was combined with the I_{APCW} score data to generate a new raster dataset. The new dataset was clipped in ArcMap so that the image showed the same areal extent used in the AnnAGNPS model. Figure 2.9 shows the resultant image where areas of high sediment load overlap with a low $I_{APCW.}$ The data was sorted into a frequency distribution table where the class interval was kept three for the first two classes. This was done to create a comparison scale between I_{APCW} and sediment load within the area that categorises the sediment load into high, moderate and low regions. Within the Pipestem Creek, areas yielding less cropland correspond to areas with increased sediment load. Sediment predicted by the

AnnAGNPS model showed limited values, but comparing average annual sediment loads to the land use data layer, it appeared that non-cropland areas did not significantly contribute to any sediment loads. The spatial distribution generated by the AnnAGNPS study showed fallow areas produce significant amounts of sediment loads from the watershed. These same locations in this study generated a low $I_{APCW.}$ Nitrogen and Phosphorus levels were not compared since the model predicted low nutrient levels due to surface runoff. AnnAGNPS is a distributed model which uses explicit data thus it is not compatible with small watersheds (Yu et al., 2013).

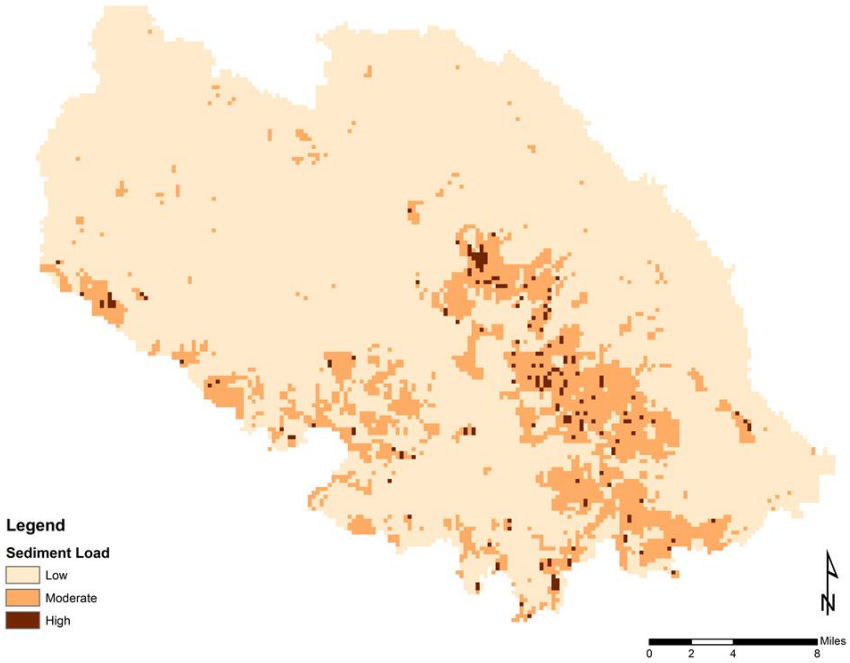

Figure 2.9. Comparative study with ANNAGNPS model: Part of the Pipestem Creek watershed (dark brown) showing highlighted areas of increased average annual sediment load. Sediment load data was derived from Pease et al., 2010.

2.6. Conclusion

The levels of TDS are much higher in the data of years 2014 to 2016 compared to the year 2011 which points out to the fact that the alkalizing effect within the watershed has increased. The goal of the study was used to identify NPS Pollution areas under limited data

conditions which to be practical and credible. Considering the costs incurred and time constraints in monitoring a watershed, this provides a cost effective preliminary method to identify impairment within a small agricultural watershed such as the Pipestem Creek. This paper combined geographic and anthropogenic variables to locate impaired areas within a watershed. We can adduce that the method employed in this study can be applied to a small watershed albeit the model is limited to data type and does not take into account a temporal component. The fusion of real time data within ArcGIS environment has improved the reliability of the Index of the ability to produce clean water output, and extended existing GIS functionalities. The implementation of the tool enables decision makers to follow a comprehensive yet easy-to-use procedure to examine weight sensitivity in both criteria and geographic space. The Pipestem reservoir is listed as having a high Total Maximum Daily Load (TMDL) for nutrients and eutrophication. Conservation practices that can be used to address these water quality issues include grazing management, erosion control, nutrient and pest management, as well as, agricultural waste management, and riparian buffers.

2.7. References

American Public Health Association, American Water Works Association, Water Pollution Control Federation and Water Environment Federation, 1915. *Standard methods for the examination of water and wastewater* (Vol. 2).

Anderson, J.R., 1976. A land use and land cover classification system for use with remote sensor data (Vol. 964). *US Government Printing Office.*

Ayers, R.S. and Westcot, D.W., 1985. Water quality for agriculture (Vol. 29). Rome: *Food and Agriculture Organization of the United Nations.*

Barnes, M., Todd, A., Lilja, R.W. and Barten, P., 2009. Forests, Water and People: Drinking water supply and forest lands in the Northeast and Midwest United States. *United States Department of Agriculture Forest Service.*

Bender, O., Boehmer, H.J., Jens, D. and Schumacher, K.P., 2005. Analysis of land-use change in a sector of Upper Franconia (Bavaria, Germany) since 1850 using land register records. *Landscape Ecology,* 20(2*),* pp. 149-163.

Brown, E., Skougstad, M.W. and Fishman, M.J., 1970. *Methods for collection and analysis of water samples for dissolved minerals and gases* (No. 05-A1). US Govt. Print. Off.

Ceballos-Silva, A. and Lopez-Blanco, J., 2003. Delineation of suitable areas for crops using a Multi-Criteria Evaluation approach and land use/cover mapping: a case study in Central Mexico. *Agricultural Systems,* 77(2), pp. 117-136.

Chen, Y.C., Lien, H.P. and Tzeng, G.H., 2010. Measures and evaluation for environment watershed plans using a novel hybrid MCDM model. *Expert systems with applications,* 37(2), pp. 926-938.

Chin, A., 2006. Urban transformation of river landscapes in a global context. *Geomorphology,* 79(3), pp. 460-487.

Chislock, M.F., Doster, E., Zitomer, R.A. and Wilson, A.E., 2013. Eutrophication: causes, consequences, and controls in aquatic ecosystems. *Nature Education Knowledge,* 4(4), p. 10.

Clark, M.L. and Mason, J.P., 2006. Water-quality characteristics, including sodium-adsorption ratios, for four sites in the Powder River drainage basin, Wyoming and Montana, water years 2001-2004. *USGS Scientific Investigations Report 2006-5113*.

Davis, M.L. and Masten, S.J., 2004. Principles of environmental engineering and science. *McGraw-Hill, New York*, p. 704.

DeZuane, J., 1997. Handbook of drinking water quality. *John Wiley & Sons*, pp. 36-58.

Fei,Z., Tiyip, T., Kung, H. and Jianli, D., 2010. The change of land use/cover and characteristics of landscape pattern in arid areas oasis: An application in Jinghe, Xinjiang. *Geo-Spatial Inf. Sci*., 13 (3), pp. 174–185.

Hosono, T., Nakano, T., Igeta, A., Tayasu, I., Tanaka, T. and Yachi, S., 2007. Impact of fertilizer on a small watershed of Lake Biwa: use of sulfur and strontium isotopes in environmental diagnosis. *Science of the Total Environment*, 384(1), pp. 342-354.

Iqbal, M.F. and Khan, I.A., 2014. Spatiotemporal land use land cover change analysis and erosion risk mapping of Azad Jammu and Kashmir, Pakistan. *The Egyptian journal of remote sensing and space science*, 17(2), pp. 209-229.

Jansen, L.J. and Di Gregorio, A., 2002. Parametric land cover and land use classifications as tools for environmental change detection. *Agriculture, ecosystems & environment*, 91(1), pp. 89-100.

Jensen, J.R., 2005. Introductory Digital Image Processing: a Remote Sensing Perspective (third ed) *Prentice Hall. Upper Saddle River*, NJ. pp. 505–512.

Kollias, V.J. and Kalivas, D.P., 1998. The enhancement of a commercial geographical information system (ARC/INFO) with fuzzy processing capabilities for the evaluation of land resources. *Computers and Electronics in Agriculture*, 20(1), pp. 79-95.

Leh, M., Bajwa, S. and Chaubey, I., 2013. Impact of land use change on erosion risk: an integrated remote sensing, geographic information system and modeling methodology. *Land Degradation & Development*, 24(5), pp. 409-421.

Madurapperuma, B., Rozario, P., Oduor, P. and Kotchman, L., 2015. Land use and land cover change detection in Pipestem Creek watershed, North Dakota. *International Journal of Geomatics and Geosciences*, 5(3), pp. 416-426.

Martin, T.D., Brockhoff, C.A., Creed, J.T. and Long, S.E., 1992. Determination of metals and trace elements in water and wastes by inductively coupled plasma-atomic emission spectrometry. *Methods for determination of metals in environmental samples*. CRC Press Inc., Boca Raton, pp. 33-91.

Mendoza, M.E., Granados, E.L., Geneletti, D., Pérez-Salicrup, D.R. and Salinas, V., 2011.Analysing land cover and land use change processes at watershed level: a multitemporal study in the Lake Cuitzeo Watershed, Mexico (1975–2003). *Applied Geography*, 31(1), pp. 237-250.

Mohammadi, M., Sheikh, V. and Saddodin, A., 2013. Development and Application of the GFHM Distributed Hydrologic Model for Flood Hydrograph Modeling (Case Study: The Jafarabad Watershed, Golestan Province, IRAN). *Journal of water engineering*. vol 5/no 15/ Winter 2012. pp. 13-30.

NRCS, 2005. SSURGO data layers.
Available at http://soils.usda.gov/survey/geography/ssurgo/

NRCS, 2007. SSURGO data layers.
Available at http://soils.usda.gov/survey/geography/ssurgo/

Osborne, L.L. and Wiley, M.J., 1988. Empirical relationships between land use/cover and stream water quality in an agricultural watershed. *Journal of Environmental Management*, 26(1), pp. 9-27.

Ouyang W, Jiao W, Li X, Giubilato E, Critto A., 2016. Long-term agricultural non-point source pollution loading dynamics and correlation with outlet sediment geochemistry. *Journal of Hydrology*. 540, pp. 379-385.

Pelorosso, R., Leone, A. and Boccia, L., 2009. Land cover and land use change in the Italian central Apennines: A comparison of assessment methods. *Applied Geography*, 29(1), pp. 35-48.

Pease, L.M., Oduor, P. and Padmanabhan, G., 2010. Estimating sediment, nitrogen, and phosphorous loads from the Pipestem Creek watershed, North Dakota, using AnnAGNPS. *Computers & Geosciences*, 36(3), pp. 282-291.

Pfaff, J.D., 1993. Method 300.0 Determination of inorganic anions by ion chromatography. *US Environmental Protection Agency, Office of Research and Development, Environmental Monitoring Systems Laboratory*, pp. 1-28.

Ritter, W.F. and Shirmohammadi, A. eds., 2000. Agricultural nonpoint source pollution: *Watershed management and hydrology*. CRC Press, pp. 5-9.

Station, M.A.E., 1998. Recommended chemical soil test procedures for the North central region. *North Central Regional Research Publication*, p. 221.

Sekar, I. and Randhir, T.O., 2007. Spatial assessment of conjunctive water harvesting potential in watershed systems. *Journal of Hydrology*, 334(1), pp. 39-52.

Soranno, P.A., Hubler, S.L., Carpenter, S.R. and Lathrop, R.C., 1996. Phosphorus loads to surface waters: a simple model to account for spatial pattern of land use. *Ecological Applications*, 6(3), pp. 865-878.

Turner, M.G., O'Neill, R.V., Gardner, R.H. and Milne, B.T., 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape ecology*, 3(3-4), pp. 153-162.

Weidner, E. and Todd, A., 2011. From the forest to the faucet: Drinking water and forests in the US, *Methods Paper. Washington, DC: USDA Forest Service*.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y. and Huang, Z., 2006. Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and urban planning*, 75(1), pp. 69-80.

Xia, Y., Ti, C., She, D. and Yan, X., 2016. Linking river nutrient concentrations to land use and rainfall in a paddy agriculture–urban area gradient watershed in southeast China. *Science of the Total Environment*, *566*, pp. 1094-1105.

Yu, X., Bhatt, G., Duffy, C. and Shi, Y., 2013. Parameterization for distributed watershed modeling using national data and evolutionary algorithm. *Computers & Geosciences*, *58*, pp. 80-90.

CHAPTER 3. TRANSITION MODELING OF LAND-USE DYNAMICS IN THE PIPESTEM CREEK, NORTH DAKOTA, USA[2]

3.1. Abstract

Significant land use changes in North Dakota have been reported and are widespread over the entire state. Such changing patterns may portend localized impairment to agricultural watersheds. In this study, Land Use Land Cover (LULC) change was modeled using geostatistics. The study area was within the Pipestem Creek watershed, a part of the Missouri Watershed James Sub-region of North Dakota, USA. Landsat Thematic mapper images from the years 2007, 2011, and 2015 were used as preliminary data. LULC information for these datasets were acquired from the Global Land cover facility and Landsat Program. Data analysis, spectral classification and post classification techniques were applied on the datasets. A transition matrix was derived using a Markov chain Monte Carlo (MCMC) model. This study demonstrates that the integration of satellite remote sensing, GIS and statistics may be an effective approach for analyzing the direction, rate, and spatial pattern of land use change.

3.2. Introduction

Over the last decade, a range of models of land use change have been developed to meet land management needs, and to better assess and project the future role of LULC change in the functioning of the Earth system. Modeling, especially if done in a spatially explicit, integrated and multi-scale manner, is an important technique for the projection of alternative pathways into

[2] The material in this chapter was co-authored by Papia F. Rozario, Dr.Peter Oduor, Larry Kotchman and Michael Kangas (Published in Journal of Geoscience and Environment Protection, *5*(03), p.182. https://doi.org 10.4236/gep.2017.53013).
Papia F. Rozario had primary responsibility for conducting this research including collecting field samples. Papia F. Rozario was the primary developer of the conclusions that are advanced here. Papia F. Rozario also drafted and revised all versions of this chapter. Dr.Peter Oduor served as proofreader and checked the math in the statistical analysis conducted by Papia F. Rozario.

the future, for conducting experiments that test our understanding of key processes, and for describing the latter in quantitative terms (Velkamp et al., 2001; Lambin et al., 2000; Lambin et al., 2001). Satellite remote sensing, in conjunction with geographic information systems (GIS), has been widely applied and been recognized as a powerful and effective tool in detecting LULC change (Ehlers et al., 1990; Meaille and Ward, 1990; Treitz et al., 1992; Westorland and Stow 1992; Harris and Ventura, 1995; Yeh and Li, 1996; Yeh and Li, 1997; Yeh and Li, 1999). Multispectral satellite data is cost effective and the information obtained from them can be used as inputs to build LULC datasets. GIS technology provides a flexible environment for spatial and statistical analyses coupled with modeling. Satellite imagery has been used to monitor discrete land cover types by spectral classification or to estimate biophysical characteristics of land surfaces via linear relationships with spectral reflectance or indices (Weng, 2002; Steininger, 1996). With easy accessibility of upgraded remote sensing software and readily available satellite imagery, the change in LULC can be assessed over a period of time (Madurapperuma et al., 2015). Particularly for applications that link remote sensing with human activity, this differentiation is important because land use emphasizes the functional role of land in economic activities while land cover does not (Kamusoko et al., 2009). Therefore, confounding land cover with land use may generate biased results in these studies (Seto et al., 2002). The models of LULC change process fall into two groups: regression-based and spatial transition-based models (Weng, 2001). The majority of research utilizes regression-based approach, which relates the locations of LULC change to a set of spatially explicit variables, and uses models such as logistic (Landis, 1994; Turner et al., 1996; Wear et al., 1998), and hedonic price models (Geoghegan et al., 1997). Cellular automaton simulation models are a type of spatial transition based models, which allow for predicting future land development based on

probabilistic estimates with Monte Carlo or other methods (Acevedo et al., 1995; Kokkinos and Maras, 1997). One crucial limiting factor to the development of process models is the lack of smart modeling tools for change processes in most software platforms. Equally important is the issue of data availability (Baker, 1989). Very few studies have attempted to link satellite remote sensing and GIS to stochastic modeling methods in LULC change studies. This paper presents a method that combines satellite remote sensing, GIS, and MCMC modeling to analyze and predict LULC changes in the Pipestem Creek, a part of the Missouri Watershed James Sub-region in North Dakota, USA between 2007 and 2015.

MCMC models are used to examine the stochastic nature of the LULC change data and to prioritize areas of impairment within an agricultural watershed. It is used as a descriptive and interrogative tool to quantify the change in land use occurring over a human - dominant landscape (Muller and Middleton, 1992). MCMC simulation models of LULC change aid in the understanding and analysis of interaction between impacts and natural resource management strategies (Brown et al., 2000; Oduor et al., 2012). Markov analysis of vegetation types tends to focus on a small area of less than a few hectares or on a single small plot (Jahan, 1986). When a few hundred hectares of land are involved, data sampling is usually applied to limit the workload to scattered plots or transects (Turner and Meyer, 1991). On the other hand, land use studies using MCMC models tend to focus on a much larger spatial scale, and involve both urban and non-urban covers (Drewett, 1969; McCauley, 2007; Bell and Hinojosa, 1977; Robinson, 1978; Jahan, 1986). Most of the studies utilizing MCMC models have used the first order of MCMC which was studied to be most suitable. MCMC have several assumptions. A primary assumption is to consider LULC change as a stochastic process (Bell and Hinojosa, 1977). Different categories are the states of the change which is defined as a stochastic process having

the property that the value of the process at time t, X_t, depends only on its value at time $t-1$, X_{t-1}, and not on the sequence of values $X_{t-2}, X_{t-3}, \ldots, X_0$ that the process passed through in arriving at X_{t-1} (Hunter, 2016; Weng, 2001). For $\{X_n, n \geq 0\}$ and $P\{X_t=i|X_{t-1}=j\} \forall\ i,j \in S$ where $S=\{1,2,\ldots,m\}$ then $P = [p_{ij}]$ (Hunter, 2016; Weng, 2001; Weng, 2002). Weng (2001) regarded the change process to be discrete for convenience with incremental time t (t = 0, 1, 2, 3,...) values.

Likewise, $P\{X_t=a_j \mid X_{t-1}=a_i\}$ is the transitional probability that makes the transition from state a_i to state a_j in one period of time . The MCMC Model used in this study was of first order homogeneous type. Therefore, $P\{X_t = i|X_{t-1} = j\} = p_{ij}$ can be applied (Wu et al., 2006; Yang et al., 2014). Here, p_{ij} can be calculated from observed data by estimating the number of times the particular observed data went from state i to j by adding the number of times the former state occurred. LULC change in itself is very dynamic, thus we cannot expect stationarity in it. However, stationary and discrete time has been used in various studies involving forest stands. Thus, MCMC models assumes two factors namely time stationarity/homogeneity and time independence. The concept of time stationarity or time homogeneity implies that equal interval in time or consistency between two states is considered within the timeline. Within a stationary MCMC and a set order, the transitional probabilities can be set through maximum likelihood estimation. The probabilities estimated are obtained by maximizing this function (Anderson et al., 1957). This estimate is just the relative frequency of transitions observed over the entire time. If the land use change sequence is a Markov process of order 0, the probability of the random variable X being in state j at time t can be determined. If the MCMC is of order 1, the probability of the random variable X being in state j at time t depends only on the last movement as aforementioned (Wu et al., 2006). Thus, testing Markov property is equivalent to testing that

the Markov process is of order 1. There are two steps in the testing procedure. First, the null hypothesis that the MCMC is of order 0 versus order 1 is tested; then the order 1 versus order 2 is tested. If the test of order 0 against order 1 is rejected, and the test of order 1 against order 2 is accepted, the process then may be assumed to be of order 1 (Anderson et al.,1957; Wu et al., 2006). In this study, spectral image classification and stochastic methods were utilized to address LULC changes by employing a finite first-order MCMC with stationary transition probabilities.

3.3. Materials and methodology

3.3.1. Image classification

Landsat 7 ETM+ images (Worldwide Reference System 2, row 027 and path 031) of the study area were acquired from the Global Land cover Facility and the Landsat Program. These images covered a span of three different years - 2007, 2011, and 2015. This freely available data has a ground resolution of 30 m. A thematic RGB band combination of bands 7, 4, and 2 was used. Data obtained was in the form of individual bands ranging from 1 to 7. Layer stacking created a new multiband file from the input bands which were resampled and re-projected. These datasets were resampled using the nearest neighbor algorithm so that their pixel brightness values were preserved. Upon layer stacking of the Landsat scenes; they were georeferenced to UTM Zone 14 North, WGS-84 Datum, then data calibration was applied which converted the digital numbers of the image to reflectance as a way of preprocessing the data. Supervised classification was performed using Maximum Likelihood algorithm. We used supervised classification because the data of the study area was available and we had a prior knowledge of the study area. Training site data was derived using digitization, which was then converted to polygonal training data. A total of 50 training sites were chosen for each image to ensure that all

spectral classes constituting each LULC category were adequately represented in the training statistics. Ten training sites for each of the five classes were chosen as ROIs (regions of interest). The reference data was collected from existing LULC maps that had been field-checked. Change detection statistics was performed on a remote sensing platform to generate a change matrix for the years 2007-2011, 2011-2015, and 2007-2015. The statistics are presented in a cross tabulation format that compares the change between the two maps, that is, the base map and the final map. We generated change matrices for years 2007 to 2011, 2011 to 2015, and 2007 to 2015. Pixel count and areal extent data for each class within each period was also generated. The maps generated from the change detection statistics were used as inputs to generate confusion matrices as a measure of accuracy assessment. The final change images were overlain with a vector file of the study areas and final maps were generated.

3.3.2. Spatial analyses

Reclassified images of 2007-2011, 2011-2015, and 2007-2015 were imported as raster datasets. The datasets were converted to ASCII using the raster conversion tool to generate the attribute table containing cell values. To estimate the land use transition from forested land to non-forested land, the datasets were reclassified using Spatial Analyst. The USGS Anderson Land Classification Scheme (Anderson, 1976) was used to classify all the remotely sensed datasets into new values. The categories included: (1) bare soil or barren land, (2) open waters, (3) cropland, (4) forested land, and (5) urban/built-up area. Tables 3.1, 3.2, and 3.3 represent the areal extent of change within each land use class for the study period 2007 to 2015. The amount of positive or negative change for each land use class is shown in the tables. Figures 3.1, 3.2, and 3.3 are graphical representations of the confusion matrix generated for the years 2007 to 2015.

Table 3.1. Land use/cover change matrix of Pipestem Creek from years 2007 to 2011 (in square km). Data was generated from Landsat imagery on a remote sensing platform, downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: August 26, 2016.

	Bare soil	**Open water**	**Cropland**	**Forested land**	**Urban built-up**	**2011 total**
Bare soil	29371.9	1107.3	3475.2	546.5	162.8	34663.6
Open water	6848.2	558.1	1740.2	260.5	4.7	9411.7
Cropland	432.7	19.0	82.5	17.4	220.5	772.1
Forested land	2288.3	446.9	1414.3	265.4	0.1	4415.1
Urban built-up	2872.5	1013.6	2182.8	644.7	4.4	6718.0
2007 total	39925.3	2698.2	7486.7	1469.5	1227.4	
Change %	-13.18	+28.81	+89.69	-77.16	+1.16	

Table 3.2. Land use/cover change matrix of Pipestem Creek from years 2011 to 2015 (in square km). Data was generated from Landsat imagery on a remote sensing platform, downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: August 26, 2016.

	Bare soil	**Open water**	**Cropland**	**Forested land**	**Urban built-up**	**2015 total**
Bare soil	829.5	4640.4	27829.5	2619.6	35520.4	35702.7
Open water	3385.5	2794.6	220.1	2341.7	8542.0	8542.3
Cropland	1055.1	453.2	4473.1	596.4	2177.8	2178.8
Forested land	1766.5	1490.1	14.9	1144.4	4415.9	4416.0
Urban built-up	261.7	34.1	233.0	16.0	544.7	1602.5
2011 total	468.8	617.7	6699.1	4573.6	5241	
Change %	+4.09	-9.24	+71.17	-34.27	+2.04	

Table 3.3. Land use/cover change matrix of Pipestem Creek from years 2007 to 2015 (in square km). Data was generated from Landsat imagery on a remote sensing platform, downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: August 26, 2016.

	Bare soil	**Open water**	**Cropland**	**Forested land**	**Urban built-up**	**2015 total**
Bare soil	920.2	944.5	30020.2	6543.7	281.9	35312.1
Open water	139.8	99.0	5880.5	520.0	0.6	8539.9
Cropland	1192.6	204.8	2641.9	138.3	1.1	2178.7
Forested land	2288.3	446.9	1414.3	265.4	0.1	4415.1
Urban built-up	625.4	2.9	28.3	2.1	943.6	1602.3
2007 total	1166.3	298.2	9486.7	1469.5	1227.4	
Change %	-9.84	-6.50	+89.90	-77.44	+10.55	

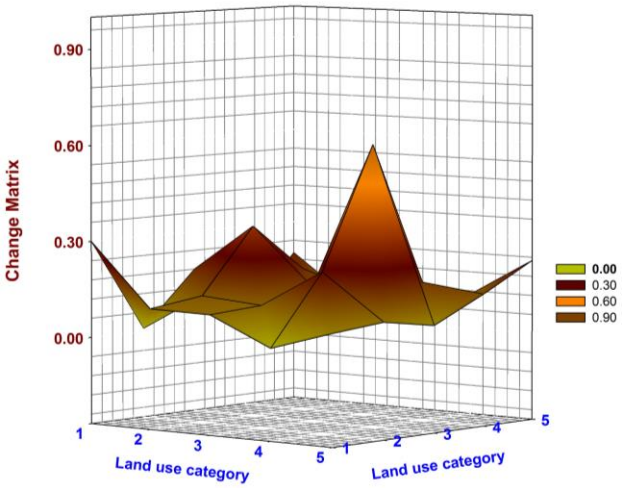

Figure 3.1. Change matrix data of Pipestem Creek for years 2007 to 2011. The peak signifies the change from forested land to agricultural land.

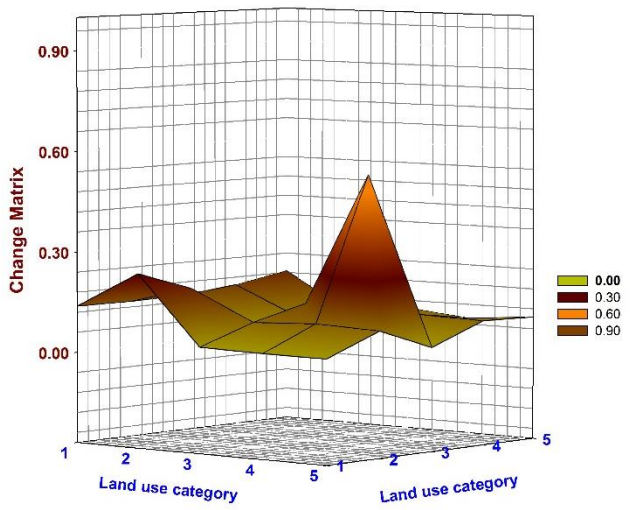

Figure 3.2. Change matrix data of Pipestem Creek for years 2011 to 2015. The peak signifies the change from forested land to agricultural land.

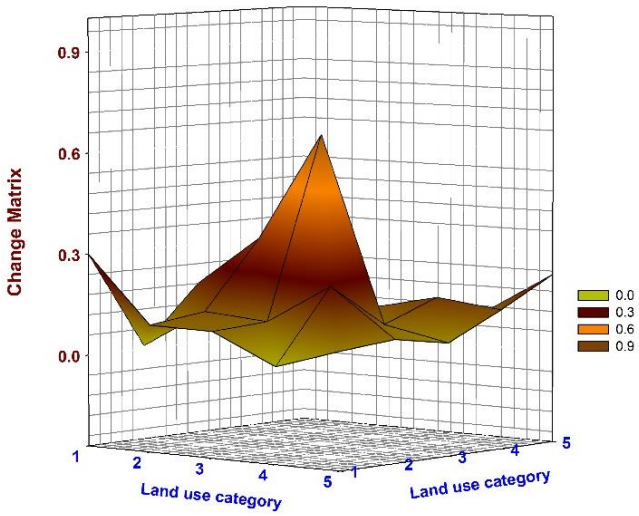

Figure 3.3. Change matrix data of Pipestem Creek for years 2007 to 2015. The peak signifies the change from forested land to agricultural land.

3.3.3. Markov analysis

The MCMC process was used for describing and projecting land use information within the watershed from satellite imagery. Data compatibility, stationarity and statistical independence were analyzed using the aforementioned MCMC process. The ASCII data generated for each year in the ArcGIS interface was imported into SemGRID 1.6.1 interface as a

layer. SemGRID 1.6.1 was used to generate the transitional probabilities of LULC data for years 2007 to 2011, 2011 to 2015, and finally 2007 to 2015. These probabilities were projected as 3D mesh plots using SigmaPlot® 10.0.

The testing of statistical independence hypothesis involved a procedure that compared the expected change with the actual change. If the number of LULC categories is M, then the statistic to be computed is Pearson's $\chi 2$ with $(M-1)^2$ degrees of freedom (Mondal et al., 2016). According to the MCMC hypothesis, the transition probability matrix governing the period 2007–2015 can be obtained by multiplying the 2007–2011 and 2011–2015 matrices. Table 3.4 shows the subset areal extent derived from the Markov Chain model where transition from each class is quantified in hectares as well as percentage value. Transition probability data derived is shown in Tables 3.5, 3.6 and 3.7. Kappa coefficient of agreement was derived based on (Jensen, 2005; Oduor et al., 2012). Reclassified Landsat datasets from the years 2007, 2011, and 2015 were combined using the combine function in ArcGIS-Spatial Analyst. Datasets from 2007 and 2011, 2011 and 2015, and finally 2007 and 2015 were combined. The attribute table for each dataset contained the pixel count information for each land use class. These attribute tables for each datasets were then exported to a database table file in MS-Excel. These were converted to text files and imported into MS-Assess where a crosstab query was performed to format and compress the data such that the row and column headings had the same class description as the reclassified datasets (Oduor et al.,2012). These tables were imported into Excel to derive the Kappa coefficients of agreement for each dataset.

Table 3.4. Subset of areal extent of Pipestem Creek in North Dakota generated by the Markov Chain Monte Carlo (MCMC) model.

State (from, to)	2007 to 2015			2011 to 2015			2007 to 2015		
	Area (%)	# of cells	Area (ha)	Area (%)	# of cells	Area (ha)	Area (%)	# of cells	Area (ha)
1,1	9.8	16163	1454.67	3.8	8362	752.58	5.5	26642	4247.10
1,2	11.7	19286	1735.74	3.7	8142	732.78	4.9	8008	811.35
1,3	12	19730	1775.70	4.0	8755	662.22	4.5	7496	674.64
1,4	9.4	15505	1395.45	6.9	14965	1346.85	6.7	10988	988.92
1,5	13.8	22701	2043.09	15.7	13412	3071.34	1.2	9015	2397.78
2,1	10.7	26899	2420.91	14.5	23708	3337.20	2.2	13946	5007.51
2,2	12.8	31944	2874.96	16.2	67118	6040.62	3.5	47190	788.31
2,3	13.3	33315	2998.35	10.9	53521	4816.89	11.4	27135	2578.14
2,4	17.7	44438	3999.42	10.3	26314	2368.26	11.8	28646	2654.73
2,5	9.4	23535	2118.15	2.9	7358	662.22	10.8	60232	1255.14
3,1	10.2	29155	2623.95	13.7	7004	3758.58	4.9	5639	3886.11
3,2	11.1	31810	2862.90	16.2	65581	7151.67	8.4	9372	3918.96
3,3	9.2	26289	2366.01	21.6	79463	5902.29	1.2	24651	6853.05
3,4	14.9	42715	3844.35	9.5	41762	630.36	10.6	183108	5420.88
3,5	13.2	37905	3411.45	2.3	28829	3164.40	11.0	53764	843.48
4,1	9.3	20957	1886.13	12.5	35160	6728.85	4.2	43179	2218.59

Table 3.4. Subset of areal extent of Pipestem Creek in North Dakota generated by the Markov Chain Monte Carlo (MCMC) model (Continued).

State (from, to)	2007 to 2015			2011 to 2015			2007 to 2015		
	Area (%)	# of cells	Area (ha)	Area (%)	# of cells	Area (ha)	Area (%)	# of cells	Area (ha)
4,2	11.3	25353	2281.77	20.6	74765	5619.96	11.0	22613	2658.06
4,3	14.4	25541	2298.69	22.3	4876	1992.42	13.1	47241	4838.76
4,4	10.9	38070	3426.30	7.9	22138	5268.24	23.9	86292	2170.17
4,5	11.5	25753	2317.77	1.5	4323	11552.31	19.2	2090	4251.69
5,1	7.5	47417	4267.53	12.6	58536	9847.98	3.6	29534	16479.72
5,2	9.1	57342	5160.78	17.6	12835	3078.63	7.5	9372	7766.28
5,3	8.6	54554	4909.86	13.6	10942	389.07	13.7	2465	18815.22
5,4	11.8	93284	8395.56	7.4	34207	438.84	33.1	1831	2035.17
5,5	10.5	97990	8819.10	1.0	62444	12541.1	29	53764	4251.69

Table 3.5. Transition probabilities matrix for states 1 to 5 for years 2007 to 2011 generated by the Markov Chain Monte Carlo (MCMC) model

	1	2	3	4	5
1	0.02999	0.01	0.4801	0.01	0.12
2	0.1044	0.01271	0.43	0.1785	0.106
3	0.101	0.1119	0.25517	0.1201	0.01153
4	0.0141	0.0014	0.323101	0.1643	0.1101
5	0.0751	0.0908	0.1623	0.1417	0.2301

Table 3.6. Transition probabilities matrix for states 1 to 5 for years 2011 to 2015 generated by the Markov Chain Monte Carlo (MCMC) model.

	1	**2**	**3**	**4**	**5**
1	0.0392	0.1387	0.0455	0.1568	0.2198
2	0.0375	0.0496	0.0419	0.1773	0.0537
3	0.0109	0.1019	0.2295	0.1259	0.0318
4	0.0393	0.0114	0.4371	0.5001	0.1121
5	0.0331	0.2025	0.0321	0.197	0.0353

Table 3.7. Transition probabilities matrix for states 1 to 5 for years 2007 to 2015 generated by the Markov Chain Monte Carlo (MCMC) model.

	1	**2**	**3**	**4**	**5**
1	0.1174	0.0699	0.1095	0.2744	0.1288
2	0.1515	0.00198	0.2051	0.2025	0.2211
3	0.0583	0.0701	0.1401	0.1659	0.0656
4	0.0499	0.1487	0.4127	0.0889	0.1998
5	0.2309	0.1488	0.0101	0.2301	0.2801

3.4. Results and discussion

3.4.1. LULC change statistics

The overall accuracy values based on the post classified images generated for 2007 - 2011, 2011 – 2015, and 2007 - 2015 using change detection statistics in ENVI 4.5® were 91.59 percent, 88.30 percent, and 89.43 percent respectively. Table 3.1 shows a representation of LULC change matrix from 2007 to 2011 where a likelihood of increment in 89.69% cropland,

28.81% open waters, and 1.16% urban area was prominent. The likelihood of decrement in forested land appeared to be at 77.16% along with a decrement of 9.06% for bare soil or barren land. Figure 3.1 shows the graphical representation of LULC change matrix from 2007 to 2011. Table 3.2 represents the data showing LULC change matrix from years 2011 to 2015 where a likelihood of increment in cropland, urban area and bare soil or barren land at 71.17%, 4.09%, and 2.04% respectively was noted. A net negative change in forested land and open waters was observed at 34.27% and 9.24% respectively. Forested land to non-forested land transition was found to be high for 2007 to 2011 period but was slightly lower from 2011 to 2015 period. Positive change in croplands was noted and could be a probable cause to the growing demand for food grains and agricultural products in this area. Figure 3.2 is a graph showing LULC change matrix from 2011 to 2015. Table 3.3 represents the data for change matrix from 2007 to 2015. About 89.90% likelihood of increment in agricultural land leading to a 77.44% likelihood of decrement in forested land in the area was noted. Figure 3.3 represents the graph for change matrix from 2007 to 2015.

3.4.2. Validation of LULC change process using Markov Chain

Markov Chain simulations for the Pipestem Creek watershed showed continuity between the trends in change from forested land to other land use classes, which was similar to the study results generated by Oduor et al., (2012). The range of forested land to cropland change was significant. Table 3.5. represents the transition probabilities of years 2007 to 2011 derived by Markov chain simulation. It represents cropland to cropland transition at 0.36 and forested land to cropland transition at 0.52. Figure 3.4 is graphical representations of the transition probability from the year 2007 to 2011 showing significant peak for categories 3 followed by 4 which represent cropland and forested land respectively. Urban or built-up land generated a low

transition probability of 12% from barren land. Table 3.6. represents the transition probabilities of years 2011 to 2015 derived by Markov chain simulation. Transition from forest to cropland was 44% and cropland to cropland was approximately 50%. Forested land transiting to cropland was relatively higher than that exhibited for the years 2007 to 2011. Barren land to urban area transition probability was high at 22%. Open waters did not show much transition throughout 2007 to 2015. Figure 3.5 is a visual representation of the tabular data from 2011 to 2015 where the only significant peak that can be seen is for cropland. Table 3.7 represents the transition probability data for years 2007 to 2015.in LULC over a period of 9 years where a significant transition probability from forested land to cropland was 41%. Figure 3.6 is a graphical representation of the same data where cropland showed up as a significant peak followed by forested land and urban land, indicating a likelihood of increment. The Kappa coefficients of agreement for the datasets were at 53.28% for the years 2007 to 2011, 55.74% for the years 2011 to 2015 and 60.24% for the years 2007 to 2015.

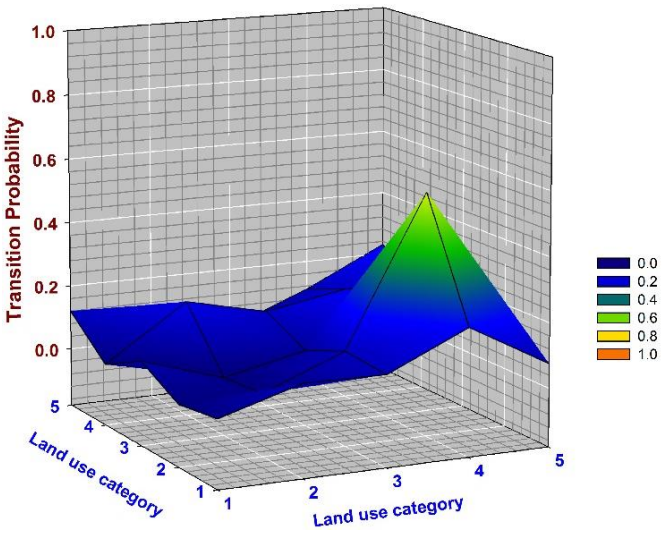

Figure 3.4. Transition probabilities for states 1 to 5 for years 2007 to 2011. The peak signifies the change from forested land to agricultural land within Pipestem Creek watershed.

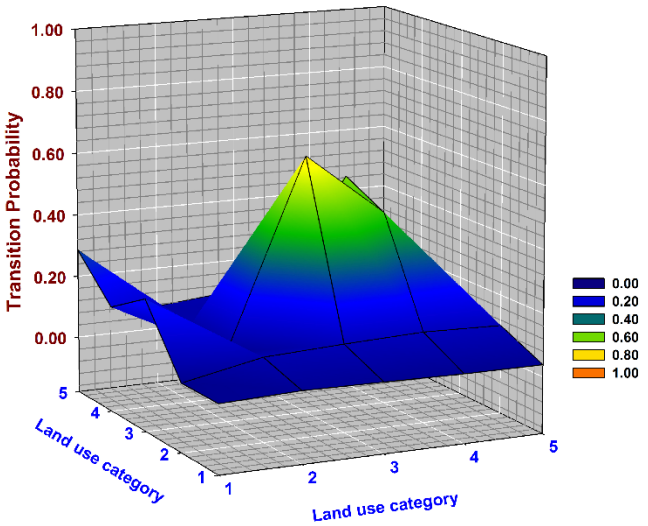

Figure 3.5. Transition probabilities for states 1 to 5 for years 2011 to 2015. The peak signifies the change from forested land to agricultural land within Pipestem Creek watershed.

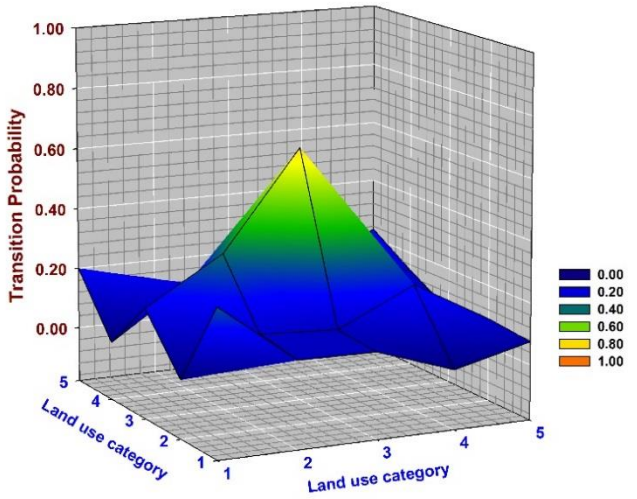

Figure 3.6. Transition probabilities for states 1 to 5 for years 2007 to 2015. The peak signifies the change from forested land to agricultural land within Pipestem Creek watershed.

Figures 3.7 to 3.9 represent LULC maps of the Pipestem Creek in North Dakota for the years 2007-2011, 2011-2015, and 2007-2015 respectively. The western parts of Stutsman and Wells County showed significant transition from forest to cropland. The southeastern part of Stutsman County showed a significant increase in urban or built-up land attributing to the

location of small towns like Jamestown. Oduor et al., (2012) used error matrices to study similarity between two datasets. A Kappa coefficient value would determine the similarity or dissimilarity between two images (Oduor et al., 2012). As such, low Kappa value may be indicative of low similarities between the datasets, which could also imply significant transition between various LULC classes. This prediction assumes spatial independence of the area units (Cabral and Zamyatin, 2009). Although Markov chains constitute a good tool for describing and projecting LULC quantities, they are insufficient for spatially explicit LULC predictions, because they also assume statistical independence of spatial units. In a similar study made by Cabral and Zamyatin (2009), it was suggested that Markov transitions can be used coupled with spatially explicit models like cellular automata and/or linear extrapolation models. The methodology presented in this study incorporated a spatial element along with a temporal model. The projections of future LULC changes on the basis of a MCMC model showed a continuing trend of increase in urban and agriculture land acreages, and the decline in forests and other natural vegetation covers. A similar study was conducted by Oduor et al., 2012, of statewide North Dakota, which depicted greater likelihood of forest to non-forested land transition especially along north central North Dakota. The prioritization map of North Dakota derived in Oduor et al., 2012 showed that part of the Pipestem creek watershed was a high priority area for Forest Stewardship Program. The MCMC model transition probabilities estimated from 2007 to 2015 showing change from forested land to other land use classes depicted a probability of future change and loss of forest acreage. The test of time homogeneity was performed to see that the process of land use change was stable throughout the full study period. This was done by comparing the data from each sub-period transition matrix i.e. 2007 to 2011, 2011, to 2015 to the full period transition matrix and then comparing the full period of 2007 to 2015.

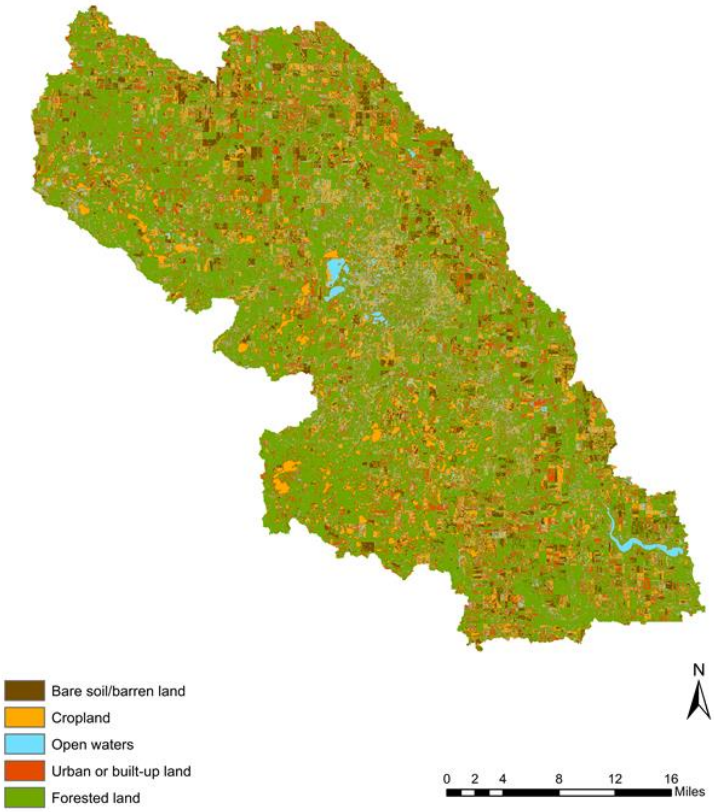

Figure 3.7. LULC map of the Pipestem Creek in North Dakota for years 2007 to 2011. Data was derived from Landsat imagery downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: August 26, 2016.

The test of time dependence in this dataset is evident in Tables 3.5, 3.6, and 3.7 where the change from forested land to agricultural land was significant from the base year to the transitioning year. The transition probabilities estimated from the full study period with an interval of 4 years was assumed to be time-stationary Markov transition matrix. This may be used to predict the future land use category distribution to provide answers to management problems as in Hall et al., (1991) study. Hall et al., (1991) also points out that early MCMC models were parameterized using data observed and measured from field surveys and air photography.

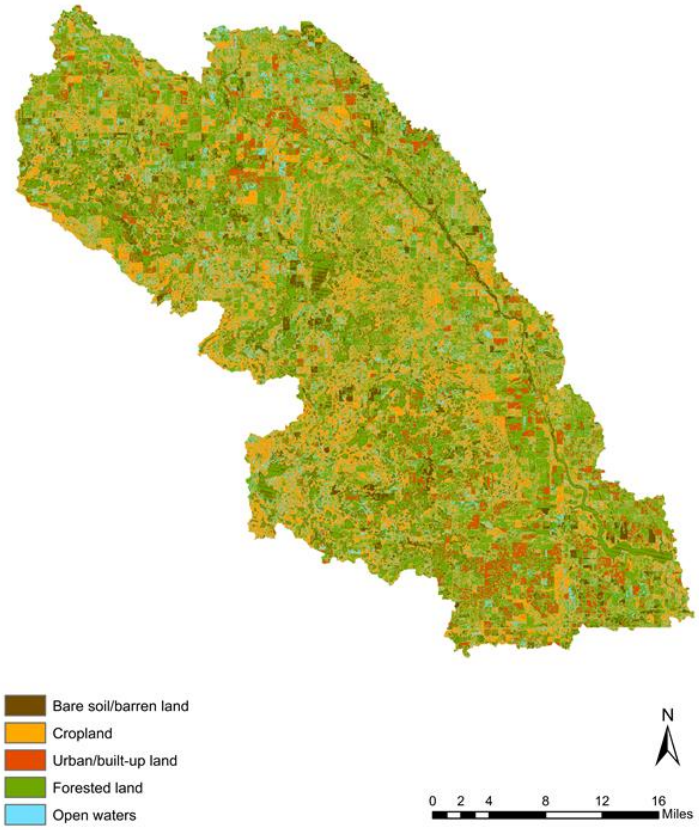

Figure 3.8. LULC map of the Pipestem Creek in North Dakota for years 2011 to 2015. Data was derived from Landsat imagery downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov) . Accessed: August 26, 2016.

These data tended to be biased and costly. The use of satellite remote sensing has enabled us to calculate less biased training sites from the full extent of the landscape as in Peterson et al., (2009). Thus, the issue of obtaining observed training sites is crucial (Peterson et al., 2009). The Markov chain models have shown the capabilities of descriptive power and simple trend projection for LULC change, regardless of whether or not the trend actually persists. The analysis can serve as an indicator of the direction and magnitude of change in the future as well as a quantitative description of change in the past. However, there are several limitations in LULC change applications. First, the model may vary with spatial resolution of the datasets and can be by rescaled (Santosh et al., 2017). Second, higher order models that, for

instance, include j2 states instead of j states in a first-order model would generate results that are less accurate (Mubea et al., 2011).

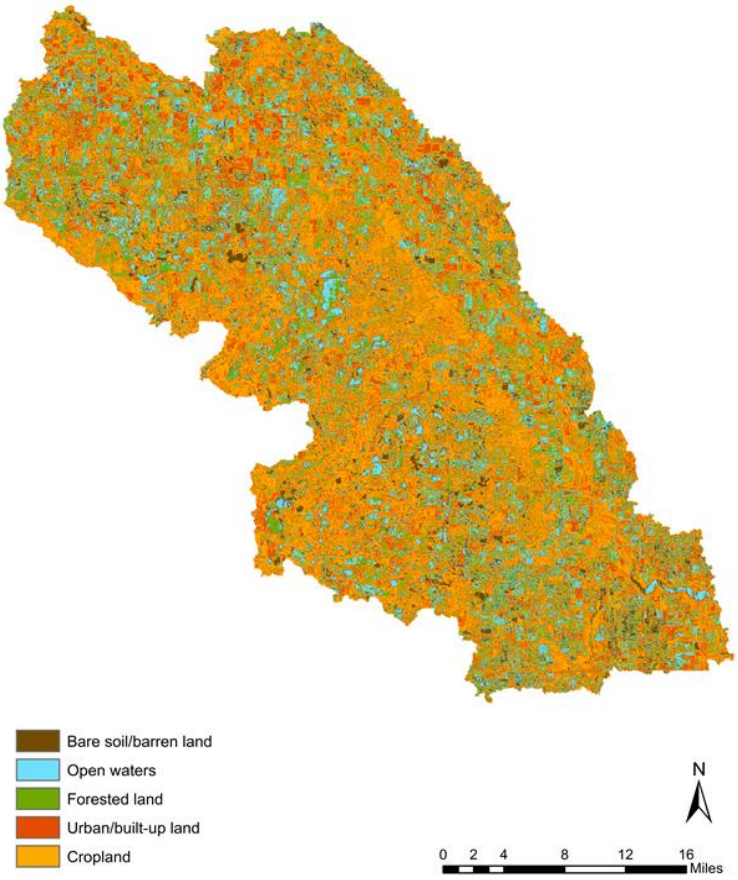

Figure 3.9. LULC map of the Pipestem Creek in North Dakota for years 2007 to 2015. Data was derived from Landsat imagery downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: August 26, 2016.

Eastman et al., (2005) shows a method to conditioning changes to the initial states in different sites, as well as in the final states, and therefore introduces spatial dependence into Markov modeling. So higher ordered effects can be studied once the spatial adjustments are made. Table 3.4. shows results generated with transition states which was corroborated in the transition matrix shown in Tables 3.5, 3.6, and 3.7 for years 2007 to 2015. Overall, areas used for all forms of agriculture increased by more than 50% of its original level at the first stage.

The built-up areas doubled in area compared to the initial stage. The Markov probabilities estimated from the full study period proved to be useful to analyze and predict the distribution of land use categories, though the land use change process cannot always be assumed static because of its dynamic nature (Eastman et al., 2005; Bell, 1974). The MCMC model, combined with the geospatial analysis proved to capacitate trend projections of the changing land use just like Sathees et al., (2015) study. This spatio-temporal model provides not only a quantitative description of change in the past but also the direction and magnitude of change in LULC in the future (Nurmiaty and Arif, 2014; Han et al., 2015; Vazquez-Quintero et al., 2016).

In terms of the quantitative accuracy, error rates for forest and agricultural land are particularly low indicating that nonparametric models can be successfully implemented in a further study of similar agricultural watersheds. In terms of the spatial accuracy for forest, and agricultural land, a confusion matrix generated low accuracy assessment as discussed earlier which were acceptable indicators. This indicates that the MCMC model can predict land use patterns objectively.

3.5. Conclusion

The MCMC model performance in predicting LULC distribution from 2007 to 2015 showed that it is possible to project land use change patterns with small deviations within acceptable limits. This study integrated land use pattern changes, MCMC model for the simulation of land use change. Landscape patterns depict complexities of spatial heterogeneity, and researchers have shown that these patterns can influence a variety of ecological phenomena. This methodology combines MCMC model and integrates it with post classification remote sensing techniques. We were able to predict LULC changes from 2007 to 2015. Future research includes the experimentation of spatially explicit models to better understand the LULC

dynamics of this area. When facing such severe and rapid LULC changes, one requirement for resource managers is to be able project future changes under certain assumptions to increase the awareness of ecological consequences. The MCMC model performance reveals the potential and the merit of using this approach for assessing future land use change in similar arid regions. Future analysis studies are recommended to use more detailed socio-environmental variables to improve the understanding of trends of LULC changes within such agricultural watersheds.

3.6. References

Anderson, J.R.1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Vol. 964, *US Government Printing Office*.

Anderson, T.W. and Goodman, L.A., 1957. Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, pp. 89-110.

Acevedo, M.E., Urban, D.L., Alban, M., 1995. Transition and gapmodels of forest dynamics. *Ecological Applications* 5, pp. 1040–1055.

Baker, W.L., 1989. A review of models of landscape change. *Landscape ecology*, 2(2), pp. 111-133.

Brown, D.G., Pijanowski, B.C. and Duh, J.D., 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of environmental management*, 59(4), pp. 247-263.

Bell, E.J. and Hinojosa, R.C., 1977. Markov analysis of land use change: continuous time and stationary processes. *Socio-Economic Planning Sciences*, 11(1), pp. 13-17.

Bell, E.J., 1974. Markov analysis of land use change—an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Sciences*, 8(6), pp. 311-316.

Cabral, P. and Zamyatin, A., 2009. Markov processes in modeling land use and land cover changes in Sintra-Cascais, Portugal. *Dyna*, 76(158), pp. 191-198.

Drewett, J.R., 1969. A stochastic model of the land conversion process: An interim report. *Regional studies*, 3(3), pp. 269-280.

Eastman, J.R., Van Fossen, M.E. and Solarzano, L.A., 2005. Transition potential modeling for land cover change. *GIS, spatial analysis and modeling*, pp. 357-386.

Ehlers, M., Jadkowski, M.A., Howard, R.R. and Brostuen, D.E., 1990. Application of SPOT data for regional growth analysis and local planning. *Photogrammetric Engineering and Remote Sensing*, 56(2), pp. 175-180.

Geoghegan, J., Wainger, L.A. and Bockstael, N.E., 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological economics*, 23(3), pp. 251-264.

Han, H., Yang, C., and Song, J., 2015. Scenario Simulation and the Prediction of Land Use and Land Cover Change in Beijing, China. *Sustainability*, 7, pp. 4260-4279.

Hall FG, Botkin DB, Strebel DE, Woods KD, Goetz SJ., 1991. Large-scale patterns of forest succession as determined by remote sensing. *Ecology*; 72(2), pp. 628-640.

Harris, P.M. and Ventura, S.J. 1995. The Integration of Geographic Data with Remotely Sensed Imagery to Improve Classification in an Urban Area. *Photogrammetric Engineering and Remote Sensing*, 61, pp. 993-998.

Hunter, J.J., 2016. The Computation of Key Properties of Markov Chains via Perturbations. *Linear Algebra and its Applications*, pp. 176-202.

Jahan, S., 1986. The determination of stability and similarity of Markovian land use change processes: a theoretical and empirical analysis. *Socio-Economic Planning Sciences*, 20(4), pp. 243-251.

Jensen, J.R., 2005. Introductory Digital Image Processing: a Remote Sensing Perspective (third ed) *Prentice Hall. Upper Saddle River, NJ*. pp. 505–512.

Kamusoko, C., Aniya, M., Adi, B. and Manjoro, M., 2009. Rural sustainability under threat in Zimbabwe–simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. *Applied Geography*, 29(3), pp. 435-

Kokkinos, E.A., Maras, A., 1997. A first-order stationary Markov Class A transition density. *Journal of the Franklin Institute* 334B (4), pp. 525–537.

Lambin, E.F., Rounsevell, M.D.A. and Geist, H.J., 2000. Are agricultural land use models able to predict changes in land use intensity?. *Agriculture, Ecosystems & Environment*, 82(1), pp. 321-331.

Lambin, E., Turner, B., Geist, H., Agbola, S., Angelson, A., Bruce, J., Coomes, O., Dirzo, R., Fischer, G., Folke, C. and George, P., 2001. Our emerging understanding of the causes of land use and cover change. *Global Environmental Change*, 11(4), pp. 261-269.

Landis, J.D., 1994. The California Urban Futures Model: a new generation of metropolitan simulation models. *Environment and planning B: planning and design*, 21(4), pp. 399-420.

Madurapperuma, B., Rozario, P., Oduor, P. and Kotchman, L., 2015. LULC change detection in Pipestem Creek watershed, North Dakota. *International Journal of Geomatics and Geosciences*, 5(3), pp. 416-426.

McCauley, J. L. (2007). A comment on the paper "Stochastic feedback, nonlinear families of Markov processes, and nonlinear Fokker–Planck equations" by TD Frank. *Physica A: Statistical Mechanics and its Applications*, 382(2), pp. 445-452.

Méaille, R. and Wald, L., 1990. Using geographical information system and satellite imagery within a numerical simulation of regional urban growth. *International Journal of Geographical Information System*, 4(4), pp.445-456.

Mondal, M.S., Sharma, N., Garg, P.K. and Kappas, M., 2016. Statistical independence test and validation of CA Markov land use land cover (LULC) prediction results. *The Egyptian Journal of Remote Sensing and Space Science*, 19(2), pp.259-272.

Mubea, K.W., Ngigi, T.G. and Mundia, C.N., 2011. Assessing application of Markov chain analysis in predicting land cover change: a case study of Nakuru Municipality. *Journal of Agriculture Science and Technology*, 12(2), pp. 126-143.

Muller, M.R. and Middleton, J., 1994. A Markov model of land use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology*, 9(2), pp. 151-157.

NRCS, 2007. SSURGO Data Layers. Available online at http://websoilsurvey.nrcs.usda.gov/. Assessed [06/16/2016].

Nurmiaty, B. S., Arif S., 2014. GIS-Based Modelling of Land Use Dynamics Using Cellular Automata and Markov Chain. *Journal of Environment and Earth Science* Vol.4, No.4, pp. 61-66.

Oduor, P.G., Kotchman, L., Nakamura, A., Jenkins, S. and Ale, G., 2012. Spatially constrained forest cover dynamics using Markovian random processes. *Forest Policy and Economics*, 20, pp. 36-48.

Peterson, L.K., Bergen, K.M., Brown, D.G., Vashchuk, L., Blam, Y., 2009. Forested land cover patterns and trends over changing forestmanagement eras in the Siberian Baikal region. *Forest Ecology and Management* 257, pp. 911–922.

P. M. Harris and S. J. Ventura, "The Integration of Geographic Data with Remotely Sensed Imagery to Improve Classification in an Urban Area," Photogrammetric Engineering and Remote Sensing, Vol. 61, No. 8, 1995, pp. 993-998.

Robinson, V.B., 1978. Information theory and sequences of land use: an application. The Professional Geographer, 30(2), pp.174-179.https://doi.org/10.1111/j.0033-0124.1978.00174.x.

Rozario, P.F., Oduor, P., Kotchman, L. and Kangas, M., 2016. Quantifying Spatiotemporal Change in LULC and Assessing Water Quality: A Case Study of Missouri Watershed James Sub-Region, North Dakota. *Journal of Geographic Information System*, 8(06), p. 663.

Santhosh S., Pandey A., Mishra S., 2017. Modelling spatiotemporal land dynamics for a trans-boundary river basin using integrated Cellular Automata and Markov Chain approach. *Applied Geography*, Vol. 82, pp. 11-23.

Sathees K., Nisha R. and Samson M., 2015. Land use change modelling using a Markov model and remote sensing. *Geomatics, Natural Hazards and Risk*, Vol. 5, No. 2, pp. 145–156.

Seto, K.C., Woodcock, C.E., Song, C., Huang, X., Lu, J. and Kaufmann, R.K., 2002. Monitoring land use change in the Pearl River Delta using Landsat TM. *International Journal of Remote Sensing*, 23(10), pp. 1985-2004.

Steininger, M.K., 1996. Tropical Secondary Forest Regrowth in the Amazon: Age, Area and Change Estimation with Thematic Mapper Data. *International Journal of Remote Sensing*, 17, pp. 9-27.

Treitz, P.M., Howarth, P.J. and Gong, P., 1992. Application of satellite and GIS technologies for land cover and land use mapping at the rural-urban fringe: a case study. *Photogrammetric engineering and remote sensing*, 58(4), pp. 439-448.

Turner, M.G., Wear, D.N. and Flamm, R.O., 1996. Land Ownership and Land-Cover Change in the Southern Appalachian Highlands and the Olympic Peninsula. *Ecological applications*, 6(4), pp. 1150-1172.

Turner, B. and Meyer, W.B., 1991. Land use and land cover in global environmental change. *International Social Science Journal*, 43(130), pp. 669-679.

Vázquez-Quintero, G., Solís-Moreno, R., Pompa-García, M., Villarreal-Guerrero, F., Pinedo-Alvarez, C., & Pinedo-Alvarez, A. (2016). Detection and projection of forest changes by using the Markov Chain Model and cellular automata. *Sustainability*, *8*(3), p. 236.

Veldkamp, A. and Lambin, E.F., 2001. Predicting land use change. *Agriculture, ecosystems & environment*, 85(1), pp. 1-6.

Wear, D.N., Turner, M.G. and Naiman, R.J., 1998. Land cover along an urban–rural gradient: implications for water quality. *Ecological Applications*, 8(3), pp. 619-630.

Westorland, S. and Stow, D.A., 1992. Category identification of changes land use polygons in an integrated image processing geographic information system. *Photogrammetric Engineering and Remote Sensing*, 58(11), pp. 1593-1599.

Weng, Q., 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of environmental management*, 64(3), pp. 273-284.

Weng, Q., 2001. A remote sensing? GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. *International journal of remote sensing*, 22(10), pp. 1999-2014.

Wu, Q., Li, H.Q., Wang, R.S., Paulussen, J., He, Y., Wang, M., Wang, B.H. and Wang, Z., 2006. Monitoring and predicting land use change in Beijing using remote sensing and GIS. *Landscape and urban planning*, 78(4), pp. 322-333.

Yang, X., Zheng, X.Q. and Chen, R., 2014. A land use change model: integrating landscape pattern indexes and Markov-CA. *Ecological Modelling*, 283, pp. 1-7.

Yeh, A.G.O. and Li, X., 1996. Urban growth management in the Pearl River delta: an integrated remote sensing and GIS approach. *ITC journal*, 1, pp. 77-85.

Yeh, A.G.O. and Li, X., 1997. An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pearl River Delta, China. *International Planning Studies*, 2(2), pp. 193-210.

Yeh, A.G.O. and Li, X., 1999. Economic development and agricultural land loss in the Pearl River Delta, China. *Habitat international*, 23(3), pp. 373-390.

CHAPTER 4. SPATIAL DEPENDENCE OF LAND USE/COVER PREDICTIONS WITHIN PIPESTEM CREEK, NORTH DAKOTA, USA

4.1. Abstract

A major threat to biodiversity in North Dakota is the conversion of forested land to cultivable land, especially those that act as riparian buffers. To reverse this trend of transformation, a validation and prediction model is necessary to assess the change. Spatial prediction within a Geographic Information System (GIS) using Kriging is a popular stochastic method. The objective of this study was to predict spatial and temporal transformation of a small agricultural watershed - Pipestem Creek in North Dakota; USA using satellite imagery from 1976 to 2015. To enhance the difference between forested land and non-forested land, a spectral transformation method - Tasseled Cap's Greenness Index (TCGI) was used. To study the spatial structure present in the imagery within the study period, semivariograms were generated. The Kriging prediction maps were post-classified using Remote Sensing techniques of change detection to obtain the direction and intensity of forest to non-forest change. TCGI generated higher values from 1976 to 2000 and it gradually reduced from 2000 to 2011 indicating loss of forested land.

4.2. Introduction

Sustainable use of riverine systems and riparian habitats are directly affected by changing land use patterns (Rozario et al., 2016). Modeling land use patterns is an important technique for the projection of alternative pathways into the future (Veldkamp et al., 2001; Lambin et al., 2000; Lambin et al., 2001). Geographic Information Systems (GIS) combined with satellite remote sensing has varied application and has been recognized as a powerful tool in detecting land use and land cover change (LULC) (Ehlers et al., 1990; Méaille and Wald, 1990; Treitz et

al., 1992; Westmoreland and Stow, 1992; Harris and Ventura, 1995; Yeh and Li, 1996, 1997, 1999; Weng, 2001). Satellite data is cost effective and the information obtained from them can be used as inputs to build land use and land cover datasets (Singh, 1989). Spatial representation of the LULC change is essential for regional planners and management (Piccini et al., 2014). To elucidate the optimal use of land and to provide input data for watershed models, it is necessary to have information on existing LULC change patterns (Madurapperuma et al., 2015).

Geostatistics deals with problems pertaining to spatial serial data, mapping and interpolation of the data on a statistical platform that are related to a time analysis (Ripley 1981). It has an ability of distinguishing the continuous nature of LULC and is able to detect random variations during modeling, dependent on the spatial correlation within the ecosystem (McBratney et al., 2003). Prediction using sample points is carried out by the spatial behavior and spatial distribution of parameters to minimize the error while doing any type of image classification (Eldeiry and Garcia, 2010). Inverse Distance Weighting and Splines are deterministic interpolation methods to analyze change in land use patterns but these methods tend to oversimplify the results, as the spatial autocorrelation of the data is not considered (Robinson and Metternicht, 2006). A geostatistical method is usually preferred where sample data points can be transformed into continuous surfaces to understand the spatial autocorrelation within the data (McBratney et al., 2003). The parameters used for any analysis can be aggregated from pixels to object class representation using image segmentation (Wang et al., 2004).

Kriging Interpolation is a very popular geostatistical method (Huang and Chen, 2007; Mishra et al., 2009). Ordinary Kriging estimates the mean as a constant in the searching neighborhood (Isaacs and Srivastava, 1989). The Kriging technique has recently become very

common for analyzing spaceborne data (Oliver et al., 2000). The values of unsampled locations are estimated by Kriging models using weighted averaging of the known sampled locations, which provide a correlation among the neighboring values that can be modelled as a function of the geographical distances between each location across the study area within the variogram (Miller et al. 2007; Eldeiry and Garcia, 2010). Global and local information in predictions can be obtained from Kriging, but the ability of the variogram in describing spatial dependence is a function of the quality and quantity of the data samples (Miller et al. 2007; Eldeiry and Garcia, 2010). According to a study by Robertson (1987), exponential models are often best-fitted semivariogram models as they use the weighted least-squares method. Curran (1988) and Woodcock et al. (1988a, b) introduced the semivariogram to remote sensing and discovered that the parameters of the variogram can be directly related to a feature in an image. The primary assumption of a geostatistical analysis when assuming spatial continuity is that samples that are located close to each other are similar than samples that are far apart (Matheron, 1971). This variation in geographic data or the spatial relation can be analyzed from a semivariogram model. An ideal semivariogram has associated features such as the lag, nugget, range and sill. The direction and distance are commonly referred to as the lag, the nugget is variability at zero distance and represents sampling and analytical errors, the range of influence in a semivariogram designates the extent beyond which autocorrelation between sampling sites is very less or zero and the sill represents the variability of spatially independent samples (Bohling, 2005).

An effective way of mapping vegetation and analyzing LULC change is using the Tasseled Cap transformation (Jin and Sader, 2005; Bauer et al., 1993; Cohen & Spies, 1992; Cohen et al., 1995; Collins & Woodcock, 1996; Dymond et al., 2002; Fiorella & Ripple, 1995a, 1995b; Franklin et al., 2002; Skakun et al., 2003). The Tasseled-Cap Transformation is a

conversion of the original bands of an image into a new set of bands with defined interpretations that are useful for vegetation mapping (Watkins, 2005). The term tasseled cap comes from the shape of the plot of the data that resembles a cap. A tasseled-cap transform is performed by taking "linear combinations" of the original image bands - similar in concept to principal components analysis (Watkins, 2005). Tasseled Cap reduces the volume of the data without any loss of information and its spectral features are directly related to land features (Jin and Sader, 2005; Crist & Cicone, 1984; Crist & Kauth, 1986; Crist et al., 1986). This transformation in remote sensing is the conversion of the readings in a set of data into composite values that is the weighted sums of separate data readings (Dymond et al., 2002). One of these weighted sums measures roughly the brightness or greenness of each pixel in the scene (Dymond et al., 2002). Cohen et al. (1995) reported that the composite values are linear combinations of the values of the separate data readings, but some of the weights are negative and others are positive. The composite values represent the degree of greenness of the pixels or the degree of yellowness of vegetation or perhaps the wetness of the soil (Cohen et al., 1995). Usually there are just three composite variables listed within a remote sensing interface. The Tasseled Cap transformation of Landsat thematic mapper (TM) consists of six multispectral features, all of which could be potentially differentiated in terms of stability and change in a multitemporal dataset (Collins and Woodcock, 1996; Crist and Kauth, 1986; Crist, 1985; Jin and Sader, 2005). The first three features, which are brightness index, greenness index, and wetness index, respectively usually account for the most variation in a single-date image (Collins and Woodcock, 1996; Crist and Kauth, 1986; Crist, 1985; Jin and Sader, 2005). Collins and Woodcock (1996) analyzed Landsat data for environmental studies and found the Tasseled Cap transformation to be a consistent indicator of assessing forest change as it captures Shortwave Infrared (SWIR). Tasseled Cap

Greenness Index (TCGI) would ideally identify forest cover but it is less sensitive to any topographic effect (Collins and Woodcock, 1996). Cohen and Spies (1992) led a study to distinguish old growth and mature forests in the Pacific Northwest using Landsat datasets. In their study, the Tasseled Cap - brightness index did not separate old growth and mature forests due to their sensitivity to topography but the greenness and the wetness index were able to identify forest disturbances.

The primary objectives of this study were: (1) to apply Ordinary Kriging (OK) Interpolation technique to smooth TCGI values, extracted from 30 m to 60 m spatial resolution Landsat images in order to assess spatio-temporal transformations; (2) to apply change detection techniques to the interpolated prediction maps to yield the intensity of the LULC change.

4.3. Materials and methodology

4.3.1. Image processing

Image classification and processing was done on a remote sensing platform - ENVI®4.5. Six Landsat images (Table 4.1) covering the study site were downloaded from the Global Land Cover Facility (GLCF, 2017). The images were acquired by different sensors (MSS, TM, and ETM+) and were from six different time periods, as listed in Table 4.1. The Landsat images were processed by applying a dark object subtraction and then converting the image digital numbers to reflectance values. Dark object subtraction was applied to remove shadows, scattering and electrical gains within the datasets (e.g. Chavez, 1996). This was done to obtain a sound quantitative analysis of the images. Reflectance values were used to calculate several vegetation indices for each image subset. These include the Normalized Difference Vegetation Index (NDVI) and Tasseled Cap indices. NDVI was calculated in ENVI® 4.5 using the equation NDVI = (NIR-Red)/(NIR+Red) (Weier and Herring, 2000). Band 3 was used as red and band 4

was used as near IR to generated NDVI and the output datasets were saved as floating point data type. Tasseled Cap was calculated in ENVI®4.5 using the *Transform* tool where the reflectance images of years 1976 to 2015 were used as inputs. The output image generated four bands – Brightness index, Greenness index (TCGI), Wetness index and a null or Non-index. These individual bands were displayed and linked to acquire regions of interest (ROI) representing forested areas. 30 training sites were acquired. Spectral separability analysis was performed using ROI Separability tool on NDVI and TCGI, incorporating mean and standard deviation values of extreme classes in each scene, to analyze the most suitable index for differentiating between forested areas and non-forested areas using methodologies of Kaufman and Remer (1994) and Lasaponara (2006). The ROI statistical results displayed univariate statistics such as minimum value, maximum value, mean, standard deviation among other values. Since the resolution of MSS and TM/ETM+ are different, for the MSS image, a 3x3 pixels window was selected and for the TM/ETM+ images, 6x6 pixels window was selected. The window values were averaged to generate new pixel values using raster calculator in ArcMap® 10.4. Thus, the resolution of the images was reduced by factors of 3 and 6, to specifically fit the MSS and TM/ETM+ images respectively.

4.3.2. Geostatistical analysis

To better understand the spatial structure of the imagery on a given date and location, an empirical semivariogram model such as Ordinary Kriging was applied to the sampled data. The six transformed datasets corresponding to the six years were imported to ArcMap® 10.4 where they were clipped to the watershed boundary shapefile. Geostatistical Analyst was used to perform OK on each dataset for each model type: Gaussian, Exponential, J-Bessel, K-Bessel, Circular, and Spherical (Appendix C). To consider the model that best fitted the study, certain

parameters were considered - (1) Cross-validation scatter plot where the measured and predicted values were compared by using the difference between them, (2) Mean estimation error where the difference between the estimated and the known point values were considered and (3) Mean standardized squared estimation error. These parameters were generated in ArcMap® 10.4.1 using Geostatistical Analysis (Appendix C). Based on these parameters, among other Kriging models, the Exponential model was found to be the best fit for this study.

Table 4.1. Landsat time series scenes for years 1976 to 2015 used in the study, downloaded from EarthExplorer portal (http://earthexplorer.usgs.gov). Accessed: October 15, 2016.

Satellite/Sensor	**Location**	**Date**	**Path/row***
Landsat - 2 MSS	North Dakota	06/08/1976	34/27
Landsat - 4 TM	North Dakota	31/08/1991	31/27
Landsat - 7 ETM+	North Dakota	30/08/2000	31/27
Landsat - 7 ETM+	North Dakota	18/08/2005	31/27
Landsat - 7 ETM+	North Dakota	22/08/2011	31/27
Landsat - 8 ETM+	North Dakota	19/08/2015	31/27

*Path/row of the MSS image is listed according to Worldwide Reference System-1 (WRS-1) while those of TM and ETM+ are according to WRS - 2.

A smoothing factor of 0.2 was applied to the search neighborhood type for all the datasets. The scatterplots derived from the model for each datasets is shown in Fig.4.3. The resultant interpolation images are shown in Fig.4.2. Isotropic distribution was assumed in all cases, similar to the study by Friedland et al., (2016).

4.4. Results and discussion

4.4.1. Image processing analysis

Figure 4.1 is the visual representation of TCGI for the Pipestem creek watershed. The greener areas represent forest cover and the light green to white areas represents non - forests to barren areas. TCGI is high for the years1976 to 1991 and it gradually decreases from 2000 to 2015. The images generated for NDVI for years 1976 to 2015 (Appendix C1) showed similar results to TCGI when compared visually, but the separability values were better than that for NDVI. TCGI efficiently determined each class on what it resembled most in the base image. TCGI was originally designed to examine vegetation properties, its advantage lies in its ability to compare different sensors with different spectral bands, as it subsets different spectral bands to one normalized layer of TCGI values (Crist and Cicone, 1984; Huang et al., 2002; Kauth and Thomas, 1976).

4.4.2. Semivariogram analysis

All six TCGI images were used for the geostatistical analysis. First, empirical semivariograms for the six periods were established. The rationale for using a semivariogram model was the similarity in spatial structure of most of its variables, gradually increasing or decreasing as a function of the increasing distance from the river until the boundary of the watershed, and the typical shape of the variogram. In the current case, the presence or absence of *sill* may be an indicator of presence or absence of forested areas. The level of *sill* may be related to the level of spatial correlation within the watershed. Therefore, if semi-variance reaches its maximum point (*sill*), beyond that, the data may not be correlated. The range may be the defining boundary of the watershed since it incorporates all the pixel values within the image that are correlated. Results of the cross-validation analyses of the exponential model are

presented in Table 4.2. For an ideal model, the Mean prediction error should be near 0 (this investigates bias), Root Mean Square (RMS) prediction error should be small, average standard error should be close to RMS error, Mean-standardized prediction error should be near 0 and RMS standardized prediction error should be near 1, indicating that the estimated prediction uncertainty is consistent (Stein, 2012). In the current study, the RMS ranged from 0.017 to 0.031, average standard error ranged from 0.016 to 0.063, which is quite close to the RMS values. Mean-standardized prediction error values ranged from 0.007 to 0.041. RMS standardized prediction error values were closer to 1 ranging from 0.591 to 0.681. The slope coefficient was very close to unity and the intercept coefficient was very close to zero, proving the ability of the chosen exponential model to reproduce the observed values (Oliver et al., 2000). The least-squares measure of fit was used, incorporating exponential models, as shown in Fig 4.2. All semivariograms were processed with 20 *lags* of 1,000 m each. Because of the irregular distribution of forests and non-forests, data values exactly separated by 1,000 m could not be expected, thus the *range* of 1,000 m to 20,000 m was selected. Lag values were determined by trial and error process to optimize the above-mentioned criteria. OK interpolation maps were produced based on the exponential models with the parameters presented in Table 4.2. Figure 4.3 shows the scatterplots generated for cross validation of the model. These graphs show the spatial correlation of the data. Most of the data values lie along the line in the scatterplots indicating how closely related the data is. During groundtruthing and field observations, no evidence was found to support an anisotropic pattern, as in Friedland et al., (2016) study, that may explain the direction of forest reduction or the direction in which the agricultural lands are increasing. So, isotropic distribution was assumed in all cases. Figure 4.2 depicts the final results for the distribution of the TCGI values for the six periods.

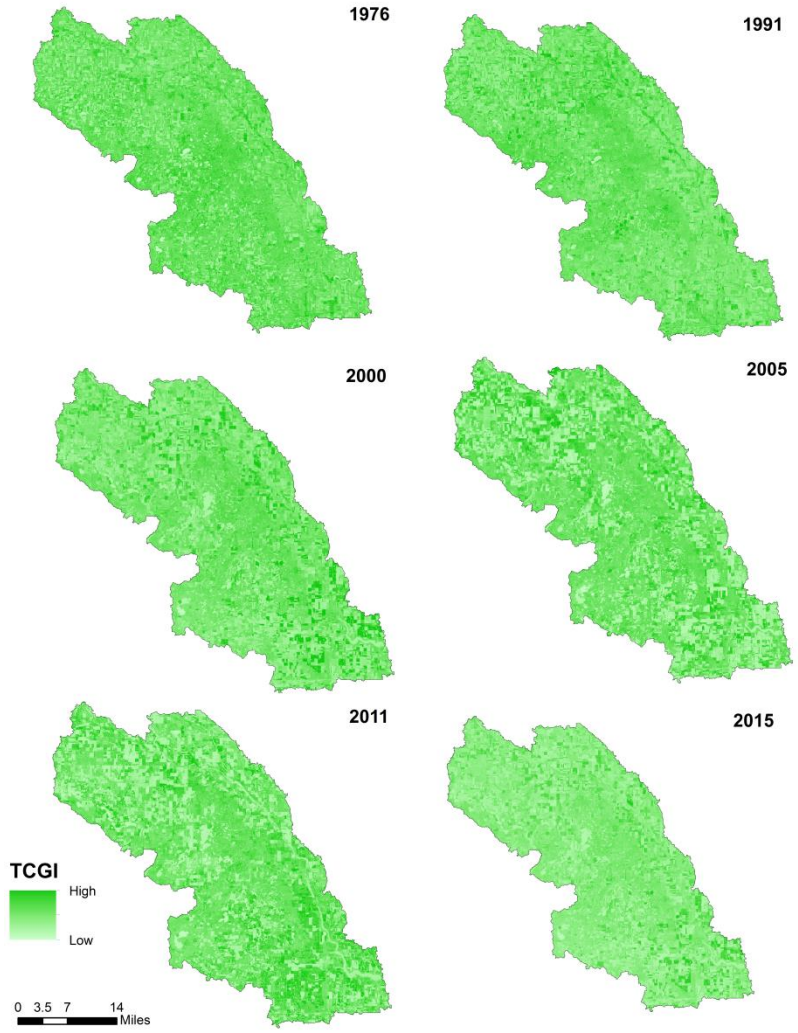

Figure 4.1. TCGI images of Pipestem Creek watershed for years 1976 to 2015. Data was generated from Landsat imagery downloaded from the Global Land Cover Facility (http://www.landcover.org). Accessed: February 12, 2016. High TCGI values depict areas with dense forest cover and lower values of TCGI depict areas devoid of forests.

The dark-red areas in the images are related to forested areas. The surrounding light red and yellow belts represent a mixed zone where forested areas and non-forested areas overlay each other or create a stable spectral balance. The zone colored by blue tones is considered to be non-forested areas that include mostly agricultural land. Forested areas are more concentrated in the 1976 image and is gradually seen to reduce for the rest of the images through 2015.

Table 4.2. Semivariogram parameters for the exponential model fitting the TCGI products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m) for years 1976 to 2015. Data was generated from TCGI images.

	1976	**1991**	**2000**	**2005**	**2011**	**2015**
Sill	0.0015	0.0012	0.0007	0.0011	0.0008	0.0010
Range (m)	742.5	762.5	797.9	1030.7	1079.1	1101
Root-mean-square (RMS)	0.031	0.038	0.017	0.017	0.028	0.023
Average standard error	0.062	0.056	0.016	0.060	0.063	0.054
Mean standardized	0.041	0.033	0.021	0.027	0.007	0.015
Root-mean-square-standardized	0.591	0.681	0.657	0.613	0.672	0.651

4.5. Conclusion

TCGI was selected to describe the spatial surface patterns since it produced the best contrast in terms of separability among the spectral indices. It produced the best contrast in terms of separability among all examined spectral indices. TCGI was able to compare between the different sensors with different spectral bands, as it reduced their different spectral bands to one normalized layer of TCGI values. The semivariance analysis was found to be a suitable method for gaining insight to the spatial structure present in the imagery for a given date and location. The similarity between the shape of the semivariogram and the directional change of the environmental variables is a logical reason for using this method. The Kriging interpolation

technique was used as a smoothing filter in which each pixel was being replaced with the solution for the semivariogram equation (exponential model in the current case) calculated from all other pixels in the image. As a result, it reduced spatial errors and fine scale variability and helped to better identify the transition from forest to non-forested areas.

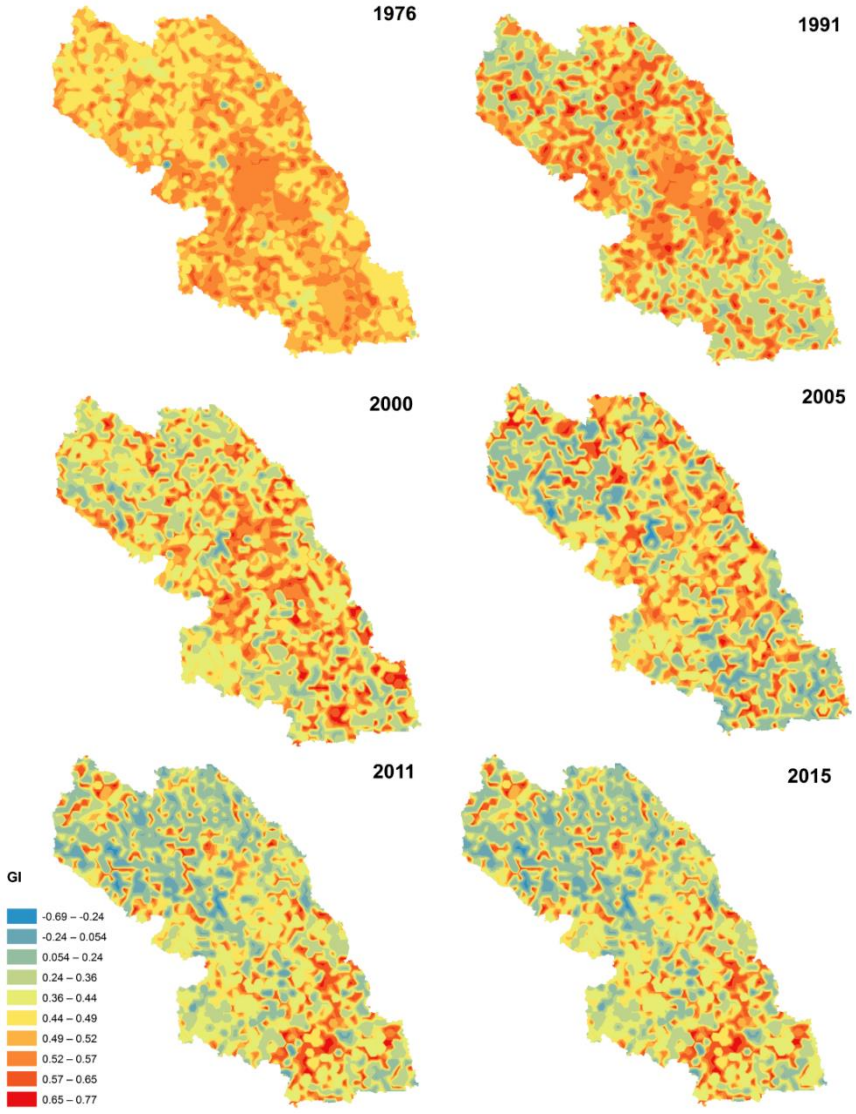

Figure 4.2. Kriging Interpolation maps of Pipestem Creek watershed based on TCGI values for years 1976 to 2015. Data was generated from Landsat imagery downloaded from the Global Land Cover Facility (http://www.landcover.org). Accessed: February 12, 2016

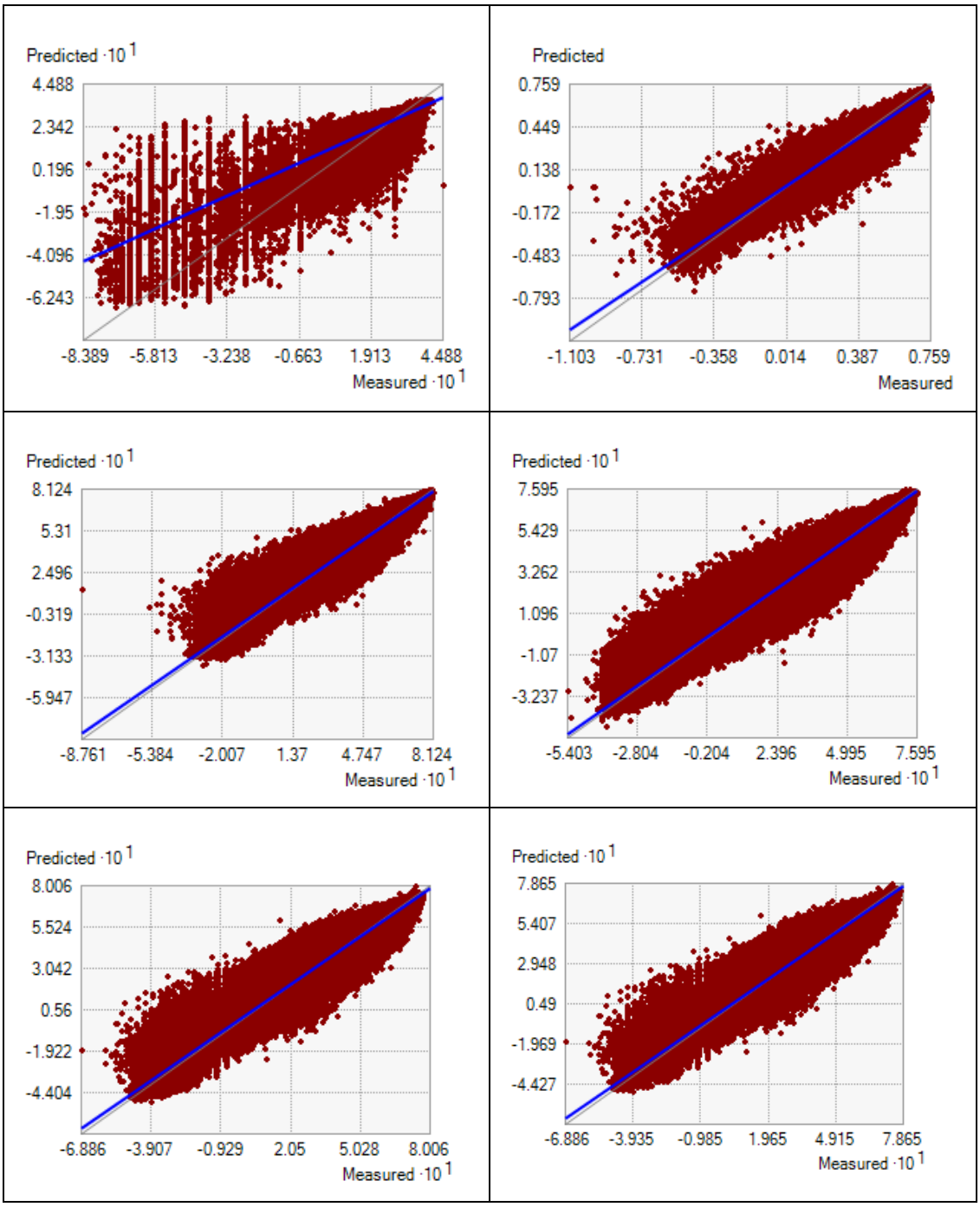

Figure 4.3. Cross validation scatterplots of the semivariogram model of Pipestem Creek watershed datasets for years 1976 to 2015. The red dots are the data points. The graphs show that the data are highly correlated as the predicted values and the measured values are close to each other.

4.6. References

Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Befort, W. and Heinzen, D.F., 1993. Satellite inventory of Minnesota forest resources. *National Aeronautics and Space Administration*, Grant NAGW-1431. NRA, pp. 87-106.

Bohling, G., 2005. Introduction to geostatistics and variogram analysis. *Kansas geological survey*, pp. 1-20.

Chavez Jr., P.S., 1996. Image-based atmospheric corrections — revisited and improved. *Photogrammetric Engineering and Remote Sensing* 62, pp. 1025–1036.

Cohen, W. B., & Fiorella, M., 1998. Comparison of methods for detecting conifer forest change with Thematic Mapper imagery. Remote sensing change detection: *Environmental monitoring methods and applications*, pp. 89-102.

Cohen, W. B., Fiorella, M., Gray, J., Helmer, E., & Anderson, K., 1998. An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, 64(4), pp. 293 – 300.

Cohen, W. B., & Spies, T. A., 1992. Estimating structural attributes of Douglas-Fir /Western Hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, 41, pp. 1 – 17.

Cohen, W. B., Spies, T. A., & Fiorella, M., 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. *International Journal of Remote Sensing*, 16, pp. 721 – 746.

Collins, J. B., & Woodcock, C. E., 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment*, 56, pp. 66 – 77.

Crist, E. P., 1985. A TM Tasseled Cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment*, 17(3), pp. 301 – 306.

Crist, E. P., & Cicone, R. C., 1984. A physically-based transformation of thematic mapper data—the TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, 22(3), pp. 256 – 263.

Crist, E. P., & Kauth, R. J., 1986. The tasseled cap de-mystified. *Photogrammetric Engineering and Remote Sensing*, 52(1), pp. 81 – 86.

Crist, E.P., Laurin, R. and Cicone, R.C., 1986 . Vegetation and soils information contained in transformed Thematic Mapper data. *Proceedings of IGARSS'86 Symposium*, pp. 1465-1470.

Curran, P.J., 1988. The semivariogram in remote sensing: an introduction. *Remote sensing of Environment*, 24(3), pp. 493-507.

Dymond, C.C., Mladenoff, D.J. and Radeloff, V.C., 2002. Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sensing of Environment*, 80(3), pp. 460-472.

Ehlers, M., Jadkowski, M.A., Howard, R.R. and Brostuen, D.E., 1990. Application of SPOT data for regional growth analysis and local planning. *Photogrammetric Engineering and Remote Sensing*, 56(2), pp. 175-180.

Eldeiry, A.A. and Garcia, L.A., 2010. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images. *Journal of Irrigation and Drainage Engineering*, 136(6), pp. 355-364.

Fiorella, M. and Ripple, W.J., 1995a. Analysis of conifer forest regeneration using Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing (American Society for Photogrammetry and Remote Sensing),* v. 59, no. 9, pp. 1383-1388.

Fiorella, M. and Ripple, W.J., 1995b. Determining successional stage of temperate coniferous forests with Landsat satellite data. *Photogrammetric Engineering and Remote Sensing (American Society for Photogrammetry and Remote Sensing*), v. 59, no. 2, pp. 239-246.

Franklin, J.F., Spies, T.A., Van Pelt, R., Carey, A.B., Thornburgh, D.A., Berg, D.R., Lindenmayer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C. and Bible, K., 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. *Forest Ecology and Management*, *155*(1), pp. 399-423.

Friedland, C.J., Joyner, T.A., Massarra, C., Rohli, R.V., Treviño, A.M., Ghosh, S., Huyck, C. and Weatherhead, M., 2016. Isotropic and anisotropic kriging approaches for interpolating surface-level wind speeds across large, geographically diverse regions. *Geomatics, Natural Hazards and Risk*, pp. 1-18.

GLCF, 2017. Available online at http://www.landcover.org/data/landsat/. Accessed on [12/2/2016].

Harris, P.M. and Ventura, S.J., 1995. The integration of geographic data with remotely sensed imagery to improve classification in an urban area. *Photogrammetric engineering and remote sensing*, *61*(8), pp. 993-998.

Haugen, D.E., Harsel, R., Bergdahl, A., Claeys, T., Woodall, C.W., Wilson, B.T., Crocker, S.J., Butler, B.J., Kurtz, C.M., Hatfield, M.A. and Barnett, C.H., 2013. North Dakota's

Forests 2010. *Resour. Bull. NRS-76. U.S. Department of Agriculture, Forest Service, Northern Research Station.* p 52.

Huang, C., Wylie, B., Yang, L., Homer, C. and Zylstra, G., 2002. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing*, 23(8), pp. 1741-1748.

Isaacs, E., Srivastava, M.R., 1989. An Introduction to Applied Geostatistics. *Oxford University Press,* New York, pp. 592.

Jin, S. and Sader, S.A., 2005. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment,* 94(3), pp. 364-372.

Kaufman, Y.J., Remer, L.A., 1994. Detection of forests using MID-IR reflectance—an application for aerosol studies. *IEEE Transactions on Geoscience and Remote Sensing* 32, pp. 672–683.

Karnieli, A., Gilad, U., Ponzet, M., Svoray, T., Mirzadinov, R. and Fedorina, O., 2008. Assessing land cover change and degradation in the Central Asian deserts using satellite image processing and geostatistical methods. *Journal of Arid Environments*, 72(11), pp. 2093-2105.

Kauth, R.J. and Thomas, G.S., 1976. The tasseled cap--a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *In LARS Symposia*, p. 159.

Lambin, E.F., Rounsevell, M.D.A. and Geist, H.J., 2000. Are agricultural land use models able to predict changes in land use intensity? *Agriculture, Ecosystems & Environment*, 82(1), pp. 321-331.

Lambin, E., Turner, B., Geist, H., Agbola, S., Angelson, A., Bruce, J., Coomes, O., Dirzo, R., Fischer, G., Folke, C. and George, P., 2001. Our emerging understanding of the causes of land use and cover change. *Global Environmental Change*, 11(4), pp .261-269.

Lasaponara, R., 2006. Estimating spectral separability of satellite derived parameters for burned areas mapping in the Calabria region by using SPOT Vegetation data. *Ecological Modelling* 196, pp. 265–270.

Madurapperuma, B., Rozario, P., Oduor, P. and Kotchman, L., 2015. Land use and land cover change detection in Pipestem Creek watershed, North Dakota. *International Journal of Geomatics and Geosciences,* 5(3), pp. 416-426.

Matheron, G., 1971. The theory of regionalized variables and its applications . *École national supérieure des mines.*

McBratney, A.B., Santos, M.M. and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117(1), pp. 3-52.

Méaille, R. and Wald, L., 1990. Using geographical information system and satellite imagery within a numerical simulation of regional urban growth. *International Journal of Geographical Information System,* 4(4), pp. 445-456.

Miller, J., Franklin, J., & Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological modelling*, 202(3), pp. 225-242.

Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D. and Van Meirvenne, M., 2009. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Science Society of America Journal*, 73(2), pp. 614-621.

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S. and Mukhopadhyay, A., 2016. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Science* (In Press).

Oduor, P.G., Kotchman, L., Nakamura, A., Jenkins, S. and Ale, G., 2012. Spatially constrained forest cover dynamics using Markovian random processes. *Forest Policy and Economics,* 20, pp.36-48.

Oliver, M.A., Webster, R., Slocum, K., 2000. Filtering SPOT imagery by kriging analysis. *International Journal of Remote Sensing* 21, pp. 735–752.

Piccini, C., Marchetti, A. and Francaviglia, R., 2014. Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. *Ecological Indicators*, 36, pp. 301-314.

Ripley, B.D. 1981. Spatial statistics. *J. Wiley, New York, NY*.

Robinson, T.P. and Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and electronics in agriculture*, 50(2), pp. 97-108.

Robertson, G. P., 1987. Geostatistics in ecology: Interpolating with known variance. *Ecology*, 68(3), pp. 744–748.

Rozario, P.F., Oduor, P., Kotchman, L. and Kangas, M., 2016. Quantifying Spatiotemporal Change in Land use and Land Cover and Assessing Water Quality: A Case Study of Missouri Watershed James Sub-Region, North Dakota. *Journal of Geographic Information System*, 8(06), pp. 663-682.

Skakun, R.S., Wulder, M.A. and Franklin, S.E., 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sensing of Environment*, *86*(4), pp. 433-443.

Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, 10(6), pp. 989-1003.

Stein, M.L., 2012. Interpolation of spatial data: some theory for kriging. *Springer Science & Business Media.*

Treitz, P.M., Howarth, P.J. and Gong, P., 1992. Application of satellite and GIS technologies for land cover and land use mapping at the rural-urban fringe: a case study. *Photogrammetric engineering and remote sensing*, 58(4), pp. 439-448.

Veldkamp, A. and Lambin, E.F., 2001. Predicting land use change. *Agriculture, ecosystems & environment*, 85(1), pp. 1-6.

Wang, L., Sousa, W.P., Gong, P., 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing* 25, pp. 5655–5668.

Watkins, T., 2005. The Tasseled Cap transformation in remote sensing. Available online at http://www.sjsu.edu/faculty/watkins/tassel.htm. Assessed on [08/02/2016].

Weier, J. and Herring, D., 2000. Measuring Vegetation (NDVI & EVI). NASA Earth Observatory.

Weng, Q., 2001. A remote sensing? GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. *International journal of remote sensing*, 22(10), pp. 1999-2014.

Westmoreland, S. and Stow, D. A. 1992. Category identification of changed land-use polygons in an integrated image processing/geographic information system. *Photogrammetric Engineering and Remote Sensing*, 58: pp. 1593–1599.

Woodcock, C.E., Strahler, A.H., Jupp, D.L.B., 1988a. The use of the variogram in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environments* 25, pp. 323–348.

Woodcock, C.E., Strahler, A.H., Jupp, D.L.B., 1988b. The use of the variogram in remote sensing: II. Real digital images. *Remote Sensing of Environments* 25, pp. 349–379.

Yeh, A.G.O. and Li, X., 1996. Urban growth management in the Pearl river delta: an integrated remote sensing and GIS approach. *ITC journal*, p. 77-85.

Yeh, A.G.O. and Li, X., 1997. An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pearl River Delta, China. *International Planning Studies*, 2(2), pp. 193-210.

Yeh, A.G.O. and Li, X., 1999. Economic development and agricultural land loss in the Pearl River Delta, China. *Habitat international*, 23(3), pp. 373-390.

Yuan, D., Elvidge, C.D., Lunetta, R.S., 1998. Survey of multispectral methods for land cover change analysis. In: Lunetta, R.S., Elvidge, C.D. (Eds.), Remote Sensing. Change Detection: Environmental Monitoring. Methods and Application. Taylor and Francis Ltd., London, pp. 21–39.

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTION

5.1. Future direction

More studies to document the expected impacts of these changes are needed. These could use more detailed socio - environmental variables to improve the understanding of the causes, locations, and trends of land use changes within such watersheds. The AHP model combined with MCMC may be used to assess larger watersheds combining more parameters. Creating a database of prediction maps of LULC for all the impaired watersheds could be a beneficial asset for policymakers.

APPENDIX A. WATER AND SOIL DATA FROM PIPESTEM CREEK

Table A1. Soil analysis data of Pipestem Creek for year 2011.

Site number	Nitrate lb./acre	Phosphate ppm	Chloride lb./acre	Sulphate ppm	Zinc ppm	Iron ppm	Manganese ppm	Copper ppm	Calcium ppm	Magnesium ppm	Potassium ppm	Sodium ppm
1	14	9	5.6	77	2.85	21	8.6	1.49	2274.3	632.8	205	127.2
2	42	14	6.8	54	1.51	11.4	7.6	0.85	2965.9	946.4	395	275.2
3	54	22	15.1	136	2.3	20.3	6.7	1.23	2646.7	946.4	270	106
4	18	19	28.5	271	0.96	12.7	27	1.36	3524.5	1008	254	242.9
5	39	15	26.9	72	1.45	18.1	10.2	1.07	2726.5	694.4	415	67.2
6	13	9	9.3	138	2.03	10.5	9.0	3	1928.5	1086.4	185	83.6
7	39	20	3.1	146	2.39	60.5	15.2	1.58	2447.2	442.4	330	174
8	22	20	29	291	7.75	71	28	1.45	2899.4	700	340	168

Table A2. Soil analysis data of Pipestem Creek for year 2016.

Site number	Nitrate lb./acre	Phosphate ppm	Chloride lb./acre	Sulphate ppm	Zinc ppm	Iron ppm	Manganese ppm	Copper ppm	Calcium ppm	Magnesium ppm	Potassium ppm	Sodium ppm
1	78	24	37	354	1.55	49.5	10.5	1.12	5033.6	1694	162	638
2	85	18	15	183	0.66	41	6.1	0.45	5541.8	775.2	336	326
3	11	17	18	114	1.83	17	4.9	0.67	2565.2	895.2	163	250
4	17	19	11	106	1.62	26.5	8.2	0.98	3242.8	484	278	322
5	31	29	52	200	1.07	9.8	5.5	0.79	5662.8	1500.4	368	572
6	31	45	10.7	101	4.30	11.5	9.7	0.68	5033.6	726	942	420
7	69	75	45	122	3.39	33.2	8.9	0.56	5488.2	894.1	878	498
8	72	77	38	231	2.45	47.1	4.4	0.33	5988.1	699	657	392

Table A3. *In situ* water sample data from sampling locations within the study area for year 2016.

Site number	1	2	3	4	5	6	7	8
Nitrate-Nitrite as N (mg/l)	0.2	0.34	0.21	0.15	0.2	0.11	0.29	0.45
Phosphate (mg/l)	2.78	1.78	2.03	2.66	0.68	1.23	3.45	2.99
Chloride (mg/l)	30.1	40.1	50.7	28.9	15	44.9	57.8	44.3
Sulphate (mg/l)	954	1170	1330	369	445	547	874	955
Calcium (mg/l)	112	115	83.5	59.4	60.5	78.4	94.2	89.4
Magnesium (mg/l)	134	148	167	53.7	79.2	122	141.3	116.5
Potassium (mg/l)	29.3	28	36	34.5	32.8	22.1	39.8	42.1
Sodium (mg/l)	215	242	369	178	211	247	201	299

Table A4. *In situ* water sample data from sampling locations within the study area for year 2016 showing the sediment load.

Site number	Total Dissolved Solids (TDS) mg/l	Total Hardness (TH) mg/l	Total Alkalinity (TA) mg/l
1	1840	830	250
2	2060	898	115
3	2410	895	200
4	916	370	232
5	572	285	154
6	444	268	168
7	610	320	112
8	485	297	103

APPENDIX B. DATA GENERATED FROM SEMGRID FOR THE MARKOV CHAIN MONTE CARLO MODEL FOR YEARS 2007 TO 2015

B1. Data generated by SemGrid for years 2007 to 2011

SemGrid 1.6.1

Date : 07-12-2016, 16:00:18

Folder: E:\Pipestem landsat images

>> dataset not declared

. import ascrec2007.txt as(ArcGis) gen(ascrec2007) type(float)

> no dataset in use

> ascrec2007 layer imported from file ascrec2007.txt

. import ascrec2011.txt as(ArcGis) gen(ascrec2011) type(float)

> ascrec2011 layer imported from file ascrec2011.txt

markest ascrec2007 ascrec2011 saving(markov)

States of variable ascrec2007

state	nobs	freq %	cumul.freq %
1	164855	5.36	5.36
2	250360	8.13	13.49
3	286689	9.31	22.80
4	224694	7.30	30.10
5	631724	20.52	50.62
Total	3078178	100.00	100.00

States of variable ascrec2011

state	nobs	freq %	cumul.freq %
1	217809	7.08	7.08
2	255709	8.31	15.38
3	303734	9.87	25.25
4	280609	9.12	34.37
5	464566	15.09	49.46
Total	3078178	100.00	100.00

Transitions from states of 2007 to 2011

from	to state	%	n.of cells	area (ha)
1	1	9.8	16163	1454.67
1	2	11.7	19286	1735.74
1	3	12.0	19730	1775.70
1	4	9.4	15505	1395.45
1	5	13.8	22701	2043.09
2	1	10.7	26899	2420.91
2	2	12.8	31944	2874.96
2	3	13.3	33315	2998.35
2	4	17.7	44438	3999.42
2	5	9.4	23535	2118.15
3	1	10.2	29155	2623.95
3	2	11.1	31810	2862.90

3	3	9.2	26289	2366.01
3	4	14.9	42715	3844.35
3	5	13.2	37905	3411.45
4	1	9.3	20957	1886.13
4	2	11.3	25353	2281.77
4	3	11.4	25541	2298.69
4	4	16.9	38070	3426.30
4	5	11.5	25753	2317.77
5	1	7.5	47417	4267.53
5	2	9.1	57342	5160.78
5	3	8.6	54554	4909.86
5	4	14.8	93284	8395.56
5	5	15.5	97990	8819.10

Transition probabilities from states of 2007 to 2011.

row->col	1	2	3	4	5
1	0.1170	0.1197	0.0941	0.1377	0.2204
2	0.1074	0.1276	0.1331	0.1775	0.0940
3	0.1017	0.1110	0.0917	0.1490	0.1322
4	0.0933	0.1128	0.1137	0.1694	0.1146
5	0.0751	0.0908	0.0864	0.1477	0.1551

> markov.tpm transition matrix saved

markest help

markest - Markov chain transition probability matrix estimation

---

Purpose: estimates transition probability for first order Markov chains

Syntax: markest varname1 [varname2] [saving(filename)]

[replace] [help]

Options:

varname# names of variables (or layer) to estimate trasition matrix.

saving declares the filename for saving transition matrix. If the filename extension is not declared, the default extension is tpm.

replace the transition matrix file, without asking.

Remarks:

markest calculates the transition matrix for a first order Markov chain.

There are two ways to prepare input data: by observation, by variables.

By observation: if only varname1 is declared, the state transition is calculated considering the variable as a time series. It is to be used for a single element, changing its state in time.

By variables: when varname2 is also declared. the transition probabilities are calculated from each observation of the first variable with respect to the corresponding observation of the second one. It is to be used for a multiple elements, changing their states between two times. An example is the change of a raster of land cells from one year to another.

Examples:

. markest RainStatus saving(transmat) transition matrix from RainStatus

. markest LandUse1990 LandUse2000 saving(changeuse)

See also: marksim

---

marksim help

marksim - Markov chain simulation

Purpose: generate random values by a first order markov chain.

Syntax:

marksim [istate(varname|#)] transp(filename) gen(varname2)

[nstep(#)] [help]

Options:

istate initial state for markov chain. This is to be a number for series or a variable name for transition between variables. If a value is declared, the time sequence is generated on a single variable. If a varname is declared, the time sequence is generated in varname2, after a number of time step indicated in the nstep option. If not declared, the initial state is the first state in the list, in alphabetic order.

transp declares the transition matrix file (*.tpm) to be used for simulation.

gen declares the varname to be generated with the simulated values.

nstep declares the number of calculation step for simulation (only with the option istate(varname)). The default value is 1.)

Remarks:

- Uses the transition matrix generated by the command markest.
- The transition probability matrix is to be saved in a file with tpm as name extension and dct format. The first column contains the initial state, the second the states after transition, and the third the probability.

Probability matrix data are arranged in one column, row by row, from left to right.

Examples:

markest RainStatus saving(transmat) transition matrix from RainStatus

marksim RstatusGen transp(transmat)

markest LandUse1990 LandUse2000 saving(changeuse)

marksim istate(LandUse2000) transp(changeuse) nstep(10)

See also: markest

B2. Data generated by SemGrid for years 2011 to 2015

SemGrid 1.6.1

Date : 07-12-2016, 16:58:31

use ascrec2007.grp

> loading file ascrec2007.grp (2837952,2)

>> grp file corrupted: n. of data different from Ncols*Nrows (0* 0<> 0)

. import ascrec2007.txt as(ArcGis) gen(ascrec2007) type(float)

> no dataset in use

> ascrec2007 layer imported from file ascrec2007.txt

. import ascrec15.txt as(ArcGis) gen(ascrec15) type(float)

> ascrec15 layer imported from file ascrec15.txt

. import ascrec2011.txt as(ArcGis) gen(ascrec2011) type(float)

> ascrec2011 layer imported from file ascrec2011.txt

. markest ascrec2011 ascrec15 saving(markov11to15)

States of variable ascrec2011

state	nobs	freq %	cumul.freq %
1	217809	7.08	7.08
2	255709	8.31	15.38
3	303734	9.87	25.25
4	280609	9.12	34.37
5	464566	15.09	49.46
Total	3078178	100.00	100.00

States of variable ascrec15

state	nobs	freq %	cumul.freq %
1	113748	3.70	3.70
2	223630	7.27	10.96
3	111677	3.63	14.59
4	254117	8.26	22.84
5	402721	13.08	35.93
Total	3078178	100.00	100.00

Transitions from states of 2011 to 2015

from	to state	%	n.of cells	area (ha)
1	1	3.8	8362	752.58
1	2	3.7	8142	732.78
1	3	4.0	8755	787.95
1	4	6.9	14965	1346.85
1	5	15.7	34126	3071.34
2	1	14.5	37080	3337.20
2	2	26.2	67118	6040.62
2	3	20.9	53521	4816.89
2	4	10.3	26314	2368.26
2	5	2.9	7358	662.22
3	1	13.7	41762	3758.58
3	2	26.2	79463	7151.67

3	3	21.6	65581	5902.29
3	4	9.5	28829	2594.61
3	5	2.3	7004	630.36

4	1	12.5	35160	3164.40
4	2	26.6	74765	6728.85
4	3	22.3	62444	5619.96
4	4	7.9	22138	1992.42
4	5	1.5	4323	389.07

5	1	12.6	58536	5268.24
5	2	27.6	128359	11552.31
5	3	23.6	109422	9847.98
5	4	7.4	34207	3078.63
5	5	1.0	4876	438.84

Transition probabilities from states of 2011 to 2015

---

row->col	1	2	3	4	5
1	0.0384	0.0374	0.0402	0.1567	0.2855
2	0.0475	0.0883	0.0419	0.0738	0.0737
3	0.0437	0.1004	0.0413	0.0816	0.0969
4	0.0393	0.1099	0.0371	0.1052	0.0784
5	0.0384	0.1025	0.0375	0.0997	0.0353

> markov11to15.tpm transition matrix saved.

markest help

markest - Markov chain transition probability matrix estimation

---

Purpose: estimates transition probability for first order Markov chains.

B3. Data generated by SemGrid for years 2007 to 2015

SemGrid 1.6.1

Date : 07-12-2016, 16:50:02

. import ascrec15.txt as(ArcGis) gen(ascrec15) type(float)

> no dataset in use

> ascrec15 layer imported from file ascrec15.txt

. import ascrec2007.txt as(ArcGis) gen(ascrec2007) type(float)

> ascrec2007 layer imported from file ascrec2007.txt

. markest ascrec2007 ascrec15 saving(markov07to15)

States of variable ascrec2007

state	nobs	freq %	cumul.freq %
1	164855	5.36	5.36
2	250360	8.13	13.49
3	286689	9.31	22.80
4	224694	7.30	30.10
5	631724	20.52	50.62
Total	3078178	100.00	100.00

States of variable ascrec15

state	nobs	freq %	cumul.freq %
1	113748	3.70	3.70
2	223630	7.27	10.96
3	111677	3.63	14.59

4	254117	8.26	22.84
5	402721	13.08	35.93
Total	3078178	100.00	100.00

Transitions from states of 2007 to 2015

from	to state	%	n.of cells	area (ha)
1	1	5.5	9015	811.35
1	2	4.9	8008	720.72
1	3	4.5	7496	674.64
1	4	6.7	10988	988.92
1	5	16.2	26642	2397.78
2	1	22.2	55639	5007.51
2	2	3.5	8759	788.31
2	3	11.4	28646	2578.14
2	4	11.8	29497	2654.73
2	5	18.8	47190	4247.10
3	1	4.9	13946	1255.14
3	2	8.4	24113	2170.17
3	3	15.2	43544	3918.96
3	4	26.6	76145	6853.05
3	5	21.0	60232	5420.88
4	1	4.2	9372	843.48

4	2	11.0	24651	2218.59
4	3	13.1	29534	2658.06
4	4	23.9	53764	4838.76
4	5	19.2	43179	3886.11
5	1	3.6	22613	2035.17
5	2	7.5	47241	4251.69
5	3	13.7	86292	7766.28
5	4	33.1	209058	18815.22
5	5	29.0	183108	16479.72

Transition probabilities from states of 2007 to 2015

row->col	1	2	3	4	5
1	0.1616	0.2592	0.2021	0.1341	0.0277
2	0.2222	0.1178	0.1885	0.1475	0.0224
3	0.0584	0.0577	0.0841	0.1519	0.2656
4	0.0451	0.1417	0.0417	0.1314	0.2393
5	0.0309	0.0360	0.0358	0.3309	0.2899

> markov07to15.tpm transition matrix saved

markest help

markest - Markov chain transition probability matrix estimation

---

Purpose: estimates transition probability for first order Markov chains

Syntax: markest varname1 [varname2] [saving(filename)]

[replace] [help]

APPENDIX C. IMAGE PROCESSING AND GEOSTATISTICAL ANALYSIS

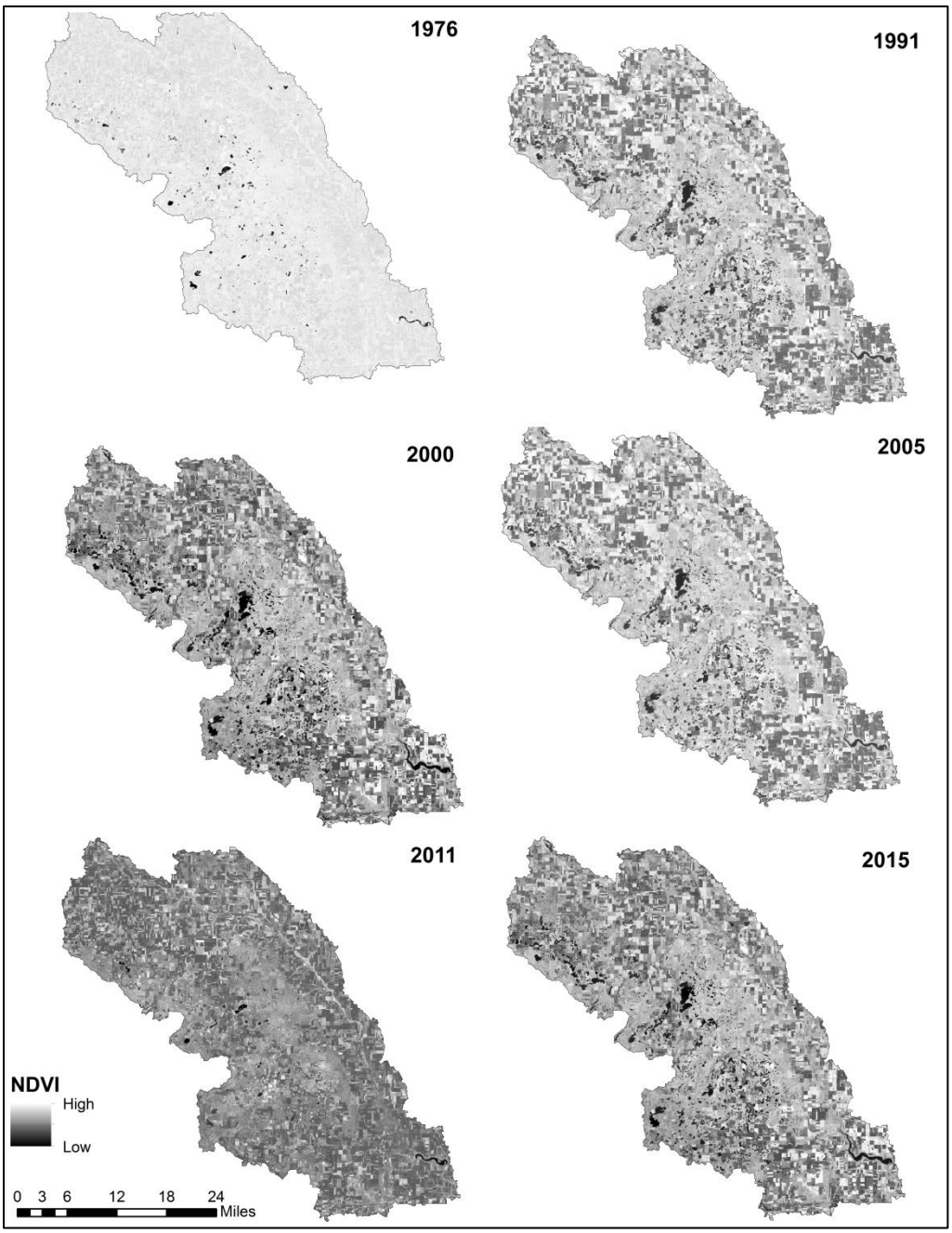

Figure C1. NDVI datasets for years 1976 to 2015 for Pipestem Creek watershed.

C1. Geostatistical analysis methods

i) Gaussian Geostatistical Simulation model (GGS)

This model first creates a grid of randomly assigned values from a standard distribution and then applies the covariance model to the input raster (Dietrich and Newsam, 1993; ArcGIS Help 10.4.1).

Assumptions (Dietrich and Newsam, 1993; ArcGIS Help 10.4.1)

- GGS assumes that the data is normally distributed.
- Clustered data needs to be declustered so that the input histogram accurately represents the sampled population.
- Gaussian process realizations are continuous.

Advantages

- GCS replicates the mean, variance, and semivariogram of the data, on average (that is, averaged over many realizations) (Chiles and Delfiner, 1999).
- Increased use of GGS follows a trend in geostatistical practice that emphasizes the characterization of uncertainty for decision and risk analysis, rather than producing the best unbiased prediction for each unsampled location (as is done with Kriging), which is more suited to showing global trends in the data (Deutsch and Journel 1998; Goovaerts 1997).

Disadvantage

- GCS avoids discontinuities in the simulated surfaces due to changes in the local neighborhood used in Kriging (Aldworth, 1998; Gribov and Krivoruchko, 2004).

Table C1. Semivariogram parameters for Gaussian Geostatistical Simulation models fitting the Tasseled Cap Greenness Index (TCGI) products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m).

Gaussian model	1976	1991	2000	2005	2011	2015
Sill	0.0023	0.0018	0.0021	0.0021	0.0024	0.0017
Range (m)	765.5	762.5	797.9	1030.7	1079.1	1101
Root-mean-square (RMS)	0.051	0.088	0.067	0.047	0.048	0.043
Average standard error	0.052	0.066	0.036	0.070	0.073	0.084
Mean standardized	0.044	0.043	0.041	0.057	0.047	0.045
Root-mean-square-standardized	1.690	1.721	0.457	0.513	0.72	0.451

Discussion

For an ideal model, the Mean prediction error should be near 0 (this investigates bias), Root Mean Square (RMS) prediction error should be small, average standard error should be close to RMS error, Mean-standardized prediction error should be near 0 and RMS standardized prediction error should be near 1, indicating that the estimated prediction uncertainty is consistent (Stein, 2012). In Table C1, the RMS ranged from 0.051 to 0.088, average standard error ranged from 0.036 to 0.084, which is not very close to the RMS values. Mean-standardized prediction error values ranged from 0.044 to 0.057. RMS standardized prediction error values were greater than 1 for years 1976 to 1991, ranging from 1.690 to 1.721. This model did not prove the ability to reproduce the observed values accurately.

ii) Spherical and Circular Geostatistical Simulation model

These models shows a progressive decrease of spatial autocorrelation (equivalently, an increase of semivariance) until some distance, beyond which autocorrelation is zero. The spherical model is one of the most commonly used models (Matheron, 1963).

Assumption

- Both models assume that the data is normally distributed (Matheron, 1963; ArcGIS Help 10.4.1).

Advantage

- These models are useful for phenomena within an enclosed boundary or perimeter (Matheron, 1963).

Disadvantage

- Datasets with irregular enclosures are not suited for these models since the spatial autocorrelation decreases with after a certain distance (Matheron, 1963).

Table C2. Semivariogram parameters for Spherical Geostatistical Simulation models fitting the Tasseled Cap Greenness Index (TCGI) products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m).

Spherical model	**1976**	**1991**	**2000**	**2005**	**2011**	**2015**
Sill	0.0015	0.0012	0.0007	0.0011	0.0008	0.0010
Range (m)	700.5	766.5	799.9	1040.7	1099.1	1101
Root-mean-square (RMS)	0.031	0.008	0.017	0.017	0.028	0.023
Average standard error	0.002	0.056	0.016	0.060	0.063	0.054
Mean standardized	0.021	0.003	0.001	0.027	0.007	0.015
Root-mean-square-standardized	1.501	1.281	1.057	1.213	1.172	0.951

Table C3. Semivariogram parameters for Circular Geostatistical Simulation models fitting the Tasseled Cap Greenness Index (TCGI) products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m).

Circular model	1976	1991	2000	2005	2011	2015
Sill	0.0015	0.0012	0.0007	0.0011	0.0008	0.0010
Range (m)	700.5	762.5	797.9	1100.7	1079.1	1200.6
Root-mean-square (RMS)	0.031	0.038	0.017	0.017	0.028	0.023
Average standard error	0.002	0.006	0.016	0.060	0.013	0.054
Mean standardized	0.001	0.033	0.021	0.017	0.007	0.015
Root-mean-square-standardized	1.591	1.681	1.657	1.613	1.672	1.651

Discussion

In Table C2, the RMS ranged from 0.008 to 0.031, average standard error ranged from 0.002 to 0.056, which is not very close to the RMS values. Mean-standardized prediction error values ranged from 0.001 to 0.021. RMS standardized prediction error values were greater than 1 for years 1976 to 2011, ranging from 1.057 to 1.501. This model did not prove the ability to reproduce the observed values accurately.

In Table C3, the RMS ranged from 0.017 to 0.038, average standard error ranged from 0.002 to 0.054, which is not very close to the RMS values. Mean-standardized prediction error values ranged from 0.001 to 0.033. RMS standardized prediction error values were greater than 1 for years 1976 to 2015, ranging from 1.591 to 1.681. This model was not found to be a best fit.

iii) J-Bessel and K-Bessel Geostatistical Simulation models

J-Bessel and K-Bessel Geostatistical Simulation models are based on Bessel's equation. This equation arises when finding separable solutions to Laplace's equation and the Helmholtz

equation in cylindrical or spherical coordinates (Curran, 1988). Bessel functions are therefore especially important for many problems of wave propagation and static potentials (Curran, 1988).

Assumption

- Both models assume that the data is normally distributed (Curran, 1988; Matheron, 1963; ArcGIS Help 10.4.1).
- They also assume singularity at the origin (Curran, 1988).

Table C4. Semivariogram parameters for J-Bessel Geostatistical Simulation models fitting the Tasseled Cap Greenness Index (TCGI) products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m).

J-Bessel model	1976	1991	2000	2005	2011	2015
Sill	0.0009	0.0006	0.0008	0.0008	0.0003	0.0004
Range (m)	842.5	762.5	797.9	1030.7	1079.1	1100
Root-mean-square (RMS)	0.001	0.008	0.007	0.007	0.008	0.013
Average standard error	0.012	0.066	0.056	0.070	0.073	0.024
Mean standardized	0.011	0.053	0.051	0.007	0.007	0.005
Root-mean-square-standardized	1.091	1.081	1.007	1.613	0.772	0.858

Table C5. Semivariogram parameters for K-Bessel Geostatistical Simulation models fitting the Tasseled Cap Greenness Index (TCGI) products for the Pipestem Creek watershed (Nugget = 0; Lag = 1,000m).

K-Bessel model	1976	1991	2000	2005	2011	2015
Sill	0.0005	0.0002	0.0007	0.0001	0.000	0.000
Range (m)	893.5	762.5	797.9	1030.7	1079.1	1002
Root-mean-square (RMS)	0.001	0.018	0.027	0.117	0.128	0.123
Average standard error	0.042	0.046	0.036	0.010	0.013	0.014
Mean standardized	0.021	0.033	0.011	0.027	0.007	0.005
Root-mean-square-standardized	1.001	1.001	1.105	1.214	1.672	1.651

Discussion

In Table C4, the RMS ranged from 0.008 to 0.031, average standard error ranged from 0.012 to 0.073, which is not very close to the RMS values. Mean-standardized prediction error values ranged from 0.001 to 0.013. RMS standardized prediction error values were greater than 1 for some datasets, ranging from 0.772 to 1.613. This model did not prove the ability to reproduce the observed values accurately.

In Table C5, the RMS ranged from 0.001 to 0.128, average standard error ranged from 0.010 to 0.042, which is not very close to the RMS values. Mean-standardized prediction error values ranged from 0.005 to 0.033. RMS standardized prediction error values were greater than 1 for years 1976 to 2015, ranging from 1.001 to 1.672. This model was not found to be a good fit.

Thus, the exponential model was chosen. This model is usually applied when spatial autocorrelation decreases exponentially with increasing distance, disappearing completely only at an infinite distance (Curran, 1988).

C2. References cited

Aldworth, J. 1998. Spatial Prediction, Spatial Sampling, and Measurement Error. Ph.D. Thesis, *Iowa State University*.

ArcGIS Help 10.4.1. Available online at http://resources.arcgis.com/en/help/main/10.1/index.html#/Key_concepts_of_geostatistical_simulation/. Accessed on 03/15/2017.

Chiles, J. P., and P. Delfiner. 1999. Geostatistics: Modeling Spatial Uncertainty. *New York: John Wiley & Sons*, pp. 449–471.

Curran, P.J., 1988. The semivariogram in remote sensing: an introduction. *Remote sensing of Environment*, *24*(3), pp. 493-507.

Deutsch, C.V., and A. G. Journel. 1998. GSLIB Geostatistical Software Library and User's Guide. 2nd Ed. *Oxford University Press, New York*, pp. 119–122.

Dietrich, C.R. and Newsam, G.N., 1993. A fast and exact method for multidimensional Gaussian stochastic simulations. *Water Resources Research*, *29*(8), pp. 2861-2869.

Goovaerts, P. 1997. Geostatistics for Natural Resource Evaluation. *Oxford University Press, New York*, pp. 369–376.

Gribov, A., and K. Krivoruchko. 2004.Geostatistical Mapping with Continuous Moving Neighborhood. *Mathematical Geology* 36 (2), pp. 267–281.

Matheron, G., 1963. Principles of geostatistics. *Economic geology*, *58*(8), pp.1246-1266.

Stein, M.L., 2012. Interpolation of spatial data: some theory for kriging. *Springer Science & Business Media.*

APPENDIX D PUBLICATIONS ARISING FROM THIS RESEARCH

Rozario P.F, Oduor, P.G, Kotchman, L. and Kangas, M. 2016. Quantifying Spatiotemporal Change in Land use and Land Cover and Assessing Water Quality: A Case Study of Missouri Watershed James Sub-Region, North Dakota. *Journal of Geographic Information System*, 8, pp. 663-682.

Rozario, P.F., Oduor, P., Kotchman, L. and Kangas, M., 2017. Transition Modeling of Land-Use Dynamics in the Pipestem Creek, North Dakota, USA. *Journal of Geoscience and Environment Protection*, *5*(03), p.182.

Rozario P.F, Oduor, P.G, Kotchman, L. and Kangas, M. 2017. Uncertainty Analysis of Spatial Autocorrelation of Land-use and Land-cover Data within Pipestem Creek in North Dakota. *Journal of Geoscience and Environment Protection* (In press).

Madurapperuma B.D., Rozario P.F, Oduor, P.G. & Kotchman, L.A. 2015.Land use and Land cover Change Detection in Pipestem Creek watershed, North Dakota, U.S.A. *International Journal of Geomatics and Geosciences* Volume 5 No.3 2015.

Oduor, P.G., Kangas, M., Kotchman, L., Claeys, T., Rozario, P.F., Anar, M.J., Wamono, A.W. and Madurapperuma, B.D. 2011. A Map Odyssey on 2011 Record Flooding in North Dakota: Ramifications on Riparian Forests. *NDFS online publications*.

Rozario P.F, Oduor, P.G, Kangas, M, Kotchman, L.A. 2017. Spatial dependence of Land use and Land cover Data within Pipestem Creek watershed in North Dakota. *Proceedings of AAG Annual Meeting 2017, Boston*.

Rozario P.F, Oduor, P.G., Kangas, M, Kotchman, L.A. 2016. Quantifying Spatiotemporal Change in Land use and Land cover and assessing water quality: A Case study of Missouri Watershed James Sub-region, North Dakota, *World Congress of GIS & Remote*

Sensing 2016, New Orleans, Louisiana.

Rozario P.F., Oduor, P.G., Kotchman, L. 2011.Assessing Forest Cover Dynamics for Missouri Watershed Region: Ramifications on Watershed Water Quality. *NDSU Graduate Research and Arts Forum session*, North Dakota State University.

Rozario P.F, Oduor, P.G., Kotchman, L.A. 2011. Change Detection in Land use and Land cover Using Remote Sensing - A Case Study of Missouri Watershed James Sub-region, North Dakota, *ND-SD 2011 Joint EPSCoR Conference*, North Dakota State University.

Rozario P.F, Oduor, P.G., Kotchman, L.A. 2011. Impact of Forest Cover Dynamics on water quality of Missouri Watershed Region in North Dakota, *14th World Lake Conference 2011*, Austin, Texas.

Rozario, P.F., Anar, M.J., Wamono, A.W. and Madurapperuma, B.D. 2011. An assessment of inundated riparian forests along Missouri River at Bismarck - Mandan Wildland Urban Interface in North Dakota. *Proceedings of the North Dakota GIS Users Conference 2011*, Grand Forks, ND.

Oduor, P.G., Kangas, M., Kotchman, L., Claeys, T., Rozario, P.F., Anar, M.J., Wamono, A.W. and Madurapperuma, B.D. 2011. A Map Odyssey on 2011 Record Flooding in North Dakota: Ramifications on Riparian Forests. *Proceedings of the North Dakota GIS Users Conference 2011*, Grand Forks, ND.