12-2013

Mapping Spatial Thematic Accuracy Using Indicator Kriging

Maria I. Martinez
University of Tennessee - Knoxville, mmarti85@utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Maria I. Martinez entitled "Mapping Spatial Thematic Accuracy Using Indicator Kriging." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geography.

Liem T. Tran, Major Professor

We have read this thesis and recommend its acceptance:

Nicholas Nagle, Budhendra L. Bhaduri

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Mapping Spatial Thematic Accuracy Using Indicator Kriging

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Maria I. Martinez
December 2013

DEDICATION

I dedicate this thesis to my son Adrian whose smile and love lights up my heart. Also to my family, who always supports me and encourages me to work hard and be my best.

I also dedicate this work to my many friends who have supported me through this process and have cheered me along – and most importantly have remained my friends after all! I want to give special thanks to Connie and Margarita for all the advice you have given me.

ACKNOWLEDGEMENTS

I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Liem Tran for serving my committee chair and for his guidance, his support and most of all patience throughout the entire process. Also I would like to Dr. Nicholas Nagle and Dr. Budhu Bhaduri for agreeing to serve on my committee.

ABSTRACT

Thematic maps derived from remote sensing imagery are increasingly being used in environmental and ecological modeling. Spatial information in these maps however is not free of error. Different methodologies such as error matrices are used to assess the accuracy of the spatial information. However, most of the methods commonly used for describing the accuracy assessment of thematic data fail to describe spatial differences of the accuracy across an area of interest. This thesis describes the use of indicator kriging as a geostatistical method for mapping the spatial accuracy of thematic maps. The method is illustrated by constructing accuracy maps for the forest land-cover classes in the 2001 National Land Cover Dataset (NLCD) extent covering the conterminous United States. Independent reference data collected for the accuracy assessment of the 2001 NCLD was used. This thesis also describes the use of indicator cokriging for improving the thematic accuracy of the forest land-cover classes by adding information from other land-cover classes as additional variables. Finally, probability surfaces resulted from indicator kriging and indicator cokriging will be used to generate alternate realizations of the forest land-cover class through stochastic simulation. Such realizations could serve as input parameters to spatially explicit models. Result show how thematic accuracy varies across regions and it outlines differences between land-cover estimates by NLCD and those created through indicator kriging.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction And General Information

1.1 Problem Statement

The production of thematic land-cover maps is one of the most common applications of remote sensing (Foody 2002). These land-cover maps support a large range of research efforts studying characteristics of the earth's surface, especially land use planning and environmental studies (Yang et al. 2001). These maps are also used as input parameters to model in spatially explicit ecological and environmental models. No map created from remote sensing can be completely accurate (Steele, Winne and Redmond 1998). The development of these maps isn't completely accurate and errors occur between actual ground-based information and image-derived data. Errors can be introduced in many ways: approximating the extent of vegetation cover as a crisp boundary, scale resolution from reality to a map, difficulties in differentiating between land-cover classes or badly defined classification schemes among other reasons (Steele et al. 1998, Zhang and Goodchild 2002). Understanding the thematic map accuracy is important to be able to use the information correctly when thematic maps are used for decision-making.

Thematic mapping accuracy is not spatially uniform and varies across landscapes as a result of several factors such as sensor resolution, spectral overlap, preprocessing algorithms and classification procedures (Campbell 1983, Tran et al. 2005). Accuracy assessment is becoming an important topic with the increased use of remotely sensed imagery and mapping, environmental modeling and other applications (Carlotto 2009). Many methods have been used in the assessment of the thematic map accuracy (Foody 2004). Of the most common approaches to describe thematic map accuracy is using a confusion or error matrix which identifies differences between the land-cover map and a reference data set. A confusion matrix provides an overall measure of classification accuracy and it provides summary of omission and commission error, and the Kappa coefficient. Errors of omission (producer's accuracy) refer to the proportion of cases where a land-cover class is correctly classified. Errors of commission (user's accuracy) identifies the frequency a land-cover class is correctly classified. These two measures are calculated as the proportion of correctly classified locations (McGwire and Fisher 2001). However, error matrices do not provide information about the spatial pattern of the distribution and variation of mapping errors as these matrices do not include information about what real areas on the ground are most likely misclassified. Also, the confusion matrix assumes the relationship between different land-cover classes in the confusion matrix do not vary across the region (Hession, Shortridge and Torbick 2006). An aspect of models and analysis using thematic data is the use of spatially explicit inputs and outputs. Therefore, it is important to document the

uncertainty of input data in a way where it includes the distribution of error across the surface map (Kyriakidis 2001).

The role of uncertainty in spatial decision support has been the focus of many studies. Much of the research has been focused on the propagation of uncertainty in spatial data and how it can affect the result of a model looking at land use allocation (Aerts, Goodchild and Heuvelink 2003). Hession et al. (2006) developed a comparative matrix summarizing different model requirements and characteristics of methods used to characterize the distribution of categorical data such as land-cover by alternative generating realizations.

1.2. Research Question and Hypothesis

The objectives of this thesis are to:

- To assess the spatial accuracy of thematic maps using indicator kriging applied to independent reference data. The method is illustrated by constructing accuracy maps for the forest land-cover classes in the 2001 National Land Cover Dataset (NLCD) extent covering the conterminous United States.
- Improving the thematic accuracy of the forest land-cover classes by adding other land-cover classes as additional variables through indicator cokriging.
- To generate alternate realizations from of the forest land-cover class probability surfaces from indicator kriging and indicator cokriging.

The NLCD, developed by the Multi-Resolution Land Characteristics (MLRC) Consortium, is a primary source of land-cover data in the United States. A nationwide land-cover accuracy assessment was conducted for the 2001 NLCD which included a collection of ground truth data used to develop a confusion matrix that outlines the accuracy of the dataset. This thesis will use these ground truth data to answer the following research questions:

- Can the spatial accuracy of the forest land-cover classes in the 2001 NLCD be mapped using indicator kriging?

- Does including spatial accuracy information from other land-cover classes improve the accuracy of the forest land-cover classes?

- What realizations of land-cover does Monte Carlo simulation provide using information from indicator kriging?

In this thesis, indicator kriging is used as the approach to map the spatial accuracy of forest land-cover classes in the thematic land-cover maps. Level II classification for forest includes values 41, 42 or 43. This method will explore whether the thematic map accuracy is spatially uniform or changes across regions and if different land-cover types show different spatial patterns of accuracy. Cokriging is used as a multi-variable approach to determine if the accuracy of the forested land-cover class can be improved. This method will look at the effects of adding accuracy information from other spatially correlated data. Lastly, Monte Carlo simulation is used to generate alternative realizations of the forest land-cover classes creating a sample distribution of the uncertainty on the data.

Chapter 2 Literature Review

2.1 Geographic Information and Uncertainty

"Geographic information can be defined as information about features and phenomena located on or near the surface of the Earth" (Goodchild et al. 1999). Geographic information deals with the need to solve geographical problems relating to many fields of studies such as environmental problems, biological conservation, or understanding demographics. The gathering of geographic information has been taken on by many government organizations such as the United States Geological Survey (USGS) or the United States Census Bureau and many private organizations (Zhang and Goodchild 2002).

The process of gathering geographic information was traditionally done by field scientists, engineers and other professionals by recording information into paper maps. The conversion of these paper maps into digital format has been an ongoing process for many organizations. Advances in geographic information technologies such as remote sensing have transformed the way data is gathered and processed (Zhang and Goodchild 2002). Remote sensing data from satellites is being used in many applications such as land cover classification (Campbell 1996, Jensen 1979, Fuller, Wyatt and Barr 1998). Geographic Information systems (GIS) along with remote sensing are common technologies used for managing and handling large amounts of data relevant to many disciplines such as geography, agriculture, hydrology, among others (Burrough and McDonnell 1998). The common assumption is that output geographic information created from any of these processes is error-free. Errors are defined as a deviation of a measurement from its true value (Rabinovich, 2005). In reality, uncertainty exists throughout the entire process from data acquisition, geoprocessing to the use of the use of the data (Zhang and Goodchild 2002). Uncertainty is the degree to with a measured value is estimated to vary from its true value. Uncertainty can have many sources such as the accuracy of an instrument, measurement error or the use of data at different scales in the geoprocessing procedure (Sommer and Wade, 2006). Sources of uncertainty will be further described in the following section.

The overall process of data abstraction is selective as it involves the approximation of real features in order to produce a conceptual model of the real world. The complexity of geographic features cannot be reproduced with perfect accuracy. The discrepancies between the modeled and real world contribute inaccuracy and uncertainty which can turn into imperfect decision-making (Goodchild 1989).

2.1.1 Sources of Uncertainty

Uncertainty can be found at different steps through the entire process from data acquisition to the use of data through geoprocessing. Data acquisition as a first step to obtain the raw components that make up the real world is the subject to uncertainty. Data acquisition has several levels of error depending on the skill of the data analysis and the instruments used. Though there have been many technological advances in such as Global Positioning system and digital image processing which have improved the quality of spatial data, it is not possible to eliminate all errors during data acquisition (Zhang and Goodchild 2002).

Another form of uncertainty occurs during geoprocessing errors. Many of the data acquired are catered to specific research purposes and there are often many sources of data representing the same phenomena in different manners. Also, it is common to find different GIS software systems using different standards and tools. The process to transfer data between one system to another is not straightforward and it can be time consuming (Včkovski 1998). This is where errors of geoprocessing are found. An example of such errors is the conversion of data types from vector to raster structure.

As more sources of data become available, it is likely that various data at different accuracy levels will be involved in a decision-making process. Methods for combining data sets need to take these accuracy differences into account so data combination can produce the best possible output that is scientifically sound. Hutchinson (1982) investigated such methods for improving data acquired through remote sensing by incorporating additional data. Spatial data however is complex and requires special processing (Atkinson and Tate 2000) and often data integration procedures do not take into account different levels of accuracy and detail on the data and assume scales and precise data during geoprocessing tools.

Uncertainty is not only a property of the data itself but also a measure of the difference between the actual content of the data and the content users would have created through perfect observation of reality. There is no check for potential misuse of a spatial data set. Users typically rely on published data by government organization. These users do not normally have access to the primary source of data from which these final products were derived from and must take the data as it is offered (Zhang and Goodchild 2002).

2.1.2 Measuring Uncertainty

According to the University Consortium for Geographic Information Science (1996) the goals of research on uncertainty in geographic data are to investigate how uncertainties are created and propagated in the GIS process and

what effects these uncertainties might have on the results. Obtaining information on uncertainties during geoprocessing and tracking the propagation of these uncertainties would allow users to assess the accuracy obtainable of a specific map analysis (Goodchild and Muller 1991). To accomplish this goal, there is a need to effectively identify sources of uncertainty.

As mentioned earlier, there are different sources of uncertainty which can affect geographic information and its analysis (Chrisman 1991). Uncertainty may be associated with the process of generalizing the real world or uncertainty may occur to errors in the measurement of positions and attributes. Inappropriate processing of source data may also be prone to errors if data are not used correctly. Aronoff (1989) provides a summary of common sources of error found in GIS applications showing how errors occur from the data source, through data manipulation to data output and use of GIS results.

Once uncertainty has been identified, the next step is to measure it. There have been many techniques for uncertainty assessment developed through land surveying which are useful in GIS mapping. Classification accuracy through remote sensing is usually evaluated by measuring percentage of correctly classified pixels as a method to measure nominal scale data (Cohen 1960).

Following the identification and evaluation of uncertainty, researches must deal with uncertainty propagation when different data types are integrated. An analysis of uncertainty propagation takes information from the uncertainty in the input data, knowledge about the process and predicts the uncertainty associated with an output. A general approach to uncertainty modeling is to use a technique with equal-probably realizations of spatial data which allows for evaluation the inherent distribution of the data and examining the range of possible outputs (Zhang and Goodchild 2002). Goodchild (1989) defined some of the basic concepts of uncertainty modeling.

Having an understanding on how spatial features are modeled is important when looking at the uncertainty of data. There are two perspectives on how to model geographic data: the object-model and the field-model. Field-based models are more common in physical geography and environmental applications while object-based models are better for cartography and facilities management. The choice between either model depends on the underlying phenomenon being studied and whether the geographic feature is considered a set of single-valued functions defined everywhere or a collection of discrete objects (Goodchild 1989). Objects represent spatial data in the form of discrete objects and their associated attributes. Fields represent phenomena that are continuous over space and time. The choice will have an effect on the possibilities for uncertainty modeling (Goodchild 1993).

2.1.3 Statistical Methods for Quantifying Uncertainty

Statistical estimation and inference usually assumes data homogeneity and independence. This is problematic when dealing with spatial data which is characterized by data dependency and heterogeneity. A characteristic of spatial data is spatial dependence since values at individual locations tend to be correlated. Using Tobler's law of geography, things closer to one another tend to be more similar than those further apart (Tobler 1970). Spatial dependence is fundamental to geographical distributions. Spatial dependency allows the inference about a variable and its spatial variation from measured data.Moran (1948) evaluated the dependency on the distribution of geographic data and derived an assessment of whether the presence of some characteristics in a location makes their presence in neighboring locations more likely. Campbell (1981) also evaluated the continuity of values in adjacent pixels looking at rural land cover classification using Landsat MSS data.

Some well-established statistical methods and tools for error analysis may be used for handling geographical uncertainty (Hession et al. 2006). Gaussian models provide a starting point for describing and modeling quantitative data sets (Bierkens and Burrough 1993). Gaussian models are commonly used in large data samples that follow a normal distribution. The normal distribution of a random variable is specified as a probability density function. Based on that function, it is possible to evaluate the errors within a certain interval ((Zhang and Goodchild 2002). Error analysis on variables measured on continuous scale such as elevation or precipitation can be evaluated through standard statistical metrics such as root-mean-square error. Categorical variables however take discrete values and error analysis must be performed using other methods. The results of accuracy assessment in categorical data tend to be reported by identifying agreement/disagreement between classified and referenced categories. Accuracy may be evaluated by measures such as the percentage of correctly classified objects, and kappa coefficient. This topic was reviewed by Congalton (1991), Goodchild (1994) and Congalton and Green (2008).

The remaining of this chapter will focus on describing methods to evaluate uncertainty in categorical data, which is the focus of this thesis.

2.2 Uncertainty in Categorical Data

Many real-world phenomena are modeled as categorical information. Categorical variables take discrete values as the outcomes of classification processes. A commonly known example of categorical data is land-cover which is differentiated into nominal categories such as residential, forest or shrub land. Such information is usually derived from remote sensing information. Thematic mapping for remote sensing data is based on image classification where classes

are grouped together based on their spectral similarities (Foody 2002).Uncertainty in remote sensing-derived information can occur for several reasons.

Misclassification can be due to the generalization of data into a pixel. This is due to the data model used to represent the complexities of the real-world. Categorical variables are modeled as fields. There are four field models for describing phenomena using categorical variables. These are irregular points, regular points, polygons and grid. Two of these are used to represent areas: the grid model, where values are represented into rectangular cells; and the polygon model, where information is divided into irregular polygons. In both models, the spatial variation within a cell or a polygon is ignored, and the cell or polygon are assigned the value of the dominant class. Uncertainty is therefore found in categorical variable when spatial variation within cells and polygons are generalized during the mapping process. Uncertainty can also occurred from the resolution used for data acquisition, classification systems not properly identified, or the difficulty of differentiating between different categories such as land-cover types or soil types. Another factor is the subjective interpretation and human judgment used during the classification process (Zhang & Goodchild, 2002).

Categorical data cannot be interpolated using traditional geostatistical methods because those techniques are focused mainly on continuous data. Instead, to account for uncertainty in categorical data, probabilistic approaches can used to model each field value as a random variable. The cumulative probability distribution is used to characterize a location in the field. This approach is the basis of geostatistical analysis and exploration of geographical data (Journel 1986). The result of an accuracy assessment is usually evaluated by measures such as the percentage of correctly classified objects. In data created through remote sensing, such as land-cover classification, ground truth data are often used to identify the percentage of pixels were correctly classified or misclassified. The general pattern of uncertainty will be spatially dependent with locations near ground truth points having smaller inaccuracy in classification that those further away.

2.2.1 Uncertainty in Thematic Land Cover Maps

In statistics, accuracy includes bias and precision (Campbell 1996). In thematic mapping, the term ‘accuracy’ is used to express the degree of correctness in the classification. As with other categorical data, classification accuracy is reported as the degree to which the derived image classification agrees with reality – represented by ground truth information. Classification error therefore occurs when the image classification does not match the ‘truth’ (Foody 2002).

Many methods of accuracy assessment have been discussed in literature (Koukoulas and Blackburn 2001, Kyriakidis 2001). The most widely method of assessment of classification accuracy for categorical data however is the construction of an error matrix. The matrix consists on the cross-tabulation of the mapped classes against a set of observed reference data (Campbell 1996). One of the most common measurements of accuracy provided by the matrix is a comparison of percentages the recorded class agrees with the reference source. This is done through two concepts: user's and producer's accuracy. User's accuracy is the probability that a location labeled as category k actually belongs the category k in the reference source. This is measured as commission error. Producer's accuracy is the probability that a location known to belong to category k is correctly labeled. This is a measured of omission error. The overall accuracy is calculated by dividing the sum of the diagonal entries in the error matrix by the total of all matrix elements. The overall classification accuracy is reported as the percent correctly classified (Foody 2002) .

One problem with the use of this approach to estimate accuracy is that it ignores those cases where a pixel may have been allocated to the correct land-cover class by chance. The Kappa Coefficient is included in the calculation to account for the effects of chance agreement (Foody 2002).

In the building of a confusion matrix there is an assumption that the agreement between classes is stationary; that is class agreement does not vary across a region. Often there is a pattern to the spatial distribution of thematic errors which are derived from the sensor used, or ground conditions. Examples are errors found at the boundaries between different land-cover classes (Congalton 1988b). Although the confusion matrix provides useful accuracy information, it does not provide information on the spatial distribution of error (Steele et al. 1998). Users of thematic maps derived from remote sensing data may benefit from a spatial representation of classification accuracy. Various approaches have been used to provide this information.

An earlier study by Fisher (1994) proposed a visual method of displaying classification errors via animation where accuracy measures such as overall accuracy, user and producer accuracy and classification accuracy are embedded in the display of the classified image. Moisen, Cutler and Edwards (1996) developed a generalized linear model approach to study misclassification errors in relationship to variables such as distance to road, slope and land-cover heterogeneity. Others such as Campbell (1987) and Townsend et al. (2000) focused on evaluating the misclassification of pixels along boundaries of homogeneous patches. More recent methods include extrapolation from training dataset (Steele et al. 1998) and the used of geostatistical approach to model the variation in accuracy over a mapped region (Kyriakidis 2001). It is the use of geostatistical approaches and its use within GIS that are providing a framework for integrating data that provides location-dependent models of spatial

uncertainty (Kyriakidis and Dungan 2001). Geostatistical studies on mapping thematic accuracy follow one of two frameworks: probabilistic or fuzzy theory methods.

2.2.2 Fuzzy Methods for Quantifying Uncertainty of Land-Cover Maps

Fuzzy set theory methods have been followed by researches in the GIS community to evaluate classification accuracy using fuzzy error matrices and fuzzy categorical data presented in terms of levels of fuzzy class memberships. Fuzzy approaches are a subject of continuous research (Foody 1996) and have been used to evaluate accuracy in classification using fuzzy error matrices and fuzzy class memberships.

Spatial categories are treated as crisp set where each location is exactly identified as a single category. Therefore a location is either correctly classified or totally misclassified resulting in a crisp binary set. Several types of spatial categories are vaguely defined. For example, the categorization of low residential areas into open space areas is inheritably vague. Fuzziness is used as a way to describe this inexactness and vagueness. Vagueness occurs because the classification system is vaguely defined. Inexactness occurs because the residential area may be incorrectly describe as open space.

The fuzzy theory approach was explored by Zhang and Foody (1998) as an application to assess fuzzy mapping of suburban land cover. Another example of using fuzzy sets for mapping spatial accuracy using thematic land-cover maps was presented by Tran et al. (2005). They defined a multi-level agreement between reference data and the corresponding pixel on the map and developed six levels of agreement. The multi-level agreement fuzzy set regarding a particular land-cover type for each pixel on the map was defined as a discrete fuzzy set. Finally, spatial accuracy maps were developed based on information from the multi-level agreement fuzzy sets. Their results showed different land-cover types had different spatial patterns of accuracy.

Critics of the fuzzy methodology however highlight the arbitrary definition of membership functions and the short development, less established, history of these approaches (Zhang and Goodchild 2002).

2.2.3 Probabilistic Methods for Quantifying Land-Cover Uncertainty

Probabilistic methods of assessing accuracy have long been established for use in a non-spatial context. Such techniques are available in GIS applications for the modeling of geographic data. Probabilities methods follow the framework of geostatistical analysis where geographies of interest are modeled as random variables. A random variable is a variable that can take a variety of outcome values according to a probability frequency distribution (Zhang and

Goodchild 2002). The conventional probabilistic approach to accuracy assessment evaluates the classification in terms of percent of locations correctly classified. Comparison can be made between the classified data and a reference data set. These agreements or disagreements are considered to be random events. For a discrete variable such as land-cover, the probability function is the probability that a specific event will occur.

The idea is that each location on the land-cover map is correctly classified or it has a probability of being correctly classified as a particular class. Steele et al. (1998) used this approach to review misclassification probabilities at training observation locations and then interpolated these estimates using kriging to create accuracy maps for thematic land-cover maps. The accuracy map showed misclassification percent of each map point if the process of sampling and classification rule were repeated infinitely. This method used a straightforward application of kriging to predict spatial accuracy; however, the study utilized training data used to create the land-cover dataset in the kriging process instead of independent reference data collected after the thematic map had been developed to create the accuracy assessment of the dataset. Probabilistic methods are also applied in other geostatistical approaches such as indicator kriging and indicator cokriging. Both methods are being increasingly used for mapping uncertainty in categorical data (Cao, Kyriakidis and Goodchild 2011).

Indicator Kriging is a non-parametric approached that can be used for categorical data and was initially presented by Bierkens and Burrough (1993) in their work on probabilistic soil mapping. They demonstrated indicator kriging as a method to estimate the conditional probability of occurrences of classes of categorical data given classes found at observation locations (Bierkens and Burrough 1993). Indicator kriging was also used by Kyriakidis and Dungan (2001) for mapping thematic accuracy by integrating image-derived data and higher accuracy class labels. Their process used indicator coding methods to determine the probability of observing a land-cover class on the ground and to update these probabilities into posterior probabilities. Stochastic simulation was then used to propagate classification uncertainty to ecological model predictions. Indicator kriging overcomes the limits from other geostatistical methods in which is designed to be used for data that does not follow two assumptions of conventional statistics: data follows a normal distribution and samples are independent.

In indicator kriging, data are pre-processed where indicator values are assigned to each data point using the following criteria: an indicator is set to 0 if the data value at a location ***s*** is below a set threshold, and 1 if otherwise:

$$I(s) = I(Z(s) < \text{threshold}) = \begin{cases} 0, & Z(s) < \text{threshold}; \\ 1, & Z(s) > \text{threshold}; \end{cases}$$

These indicator values are used input to ordinary kriging. Ordinary produces continuous surface predictions where the unsampled locations are given a value between 0 and 1. These predictions are interpreted as the probability that the threshold is exceeded at location ***s***. This prediction surface can be considered as a probability map of the threshold being exceeded (Krivoruchko and Crawford 2005).

Indicator kriging can be applied for mapping individual land-cover classes. However, the distribution of land-cover classes in the landscape is not random but classes are spatially correlated. The land-cover class of interest (primary variable) usually contains one or more correlated land-cover classes (secondary variables). The addition of the cross-correlated information in the secondary variable helps reduce the variance of the estimation error (Isaaks and Srivastava 1989) . This method is useful when two variables are known to be spatially associated such as in the case of land-cover data. The accuracy on a single land-cover type can be improved by combining data of varying accuracies into a single data set using cokriging.

2.3 Generating Alternative Land-cover maps

Almost all data stored in a GIS system are uncertain and when data stored in a GIS database are used as input to GIS operations, the uncertainty will propagate to the output of the operation (Aerts et al. 2003). Much research has been done on reviewing geostatistical tools are available that map the spatial accuracy of thematic data. Much research has also been done on spatial uncertainty model describing methods for generating land cover realizations that can then be used as inputs for error propagation analysis through stochastic simulation (Heuvelink 1998).

Stochastic simulation can be used to evaluate uncertainties in the spatial distribution of the error in categorical data by generating multiple maps of land cover classifications. Hession et al. (2006) provides information on several methods capable of generating multiple realizations of categorical data such as confusion matrix simulation (Fisher 1994), indicator cokriging with sequential indicator simulation (Boucher and Kyriakidis 2006) or simple indicator kriging with varying local means and SIS simulation among others. A growing volume of research focuses on the use of Monte Carlo simulation because of its simplicity for implementation (Kyriakidis and Dungan 2001). Also, the Monte Carlo method does not require knowledge on how the data will be used and it is therefore suitable for many applications.

The idea of the Monte Carlo method used for categorical data follows these basic steps (Krivoruchko and Crawford 2005):

1) The probability of a pixel occurring at a specific category is determined using indicator kriging. All simulated values range between 0 and 1.
2) Compute a statistics of interest.
3) Repeat process 1000+ times.
4) Compare the observed statistic for a given dataset to the distribution of the simulated statistics.

Chapter 3 Materials and Methods

3.1 2001 National Land Cover Dataset

The NLCD was developed by MLRC as a source for land cover information which has been used by a broad spectrum of scientific and governmental applications (Homer et al. 2007). The objective of this dataset is to provide a consistent land-cover layer for all 50 States and provide a dataset that can be applied to a variety of applications.

MLRC first published the 1992 NLCD as a 30-meter resolution land cover data layer over the conterminous United States from 1992 Landsat Thematic Mapper (TM). NLCD 1992 consisted of a 21-class land cover classification scheme that has been applied across the lower 48 United States at a spatial resolution of 30 meters. The classification approach was a modified level 2 version of the Anderson land-use and land-cover classification system (Anderson et al. 1972). Research was pursued to update the 1992 NLCD into a full scale land-cover dataset to produce a layer covering all 50 States and Puerto Rico (Homer et al. 2004).

The 2001 NLCD was created as an update to the previously published 1992 NLCD product. The 2001 NLCD was derived from Landsat 5 and Landsat 7 images according to methodology outlined in Homer et al. (2004). NLCD 2001 improves on NLCD92 in that it consists of three elements: land cover, percent developed impervious surface and percent tree canopy density. The 2001 NLCD maps 16 land-cover classes (Table 3.1 and Figure 3.1) across all 50 States and Puerto Rico at a 30-m by 30-m resolution (Homer et al. 2004).

Table 3.1. 2001 NLCD Land-Cover classes

Code	Description	Code	Description
11	Open Water	42	Evergreen forest
12	Perennial ice/snow	43	Mixed forest
21	Developed, open space	52	Shrub/scrub
22	Developed, low intensity	71	Grassland/herbaceous
23	Developed, medium intensity	81	Pasture/hay
24	Developed, high intensity	82	Cultivated crops
31	Barren land (rock/sand/clay)	90	Woody wetlands
41	Deciduous forest	95	Emergent vegetation wetlands

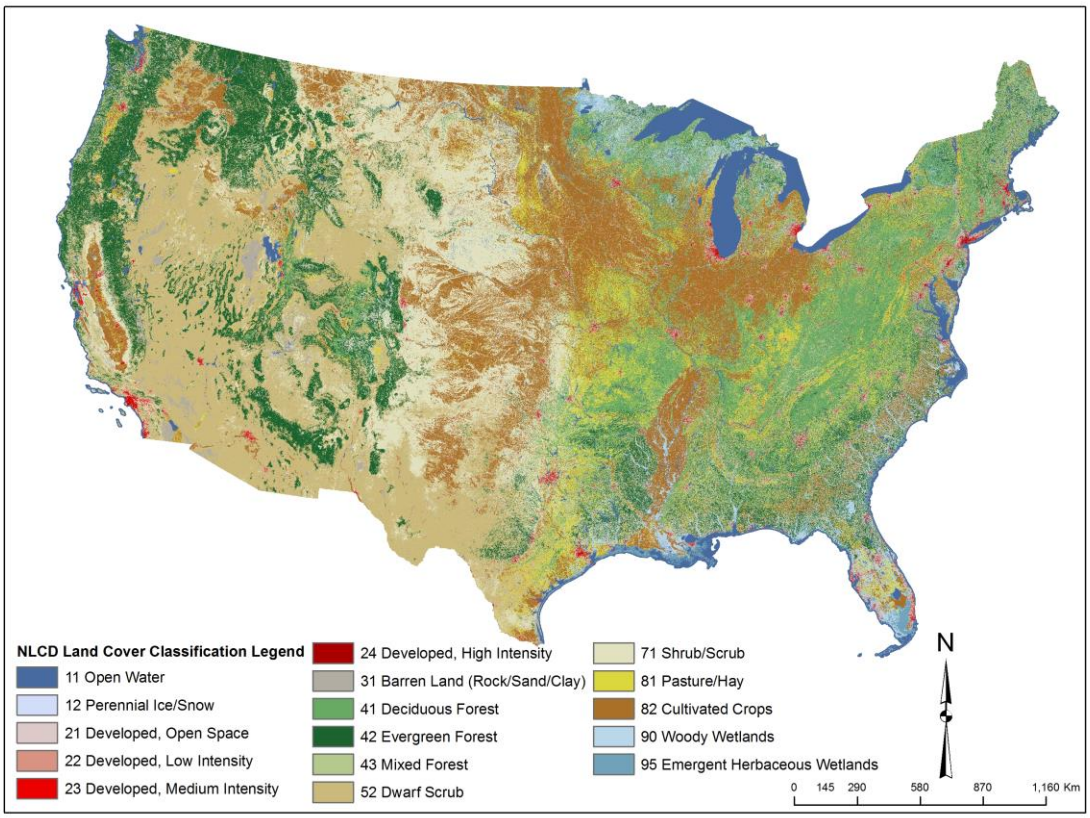

Figure 3.1. 2001 NLCD Land-Cover Classification

The latest version of the NLCD is the 2006 NLCD. Similar to the 2001 version, the 2006 NLCD consists of a 16-class land cover classification scheme that has been applied across the conterminous United States at a spatial resolution of 30 meters. NLCD2006 is based on classification of Landsat Enhanced Thematic Mapper+ (ETM+) circa 2006 satellite data. NLCD2006 also quantifies land cover change between the years 2001 to 2006.

The 2001 NLCD was used in this thesis. Analysis was focuses on forest land-cover classes (Figure 3.2). Forest categories include pixels classified with codes 41, 42 and 43.

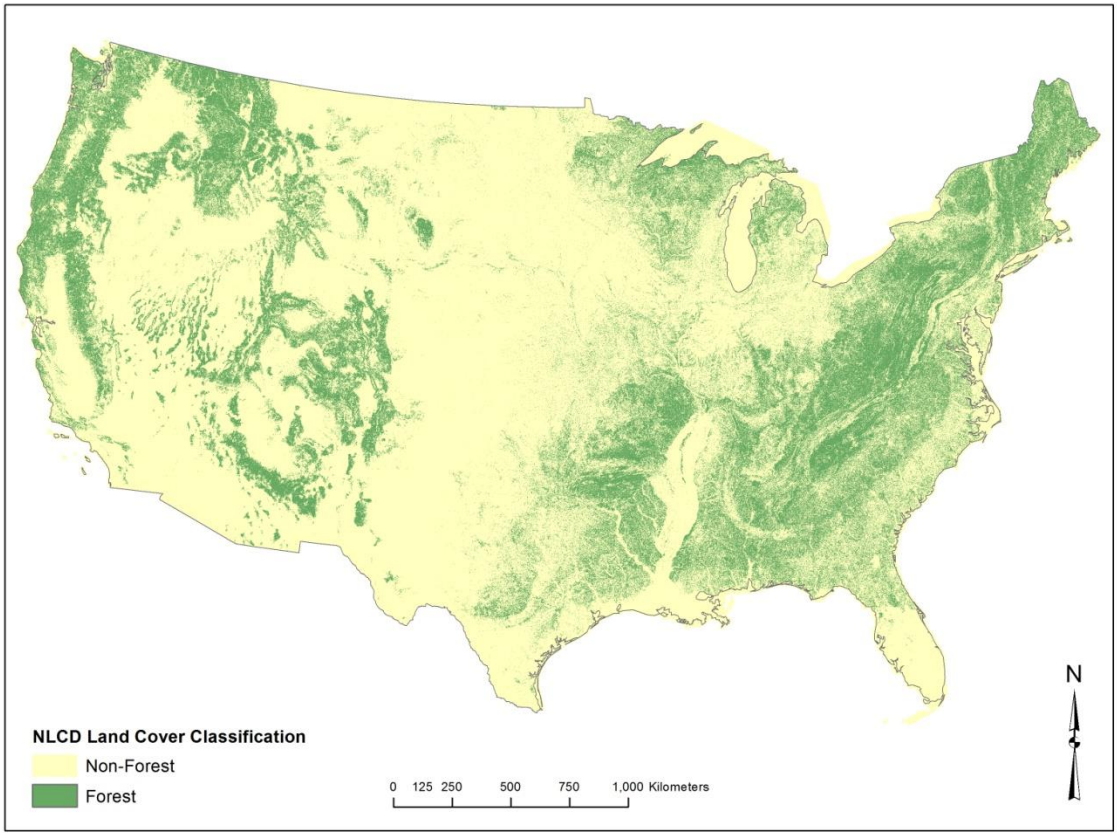

Figure 3.2. 2001 NLCD Forest/Non-Forest Classification

3.2 Ground Truth Dataset

The accuracy assessment of the 2001 NLCD was published in 2010 by Wickham et al. (2010). The conceptual framework for the assessment was outlined by Stehman et al. (2008) and it is based on the sampling design implemented for the 1992 NLCD.

The sampling design of the 2001 NLCD accuracy assessment was a two-stage level cluster sample with three levels of stratification. The first level of stratification divided the conterminous United States into 10 regions which were then further divided into frame cells that were 120-km by 120-km – these frame cells are the second level of stratification. The first stage sample was created by randomly selecting 12-km by 12-km sampling units within each sampling region. Fifty five sampling units were selected within each region. The last layer of stratification was a land-cover class level where 100 sample pixels of each land-cover class were sampled within each of the 10 regions using stratified random sampling from the sampling units. The perennial ice/snow land-cover category was excluded from the sampling. A total of 15,000 samples were collected for the assessment (Wickham et al. 2010). Figure 3.3 shows the location of the ground truth points within each of the 10 regions. A copy of the reference dataset was obtained from personal contact with Jim Wickham.

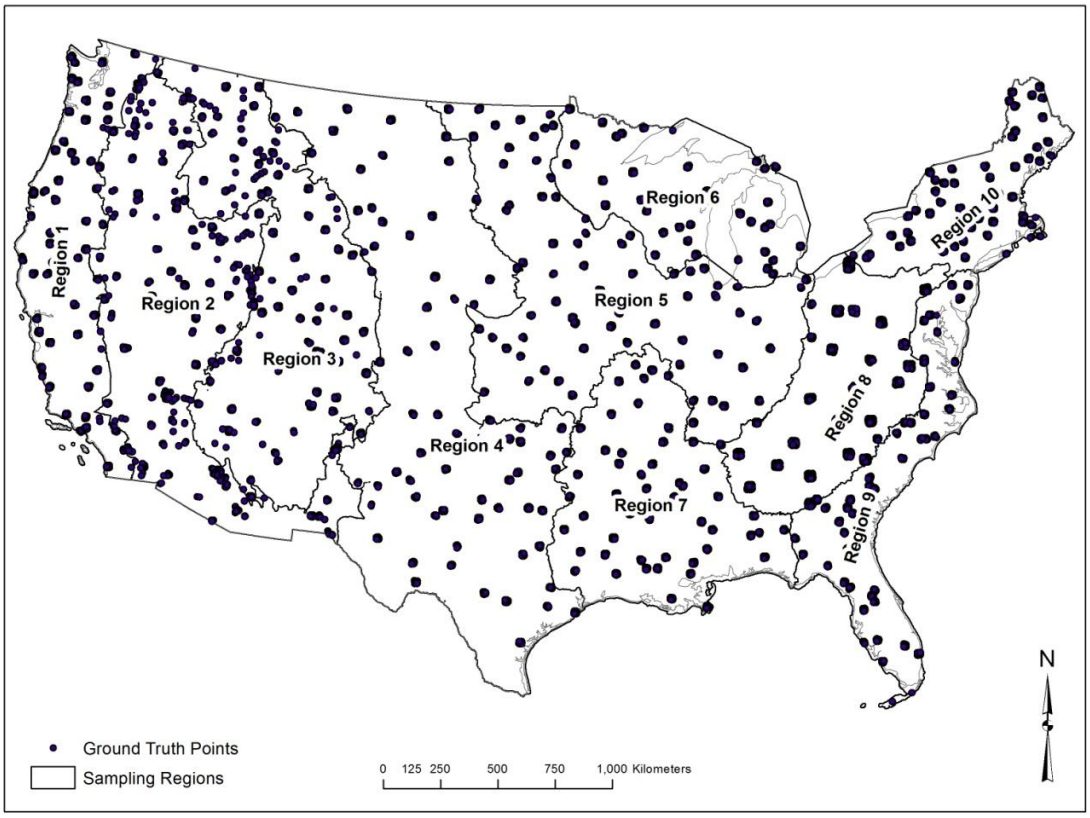

Figure 3.3. Location of All Sampling Points within the 10 Regions

This reference dataset includes the following primary attributes derived from the reference data: 1) the primary label, 2) the alternate label, 3) photointerpreter confidence. Attributes derived from the map included 4) the sample center pixel map label, 5) the modal map label for a 3x3 pixel window surrounding the sample pixel, and 6) boolean field listing agreement between reference point and pixel. Primary and alternate labels are based on codes listed in Table 3.1. The primary definitions of agreement used by Wickham et al. (2010) were either the mapped land-cover class of the sample pixel matched either the primary or alternate reference land-cover label or a modal land-cover class matched either the primary or alternate reference land-cover label. The model map classes were determined on a 3x3 pixel window centered on the sampled pixel. Two new definitions of agreements were established for this study.

The first definition of agreement was based on the case where both the map label and reference label attributes were attributed with any combination of forest land-cover class categories (i.e. 41, 42 or 43). Based on this definition, a pixel where, for example, the map label was calculated as 41 and the reference label was calculated as either 41, 42 or 43 would be considered to be in agreement. In addition, all points where the map label was attributed in any of the non-forest categories and the 'Agree' field was attributed as 1 was changed to 0. This added information about the known location of non-forest points. A field called 'New Agree' was added to the reference dataset to calculate the agreement values as either '1' or '0'. A point attributed as '1' in the 'New Agree' field indicates the points meets the agreement definition. A point attributed as '0' when the point did not meet the agreement definition. This definition of agreement was used during the indicator kriging analysis.

The second definition of agreement was based on the case where points with map label attributed as forest (values 41, 42 or 43) did not match the reference label but did match the both alternate label. This definition of agreement was used during the indicator cokriging analysis.

Accuracy results for the 2001 NLCD were reported for the 10 geographic regions in a series of error matrices listing the user's and producer's accuracies per region for each land-cover class. Overall nationwide accuracy values are also provided. Additional details about the methodology and results of the accuracy assessment have been described in Wickham et al. (2010) .

3.3 Probabilistics Methods

3.3.1 Indicator Kriging

Indicator kriging was used to predict the probability of a pixel at unvisited locations belonging to a forest or non-forest land-cover class. A 'New Agree' field was added to the attribute table and was calculated as 0 or 1 based on the first

agreement definition described in the previous section. A point attributed as ‘1’ indicated the map label belonging to any of the forest land-cover classes (i.e. 41, 42 or 43) matches the reference label. This established agreement at the Level 1 NLCD classification. Any point attributed as ‘0’ in the dataset indicates map label information does not match the Level 1 definition of agreement and those pixels does not belong to any of the land-cover categories under review. Figure 3.4 shows the location of points labeled where there was/wasn’ an agreement. Indicator kriging was done using all ground truth points within the conterminous U.S. and also dividing points for each mapping region.

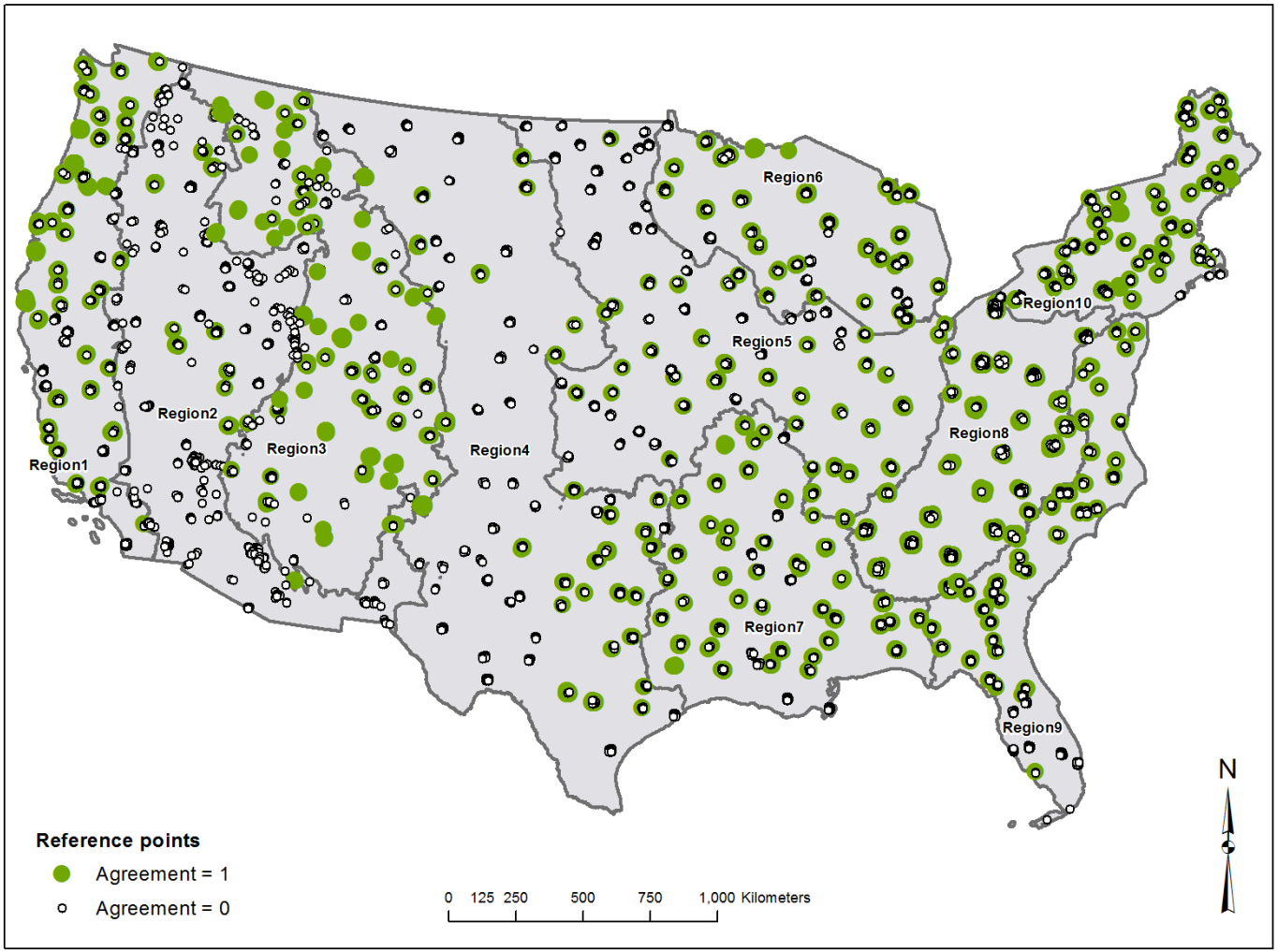

Figure 3.4. Location of Sampling Points Used for Indicator Kriging

Geostatistical Analysis tools such as indicator kriging were developed for examining statistical analysis to examine the spatial correlation of the data. If data are spatially independent, we cannot predict values at unknown locations. The semivariogram and covariance functions can be used to quantify the strength of the spatial correlation between data points as a function of (ESRI 2003). The semivariogram model available in the Geostatistical Analyst Extension in ArcGIS 10.0 was used. Using ground truth information, the semivariogram was fitted to match a particular model that appropriately defines the mathematical properties of the data (O'Sullivan and Unwin 2010). Three model parameters were evaluated: model chosen, the effect of anisotropy and lag size.

The process of creating the best suited semivariogram for a dataset may involve adding information about the effect of direction in the data. This is known as anisotropy. Anisotropy accounts for the semivariogram to change not only with distance between the points but also with direction. Isaaks and Srivastava (1989) described the effect of anisotropy in ordinary kriging. The idea is that in many data sets, data values are more continuous in certain directions than in others. Local fluctuations may be found in the direction of the anisotropy. ArcGIS Geostatistical Analyst extension allows accounting for anisotropy by flagging 'Anisotropy' as 'True' in the Semivariogram/Covariance Modeling window. Anisotropy was evaluated using the 2001 NLCD ground truth points.

The lag size chosen in the semivariogram modeling process can also have an effect on the results. If the lag distance is too large, autocorrelation at short distances may be missed. A small lag size may create empty bins in the data (ESRI 2003). Selecting the appropriate lag size is not straightforward when the data uses an irregular random sampling scheme. The ground truth data used in this thesis has an irregular sampling scheme. Different approaches were used to determine the most appropriate lag size to use on the NLCD ground truth data. One approach was using the 'Average Nearest Neighbor' tool available in ArcGIS 10.0 to determine the average distance between points and their closest neighbors. Other approaches used included the default lag size assigned by the semivariogram modeling tool and running the 'Optimized' tool which focuses on minimizing the mean square error of the data (ESRI 2010).

ArcGIS provides several functions to model the semivariogram including circular, spherical, Gaussian, and stable among others. The model selected will influence the prediction of the unknown values and determine the shape of the curve near the origin. The steeper the curve is near the origin, the more influence the closest neighbors have on the prediction. The result is a less smooth output surface. Also, as local variation in the surface increases, the range decreases and the nugget value increases. The spherical, exponential and Gaussian models were evaluated as part of the analysis to identify which model best fits the data. The exponential model is applied when spatial correlation decreases

exponentially with increasing distance, i.e., correlation disappears at a short distance. This spherical model shows a progressive decrease of spatial autocorrelation (equivalently, an increase of semivariance) until some distance, beyond which autocorrelation is zero. The Gaussian model represents very smooth behavior at short distances (Kanevski and Maignan 2004).

The final cross-validation process provides information about which model provides the best prediction by calculating a series of statistics. The goal of the kriging process is to minimize the variance of the error and to create a mean residual of the error which is closest to 0. Many estimation models aim to minimize the mean residual error and the distinctive property of ordinary kriging method used with indicator kriging is to minimize the variance (Isaaks and Srivastava 1989). ArcGIS Geostatistical Analyst tool provides several errors estimators on the results which can be reviewed to identify which model will produce the best results. Appendix A.1 includes a list of all the error estimators described below and the formula used to calculate each prediction.

The 'mean prediction error' provides a measurement of bias in the model. This value is calculated as the averaged difference between the measured and the predicted values. The results looking for are those where the prediction is unbiased and has a mean residual error closest to 0.The 'mean prediction error' calculated depends on the scale of the data. To standardize these results, ArcGIS kriging outputs also provide the 'mean standardized', which is calculated as the 'mean predictor error' divided by the prediction standard error at each location. These results should also be near 0.

Bias is not the only source for estimating error. An unbiased estimator does not imply an estimated value is equal to the true value, it only implies the errors are zero on average. In addition to the estimator being unbiased, we aim for individual errors and variances to be small. This is reported through the 'root-mean-square error' (RMSE), which is calculated as the square root of the square 'mean prediction error.' The RMSE indicates how closely the model predicts the measured values. RMSE is a better indicator of error than the mean prediction of error because it includes an estimation of both bias and variance and it indicates predictions do not deviate much from the measured values. Taking the square differences of the mean values avoid positive and negative values from canceling each other out and it provides a more reliable source of error (Burt, Barber and Rigby 2009).

Other estimators of errors provided by ArcGIS are the 'average standard error', and 'root-mean-square standardized error' (RMSSE). The average standard error is the average of the prediction standard error. The RMSSE is the square root of the squared mean standardized error. This predictor indicates whether the variability of model predictions is being over or under estimated. An 'average standard error' with a value close to RMSE indicates the variability in

the prediction is being correctly identified. If the ‘average standard error’ is greater than RMSE, the variability in the prediction is being overestimated. The RMSSE value should be close to 1 to indicate prediction standard values are correct. If the RMSSE is greater than 1, the variability of the estimators is being underestimated. A RMSSE value less than 1 indicates the variability of the estimators is being overestimated (ESRI 2003).

Indicator kriging produced a probability surface indicating the chance that a pixel is classified in the forest or non-forest land-cover category.

3.3.2 Indicator Cokriging

Indicator cokriging was used as a way to minimize the variance and residual error estimated from the indicator kriging of the forest land-cover class using other land cover categories as secondary variables. The methodology used for indicator co-kriging process followed the same as with indicator kriging analyzing the effect of the model, lag size and anisotropy in the data. The co-kriging tools available in ArcGIS allows for using up to four variables. The primary variable used was points classified in the forest land-cover category. That includes points where both the map and reference labels were attributed as 41, 42 or 43. The secondary variables used were ground truth points attributed as a forest land cover category where the map and reference labels do not match, but the map and alternate labels do. Like indicator kriging, cokriging was done using all ground truth points within the conterminous U.S. and also dividing points for each mapping region. Figure 3.5 shows points used for indicator cokriging..

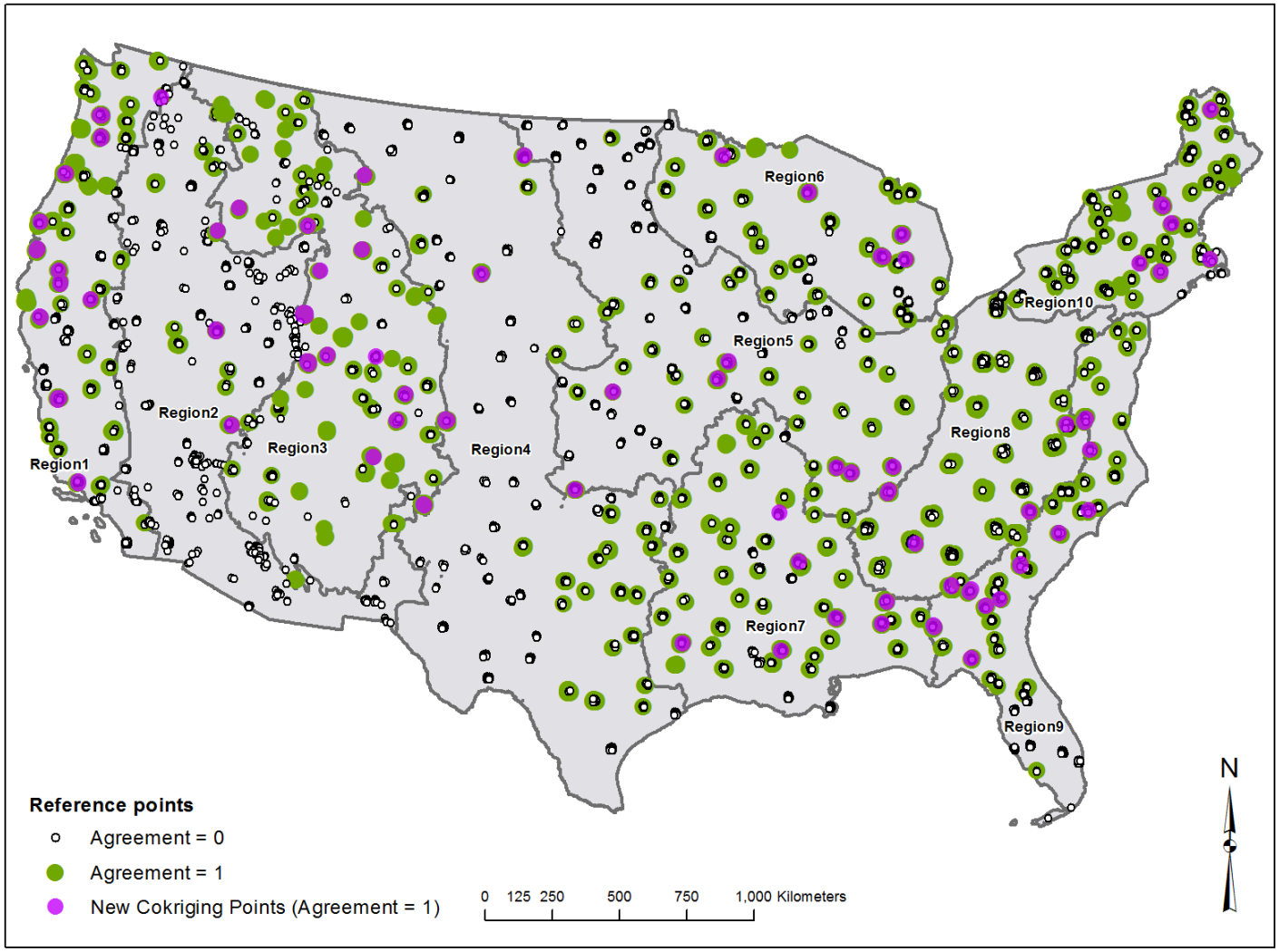

Figure 3.5. Location of Sampling Points Used for Indicator Cokriging.

3.3.3 Monte Carlo Simulation

There are many methods which can used for evaluating the propagation of quantitative errors in spatial operations. Examples are Taylor series approximation, Rosenblueth's method, and Monte Carlo simulation (Aerts et al. 2003). Stochastic simulations such as Monte Carlo are used to generate alternative outcomes of an unknown parameter over an area of interest and are valuable methods for assessing the uncertainty resulting from geoprocessing operations in GIS (Kyriakidis 2001, Krivoruchko and Crawford 2005). The idea is to repeat a process many times with input values that are randomly sampled so that statistics such as mean, variance, etc, can be computer from a sample distribution. Monte Carlo simulations proceed as follows:

1. Simulated values are created for the probability distribution of the input data under study.
2. Generate a set of random inputs
3. Evaluate the model and store results.
4. Process is repeated many times (i.e. 1,000 times)
5. Compute and analyze the results using histograms, summary statistics, etc.

Monte Carlo simulation was the method used in this thesis to create multiple realizations of the forest and non-forest land-cover classes using the probability surfaces created through the kriging process outlined in the previous section of this chapter.

Figure 3.6 outlines the workflow followed. The process involved creating a random raster and comparing it with coincident pixels in the probability surfaces created from the indicator kriging and indicator cokriging analysis. Pixels where the value in the random raster was less or equal to the probability surfaces were considered to become a forest pixel. When every location in the raster was visited, the realization was stored and the procedure was repeated by following a new random path through all cells, generating a new realization. This process was run 1,000 times. With geostatistics simulation, multiple realizations of the probability surface were created resulting in a distribution of the spatial uncertainty in our data. The multiple realizations are possible versions of the NLCD forest land-cover classes (Krivoruchko and Crawford 2005). The Monte Carlo simulation was run for the probability surface created through indicator kriging, and the three probability surface scenarios created through indicator cokriging. This workflow was done using Model Builder in ArcGIS 10.0 (See details in Appendix A2).

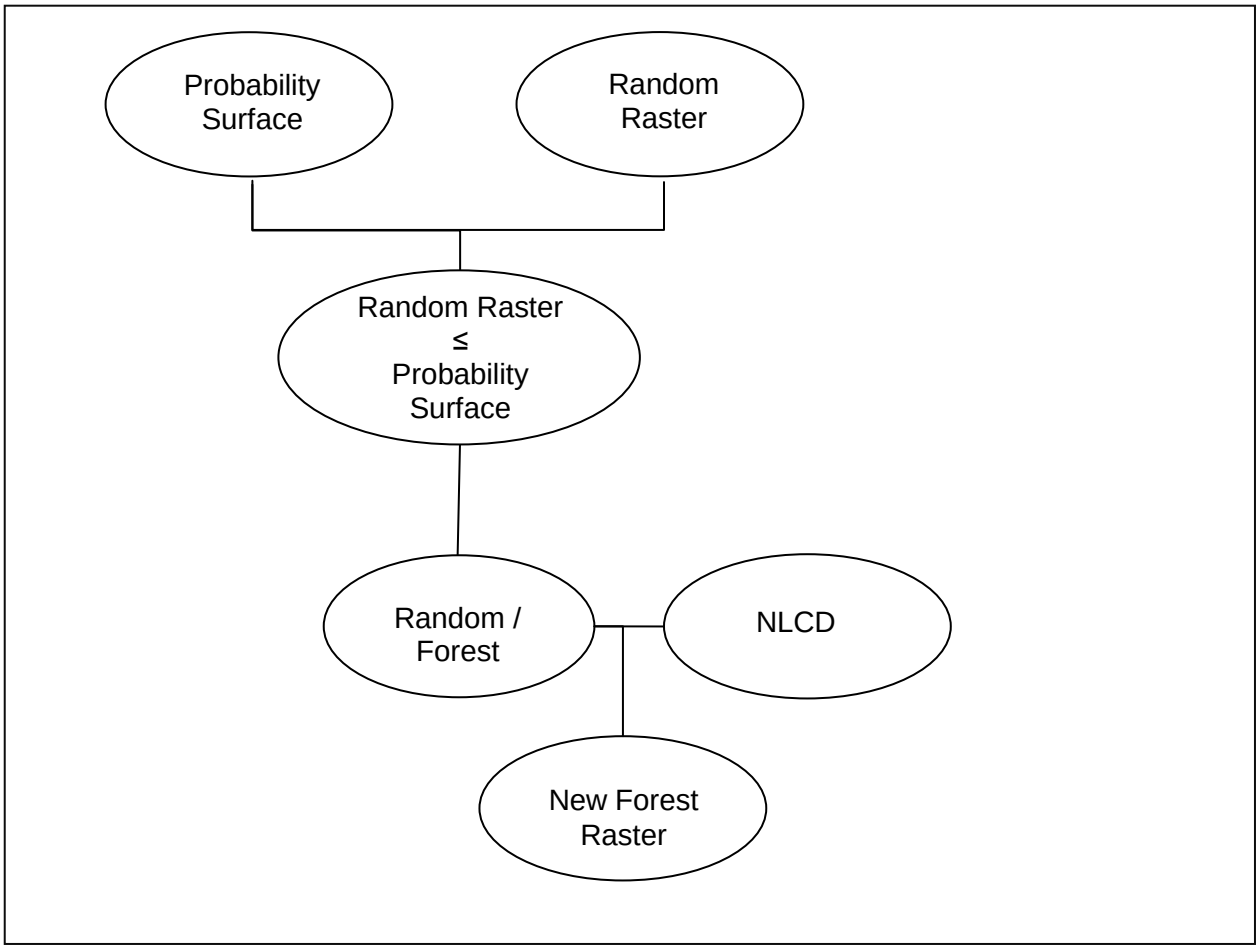

Figure 3.6. Monte Carlo Simulation Decision Workflow.

Chapter 4 Results and Discussion

4.1 Results

4.1.1 Indicator Kriging

Indicator kriging was done using ground truth points where the map label was a forest land-cover class (values 41, 42 or 43). The ‘New Agree’ field in those points was classified as either 1 or 0 depending if the map and reference label matched. Ground truth points for other land-cover classes where the map and reference labels agreed (‘Agree’=1) were included in the indicator kriging process as well but the ‘Agree’ value was recalculated as ‘0’ to indicate these points are not classified as forest. Indicator kriging was done for points within the conterminous U.S. and also for points in each mapping region.

Nationwide Results

Figure 4.1 shows the isotropic semivariograms resulting from indicator kriging using Gaussian, spherical, and exponential models using all points within the U.S. The models shown assume the covariance between the data depend only on the distance between points and not on the direction. Default ‘lag size’ of 107,000 meters was used. The semivariograms indicate spatial correlation reach a plateau between 16 km and 27 km (Table 4.1). All graphs show a constant pattern after the range distance. Information listed in Table 4.1 indicates the Gaussian model shows greater variation in the data as it has the smallest range and larger nugget. The exponential model resulted in the smallest nugget and largest range, indicating correlation in the data at longer distance.

Table 4.1. Output Parameters of Isotropic Semivariogram Models

Model	C_o	a (km)	$C_o + C_1$
Spherical	0.070	26.5	0.16
Gaussian	0.070	15.8	0.16
Exponential	0.064	28.6	0.16

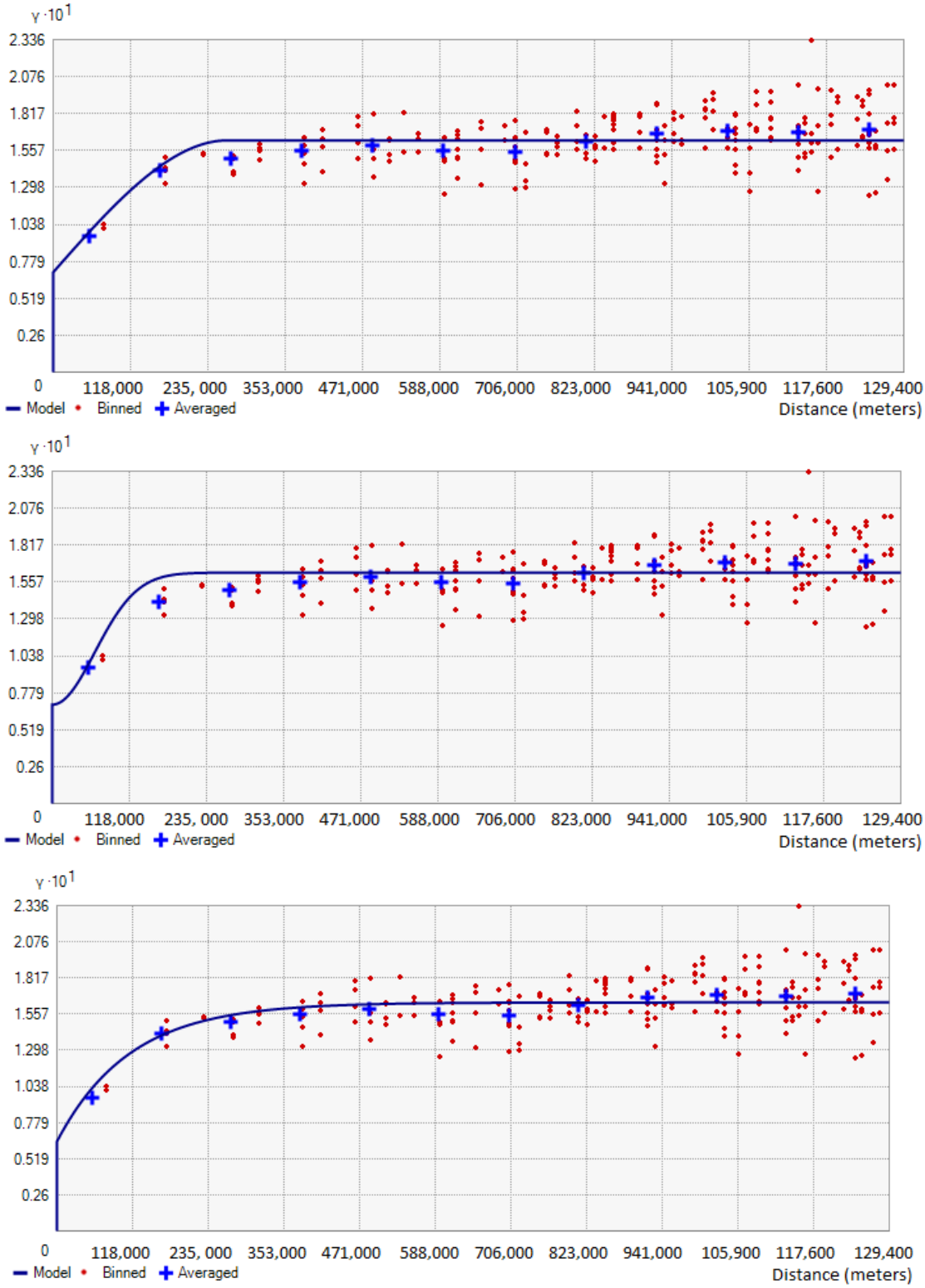

Figure 4.1. Semivariogram for Isotropic Spherical, Gaussian and Exponential Models (top to bottom).

The effect of direction (anisotropy) was evaluated next. Results using all ground truth points are shown on the semivariograms in Figure 4.2. Output parameters are listed in Table 4.2. Default lag size was used for all models. Anisotropy accounts for the fact that values in many datasets may be more continuous along certain directions than along others. Using an anisotropic model puts more of the weight to sample points in the direction of maximum continuity (Isaaks and Srivastava 1989). The semivariogram map (Figure 4.2) shows data has maximum continuity is on the SW-NE direction. Anisotropic semivariograms were fitted to that direction. Output parameters (Table 4.2) show the three models reached range between 18 to 34 kilometers. The range increased in the three models when anisotropy was included. Graphs show a constant pattern after reaching the range. The nugget also increased in all three models when anisotropy was considered. The sill decreased, indicating incorporating direction effect decreases the variance on the results.

Table 4.2. Output Parameters of Anisotropic Models.

Model	**C_o**	**a (km)**	**$C_o + C_1$**
Spherical	0.083	27.4	0.13
Gaussian	0.082	17.6	0.13
Exponential	0.084	34.2	0.13

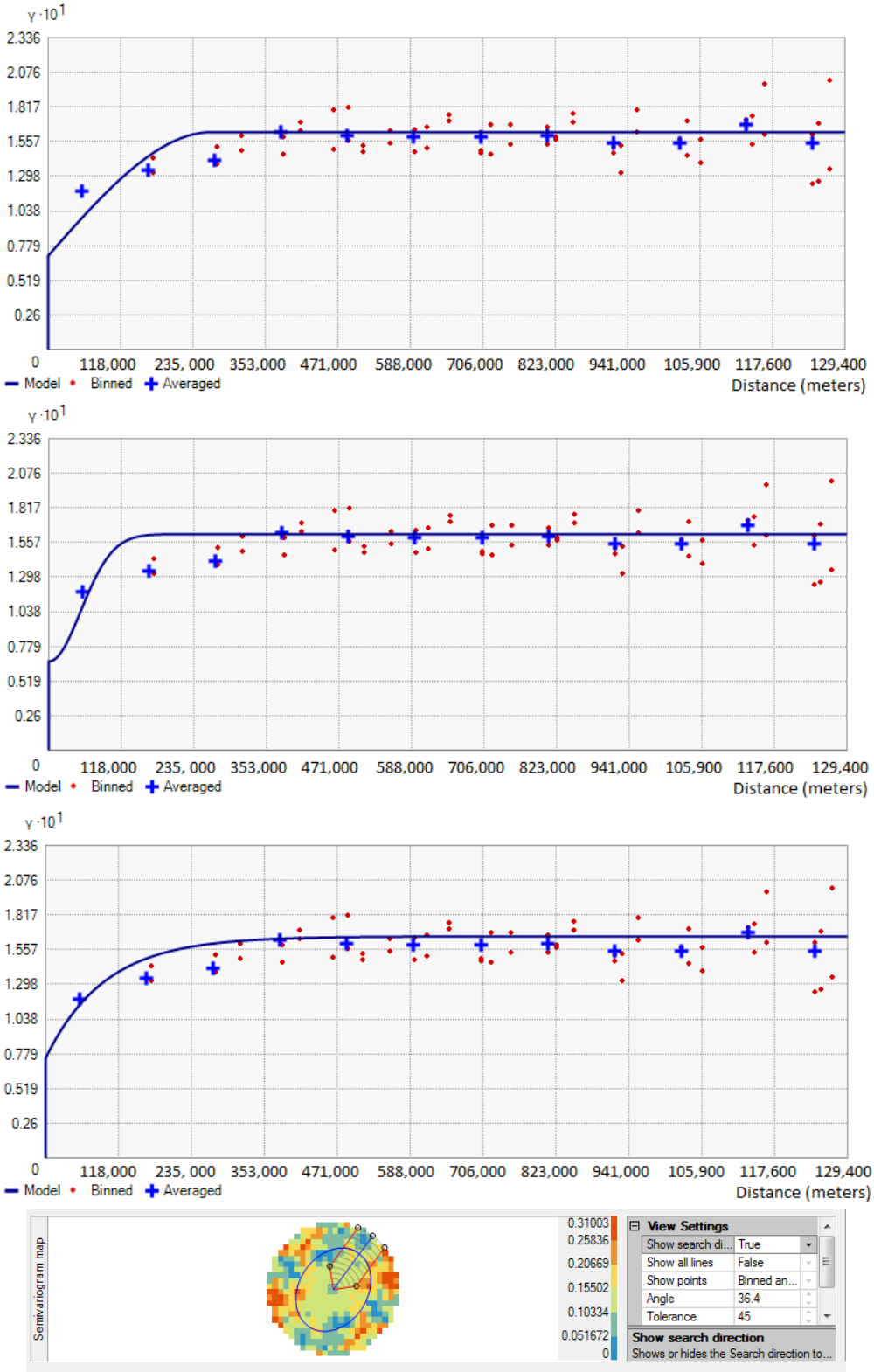

Figure 4.2. Semivariogram of Anisotropic Spherical, Gaussian and Exponential Models.

The effect of the lag size selected was also evaluated. Results chosen are shown on Figure 4.3 and Table 4.3. Selecting a lag size that is too big may mask correlation at short distances. However, selecting a lag size that is too small may create empty bins or bins too small to be representative of the data. Different lag sizes were tested on the data including the lag size selected using the 'Optimize model' option in ArcGIS. This tool allows the program to select the ideal lag distance for each model that would minimize the error and variance. Optimizing the data decreased the lag distance used for all models. Anisotropy was also included in these results. The range (Table 4.3) was reached between 34 km to 47 km. Using a smaller lag size in combination with an anisotropic model reduced the variance in all scenarios and increased both the range and the nugget. The semivariograms in Figure 4.3 show how data fluctuates as the range becomes larger. The spherical model resulted in the smallest variance and range among the three models.

Table 4.3. Output Parameters of Optimized Models

Model	**C_o**	**a (km)**	**$C_o + C_1$**
Spherical	0.053	46.6	0.11
Gaussian	0.058	34.3	0.12
Exponential	0.083	34.2	0.11

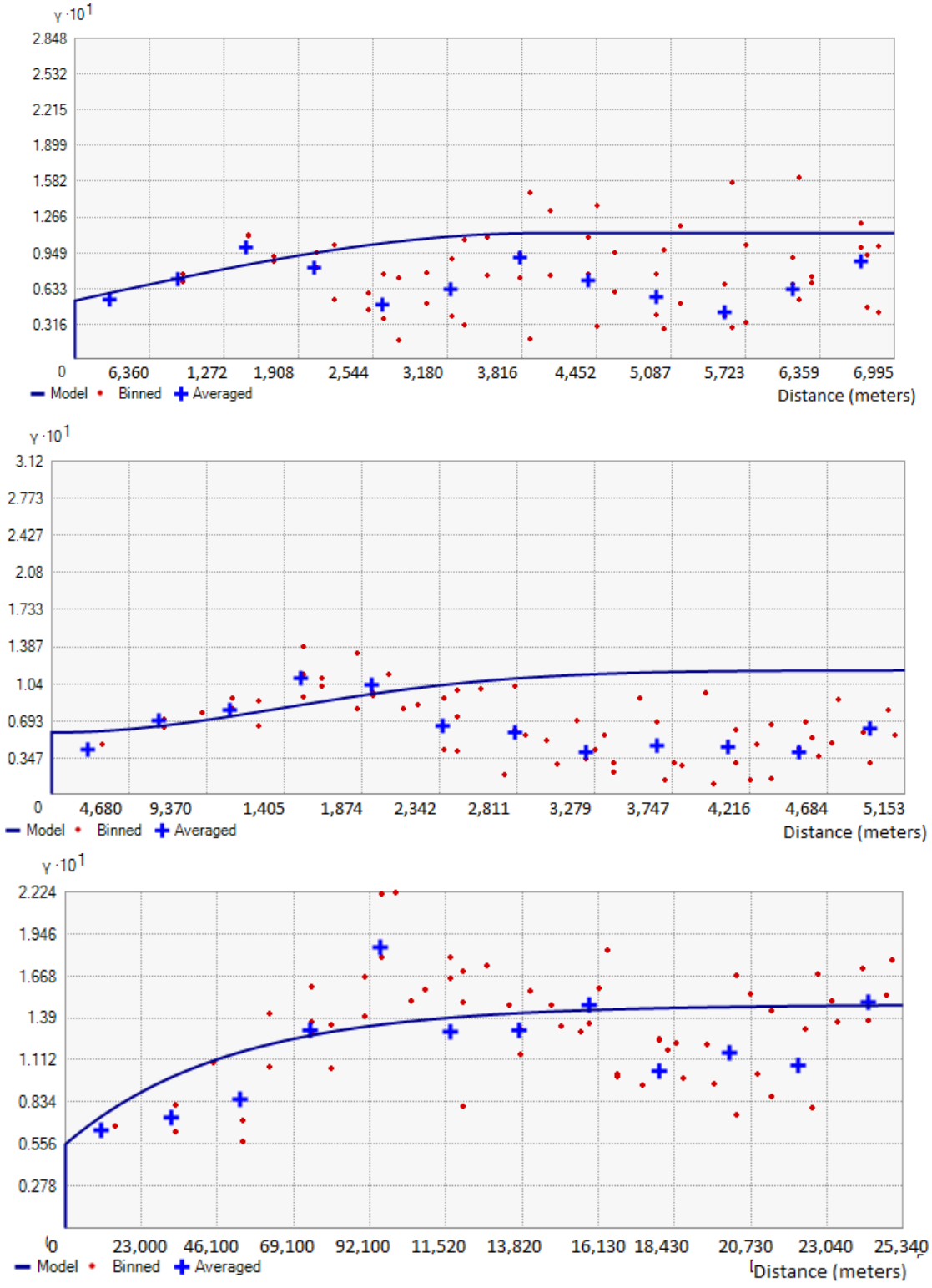

Figure 4.3. Semivariogram for Spherical, Gaussian and Exponential (top to bottom) models using lag distance calculated by the Optimized Model.

Results indicated the optimized anisotropic semivariogram model created the smallest variance for all models. The next step was to use the optimized anisotropic scenario within the indicator kriging tool to identify which model is better suited for the data. Results using Gaussian, exponential and spherical models were compared. Table 4.4 includes prediction errors created by each model. Both the mean prediction error and the mean standardized are smallest using the exponential model. This indicates the exponential model provides the most unbiased results. The exponential model also has the smallest RMSE. The average standard error is smaller than the RMSE for all models, indicating all models underestimate the variability of the values. The RMSSE is greater than 1 in all models; also an indication that all models underestimate the variability of the datasets.

Table 4.4. Indicator Kriging Prediction Errors for Conterminous United States

Optimized Model					
Model	**Mean Prediction Error**	**Root-Mean-Square Error**	**Mean Standardized**	**Root-Mean Square Standardized Error**	**Average Standard Error**
Spherical	0.00047	0.36	0.0020	1.10	0.32
Gaussian	0.00047	0.36	0.0020	1.13	0.32
Exponential	0.00032	0.35	0.0016	1.08	0.33

Note that the goal of the kriging process is to create an estimation that is both unbiased (mean residual error is equal to 0) and the variance is minimized (Isaaks and Srivastava 1989). Furthermore, RMSE is the estimator which provides the most accurate estimation of error as it takes into account both the bias and variance factors. This information suggests the optimized anisotropic exponential model better fits the data and will create a more accurate surface since it resulted in the smallest RMSE.

The probability surface created by fitting the optimized anisotropic exponential model for the conterminous U.S. is shown in Figure 4.4. The probability of a pixel to be a forest cell increases from light yellow to dark brown areas. The image shows the mapping regions used by Wickham et al. (2010) in the NLCD suggest hard boundaries between forest and non-forest areas at region boundaries. For example, region 3 shows forest land along its boundaries that seem to end abruptly and do not transition to regions 2 and 4. The probability surface however shows, for example, the forest land transitions from the south-eastern corner of region 3 through region 4. Reviewing the spatial distribution of these probability values along boundary lines would allow to better delineate where mapping regions should be placed.

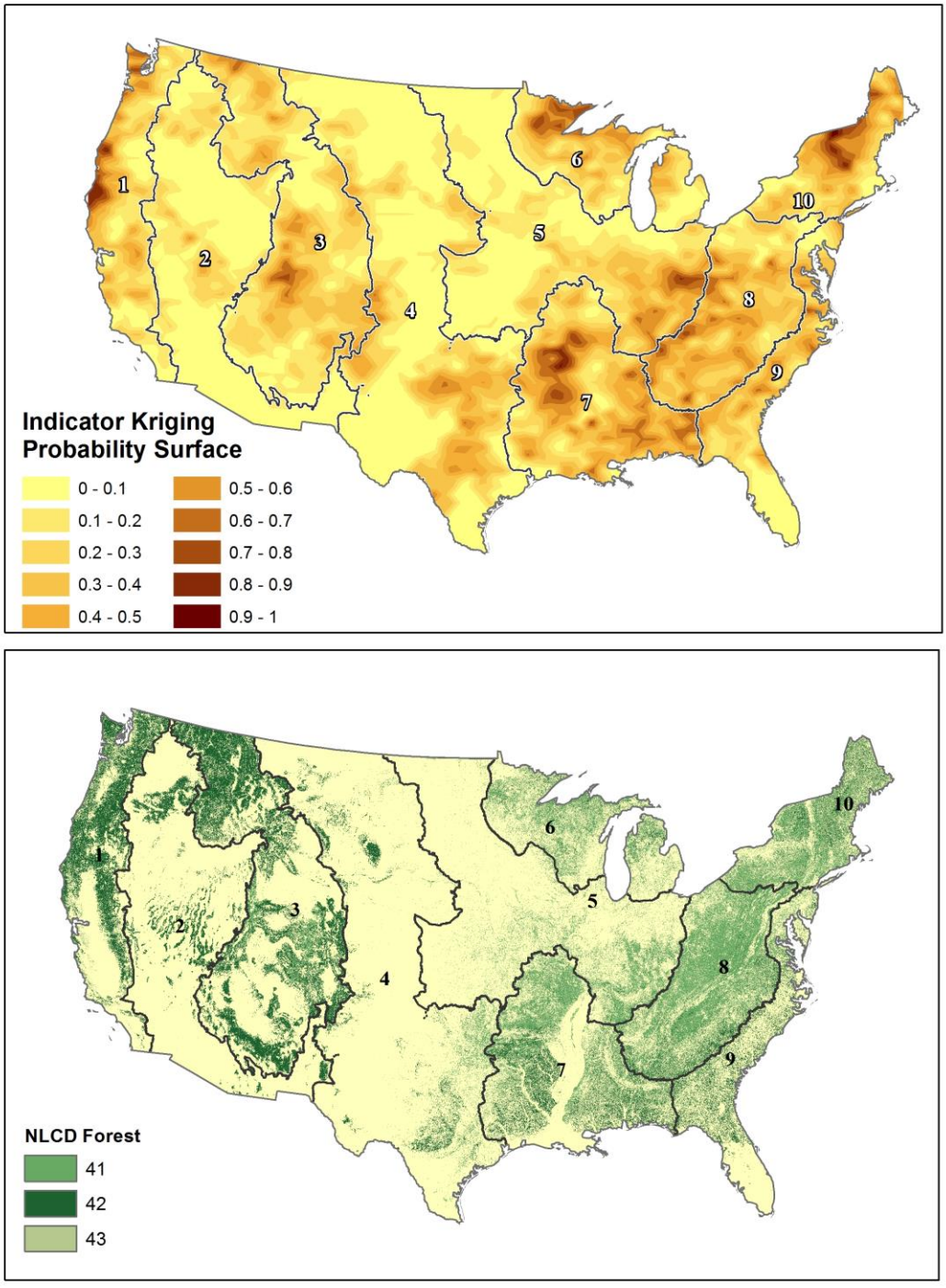

Figure 4.4. Indicator Kriging Probability surface (top) and 2001 National Land Cover Dataset Forest Land-Cover (bottom)

Results show that approximately 10% percent of pixels (196 million acres) in the probability surface created for the entire United States have a probability equal or greater than 50% of being classified as forest (Table 4.6). Approximately 26% of pixels (460 million acres) nationwide are classified as forest in the 2001 NLCD.

Tables 4.7 and 4.8 compare the user's and producer's accuracies for the U.S. in the 2001 NLCD and the probability surface created through indicator kriging. Overall accuracy for the probability surface is also reported. NLCD's accuracies are available from Wickham et al. (2010). The overall accuracy of the probability surface is 66%. The producer's accuracy is "good" (86%), but smaller than the producer's accuracy of the 2001 NLCD (88.5%). Even though the overall accuracy of the classification by the probability surface is poor, it is adequate for the purpose of mapping the forest class. However, the user's accuracy of the probability surface is only 28%. This indicates that even though 86% of the forested areas have been correctly identified as forest, only 27% percent of the areas are truly of forest category. On the other hand, the 2001 NLCD did a better job at classifying forest land-cover across the U.S as 87% of the areas classified as forest are truly of that category.

Figure 4.4 shows the overall spatial distribution of forest land in the probability surface coincides that of the forest land in NLCD. The producer's accuracy of both surfaces would indicate that both layers reasonably are able to classify forest land within the conterminous U.S. Using a probability greater than 50% of a cell classified as forest, areas where NLCD and kriging surfaces differ can be identified. One example of the differences between both surfaces can be found in region 1 where NLCD shows mostly a continuous band of forest land along a north-south band in the eastern part of region 1. The kriging results show the probability of this area of being forest is less than 50% through the entire area. Other differences are found in regions 8 and 10 where NLCD also shows consistent forest land coverage but the probability surface shows the change of the area along the boundary between those two regions have a probability less than 40% of being classified as forest. Table 4.8 shows both the user' and producer' accuracies in regions 8 and 10 are very small. The kriging process in these two regions both was not able to properly map the forest class and those pixels classified as forest are not truly forest land.

Results from the exponential model listed in Table 4.4 shows the mean prediction error and mean standardized prediction errors as being 0.0032 and 0.0016 respectively. These results indicate there is bias in the results as the values are not zero. The error in the individual points ranges from -1 to 1, indicating points with an indicator value of 1 were predicted to have a value of 0, and points with an indicator value of 0 were predicted to have a value of 1. These points are scattered throughout all regions. The mean error was 0.00045 while

the standard deviation of the error was 0.416 indicating error values are dispersed across the country.

The average standard error of all points was 0.33, which is smaller than the RMSE of 0.35. This indicates the variability of the predictions from the true values is being underestimated by the indicator kriging process. The standard error in all points ranges from 0.29 to 0.42. It is smaller around data points and increases with distance from sampling data points (Figure 4.5). Additionally, the RMSSE is greater than 1 which is also an indication the predictions are being underestimated.

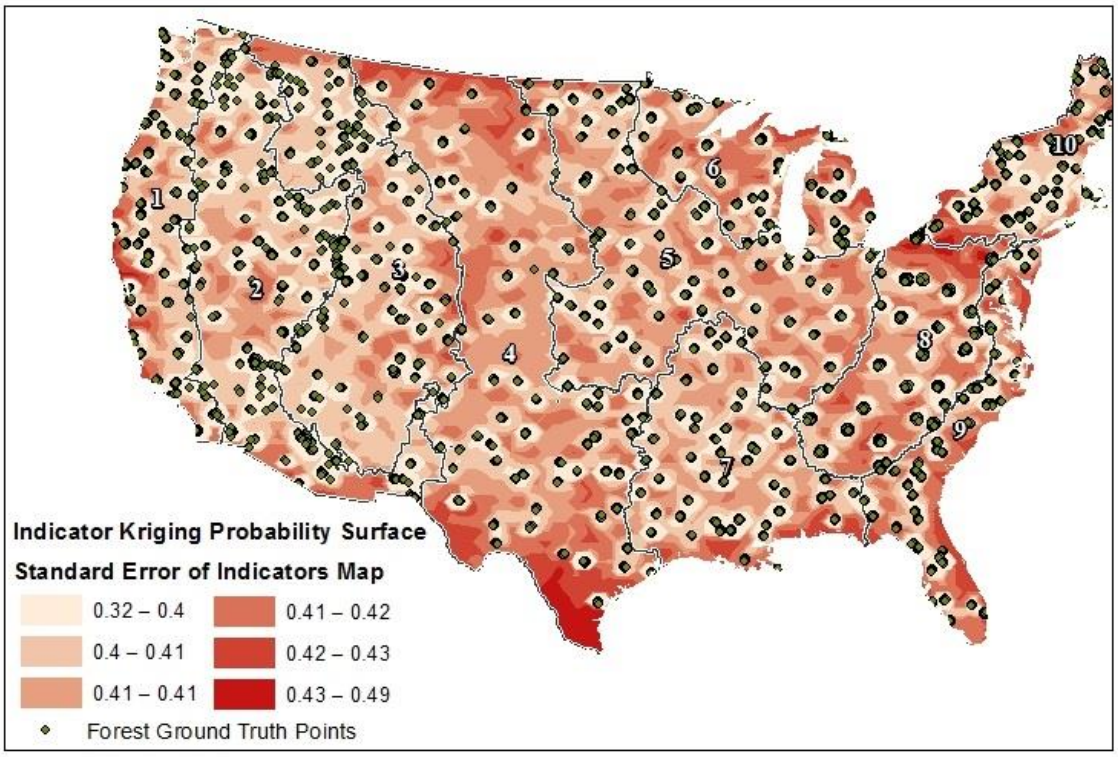

Figure 4.5. Standard Error Surface from Indicator Kriging Surface.

Region Results

The same exponential model scenario was repeated dividing ground truth points per the 10 mapping regions used by Wickman et al (2010) in the accuracy assessment of the 2001 NLCD. Prediction errors per region are listed in Table 4.5. The probability surfaces created by fitting the optimized anisotropic exponential model mapping region are shown in Figure 4.6. Each region in Figure 4.6 is a separate raster. The Figure shows areas in some regions, such as region 10, where there were not enough reference points to interpolate information about the forest land in these sections. The probability surface does not have any values in these areas. Additionally, as it was seen in the nationwide probability surface shown in Figure 4.4, the hard boundaries used by Wickham et al. (2010) to delineate mapping regions do not always coincide where boundaries between regions would be delineated using information in the probability surface. Both Figures 4.4 and 4.6 show the same transition area in the probability surface between the forest land in the south-eastern corner of region 3 and region 4. The NLCD shows the forest land ends at the region 3-side of the boundary between. Region 4 does not show any forest land. The mapping regions may not correctly divide regions at this location. An example where the mapping regions may be accurate can be seen in the boundaries between regions 1 and 2. The NLCD shows a band of forest land along the eastern side of region 1. This forest land does not appear in region 2. The transition between these two regions however is delineated by the Rocky Mountains. The mapping regions are correctly delineated.

Figure 4.7 shows the standard error surface per region. The probability of a pixel to be a forest cell increases from light yellow to dark brown areas. Table 4.6 compares the percentage of pixels in each mapping region with a probability higher than 50% of being classified as forest and the percentage of pixels classified as forest in the 2001 NLCD. Area in acreage is also included. Overall, all regions show the percentage of the pixels with a probability higher than 50% of being classified as forest is less than the percentage of forest pixels estimated during NLCD classification process. These differences are found throughout all regions.

Tables 4.7 and 4.8 compare the user's and producer's accuracies for the ten mapping regions in the 2001 NLCD and the probability surface created through indicator kriging. The overall accuracy of the probability surfaces ranges from 23% in region 4 to 98% in region 2. Overall, the producer's accuracy is smaller in the probability surfaces created for all regions except region 4, which had a producer's accuracy of 95.5%. The producer's accuracies for the remaining regions varied from 13% to 75%. All regions however had poor user's

accuracies. We can conclude that, in general, indicator kriging was poor at mapping the forest land cover areas, and worst at correctly classifying areas that area truly of the forest category. Overall, the 2001 NLCD did a better job at classifying forest land-cover across all regions, with the exception of region 2 where the user's accuracy is only 39% in the 2001 NLCD.

Table 4.5. Indicator Kriging Prediction Errors for all Mapping Regions in the United States

Optimized Model					
Region	**Mean Prediction Error**	**Root-Mean-Square Error**	**Mean Standardized**	**Root-Mean Square Standardized**	**Average Standard Error**
1	0.0058	0.34	0.0040	1.35	0.25
2	0.0015	0.20	0.0035	1.00	0.21
3	-0.0062	0.41	-0.014	0.95	0.43
4	0.0016	0.30	0.0069	1.55	0.25
5	0.0017	0.27	0.0059	0.93	0.29
6	-0.0023	0.39	-0.0051	1.18	0.33
7	0.0020	0.42	0.0095	1.21	0.34
8	0.0022	0.40	0.010	1.22	0.32
9	0.0055	0.37	0.016	1.06	0.35
10	0.00031	0.39	0.0079	1.33	0.28

Table 4.6. Percentage of Pixels in Indicator Kriging Surfaces with a Probability Higher than 50% of Forest Classification

Region	**Indicator Kriging**	**Indicator Kriging Acreage**	**2001 NLCD**	**2001 NLCD Acreage**
1	16.28%	17,555 849	41.33%	54,355,771
2	0.59%	851,525	11.45%	29,167,825
3	7.34%	17,800,730	37.98%	95,063,664
4	3.35%	13,461,066	6.33%	26,065,911
5	5.05%	17,015,429	11.40%	32,303,345
6	21.97%	16,839,985	26.39%	36,238,993
7	37.73%	52,392,271	39.40%	72,865,941
8	23.76%	29,128,782	61.92%	87,591,692
9	9.55%	9,632,661	22.94%	26,350,329
10	27.52%	21,256,252	56.72%	52,049,783
Nationwide	10.20%	196,055, 387	25.28%	512,053,254

Table 4.7. Commission and Omission Error in the 2001 NLCD (Wickham et al. 2010)

	User's accuracy	Producer's accuracy
Region 1	86%	85.3%
Region 2	39%	97.7%
Region 3	78.5%	82.6%
Region 4	77.1%	68.3%
Region 5	79.6%	73.1%
Region 6	78.4%	79.8%
Region 7	80.7%	72.2%
Region 8	86%	90.5%
Region 9	96.0%	85.8%
Region 10	89.7%	90.6%
U.S.	87%	88.5%

Table 4.8 Commission and Omission Error in the Indicator Kriging Probability Surface

	User's accuracy	Producer's accuracy	Overall Accuracy
Region 1	19.62%	32.53%	87.65%
Region 2	7.84%	27.69%	97.59%
Region 3	18.68%	75%	77.32%
Region 4	6.59%	95.50%	22.67%
Region 5	31.97%	44.30%	90.05%
Region 6	20.28%	12.79%	85.35%
Region 7	1.76%	7.31%	96.70%
Region 8	2.92%	35.46%	88.88%
Region 9	41.56%	75.14%	80.71%
Region 10	1.46%	51.09%	85.67%
U.S.	27.61%	86.41%	65.84%

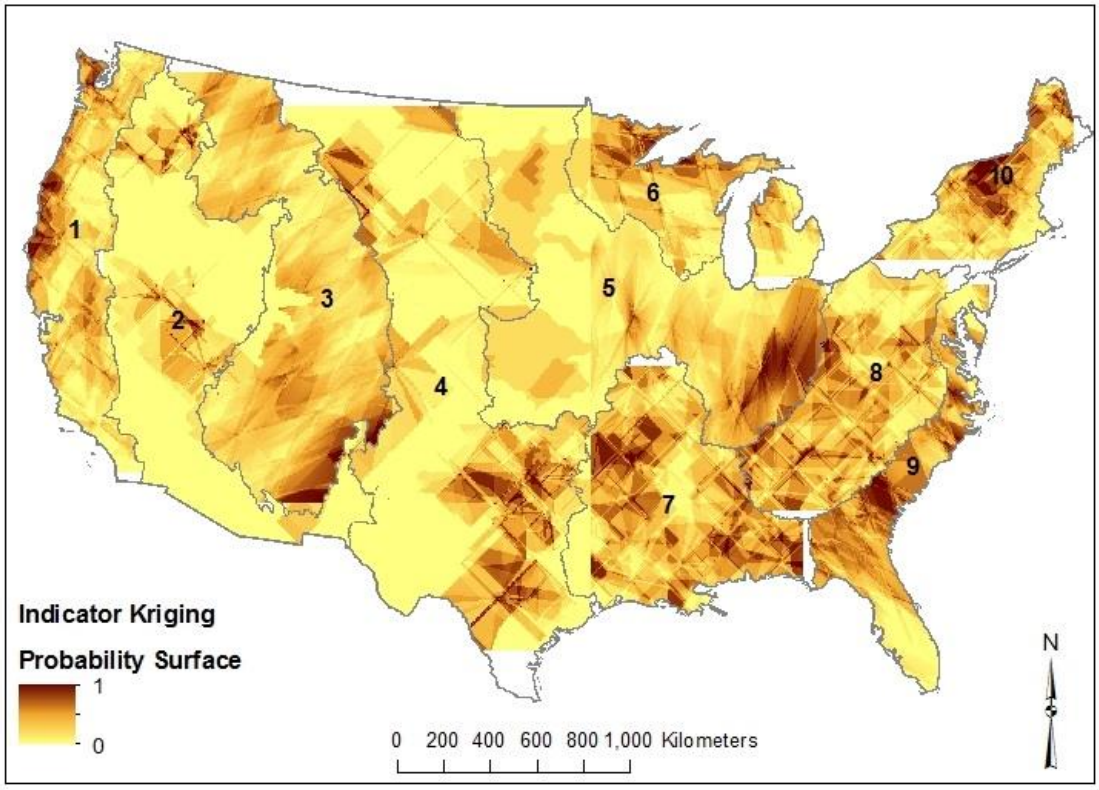

Figure 4.6. Indicator Kriging Probability surface per Region

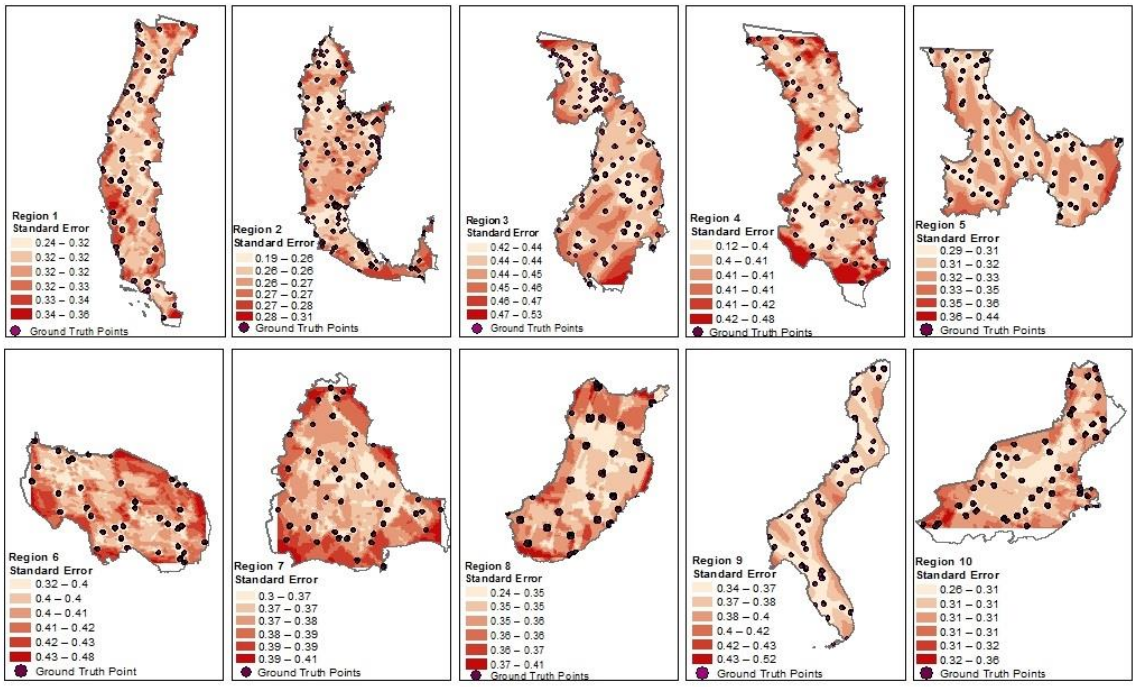

Figure 4.7. Standard Error Per Region using Exponential Model

Region 7 showed the smallest discrepancy between the percentage of pixels in the probability surface with a probability of being classified as forest being greater than 50% and the percentage of pixels classified as forest in the 2001 NLCD. The difference was 28%. The overall accuracy of region 7 in the probability surface is also the second highest near 97% (Table 4.8). The overall distribution of pixels classified as forest coincides with that shown on the NLCD. The user's and producer's accuracies for region 7 however indicate that the probability surface did a poor job a mapping forest land-covers and correctly classifying the pixels as well. Region 7 also has the second smallest producer's accuracy in the 2001 NLCD (Table 4.7) at 7.3%. The majority of the forest land are located along the west and east areas of the region; the length of the Mississippi River along the center of the region is the less forested area.

Other regions in the Midwest (4, 5 and 6) showed fewer discrepancies between the percentage of pixels in the probability surface with a probability of being classified as forest being greater than 50% and the percentage of pixels classified as forest in the 2001 NLCD. The differences were 48%, 47% and 54% respectively. The overall spatial distribution of forest lands also coincided with the

forest land distribution in NLCD. The producer's accuracy in the 2001 NLCD for regions 4, 5 and 6 was 68%, 73% and 80% respectively. The producer's accuracy were 95.5%, 44.3% and 13% respectively. Region 4 had the smallest overall accuracy of all regions (23%) but highest producer's accuracy (95.5%). These results indicate the probability surface does a better job at mapping forest land in region 4 than NLCD does; however the user's accuracy of region 4 in the probability surface is only 7%, and 77% percentage in the 2001 NLCD. As it occurred with all regions, the probability surface did a poor job at classifying pixels as forest that are truly of that category. Most of the forest land in region 4 is located on the east corner along the boundary with region 3 and along the boundary with region 7. The forest lands in region 5 are located mainly along the boundary with region 8; forested areas in region 6 are found throughout the region, especially around the edges of the lakes.

Cross validation results show regions 4, 5, 6, and 7 had large mean prediction errors among all 10 regions. Despite these regions showed the smallest difference between the probability surface and NLCD, the predictions varied more from the true values in these areas than regions with larger differences with NLCD. Region 7 had the largest RMSE of all 10 regions with a RMSE value of 0.42. The RMSSE varies among the five areas. Regions 5 and 3 are the only regions that had RMSSE under 1 and an average standard error greater than the RMSE. This region is overpredicting the variability of the predicted values. Regions 4, 6 and 7 are all underpredicting the variability of the predicted values.

The regions with largest discrepancies between the probability surface and NLCD were regions 2 and 3. The probability surface created through indicator kriging indicated that less than 1% of pixels in region 2 have a probability higher than 50% of being classified as forest. In contrast, the percentage of pixels classified as forest in regions 2 in NLCD is approximately 11%. The difference in region 2 was as much as 97%; the highest among all regions. Tables 4.7 and 4.8 show region 2 was poorly mapped in both the 2001 NLCD and the probability surfaces as user's accuracies were very small at 39% and 8% respectively. The producer's accuracy for this region in the 2001 NLCD however is high at approximately 98%. NLCD shows patches of forest land along the center of the region, and a larger area in the north. These areas coincide with pixels in the probability surface where the probability values are greater than 0.5. However, the probability values mostly range between 0.32 and 0.49. Using a threshold of 0.5 leaves most of this region unforested. Despite region 2 showing the greatest discrepancy with NLCD, cross validation results show (Table 4.5) suggest region 2 has the most accurate results. The mean prediction error for region 2 was closest to zero among all regions which would indicate region 2 created the most unbiased results. Region 2 also had the smallest RMSE, the average error was the closest to its RMSE, and this region also showed the smallest error in the prediction variability. The average standard error for region 2

was also greater than RMSE, however the RMSSE is equal to 1 which indicated the prediction is valid. As stated earlier, the difference between both surfaces was the highest among all regions at 97%.

Results in region 3 also differed a lot with NLCD. The difference between the probability surface and the NLCD was approximately 81%. Similar to region 2, the distribution of forest land seen in the NLCD and that seen on the probability surface is very similar. However, the values of these pixels mostly fall below the 0.5 threshold value used. In contrast to region 2, the prediction error in region 3 was the farthest from zero among all regions, indicating results in this region are the most biased. The RMSE of region 3 is also the second largest from all regions. The variability of the estimations is being overestimated in this region. This is indicated by the average standard error being larger than the RMSE and also the RMSSE being smaller then 1. Figure 4.6 shows forest cover in NLCD for region 3 is found along the northern area of the region and some areas along the western and southwestern boundaries with region 2 as well as eastern boundary with region 4. Figure 4.6 shows the central and northern areas of Region 3 includes pixels where a probability of being forest land is 50% or greater. There are also some patches on the south eastern corner of region 3 where the probability of the pixels for being classified as forest is greater than 50%. As it occurred in region 2, the overall extent of forest land shown in region 3 of the NLCD is larger than that of the probability surface; 38% of pixels area classified as forest by NLCD in comparison with 7% of pixels by the probability surface (Table 4.6).

The percentage of pixels in the probability surface with a probability of being classified as forest being greater than 50% in regions 1, 8, 9 and 10 differ from the percentage of pixels classified as forest in the 2001 NLCD between 63% and 68%. Overall, the distribution of forest land in the probability surface in all regions and the 2001 NLCD is very similar (Figures 4.4 and 4.5). The differences between the results and the NLCD were found in the extent of these forest lands. This is true to all 10 regions.

Region 8 also shows one of the biggest discrepancies between NLCD and the probability surface. The probability surface created through indicator kriging indicates that close to 24% of pixels has a probability higher than 50% of being classified as forest. In contrast, NLCD shows the percentage of pixels classified as forest is approximately 62%. The difference between both surfaces was about 67%. Region 8 has the highest percentage of forest land according to NLCD, and it shows a consistent extent across the region (Figure 4.4). The distribution of forest land in the probability surface coincides with that of the NLCD. The RMSE for this region was 0.40, which was the third highest among all regions (Table 4.5). The variability of the predictions is being underestimating as the RMSS is greater than 1 and the average standard error is less than the RMSE. The variability of the estimations is being underestimated in regions 1, 4, 6, 7, 8, 9

and 10. The average standard error in all these regions is smaller than RMSE. Also, the RMSSE is greater than 1 in all these regions.

The percentage of pixels with a probability greater than 50% to be classified as pixels in region 1 is 16%. The NLCD has 41% of the pixels classified as forest. The difference (68%) in the spatial distribution of forest lands between these two surfaces is seen more clearly in this region than any other region (Figures 4.4 and 4.5). NLCD shows two predominant bands of forest land in region 1, one along the western coast and one right along the boundary with region 2. The distribution of forest land in the probability surface coincides along the western band of the region. However, the extent of forest along the boundary with region 2 is not as extensive as is seen in the NLCD. This region has the second largest mean prediction error with a value of 0.0058. The RMSS is greater than 1 (1.35) and the average standard error is 0.25, smaller than the RMSE of 0.34. The variability of the predictions in this region is being underestimated.

Region 10 has 57% of pixels classified as forest according to the 2001 NLCD. However, the probability surface only estimates 28% of pixels have a probability greater than 50% of being classified as forest. The difference between the two surfaces is approximately 66%. Region 10 has the smallest mean prediction error. NLCD shows region 10 is forested throughout most of its extent. The extent coincides with the extent shown by the indicator kriging surface. The probability value in this region ranges from 0 to1 but the mean value is 0.35.

Figure 4.7 shows the standard error surface in all regions. Similarly to the nationwide distribution of the standard error (Figure 4.5), the standard error per region is smaller around the ground truth points and it increases with distance from the points. The map represents data density.

4.1.2. Indicator Cokriging

Indicator cokriging was used to determine if the variance and residual error estimated from indicator kriging on forest land-cover class are minimized using a secondary variable. The secondary variable used included ground truth points where the 'MAP' field was classified as forest (41, 42, 43) matched the alternate label ('ALT' field) and not the reference label ('REF' field. Like the indicator kriging analysis, indicator cokriging was done using all ground truth points within the conterminous U.S. and also dividing points for each mapping region.

Nationwide Results

Figure 4.8 shows the semivariogram resulting from indicator cokriging using an anisotropic exponential model was used in the three models following results in Section 4.1.1. The lag distance was changed to 1,000 meters.

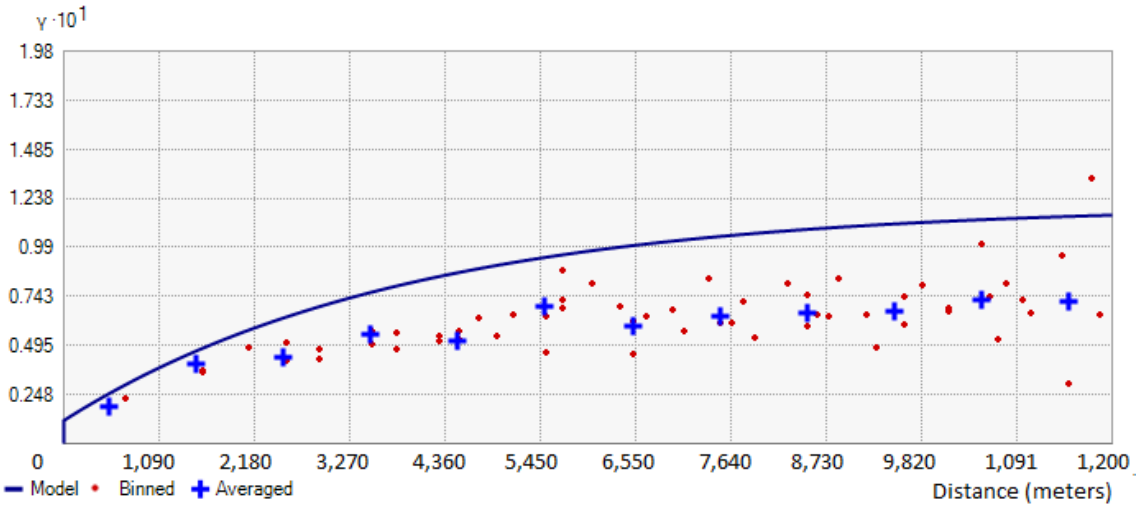

Figure 4.8. Semivariogram for Indicator Cokriging Exponential

As seen in Figure 4.1, the semivariogram created from the cokriging process showed correlation of the data at a short distance. The data shows a constant pattern after reaching the range at approximately 7,000 meters. The range was reached at a shorter distance when a secondary variable was used, indicating the points are spatially correlated at a longer distance with indicator kriging than cokriging.

Table 4.9 includes a comparison of the estimators of errors from both indicator kriging and indicator cokriging scenarios using ground truth points in the conterminous United States. The RMSE was smaller in indicator cokriging results than indicator kriging meaning the cokriging model creates less unbiased results. The mean prediction error and the mean standardized are both larger when using a secondary variable. Even though the RMSE is smaller, the predictions are farther away from the measure values.

The errors in the individual points range from -0.83 to 0.76. The mean error was 0.00138 and the standard deviation of the error was 0.15; the errors are not as dispersed across the country as the error during indicator kriging.

The average standard error of all points was 0.18, which is larger than the RMSE of 0.15, and indicates the model, unlike it was seen in indicator kriging results, is overestimating the variability of the dataset. Additionally, the RMSSE in the cokriging results was less than 1, which is also an indication of values being overestimated. The standard error in all points ranges from 0.12 to 0.26. As it was seen during indicator kriging, the standard error is smaller around data points and it increases with distance (Figure 4.10).

The probability surface created by fitting the optimized anisotropic exponential model for the conterminous U.S. is shown in Figure 4.9. The probability of a pixel to be a forest cell increases from light to yellow to dark brown. Figure 4.10 shows the standard error surface created from the standard error of interpolated values from the indicator kriging.

Table 4.9. Comparison of Indicator Kriging and Indicator Cokriging Prediction Errors for the Conterminous United States.

	Indicator Kriging	**Indicator Cokriging**
Mean Prediction Error	0.00032	0.00138
Root-Mean-Square-Error	0.35	0.15
Mean Standardized	0.0016	0.0081
Root-Mean-Square-Standardized-Error	1.08	0.82
Standard Error	0.33	0.18

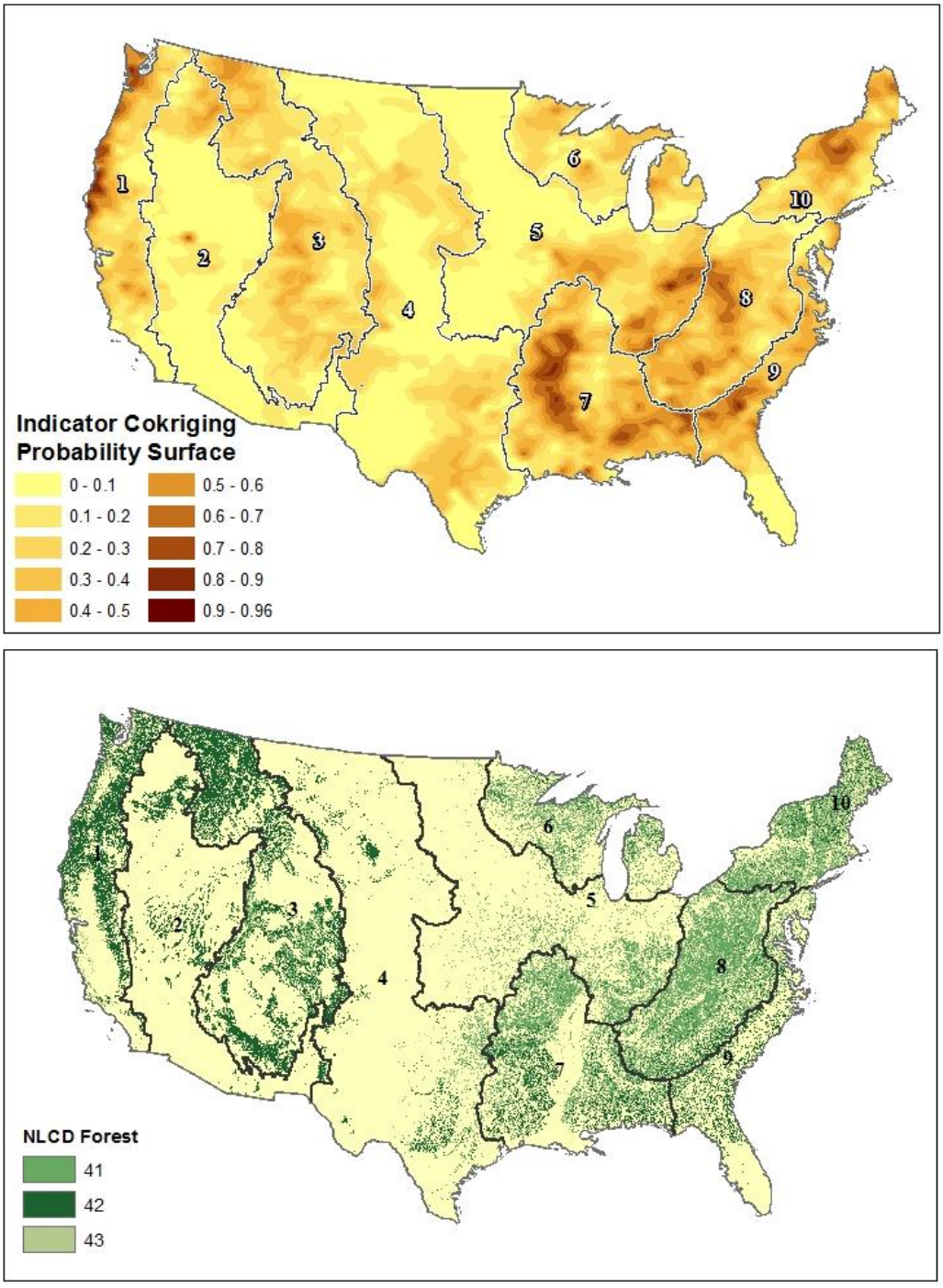

Figure 4.9. Probability surfaces from indicator cokriging.

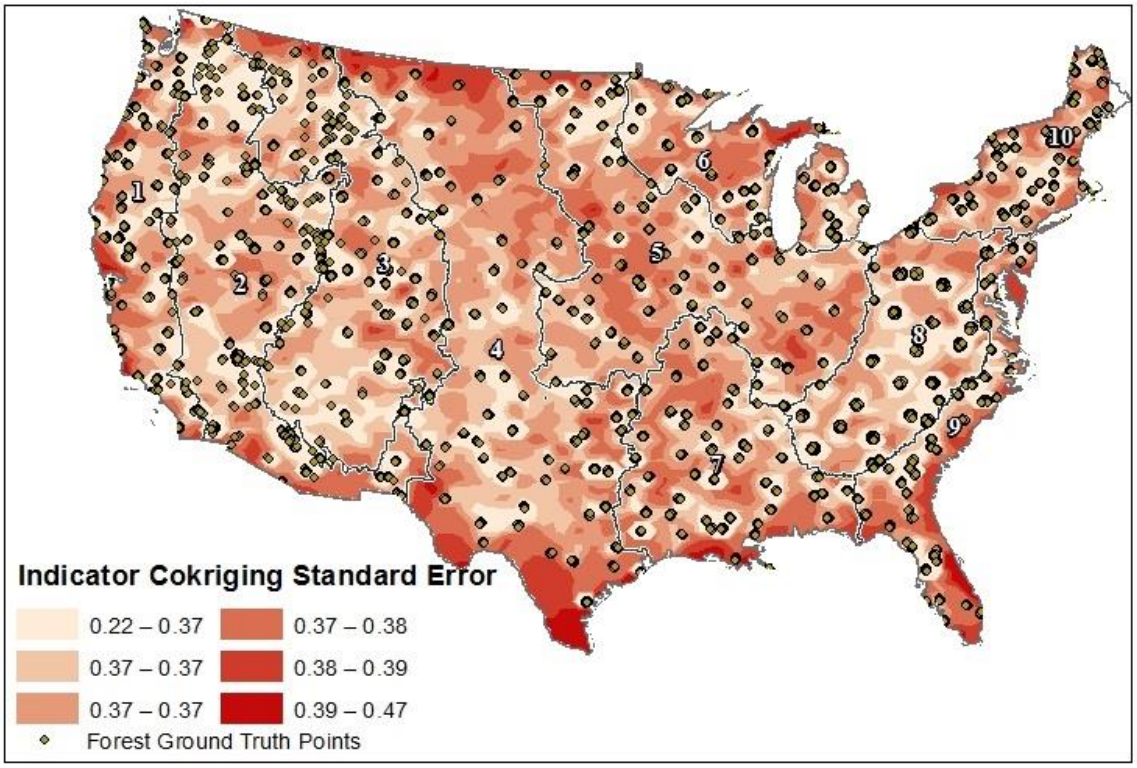

Figure 4.10. Standard error surfaces from indicator cokriging.

Results shown that approximately 8% of the pixels (159 million acres) in the probability surface created for the entire United States have a probability equal or greater than 50% of being classified as forest. This is less than the results indicator kriging which classified approximately 10% of pixels as forest.

The results shown in the probability surface created through cokriging are similar to the results shown in indicator kriging. The overall spatial distribution of pixels classified as forest in NLCD coincides with pixels with higher probability of being classified as forest. The differences are again mostly found in the extent of the distribution of forest land shown in all regions, especially in regions 1, 8 and 10. Overall, the probability of pixels to be forest narrows down in their extent when the secondary variable was added.

Table 4.12 compare the user's and producer's accuracies for the U.S. in the 2001 NLCD and the probability surface created through indicator cokriging. The overall accuracy of the probability surface is 68%, slightly better than the overall accuracy of the probability surface created from indicator kriging (Table 4.8). The producer's accuracy is "good" (81%), but again smaller than the producer's accuracy of the 2001 NLCD (88.5%), and also smaller than the producer's accuracy of indicator kriging. Even though the overall accuracy of the classification by the probability surface is again poor, it is adequate for the purpose of mapping the forest class. However, the user's accuracy of the probability surface is only 22%. This indicates that even though 81% of the forested areas have been correctly identified as forest, only 22% percent of the areas are truly forest land.

Region Results

The same exponential model scenario was repeated diving ground truth points per the 10 mapping regions used by Wickman et al (2010) in the accuracy assessment of the 2001 NLCD. Prediction errors per region are listed in Table 4.10. The probability surfaces created by fitting the optimized anisotropic exponential model mapping region are shown in Figure 4.11. The probability of a pixel to be a forest cell increased from light yellow to dark brown areas. Table 4.11 compares the percentage of pixels in each mapping region with a probability higher than 50% of being classified as forest and the percentage of pixels classified as forest and the percentage of pixels classified as forest in the 2001 NLCD. Area is acreage for each region is also included. Overall, all regions show the percentage of the pixels with a probability higher than 50% of being classified as forest is less than the percentage of forest pixels estimated during the NLCD classification process. These differences are found throughout all regions.

Figure 4.12 shows the standard error map produced from the standard errors of interpolated values. As it was seen in the nationwide results, the standard error is smaller around data points and it increases with distance.

Table 4.12 compares the user's and producer's accuracies for the ten mapping regions in the probability surface created through indicator cokriging. The overall accuracy has improved for all regions except regions 1 and 2 in comparison with the overall accuracy from indicator kriging (Table 4.8). The producer's accuracies have decreased for all regions except regions 1 and 2. Cokriging does not do as good job at mapping the forest class. With the exception of region 4, which had a producer's accuracy of 96%, the accuracy in other regions ranged from 7% to 51%. Furthermore, the user's accuracy for all regions ranged from 2% to 22%. A small percentage is all regions of the areas identifies as forest within the classification are truly of that category. As it occurred when comparing results from indicator kriging and NLCD, the 2001 NLCD did a better job at classifying forest land-cover across all regions.

Table 4.10. Indicator Cokriging Predictions Per Region

Region	Mean	Root-mean-square	Mean Standardized	Root-Mean Square Standardized	Average Standard Error
1	0.00021	0.33	0.0012	1.33	0.25
2	0.0034	0.10	0.02	0.59	0.17
3	0.0028	0.32	0.0054	0.71	0.46
4	0.00089	0.12	0.0044	0.57	0.22
5	0.0021	0.093	0.017	0.77	0.12
6	0.00050	0.19	0.0033	0.72	0.26
7	0.0011	0.17	0.0041	0.63	0.27
8	0.0013	0.13	0.0068	0.66	0.20
9	0.0010	0.18	0.0050	0.85	0.21
10	0.0029	0.081	0.046	1.29	0.061

Table 4.11. Percentage of Pixels in Indicator Cokriging surfaces with a Probability Higher than 50% of Forest Classification

Region	Indicator Cokriging	Indicator Cokriging Acreage	2001 NLCD	2001 NLCD Acreage
1	17.46%	22 428 520	41.33%	54,355,771
2	0.19%	493,717	11.45%	29,167,825
3	10.80%	26,451,890	37.98%	95,063,664
4	3.65%	14,247,849	6.33%	26,065,911
5	4.90%	14,605,164	11.40%	32,303,345
6	19.08%	25,101,706	26.39%	36,238,993
7	36.63%	63,821,884	39.40%	72,865,941
8	24.66%	34,324,420	61.92%	87,591,692
9	13.43%	12,009,322	22.94%	26,350,329
10	27.46%	17,853,117	56.72%	52,049,783
Nationwide	8.27%	158,977,718	25.28%	512,053,254

Table 4.12. Commission and Omission Error in the Indicator CoKriging Probability Surface

	User's accuracy	Producer's accuracy	Overall Accuracy
Region 1	21.68%	34.69%	86.01%
Region 2	11.11%	50.52%	94.17%
Region 3	10.11%	44.59%	85.42%
Region 4	5.20%	95.67%	22.58%
Region 5	6.62%	16.45%	96.79%
Region 6	10.19%	6.67%	94.87%
Region 7	1.72%	6.86%	96.93%
Region 8	2.62%	8.74%	96.64%
Region 9	6.06%	20.51%	94.34%
Region 10	12.84%	18.55%	91.35%
U.S.	21.72%	80.92%	68.03%

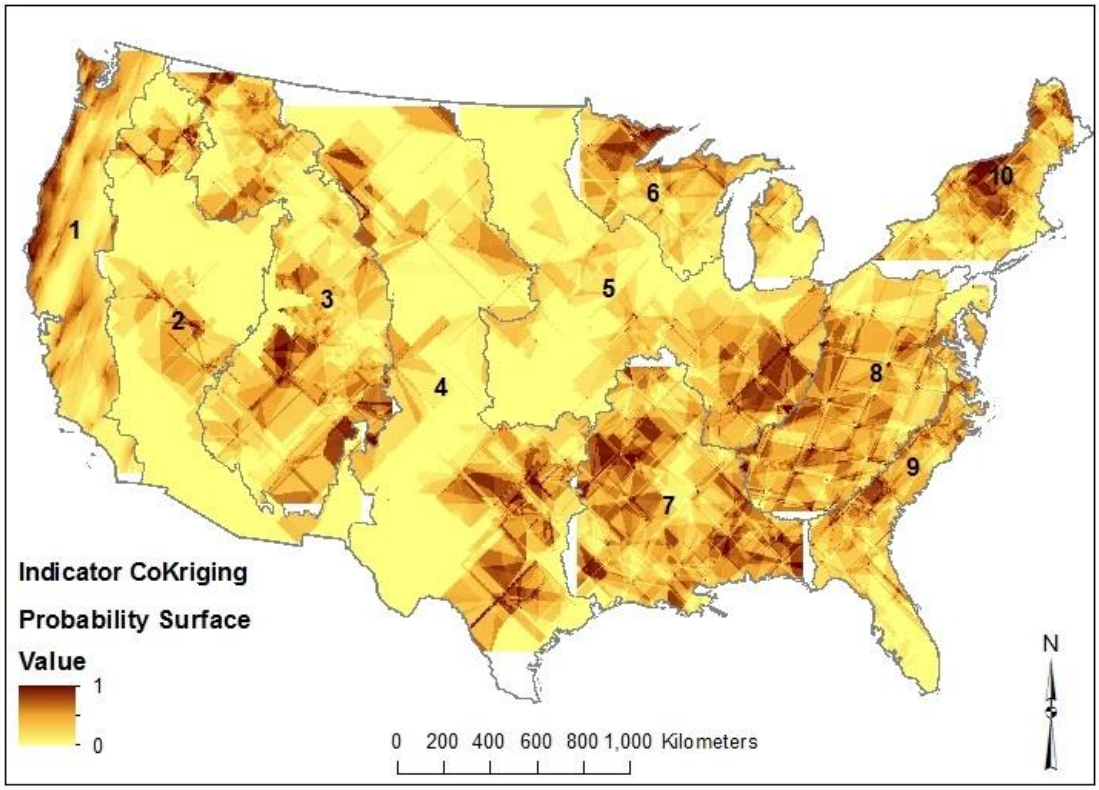

Figure 4.11. Probability Surfaces per Region from Indicator Indicator Cokriging

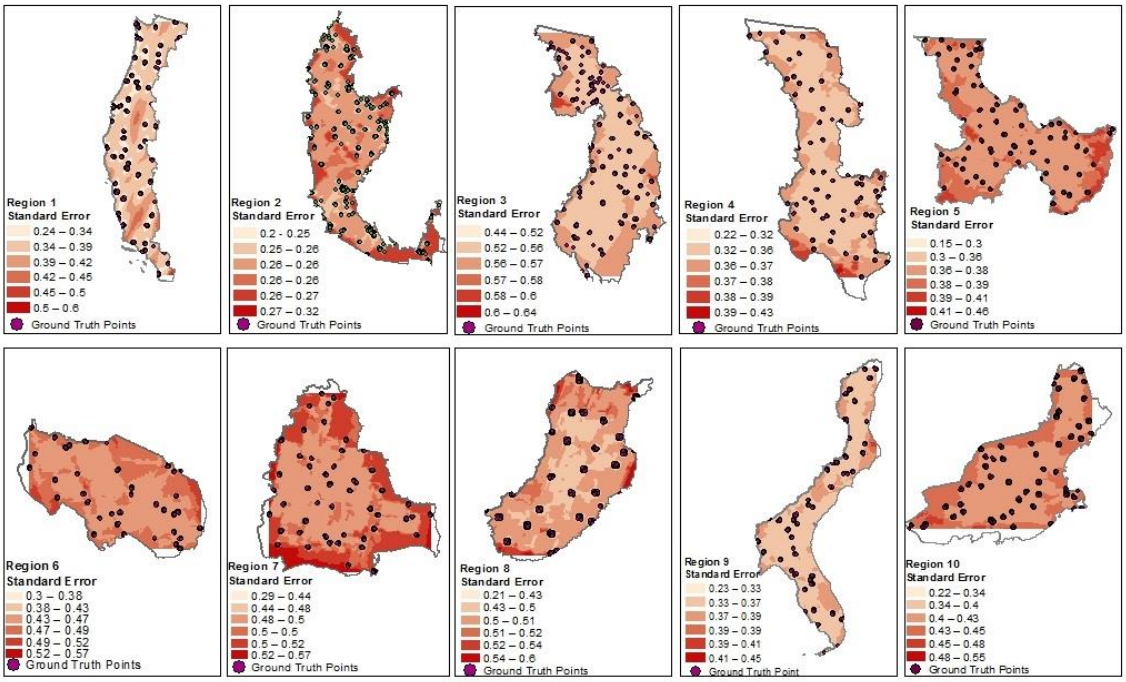

Figure 4.12. Standard Error of Surfaces per Region from Indicator Cokriging

Comparing the spatial distribution and extent between the 2001 NLCD and probability surface created through cokriging show similarities to the results found in the indicator kriging analysis. Region 7 had the smallest discrepancy between the percentage of pixels in the probability surface with a value greater than 0.5 and the percentage of pixels classified as forest in the 2001 NLCD. The difference however between the two surfaces decreased from 28% to 12% when comparing the indicator kriging and cokriging results. The spatial distribution of forest land shown in the probability surface is still similar to that in NLCD. Areas with higher probabilities of being classified as forest are found on the north-west and south-east parts of the region while the probability values in the center of the region are range from 0 to 0.40. All indicator cokriging predictions for region 7 are smaller in than the indicator kriging predictors. The RMSE decreased from 0.42 to 0.17; values predicted during cokriging results are closer to the measured values than those in indicator kriging results (Table 4.8). The mean standardized predictor was reduced to 0.0041, indicating the results are more unbiased. Also, the RMSS is now less than 1 and the average standard error is less than the RMSE; the variability of the predictions is being overestimated instead of underestimated.

Similarly to indicator kriging results, regions 2 and 3 had the largest discrepancies between the percentage of pixels in the probability surface with a

value greater than 0.5 and the percentage of pixels classified as forest in the 2001 NLCD. However, results are different for each region. The probability surface created for region 2 through indicator cokriging only classified 0.1% of the region area as having a probability greater than 50% of being classified as forest. In contrast, the percentage of pixels classified as forest in region 2 by indicator kriging and NLCD surfaces were 0.6% and 11% respectively. The difference between NLCD and cokriging increased to 98%. The probability surface showed the areas with higher probability of being classified as forest are found in two patches; one in the north and one in the center of the region. This coincides with the spatial distribution of forest land shown by NLCD in this region. However, the highest probability value in the entire region is 0.50. The majority of the pixels fall below the 0.5 threshold used in the analysis.

The difference between region 3 and NCLD on the other hand decreased from 81% to 72%. The spatial distribution of the forest land matches that of NLCD with patches of areas with high probability values across the region mixed with areas of low probability values. Some of the cross validation results (Table 4.8) for both regions differ from the indicator kriging results. The mean prediction error for region 2 was closest to zero among all regions in indicator kriging analysis; however the mean prediction error in cokriging results for this region was farther away from zero among all regions. The RMSE for region 2 was still the smallest of all regions and the average error remained still greater than the RMSE. The RMSSE however decreased from 1 to 0.59. While the RMSSE for region 2 in indicator kriging implied this region had the most valid predictions, the predictions now are being overestimated. The RMSE for region 3 is still the second largest among all regions but the value decreased from 0.41 to 0.32. The variability of the estimations is also being overestimated by indicator cokriging since the average standard error is greater than the RMSE (0.46 vs. 0.32), and the RMSSE is less than one.

The difference between the percentages of pixels classified as forest in NLCD in regions 1, 4, 6, 8 and 9 the percentage of pixels with probability greater than 50% in the cokriging surface also decreased from the results seen in the indicator kriging analysis. Cokriging results estimated larger areas of these regions having greater probability of being classified as forest. In region 4, for example, approximately 3.5% of region 4 had a probability greater than 50% of being forest while 6% of region 4 in NLCD was classified as forest. The difference between the two surfaces is 45%. The difference between NLCD classification and indicator kriging results was 48%. As in other regions, the spatial distribution of these forest lands coincides with NLCD. The majority of forested areas are located near the boundary with region 7. In regions 1, 6, 8 and 9, the differences between the percentage of pixels classified as forest in NLCD and pixels in the cokriging surface with probability greater than 50% were 59%, 31%, 61% and 54% respectively. Forested areas in region 1 are mainly found in a band extending across the western side of the region and another one along

the south-east. Forested areas in regions 6 and 8 are scattered throughout the region. Forest lands in region 9 are found mainly in the northern half of the region. Cross validation results (Table 4.8) show corkriging reduced RMSE for regions 4, 6, 8 and 9; predicted values are closer to true values. Unlike the results from indicator kriging analysis, the RMSSE in these regions resulting from corkriging was less than 1 indicating the variability of the predictions is being overestimated. Region 9 had a RMSSE closest to 1 among all regions. The average standard error in all these regions is also greater than the RMSE. Cross validation values for region 1 did not change much between indicator kriging and cokriging results, with the exception of the mean prediction error which decreased from 0.0058 to 0.00021. Region 1 has the smallest mean prediction error from all regions. The RMSE changed from 0.34 to 0.33 and the RMSSE value changed from 1.35 to 1.33. The predictions are still being underestimated in this region. The average standard error remained the same at 0.25.

Cokriging results for regions 5 and 10 resulted in smaller percentage of the pixels as having a probability greater than 50% of being classified as forest compared to the results in the indicator kriging analysis. The difference between the percentages of pixels classified as forest in NLCD and pixels with probability values greater than 0.5 increased with cokriging. Pixels with probability values greater than 0.5 in region 5 are mostly located near the boundary with region 8. Region 10 is mostly forested throughout the region but probability values increase in the center of the region. Both regions have the smallest RMSE and average standard error of all regions (Table 4.8). The predictions for region 5 were overestimated (RMSSE was smaller than 1) while predictions in region 10 were underestimated (RMSSE was greater than 1).

Figure 4.12 shows the standard error surface in all regions. Similarly to the distribution of the standard error in indicator kriging (Figure 4.7), the standard error per region is smaller around the ground truth points and it increases with distance from the points. The map represents data density.

4.1.3. Monte Carlo Simulation

Monte Carlo simulation was used to generate multiple realizations of the forest land cover class from the indicator and cokriging surfaces created in Sections 4.1.1 and 4.1.2. These probability surfaces provide a guidance of the spatial distribution and percentage of pixels that may be classified as forest. The Monte Carlo simulation process used included information of forest and non-forest areas. The process followed is shown in the workflow outlined in Figure 3.6. Model Builder in ArcGIS 10.0 was used to create the Monte Carlo methodology. The model was run 1,000 times. The results were surfaces with pixels classified as either forest or non-forest. Figure 4.13 shows an example of the surfaces created.

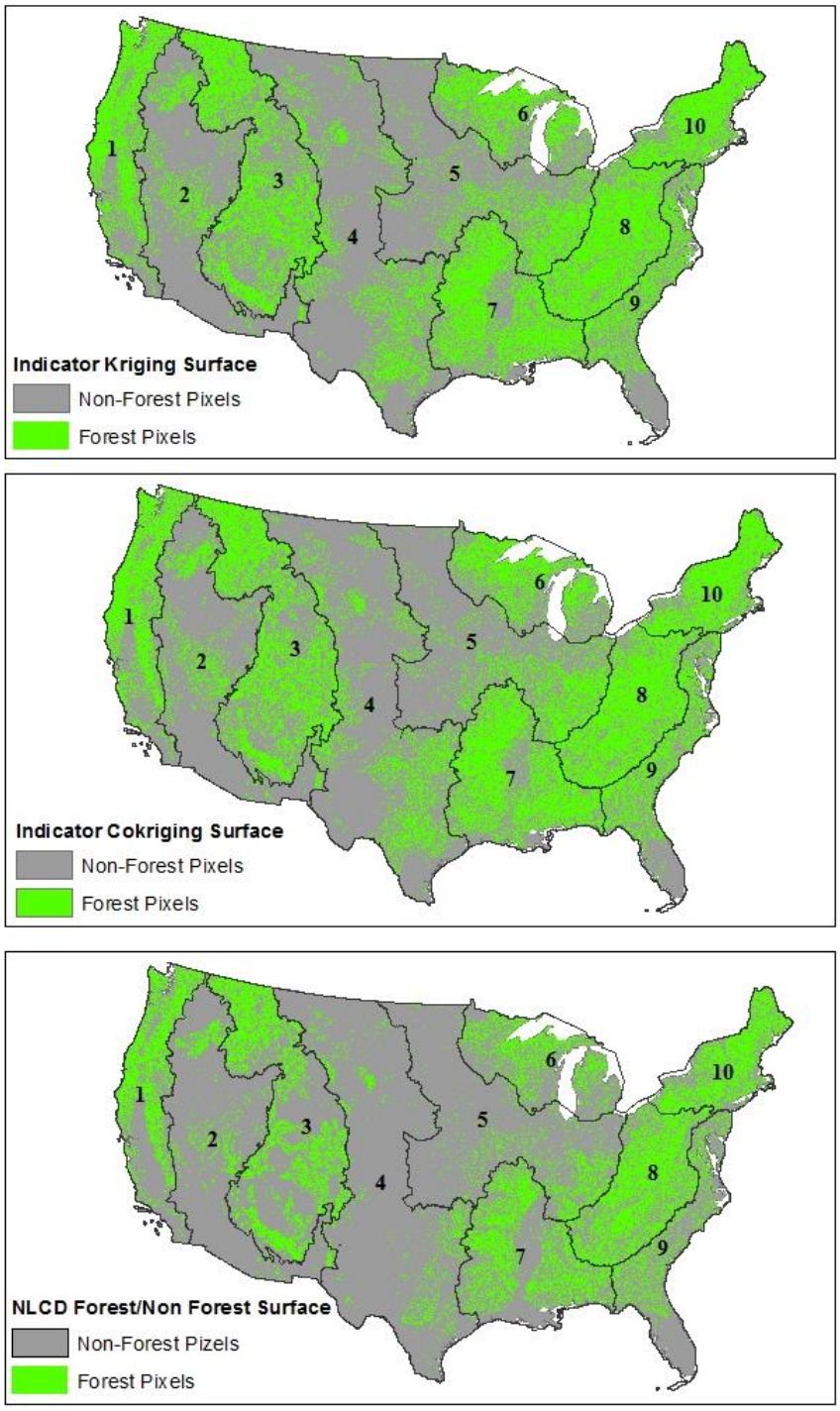

Figure 4.13. Monte Carlo Surfaces from indicator kriging (top) and indicator cokriging (middle) and NLCD Forest/Non-Forest surface (bottom).

A common method to display result from Monte Carlo simulation is to show frequency distribution in histogram format. However, results of the simulation process showed less than a 1% difference between the distribution of the number of forest and non-forest cells from surfaces created through Monte Carlo simulation.

An appropriate method for evaluating probabilistic maps is to check the consistency of the distribution of classes between kriged and sample data (Zhang and Goodchild 2002). Table 4.13 compares the percentage of pixels classified as forest from land-cover realizations created through Monte Carlo simulation and 2001 NLCD. Results are shown in nationwide and per region extents.

Table 4.13. Comparison of per region percentage of forest pixels in Monte Carlo simulation results from indicator kriging, cokriging surfaces and 2001 NLCD

Region	Indicator Kriging	Indicator Cokriging	2001 NLCD
1	56.17%	57.19%	41.33%
2	18.38%	18.53%	11.45%
3	53.01%	51.74%	37.98%
4	18.62%	17.52%	6.33%
5	23.22%	23.74%	11.40%
6	49.56%	45.82%	26.39%
7	61.00%	61.19%	39.40%
8	74.64%	73.95%	61.92%
9	71.28%	46.45%	22.94%
10	73.97%	73.24%	56.72%
Nationwide	39.32%	38.97%	25.28%

Indicator kriging and indicator cokriging created probabilistic surfaces where approximately 39% of pixels in the conterminous U.S. were classified as forest. The percentage of pixels classified as forest in the 2001 NLCD is 26%. Indicator kriging maps a higher percentage of forest lands. Results are similar when evaluating results per region. Indicator kriging and cokriging surfaces resulted in a higher percentage of pixels being classified as forest in all regions. Using cokriging classes did not change the simulated percentage of forest cells for either region. Overall, realizations of forest land cover created through Monte Carlo simulation using cokriging surfaces only changed the percentage of pixels classified in region 9.

4.2. Discussion

Chapter 2 outlined probabilistic and fuzzy methods that have been used in the literature to describe the spatial distribution of uncertainty on a thematic map. This related to two research questions of this thesis. The first question was to use indicator kriging as a method to map the spatial accuracy of the forest land-cover class in the 2001 NLCD. The second question was to determine if using secondary information through indicator cokriging would improve the results from the indicator kriging approach.

The overall results of indicator kriging analysis show that indicator kriging can provide helpful information about the spatial distribution of the forest land-cover categories but the percentage of pixels that are shown to have a high probability of being classified as forest was underestimated when comparing to the forest land-cover classification seen in the 2001 NLCD. The difference between both surfaces was 62% when data was analyzed at the nation-level and ranged from 12% to 98% when data was analyzed at the region-level. Evaluating the distribution of the forest land-cover class in the 2001 NLCD shows region boundaries are clearly marked (Figure 4.4). For example, the boundary between region 2 and regions 1 and 3 show a very distinct change between forest and non-forest land. Another example can be found along the boundary of regions 3 and 4. This is an indication of misclassification of the forest land-cover classes in the 2001 NLCD along region boundaries. Probability surfaces are however able to interpolate forest land across region boundaries where points across the entire country are used.

Differences between the probability surfaces and the classification shown in the NLCD may be due to several factors such as the sampling design and density of points of the reference dataset, the fragmentation of the landscape in each region or the decisions made during the statistical analysis of the data.

Two factors affecting the results can be derived from the statistical analysis. One of these factors is the threshold value used to determine the percentage of pixels that have a high probability of being classified as forest. A threshold of 50% was used in this thesis. This threshold resulted in approximately 10% of the pixels in the conterminous U.S. to be classified as forest. Approximately 26% of pixels in the NLCD were classified as forest. The results suggest that using a threshold of 50% forest/non-forest may underestimate the percentage of pixels that have a probability of being classified as forest. Changing the threshold used in the analysis would not change statistical results outlined in Table 4.5, but it would have an effect on the percentage of pixels that would be classified as forest. Additional analysis could be done to compare results using different threshold values and determine a more appropriate threshold to use.

A second factor influencing the results that was derived from the statistical analysis was the parameters used in the kriging process. Choosing the correct parameters in the Geostatistical Analysis process can be considered an “art.” Trial and error and an understanding of the dataset are necessary to choose the appropriate parameters. Additional work and research on what parameters should be used during the kriging process may improve the results. In addition, there are often question to whether geostatistical analysis tools available through GIS software are able to provide sound results. GIS however is being used more and more in the research community and these tools continue to be tested an improved.

The analysis of the results also raises questions about whether the characteristics of the reference data developed for the 2001 NLCD used in this thesis are appropriate for the type of analysis attempted in this study. Questions can be asked about the reference data such as whether changing the sampling design of the reference points would affect the results, or whether the complexity of the land-cover patterns in different regions are properly captured by the density of the reference points.

Stehman (2009) and Stehman and Czaplewski (1998) described basic criteria that should be considered for constructing a sampling design that can be implemented for map accuracy assessment. They stated that choosing the correct sampling design is based on the objectives of the assessment. As described in section 3.2, the sampling design of the 2001 NLCD accuracy assessment used a two-stage cluster sample with three levels of stratification. The primary objective of this sampling design was to assess per-class accuracy of land-cover classes using an approach that would result in the creation of an error matrix reporting overall, user’s and producer’s accuracies (Stehman et al. 2008). The U.S. was stratified into ten mapping regions to ensure that the sample size allocated to “rare” land cover classes would be large enough to produce precise estimates of accuracy. The goal of the second stratification was to spread the samples geographically within each region while keeping the cost down by clustering the points (Wickham et al. 2010). Cluster sampling was used because of the cost and convenience of spatially constraining the sample pixels to a limited number of clusters (Wickham et al. 2013).

When looking at the land-cover composition of the ten regions according to the 2001 NLCD, regions 1, 2, 4 and 5 have the smallest percentage of forest land among all regions. Statistically, both indicator kriging and cokriging results (Tables 4.5 and 4.8) showed these regions had the smallest RMSE among all regions. Region 2 had actually the smallest average standard error and a RMSS of 1, suggesting this region had the most accurate results. This would indicate that the sampling design was successful at minimizing errors in regions where forest land-cover is a “rare” class, therefore meeting the primary objective of the accuracy assessment for the NLCD. However, the objective of this thesis was to

map the spatial distribution of the accuracy, not just creating a confusion matrix of user's and producer's accuracy. This objective was not a priority of the accuracy assessment of the NLCD. The clustering of points in the two-stage cluster approach may have influenced the results of the interpolation of the data across all regions and reduced the extent of the pixels classified as forest in the probability surfaces. This effect was seen in all regions as the percentage of pixels with probability higher than 0.5 was smaller in all regions in comparison with the percentage of pixels classified as forest in the NLCD. This is also more evident when reference points were analyzed across the entire U.S. as opposed to per-region level. The number and density of points used to analyze the spatial distribution of the accuracy in the forest land-cover category is not sufficient to provide accurate results in large areas.

Results slightly improved when analyzing data at the region-level but the clustering on reference points still influenced the results. Regions along the west coast (1, 2 and 3) showed the greatest differences between the mapped forest land in the 2001 NLCD and the probability surfaces created through indicator kriging. These differences were described in the Results section. The percentage of pixels in forest category in NLCD and the probability surface in region 2, for example, differed by 98%. This difference was found especially in the center of the region. The NLCD shows this area as being classified mainly as shrub, not forest. Only few reference points in the forest land cover class within region 2 had an agreement between the map and reference labels and these points were spread across the entire region. Points are too far away from each other for the results to show pixels with high probability of becoming forest land at a large distance from the location of each point. Using additional ground truth points through cokriging did not improve the extent of forest land in this region (Table 4.8) and it also did not change the percentage of pixels classified as forest land from the realizations created through Monte Carlo simulations (Table 4.14).

In general, reference points where there was an agreement between the map and reference label were more uniformly spread across the entire area in regions along the east coast than on the west coast. Both indicator kriging and indicator cokriging results (Tables 4.5 and 4.8) show regions along the eastern United States (regions 7, 8 and 10) have higher percentage of pixels classified as forest land than regions in the mid-west. This correlates with the proportion of forest area in these regions in comparison with the rest of the country. These overall findings also coincide with the evaluation of 2001 NLCD accuracy assessment by Wickham et al. (2010). They stated that "deciduous forest user's accuracy decreased from east to west." The distribution of pixels with high probability of being classified as forest cell and forest areas the 2001 NLCD is more similar in these regions than other regions. The extent of the forest however still differs due to not having sufficient reference points and these points being clustered in small geographic areas. Among all regions along the eastern U.S, regions 7 and 10 show the smallest difference between the percentage of pixels in the forest

classification in NLCD and the percentage in the probability surface. Statistically, regions 7, 8 and 10 also have the highest RMSE of all regions during the indicator kriging process. However, adding additional information through indicator cokriging improved the RMSE in regions 7, 8 and 10.

The effect of point clustering in the kriging interpolation process when looking all points in the conterminous U.S. can be seen in output parameters used in both indicator kriging (Figure 4.3) and indicator cokriging (Figure 4.8). In both kriging analyses, the semivariogram showed the spatial autocorrelation between points occurred at a short distance. The range calculated by the indicator kriging was 34 km, and 7 km for indicator cokriging. The nearest distance between two sampling points in the conterminous U.S is approximately 1.5 km. The distance between points within a cluster ranges from 1.5 km to 42 km. The range is at times shorter than the distance between two points farthest away within a cluster. This indicates that the kriging process only uses information from all or most points within a group of points to interpolate the information. Information from points beyond the clusters is not being considered in the kriging process.

Semivariogram parameters for region-level analysis showed the same effect. The range used in the kriging process for all regions varied from as little as 1.5 km in regions 10 and 8 to 827 km in region 3. Regions where the range was smaller tend to have sampling points that are clustered within 1.5 km to 20 km apart. Distances between clusters are greater than the range distances; the spatial autocorrelation between points is shorter than the distances between clusters and the kriging process does not have enough information to correctly interpolate between those points.

Overall, the clustering of samples causes the variogram and the kriging results to be more representative of a particular area around the reference points. The use of the results for the entire area of the U.S. or the entire area of a region is questionable.

Two-stage cluster sampling provides a more cost-effective alternative, but this advantage occurs at the expense of being able to assess the land cover composition and landscape pattern. Congalton (1988a) stated that the complexity of a given environment dictates the appropriate sampling scheme to use. This leads to another objective to consider when choosing reference data for accuracy assessment. This objective is to assess the accuracy of landscape features and patterns (Stehman et al. 2008). This objective was not a priority in the selection of the sampling design of the accuracy assessment of the 2001 NLCD. Further analysis of the classification error should incorporate information such as landscape characteristics in the analysis (Smith et al. 2003). How do landscape characteristics affect the spatial accuracy assessment of the forest land-cover classes?

Patch size and land-cover heterogeneity are two examples of landscape characteristics that are thought to affect classification error (Campbell 1996). Both influence the classification error by introducing pixel misclassification when the land-cover map and reference data sets are misregistered causing confusion as to the land cover actually present at a location. Smith et al (2002), for example, used the 1992 NLCD to look at establishing land-cover specific relationships between classification accuracy and patch size and heterogeneity. They found that accuracy decreases as land cover heterogeneity increases and patch size decreases.

This thesis did not include a detailed study of the landscape fragmentation of the mapping regions in the NLCD. However, general calculations of the percentages composition of each land cover class within each region showed that regions 2, 5 and 8 have one predominant land-cover type among other types; shrub/scrub, crops and forest respectively. Over 60% of the area in these regions consists of these categories. Regions 2 and 5 had smallest RMSE in the indicator kriging analysis, while region 8 had the third largest RMSE. The differences between the probability surfaces in these regions and the forest classification in NLCD varies as region 2 had the largest difference, while region 5 had one of the smallest differences.

The same can be said of other region. Forest land-cover categories was the predominant classification in regions 7, 10, 6 and 1 with forest land covering less than 40% of the area in any of those regions. The most common forest land cover category in region 9 is wetlands (25% of the area). Shrub/scrub makes up about 42% of region 3 while approximately 42% of the area of region 4 consists of grassland. These calculations do not provide enough information to determine how fragmented the landscape is in any of these regions.

Chapter 5 Conclusions and Recommendations

The goal of this study was to describe a method of mapping spatial accuracy of thematic land cover map using indicator kriging and indicator cokriging. The study focused on the forest land-cover category outlined by the 2001 NLCD. The overall results indicate that indicator kriging can provide helpful information about the spatial distribution of the forest land-cover categories but the percentage of pixels that are shown to have a high probability of being classified as forest was underestimate. This percentage however could vary based on different factors.

First, the indicator kriging process only used information from one variable. The indicator cokriging used as a method to add information from secondary land-cover classes did not change the overall results. One reason may be the distribution of the ground truth points. The sampling method used by the accuracy assessment of the 2001 NLCD was a stratified/cluster method. Points in different categories tended to be clustered in a geographic area which may have influenced the results of the interpolation of the data across large regions. A systematic spatial sample collected throughout the region would ensure a regular spacing of points in the area. However, the cost involved in gathering information using such sampling method in the entire country would be very expensive and very difficult to accomplish. Cluster sampling does provide an advantage of lower collection costs. Using ground truth data from a different thematic dataset with a different sampling distribution may provide missing information in areas where sampling was not done through NLCD. Two questions could be asked about the sampling design: how many samples are required to obtain a reliable estimate of the classification accuracy and where in the study should the samples be acquired to obtain an unbiased estimate of the accuracy?

A second factor in the results was the threshold value used to determine the percentage of pixels that have a high probability of being classified as forest. A threshold of 50% was used. This threshold may be too low. Additional research should be done to determine a more appropriate threshold.

A third factor is the parameters used in the kriging process. Choosing the correct parameters in the kriging process can be considered more an “art.” Trial and error and an understanding of the dataset are necessary to choose the appropriate parameters. Additional work and research on what parameters should be used during the kriging process may improve the results. Also, there are often question to whether geostatistical analysis tools available through GIS are able to provide sound results. GIS however is being used more and more in the research community and these tools continue to be tested an improved.

Even though there are some shortcomings to the method, the overall results, however, do provide an understanding of the spatial distribution of the

forest land-cover dataset and it can provide some valuable information that is not available through a confusion matrix.

The results found in this study show indicator kriging can be used as a tool to map the probability distribution of any land cover class. However, the current sampling distribution of reference points does not provide sufficient information for the kriging interpolation process to create an accurate surface. Results could be used to provide an initial understanding of the spatial distribution of the uncertainty in the 2001 NLCD but more analysis is needed.

Future area of work remains. Sampling assumes all land-cover classes occur with equal frequency in all regions and that classification error occurs with equal frequency across all classes. In reality, few land-cover classes tend to dominate within a region.

LIST OF REFERENCES

Aerts, J. C. J. H., M. F. Goodchild & G. B. M. Heuvelink (2003) Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems. *Transactions in GIS,* 7**,** 211-230.
Anderson, J. R., E. E. Hardy, J. T. Roach & R. E. Witmer. 1972. A Land Use and Land Cover Classification System for Use with Remote Sensing. 28. U.S. Geological Survey.
Aronoff, S. (1989) Geographic information systems: a management perspective.
Atkinson, P. M. & N. J. Tate (2000) Spatial scale problems and geostatistical solutions: a review. *The Professional Geographer,* 52**,** 607-623.
Bierkens, M. F. P. & P. A. Burrough (1993) The Indicator Approach to Categorical Soil Data. I. Theory. *Journal of Soil Science,* 44**,** 361-368.
Boucher, A. & P. C. Kyriakidis (2006) Super-resolution land cover mapping with indicator geostatistics. *Remote Sensing of Environment,* 104**,** 264-282.
Burrough, P. A. & R. McDonnell. 1998. *Principles of geographical information systems*. Oxford university press Oxford.
Burt, J. E., G. M. Barber & D. L. Rigby. 2009. *Elementary Statistics for Geographers*. New York, NY: The Guildfor Press.
Campbell, J. (1981) Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing,* 47**,** 355-363.
Campbell, J. B. 1983. *Mapping the Land: Aerial Imagery for Land Use Information*. Washington, DC: Association of American Geographers.
---. 1987. *Introduction to Remore Sensing*. New York, New York: Guilford Press.
---. 1996. *Introduction to Remote Sensing* New York and London: The Guilford Press.
Cao, G., P. C. Kyriakidis & M. F. Goodchild (2011) Combining spatial transition probabilities for stochastic simulation of categorical fields. *International Journal of Geographical Information Science,* 25**,** 1773-1791.
Carlotto, M. J. (2009) Effect of errors in ground truth on classification accuracy. *International Journal of Remote Sensing,* 30**,** 4831-4849.
Chrisman, N. (1991) The error component in spatial data. *Geographical information systems,* 1**,** 165-174.
Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and psychological measurement,* 20**,** 37-46.
Congalton, R. G. (1988a) A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing,* 54**,** 593-600.
Congalton, R. G. (1988b) Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing,* 54**,** 587-592.
--- (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment,* 37**,** 35-46.

Congalton, R. G. & K. Green. 2008. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.
ESRI. 2003. ArcGIS 9. Using ArcGIS Geostatistical Analyst: ONLINE. Accessed from: http://forums.esri.com/Thread.asp?c=93&f=1727&t=257926.
---. 2010. ArcGIS Resource Center Desktop 10. ONLINE: . Accessed from: http://help.arcgis.com/en/arcgisdesktop/10.0/help/.
Fisher, P. F. (1994) Visualization of the Reliability in Classifiied Remotely Sensed Data. . *Photogrammetric Engineering and Remote Sensing (PE&RS),* 60**,** 905-910.
Foody, G. (1996) Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing,* 17**,** 1317-1340.
Foody, G. M. (2002) Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* 80**,** 185-201.
--- (2004) Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing,* 70**,** 627-633.
Fuller, R. M., B. K. Wyatt & C. J. Barr (1998) Countryside survey from ground and space: different perspectives, complementary results. *Journal of Environmental Management,* 54**,** 101-126.
Goodchild, M. 1989. Modelling errors in objects and fields. In *Accuracy in Cartography,* eds. M. Goodchild & S. Gopal, 107-113. London: Taylor & Francis.
Goodchild, M., M. Egenhofer, K. Kemp, D. Mark & E. Sheppard (1999) Introduction to the Varenius Project. *International Journal of Geographical Information Science,* 13**,** 731-745.
Goodchild, M. F. (1993) The state of GIS for environmental problem-solving. *Environmental modeling with GIS***,** 8-15.
--- (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues. *Journal of Vegetation Science,* 5**,** 615-626.
Goodchild, M. F. & J.-C. Muller. 1991. *Issues of quality and uncertainty*.
Hession, S., A. M. Shortridge & N. M. Torbick. 2006. Categorical models for spatial data uncertainty. In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.
Heuvelink, G. B. M. 1998. *Error Propagation in Environmental Modelling with GIS*. Padstow, UK: T.J. International.
Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel & J. Wickham (2007) Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing,* 73**,** 337-341.
Homer, C., C. Huang, L. Yang, B. Wylie & M. Coan (2004) Development of a 2001 National Landcover Database for the United States.pdf. *Photogrammetric Engineering & Remote Sensing,* 70**,** 820-840.

Hutchinson, C. F. (1982) Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing,* 48**,** 123-130.

Isaaks, E. H. & R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press.

Jensen, J. R. (1979) Spectral and Textural Features to Classify Elusive Land Cover at the Urban Fringe. *The Professional Geographer,* 31**,** 400-409.

Journel, A. G. (1986) Constrained Interpolation and Qualitative Information - the Soft Kriging Approach. *Mathematical Geology,* 18**,** 269-305.

Kanevski, M. & M. Maignan. 2004. *Analysis and Modeling of Spatial Environmental Data*. New York, NY: Marcel Dekker, Inc.

Koukoulas, S. & G. A. Blackburn (2001) Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. *Photogrammetric Engineering and Remote Sensing,* 67**,** 499-510.

Krivoruchko, K. & C. A. G. Crawford. 2005. Assessing the Uncertainty Resulting from Geoprocessing Operations. In *GIS, Spatial Analysis, and Modeling,* eds. D. J. Maguire, M. Batty & M. F. Goodchild. Redlands, California: ESRI Press.

Kyriakidis, P. C. 2001. Geostatistical Models of Uncertainty for Spatial Data. In *Spatial uncertainty in ecology : implications for remote sensing and GIS applications,* eds. C. T. Hunsaker, M. A. Friedl, M. F. Goodchild & T. J. Case, 175-213. New York: Springer.

Kyriakidis, P. C. & J. L. Dungan (2001) A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environmental and Ecological Statistics,* 8**,** 311-330.

McGwire, K. C. & P. F. Fisher. 2001. Spatially Variable Thematic Accuracy: Beyond the Confusion Matrix. In *Spatial Uncertainty in Ecology,* eds. C. T. Hunsaker, M. A. Friedl, M. F. Goodchild & T. J. Case, 308-329. New York: Springer.

Moisen, G. C., D. R. Cutler & T. C. J. Edwards (1996) Generalized linear mixed models for analyzing error in a satellite-based vegetation map of Utah *In Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium***,** 459-466.

Moran, P. A. (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological),* 10**,** 243-251.

O'Sullivan, D. & D. J. Unwin. 2010. *Geographic Information Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Smith, J. H., S. V. Stehman, J. D. Wickham & L. Yang (2003) Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment,* 84**,** 342-349.

Steele, B. M., J. C. Winne & R. L. Redmond (1998) Estimation and Mapping of Misclassification Probabilities for Thematic Land Cover Maps. *Remote Sensing of Environment,* 66**,** 192-202.

Stehman, S. V. (2009) Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing,* 30**,** 5243-5272.

Stehman, S. V. & R. L. Czaplewski (1998) Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment,* 64**,** 331-344.

Stehman, S. V., J. D. Wickham, T. G. Wade & J. H. Smith (2008) Designing a Multi-Objective, Multi-Support Accuracy Assessment of the 2001 National Land Cover data (NLCD 2001) of the Conterminous United States. *Photogrammetric Engineering & Remote Sensing,* 74**,** 1561-1571.

Tobler, W. R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* 46**,** 234-240.

Townsend, J. R. G., C. Huang, S. N. V. Kalluri, R. S. DeFries & S. Liang (2000) Beware of per-pixel characterization of land cover. *International Journal of Remote Sensing,* 21**,** 839-843.

Tran, L. T., J. Wickham, S. T. Jarnagin & C. G. Knight (2005) Mapping Spatial Thematic Accuracy with Fuzzy Sets. *Photogrammetric Engineering & Remote Sensing,* 71**,** 29-36.

Včkovski, A. 1998. *Interoperable and distributed processing in GIS*. CRC Press.

Wickham, J. D., S. V. Stehman, J. A. Fry, J. H. Smith & C. G. Homer (2010) Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sensing of Environment,* 114**,** 1286-1296.

Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry & T. G. Wade (2013) Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment,* 130**,** 294-304.

Yang, L., S. V. Stehman, J. H. Smith & J. D. Wickham (2001) Thematic accuracy of MRLC land cover for the eastern United States. *Remote Sensing of Environment,* 76**,** 418-422.

Zhang, J. & G. Foody (1998) A fuzzy classification of sub-urban land cover from remotely sensed imagery. *International Journal of Remote Sensing,* 19**,** 2721-2738.

Zhang, J. & M. F. Goodchild. 2002. Uncertainty in Categorical Variables. In *Uncertainty in Geographical Information*, x, 266 p. London ; New York: Taylor & Francis.

APPENDIX

A.1. Cross-Validation Summaries

Summary statistics and graphs can be made by comparing the predicted value to the actual value from cross-validation. Let $\hat{Z}(s_i)$ be the predicted value from cross-validation, let $z(s_i)$ be the observed value, and let $\hat{\sigma}(s_i)$ be the prediction standard error for location s_i. Then some of the summary statistics given by Geostatistical Analyst are:

1. Mean Prediction Error

$$\frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - z(s_i))}{n}$$

2. Root-Mean- Square Prediction Error

$$\sqrt{\frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - z(s_i))^2}{n}}$$

3. Average Standard Error

$$\sqrt{\frac{\sum_{i=1}^{n}\hat{\sigma}(s_i)}{n}}$$

4. Mean Standardized Prediction Error

$$\frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - z(s_i))/\hat{\sigma}(s_i)}{n}$$

5. Root-mean-square standardized prediction error

$$\sqrt{\frac{\sum_{i=1}^{n}[(\hat{Z}(s_i) - z(s_i))/\hat{\sigma}(s_i)]^2}{n}}$$

A.2. Monte Carlo Simulation through Model Builder

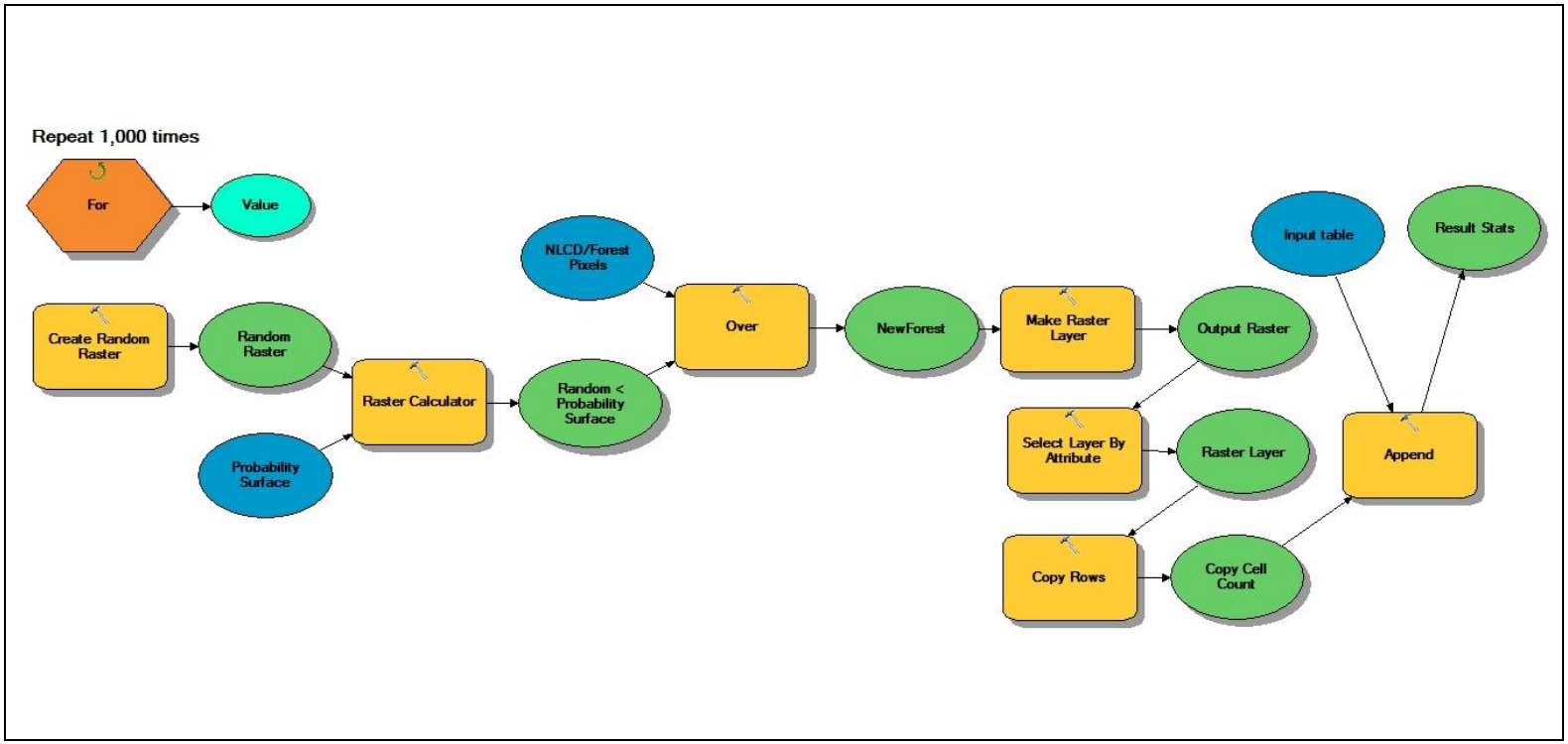

Figure A.2. Monte Carlo Simulation through Model Builder

VITA

Maria Isabel Martínez was born in Málaga, Spain. She lived in several Spanish cities until landing in the United States for the first time in 1996. She graduated *Magna Cum Laude* with a Bachelor of Science degree in Natural Resources Conservation from the University of Florida in May of 2001. She immediately began working full time as a Senior GIS Analyst with Jones Edmunds where she stayed for nine years before moving to Knoxville, Tennessee after taking a position as a GIS Analyst at Oak Ridge National Lab. She enrolled in the Department of Geography graduate program in Fall of 2011. She graduated in December of 2013 with a Master of Science degree in Geography and plans on continuing research on environmental science issues and geospatial applications.