9-9-2010

Validation of geographically based surface interpolation methods for adjusting construction cost estimates by project location

Su Zhang

Su Zhang
Candidate

Civil Engineering
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

[illegible], Chairperson

Susan Bogus Halter

for R. Matthews

VALIDATION OF GEOGRAPHICALLY BASED SURFACE INTERPOLATION METHODS FOR ADJUSTING CONSTRUCTON COST ESTIMATES BY PROJECT LOCATION

BY

SU ZHANG

BACHELOR OF CONSTRUCTION MANAGEMENT
AGRICULTURAL UNIVERSITY OF HEBEI
JUNE, 2006

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF CONSTRUCTION MANAGEMENT

The University of New Mexico
Albuquerque, New Mexico

AUGUST, 2010

DEDICATION

这篇文论献给我的家人，尤其是我的妻子和爷爷。没有他们的支持和鼓励，我就不可能完成这篇论文。同时感谢同办公室的同事和土木工程系里的朋友们。有了你们的帮助，我顺利地完成这篇论文。

This thesis dedicated to my family, especially my wife and my grandfather. My thesis will never be completed without their support and encouragement. Also, this thesis dedicated to my colleagues in the construction graduate students office and other friends in the Department of Civil Engineering at the University of New Mexico. With their help, I successfully completed my thesis.

ACKNOWLEDGEMENTS

I really appreciate Dr. Giovanni C. Migliaccio, my advisor and thesis committee chair, for giving me his trust by providing me the opportunity to pursue my Master's degree in the United States and work with him as his research assistant. I am deeply grateful to him, for his kindness, time and patience in helping me to get my Master's degree. To me, Dr. Migliaccio is not only my academic advisor, but also my friend. I sincerely hope Dr. Migliaccio will be successful in his career.

I would like to thank Dr. Susan Bogus Halter and Dr. James R. Matthews, for being a member of my thesis committee. I also appreciate their help and kindness for both of my life and academic development during two years study in the Department of Civil Engineering at the University of New Mexico

I also would like to thank Dr. Paul A. Zandbergen who comes from Netherlands and a faculty member in the Department of Geography at the University of New Mexico. I truly appreciate him that being my instructor for the ArcGIS software which is significantly useful for my research.

VALIDATION OF GEOGRAPHICALLY BASED SURFACE INTERPOLATION METHODS FOR ADJUSTING CONSTRUCTON COST ESTIMATES BY PROJECT LOCATION

BY

SU ZHANG

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF CONSTRUCTION MANAGEMENT

The University of New Mexico
Albuquerque, New Mexico

AUGUST, 2010

VALIDATION OF GEOGRAPHICALLY BASED SURFACE INTERPOLATION METHODS FOR ADJUSTING CONSTRUCTION COST ESTIMATES BY PROJECT LOCATION

By

Su Zhang

Bachelor of Construction Management, Agricultural University of Hebei, 2006

Master of Construction Management, University of New Mexico, 2010

ABSTRACT

In the construction industry, cost estimates are fundamental to the success of a construction project. Location factors are commonly used to adjust cost estimates by project location. However, not all locations have corresponding factors. Nowadays, the construction industry has employed a simple, proximity-based location factor interpolation method which is widely accepted and used. Under this method, for a location without adjustment factor, the factor of the geographically "nearest neighbor" will be selected. Although this approach was statistically substantiated by former research, it was still not sufficiently supported, considering that only one year's RSMeans City Cost Index (CCI) dataset was tested. With the help of the Global Moran's I Test in ArcGIS software, this study evaluated the spatial autocorrelation of the changes in RSMeans CCI value from year 2005 to 2009. The evaluation results substantially supported the validity of the proximity-based location factor interpolation method. In addition, evaluation of current and alternative surface interpolation methods reveals that condition nearest neighbor (CNN) method is the best rough surface interpolation method while inverse distance weighted (IDW) method is the best smooth surface interpolation method. Moreover, the Area Cost Factor (ACF) of the Department of Defense (DoD) was incorporated in this research to cross-validate all evaluations. This research is an initial step for identifying surface interpolation methods to develop spatial prediction models for location adjustment based upon several datasets, including construction cost data and socio-economical data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACF	Area Cost Factor
CCI	City Cost Index
CNN	Conditional Nearest Neighbor
DoD	Department of Defense
EID	Exclusive Identification Number
GIS	Geographic Information Systems
IDW	Inverse Distance Weighted
NN	Nearest Neighbor
RMSE	Root Mean Square Error
ST AVG	State Average
ST DEV	Standard Deviation
2D	Two-Dimensional
3D	Three-Dimensional

CHAPTER 1.0 INTRODUCTION

1.1 Overview

The construction industry is the largest industry in the United States (Gould, 1997). Every year, thousands of construction projects are built. In order to carry out a construction project successfully, one fundamental requirement is to perform accurate cost estimates. Throughout the lifecycle of a construction project, various types of cost estimates will be developed for various purposes, such as budgeting and bidding.

All the cost estimates can be classified into two main categories – conceptual (or preliminary) cost estimates and detailed cost estimates. In most cases, a conceptual cost estimate is used for programming and budgeting while a detailed cost estimate is used for bidding. During programming and budgeting, one extremely important component is to estimate the approximate cost for the intended construction project. This estimate process is called conceptual cost estimating, which is basis for successive cost estimates.

As to cost estimates, three factors will greatly affect the accuracy. One of these factors is the construction cost database. Most owners are very small and therefore, they do not have enough past construction projects to develop an in-house cost database. In practice, they will use published construction cost data from independent cost data suppliers such as RSMeans. However, some large nationwide or multinational companies will have enough past projects to develop a complete in-house cost database. It is a widespread belief that the cost estimates developed based on an in-house database would be more accurate than that developed based on an independent supplier's database.

The second factor is the level of definition of project scope. For most owners, it is impossible to develop precise project scope at the pre-design phase since they only have a general idea about the intended project. Even if there is a complete in-house cost database and the cost estimator is very experienced, without an exact project scope, the accuracy could not be greatly improved. That is because the cost estimator could not fully and accurately consider the changes and consequent risks for an undefined project scope.

The last factor that must be addressed is the method used for adjusting cost estimates. Cost estimates are developed based on historical construction cost data and adjustment factors which include location, time, size, and complexity, etc. Different adjustment methods will generate different levels of accuracy, and therefore, it is worthwhile to conduct related research.

With an effort to increase the accuracy of the cost estimates, this research specifically focuses on location adjustment method. In fact, this research is the continuation of initial research that was conducted by Martinez (2010).

Currently, the construction industry uses a simple, proximity-based location adjustment method. Although this approach was statistically substantiated by Martinez (2010), it was still not sufficiently supported, considering that only one year's RSMeans City Cost Index (CCI) was tested. In this research, the validity of the current proximity-based location adjustment method was better supported. In addition, three former surface interpolation methods were cross-validated by the Area Cost Factors from the Department of Defense (DoD ACF). Finally, three new alternative methods were established and cross-validated. Future research, however, is required to fully validate these three new surface interpolation methods.

1.2 Research Statement and Justification

Accurate cost estimating is crucial to the success of construction projects. From the owner's perspective, inaccurate estimating will be presented in two forms, overestimates and underestimates, and both of them are harmful. Overestimate means that owners need to allocate more funding than actually needed for a specific construction project. Therefore, the construction projects which are not considered important in the first place but on a "short-list" will be greatly influenced. Underestimate will put an owner in an awkward situation to seek additional funding, decrease the project scope, or terminate the project.

For construction project owners, the common method for dealing with expected inaccuracy of cost estimates is to include contingencies into their estimates. The purpose of using contingencies is to alleviate the consequences of potential errors in cost estimates. However, using contingencies will produce financial inefficiency. For project owners, especially public owners or governmental agencies, inefficient funding allocation is a big issue. While inaccurate estimating may be un-influential in periods of economic growth, currently governmental agencies and private companies are struggling with meeting needs for new construction and/or renovation of buildings and infrastructures while being subject to continuous budget and/or credit line cuts.

A cost estimate is an experience-based subjective judgment process. According to Walsh (2008), now the construction industry is experiencing an embarrassing problem – lack of experienced cost estimators in recent years. Less experienced or inexperienced cost estimators could not consider the overall parameters of a construction project to choose appropriate adjustment factors such as the location adjustment factor. Inappropriate adjustments will lead to inaccurate estimates, which may be detrimental or even fatal to a project. In addition, different cost estimators with different experiences will make different adjustments. In other words, there is no standard for a cost estimator to follow. If more factors are considered, much more errors will be created (Gould, 1997).

This research is aimed at providing an assessment of the systematic error characteristic of the currently used location adjustment techniques. In addition, this research will develop and evaluate alternative location adjustment techniques which can be successfully employed to improve the accuracy of location adjustment. In former research conducted by Martinez (2010), several location adjustment methods were developed and evaluated. However, the evaluation was based upon the 2006 RSMeans CCI dataset. The location adjustment methods validation in this research was based on RSMeans CCI and DoD ACF dataset from year 2005 to year 2009.

The expected impact of this research is to establish a more rational location adjustment method. In addition, the findings from this research could be broadly applied to different fields beyond construction. Any industry with a need for adjustment of cost data to specific locations (e.g. determination of adjustments to employee salaries due to relocation, location-specific fund distributions, etc.) would benefit from this study.

This research is an initial step for identifying a surface prediction model for location adjustments. This model will be based upon several datasets, including construction cost data and socio-economical data. With the surface prediction model, the location adjustment factor can be quickly and accurately determined for a location without a factor, even if this process is performed by an inexperienced cost estimator.

1.3 Research Questions

The geographical locations, which are usually represented in the form of cities or towns, are selected to represent the location adjustment factors. Although there are many commercial or governmental location adjustment factor datasets such as RSMeans CCI and DoD ACF available, no one of them can cover all the locations across the United States, considering there are more than 30, 000 cities. For example, in the contiguous United States (excluding the state of Hawaii and Alaska), there are 649 cities associated with RSMeans CCI and there are 337 locations associated with DoD ACF. With this fact in mind, one very important problem can be established:

How to perform location adjustment for a location without a location factor?

RSMeans explained this question by stating “For a city not listed [in the CCI], use the factor for a nearby city with similar economic characteristics” (RSMeans, 2006). However, this explanation is ambiguous and there are many possible interpretations. In the construction industry, a common interpretation to this explanation is a simple, proximity-based interpolation method. In this research, this proximity-based method is called nearest neighbor (NN) interpolation method. An example was provided to explain the NN method: if an owner wants to carry out a construction project in a city without location adjustment factors, he or she could choose the geographically nearest location’s adjustment factor. Although the NN interpolation method is commonly used in the construction industry, its validity is not substantially supported. Recent research, which was conducted by Martinez (2010), employed spatial autocorrelation analysis to support the validity of the NN interpolation method.

Another 14 location adjustment methods were identified and compared with NN method by Martinez (2010). Basically these 15 methods can be classified into two surface interpolation methods, which include nearest neighbor and local averaging. However, these two methods yield only rough surfaces. Nearest neighbor will produce a limited number of pieces of surfaces across the contiguous United States (649 surfaces for CCI and 337 surfaces for ACF). Local averaging will produce 48 pieces of surfaces across the contiguous United States. Because the amount of small pieces of surface is very small, the overall surface for contiguous United States is very rough. Another three surface interpolation methods, which include inverse distance weighted (IDW), kriging, and spline, can yield smooth surfaces. These methods are developed based on hundreds of thousands of pieces of surfaces and therefore the overall surface for continuous United States is very smooth. With this in mind, the following primary research question for this research was established:

For location adjustment factors, which is the best surface interpolation method?

There is no former relative research or idea to answer this question. In answering this primary research question, the secondary research questions were answered.

1. Can the current, industry-suggested NN interpolation method be better supported?

2. What are the possible alternatives to the current methods that may produce a smooth surface method?

3. Can these alternative methods be statistically proven to produce a more accurate construction cost estimate?

4. Can these alternative methods be visualized?

5. Can these alternative methods be cross-validated by another set of location adjustment factors such as DoD ACF?

1.4 Scope Limitations

This research focuses on location adjustment methods for cost estimates. Actual construction project data were not used in this research. The spatial autocorrelation of the changes in RSMeans CCI value from year 2005 to year 2009 was tested to better support the proximity-based interpolation method. In addition, the spatial autocorrelation of the DoD ACF and the changes in DoD ACF value from year 2005 to year 2009 were evaluated to cross-validate the proximity-based interpolation method. ArcGIS software was employed to conduct and visualize spatial autocorrelation evaluation. In addition, with the help of ArcGIS and RSMeans CCI dataset, another three surface interpolation methods for location adjustment in cost estimates were established and compared with the current two surface interpolation methods. Moreover, another set of location adjustment factors, DoD ACF, was used to cross-validate the comparison of these five surface interpolation methods.

Although mentioned in this study, the followings were not evaluated in this study:

1. Time adjustment methods
2. Scope adjustment methods
3. Actual project cost data
4. Surface prediction models

These may be topics for future research in cost data analysis.

1.5 Thesis Structure

There are six chapters in this research. Chapter 1 summarizes the research objectives, justification, and questions. In addition, research limitations are mentioned in Chapter 1. Related literature for this research is reviewed in Chapter 2. The detailed research methodology is introduced in Chapter 3. In Chapter 4, the analysis and comparison for the 5 surface interpolation methods are performed. In these 5 surface interpolation methods, 2 are rough surface interpolation methods, including nearest neighbor and local averaging. The nearest neighbor method can be subdivided into two methods based on the state boundary. One is the nearest neighbor (NN) method which does not consider the state boundary while the other one is the conditional nearest neighbor (CNN) which considers the state boundary. Another 3 are smooth interpolation methods. The discussion of the analysis and comparison results is conducted in Chapter 5. In Chapter 6, the research findings and conclusions are summarized. In addition, the detailed limitations for this research are showed.

CHAPTER 2.0 REVIEW OF RELATED LITERATURE

2.1 Overview

In this research, cost estimating is the investigated subject and the Geographic Information Systems (GIS) was used as an analysis tool. In order to get a deep understanding of the application of GIS in location adjustment for cost estimates, it is extremely important to find related literature to review. In this chapter, literature will be discussed in the following aspects:

1. Cost Estimate (see section 2.2)
2. Location Adjustment (see section 2.3)
3. Geographic Information Systems (see section 2.4)
4. Surface Interpolation Methods (see section 2.5)
5. Spatial Autocorrelation Analysis (see section 2.6)

Note: The literature review on cost estimate, location adjustment, and geographic information systems, and spatial autocorrelation analysis was a joint effort with Adam A. Martinez (2010).

2.2 Cost Estimate

According to Gould (1997), a cost estimate is an educated guess, an appraisal, an opinion, or an approximation as to the cost of a project prior to its actual construction. A cost estimate can be developed at any point throughout the lifecycle of a construction project. Therefore, it is advisable to classify cost estimates explicitly. The Association for the Advancement of Cost Engineering International (AACEI) recommended a generic cost estimate classification matrix, which was summarized in Table 1 below.

Table 1 AACE Cost Estimate Classification System (adapted from Christensen & Dysert, 2003)

	Primary Characteristic	Secondary Characteristic			
Estimate Class	**Level of Project Definiation** expressed as % of compelete definition	**End Usage** typical purpose of estimate	**Methodology** typical estimating method	**Expected Accuracy Range** typical +/- range relative to best index of 1[a]	**Preparation Effort** typical degree of effort relative to least cost index of 1 [b]
Class 5	0% to 2%	Screening or Feasibility	Stochastic or Judgment	4 to 20	1
Class 4	1% to 15%	Concept Study or Feability	Primarily Stochastic	3 to 12	2 to 4
Class 3	10% to 40%	Budget, Authorization, or Control	Mixed, but Primarily Stochastic	2 to 6	3 to 10
Class 2	30% to 70%	Control or Bid/Tender	Primarily Deterministic	1 to 3	5 to 20
Class 1	50% to 100%	Check Estimate or Bid/Tender	Deterministic	1	10 to 100

Note: [a] if the range index value of "1" represents +10/-5%, then an index value of 10 represents +100/-50%
[b] if the cost index value of "1" represents 0.005% of project cost, then an index of 100 represents 0.5%

Based on several factors, including level of project definition, end usage, estimate method, expected accuracy range, and preparation effort, Christensen and Dysert (2003) established five cost estimate classes throughout the lifecycle of a construction project. The primary characteristic used to differentiate estimate class is the level of project definition, which is expressed in the form of percentage of complete definition. According to Table 1, the accuracy range for class 1 is from positive 10% to negative 5%, which means that it is possible to overestimate 10% or underestimate 5%. For class 2, the accuracy range is positive 30% to negative 15%. For class 3, the accuracy range is positive 60% to negative 30%. For class 4, the accuracy range is positive 120% to negative 60%. Lastly, for class 5, the accuracy range is positive 200% to negative 100%.

As mentioned in Chapter 1, one of the objectives of this research is to improve the accuracy of the cost estimates. According to Carr (1989), a cost estimate must be an accurate reflection of reality. This accurate reflection of reality depends on what the cost estimators try to predict. Therefore, the more detailed the estimate is, the more accurate the cost estimate should become. However, as the level of details increases, the cost for developing cost estimate increases correspondingly. That is because higher level of detail requires more information, time and effort. According to Carr (1989), the level of details is based upon two criteria: (1) whether a particular level of uncertainty is acceptable, and (2) if it is reasonably uniform for all components of the estimate. Construction project owners should develop cost estimates based on an appropriate level of detail during the lifecycle of construction projects.

The first phase for a construction project should be the conceptual phase. During this phase, a project owner needs project cost information so that decisions as to the location and scope can be made before money is spent on design or property purchasing (Gould, 1997). This type of cost estimating mentioned above is called conceptual or rough order of magnitude (ROM) cost estimates. According to the classification system, class 4 and class 5 are conceptual cost estimates.

According to Gould (1997), conceptual cost estimates or ROM cost estimates are prepared with very little information, relying mostly on historic data and whatever descriptions are available. Typically, conceptual cost estimates are developed by establishing the gross unit cost from past similar projects which is adjusted for multiple project specific characteristic and multiplying this unit cost by the number of units intended. The unit cost could be cost per square foot for a parking lot, cost per cubic foot for a warehouse, or cost per mile for a highway.

Large owners will have their own unit price system since they have enough past projects to develop a complete construction cost database. However, for most owners, their unit prices are developed based on a national average basis. However, it is the fact that different construction projects will have different characteristics which include size, location, time, complexity, quality, and construction market conditions, etc. Therefore, it is necessary to adjust the unit price accordingly. The national average unit price must be adjusted by location since different cities will have different consumption level. In addition, the national average unit price which is developed from past projects must be adjusted to current and future dollar value. Moreover, if the intended project is larger or smaller than the standard project, the unit cost must also be adjusted. Finally, Gould (1997) pointed out that an appropriate contingency should be given to allow for later scope adjustments and economic or market condition changes.

Accurate cost estimates is fundamental to the success of a construction project. Here is an example: during construction, if more funds are needed than estimated, the project owner probably has only three options: 1) allocating additional funds to the project, 2) reducing the scope of the project, and 3) terminating the project. It is easier to understand that terminating the project means the project is unsuccessful. However, option one and option two would also greatly affect the success of a construction project. For option one, although the project owner may seek additional funds to continue the underestimated project, he or she lost the opportunity cost for the additional funds. For option two, reduced project scope would greatly affect the functionality of the project.

According to Gould (1997), detailed cost estimates are typically prepared towards the end of the design phase, as they require precise project information. Detailed cost estimates are developed based on the quantity takeoff and the unit price.

The actual cost of a construction project cannot be obtained until it is completed. From Table 1, it is clear that the expected accuracy range for cost estimates is very huge, from -100% underestimated to +200% overestimated, which are highly inaccurate projections. For that reason, there is always a striking demand to improve the accuracy of cost estimates in the construction industry.

Cost estimate consists of many important components. One of these important components is the person who is responsible for developing cost estimation, namely cost estimator. A good cost estimator should possess a combination of knowledge, managerial talents, and construction experience (Popescu & Charoenngam, 1995). In addition, according to Popescu et al. (2003), a good cost estimator should have the skills in the following aspects:

- Ability to read and understand contract documents, with special skills in reading construction drawings for all specialties and related specifications
- Ability to accurately take off the quantities of construction work for which he or she is preparing the detail estimate
- Ability to visualize the future building from drawings, which usually requires some years of construction site experience
- Knowledge of arithmetic, basic geometry, and statistics
- Familiarity with estimation software in depth and with available building cost databases
- Knowledge of building construction methods
- Knowledge of labor productivity, crew composition, and impacts of various forecasted site conditions on crew output
- Possession of office managerial skills in organizing project related cost information
- Ability to work under pressure and meet all bid requirements and deadlines (p.47)

Of all the skills mentioned above, one is familiarity with available building cost databases. This may be one of the most important features for a good cost estimator. On the other hand, the reliability of the cost data sources should be considered. In the construction industry, there are a great many commercial cost data sources available nowadays. For all types of construction projects, a common approach to use cost data is estimating by cost index. According to McCabe et al. (2002), many cost indices have been developed since it is popular to perform cost estimates by using cost index. According to McCabe et al. (2003), some examples of the cost data sources available to cost estimators are:

- RSMeans Building Construction Cost Data
- Engineering News Record
- Hanscomb-Means International Construction Cost Index
- Hanscomb's Yardsticks
- Helyar Construction Cost Guide
- KPMG International Cost Comparison Analysis
- Richardson Construction Cost Trend Reporter
- Richardson International Cost Index

As mentioned earlier, construction cost databases can be developed based on internal or external information. For very large nationwide or multinational companies such as Wal-Mart and Intel, they could have enough internal construction projects to develop a complete cost database. However, for other owners, they need to use external published database such as RSMeans CCI.

The City Cost Index (CCI) of the RSMeans Building Construction Cost Data and the Area Cost Factor (ACF) of the Department of Defense (DoD) were incorporated in this research. RSMeans CCI has been demonstrated to be very useful for commercial construction projects since it provides location adjustment factor for major cities across the United States and Canada. For military projects, DoD ACF is a common index since it provides location adjustment factor for all the cities where the bases are located. RSMeans CCI and DoD ACF index are updated and published annually.

Although Popescu et al. (2003) did not mention in their research, another important feature for a good cost estimator is the selection of method to develop cost estimates. According to Christensen and Dysert (2003), in cost estimate, the quality of the input information can greatly affect the accuracy of the output. While the methods can be considered as input, the estimate results can be considered as output.

According to Ratner (2002), who is the editor for *Walkers Building Estimators Reference Book,* there are many different cost estimate methods. Ratner assumed that if 20 different cost estimators were told to develop cost estimates based on the same drawings and plans, no more than two cost estimates would be developed on the same basis. From this, it is clear that cost estimate is a very subjective process, which would result in inaccurate estimate of construction cost. This inaccuracy will be augmented in the preliminary stage since the construction project is not completely and clearly defined. Therefore, it would be safe to say that cost estimate methods will greatly affect the accuracy.

In this chapter, various estimate methods that were published as well as how they can relate to this research were discussed. One important thing that needs to be pointed out is that not all methods are specifically applied to cost estimates. However, all of them are applied to construction or related fields.

Two of the various estimate methods studied by Duverlie and Castelain (1999) are the parametric and the case based reasoning (CBR) method. They pointed out that although the parametric method has the advantage of being made easily within a project, the obvious disadvantage is that it functions as a “black box” that does not allow users to verify the results or to ensure that they are looking for a particular case. However, the CBR method has the capability to accept unknown information and process for particular cases, which is very useful for the designer. Generally speaking, CBR is more precise than the parametric method. However, CBR is more difficult to be applied in a project since it requires a complete reasoning system based on individual projects. However, Duverlie and Castelain’s research is only applied during the design phase of a construction project. Considering this, Duverlie and Castelain’s research is partly related to this research since cost estimate methods at the conceptual phase, which is before the design phase, were not considered. The following cost estimation methods were mentioned and described in Duverlie and Castelain’s research:

- *The intuitive method* is based on the experience of the estimator. The result is always dependent on the cost estimator’s knowledge.
- *The analogical method* attempts to evaluate the cost of a set or a system from similar sets or systems.
- *The parametric method* seeks to evaluate the costs of a product from parameters characterizing the product but without describing it completely.
- *The analytical method* allows evaluation of the cost of a product from a decomposition of the work required into elementary tasks (p. 1).

In Chapter 1, the experience of cost estimators was discussed, which could be considered as an intuitive method. As mentioned earlier, Walsh (2008) pointed out that the construction industry was forced to rely on inexperienced estimators due to the current shortage of professional cost estimators. Based on this, it can be inferred that nowadays the intuitive method is jeopardized since its result is always dependent on the cost estimator's knowledge. The object of this research is to relieve this dependency problem by establishing some alternative location adjustment methods which are statistically proven. With the help of this research, instead of experienced-based justification, inexperienced cost estimators can perform location adjustments based on statistical justification.

Another research related to cost estimate method, conducted by Kim et al. (2004), compared the accuracy of the following three cost estimate techniques:

- Multiple Regression Analysis (MRA)
- Neural Networks (N-Net)
- Case-Based Reasoning (CBR)

Data of 530 residential building projects built in the year of 1997 in Seoul, Korea, were included in their research. According to Kim et al. (2004), MRA is used for explaining the phenomena and prediction of future events. In MRA, the variability of the criterion variable Y is explained by a set of predictor variables X_1, X_2,…, X_n. The N-Net, which is a computer system that is widely applied in many industries, including the construction industry, can simulate the learning process of the human brain. CBR, which is based on rule-based reason, is an alternative cost estimate technique. In CBR, experience or memory is used to develop the reasoning. One important thing that should be pointed out is that although the cost estimate techniques assessed by Kim et al. (2004) were not specifically related to this research, the methods used for measuring the performance of cost estimate techniques were specifically referenced. The method used by Kim et al. (2004) to measure the performance were respective variance and mean of the absolute error, which were also employed in this research. Four more indices, including median, standard deviation, mode, and skew of the absolute error were considered as performance measurement in this research.

The competition in today's construction industry is very intense. Accurate cost estimate means lower bidding cost. Therefore, when considering the factors that contribute to win the competition, it is clear that lower cost is as crucial as quality and functionality (Layer et al., 2002). Three types of cost estimate methods were included in the research of Layer et al. (2002), and they are:

- statistical model
- analogous model
- generative-analytical model

Based on the analysis of Layer et al. (2002), the shortcomings of the cost estimate methods mentioned above include the following:

- There is a lack in accuracy. None of the methods mentioned is able to determine the costs with the required accuracy
- The integration of cost calculation in the product development process and the possibility of design concurrent use are not solved satisfactorily
- Thus far, the product development process is only partially supported. Existing methods cover only parts of the process, interrupting the cost calculation workflow
- The increasing level of maturity during product development is not sufficiently considered. Not all the processes needed are taken into account, so that the costs calculated end up too low
- Cost estimation using statistical and analogous models can be carried out only on the basis of historic data. Innovative technologies or new resources cannot be added
- In rule-based systems, the acquisition and the maintenance of knowledge are difficult. The experience and the knowledge provided by experts do not carry enough weight (p. 507)

Accuracy, as mentioned by Layer et al. (2002), is a key component that leads to the shortcomings of cost estimate method. It would be a great contribution to the construction cost estimate if several cost estimation methods which can increase the accuracy are developed. This is also one of the most important objectives of this research.

In the preliminary stage, also known as the conceptual stage, due to the incomplete definition and limited available information, most project owners will employ rapid cost estimate methods which usually will lead to less accurate cost projection. Since sequential cost estimates are developed based on the conceptual cost estimates, it is understandable that this less accuracy will pass through the lifecycle of a construction project. Therefore, if the accuracy could be increased at the preliminary stage, it would be great beneficial to all the following sequential stages.

This research specifically focuses on the location adjustment method which is implemented by using location adjustment factors. Pietlock (2006) describes location factors as follows:

> A location factor is an instantaneous (i.e., current—has no escalation or currency exchange projection), overall total project factor for translating the total cost of the project cost elements of a defined construction project scope of work from one geographic location to another. This factor recognizes differences in productivity and costs for labor, engineered equipment, commodities, freight, duties, taxes, procurement, engineering, design, and project administration. The cost of land, scope/design differences for local conditions and codes, and differences in operating philosophies are not included in a location factor.

However, one important thing that must be addressed is that the method of using location factors is not only appropriate for conceptual cost estimates, but also for higher class of cost estimates, namely detailed cost estimates. For example, during the process of detailed cost estimates, if there are no concrete contractors available for a specific city, the cost estimator might perform location adjustment for the concrete price by using another city's concrete price with the consideration of transportation expense.

2.3 Location Adjustment

During the preliminary stage of a construction project, adjustments, which are based on specific project characteristics such as project time, size, location, and complexity, are performed for the cost estimates. According to Popescu et al. (2003), a common procedure of applying cost estimate adjustments is:

- Determine the usable area of the building, volume, or number of occupant units
- Select from the most recently published standards for the type of building that most closely matches the project, the unit area, unit volume, or occupancy unit standard cost
- Adjust selected standard costs to a projects location using regional adjustment factors (p. 59)

According to the reference mentioned above, one step of the adjustment procedure is size adjustment. In practice, the unit area (square foot), unit volume (cubic foot), or occupancy unit (number of beds or number of students) are used to adjust the size. Another step of the adjustment procedure is the location adjustment. According to Popescu et al. (2003), location adjustments are performed by using regional adjustment factors. For example, the RSMeans CCI is a published source of regional adjustment factors for commercial construction projects. DoD ACF is another example which is mainly for military projects. In fact, the RSMeans CCI and DoD ACF were important component of this research since they provide the necessary data for the evaluation of the location adjustment methods.

DoD ACF is a united facilities criteria design guide created by the Department of Defense. The following organizations are represented by the unified facilities:

- U.S. Army Corps of Engineers
- U.S. Naval Engineers Facilities command
- U.S. Air Force Civil Engineer Center

In addition, for the ACF, the Department of Defense of the United States (2005) presents the following statement:

> The ACF index is used in adjusting estimated costs to a specific geographical area. The factors reflect the average surveyed difference for each location in direct costs between that location and the national average location.

Moreover, in the Historical Air Force Construction Cost Handbook, the Air Force Support Agency describes ACF in the following statement:

> Location Factors or Area Cost Factors (ACF) are used by all DoD services to adjust average historical facility cost to a specific project location. This allows increased accuracy in identifying project costs during initial project submissions or when specific design information is not available. The area cost factor index takes into consideration the cost of construction material, labor and equipment, and other factors such as weather, climate, seismic conditions, mobilization, overhead and profit, labor availability, and labor productivity for each area (p. 73).

After analyzing both of the ACF and RSMeans CCI, it is interesting to find that various factors such as weather, climate, and labor productivity are reflected in the ACF index. However, RSMeans CCI did not consider these factors. Only construction materials, labor, and equipment are reflected in the RSMeans CCI. In addition, the ACF for each location reflects the relative relationship of construction cost at that location to the national level average of ACF=1. However, the CCI for each location reflects the relative relationship of construction cost at that location to the national level average of CCI = 100. In this research, in order to use ACF to cross-validate the various location adjustment methods, it is advised to use the same basis. Therefore, the ACF for each location is multiplied by 100 times.

As mentioned in Chapter 1, location factors are one of the three factors that will greatly affect the accuracy of cost estimates. Several location factors difficulties when creating a cost estimate were acknowledged by Popescu et al. (2003). They are listed in the following:

- Published cost standards seldom represent 100% of the project under consideration.
- The location factor of adjusting a city or community is not accounted for in the published standard.
- The time factor involved in extrapolating future construction cost variations may differ (p. 59).

This research focuses on the location adjustment component of the cost estimates. As mentioned before, one of the problems regarding the location adjustment of cost estimates is that not all cities are included in the published location factors database. In fact, this problem motivated the primary research question of former research conducted by Martinez (2010) – how should a cost estimator adjust cost estimates for locations that do not have location factors? Although they established two statistically proven surface interpolation methods, not all the surface interpolation methods were considered. Therefore, this research is the second phase of their research to statistically test all the surface interpolation methods.

The concept of an "area cost factor" as an input decision for construction expansion was introduced by Johannes et al. (1985). According to this study, the area cost factor can be described by the construction cost in an area relative to the cost in another area. The primary purpose of their research is to explore how to construction theoretically appropriate area cost factors by the economic theory of cost functions. In their research, three important sections and a conclusion sum up their findings.

In the first section, Johannes et al. (1985) described the economic theory of cost functions and regional cost differentials. To explain clearly, Johannes et al. (1985) introduced the duality principle in economics and a production technology. They claimed that it is possible to derive the minimum cost of producing any amount of output, namely a "cost function", by knowing the prices of inputs and the level of output. When the cost function is developed, the regional cost differences can be exactly determined by using a cost factoring method. For the cost function, one important assumption is the functional form of the production technology. In their research, several famous production functions which are being used in economics and engineering field were introduced and explained. Of all the production functions, the most popular one is the Cobb-Douglass function. Cobb-Douglass function allows us to break down regional differentials into regional factors, which is a very useful application. However, the regional cost factor depends on the factor prices across regions and the level of output. Johannes et al. (1985) pointed out that the area factor is dependent on the relative of labor across regions, the relative price of material across regions, and the amount of construction activity across regions.

In the second section, Johannes et al. (1985) focused on the estimation of cost differences. How the estimation of cost differences is accomplished for a sample of US military construction projects is described in the second section. In addition, Johannes et al. (1985) developed the ordinary least squares (OLS) technique which uses data, including new housing units authorized and the number of general construction contractors, to produce the cost function estimates. According to Salvatore & Reagle (2002), OLS is a simple regression analysis technique for determining the "best" line of fit. Salvatore and Reagle also describe regression analysis as a tool for testing hypothesis and for prediction (2002). Regression analysis, including OLS may be beneficial in future research related to this thesis topic.

The regional cost factors, which were determined for the year 1975, 1976, 1977, and 1978 using individual cost factors for particular locations, were explained in the third section. For each city where a set of wage data and material price data was available or could be constructed, the area cost factors are presented. With the help of the data mentioned above, a standardized city and state cost index was constructed. In addition, the standard adjustment method for locations without cost factors was described in this section. That is taking the input and then multiplying them by the cost factors for the closest city to get the cost factor for the specific location under consideration. As to the differential changes in input factor prices, it is advised to adjust the rate of inflation.

In the conclusion, Johannes et al. (1985) mentioned the goal of their study again, which is to employ economic theory of production and costs to generate construction project cost estimates based on project regions. According to Cobb-Douglass production theory, the average of the various input prices is the regional cost factor. For specific cities, the cost factors were calculated based on the available data from year 1975 to year 1978. They pointed out that inflation rate can be employed to determine future cost factors. In addition, considering that the estimated function is available in their study and assuming that the information about local factor prices and conditions is known, it is possible to construct an area cost factor for a particular construction project. It is interesting to know how the ACF index was determined, although their study did not specifically relate to this research. One important thing that must be pointed out is that this research is not going to explain how the RSMeans CCI and DoD ACF were determined, but just to use RSMeans CCI and DoD ACF as data sources to evaluate location adjustment methods, which will be explained in the section of methodology.

2.4 Geographic Information Systems

How the physical world works? It is a fundamental motivation to study in all sciences for human beings. Before discussing geographic information systems (GIS), geographic information science should be explained. According to Poku and Arditi

(2006), geographic information science is a discipline in which people try to understand how the world works by evaluating and describing human relationships with the earth, namely exploring the spatial relationships between man and the physical environments. In order to visualize and analyze spatial relationships, GIS was developed as a tool. According to Bolstad (2005), GIS have been developed since the early 1980s and were one of the fastest growing computer-based technologies of the 1990's. In addition, GIS have been used in a multitude of industries as analytical, managerial, and visualization tools. The most important characteristic for GIS is that it can incorporate database file with geographically referenced thematic data. This means that in GIS, a data layer can contain not only the geographic location such as the coordinates, but also specific attributes such as population which are related to the location. This characteristic enables GIS to be a powerful analysis tool since it allows users to not only locate a location, but also quantitatively analyze the attributes of a location. In fact, this important characteristic of GIS was utilized in this research appropriately. In this research, GIS was used to test the spatial autocorrelation of both RSMeans CCI and DoD ACF. In addition, geostatistical analyst tool in GIS was used to get the error for three surface interpolation methods. Finally, GIS was employed to visualize the spatial relationships between RSMeans CCI / DoD ACF and locations in the contiguous United States. This section will discuss the literature with regard to research involving GIS.

GIS has been widely employed in various fields. For example, GIS can be used to analyze cost data and improve cost estimate through the power of geographic management (Ashur and Crockett, 1997). An ability of GIS is to integrate geographic locations with spreadsheet information database. With the help of GIS, information such as location adjustment factor for each geographic location could be retrieved and displayed. Typically, historical bid data is used by state highway departments to estimate construction project costs. According to Ashur and Crockett (1997), a systematic information collection, organization, and storage process can be developed based on GIS to manage relevant historical cost data. Traditional data collection and storage methods, which have been done for many years, are not ideal (Ashur and Crockett, 1997). That is because a great amount of time is required to page through and assimilate compiled cost information. However, if historical data can be managed and visualized by geographic location, then data collection, storage, and retrieve could be greatly simplified. Nowadays, effective decision making is a challenge to most managers since there is an overwhelming amount of information for them to analyze. The ideal technology for managing data geographically could greatly support more effective decision making.

Although GIS has been successfully implemented in many fields for construction project management, which includes planning, scheduling, and construction material management, its application in construction cost estimate, especially at the conceptual level, is not prominent. For the application of GIS in construction project planning, Cheng & O' Connor (1996) studied application of GIS for enhanced construction site layout. For the application of GIS in construction material management, Cheng and Yang (2001) researched GIS-based integrated material layout planning and cost estimate. Based on GIS, Zhong et al. (2004) developed visual simulation methodologies and applied them in concrete dam construction processes. Oloufa et al. (1994) established the application of GIS in construction site investigation. In addition, for E-commerce applications in construction material procurement, Li et al. (2003) created an internet-based GIS. However, according to Jeljeli et al. (1993), even with all these studies mentioned above, the potential application of GIS in the construction industry has not been fully realized. Moreover, Bansal & Pal (2007) pointed out that although GIS has been widespread applied in the construction industry, construction project visualization with GIS has not yet been used to its full potential.

The effect of using the GIS environment for construction cost estimate and visualization was studied by Bansal and Pal (2007). In their research, a five – step method for quantity takeoff in cost estimate was proposed. In the first step, the user should divide a single architectural drawing into several different themes, which act as the basis of the GIS-based cost estimate. In the second step, the users convert the computer aided design (CAD) into shapefiles and then format them for ArcGIS software. In step 3, the boundaries between adjacent polygons are dissolved. In step 4, the attributes which are needed in the quantity takeoff, such as area and length, are created as new fields and entered manually into the attribute table. In the last step, a new table, also known as the bill of quantity (BOQ) is created. In the BOQ, there will be 8 fields to represent the attributes of each data theme. Despite Bansal and Pal (2007) did an excellent research about how to create GIS-aided quantity take off, it is not related to conceptual construction cost estimate. It is just an example of how to use GIS in detailed construction cost estimate.

In recent years, computer and information technology is developing rapidly as technology is evolving. According to Yu et al. (1999), the evolution of information technology and computing for architecture, engineering, construction, and facilities management fields (AEC/FM) will inevitably motivate the invention of tools that can collaborate through shared information about AEC/FM projects. It is extremely important to develop a management information system that can collect and share cost information since past cost data are very important for construction cost estimate. The Industry Foundation Classes (IFCs), which are developed by the International Alliance for Interoperability (IAI), are general models that support project information sharing and exchange among different types of computer applications. In addition, Yu et al. (1999) agree that most Building Information Modeling (BIM) packages rely on IFC to improve data interoperability and the main focus has been on representing work plans, resources and cost / schedule information. Based on the information mentioned above, it is clear that cost estimates will eventually be improved by some type of information technology since cost information is included in the list developed by Yu et al. (1999).

2.5 Surface Interpolation Methods

According to Bolstad (2005), there are many spatial interpolation methods or surface interpolation methods, but the followings are the most common:

- Thiessen Polygon (Nearest Neighbor, NN)
- Local Averaging (Fixed Radius)
- Inverse Distance Weighted (IDW)
- Kriging
- Spline

For each method, there are both inherent advantages and disadvantages and no single method has been proven to be the best (Bolstad, 2005). Of all the methods mentioned above, Bolstad (2005) conceptually defines near neighbor or thiessen polygon as the simplest method. That is because the mathematical function used in thiessen polygon is simple equality function and the nearest point is used to assign a value to a location without value. The important characteristic of the nearest neighbor interpolation method is that it defines a set of polygons, knows as thiessen polygons and all locations within a given thiessen polygon have an identical value for the Z variable (Z variable is used to denote the value of a variable of interest at an X and Y sample location). Z could be any variable we can measure at a point, such as elevation, size, and production in pounds per acre. In this research, Z variable is the RSMeans CCI or DoD ACF value. Thiessen polygons define a region around each sampled point that has an equal value to the sampled point.

One important thing should be pointed out is that the transition between polygon edges is abrupt, that is to say the variable change suddenly from one value to the next across the thiessen polygons boundary. Based on the sample points, we can develop thiessen polygons. Within thiessen polygons, the values of other points are estimated to be equal to the sample point located at the center of the polygon. Thiessen polygons provide an exact interpolator. For an exact interpolator, the interpolated surface equals the sampled values at the same point. The fact is that the value for each sample location is preserved, so there is no difference between the true and interpolated values at the sample points. However, exact interpolators are often not the best in an analysis. In a former research conducted by Martinez (2010), two basic location adjustment methods were analyzed based on thiessen polygon, namely NN and conditional near neighbor (CNN). The difference between NN and CNN is that CNN consider the state boundary as a criterion to select the nearest neighbor. In other words, when using the CNN method, both the location and its nearest neighbor are in the same state.

Local averaging could be considered as a slightly more complex method than Thiessen Polygon but as a less complex method than most of other spatial interpolation methods. In fact, local averaging method can also be viewed as a simple method. For local averaging, Bolstad (2005) presented the following statements:

> In a fixed radius interpolation, a raster grid is specified in a region of interest. Cell values are defined based on the average value of nearby samples. The samples used to calculate a cell value depend on a search radius. The search radius defines that size of a circle that is centered on each cell. Any sample points found inside the circle are used to interpolate the value for that cell. Points that fall within the circle are averaged, those outside the circle ignored. However, the number of samples is decided based on what search radius value is defined (p. 445 - 446).

In the research of Martinez (2010), this method was established as the state average (ST AVG) method. In their research, instead of defining the search radius traditionally, the state boundary was used to define the spatial extents of the search. To better explain this concept, there is an example. All values within a state were averaged to estimate a collective value used for every potential project location within the state. Local averaging interpolators are not exact interpolators since they may average several points in the vicinity of a sample, and therefore they are unlikely to place the measured value at sample points in the interpolated surface.

However, both of these two surface interpolation methods yield only rough surface. IDW, kriging, and spline can yield smooth surfaces, which would be the prototype for 3D spatial prediction. According to Bolstad (2005), the IDW interpolator estimates the value at unknown points using the sampled values and distance to nearby known points. In addition, his statements about IDW are:

> The weight of each sample point is an inverse proportion to the distance, thus the name. The farther away the point, the less weight the point has in helping defined the value at an unsampled location. Any greater than two points may be used, up to all points in the sample. Typically some fixed number of close points is used, for example, the three nearest sampled points will be used to estimate values at unknown locations. IDW is an exact interpolator. Interpolated values are equal to the sampled values at each sampled point. IDW results in smooth interpolated surface. The values do not jump discontinuously at edges, as occurs with Thiessen Polygons. The IDW, and all other interpolators, should be applied only after the user is convinced the method provides estimates with sufficient accuracy (p. 447 - 448).

Kriging is a statistically-based estimator of spatial variables (Bolstad 2005). If differs from the trend-surface method in that predictions are based on regionalized variable theory, which includes three main components:

> The first is the spatial trend, an increase or decrease in a variable that depends on direction. The local spatial autocorrelation, which is the tendency for points near each other to have similar values, is described in the second component. The last component is random, stochastic variation. These three components are combined in a mathematical model to develop an estimation function, which is then applied to the measured data to estimate values across the study area (p. 457).

Like IDW, weights in kriging are used with measured sample variables to estimate values at unknown locations. With kriging, the weights are chosen in a statistically optimal fashion, given a specific kriging model and assumptions about the trend, autocorrelation, and stochastic variation in the predicted variable (Bolstad, 2005).

Kriging uses the concept of a lag distance. About the lag distance, Bolstad (2005) presented the following statements:

> Lag distances often are applied with an associated lag tolerance. A lag tolerance is required because the individual lag distances typically are not repeated in the sample data, and the reason is most or all distances between sample points are different and there is no replication to calculate the variability at each lag. The fact is that some distances may be quite similar, but usually will differ in the smallest decimal places. The lag tolerance can circumvent this problem by combing observations for subsequent calculations (p. 458).

A spline is a flexible ruler that was commonly used by draftsmen to create smooth curves through a set of points. A road location may have been surveyed at a set of points. To produce a smoothly bending line to represent the road, the draftsman carefully plotted the points, and the spline ruler was bent along a path defined by the set of points. A smoothly curving line was then drawn along the edge of the spline (Bolstad, 2005). For spline functions, Bolstad presented the following statements:

> Spline functions, also referred to as Spline, are used to interpolate along a smooth curve. These functions serve the same purpose as the flexible ruler in that they force a smooth line to pass through a desired set of points. Spline functions are more flexible because they may be used for lines or surfaces and they may be estimated and changed rapidly. The sample points are analogous to the drafted points in that these points serve as the "guide" through which the spline passes (p.450).

2.6 Spatial Autocorrelation Analysis

It is the fact that there are few published studies which are similar to the subject of that incorporating GIS with construction cost estimate. However, many published studies are related to two important aspects of this research, namely spatial pattern and spatial autocorrelation. Both spatial pattern and spatial autocorrelation are important parts of the spatial analysis in GIS and spatial analysis in GIS can contribute to the cost estimates and this contribution is the most important part of this research.

Messner et al. (2003) use exploratory spatial data analysis (ESDA) to examine the distribution of homicides in 78 counties in, or around, the St. Louis metropolitan area for two time periods: a period of relatively stable homicide (1984–1988) and a period of generally increasing homicide (1988–1993). ESDA is a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity. In addition, the core of the ESDA is the formal treatment of the notion of spatial autocorrelation. The research conducted by Messner et al. (2003) is very useful and it provides many clues and good ideas about how to examine the distribution of the changes in CCI value and how to analyze the spatial pattern of the changes in CCI value over time.

What is spatial pattern? According to Unwin (1996), spatial pattern is the characteristics of the spatial arrangement of objects by their spacing relation to each other. In addition, Unwin (1996) specifically addresses visualization as a necessary first step in all spatial data analysis.

According to Monmonier (1990), the strategies for the visual display and analysis of geographic time-series data are spatial or non-spatial, single-view or multiple-view, static or dynamic. However, Monmonier addresses the graphic portrayal of geographic time-series data. Monmonier explores a variety of graphic strategies for the simultaneous symbolic representation of time and space, and further summarizes these strategies in a potential use conceptual framework which is useful for cartographers, geographers, and graphic designers. These strategies include statistical diagrams, maps, video animations, and interactive graphics systems, which can manipulate time as a variable.

In this research, the spatial autocorrelation was measured with the help of GIS. What is spatial autocorrelation? According to Bolstad (2005), spatial autocorrelation is the tendency of nearby objects to vary in concert, which means that high values are found near high values, and low values are found near low values. In other words, spatial autocorrelation is a phenomenon where the values of a variable located within certain geographic area show a similar pattern (Suriatini, 2006). The occurrence of spatial autocorrelation can be examined in the positive and negative form. According to Lee and Wong (2001), positive spatial autocorrelation is said to occur when high or low values for a random variable tend to cluster in space. Negative spatial autocorrelation occurs when locations tend to be surrounded by neighbors with very dissimilar values. Zero means the observed values are arranged randomly and independently over space.

Bivand (1998) conducted a review of spatial statistical techniques for location studies. Of Bivand’s research results, the most useful part for my project is the global and local measurement of the spatial autocorrelation. Global spatial autocorrelation is a measure of the overall clustering of the data and it yields only one statistic to summarize the whole study area. But if there is no global autocorrelation or no clustering, there is still a way to find clusters at a local level using local spatial autocorrelation. For this research, the spatial autocorrelation of the changes in CCI value at national level will be tested with the help of global spatial autocorrelation. According to Bivand (1998), the spatial autocorrelation can be developed with the help of global and local Moran's I test.

How to use the global and local Moran's I test? Rosenberg et al. (1999) use local Moran's I test result to assess the spatial autocorrelation of cancer mortalities in Western Europe. In addition, Borden and Cutter (2008) uses global and local Moran's I test result to assess the spatial autocorrelation of natural hazards mortality in the United States. Martinez (2010) employed the global and local Moran's I test results to assess the spatial autocorrelation of RSMeans' CCI value. Moran's I test result includes Moran's I index, Z score, and p-value. In general, a Moran's index value near +1.0 indicates clustering (positive spatial autocorrelation) while an index value near -1.0 indicates dispersion (negative spatial autocorrelation). However, without looking at statistical significance there is no basis for knowing if the observed pattern is just one of many possible versions of random. Therefore, there is also need to check the Z score and p-value. The Z score is a test of statistical significance that helps us decide whether or not to reject the null hypothesis. The p-value is the probability that we have falsely rejected the null hypothesis. When the p-value is small and the absolute value of the Z score is large enough that it falls outside of the desired confidence level, the null hypothesis can be rejected.

CHAPTER 3.0 RESEARCH METHODOLOGY

3.1 Overview

This research specifically focused on assessing surface interpolation methods for adjusting construction cost estimates by project location. However, this research was not related to any other adjustment parameters which will also affect cost estimates such as project size, time, and complexity. No actual construction project data was collected in this research. An overview of the research framework implemented in this study can be explained using the flowchart in Figure 1.

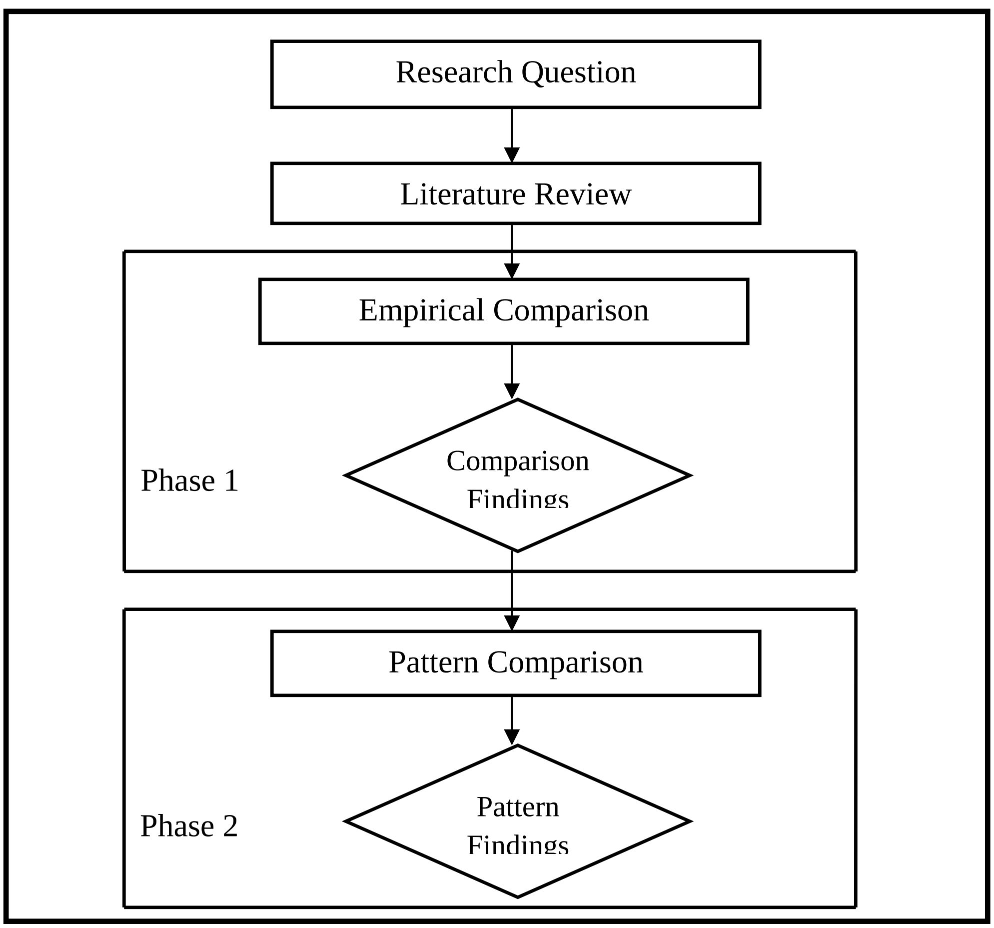

Figure 1 Flowchart of Research Steps

According to the flowchart of research steps, the first step is the research question. This step has been described in Chapter 1, and included the tasks of setting up research objectives, establishing research justification, and identifying research scope limitations. After defining research questions, the next step was to perform a comprehensive literature review. This step was described in Chapter 2. According to Figure 1, there are two analysis phases incorporated in this research. In phase 1, an exploratory empirical comparison of surface interpolation methods were conducted and some statistical findings were produced. In the second analysis phase, dissimilar to phase 1, a pattern comparison was performed and some visual findings were presented. In phase 1, statistical comparison, 5 surface interpolation methods were evaluated and they are:

- Nearest Neighbor
- Local Averaging
- Inverse Distance Weighted (IDW)
- Kriging
- Spline

However, In order to be consistent with the research conducted by Martinez (2010), the nearest neighbor method was divided into two different methods based on the state boundary. In addition, a new name, state average was assigned to local averaging. The 6 methods actually evaluated in this research are listed as follows:

- Condition Nearest Neighbor (CNN, nearest neighbor with state boundary)
- Nearest Neighbor (NN, nearest neighbor without state boundary)
- State Average (ST AVG, local averaging)
- IDW
- Spline
- Kriging

In phase 2, pattern comparison, three surface interpolation methods which can be used to develop smooth surface were assessed. These three methods are:

- IDW
- Spline
- Kriging

In fact, this research is the second phase of the research conducted by Martinez (2010) in the field of location adjustment methods for construction cost estimates. In this previous research, 15 different location adjustment methods were evaluated and compared, including the method currently adopted by the construction industry, the NN method. However, as mentioned in Chapter 1, all these 15 methods were based on two surface interpolation methods – nearest neighbor and local averaging interpolation method, plus considering various criteria.

The goal of this research is to compare the surface interpolation methods and then develop a prototype surface cost function for spatial prediction models.

In phase 1, the proximity-based interpolation method, NN or CNN method, was better supported. The former research conducted by Martinez (2010) was conducted based on 649 cities' RSMeans City Cost Index (CCI) value in the year of 2006. However, RSMeans will publish the CCI for the 649 cities annually and it is the fact that most of the CCI value changes. Although some cities keep the same CCI value for two or three years, each city's CCI value changes in the long term. Therefore, to better support the validity of the proximity-based interpolation method, it is very important to assess the spatial autocorrelation of the changes in CCI value. The spatial autocorrelation analysis was based on the technique of Global Moran's I test. Moreover, the spatial autocorrelation of the 2006 Area Cost Factors from the Department of Defense (DoD ACF) and the changes in DoD ACF value from year 2005 to year 2009 were assessed to cross-validate the result obtained based on RSMeans CCI dataset.

In phase 1 analysis and comparison, all the surface interpolation methods, including CNN, NN, ST AVG, IDW, kriging, and spline were evaluated and compared by the following techniques:

- Comparison of Overestimates and Underestimates
- Best Performance Comparison
- Comparison of Error Percentage
- Descriptive Statistics
- Pattern Comparison

Analysis and comparison findings obtained from these techniques were interpreted in Chapter 5. The errors for three surface interpolation methods, including IDW, kriging, and spline, were obtained through the “Geostatistical Analyst” in the ArcGIS software.

In pattern comparison, both 2 D and 3 D surface interpolation models were developed and compared for IDW, kriging, and spline.

In both phase 1 and phase 2 analyses and comparison, RSMeans CCI was used to test the result while DoD ACF was used to cross-validate the result.

3.2 Research Hypothesis

There are four research hypotheses for this research. They are:

(1) The interpolation method for location factors is valid.

(2) Of all the three rough surface interpolation methods (CNN, NN, and ST AVG), CNN is expected to perform the best (producing less error).

(3) Of all the three smooth surface interpolation methods (IDW, Kriging, and Spline), Kriging is expected to perform the best (producing less error).

(4) Smooth surface interpolation methods are expected to outperform rough surface interpolation methods.

3.2.1 The Research Hypothesis for the Validity of Interpolation Methods

The validity of interpolation method is affected by the spatial autocorrelation result of the location factors. In another word, the interpolation method is valid if the location factors are spatially autocorrelated. In order to test this hypothesis, the spatial autocorrelation of the RSMeans CCI 2006 dataset, the changes in RSMeans CCI dataset from year 2005 to 2009, DoD ACF 2006 dataset, and the changes in DoD ACF dataset from year 2005 to 2009 were assessed. In addition, the spatial autocorrelation can be tested with help of Moran's I Test in the ArcGIS software. Three indices are used for displaying the test result, including Moran's I index, Z-score, and p-value.

The expected result is the interpolation method for location factors is a valid method. In the former research conducted by Martinez (2010), the 2006 RSMeans CCI dataset is spatially autocorrelated and then the validity of the interpolation method for location factors was supported. Considering that both RSMeans CCI and DoD ACF datasets are systematic datasets, the location factors and the changes in location factors should be spatially autocorrelated, which means that high values are found near high values while low values are found near low values. If the location factors are spatially autocorrelated, it is safe to conclude that interpolation method for location factors is valid.

3.2.2 The Research Hypothesis for the Best Rough Surface Interpolation Methods

The second research hypothesis was explained by Figure 2.

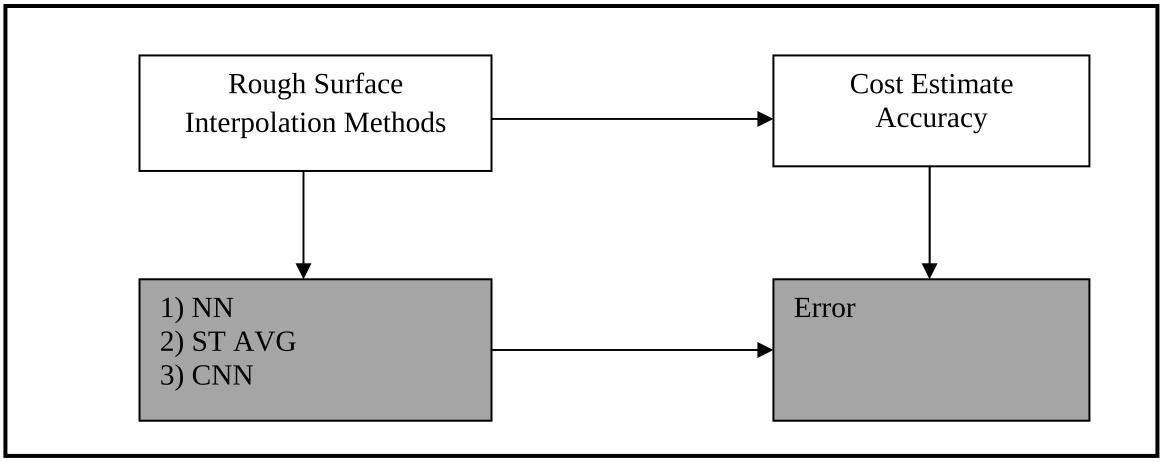

Figure 2 The Research Hypothesis for the Best Rough Surface Interpolation Methods

The accuracy of cost estimates is affected by the rough surface interpolation methods. Depending on which of the three rough surface interpolation methods (including NN, ST AVG, and CNN) is selected, the error will change.

The expected result is that CNN performs the best, which means that CNN has the least error. In the research conducted by Martinez (2010), with the help of 2006 RSMeans CCI dataset, CNN interpolation method was validated as the best method. RSMeans CCI dataset is a scientific and systematic dataset. Therefore, based on the data from year 2005 to year 2009, the result should be the same. In addition, CNN method considers the state boundary criteria and state boundary has great effect on the location factors.

3.2.3 The Research Hypothesis for the Best Smooth Surface Interpolation Methods

The third research hypothesis was explained by Figure 3.

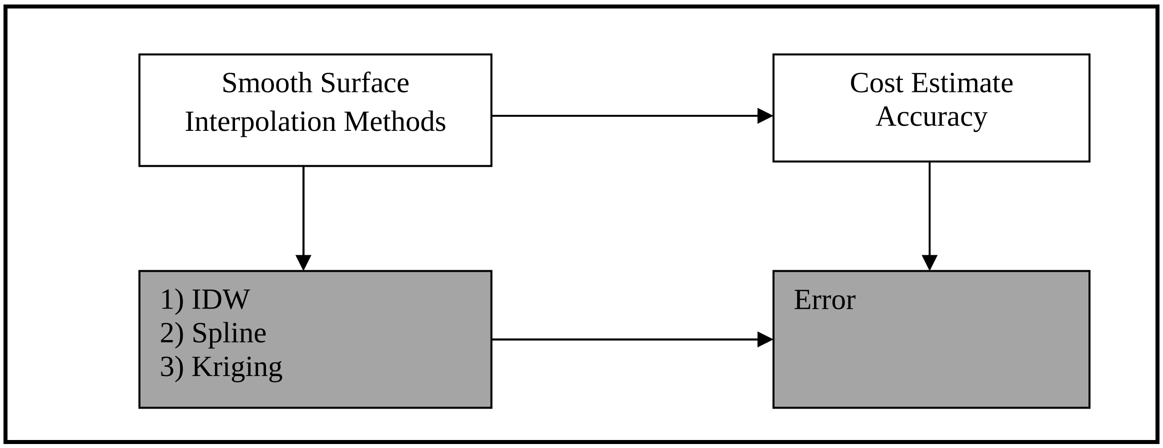

Figure 3 The Research Hypothesis for the Best Smooth Surface Interpolation Methods

The accuracy of cost estimates is affected by the smooth surface interpolation methods. Depending on which of the three smooth surface interpolation methods (including IDW, spline, and kriging) is selected, the error will change.

The expected result is that kriging performs the best, which means that kriging has the least error. Kriging is a more complex method with lots of tweeking. Kriging overcomes many shortcomings of the traditional interpolation methods. The kriging weights are determined by the semivariogram and the configuration of the data set. Kriging is an optimal interpolator in the sense that the estimates are unbiased and have known minimum variances. Since the estimation variances can be determined and mapped like the estimates, and assuming a particular distribution, it is possible to calculate the confidence which can be placed in the estimates. This makes kriging uniquely different from other interpolation methods.

3.2.4 The Research Hypothesis for the Best Surface Interpolation Methods

The last research hypothesis was explained by Figure 4.

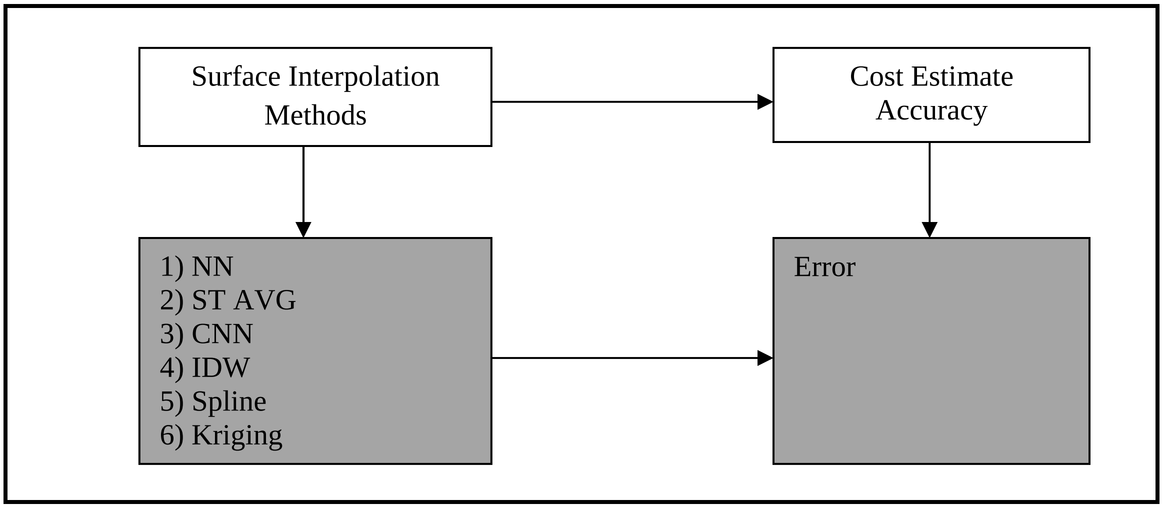

Figure 4 The Research Hypothesis for the Best Surface Interpolation Methods

The accuracy of cost estimates is affected by the various surface interpolation methods. Depending on which of the six surface interpolation methods (including NN, ST AVG, CNN, IDW, kriging, and spline) is selected, the error will change.

The expected results are that kriging performs the best among all the surface interpolation methods. In addition, the smooth surface interpolation methods outperform the rough surface interpolation methods. All smooth surface interpolation methods are developed based on complex functions and hundreds of thousands of cells, and hence they will outperform the rough surface interpolation methods are developed based on simple functions. Since the hypothesis is that kriging is the best smooth surface interpolation method, it will also be the surface interpolation method.

3.3 Spatial Interpolation

The function of spatial estimating in GIS was employed in this research. In geographic information science, spatial estimation incorporates interpolation and prediction techniques. Both interpolation and prediction techniques can be used to estimate variables for locations where the variables have not been measured. However, it is important to understand that spatial prediction differs from spatial interpolation. According to Bolstad (2005), spatial prediction is different than interpolation because it uses a statistical fitting process rather than a set algorithm, and because spatial prediction uses independent variables as well as coordinate locations to estimate unknown variables. Bolstad (2005) admits, “Our distinction between spatial prediction and interpolation is artificial, but it is useful in organizing our discussion, and highlights an important distinction between our data-driven models and our fixed interpolation methods” (p. 409).

In this research, although the ambiguous distinction between the two terms existed, they are used distinctly. The current phase of this research will focus on spatial interpolation. Spatial prediction will be a good research topic as a continuation of this research.

According to Bolstad (2005), nearest neighbor, local averaging, inverse distance weighted (IDW), kriging, and spline are 5 common surface interpolation methods. These five methods were analyzed and compared in this research.

However, there are another five surface interpolation methods in ArcGIS software, including natural neighbor, spline with barriers, topo to raster, topo to raster by file, and trend. According to Fan et al. (2005), the natural neighbor suffers from the disadvantage of being computationally costly, especially when the number of sites is large. Therefore, the natural neighbor method is not considered in this research. For spline with barriers, it is similar to spline except that it considered the barriers. This is useful at the state level analysis, but not for the national level. Topo to raster and topo to raster by file are exclusively used for topography. As to trend, the disadvantage is that this method is highly affected by the extreme values and uneven distribution of observational data points. The problem is further complicated by the fact that some of the data points are more informative than others. In addition, all these five methods could not be analyzed with the help of Geostatistical Analyst tool in ArcGIS. Therefore, they are not considered in this research.

3.4 Performance Measurement: Error

This subsection will discuss how to measure the performance of each surface interpolation method. In this research, performance was evaluated in the form of an "error" value. According to Taylor (1997):

> All measurements, however careful and scientific, are subject to some uncertainties. Error analysis is the study and evaluation of these uncertainties, its two main functions being to allow the scientist to estimate how large his uncertainties are, and to help him to reduce them when necessary. The analysis of uncertainties, or "errors," is a vital part of any scientific experiment (p. xv).

Various methods can be used to calculate the error. In this research, RSMeans CCI dataset from year 2005 to year 2009 were used. In the RSMeans dataset, a total of 649 cities in the contiguous United States were plotted as points on a map by ArcGIS software. The individual point is the actual location of the city with CCI value. In addition, the CCI value from year 2005 to year 2009 for each CCI city were uploaded in an Excel spreadsheet and then were joined as attributes to spatially associate with each corresponding city. The cities with CCI values in the state of Alaska and Hawaii were excluded from this research since they are not part of the contiguous United States. In other words, they do not have any neighbor state. The cities with the CCI value attributes were then exported as a new data layer and a map was created to display the CCI cities throughout the contiguous United States. In order to identify a CCI city quickly and correctly, an exclusive identification number (EID) will be given to each city. The map mentioned above is shown in Figure 5.

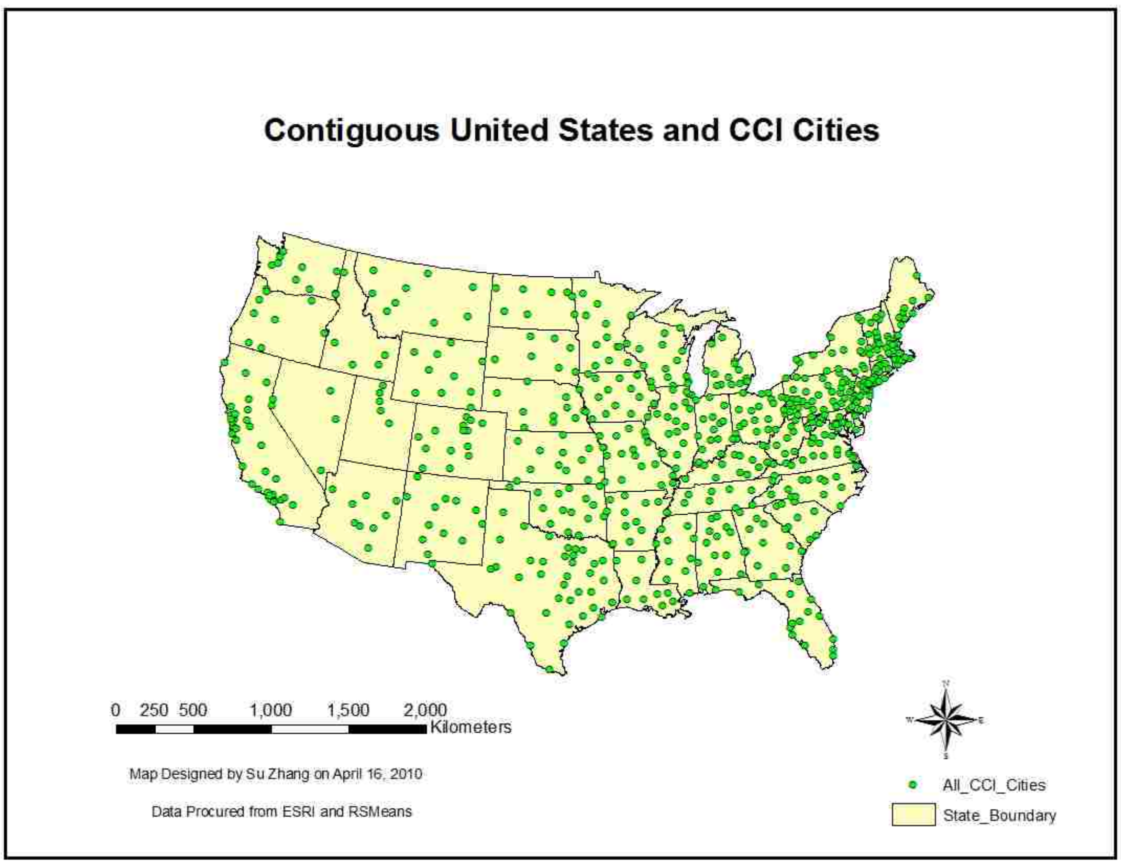

Figure 5 RSMeans CCI Cities

In addition, the attribute table for RSMeans CCI cities is shown in Figure 6.

Attributes of All_CCI_Cities

FID	Shape *	NAME	ST	STATE	EID	CCI_05	CCI_06	CCI_07	CCI_08	CCI_09
0	Point	BELLEVILLE	KS	KANSAS	0	79	83.5	83	83.2	82.7
1	Point	EMPORIA	KS	KANSAS	1	80.1	79.8	80.3	81.6	81.2
2	Point	TOPEKA	KS	KANSAS	2	85.2	85.6	85.8	84.9	85.1
3	Point	CHARLESTON	SC	SOUTH C	3	75.1	75.3	78.7	83.3	83.2
4	Point	SAVANNAH	GA	GEORGIA	4	79.7	79.6	82.3	82.4	83.3
5	Point	STATESBORO	GA	GEORGIA	5	73.4	73.4	76.4	76.8	77
6	Point	MIDLAND	TX	TEXAS	6	77.4	77	77.7	77.4	77.9
7	Point	ODESSA	TX	TEXAS	7	75.4	75.2	75.8	75.8	76.3
8	Point	NEWCASTLE	WY	WYOMIN	8	73.1	77.6	81.2	81.4	81.3
9	Point	RAPID CITY	SD	SOUTH D	9	78	78.4	78.3	79.4	81.5
10	Point	CHILLICOTHE	MO	MISSOURI	10	85.1	85.4	84.1	84.8	83.9
11	Point	HARRISONVILLE	MO	MISSOURI	11	97.2	98.6	96.6	98	96.2
12	Point	KANSAS CITY	MO	MISSOURI	12	103.4	103.5	103.1	102.4	102.1
13	Point	KIRKSVILLE	MO	MISSOURI	13	88.3	89	89	88.6	88
14	Point	SEDALIA	MO	MISSOURI	14	90.7	91.2	90.8	90.9	89.6
15	Point	ST. JOSEPH	MO	MISSOURI	15	95.3	97	96.3	96.1	94.9
16	Point	KANSAS CITY KANSAS	KS	KANSAS	16	95.5	97.7	98.1	96.8	96.4
17	Point	LIMA	OH	OHIO	17	92.9	92.8	92.9	93.7	92.1
18	Point	FORT WAYNE	IN	INDIANA	18	88.5	89.9	88.7	88.6	88.2
19	Point	SOUTH BEND	IN	INDIANA	19	89.4	91.5	89.7	89.5	90.4
20	Point	HUTCHINSON	KS	KANSAS	20	74.7	74.5	79.2	80.5	80.1
21	Point	INDEPENDENCE	KS	KANSAS	21	78.1	85.1	84.5	84.1	83.5
22	Point	SALINA	KS	KANSAS	22	80.1	83.5	83	83.4	83.1
23	Point	GREENVILLE	SC	SOUTH C	23	72.9	73.4	75.2	78.2	78.3
24	Point	ROCK HILL	SC	SOUTH C	24	67.2	67.7	69.4	77.2	77.1
25	Point	SPARTANBURG	SC	SOUTH C	25	73	73.5	75.3	78.3	78.4
26	Point	ASHEVILLE	NC	NORTH C	26	74.5	75.5	77.8	77.1	77.5
27	Point	GASTONIA	NC	NORTH C	27	74.6	76	78.3	77.6	78
28	Point	HICKORY	NC	NORTH C	28	67.5	67.8	75.1	74.5	74.7
29	Point	CHARLOTTE	NC	NORTH C	29	75.3	76.5	80.4	79.5	80
30	Point	CHEYENNE	WY	WYOMIN	30	78	81.4	86.1	85.5	85.7
31	Point	RAWLINS	WY	WYOMIN	31	74.3	78.8	82.5	82.7	82.6
32	Point	WHEATLAND	WY	WYOMIN	32	75.4	79.4	81.8	81.9	81.8
33	Point	BECKLEY	WV	WEST VIR	33	93.3	93.1	92.8	92.8	92.6
34	Point	BLUEFIELD	WV	WEST VIR	34	89.1	89.1	89	89.1	89
35	Point	BUCKHANNON	WV	WEST VIR	35	95.8	96	95.5	95	94.7
36	Point	CHARLESTON	WV	WEST VIR	36	95.9	95.7	95.3	94.8	95.1
37	Point	CLARKSBURG	WV	WEST VIR	37	95.5	96.2	95.8	95	94.5
38	Point	GASSAWAY	WV	WEST VIR	38	95.3	95.4	95	95.3	94.5
39	Point	HUNTINGTON	WV	WEST VIR	39	95.5	96.9	98.8	96.3	96.3
40	Point	LEWISBURG	WV	WEST VIR	40	91.6	91.6	91.5	92.1	92
41	Point	MARTINSBURG	WV	WEST VIR	41	89.8	90.1	89.9	89.4	89.5
42	Point	MORGANTOWN	WV	WEST VIR	42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Record: 1 Show: All Selected Records (0 out of 649 Selected) Options

Figure 6 Attributes of the Data Layer of All CCI Cities

Similarly, DoD ACF dataset from year 2005 to year 2009 were used. A total of 337 locations in the contiguous United States were referenced as points on a map with the help of ArcGIS software. The point is the actual location with ACF value. For each point, the ACF value is also from year 2005 to year 2009. The location with ACF value in the state of Alaska and Hawaii were excluded with the same reason as CCI. In order to identify an ACF location quickly and correctly, an exclusive identification number (EID) will also be given to each location. The map mentioned above is shown in Figure 7.

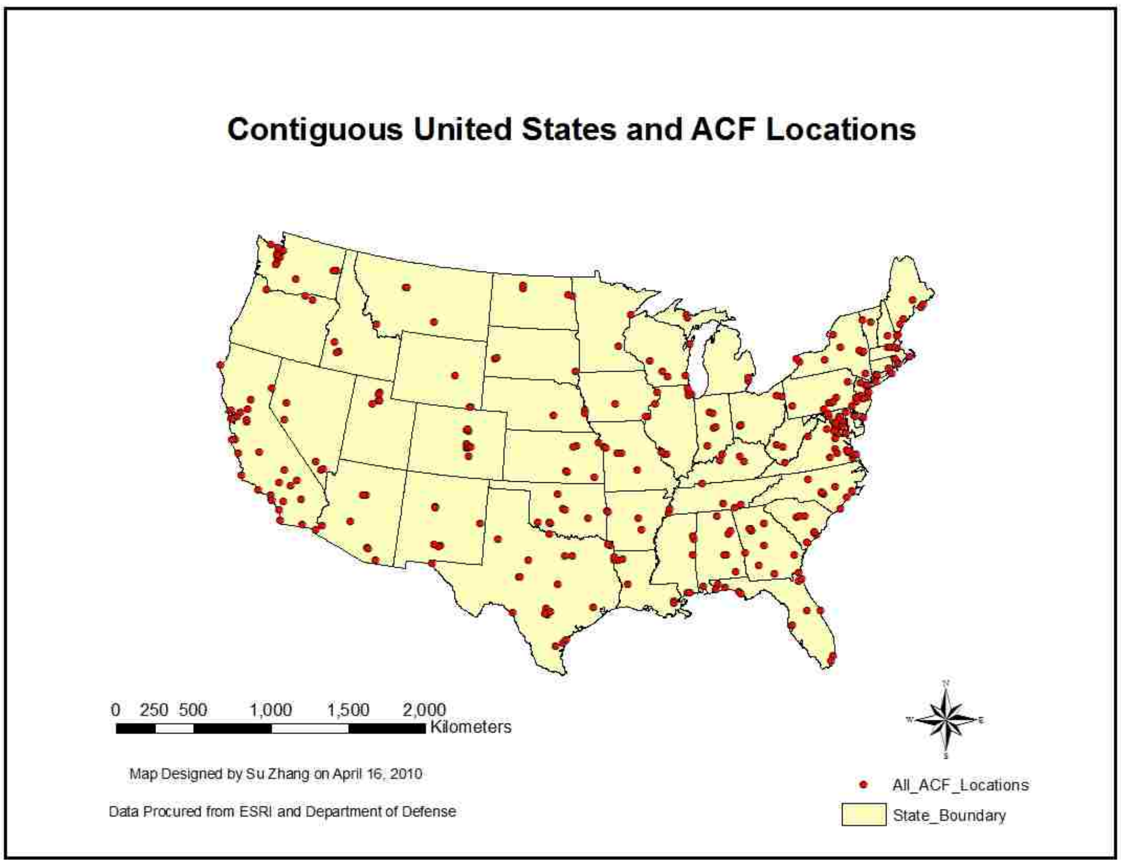

Figure 7 DoD ACF Locations

In addition, the attribute table for DoD ACF locations is shown in Figure 8.

Attributes of All_ACF_Location

FID	Shape *	EID	Name	ST	State	MILCON_05	MILCON_06	SUSTAIN_06
0	Point	0	MOBILE	AL	ALABAMA	0.8	0.84	0.8
1	Point	1	MONTGOMERY	AL	ALABAMA	0.74	0.81	0.77
2	Point	2	ANNISTON ARMY DEPOT	AL	ALABAMA	0.77	0.7	0.67
3	Point	3	FORT MCCLELLAN	AL	ALABAMA	0.76	0.72	0.67
4	Point	4	FORT RUCKER	AL	ALABAMA	0.78	0.75	0.66
5	Point	5	MAXWELL AIR FORCE BASE	AL	ALABAMA	0.74	0.81	0.77
6	Point	6	REDSTONE ARSENAL	AL	ALABAMA	0.78	0.84	0.84
7	Point	7	FLAGSTAFF	AZ	ARIZONA	1	0.978	0.94
8	Point	8	TUCSON	AZ	ARIZONA	0.9	0.96	0.94
9	Point	9	DAVIS-MONTHAN AIR FORCE BASE	AZ	ARIZONA	0.96	0.99	0.97
10	Point	10	FORT HUACHUCA	AZ	ARIZONA	1.12	1.1	1.08
11	Point	11	LUKE AIR FORCE BASE	AZ	ARIZONA	1	0.99	0.97
12	Point	12	NAVAJO ARMY DEPOT	AZ	ARIZONA	1	0.98	0.94
13	Point	13	MARINE CORPS AIR STATION YUMA	AZ	ARIZONA	1.11	1.25	1.32
14	Point	14	YUMA PROVING GROUND	AZ	ARIZONA	1.11	1.25	1.32
15	Point	15	FORT SMITH	AR	ARKANSAS	0.92	0.85	0.82
16	Point	16	PINE BLUFF	AR	ARKANSAS	0.78	0.85	0.82
17	Point	17	FORT CHAFFEE	AR	ARKANSAS	0.92	0.85	0.82
18	Point	18	LITTLE ROCK AIR FORCE BASE	AR	ARKANSAS	0.8	0.87	0.84
19	Point	19	SAN DIEGO	CA	CALIFORNIA	1.16	1.13	1.17
20	Point	20	SAN FRANCISCO	CA	CALIFORNIA	1.37	1.23	1.31
21	Point	21	29 PALMS MARINE CORPS BASE	CA	CALIFORNIA	1.38	1.29	1.34
22	Point	22	BARSTOW	CA	CALIFORNIA	1.24	1.28	1.33
23	Point	23	BEALE AIR FORCE BASE	CA	CALIFORNIA	1.24	1.27	1.33
24	Point	24	CAMP PENDLETON	CA	CALIFORNIA	1.18	1.12	1.16
25	Point	25	CENTERVILLE BEACH	CA	CALIFORNIA	1.37	1.02	0.99
26	Point	26	ALAMEDA	CA	CALIFORNIA	1.24	1.23	1.31
27	Point	27	PETALUMA	CA	CALIFORNIA	1.24	1.29	1.38
28	Point	28	CHINA LAKES NAVAL WEAPONS CENTER	CA	CALIFORNIA	1.4	1.26	1.31
29	Point	29	CONCORD	CA	CALIFORNIA	1.24	1.23	1.31
30	Point	30	EDWARDS AIR FORCE BASE	CA	CALIFORNIA	1.38	1.29	1.34
31	Point	31	EL CENTRO NAVAL AIR FACILITY	CA	CALIFORNIA	1.21	1.29	1.33
32	Point	32	FORT HUNTER LIGGETT	CA	CALIFORNIA	1.44	1.22	1.27
33	Point	33	FORT IRWIN	CA	CALIFORNIA	1.3	1.28	1.33
34	Point	34	FORT ORD	CA	CALIFORNIA	1.21	1.21	1.26
35	Point	35	LEMOORE	CA	CALIFORNIA	1.24	1.2	1.24
36	Point	36	LOS ANGELES AREA	CA	CALIFORNIA	1.24	1.13	1.19
37	Point	37	MARCH AIR RESERVE STATION	CA	CALIFORNIA	1.26	1.09	1.15
38	Point	38	MONTEREY AREA	CA	CALIFORNIA	1.2	1.21	1.26
39	Point	39	PORT HUENEME AREA	CA	CALIFORNIA	1.18	1.14	1.19
40	Point	40	SACRAMENTO AREA	CA	CALIFORNIA	1.12	1.13	1.18

Record: 9 Show: All Selected Records (0 out of 337 Selected) Options

Figure 8 Attributes of the Data Layer of All ACF Locations

As mentioned earlier, RSMeans CCI dataset was used to perform internal validation to test which surface interpolation method has the most accurate result, while DoD ACF was used to cross-validate the results. In addition, if some actual project cost data will be available in the future, they could be used to externally validate the same surface interpolation methods. The application of actual cost data as external validation could be the next phase of this research.

In each surface interpolation method, there are two location factor values for each location. One is the measured value which is procured from the RSMeans CCI or DoD ACF database while the other is the predicted value. These two values can be seen as "twin value". It is very important to understand what is meant by twin value. In a pair of twin value, one value is the ideal alternative value to another value. In other words, the twin value is what would be used if the original location did not have a location factor value. The difference of the twin value for a location varies depending on which surface interpolation method is selected. That is because the predicted value is different when different surface interpolation method is used. The predicted value is referred to the estimated value. Since another value in the twin value is the actual value, the difference between predicted and actual value is what distinguishes the performance of different surface interpolation methods. The calculation of the difference between predicted and actual value produces an "error". The following general remarks with regard to error analysis are referenced from Ito (1987):

> The data obtained by observations or measurements in astronomy …. and other sciences do not usually give exact values of the quantities in question. The error is the difference between the approximation and the exact value (p. 547).

Based on the general remarks referenced above, it is clear that this type of error calculation was a common practice in many scientific research studies. In this research, error has two forms. One is an overestimate while the other one is an underestimate. It means overestimate if the difference between predicted value and actual value was positive. Similarly, it means underestimate if the difference between predicted value and actual value was negative. For each RSMeans CCI city and DoD ACF location, error was calculated based on various surface interpolation methods. This was included in the empirical comparison phase. The following formulas were used to calculate relative and absolute errors for each surface interpolation method.

Data Source: RSMeans CCI

$$E_{k,j,i} = P_{k,j,i} - A_{k,j,i};$$

$$ER_{k,j,i} = E_{k,j,i};$$

$$EA_{k,j,i} = |E_{k,j,i}|$$

$$i\ (1\ to\ 649) = location\ ID$$

$$j\ (1\ to\ 6) = interpolation\ method\ ID$$

$$k\ (1\ to\ 5) = RSMeans\ CCI\ dataset\ year\ \ ID$$

$$ER_{k,j,i} = Relative\ Error\ for\ method\ j\ in\ year\ k$$

$$EA_{k,j,i} = Absolute\ Error\ for\ method\ j\ in\ year\ k$$

Data Source: DoD ACF

$$E_{k,j,i} = P_{k,j,i} - A_{k,j,i};$$

$$ER_{k,j,i} = E_{k,j,i};$$

$$EA_{k,j,i} = |E_{k,j,i}|$$

$$i\ (1\ to\ 337) = location\ ID$$

$$j\ (1\ to\ 6) = interpolation\ method\ ID$$

$$k\ (1\ to\ 5) = RSMeans\ CCI\ dataset\ year\ \ ID$$

$$ER_{k,j,i} = Relative\ Error\ for\ method\ j\ in\ year\ k$$

$$EA_{k,j,i} = Absolute\ Error\ for\ method\ j\ in\ year\ k$$

In these equations, $E_{k, j, i}$ depicts error for location i when using method j in year k. $P_{k, j, i}$ depicts estimated value for location i when using method j in year k while $A_{k, j, i}$ depicts actual value for location i. One important thing that needs to be pointed out is that $A_{k, j, i}$ is independent from any interpolation method but dependent on the year. $ER_{k, j}$ is the average relative error and $EA_{k, j}$ is the average absolute error when using method j in the year of k.

3.5 Error Calculation

The error calculation for CNN, NN, and ST AVG is very straightforward. Based on their definitions, with the help of an Excel spreadsheet, the errors can be calculated easily.

For a location, the nearest neighbor is the location with the shortest linear distance. For CNN, the nearest neighbor must be in the same state. However, for NN, the nearest neighbor could be in another state. Linear distance was employed since it can simplify the calculation for multiple geographic locations throughout the contiguous United States. Transportation distance which considers other factors such as highway and road travel could be a good topic for future research. With the help of the tool of “Near” in ArcGIS, the nearest neighbor for a location could be identified easily. For CNN, this tool was used at the state level. For NN, this tool was used at the national level. Each location has an EID, and with the help of the tool of “Join” in ArcGIS, the nearest neighbor for a location and its corresponding location factor could be related. Exporting the attributes table and then opening it in an Excel spreadsheet, the error for each location could be calculated. One important thing that should be addressed is that in CNN or NN, the nearest neighbor’s location factor is the predicted location factor.

For ST AVG, it is pretty straightforward to calculate the error. The predicted location factor is the average location factor for all the locations within a specific state. With the help of an Excel spreadsheet, the error for each location could be calculated.

However, for IDW, kriging, and spline, a great amount of efforts are needed to calculate the error since all of them are developed based upon complex functions.

However, this problem can be resolved with the help of the Geostatistical Analyst in ArcGIS.

The Geostatistical Analyst provides dynamic environment to help solve such spatial problems as improving estimation, assessing environmental risks, or predicting the existence of any geophysical element. In addition, the Geostatistical Analyst provides a wide variety of tools for exploration of spatial data, identification of data anomalies, and evaluation of error in prediction surface models, statistical estimation, and optimal surface creation.

Geostatistical Analyst can be used to create statistical interpolated continuous surfaces from measured samples. These surfaces represent a statistical estimation or prediction of where a certain phenomenon may occur. Not only are interpolated surfaces created, but also a wide range of analytical and exploratory tools are incorporated to extract useful information from the data. In addition, Geostatistical Analyst can provide a cost-effective, logical solution for analyzing a variety of data sets that would otherwise cost an enormous amount of time and money to accomplish.

In Geostatistical Analyst, there is a geostatistical wizard for getting the predicted and actual location factors. In this wizard, there are many parameters that need to be selected for each surface interpolation method, and different selection will lead to different predicted location factor values. Therefore, it is necessary to test the various combinations of these parameters to get the most accurate results. This test is defined as effect analysis in this research.

In effect analysis, the first step is to find out the most important parameters for IDW, kriging, and spline. The selection is based on three criteria. First, when developing the 2D distribution in ArcGIS for IDW, kriging, and spline method, some parameters need to be selected. If a parameter in 2D distribution was identical to the parameters in Geostatistical Analyst, it would be selected as a most important parameter. Second, the characteristic of the RSMeans CCI and DoD ACF value will be considered. Both of CCI and ACF value has the potential to be spatially autocorrelated. Third, the characteristic of IDW, kriging, and spline method will be considered. For example, lag is very important to kriging. The most important parameters for IDW, kriging, and spline are summarized in Table 2.

Table 2 Most Important Parameters for IDW, Kriging, and Spline

Method	Most Important Parameter
IDW	Power, Neighbors to Include
Kriging	Semivariogram Model, Number of Lags , Anisotropy, and Neighbors to Include
Spline	Kernel Function, Neighbors to Include

For IDW, power controls the significance of surrounding points on the interpolated value. A higher power results in less influence from distant points. Power can be any real number greater than zero, but the most reasonable results will be obtained using values from 0.5 to 3. The default value is 2. In this research, the powered is tested from 1 to 5 to cover the most possible value. Neighbors to include is an integer value specifying the number of nearest input sample points to be used to perform interpolation. The default is 12 points. In this research, the number of neighbors is tested for 5, 10, 15, 20, and 25 to cover the most possible range of value.

For kriging, semivariogram modeling is a key step between spatial description and spatial prediction. The main application of kriging is the prediction of attribute value at unsampled location. There are 11 semivariogram models in total but only 4 of them were tested. That is because only 4 are used in 2D distribution. The 4 semivariogram models are spherical, circular, exponential, and Gaussian. The selection of a lag size has important effects on the empirical semivariogram. For example, if the lag size is too large, short-range autocorrelation may be masked. If the lag size is too small, there may be many empty bins, and sample sizes within bins will be too small to get representative “averages” for bins. The number of lags was tested from 7 to 11 since less than 7 or greater than 11 could not produce error. The selection of number of lags is very flexible, which means the selection depends on the characteristic of the dataset. If another location factor dataset other than RSMeans CCI or DoD ACF was selected to perform this test, the number of lags may be different. Anisotropy is the property of being directionally dependent, which implies homogeneity in all directions. The option for anisotropy is “yes” or “no”. If yes, anisotropy will be considered in the calculation of predicted location factors value. For neighbors to include, it is similar to that in IDW, and the number of neighbors to include is tested for 5, 10, 15, 20, and 25.

For spline, kernel function is the selection of different spline type. For example, completely regularized spline yields a smooth surface and smooth first derivatives while spline with tension tunes the stiffness of the interpolant according to the character of the modeled phenomenon. There are five kernel functions available but only 2 of them are tested. That is because only 2 functions are used in 2D distribution. The 2 functions are completely regularized spline and spline with tension. For neighbors to include, it is similar to that in IDW, and the number of neighbors to include is tested from 5 to 25.

One index, the root-mean-square-error (RMSE), was selected to compare the result of the various combinations. That is because at the end of each wizard, the RMSE will be presented automatically. RMSE is a statistical measure of the magnitude of a varying quantity. It is especially useful when variants are positive and negative. The RMSE was calculated for every parameter combination for each surface interpolation method based upon both RSMeans CCI and DoD ACF dataset.

An Excel spreadsheet will be used to record RMSE value for each combination for each method and find out the combination with the lowest RMSE. Table 3 to Table 5 showed the RMSE calculation for IDW, kriging, and spline. One important thing that needs to be addressed is that calculation in these figures were based on 2009 RSMeans CCI dataset. This type of calculation will be performed from year 2005 to year 2009 for both RSMeans CCI and DoD ACF.

Table 3 RMSE Calculation for IDW Parameter Combination

RSMeans CCI 2009 IDW						
	Power					
Neighbors to Include		1	2	3	4	5
	5	4.224	4.264	4.366	4.476	4.575
	10	4.206	4.163	4.264	4.394	4.515
	15	4.362	4.187	4.245	4.371	4.498
	20	4.527	4.221	4.24	4.361	4.491
	25	4.654	4.263	4.246	4.359	4.489

Table 4 RMSE Calculation for Kriging Parameter Combination

RSMeans CCI 2009 Kriging								
	Circular	Spherical	Exponential	Guassian	Circular with Anisotropy	Spherical with Anisotropy	Exponential with Anisotropy	Guassian with Anisotropy
Lag 7 Neighbor 1	5.386	5.386	5.386	5.386	5.468	5.376	5.413	5.413
Lag 7 Neighbor 5	4.33	4.326	4.277	4.364	4.35	4.323	4.259	4.378
Lag 7 Neighbor 10	4.422	4.41	4.295	4.519	4.47	4.42	4.291	4.599
Lag 7 Neighbor 15	4.62	4.598	4.4	4.814	4.614	4.567	4.331	4.846
Lag 7 Neighbor 20	4.769	4.734	4.451	5.102	4.739	4.696	4.361	5.113
Lag 8 Neighbor 1	5.386	5.386	5.386	5.386	5.43	5.412	5.412	5.477
Lag 8 Neighbor 5	4.333	4.328	4.284	4.364	4.369	4.331	4.265	4.402
Lag 8 Neighbor 10	4.428	4.416	4.309	4.521	4.49	4.459	4.311	4.603
Lag 8 Neighbor 15	4.632	4.61	4.423	4.82	4.646	4.593	4.359	4.851
Lag 8 Neighbor 20	4.788	4.753	4.481	5.115	4.772	4.718	4.4	5.14
Lag 9 Neighbor 1	5.386	5.386	5.386	5.386	5.44	5.475	5.425	5.44
Lag 9 Neighbor 5	4.334	4.33	4.288	4.355	4.369	4.348	4.267	4.422
Lag 9 Neighbor 10	4.431	4.42	4.319	4.523	4.493	4.471	4.326	4.623
Lag 9 Neighbor 15	4.638	4.616	4.438	4.824	4.654	4.599	4.381	4.899
Lag 9 Neighbor 20	4.798	4.763	4.502	5.124	4.786	4.734	4.427	5.179
Lag 10 Neighbor 1	5.386	5.386	5.386	5.386	5.452	5.442	5.435	5.439
Lag 10 Neighbor 5	4.334	4.332	4.293	4.364	4.381	4.37	4.277	4.394
Lag 10 Neighbor 10	4.432	4.426	4.33	4.523	4.491	4.484	4.34	4.621
Lag 10 Neighbor 15	4.639	4.627	4.458	4.825	4.7	4.645	4.403	4.911
Lag 10 Neighbor 20	4.799	4.781	4.529	5.126	4.817	4.759	4.46	5.198
Lag 11 Neighbor 1	5.386	5.386	5.386	5.386	5.357	5.452	5.46	5.409
Lag 11 Neighbor 5	4.333	4.332	4.299	4.364	4.422	4.379	4.307	4.435
Lag 11Neighbor 10	4.43	4.427	4.343	4.522	4.555	4.486	4.367	4.626
Lag 11 Neighbor 15	4.635	4.629	4.48	4.823	4.74	4.686	4.433	4.978
Lag 11 Neighbor 20	4.794	4.785	4.559	5.122	4.866	4.792	4.485	5.248

Table 5 RMSE Calculation for Spline Parameter Combination

RSMeans CCI 2009 Spline						
	Kernel Function					
Neighbors to Include		1	2	3	4	5
	5	4.559	4.559	4.5	4.691	6.678
	10	5.242	4.552	4.514	5.068	4.998
	15	5.432	4.512	4.467	5.805	4.901
	20	5.559	4.506	4.466	6.004	4.834
	25	5.538	4.508	4.46	5.989	4.825

In these tables, the values are the RMSE values for each combination. The value with the yellow color is the lowest RMSE value. The parameters in the combination with the lowest RMSE value will be used in Geostatistical Analyst Wizard to calculate the predicted location factors and errors. The actual location factor, predicted location factors, and errors were stored in an attribute table associated with a new data layer. This attribute table could be exported and opened in an Excel spreadsheet.

3.6 Empirical Comparison

In this section, the spatial autocorrelation of the changes in RSMeans CCI value from year 2005 to year 2009 will be evaluated based on Global Moran's I Test in ArcGIS software. In addition, the spatial autocorrelation of the DoD ACF value in the year of 2006 and the spatial autocorrelation of the changes in DoD ACF value from year 2005 to year 2009 were evaluated to cross-validate the validity of the proximity-based interpolation method.

Six surface interpolation methods, which include CNN, NN, ST AVG, IDW, kriging, and spline will be evaluated based on the following techniques:

- Comparison of Overestimates and Underestimates
- Best Performance Comparison
- Comparison of Error Percentage
- Descriptive Statistics

3.6.1 Global Moran's I Test

According to Bivand (1998), there are two ways to measure the spatial autocorrelation. One is the global spatial autocorrelation while the other is local spatial autocorrelation. Global spatial autocorrelation is a measure of the overall clustering of the data and it yields only one statistic to summarize the whole study area. However, if there is no global autocorrelation or no clustering, there is still a way to find clusters at a local level using local spatial autocorrelation. For this research, only global spatial autocorrelation was considered. That is because the aim of this research is to measure the overall clustering of the changes in RSMeans CCI value and DoD ACF value from year 2005 to year 2009 and the overall clustering of DoD ACF value in year 2006, namely measure the clustering at the national level. The Global Moran's I test, which is an available function in the ArcGIS software, was employed to evaluate the degree of the spatial autocorrelation. The Global Moran's I test was specifically selected since it was an established method for measuring global spatial autocorrelation. According to Banerjee et al. (2004), two standard statistics can be used to measure the strength of spatial autocorrelation, including Moran's I and Geary's C. Moran's I test is available in ArcGIS while Geary's C is not available. Based on this, a possible future research topic could be to test spatial autocorrelation by Geary's C statistics and compare the results with those of the Global Moran's I test.

As mentioned earlier, the Global Moran's I test was conducted at the national level. After running the Global Moran's I test in the ArcGIS software, only 1 result was displayed. According to Bolstad (2005):

> Moran’s I values approach a value of +1 in areas of positive spatial correlation, meaning large values tend to be clumped together, and small values clumped together. Values near zero occur in areas of low spatial correlation (pg. 412).

As mentioned in Chapter 2, Global Moran’s I test result includes Moran’s I index, Z score, and p-value. For Moran's index value, an index value approaches +1.0 means positive spatial autocorrelation (clustering) while an index value approaches -1.0 means negative spatial autocorrelation (dispersion). However, without looking at the statistical significance there is no basis for knowing if the observed pattern is just one of many possible versions of randomness. Therefore, there is still a need to check the Z score and p-value. The Z score is a test of statistical significance that helps us decide whether or not to reject the null hypothesis. In other words, the Z-score evaluated if the null hypothesis should be rejected. The p-value is the probability that we have falsely rejected the null hypothesis. When the p-value is small and the absolute value of the Z score is large enough that it falls outside of the desired confidence level, the null hypothesis can be rejected. In order to reject the null hypothesis with statistically significant confidence, the Z-score must be less than –1.96 or greater than 1.96 when using a 95% confidence level (0.05 significance level). If evidence of significant spatial autocorrelation results from the Global Moran’s I tests, it will ultimately substantiate the validity of proximity based spatial interpolation method.

3.6.2 Comparison of Overestimates and Underestimates

A comparison of overestimates and underestimates for each surface interpolation method was conducted. For RSMeans CCI dataset, an Excel spreadsheet was created to show the actual number of overestimates and underestimates for each surface interpolation method from year 2005 to year 2009. DoD ACF dataset was used to cross-validate the result obtained from RSMeans CCI dataset. For DoD ACF dataset, one Excel spreadsheet was also created to show the actual number of overestimates and underestimates for each surface interpolation method from year 2005 to year 2009. One important thing that needs to be pointed out is that relative error was used in the comparison of overestimates and underestimates. The comparison over overestimates and underestimates was performed to test whether a pattern could be observed. It could possibly assist future research which involves location adjustment methods if an obvious pattern was observed.

3.6.3 Best Performance Comparison

In order to evaluate the six surface interpolation methods, a best performance comparison was also conducted. As mentioned earlier, there were 649 RSMeans CCI cities and 337 DoD ACF locations from which error was calculated. As a matter of fact, each city or each location will produce an error value depending on which surface interpolation method was used, which year's data was selected, and which dataset was chosen. In order to find which surface interpolation method works the best, performance was quantified out of the 6 surface interpolation methods. In a series of Excel spreadsheets, a count of this measurement of performance was performed. Absolute error was employed in this technique.

3.6.4 Comparison of Error Percentages

Another technique used for analysis is the comparison of error percentages. Different levels of error were classified as the following: very low, low, medium, high, and very high.

If the error is between 0 and 1%, it was concluded that the error is very low. If the error is between 1% and 3%, it was concluded that the error is low. If the error is between 3% and 5%, it was concluded that the error is medium. In addition, if the error is between 5% and 10%, it was concluded that the error is high. Finally, if the error is greater than 10%, it was concluded that the error is very high. A count of how many cities or locations were included in these levels and a corresponding percentage were calculated and displayed in a series of Excel spreadsheet. Absolute error was used in this technique. This comparison was chosen to evaluate which surface interpolation method can produce the highest accuracy. High accuracy is defined as more errors in very low and low category.

3.6.5 Descriptive Statistics

In this research, descriptive statistics include mean, median, standard deviation, were calculated for the six interpolation methods from year 2005 to year2009. Absolute error values were considered in all calculations. Various types of tables and charts were developed to summarize and compare the statistics. Descriptive statistical comparisons were used in this research to determine whether a surface interpolation method could be statistically proven to outperform another one.

Moreover, the average value of mean, median, and standard deviation value for each method were calculated and compared. In order to find out which surface interpolation method is the best, a ranking method was employed. The rank for the method with the lowest average value of mean, median, or standard deviation is one, while the rank for the method with the largest average value of mean, median, or standard deviation is six, considering there are six surface interpolation methods. The method with the lowest rank value is the best surface interpolation method.

3.7 Pattern Comparison

In pattern comparison, IDW, kriging, and spline, which can produce smooth surface function, were compared both in 2D and 3D format. This analysis and comparison is the initial step for developing a 3D surface construction cost function.

3.7.1 2D Comparison

2D format of the RSMeans CCI and DoD ACF distribution was developed for each smooth surface interpolation method in each year.

3.7.2 3D Comparison

3D format of the RSMeans CCI and DoD ACF distribution was developed for each smooth surface interpolation method in each year.

CHAPTER 4.0 ANALYSIS AND COMPARISON

4.1 Overview

This section will present the results from analysis and comparison. As an overview, the following analyses and comparisons were performed.

- Global Moran's I Test
- Error Calculation
- Comparison of Overestimate and Underestimate
- Best Performance Comparison
- Comparison of Error Percentages
- Descriptive Statistics
- Pattern Comparison

4.2 Global Moran's I Test

In former research conducted by Martinez (2010), Global Moran's I test was conducted for 2006 RSMeans City Cost Index (CCI) at both national and state level. According to their result, there was evidence of positive, statistically significant spatial autocorrelation between proximity and RSMeans CCI values, at both national level and several states.

The research mentioned above was conducted based on the RSMeans CCI value of the 649 cities in the contiguous United States in the year of 2006. However, the CCI value is published every year and most cities' CCI value changes from year to year. Although some cities keep the same CCI value for several years, each city's value changes in the long term. Therefore, in order to better support the validity of the proximity-based interpolation location adjustment method, it is essential to conduct spatial analysis for the changes in temporal CCI value, namely from year 2005 to year 2009. The changes in temporal CCI value were defined as the changes between 2005 and 2006, between 2006 and 2007, between 2007 and 2008, and between 2008 and 2009.

The null hypothesis for this analysis is: for the changes in CCI value from year 2005 to year 2009, there is no spatial clustering (strong spatial autocorrelation) associated with the 649 cities. In addition, the analysis assumptions are described as follows:

- All locations impact/influence all other locations
- The farther away a feature is, the smaller impact it has, but the influence does not drop off quickly
- The distribution of features is not potentially biased due to sampling design or an imposed aggregation scheme

The Global Moran's I test's results were displayed in dialogues by the ArcGIS software. All the dialogues were shown in Exhibit A1 in Appendix A. The Global Moran's I test results were summarized in Table 6.

Table 6 Spatial Autocorrelation Analysis Summary for the Changes in CCI Value

Year	Moran's I Index	Z Score	P-value	Clustered
2005-2006	0.233824	12.208802	0.000001	Yes
2006-2007	0.491619	25.576931	0.000001	Yes
2007-2008	0.199904	10.504333	0.000001	Yes
2008-2009	0.197741	10.527437	0.000001	Yes

From Table 6, it is clear that Moran's I index for the four tests are positive and we can infer that positive spatial autocorrelation may exist. However, without looking at statistical significance, there is no basis for knowing if the observed pattern is just one of many possible versions of random. Therefore, there is still a need to look at the Z score (standard deviation) and p-value. To give an example to explain how to use Z score and p-value: the corresponding Z score values when using a 95% confidence level (this is the minimum confidence level in which we can reject the null hypothesis) are -1.96 and + 1.96 standard deviations. The p-value associated with a 95% confidence level is 0.05, which means that there is less than 5% likelihood that this clustered pattern could be the result of random change. If the Z-score is between -1.96 and + 1.96, the p-value will be larger than 0.05, and then we cannot reject the null hypothesis. If the Z-score falls outside the range (-1.96 to +1.96) then we can reject the null hypothesis. Similarly, in a 99% confidence level, the range is -2.58 to +2.58. The p-value of the four tests is 0.000001 and at the same time the Z-score is greater than 2.58. Therefore, the conclusion is that the changes in CCI value for the 649 cities from year 2005 to year 2009 are spatially autocorrelated. In addition, since the Z-score of the four tests is much greater than 2.58, we can conclude that the changes in CCI value for the 649 cities from year 2005 to year 2009 are highly spatially autocorrelated.

In addition, with the help of the ArcGIS software, it is possible to visualize the spatial patterns of the changes in CCI value from year 2005 to year 2009. A series of maps which show the changes in CCI value between year 2005 and 2006, year 2006 and 2007, year 2007 and 2008, and year 2008 and 2009 can be created by using the thematic map techniques.

One important thing that needs to be pointed out is that the changes in CCI value will be shown in the format of percentage. Therefore, it is necessary to perform normalization. For example, the changes in CCI value between year 2007 and year 2008 will be normalized by the CCI value in the year of 2007. For the classification, the range is defined as follows:

- -1% to +1% (no change)
- +1.1% to 3% (small increase); -3% to -1.1% (small decrease)
- + 3.1 to 5% (medium increase); -5% to -3.1% (medium decrease)
- Greater than 5% (large increase); less than -5% (large decrease)

All the maps that show the spatial patterns of the changes in CCI value from year 2005 to year 2009 are displayed in Exhibit B1 in Appendix B.

For this research, besides the RSMeans CCI dataset, there is another dataset, namely Area Cost Factor of the Department of Defense (DoD ACF). With the help of the

DoD ACF dataset, it is possible to cross-validate the proximity-based location adjustment method.

In the research conducted by Martinez (2010), the 2006 RSMeans CCI dataset was used to assess the validity of the proximity-based interpolation method. In order to be comparable with the 2006 RSMeans CCI dataset, the 2006 DoD ACF dataset was selected to cross-validate the validity of the proximity-based interpolation method. For DoD ACF dataset, Global Moran's I test were conducted nationally and then for each individual state, which created both national and state level results.

At the national level, the null hypothesis is that there is no spatial clustering (strong spatial autocorrelation) associated with the 337 DoD ACF locations. The results of the Global Moran's I test at the national level were displayed in Figure 9 and Figure 10.

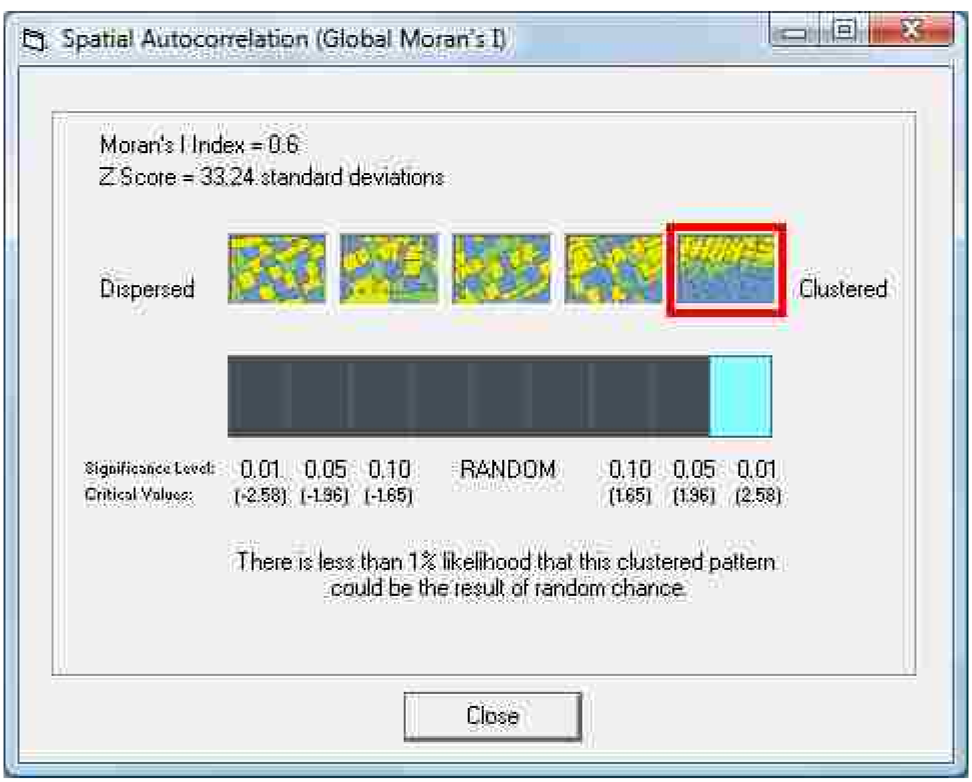

Figure 9 Global Moran's I Test Dialogue for 2006 DoD ACF

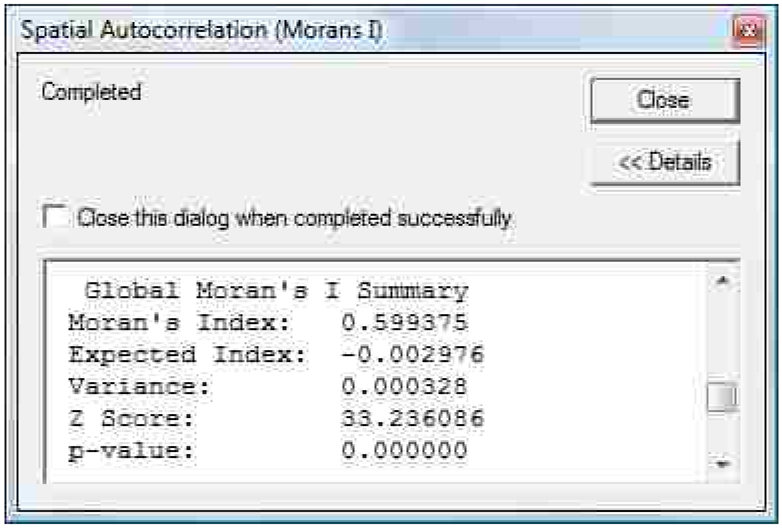

Figure 10 Global Moran's I Test Summary for 2006 DoD ACF

The results above returned a positive Moran's I index. With a Z score of 33.236086 and a p-value of 0.000001, we can reject the null hypothesis. The Z-score of 0.6 indicated that the DoD ACF values are spatially clustered across the contiguous United States.

Table 7 summarizes the Global Moran's I test at the state level. There were 15 instances in which the test did not successfully determine a Moran's Index or a Z score. This was primarily because of the lack of enough input data for the test. In another words, there were not enough cities within the same state or district to effectively measure the level of spatial autocorrelation. As mentioned earlier, there are only 337 ACF locations across the contiguous United States. That means for each individual state, there are fewer locations to be assessed than CCI cities. These instances include Connecticut, Delaware, District of Columbia, Idaho, Iowa, Minnesota, Montana, Nebraska, New Hampshire, Oregon, Rhode Island, South Dakota, Vermont, West Virginia, and Wyoming. For these instances, a "not applicable" was noted in Table 7.

Table 7 Global Moran’s I Tests Results for 2006 ACF at State Level

Global Moran’s I Test Results for 2006 ACF at State Level				
State/District	Moran's Index	Clustered	Z Score	Significant
ALABAMA (AL)	-0.141971		0.248385	
ARIZONA (ZA)	0.62	YES	2.04	YES
ARKANSAS (AR)	-0.333333		0.000000	
CALIFORNIA (CA)	0.009745		0.728871	
COLORADO (CO)	0.393035	YES	3.312554	YES
CONNECTICUT (CT)	NOT APPLICABLE			
DELAWARE (DE)	NOT APPLICABLE			
DISTRICT OF COLUMBIA (DC)	NOT APPLICABLE			
FLORIDA (FL)	0.391831	YES	1.673357	NO
GEORGIA (GA)	-0.049397		0.773887	
IDAHO (ID)	NOT APPLICABLE			
ILLINOIS (IL)	0.860026	YES	3.523186	YES
INDIANA (IN)	-0.056419		1.234245	
IOWA (IA)	NOT APPLICABLE			
KANSAS (KS)	-0.158771		0.235143	
KENTUCKY (KY)	-0.120536		0.776602	
LOUISIANA (LA)	0.895826	YES	2.833735	YES
MAINE (ME)	-0.084507		0.289794	
MARYLAND (MD)	0.049662		0.852839	
MASSACHUSETTS (MA)	-0.240687		-0.328569	
MICHIGAN (MI)	1.000000	YES	1.414214	NO
MINNESOTA (MN)	NOT APPLICABLE			
MISSISSIPPI (MS)	0.607908	YES	1.998708	NO
MISSOURI (MO)	0.073544	YES	1.522473	NO
MONTANA (MT)	NOT APPLICABLE			
NEBRASKA (NE)	NOT APPLICABLE			
NEVADA (NV)	0.455838	YES	1.592492	NO
NEW HAMPSHIRE (NH)	NOT APPLICABLE			
NEW JERSEY (NJ)	-0.090218		0.583692	
NEW MEXICO (NM)	-0.035241		1.074521	
NEW YORK (NY)	0.680099	YES	3.311703	YES
NORTH CAROLINA (NC)	-0.044788		0.432661	
NORTH DAKOTA (ND)	1.000000	YES	1.414214	NO
OHIO (OH)	0.807487	YES	1.414214	NO
OKLAHOMA (OK)	-0.039477		0.763537	
OREGON (OR)	NOT APPLICABLE			
PENNSYLVANIA (PA)	0.193870	YES	3.843932	YES
RHODE ISLAND (RI)	NOT APPLICABLE			
SOUTH CAROLINA (SC)	0.555415	YES	2.458957	YES
SOUTH DAKOTA (SD)	NOT APPLICABLE			
TENNESSEE (TN)	0.553724	YES	1.373528	NO
TEXAS (TX)	0.031262	YES	1.488646	NO
UTAH (UT)	0.069930	YES	2.756862	YES
VERMONT (VT)	NOT APPLICABLE			
VIRGINIA (VA)	0.628024	YES	5.646973	YES
WASHINGTON (WA)	0.036932	YES	2.486508	YES
WEST VIRGINIA (WV)	NOT APPLICABLE			
WISCONSIN (WI)	-0.271201		-0.085375	
WYOMING (WY)	NOT APPLICABLE			

From Table 7, it is clear that no test showed obvious evidence of negative spatial autocorrelation, namely significant dispersion pattern. That is because there are no instances with a large negative Moran's I index value with a Z score less than -1.96. Furthermore, 19 of 34 states or district showed result of positive Moran's I index value. In addition, 10 of these 19 states showed results of significant spatial autocorrelation. The 10 highlighted states in Table 8 showed the result of positive, statistically significant spatial autocorrelation. For these 10 states, there was evidence to reject the hull hypothesis which stated that DoD ACF values were not spatially autocorrelated. Therefore, there was evidence of positive, statistically significant autocorrelation for DoD ACF values in some states. These states were compiled and were shown in Table 8.

Table 8 Positive Moran's I Index and Significant Z-Score States for 2006 ACF

State/District	Moran's Index	CLUSTERED	Z Score	SIGNIFICANT
ARIZONA	0.62	YES	2.04	YES
COLORADO	0.393035	YES	3.312554	YES
ILLINOIS	0.860026	YES	3.523186	YES
LOUISIANA	0.895826	YES	2.833735	YES
NEW YORK	0.680099	YES	3.311703	YES
PENNSYLVANIA	0.193870	YES	3.843932	YES
SOUTH CAROLINA	0.555415	YES	2.458957	YES
UTAH	0.069930	YES	2.756862	YES
VIRGINIA	0.628024	YES	5.646973	YES
WASHINGTON	0.036932	YES	2.486508	YES

In the research conducted by Martinez (2010), which was based on 2006 RSMeans CCI dataset, there were also no instances of a negative Moran's I index with a Z-sore much less than -1.96. In addition, 24 of 46 states showed results of positive Moran's I index. Furthermore, 19 of these 24 states showed results of positive Moran's I index, and Z-scores greater than 1.96. The Global Moran's I test for 2006 ACF and CCI at state level were compared in Table 9.

Table 9 Comparison of Global Moran's I Test for 2006 ACF and CCI at State Level

Global Moran's I Test Results for 2006 ACF and CCI at State Level								
	2006 ACF				2006 CCI			
State/ District	Moran's Index	Clustered	Z Score	Significant	Moran's Index	Clustered	Z Score	Significant
AL	-0.141971		0.248385		-0.106245		-0.476793	
ZA	0.62	YES	2.04	YES	-0.178088		-0.473334	
AR	-0.333333		0.000000		-0.115634		-0.426931	
CA	0.009745		0.728871		0.820966	YES	14.042334	YES
CO	0.393035	YES	3.312554	YES	-0.115919		-0.560558	
CT	NOT APPLICABLE				0.041086	YES	1.825827	NO
DE	NOT APPLICABLE				NOT APPLICABLE			
DC	NOT APPLICABLE				NOT APPLICABLE			
FL	0.391831	YES	1.673357	NO	0.101853	YES	2.225169	YES
GA	-0.049397		0.773887		-0.048505		0.773887	
ID	NOT APPLICABLE				-0.026602		0.861344	
IL	0.860026	YES	3.523186	YES	0.498979	YES	8.72939	YES
IN	-0.056419		1.234245		-0.028556		0.875952	
IA	NOT APPLICABLE				0.050944	YES	2.459729	YES
KS	-0.158771		0.235143		-0.047728		0.669308	
KY	-0.120536		0.776602		0.14384	YES	3.418235	YES
LA	0.895826	YES	2.833735	YES	0.028915	YES	1.928826	NO
ME	-0.084507		0.289794		-0.14854		-0.491829	
MD	0.049662		0.852839		0.057158	YES	1.255529	NO
MA	-0.240687		-0.328569		0.135946	YES	2.897366	YES
MI	1.000000	YES	1.414214	NO	0.563173	YES	5.419719	YES
MN	NOT APPLICABLE				0.323685	YES	2.518224	YES
MS	0.607908	YES	1.998708	NO	-0.054603		0.828869	
MO	0.073544	YES	1.522473	NO	-0.019038		0.754391	
MT	NOT APPLICABLE				-0.076468		0.53581	
NE	NOT APPLICABLE				0.016878	YES	1.21904	NO
NV	0.455838	YES	1.592492	NO	0.031196	YES	0.658152	NO
NH	NOT APPLICABLE				0.313499	YES	3.296673	YES
NJ	-0.090218		0.583692		0.093083	YES	2.015206	YES
NM	-0.035241		1.074521		0.022073	YES	2.035047	YES
NY	0.680099	YES	3.311703	YES	0.625865	YES	8.04446	YES
NC	-0.044788		0.432661		0.071429	YES	0.739861	NO
ND	1.000000	YES	1.414214	NO	-0.022863		0.428601	
OH	0.807487	YES	1.414214	NO	0.14973	YES	3.909278	YES
OK	-0.039477		0.763537		-0.030998		0.940727	
OR	NOT APPLICABLE				0.061146	YES	2.3324	YES
PA	0.193870	YES	3.843932	YES	0.144825	YES	5.507111	YES
RI	NOT APPLICABLE				NOT APPLICABLE			
SC	0.555415	YES	2.458957	YES	-0.128443		0.157375	
SD	NOT APPLICABLE				-0.320901		-1.823682	
TN	0.553724	YES	1.373528	NO	-0.201274		-0.784914	
TX	0.031262	YES	1.488646	NO	-0.012473		0.489588	
UT	0.069930	YES	2.756862	YES	-0.080554		1.387565	
VT	NOT APPLICABLE				-0.028159		1.033033	
VA	0.628024	YES	5.646973	YES	0.5849	YES	5.595534	YES
WA	0.036932	YES	2.486508	YES	0.326007	YES	3.852031	YES
WV	NOT APPLICABLE				0.085484	YES	2.788805	YES
WI	-0.271201		-0.085375		0.323674	YES	4.112459	YES
WY	NOT APPLICABLE				-0.089586		0.353008	

To sum up, the Global Moran's I test for 2006 DoD ACF, both at the national and state level, acts as a cross-validation supporting proximity-based interpolation methods. However, similar to the RSMeans CCI dataset, the DoD ACF value is published every year and most locations' ACF value changes. Although some cities keep the same ACF value for several years, each location's ACF value changes in the long term. Therefore, in order to better support the validity of the proximity-based location adjustment method, it is essential to conduct spatial analysis of the changes in temporal ACF value (from year 2005 to year 2009), which is similar to RSMeans CCI dataset. Therefore, the null hypothesis and analysis assumption are not repeated here.

The results of the Global Moran's I test for the changes in ACF value from year 2005 to year 2009 were displayed in dialogues by the ArcGIS software. All the dialogues were shown in Exhibit A2 in Appendix A. The test results were summarized in Table 10.

Table 10 Spatial Autocorrelation Analysis Summary for the Changes in ACF Value

Year	Moran's I Index	Z Score	P-value	Clustered
2005-2006	0.208276	11.684750	0.000001	Yes
2006-2007	0.112900	6.564342	0.000001	Yes
2007-2008	0.114060	6.751173	0.000001	Yes
2008-2009	0.083788	4.790619	0.000002	Yes

According to the explanation for Global Moran's I test result mentioned above, we can conclude that the changes in ACF values from year 2005 to year 2009 are highly spatially autocorrelated. This result cross-validates the validity of the current proximity-based location adjustment method in the construction industry. Therefore, the underlying assumption for the current proximity-based location adjustment methods has been completely validated by the internal cost information.

Similar to the CCI dataset, the changes in ACF value will be shown in the format of percentage. Therefore, the normalization will also be employed. For example, the changes in ACF value between year 2007 and 2008 will be normalized by the ACF value of 2007. For the classification, the range is defined as follows:

- -1% to +1% (no change)
- +1.1% to 3% (small increase); -3% to -1.1% (small decrease)
- + 3.1 to 5% (medium increase); -5% to -3.1% (medium decrease)
- Greater than 5% (large increase); less than -5% (large decrease)

All the maps to show the spatial patterns of the changes in ACF value from year 2005 to year 2009 are displayed in Exhibit B2 in Appendix B.

4.3 Error Calculation

The underlying concept for error calculation is that for a location with a location factor, we assume there is no location factor and a predicted location factor can be developed by interpolation. The predicted location factor minus the actual one is the error.

As mention in section 3.4, the error calculation for conditional nearest neighbor (CNN), nearest neighbor (NN), and state average (ST AVG) can be achieved with the help of an Excel spreadsheet. Inverse distance weighted (IDW), kriging, and spline are very complex functions and then it is too time-consuming to calculate the error manually. However, this problem can be solved with the Geostatistical Analyst in ArcGIS

As mentioned in Chapter 3, in order to get the most accurate error result for IDW, kriging, and spline, it is necessary to perform effect analysis to get the parameter combination with the lowest root-mean-square-error (RMSE) value.

With the help of the effect analysis, the combinations of parameters which will lead the lowest RMSE for IDW, kriging, and spline were summarized in Table 11 to 13. The effect analysis was performed for both RSMeans CCI and DoD ACF.

Table 11 Lowest RMSE Parameter Combination for IDW

Dataset	Power	Neighbors to Include	RMSE
CCI 2005	2	10	4.438
CCI 2006	2	10	4.416
CCI 2007	2	10	4.084
CCI 2008	2	10	4.053
CCI 2009	2	10	4.163
ACF 2005	2	5	6.398
ACF 2006	2	15	5.918
ACF 2007	2	25	6.353
ACF 2008	2	25	6.439
ACF 2009	2	25	6.309

Table 12 Lowest RMSE Parameter Combination for Kriging

Dataset	SM	Number of Lags	Anisotropy	Neighbors to Include	RMSE
CCI 2005	Exponential	7	Yes	10	4.561
CCI 2006	Exponential	7	Yes	10	4.57
CCI 2007	Exponential	7	Yes	5	4.18
CCI 2008	Exponential	7	Yes	5	4.125
CCI 2009	Exponential	7	Yes	5	4.259
ACF 2005	Exponential	7	Yes	25	6.311
ACF 2006	Exponential	9	Yes	25	6.207
ACF 2007	Exponential	7	No	15	6.831
ACF 2008	Exponential	7	Yes	25	6.824
ACF 2009	Exponential	7	Yes	25	6.637

Table 13 Lowest RMSE Parameter Combination for Spline

Dataset	Kernel Function	Neighbors to Include	RMSE
CCI 2005	Tension	25	4.633
CCI 2006	Regularized	5	4.593
CCI 2007	Regularized	5	4.298
CCI 2008	Tension	25	4.372
CCI 2009	Tension	20	4.506
ACF 2005	Tension	15	6.298
ACF 2006	Regularized	15	6.003
ACF 2007	Regularized	15	6.414
ACF 2008	Regularized	15	6.526
ACF 2009	Regularized	15	6.372

After getting the lowest RMSE parameter combination for IDW, kriging, and spline surface interpolation method, Geostatistical Analyst Wizard was used to get the error for each method based on both CCI and ACF value from year 2005 to year 2009. The errors were exported in Excel spreadsheets for later analyses.

4.4 Comparison of Overestimate and Underestimate

A comparison of overestimate and underestimate for CNN, NN, ST AVG, IDW, kriging, and spline surface interpolation methods was conducted for RSMeans CCI dataset from year 2005 to year 2009. Error was presented in the form of the difference between estimated values minus actual values for each of the 649 CCI cities, which created positive, negative, and zero differences. To state it simply, overestimate, underestimate, and accurate estimate are produced.

Error classifications, which include overestimates, underestimates, and accurate estimates, were calculated for each interpolation method for both CCI and ACF dataset from year 2005 to year 2009. Overestimate means that the relative error is greater than zero while underestimate means that the relative error is less than zero. Accurate estimate means the error is zero. These results were shown Exhibit C1 (based on RSMeans CCI) and Exhibit C2 (based on DoD ACF) in Appendix C.

Analyzing the results reported in each method can help us understand these tables. For example, let us analyze Table C1.4. At the bottom of each column, it shows that a total of 649 observations were calculated. For the CNN method, out of the 649 observations, 312 were underestimated, 325 were overestimated, 11 were accurately estimated, and 1 was inconclusive. For the ACF dataset, let us analyze Table C2.4. At the bottom of each column, it shows that a total of 337 observations were calculated. For the CNN method, out of the 337 observations, 100 were underestimated, 97 were overestimated, 140 were accurately estimated, and nothing was inconclusive.

4.5 Best Performance Comparison

Errors of CNN, NN, ST AVG, IDW, kriging, and spline methods were compared with each other's, namely a bi-variable comparison of these methods were performed at the national level. Absolute values of error were calculated to quantify the performance. For these calculations, two series of spreadsheets were developed. One series was used to determine which method provided more accuracy for each of the 649 CCI cities while the other one was used to determine which method provided more accuracy for each of the 337 ACF locations. The results are shown in Exhibit D1 (based on RSMeans CCI) and Exhibit D2 (based on DoD ACF) in Appendix D.

4.6 Comparison of Error Percentages

Continuing with the six surface interpolation methods comparison, the actual count of how many cities or locations were included in very low, low, medium, high, and very high levels and corresponding percentage were showed in Exhibit E1 (based on RSMeans CCI) and Exhibit E2 (based on DoD ACF) in Appendix E. This type of comparison was conducted for RSMeans CCI dataset from year 2005 to year 2009. In addition, the results obtained from the DoD ACF dataset from year 2005 to year 2009 were used to cross-validate the results obtained from RSMeans CCI dataset.

4.7 Descriptive Statistics

Median, mean, and standard deviation of the absolute error values for all six surface interpolation methods were calculated and summarized in Exhibit F1 (based on RSMeans CCI) and Exhibit F2 (based on DoD ACF) in Appendix F. In addition, the average value of median, mean, and standard deviation of the six surface interpolation methods based on RSMeans CCI dataset from year 2005 to year 2009 were summarized and compared in Table 14. Moreover, the average value of median, mean, and standard deviation of the six surface interpolation methods based on DoD ACF dataset from year 2005 to year 2009 were summarized and compared in Table 16. With the help of ranking method, the rank for each surface interpolation method was developed and displayed in Table 15 and Table 17.

Table 14 Average Value of Descriptive Statistics for Each Method Based on CCI

RSMeans CCI Dataset												
	CNN						NN					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	3.32	3.34	3.09	2.93	3.09	3.15	4.09	4.14	3.72	3.81	3.89	3.93
Mean	3.17	3.17	2.96	2.89	3.06	3.05	3.79	3.79	3.55	3.55	3.73	3.68
Median	2.10	2.00	2.00	1.90	2.20	2.04	2.40	2.40	2.40	2.40	2.60	2.44
	ST AVG						IDW					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	3.76	3.77	3.63	3.66	3.64	3.69	2.96	2.93	2.78	2.80	2.82	2.86
Mean	3.72	3.80	3.50	3.41	3.39	3.56	3.31	3.31	3.00	2.94	3.06	3.12
Median	2.50	2.60	2.30	2.20	2.10	2.34	2.40	2.50	2.20	2.20	2.30	2.32
	Kriging						Spline					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	2.98	2.98	2.82	2.82	2.87	2.89	3.09	3.10	2.89	3.00	3.04	3.02
Mean	3.46	3.47	3.08	3.01	3.15	3.23	3.45	3.39	3.16	3.14	3.26	3.28
Median	2.60	2.70	2.20	2.20	2.40	2.42	2.50	2.50	2.30	2.30	2.40	2.40

Table 15 Ranking for Each Method Based on CCI

Method	STDEV	Mean	Median	Score	Rank
CNN	4	1	1	6	2
NN	6	6	6	18	6
ST AVG	5	5	3	13	5
IDW	1	2	2	5	1
Kriging	2	3	4	9	3
Spline	3	4	5	12	4

Table 16 Average Value of Descriptive Statistics for Each Method Based on ACF

DoD ACF Dataset												
	CNN						NN					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	5.25	5.10	5.60	5.71	5.90	5.51	8.21	6.77	7.46	7.50	6.79	7.35
Mean	4.29	3.57	3.89	4.02	3.97	3.95	6.90	5.76	5.97	6.29	5.88	6.16
Median	2.00	1.00	2.00	2.00	1.00	1.60	4.00	3.00	3.00	4.00	4.00	3.60
	ST AVG						IDW					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	3.89	4.57	4.70	4.68	4.89	4.55	4.92	4.66	4.91	4.85	4.92	4.85
Mean	4.59	4.87	5.07	5.36	4.92	4.96	4.16	3.65	4.04	4.25	3.96	4.01
Median	3.50	4.00	3.90	4.30	3.20	3.78	2.50	2.10	2.50	3.00	2.10	2.44
	Kriging						Spline					
	05	06	07	08	09	Mean	05	06	07	08	09	Mean
STDEV	4.52	4.77	4.75	4.77	4.87	4.74	4.56	4.61	4.89	4.79	4.84	4.74
Mean	4.41	3.98	4.91	4.89	4.52	4.54	4.25	3.86	4.16	4.44	4.15	4.17
Median	2.90	2.40	3.90	3.80	3.00	3.20	2.60	2.50	2.70	3.10	2.40	2.66

Table 17 Ranking for Each Method Based on ACF

Method	STDEV	Mean	Median	Score	Rank
CNN	4	1	1	6	1
NN	5	6	5	16	6
ST AVG	1	5	6	12	5
IDW	3	2	2	7	2
Kriging	2	4	4	10	4
Spline	2	3	3	8	3

4.8 Pattern Comparison

Patten comparison is the second phase of this research. In this section, both 2D and 3D distribution for IDW, kriging, and spline smooth surface interpolation method were developed. 2D distribution was achieved with the help of Spatial Analyst in ArcGIS while 3D distribution was developed with the aid of the ArcScene. When using Spatial Analyst, the parameters selection was based on the result of effect analysis. In addition, the cell size of 10,000 meters by 10,000 meters was used for each method, considering that the contiguous United State is very large. The selection of cell size is a subject process and it might be a good topic for future study.

4.8.1 2D Distribution Visualization

The 2D distribution visualization for IDW, kriging, and spline surface interpolation methods was showed in Exhibit G1 (based on RSMeans CCI) and Exhibit G2 (based on DoD ACF) in Appendix G.

4.8.2 3D Distribution Visualization

The 3D distribution visualization for IDW, kriging, and spline were performed. One important thing that needs to be addressed is that for the 3D comparison, only 2009 RSMeans CCI dataset and 2009 DoD ACF dataset were used since the aim of this type comparison is to test whether these two cost dataset have the same trend.

All the 3D distribution models were showed in Figure 11 to 13. In these figures, red color models are ACF while green color models are CCI. There is a gap between CCI and ACF, which means that ACF values are higher. As stated before, 3D distribution visualization is to test whether CCI and ACF model have the same trend. In these three figures, the x-axis is the longitude while y-axis is the CCI or ACF value.

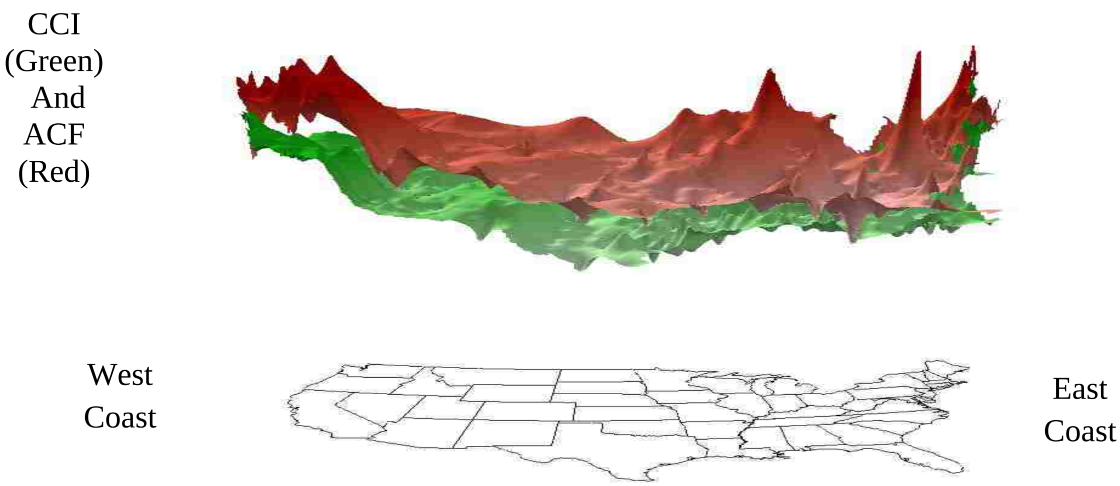

Figure 11 3D IDW Distribution for 2009 CCI and ACF

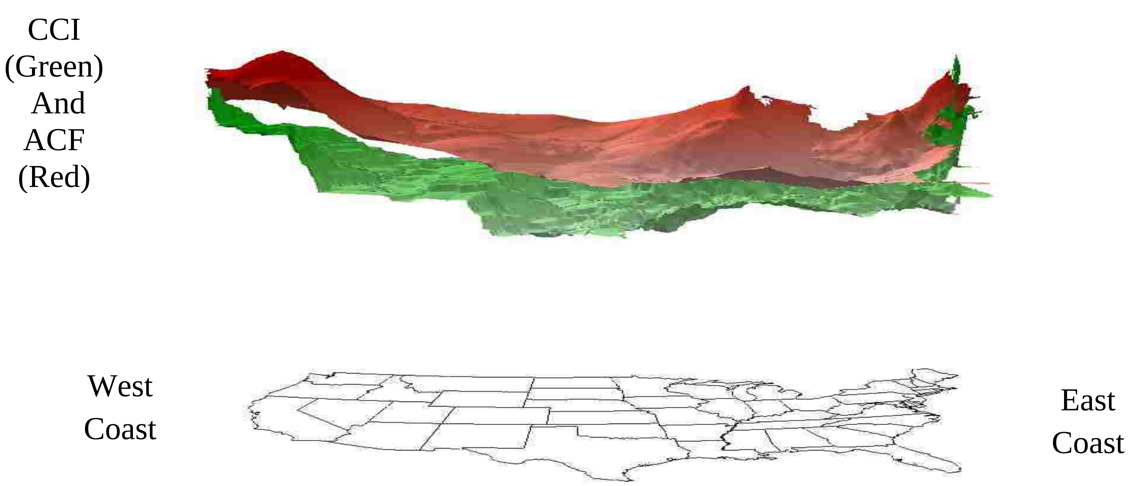

Figure 12 3D Kriging Distribution for 2009 CCI and ACF

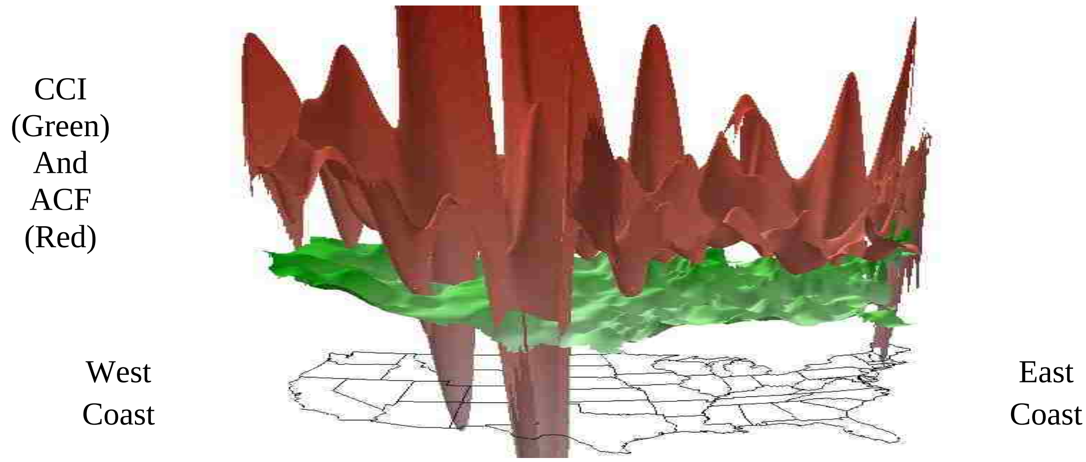

Figure 13 3D Kriging Distribution for 2009 CCI and ACF

CHAPTER 5.0 DISCUSSION

5.1 Overview

In this chapter, a discussion of the following results will be addressed.

- Global Moran's I Test
- Error Calculation
- Comparison of Overestimate and Underestimate
- Best Performance Comparison
- Comparison of Error Percentages
- Descriptive Statistics
- Pattern Comparison

One important thing that should be addressed is that for the results mentioned above, RSMeans CCI dataset from year 2005 to year 2009 were used to validate them while DoD ACF dataset from year 2005 to year 2009 were used to cross-validate them.

5.2 Discussion of Global Moran's I Test Results

In the former research conducted by Martinez (2010), it was determined that the RSMeans CCI values were strongly spatially autocorrelated. He concluded that the current, industry-adopted interpolation method, namely the proximity-based location adjustment method was statistically valid. In this research, based on the results from the strong spatial autocorrelation of the changes in CCI value from year 2005 to year 2009, the proximity-based location adjustment method was better supported. Based on the maps for the changes in CCI value from year 2005 to year 2009, we can visually identify that the changes in CCI value are spatially autocorrelated.

The spatial autocorrelation analysis of the DoD ACF dataset cross-validated the validity of the proximity-based location method. Based on the results from the national and state level Global Moran's I test for the DoD ACF 2006 dataset, it was determined that DoD ACF values were strongly spatially autocorrelated. In addition, the changes in ACF value from year 2005 to year 2009 were also significantly spatially autocorrelated. Similar to CCI values, based on the maps for the changes in ACF value from year 2005 to year 2009, we can visually identify that the changes in ACF value are spatially autocorrelated.

Based on the results from both RSMeans CCI and DoD ACF, it is very safe to conclude that the proximity-based location adjustment method which is adopted broadly by the construction industry is valid.

5.3 Error Calculation Results

In the error calculation for CNN and NN, two tools in ArcGIS, namely "Near" and "Join", were employed. They greatly accelerated the error calculation, considering that both RSMeans CCI and DoD ACF datasets are from year 2005 to year 2009.

As to the error calculation for IDW, kriging, and spline, since there were many parameters to choose for each method, an analysis called effect analysis was created. The underlying concept of effect analysis was to find the most important parameters for each interpolation method and then develop all the possible combinations of these parameters. RMSE was used to measure which combination will produce the most accurate result for each method. As this stage, it seems that effect analysis works very well and for future more detailed comparison for IDW, kriging, and spline methods, effect analysis could be considered as one good method to choose the parameters. However, no matter what parameter combination is selected, the error difference is not very huge. Therefore, for future studies, it is doable that we just employ the default parameters to obtain the error for IDW, kriging, and spline methods.

In addition, if parameter combination analysis is necessary and there are more parameters must be considered, it is recommended that a new analysis method should be developed, considering the effect analysis is time-consuming.

5.4 Comparison of Overestimate and Underestimate Results

For the RSMeans CCI dataset from year 2005 to year 2009, the comparison of overestimate and underestimate revealed a slight increase in overestimates for all methods. For each method, there are more overestimates than underestimates. However, the differences between the number of overestimate and underestimate for each method were not relatively significant or extreme. This implied that for RSMeans CCI, CNN, NN, ST AVG, IDW, kriging, and spline method might have a slight tendency to be overestimated. When comparing the number of accurate estimates based on the results from RSMeans CCI, it was not obvious to determine which method significantly outperforms the other methods. However, based on the DoD ACF dataset from year 2005 to year 2009, only ST AVG and IDW might have a slight tendency to be overestimated. In addition, when comparing the number of accurate estimates, it was determined that CNN significantly outperformed the other surface interpolation methods.

5.5 Best Performance Comparison Results

A best performance comparison was conducted for each method. Here the best performance was defined as the lower absolute error value for a location. For the RSMeans CCI dataset from year 2005 to year 2009, CNN and IDW were determined as the best two surface interpolation methods. In addition, considering that CNN is the rough surface interpolation method while IDW is the smooth surface interpolation method, we can assume that IDW is the best option for developing a surface cost function. However, this result could not be cross-validated by the DoD ACF dataset from year 2005 to year 2009. One possible reason is that there are not enough data points in the DoD ACF dataset.

5.6 Comparison of Error Percentages Results

Based on the results from RSMeans CCI dataset from year 2005 to year 2009, it was not confirmed that CNN outperformed all the other surface interpolation methods. In addition, IDW did not outperform all the other smooth surface interpolation method. Sometimes kriging or spline is better than IDW while sometimes not. The comparison of error percentages based on DoD ACF dataset from year 2005 to year 2009 provide the same result. Due to the research deadline, an appropriate method to analyze the error percentage results was not identified. Even though there is no appropriate analysis method, the information provided by the table of and figure of implied that CNN had the potential to be the most accurate surface interpolation method while IDW had the potential to be the most accurate smooth surface interpolation method. To fully conclude this implication, relevant statistical assessments needed to be performed.

5.7 Descriptive Statistics Results

From the results obtained from RSMeans CCI dataset from year 2005 to year 2009, it was confirmed that IDW and CNN are the best two surface interpolation methods. CNN outperformed all the other rough surface interpolation methods. In addition, IDW outperformed all the other smooth surface interpolation method. The ranks for the 6 surface interpolation methods from best to worst are: 1) IDW 2) CNN 3) Kriging 4) Spline 5) ST AVG 6) NN. Moreover, descriptive statistics results obtained from DoD ACF dataset from year 2005 to year 2009 cross-validated this conclusion. The ranks for the 6 surface interpolation methods from best to worst are: 1) CNN 2) IDW 3) Kriging 4) Spline 5) ST AVG 6) NN. One interesting finding is that the current, industry-adopted NN interpolation method is the worst surface interpolation method. However, the problem of which surface interpolation method is the best one was not solved int this research. It might be a topic for future study.

5.8 Pattern Comparison Result

This research is the initial phase for developing a smooth surface cost function. Based on the pattern comparison results, we can visually identify which surface interpolation method will produce the best result.

Based on the 2D distribution results developed from RSMeans CCI dataset from year 2005 to year 2009, it is difficult to distinguish the best method from the others. From the maps, it is clear that all these three methods could produce a smooth distribution.

However, based on the 2D distribution results developed from DoD ACF dataset from year 2005 to year 2009, it is obvious that IDW produces the smoothest distribution since there are some bulks in both kriging and spline.

In the 3D distribution comparison, it is also clear that IDW is the best smooth surface interpolation method. That is because by using IDW, both RSMeans CCI and DoD ACF display the same trend.

One possible reason that kriging and spline did not work well in DoD ACF dataset is not enough sample locations exist. It implies that when there are enough sample locations in a cost dataset, especially at the individual state level, any smooth surface interpolation method can be selected. However, when there are not enough sample locations available, the best option is IDW method.

5.9 Comprehensive Results Discussion

For CNN, NN and ST AVG, all of them are developed based on simple functions. However, CNN outperforms the other two methods. Both CNN and ST AVG consider the state boundary and both of them outperform the NN method. Therefore, state boundary criteria may play an important role for interpolation. In addition, CNN considers the linear distance while ST AVG considers the average value across the same state. Therefore, linear distance function may be better than average value function.

IDW outperforms the kriging and spline methods. Although kriging is the more complex method, it does not provide the more accurate result. One possible reason is that not enough parameters were appropriately selected. Another problem associated with kriging is the estimation of semivariogram. It is not always easy to ascertain whether a particular estimate of the semivariogram is in fact a true estimator of the spatial correlation in an area. Finally, kriging is not a suitable method for data sets which have anomalous pits or spikes, or abrupt changes such as break lines.

Spline also does not outperform IDW. One possible reason is that spline is best for gently varying surfaces where change in physiography or other phenomenon is not abrupt. It is not appropriate if there are large changes in the surface within a short horizontal distance because it can overshoot estimated values.

For RSMeans CCI dataset, IDW is the best surface interpolation method while for DoD ACF dataset CNN is the best method. However, CNN is a fast and easy method while IDW needs a great amount of calculations. The commercial software for location adjustment can be developed based on IDW surface interpolation method.

CHAPTER 6.0 CONCLUSION

6.1 Summary of Research Results

Global Moran's I test results provided evidence of strong spatial autocorrelation existed for both RSMeans CCI values and DoD ACF values. Therefore, the current, industry-adopted proximity-based location adjustment method was completely validated by the internal cost information.

Based on the comparison of overestimate and underestimate, best performance comparison, and comparison of error percentages, and descriptive statistics, it was determined that CNN is the best surface interpolation method while IDW is the best smooth surface interpolation method. With the 2D and 3D pattern comparison, the result that IDW is the best smooth surface interpolation method was visually supported.

6.2 Research Questions

The following questions were addressed throughout this research and the answers were summarized in the following sections:

1. Can the current, industry-suggested NN interpolation method be better supported?

2. What are the possible alternatives to the current methods that may produce a smooth surface method?

3. Can these alternative methods be statistically proven to produce a more accurate construction cost estimates?

4. Can these alternative methods be visualized?

5. Can these alternative methods be cross-validated by another set of location adjustment factors such as DoD ACF?

6.3 Research Rationale and Findings

All the five questions mentioned in section 6.2 were evaluated. The following sections will discuss the findings of these questions. In addition, the research rational behind each finding was also explained.

6.3.1 Research Rationale and Findings for Question 1

An understanding of “current method” is needed to answer the first question. In this research, the current method is referred to “nearest neighbor” (NN) location adjustment method, which is spatial interpolation method based on linear distance, namely proximity. For this proximity-based method, the estimation of a variable for a location completely relies on the same variable of the closest location. The variable in this research are RSMeans CCI and DoD ACF. However, based on the analyses in this research, the NN method does not perform well when compared with the other 5 methods.

The Global Moran’s I test was conducted to test the spatial autocorrelation of the CCI and ACF. Results indicates that the both the CCI values and ACF values were highly spatially autocorrelated. In addition, the changes in CCI value and the changes in ACF values were also highly spatially autocorrelated, which means that changes in values have the tendency to vary in concert. High values are found near high values, and low values are found near low values. Therefore, the underlying assumption for proximity-based location adjustment method was fully validated.

6.3.2 Research Rationale and Findings for Question 2

For the second question, besides CNN, NN, ST AVG, another three possible alternatives were identified. They are IDW, kriging, and spline. These six different surface interpolation methods can be classified into two categories. CNN, NN, and ST AVG are rough surface interpolation methods while IDW, kriging, and spline are smooth surface interpolation methods.

6.3.3 Research Rationale and Findings for Question 3

The third question is to statistically compare the six surface interpolation methods mentioned above. The error for each method was used to measure performance. Error is defined as the result of estimated factor minus actual factor. Based on the error, comparison of overestimate and underestimate, comparison of best performance, and comparison of error percentages were performed. The statistical testing technique employed in this research is descriptive statistics, which includes median, mean, standard deviation, mode, skewness, and kurtosis (see the tables in Appendix F).

The results of the comparison and statistical analysis mentioned above demonstrated that for RSMeans CCI dataset, IDW is the best smooth surface method while CNN is the best rough surface method. For DoD ACF dataset, IDW is also the best smooth surface interpolation methods while CNN is the best rough surface interpolation method. However, as to which one is the best surface interpolation method, the answer is not consistent. For RSMeans CCI dataset, IDW is the best method. But for DoD ACF, CNN is the best method. Due to the research deadline, the best surface interpolation method was not found. This problem might be a good topic for future research. One point that needs to be addressed is that CNN is a quick and easy method which does not need a lot of time and experiences. However, IDW is a complex function and its usage needs the help of specific software such as ArcGIS. If there is commercial software for location adjustment based on IDW, it will greatly improve the efficiency.

6.3.4 Research Rationale and Findings for Question 4

This research is the initial step to develop a smooth 3D surface cost function for spatial prediction. Therefore, it is necessary to visualize the spatial distribution of the location factors to compare the results.

With the aid of the ArcGIS software, both 2D and 3D distribution were developed and the results showed that when there are enough sample locations in a cost dataset, especially at the individual state level, any smooth surface interpolation method can be selected. However, when there are not enough sample locations available, the best option is IDW method.

6.3.5 Research Rationale and Findings for Question 5

If all the analyses were based on RSMeans CCI dataset, they may be random results. Therefore, it is necessary to employ another published dataset to cross-validate the results. In this research, DoD ACF was selected to perform the cross-validation.

All the comparisons and analyses in this research were successfully cross-validated by the DoD ACF dataset.

6.4 Limitations of the Research

There are several limitations for this research and they are:

- RSMeans CCI Dataset and DoD ACF Dataset
- Parameter Selection for IDW, Kriging, and Spline
- External Validation

Each of these topics will be discussed in the following section.

6.4.1 RSMeans CCI Dataset and DoD ACF Dataset

The CCI is published by the RSMeans while the ACF is published by the Department of Defense annually. Both the RSMeans CCI and the DoD ACF were assumed to be a valid predictor of construction costs. For RSMeans CCI, the types of projects are limited to commercial or industrial projects that cost at least $ 1,000,000.00. For DoD ACF, the types of projects are limited to military projects without cost limitation. In addition, both of RSMeans CCI and DoD ACF were limited to new construction which did not include renovations or minor modifications. These limitations apply to the research findings.

6.4.2 Parameter Selection for IDW, Kriging, and Spline

As mentioned earlier, there is a geostatistical wizard for getting the predicted and actual location factors in Geostatistical Analyst. In this wizard, there are many parameters that need to be selected for each surface interpolation method, and different selection will lead to different predicted location factor values. In this research, there is a limitation that only a few of most important parameters were selected and tested. These limitations apply to the research findings.

6.4.3 External Validation

As mentioned earlier, both RSMeans CCI dataset and DoD ACF dataset are internal validation data sources. Therefore, actual construction projects cost data could be a possible continuation of this research. Theoretically, CNN is the best rough surface interpolation method and IDW is the best smooth surface method. It is necessary to test whether the result is the same when using actual cost data.

6.5 Implication for Future Research

Future research with regard to location adjustment method is to follow. One possible future research topic is to use Geary's C instead of Global Moran's I test to assess the degree of spatial autocorrelation.

In addition, linear distance can be substituted by the transportation distance. For CNN and NN surface interpolation method, proximity is measured by the linear distance. A possible and recommended alternative to linear distance could be the actual road transportation distance. The logic behind this idea is that transportation cost can affect greatly the construction cost. For example, if city A and city B has a closer transportation distance, while city A and city C has a closer linear distance, probably city A is affected more by city B instead of city C. The use of transportation distance instead of the linear distance could have a significant effect on the error calculation.

To develop a new method to test various parameter combinations is also a good future research topic. Moreover, a recommended research topic is to evaluate the six surface interpolation methods at the state level. It is possible that in some specific states, IDW is the best surface interpolation method. After this initial step, future research topic is to develop a complete smooth surface cost function based on IDW method. In the cost function, several criteria such as income and house values will be included. Then this cost function could be used for spatial prediction.

Finally, actual construction project cost data should be collected and used to test whether the same result could be developed as using RSMeans CCI and DoD ACF.

REFERENCES

Air Force Civil Engineering Support Agency. (2005). *Historical Air force Construction Cost Handbook*. Retrieved October 30, 2009, from http://www.wbdg.org/ccb/AF/AFH/histbook.pdf

Ashur, S. A., & Crockett, B. (1997). Geographic information systems as a support tool in construction cost estimating in state DOTs. *Transportation Research Record, 1575*, 112-115.

Bivand, R. (1998). A review of spatial statistical techniques for location studies. Retrieved October 25, 2009, from http://bora.nhh.no/bitstream/2330/2091/1/rober%20bivand221.pdf

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004) *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL: Chapman & Hall/CRC

Bolstad, P. (2005). *GIS Fundamentals, a first text on Geographic Information Systems* (2nd Ed.). White Bear Lake, MN: Eider Press.

Bansal, V.K., & Pal, M. (2007). Potential of geographic information systems in building cost estimation and visualization. *Automation in Construction*, *16*, 311–322.

Borden, A. K., & Cutter, S. L. (2008). Spatial patterns of natural hazards mortality in the United States. Retrieved October 25, 2009, from http://www.ij-healthgeographics.com/content/7/1/64

Carr, R.I. (1989) Cost estimating principles, *Journal of Construction Engineering and Management*, *115(4)*, 545-551.

Cheng, M. Y., & O’Connor, J. T. (1996). ARCSite: Enhanced GIS for construction site layout. *Journal of Construction Engineering and Management, 122(4)*, 329-336.

Cheng, M. & Yang, S. (2001). GIS-based cost estimates integrated with material layout planning. *Journal of Construction Engineering and Management, 127(4),* 291-299.

Christensen, P., & Dysert, L.R. (2003). *AACE International Recommended Practice No. 17R-97 Cost Estimate Classification System*. Retrieved March 22, 2008, from http://www.aacei.org/technical/rps/17r-97.pdf

Duverlie, P., & Castelain, J.M. (1999). Cost estimation during design step: parametric method versus case based reasoning. *The International Journal of Advanced Manufacturing Technology, 15*, 895-906.

Gould, F. E. (1997). *Managing the Construction Process*. Columbus, OH: Prentice Hall.

Ito, K. (1987). *Encyclopedic Dictionary of Mathematics* (2^{nd} Ed.). Cambridge, MA: MIT Press.

Johannes, J. M., Koch, P.D., Rasche, R.H. (1985). Estimating Regional Construction Cost Differences: Theory and Evidence. *Managerial and Decision Economics, 6(2)*, 70-79.

Jeljeli, M. N., Russell, J. S., Guy Meyer, H. W., & Vonderohe, A. P. (1993). Potential application of geographic information systems to construction industry. *Journal of Construction Engineering and Management, 119(1),* 72-86.

Kim, G. H., An, S.H., & Kang, K.I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39 (10)*, 1235-1242.

Lee, J. & Wong, D.W.S. (2001). *Statistical Analysis with ArcView GIS.* New York: John Wiley & Sons Inc.

Layer, A., Brinke, E., Van Houten, F., Kals, H., & Haasis, S. (2002). Recent and future trends in cost estimation. *International Journal of Computer Integrated Manufacturing, 15*(6), 499-510.

Li, H., Kong, C. W., Pang, Y. C., Shi, W. Z., & Yu, L. (2003). Internet-based geographical information systems system for E-commerce application in construction material procurement. *Journal of Construction Engineering and Management, 129(6),* 689-697.

Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Journal of* Cartographica, 27(1), 30-45.

Messner S. F., Anselin, L., Baller, R. D, Hawkins, D. F., Deane, G., & Tolnay, S. E. (1999). The spatial patterning of county homicide rates: an application of exploratory spatial data analysis. Journal of Quantitative Criminology, 15(4), 423-450.

McCabe, B., O'Grady, J., & Walker, F. (2002, June). *A study of construction cost sources.* Paper presented at the Canadian Society for Civil Engineering Annual Conference. Montreal, Canada.

Martinez A. A. (2010). Validation of Methods for Adjusting Construction Cost Estimates by Project Location. (Master's Thesis, University of New Mexico, 2010)

Oloufa, A. A., Eltaban, A. A., & Papacostas, C. S. (1994). Integrated GIS for construction site investigation. *Journal of Construction Engineering and Management, 120(1),* 211-222.

Pietlock, B.A. (2006). *AACE International Recommended Practice No. 28R-03 Developing Location Factors by Factoring – as Applied in Architecture & Engineering, Procurement & Construction.* Retrieved March 22, 2008, from http://www.aacei.org/technical/rps/28r-03.pdf

Poku, S. E., & Arditi, D. (2006), Construction scheduling and process control using geographic information systems, *Journal Computing in civil Engineering, 20 (5),* 351-360.

Popescu, C.M., Phaobunjong, K., & Ovararin, N. (2003). *Estimating Building Costs.* New York, NY: Marcel Dekker, Inc.

Rosenberg, M.S., Sokal, R. R., Oden, N. L. & DiGiovanni, D. (1999). Spatial autocorrelation of cancer in Western Europe. *European Journal of Epidemiology, 15(1)*, 15-22.

Ratner, J. (2002). *The Building estimator's reference book: a reference book setting forth detailed procedures and cost guidelines for those engaged in estimating building trades.* Lisle, IL: Frank R. Walker Co.

Salvatore, D., & Reagle, D. (2002). *Schaum's Outlines of theory and problems of Statistics and Econometrics* (2nd Ed.). Blacklick, OH: McGraw-Hill.

Suriatini, I. (2006). Spatial autocorrelation and real estate studies: a literature review. Retrieved October 25, 2009, from http://eprints.utm.my/438/1/1-13.pdf.

Taylor, J. R. *An Introduction to Error Analysis: the Study of Uncertainties in Physical Measurements* (2nd Ed.). Sausalito, CA: University Science Books.

United States of America Department of Defense. (2005). *Unified Facilities Criteria Design Guide: Construction Cost Estimating.* Retrieved October 30, 2009, from http://www.wbdg.org/ccb/DOD/UFC/ufc_3_700_02a.pdf

Unwin, D.J. (1996). GIS, spatial analysis and spatial statistics. *Journal of Progress in Human Geography, 20(4)*, 540-551.

Walsh, E. (2008). *The Estimator Shortage ~ Is There One? If There Is, What Do You Do?* Retrieved March 22, 2008, from http://commercialestimators.com/forums/p/147/258.aspx

Yu, K., Froese, T., and Grobler, F. (2000). A Development Framework for Data Models for Computer-Integrated Facilities Management. *Automation in Construction, 9(2),1*45-167.

Zhong, D., Li. J., Zhu, H., & Song, L. (2004). Geographic information system-based visual simulation methodology and its application in concrete dam construction processes. *Journal of Construction Engineering and Management, 130(5),* 742-750

APPENDIX A. GLOBAL MORAN'S I TEST RESULTS

Exhibit A1. Global Moran's I Test Results for the Changes in CCI Value

from Year 2005 to Year 2009

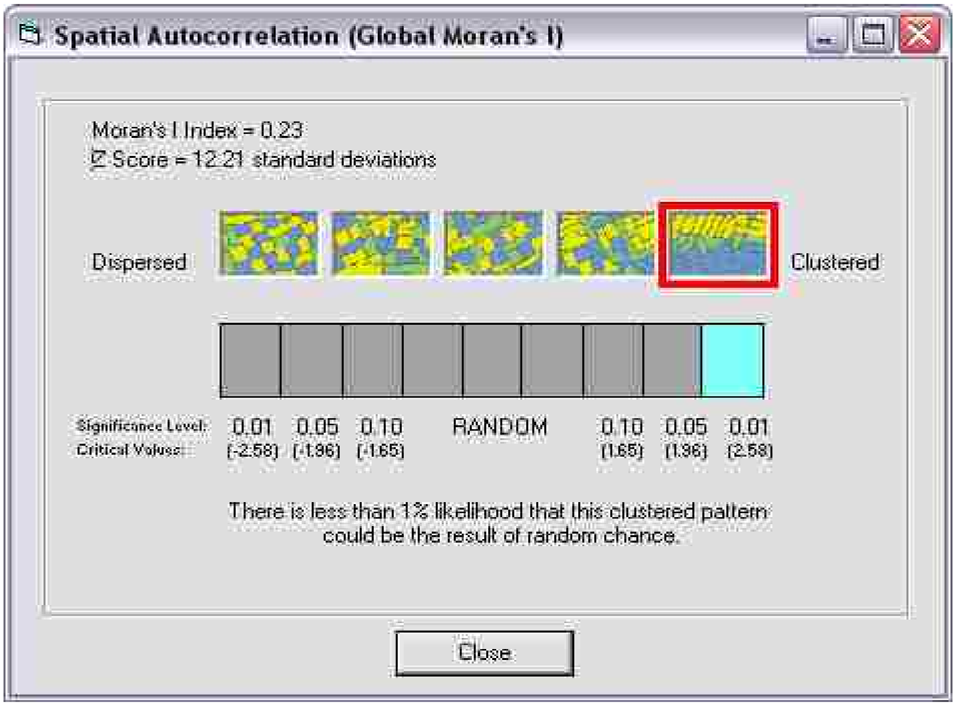

Figure A1.1 Global Moran's I test Dialogue for the Changes in CCI from 2005 to 2006

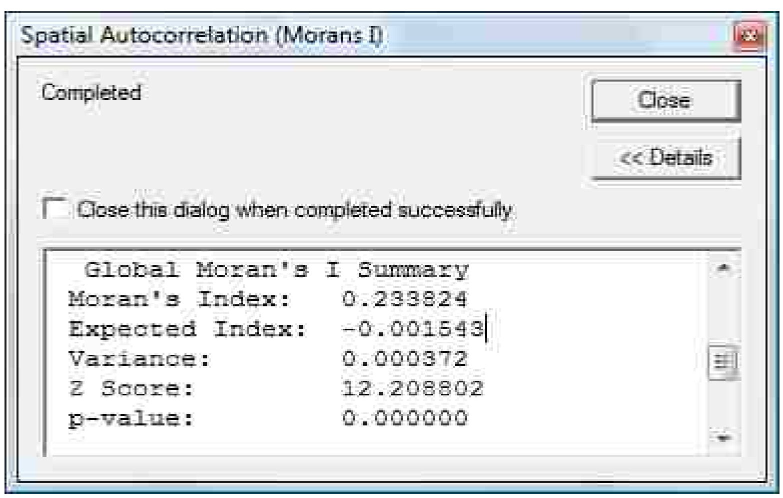

Figure A1.2 Global Moran's I test Summary for the Changes in CCI from 2005 to 2006

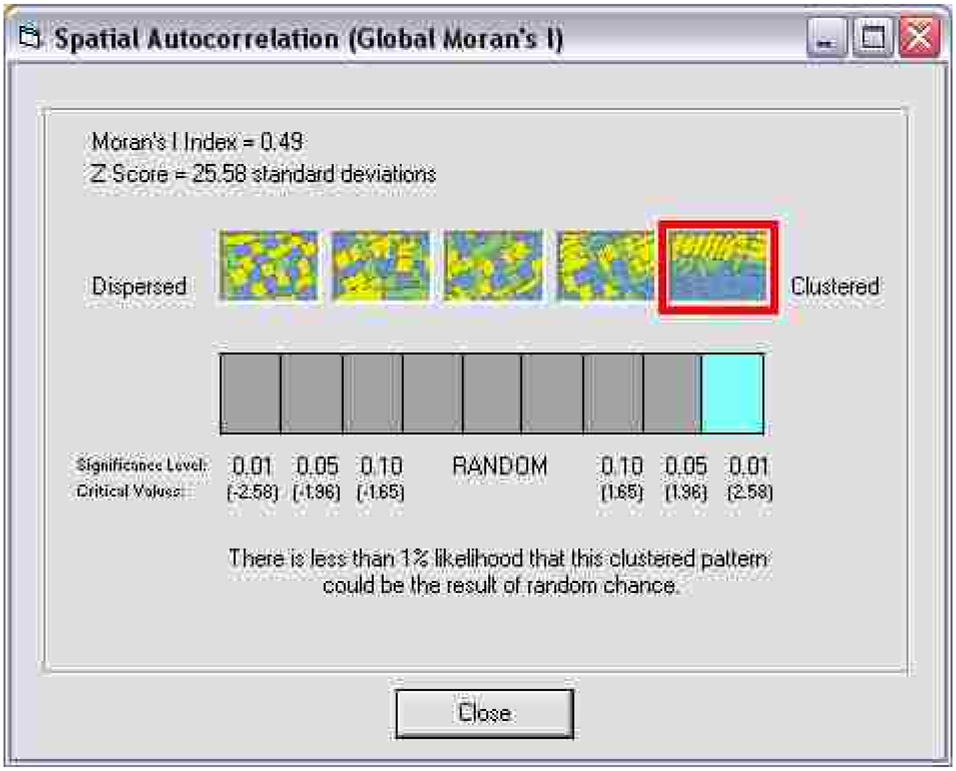

Figure A1.3 Global Moran's I test Dialogue for the Changes in CCI from 2006 to 2007

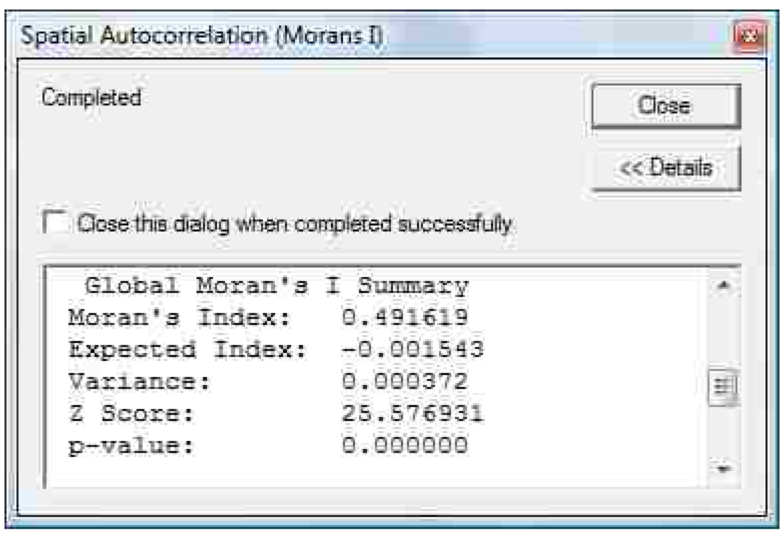

Figure A1.4 Global Moran's I test Summary for the Changes in CCI from 2006 to 2007

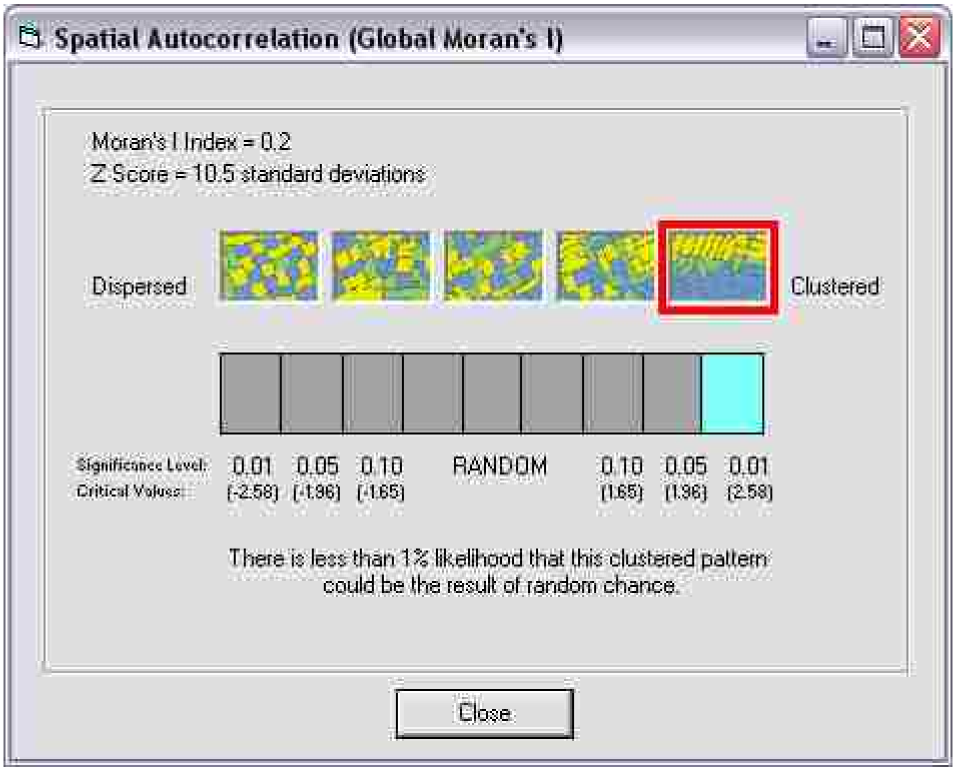

Figure A1.5 Global Moran’s I test Dialogue for the Changes in CCI from 2007 to 2008

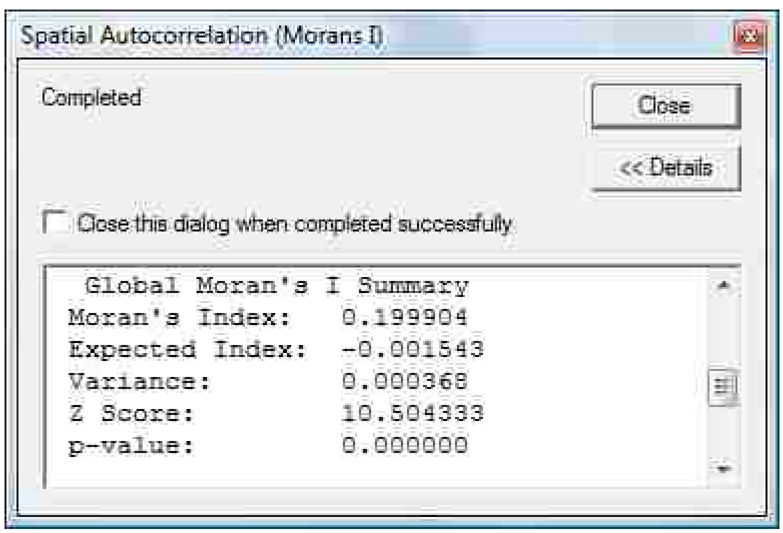

Figure A1.6 Global Moran’s I test Summary for the Changes in CCI from 2007 to 2008

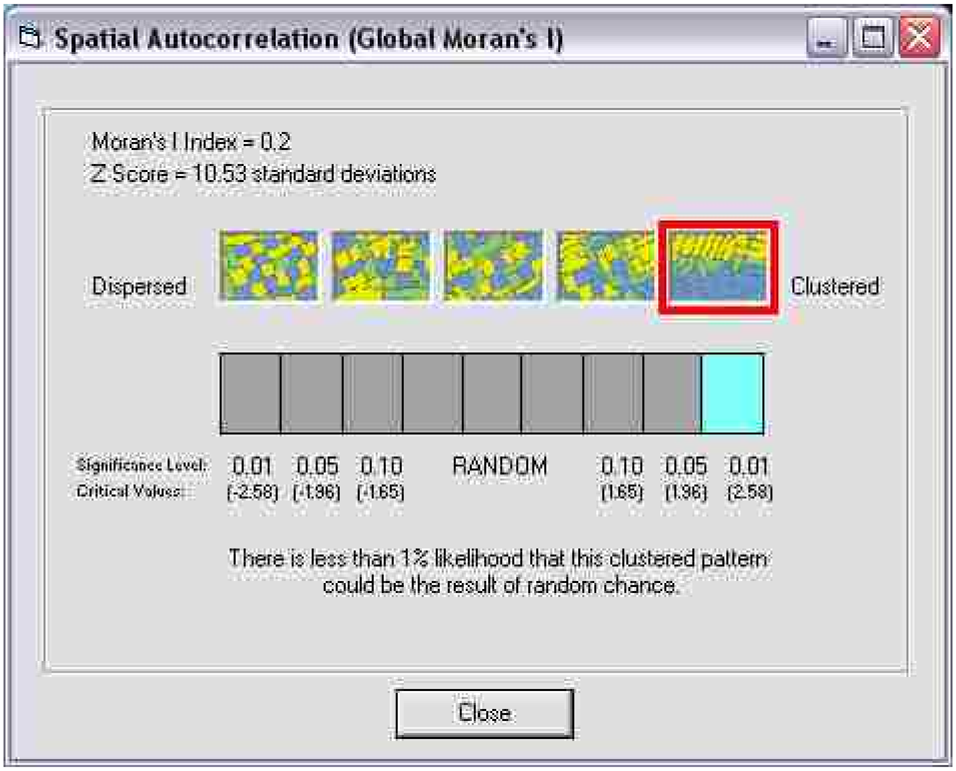

Figure A1.7 Global Moran's I test Dialogue for the Changes in CCI from 2008 to 2009

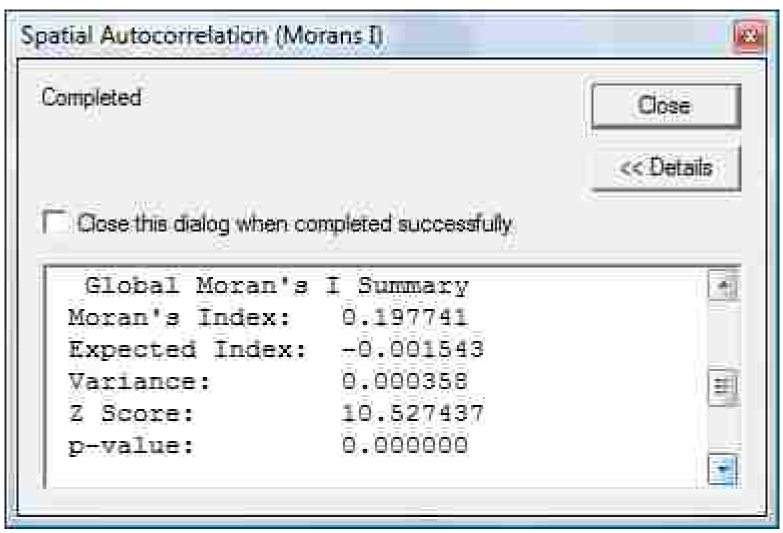

Figure A1.8 Global Moran's I test Summary for the Changes in CCI from 2008 to 2009

Exhibit A2. Global Moran's I Test Results for the Changes in ACF Value from Year 2005 to Year 2009

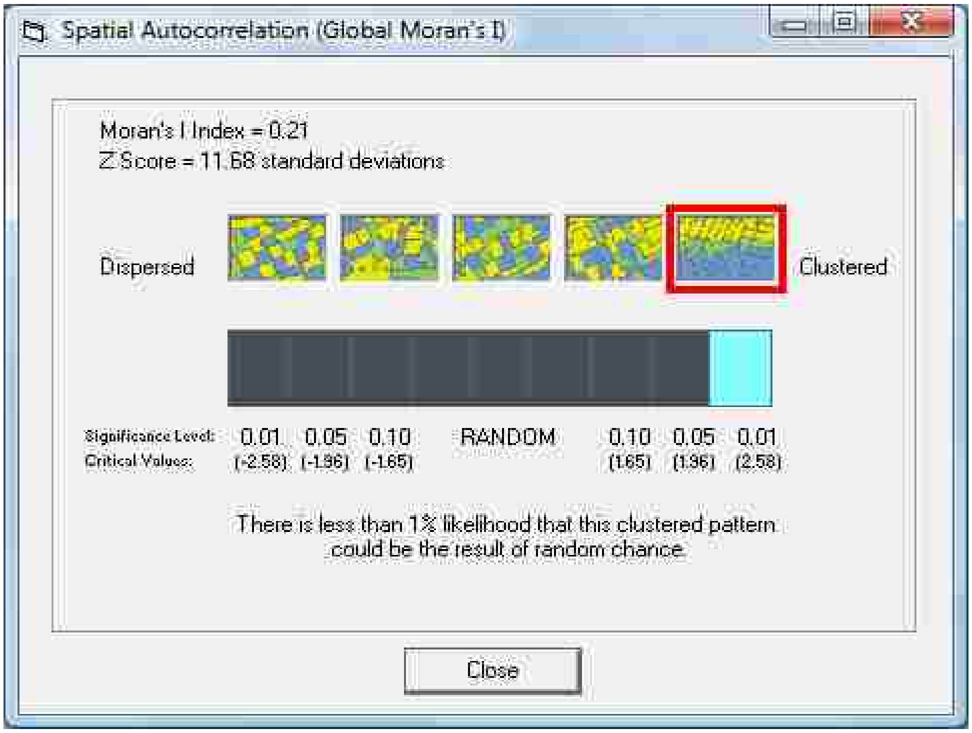

Figure A2.1 Global Moran's I test Dialogue for the Changes in ACF from 2005 to 2006

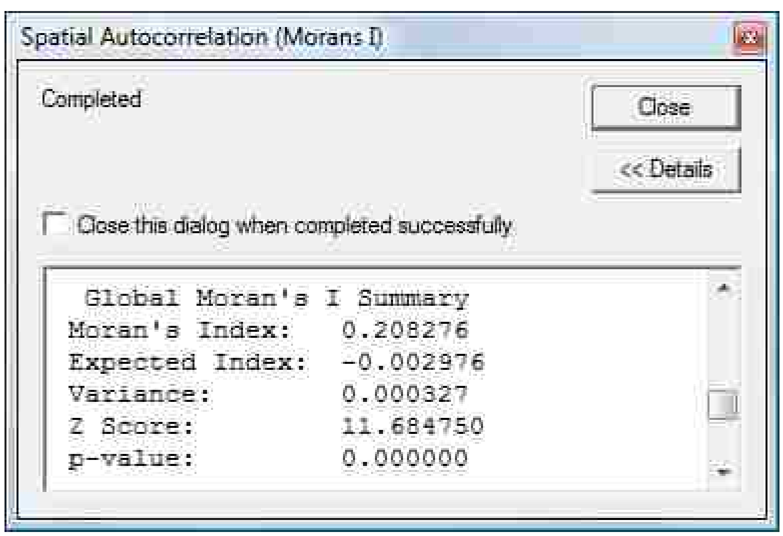

Figure A2.2 Global Moran's I test Summary for the Changes in ACF from 2005 to 2006

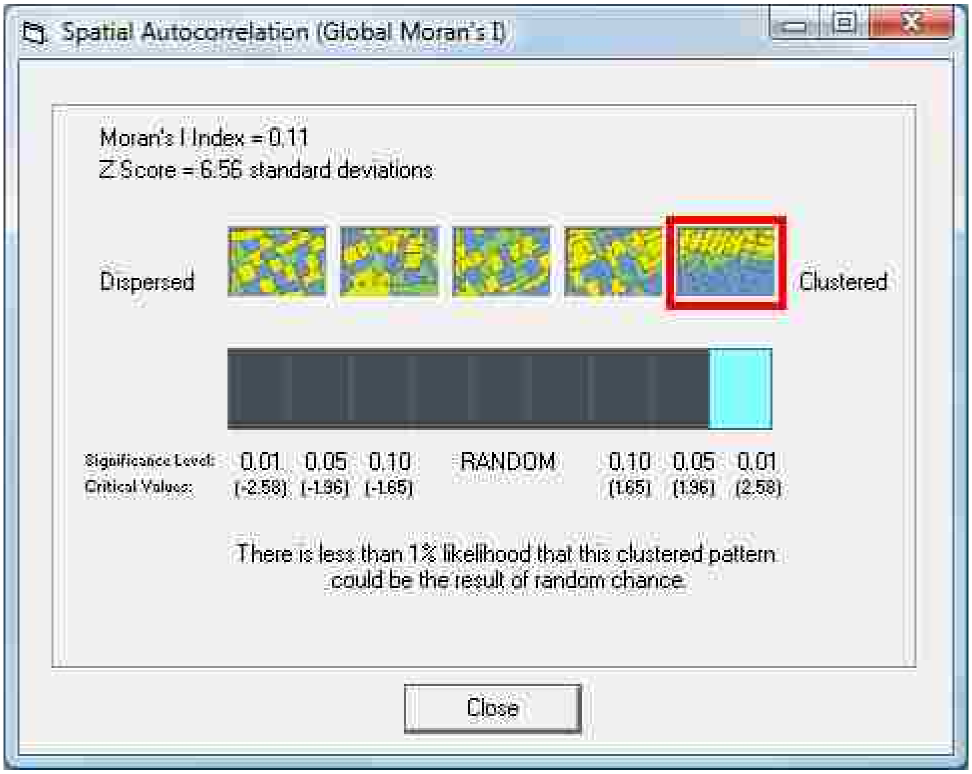

Figure A2.3 Global Moran's I test Dialogue for the Changes in ACF from 2006 to 2007

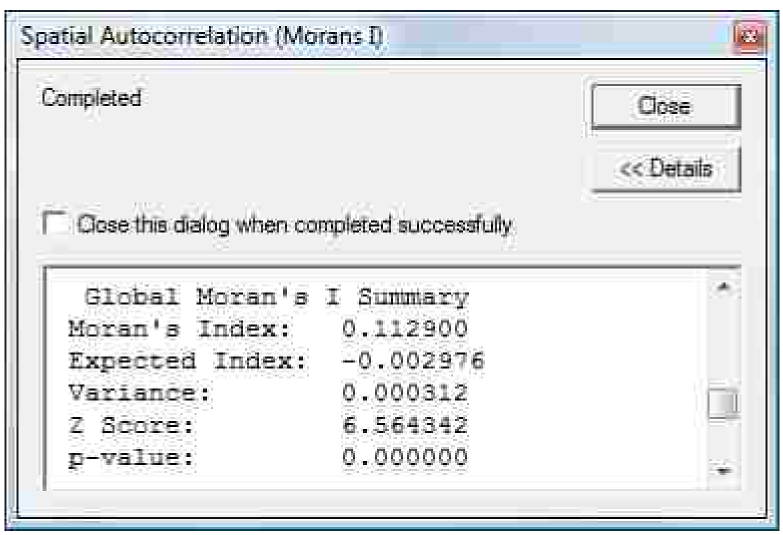

Figure A2.4 Global Moran's I test Summary for the Changes in ACF from 2006 to 2007

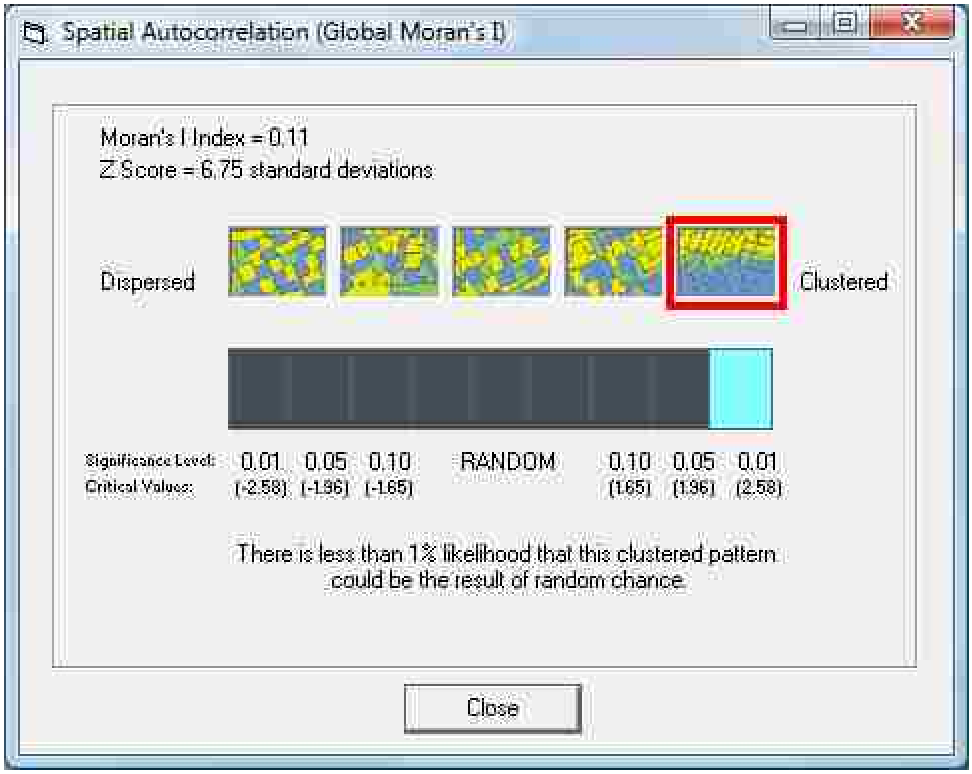

Figure A2.5 Global Moran's I test Dialogue for the Changes in ACF from 2007 to 2008

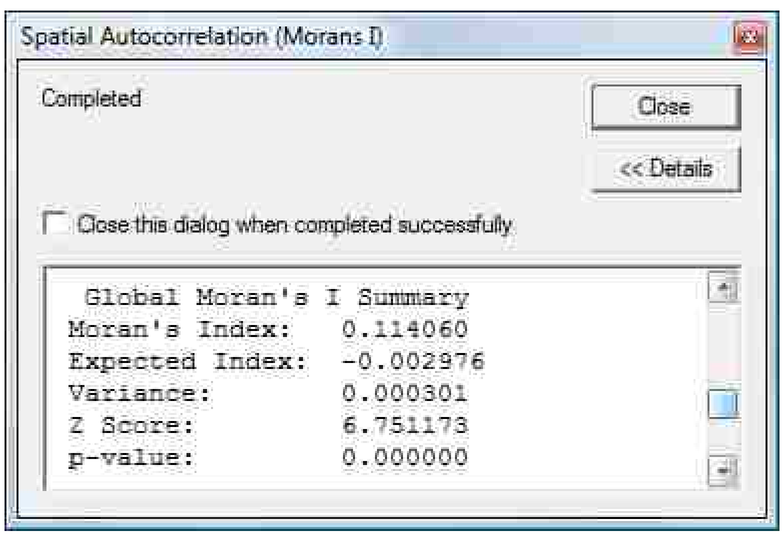

Figure A2.6 Global Moran's I test Summary for the Changes in ACF from 2007 to 2008

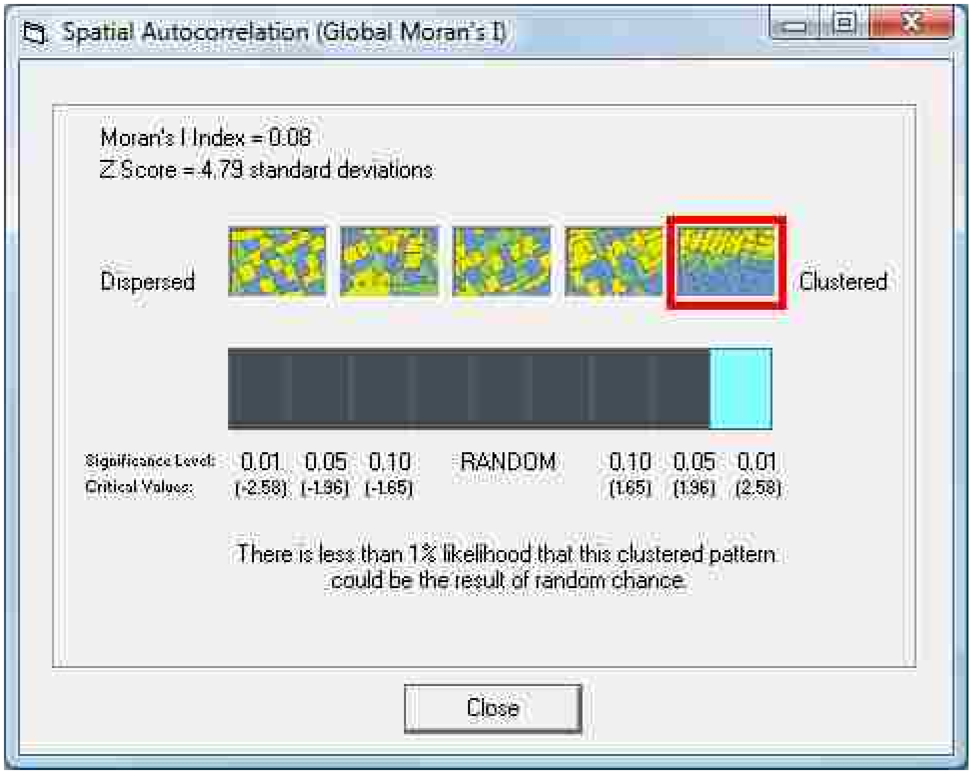

Figure A2.7 Global Moran's I test Dialogue for the Changes in ACF from 2008 to 2009

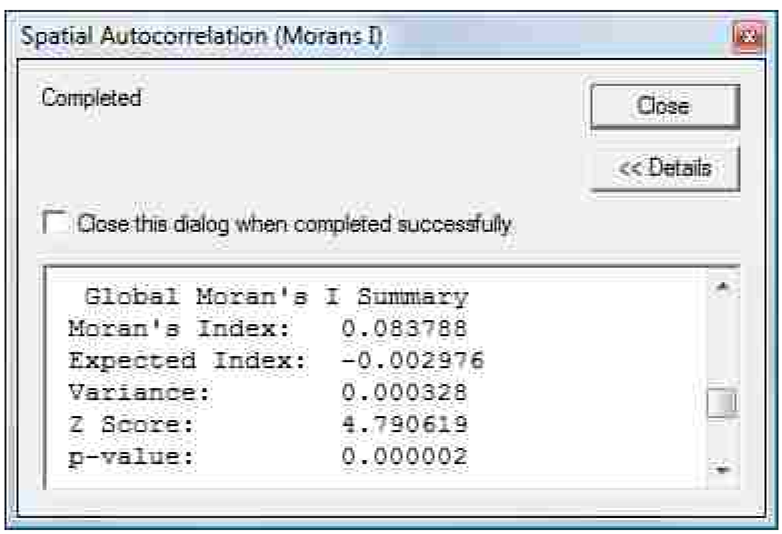

Figure A2.8 Global Moran's I test Summary for the Changes in ACF from 2008 to 2009

APPENDIX B. SPATIAL PATTERSNS OF THE CHANGES IN LOCATION FACTORS

Exhibit B1. Spatial Patterns of the Changes in CCI Value

from Year 2005 to Year 2009

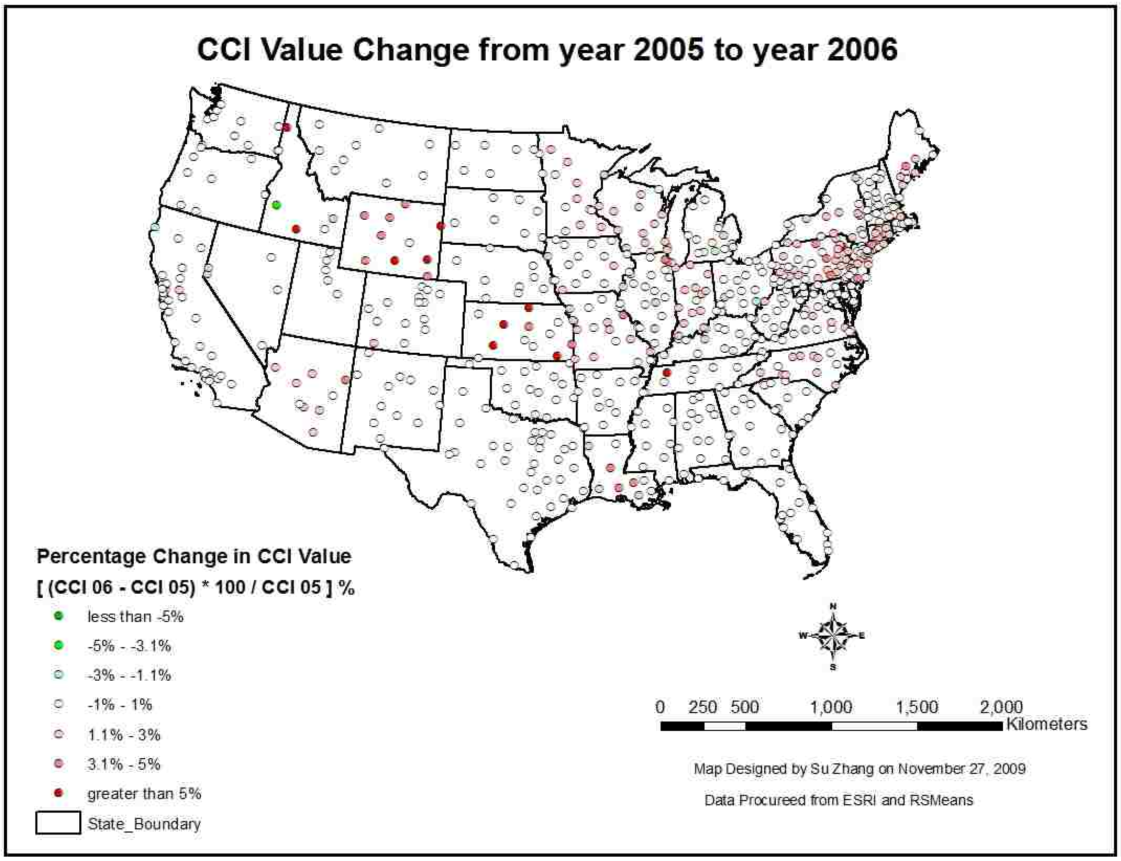

Figure B1.1 Spatial Pattern of the Changes in CCI Value from year 2005 to year 2006

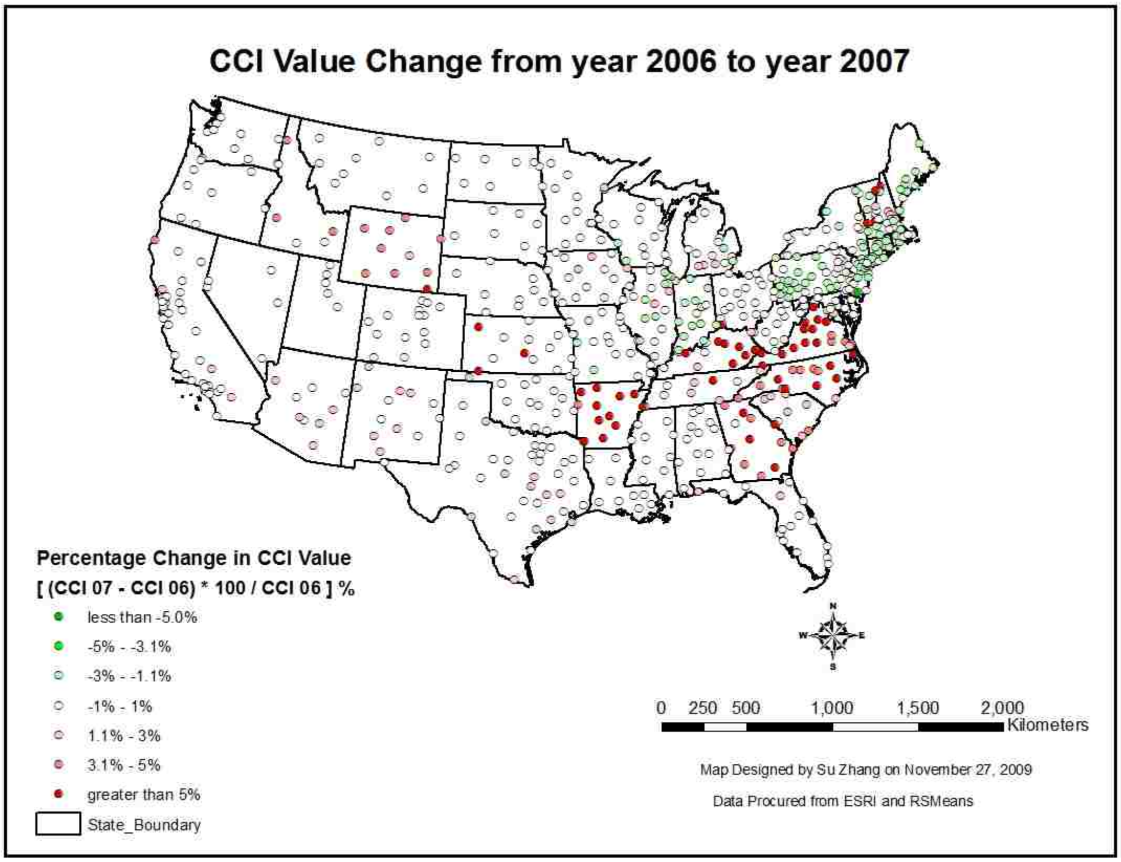

Figure B1.2 Spatial Pattern of the Changes in CCI Value from year 2006 to year 2007

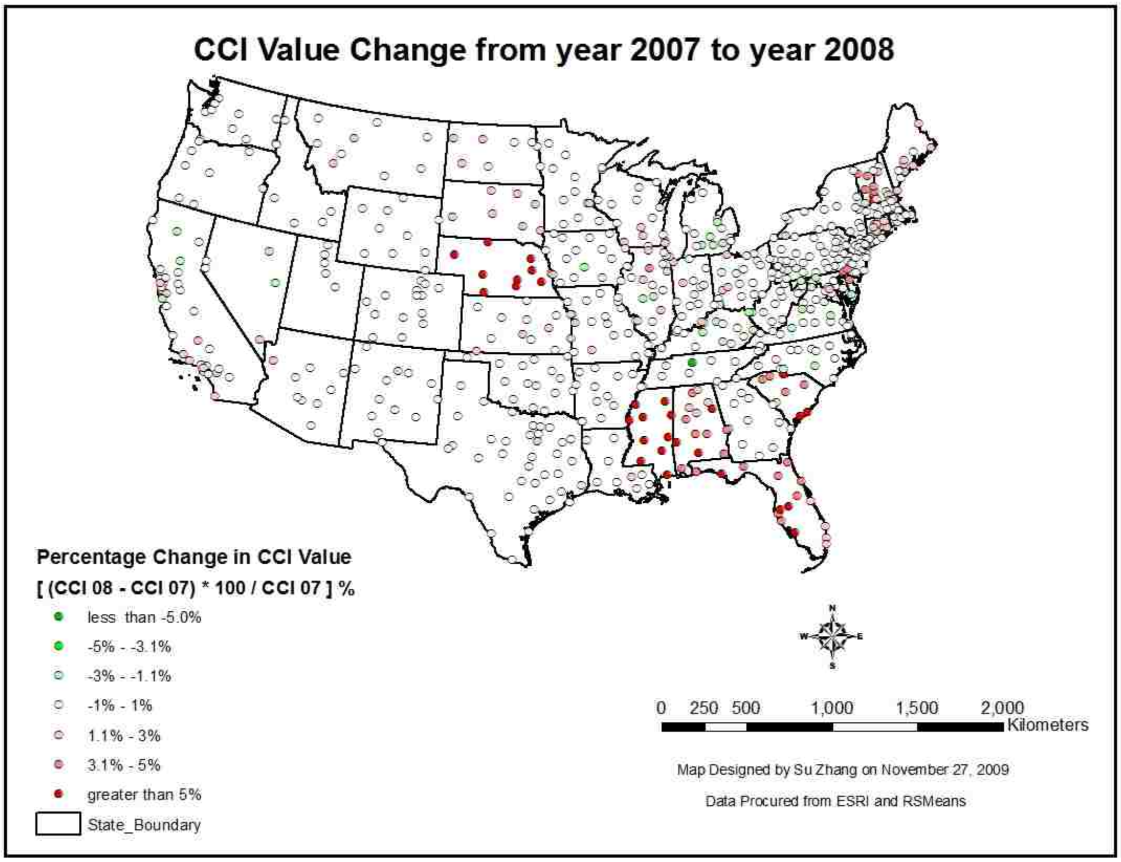

Figure B1.3 Spatial Pattern of the Changes in CCI Value from year 2007 to year 2008

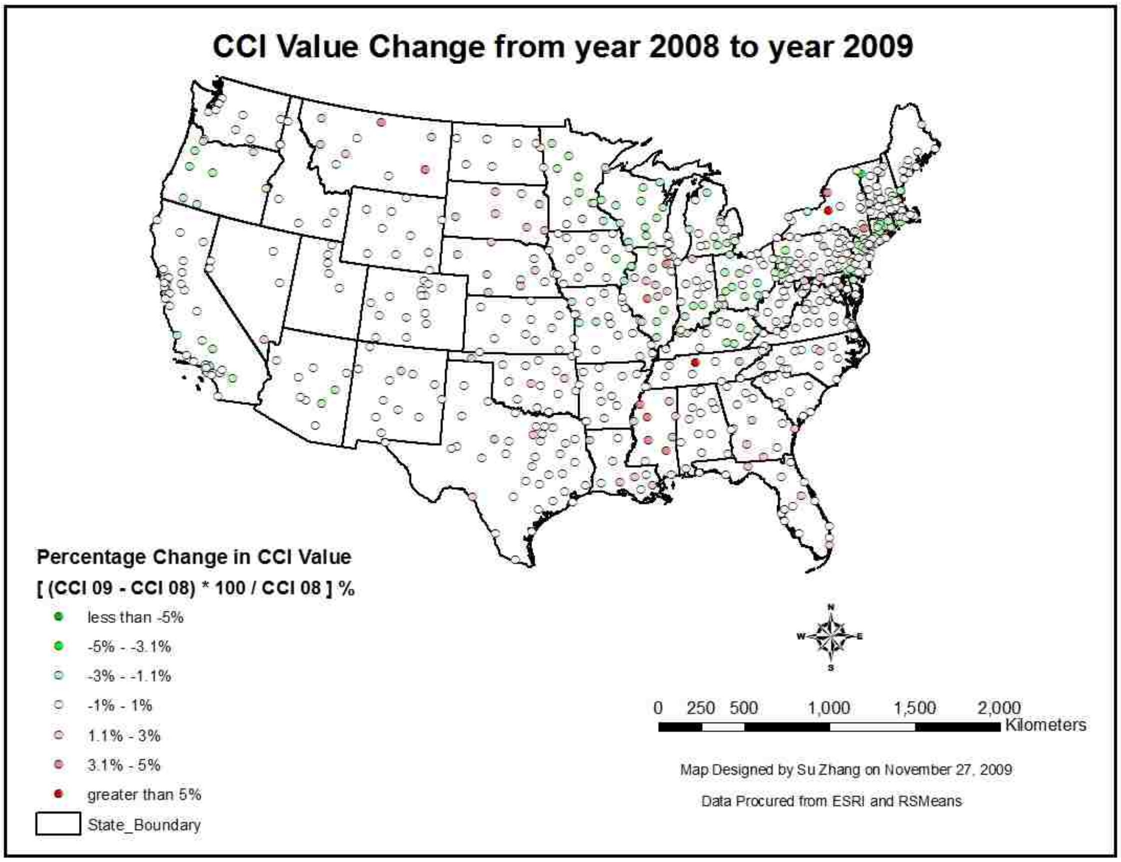

Figure B1.4 Spatial Pattern of the Changes in CCI Value from year 2008 to year 2009

Exhibit B2. Spatial Patterns of the Changes in ACF Value

from Year 2005 to Year 2009

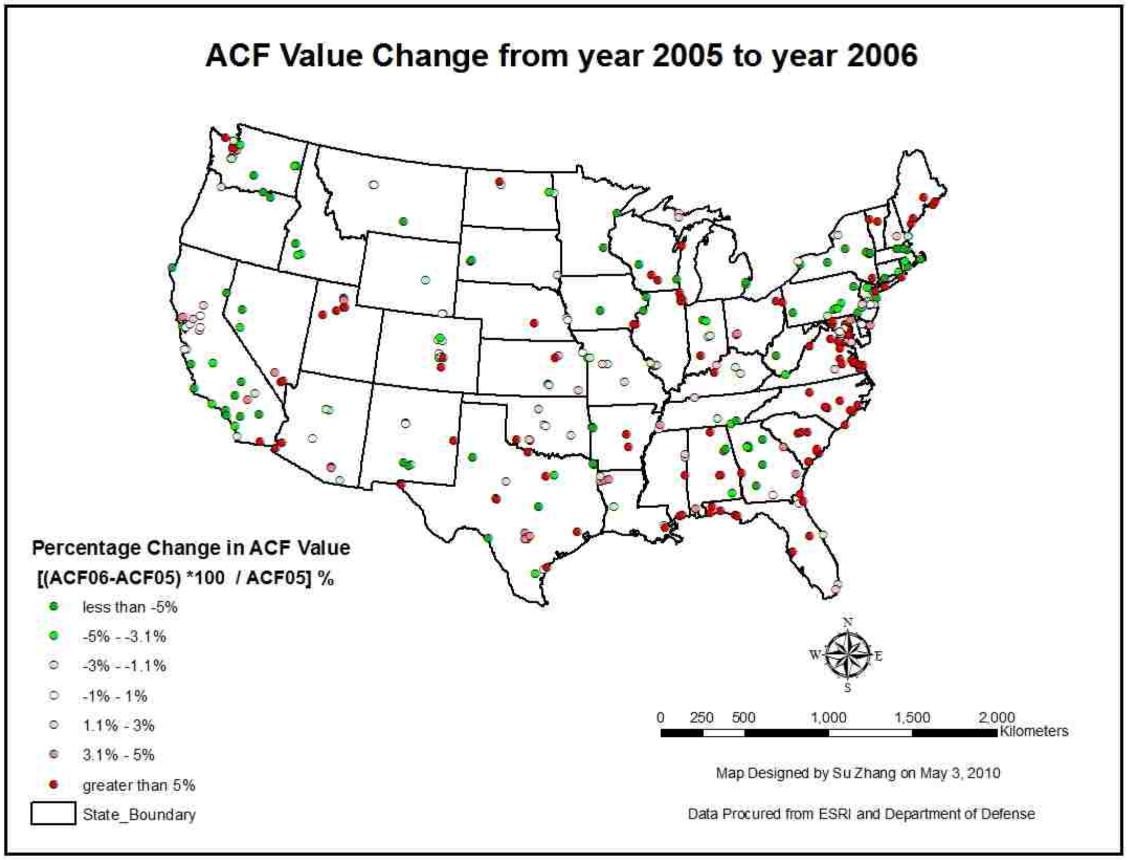

Figure B2.1 Spatial Pattern of the Changes in ACF Value from year 2005 to year 2006

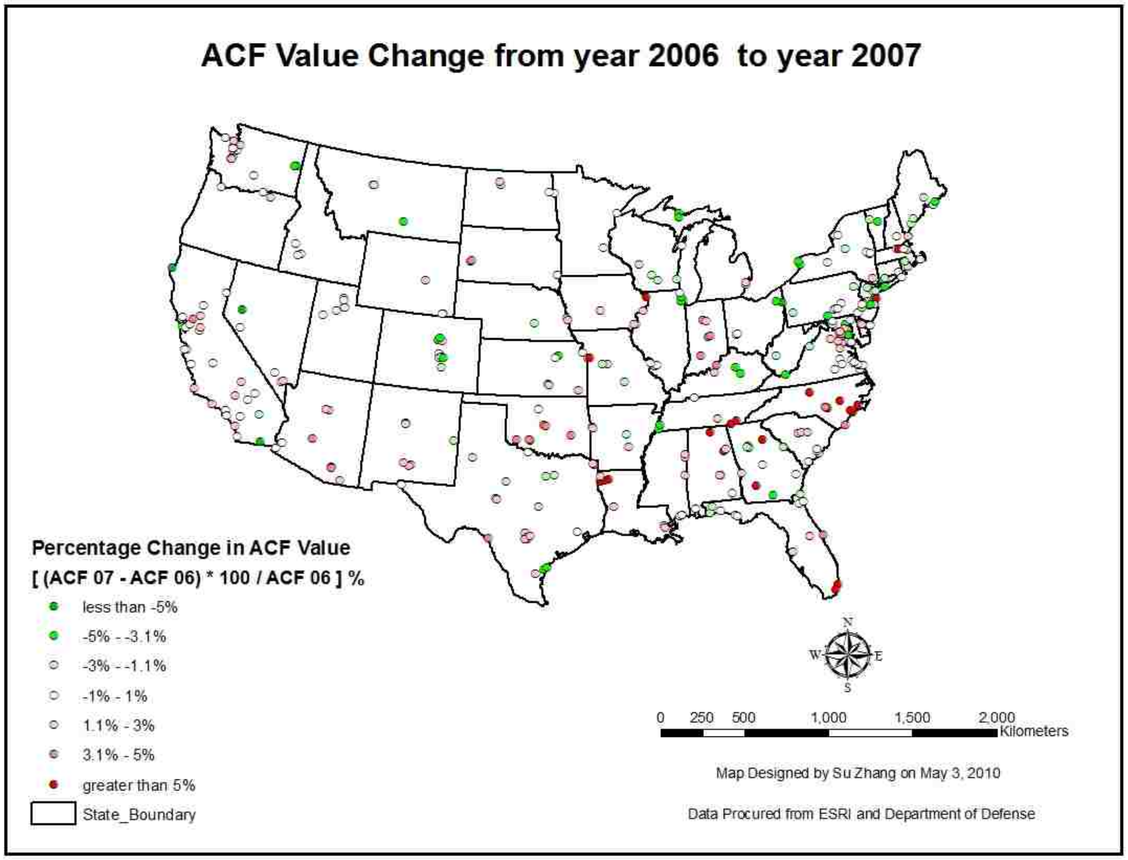

Figure B2.2 Spatial Pattern of the Changes in ACF Value from year 2006 to year 2007

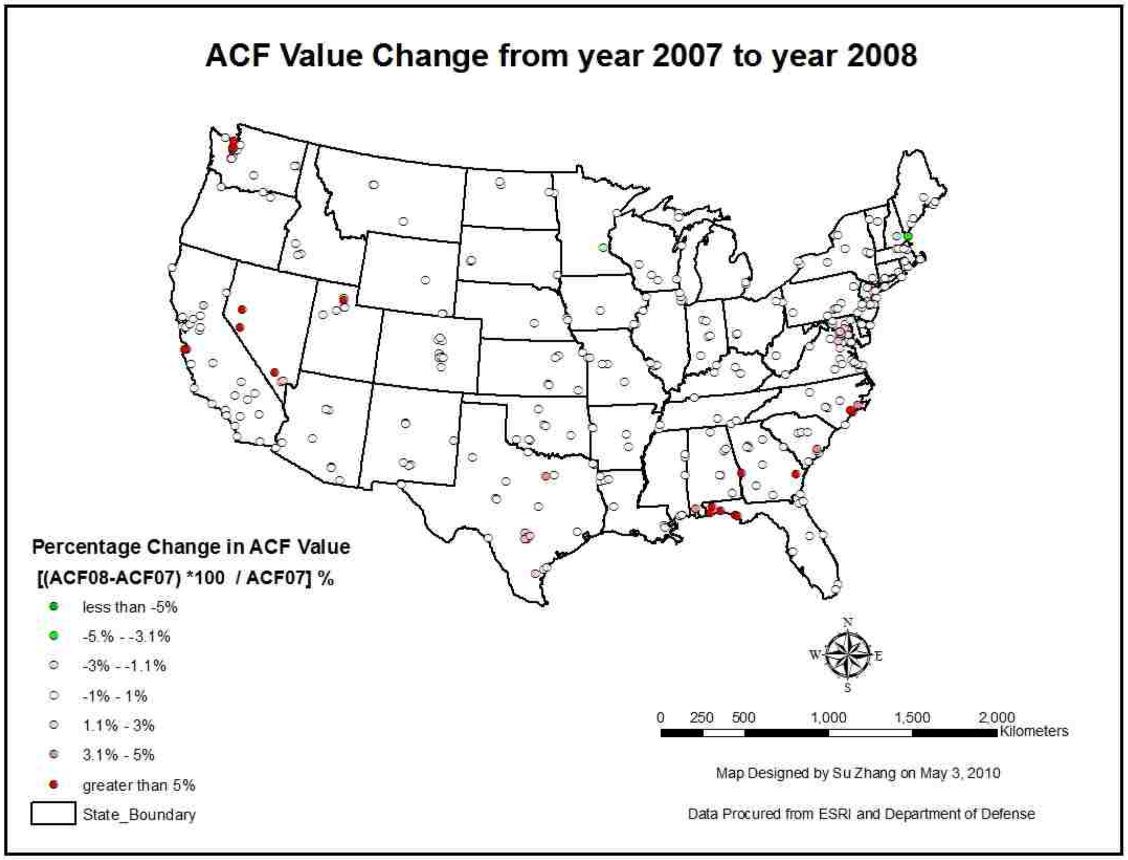

Figure B2.3 Spatial Pattern of the Changes in ACF Value from year 2007 to year 2008

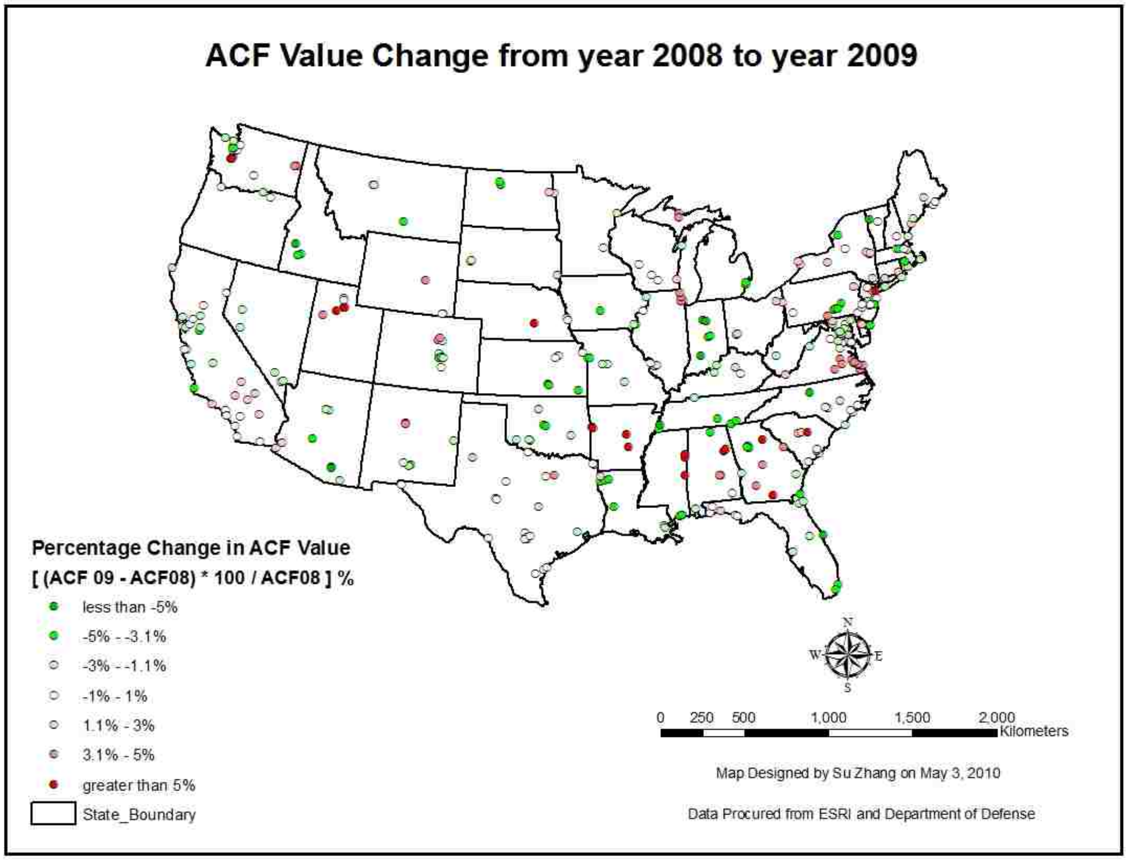

Figure B2.4 Spatial Pattern of the Changes in ACF Value from year 2008 to year 2009

APPENDIX C. ERROR CLASSIFICATION SUMMARY

Exhibit C1. Error Classification Summary for RSMeans CCI Dataset

Table C1. 1 Error Classification Summary for 2005 CCI

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	307	304	306	285	315	321
Overestimates	328	335	334	359	323	317
Perfect Estimates	13	10	8	5	11	11
Inconclusive	1	0	1	0	0	0
TOTAL	649	649	649	649	649	649

Table C1. 2 Error Classification Summary for 2006 CCI

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	301	302	313	283	322	315
Overestimates	332	334	328	356	319	325
Perfect Estimates	15	13	7	10	8	9
Inconclusive	1	0	1	0	0	0
TOTAL	649	649	649	649	649	649

Table C1. 3 Error Classification Summary for 2007 CCI

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	310	307	306	286	301	319
Overestimates	328	333	329	353	339	328
Perfect Estimates	10	9	13	10	9	2
Inconclusive	1	0	1	0	0	0
TOTAL	649	649	649	649	649	649

Table C1. 4 Error Classification Summary for 2008 CCI

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	312	312	308	278	297	314
Overestimates	325	329	331	365	345	327
Perfect Estimates	11	8	9	6	7	8
Inconclusive	1	0	1	0	0	0
TOTAL	649	649	649	649	649	649

Table C1. 5 Error Classification Summary for 2009 CCI

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	309	310	303	276	292	318
Overestimates	338	338	336	365	350	321
Perfect Estimates	1	1	9	8	7	10
Inconclusive	1	0	1	0	0	0
TOTAL	649	649	649	649	649	649

Exhibit C2. Error Classification Summary for DoD ACF Dataset

Table C2. 1 Error Classification Summary for 2005 ACF

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	113	133	145	137	159	164
Overestimates	107	115	174	140	161	162
Perfect Estimates	117	89	18	60	17	11
Inconclusive	0	0	0	0	0	0
TOTAL	337	337	337	337	337	337

Table C2. 2 Error Classification Summary for 2006 ACF

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	98	124	150	151	174	152
Overestimates	94	122	156	167	158	168
Perfect Estimates	145	91	31	19	5	17
Inconclusive	0	0	0	0	0	0
TOTAL	337	337	337	337	337	337

Table C2. 3 Error Classification Summary for 2007 ACF

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	100	116	156	145	157	149
Overestimates	95	124	160	173	176	172
Perfect Estimates	142	97	21	19	4	16
Inconclusive	0	0	0	0	0	0
TOTAL	337	337	337	337	337	337

Table C2. 4 Error Classification Summary for 2008 ACF

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	100	116	156	150	163	156
Overestimates	97	125	157	171	173	172
Perfect Estimates	140	96	24	16	1	9
Inconclusive	0	0	0	0	0	0
TOTAL	337	337	337	337	337	337

Table C2. 5 Error Classification Summary for 2009 ACF

Error Classification	CNN	NN	ST AVG	IDW	Kriging	Spline
Underestimates	99	127	148	143	164	149
Overestimates	85	116	164	172	169	170
Perfect Estimates	153	94	25	22	4	18
Inconclusive	0	0	0	0	0	0
TOTAL	337	337	337	337	337	337

APPENDIX D. NATIONAL LEVEL BI-VARIABLE COMPARISON

Exhibit D1. National Level Bi-Variable Comparison for RSMeans CCI Dataset

Table D1. 1 National Level Bi-Variable Comparison for 2005 CCI

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	116	62					470
2	CNN vs. ST AVG	343		294				11
3	CNN vs. IDW	332			302			14
4	CNN vs. Kriging	335				299		14
5	CNN vs. Spline	339				293		16
6	NN vs. ST AVG		313	325				10
7	NN vs. IDW		282		351			16
8	NN vs. Kriging		293			342		14
9	NN vs. Spline		289				342	18
10	ST AVG vs. IDW			307	325			16
11	ST AVG vs. Kriging			307		328		13
12	ST AVG vs. Spline			313			323	12
13	IDW vs. Kriging				328	286		35
14	IDW vs. Spline				322		302	25
15	Kriging vs. Spline					316	320	13

Table D1. 2 National Level Bi-Variable Comparison for 2006 CCI

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	116	63					469
2	CNN vs. ST AVG	353		283				12
3	CNN vs. IDW	340			291			17
4	CNN vs. Kriging	348				291		9
5	CNN vs. Spline	330				305		13
6	NN vs. ST AVG		318	320				10
7	NN vs. IDW		280		346			23
8	NN vs. Kriging		308			333		8
9	NN vs. Spline		273				358	18
10	ST AVG vs. IDW			306	332			10
11	ST AVG vs. Kriging			294		334		20
12	ST AVG vs. Spline			316			318	14
13	IDW vs. Kriging				328	288		33
14	IDW vs. Spline				308		302	39
15	Kriging vs. Spline					284	332	33

Table D1. 3 National Level Bi-Variable Comparison for 2007 CCI

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	117	59					472
2	CNN vs. ST AVG	347		279				22
3	CNN vs. IDW	306			323			19
4	CNN vs. Kriging	322				304		22
5	CNN vs. Spline	347				284		17
6	NN vs. ST AVG		313	321				14
7	NN vs. IDW		243		383			23
8	NN vs. Kriging		273			355		21
9	NN vs. Spline		277				350	22
10	ST AVG vs. IDW			302	333			13
11	ST AVG vs. Kriging			317		321		10
12	ST AVG vs. Spline			327			313	8
13	IDW vs. Kriging				313	294		42
14	IDW vs. Spline				327		283	39
15	Kriging vs. Spline					339	291	19

Table D1. 4 National Level Bi-Variable Comparison for 2008 CCI

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	119	59					470
2	CNN vs. ST AVG	338		296				14
3	CNN vs. IDW	308			322			18
4	CNN vs. Kriging	321				314		13
5	CNN vs. Spline	330				305		13
6	NN vs. ST AVG		311	328				10
7	NN vs. IDW		247		375			27
8	NN vs. Kriging		282			355		12
9	NN vs. Spline		274				356	19
10	ST AVG vs. IDW			302	323			23
11	ST AVG vs. Kriging			311		321		16
12	ST AVG vs. Spline			318			313	17
13	IDW vs. Kriging				321	297		31
14	IDW vs. Spline				336		277	36
15	Kriging vs. Spline					340	281	28

Table D1. 5 National Level Bi-Variable Comparison for 2009 CCI

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	113	65					470
2	CNN vs. ST AVG	330		302				16
3	CNN vs. IDW	316			315			17
4	CNN vs. Kriging	319				308		21
5	CNN vs. Spline	323				309		16
6	NN vs. ST AVG		295	339				14
7	NN vs. IDW		256		365			28
8	NN vs. Kriging		272			360		17
9	NN vs. Spline		264				363	22
10	ST AVG vs. IDW			318	313			17
11	ST AVG vs. Kriging			327		308		13
12	ST AVG vs. Spline			335			301	12
13	IDW vs. Kriging				310	293		46
14	IDW vs. Spline				334		275	40
15	Kriging vs. Spline					343	286	20

Exhibit D2. National Level Bi-Variable Comparison for DoD ACF Dataset

Table D2. 1 National Level Bi-Variable Comparison for 2005 ACF

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	114	62					161
2	CNN vs. ST AVG	186		147				4
3	CNN vs. IDW	121			119			97
4	CNN vs. Kriging	190				125		22
5	CNN vs. Spline	188				134		15
6	NN vs. ST AVG		154	179				4
7	NN vs. IDW		114		173			50
8	NN vs. Kriging		147			176		14
9	NN vs. Spline		153				173	11
10	ST AVG vs. IDW			142	189			6
11	ST AVG vs. Kriging			146		187		4
12	ST AVG vs. Spline			148			185	4
13	IDW vs. Kriging				181	131		25
14	IDW vs. Spline				182		137	18
15	Kriging vs. Spline					158	163	16

Table D2. 2 National Level Bi-Variable Comparison for 2006 ACF

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	133	52					152
2	CNN vs. ST AVG	203		118				16
3	CNN vs. IDW	191			120			26
4	CNN vs. Kriging	210				120		7
5	CNN vs. Spline	200				119		18
6	NN vs. ST AVG		160	170				7
7	NN vs. IDW		141		182			14
8	NN vs. Kriging		163			170		4
9	NN vs. Spline		147				179	11
10	ST AVG vs. IDW			130	200			7
11	ST AVG vs. Kriging			130		200		7
12	ST AVG vs. Spline			130			197	10
13	IDW vs. Kriging				188	134		15
14	IDW vs. Spline				180		131	26
15	Kriging vs. Spline					147	177	13

Table D2. 3 National Level Bi-Variable Comparison for 2007 ACF

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	119	59					159
2	CNN vs. ST AVG	202		130				5
3	CNN vs. IDW	199			110			28
4	CNN vs. Kriging	223				108		6
5	CNN vs. Spline	203				115		19
6	NN vs. ST AVG		165	166				6
7	NN vs. IDW		160		162			15
8	NN vs. Kriging		168			168		1
9	NN vs. Spline		155				170	12
10	ST AVG vs. IDW			136	198			3
11	ST AVG vs. Kriging			164		168		5
12	ST AVG vs. Spline			128			199	10
13	IDW vs. Kriging				211	118		8
14	IDW vs. Spline				169		148	20
15	Kriging vs. Spline					113	213	11

Table D2. 4 National Level Bi-Variable Comparison for 2008 ACF

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	117	57					163
2	CNN vs. ST AVG	200		134				3
3	CNN vs. IDW	204			107			26
4	CNN vs. Kriging	227				109		1
5	CNN vs. Spline	212				112		13
6	NN vs. ST AVG		170	164				3
7	NN vs. IDW		164		159			14
8	NN vs. Kriging		172			164		1
9	NN vs. Spline		161				165	11
10	ST AVG vs. IDW			136	197			4
11	ST AVG vs. Kriging			159		173		5
12	ST AVG vs. Spline			139			187	11
13	IDW vs. Kriging				205	125		7
14	IDW vs. Spline				181		133	23
15	Kriging vs. Spline					124	204	9

Table D2. 5 National Level Bi-Variable Comparison for 2009 ACF

#	Comparison	CNN	NN	ST AVG	IDW	Kriging	Spline	Equal
1	CNN vs. NN	129	55					153
2	CNN vs. ST AVG	205		124				8
3	CNN vs. IDW	192			119			26
4	CNN vs. Kriging	220				113		4
5	CNN vs. Spline	204				111		22
6	NN vs. ST AVG		159	172				6
7	NN vs. IDW		140		181			16
8	NN vs. Kriging		167			168		2
9	NN vs. Spline		150				173	14
10	ST AVG vs. IDW			123	203			11
11	ST AVG vs. Kriging			144		187		6
12	ST AVG vs. Spline			140			189	8
13	IDW vs. Kriging				207	119		11
14	IDW vs. Spline				192		120	25
15	Kriging vs. Spline					117	207	13

APPENDIX E NATIONAL-LEVEL ERROR PERCENTAGE COMPARISON

Exhibit E1. National-Level Error Percentage Comparison for RSMeans CCI Dataset

Table E1. 1 National-Level Error Percentage Comparison for 2005 CCI

Interpolation Methods Error for CCI 2005						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	177	213	96	105	58
	percentage	27%	33%	15%	16%	9%
NN	count	155	190	104	128	72
	percentage	24%	29%	16%	20%	11%
ST AVG	count	135	200	124	132	58
	percentage	21%	31%	19%	20%	9%
IDW	count	141	205	118	145	40
	percentage	22%	32%	18%	22%	6%
Kriging	count	126	201	122	162	38
	percentage	19%	31%	19%	25%	6%
Spline	count	134	196	129	139	51
	percentage	21%	30%	20%	21%	8%

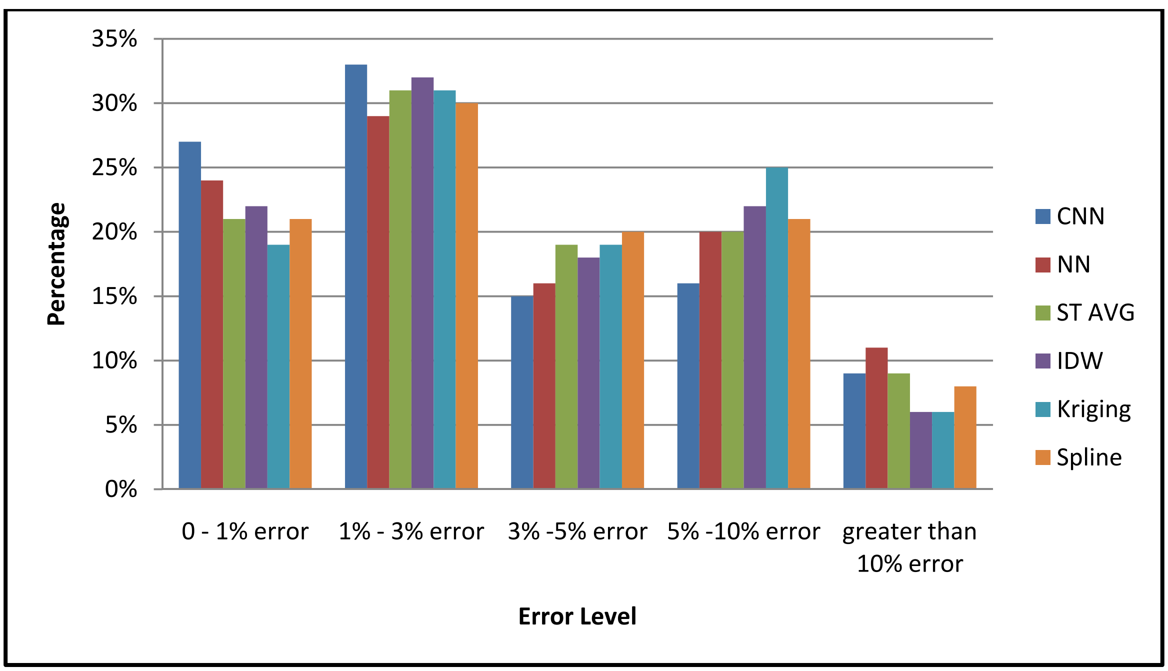

Figure E1.1 National-Level Error Percentage Comparison for 2005 CCI

Table E1. 2 National-Level Error Percentage Comparison for 2006 CCI

Interpolation Methods Error for CCI 2006						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	180	209	94	109	57
	percentage	28%	32%	14%	17%	9%
NN	count	162	184	108	123	72
	percentage	25%	28%	17%	19%	11%
ST AVG	count	123	210	122	135	59
	percentage	19%	32%	19%	21%	9%
IDW	count	146	200	125	139	39
	percentage	23%	31%	19%	21%	6%
Kriging	count	140	189	123	153	44
	percentage	21%	29%	19%	24%	7%
Spline	count	136	206	126	135	46
	percentage	21%	32%	19%	21%	7%

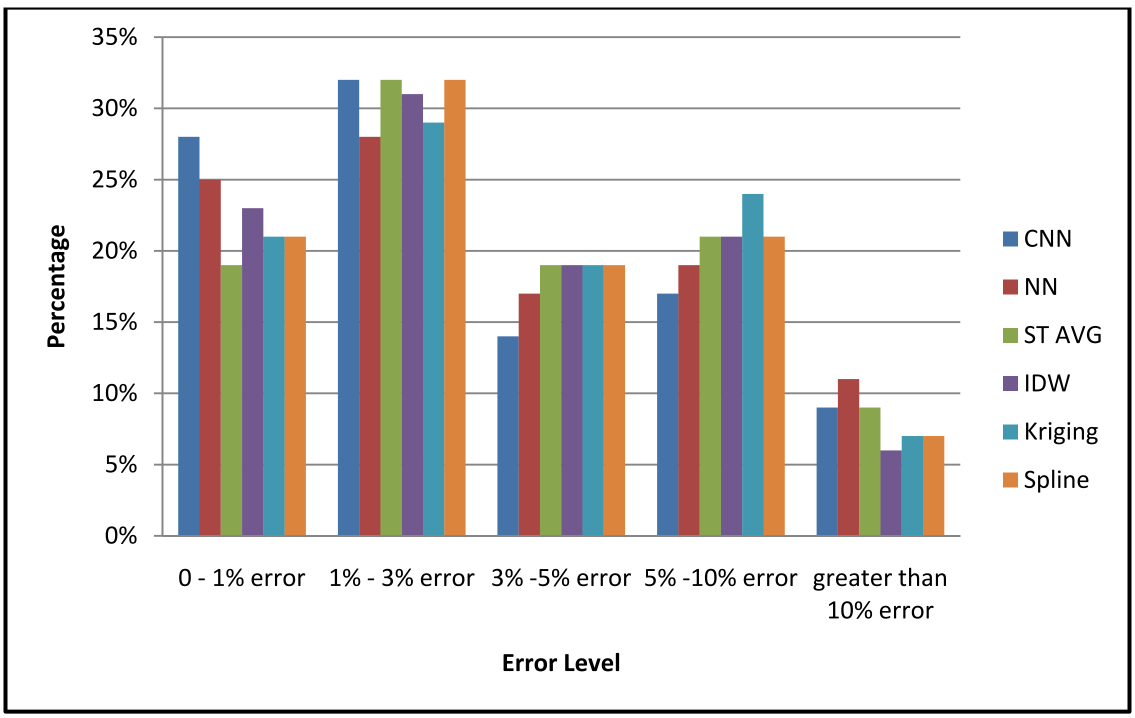

Figure E1.2 National-Level Error Percentage Comparison for 2006 CCI

Table E1. 3 National-Level Error Percentage Comparison for 2007 CCI

Interpolation Methods Error for CCI 2007						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	194	198	104	117	36
	percentage	30%	30%	16%	18%	6%
NN	count	168	179	120	126	56
	percentage	26%	28%	18%	19%	9%
ST AVG	count	140	225	113	119	52
	percentage	22%	35%	17%	18%	8%
IDW	count	163	214	123	121	28
	percentage	25%	33%	19%	19%	4%
Kriging	count	155	220	120	123	31
	percentage	24%	34%	18%	19%	5%
Spline	count	155	210	118	133	33
	percentage	24%	32%	18%	21%	5%

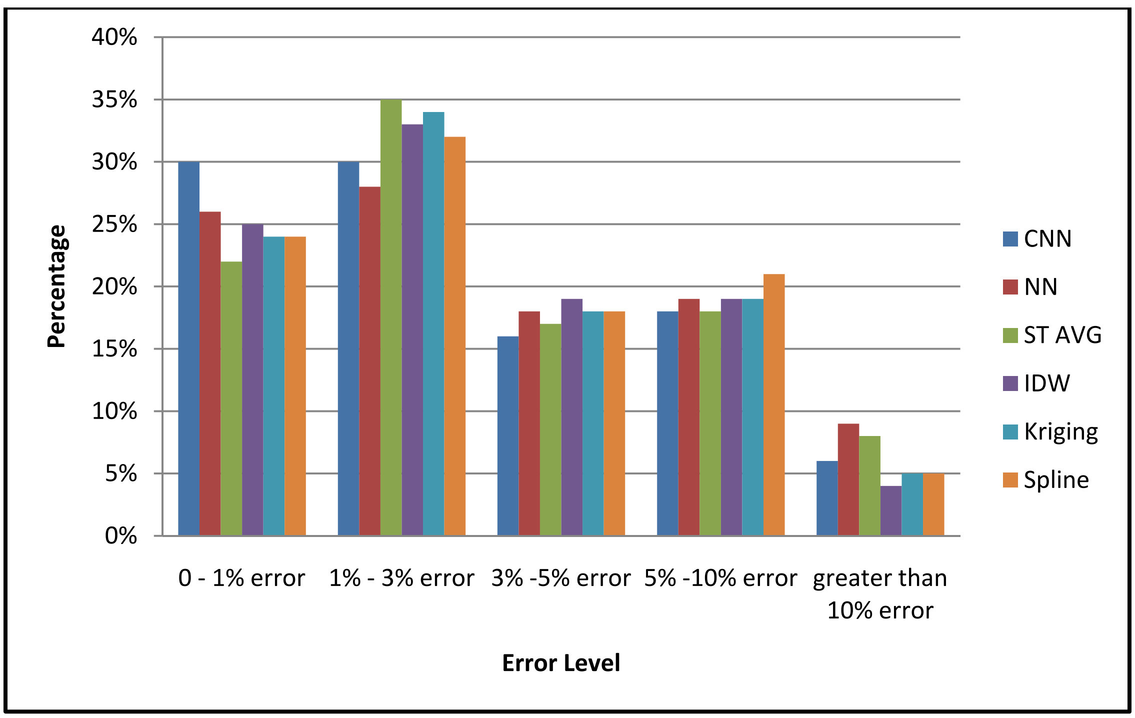

Figure E1.3 National-Level Error Percentage Comparison for 2007 CCI

Table E1. 4 National-Level Error Percentage Comparison for 2008 CCI

Interpolation Methods Error for CCI 2008						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	178	210	126	103	32
	percentage	28%	32%	19%	16%	5%
NN	count	156	194	121	128	50
	percentage	24%	30%	18%	20%	8%
ST AVG	count	139	246	100	118	46
	percentage	22%	38%	15%	18%	7%
IDW	count	168	220	112	130	19
	percentage	26%	34%	17%	20%	3%
Kriging	count	160	215	122	130	22
	percentage	25%	33%	18%	20%	4%
Spline	count	147	216	138	119	29
	percentage	23%	33%	21%	18%	5%

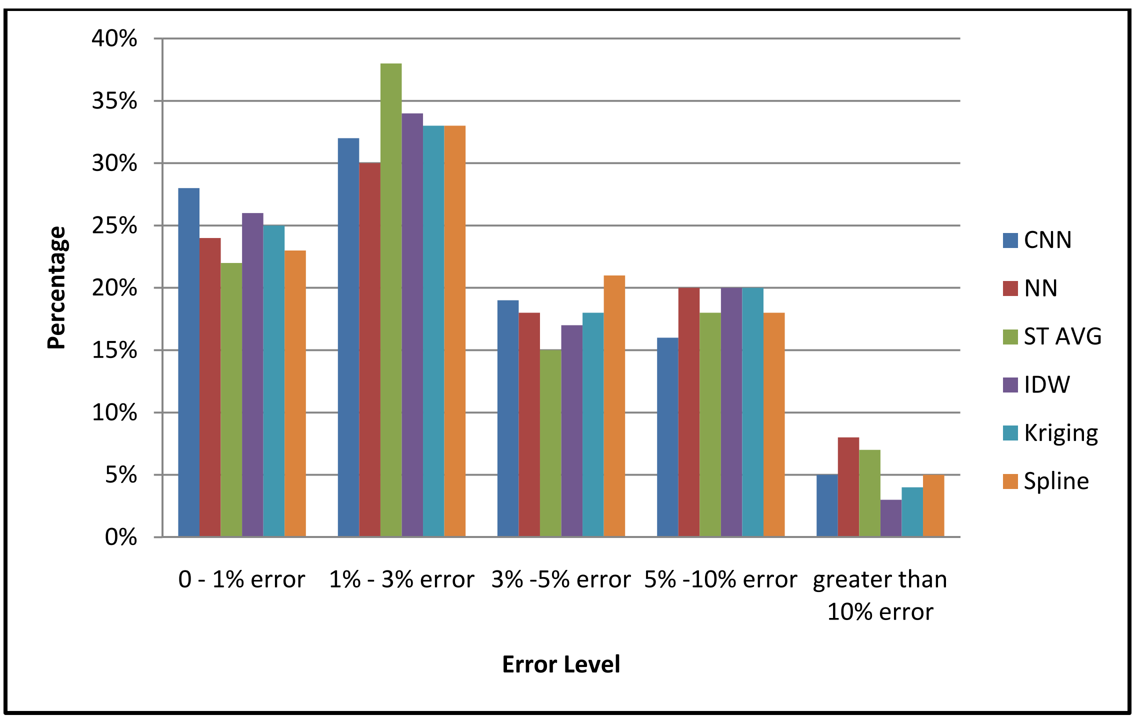

Figure E1.4 National-Level Error Percentage Comparison for 2008 CCI

Table E1. 5 National-Level Error Percentage Comparison for 2009 CCI

Interpolation Methods Error for CCI 2009						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	175	215	111	112	36
	percentage	27%	33%	17%	17%	6%
NN	count	153	193	114	131	58
	percentage	24%	30%	18%	20%	9%
ST AVG	count	153	222	109	122	43
	percentage	24%	34%	17%	19%	6%
IDW	count	149	213	133	128	26
	percentage	23%	33%	20%	20%	4%
Kriging	count	143	210	138	131	27
	percentage	22%	33%	21%	20%	4%
Spline	count	132	227	129	130	31
	percentage	20%	35%	20%	20%	5%

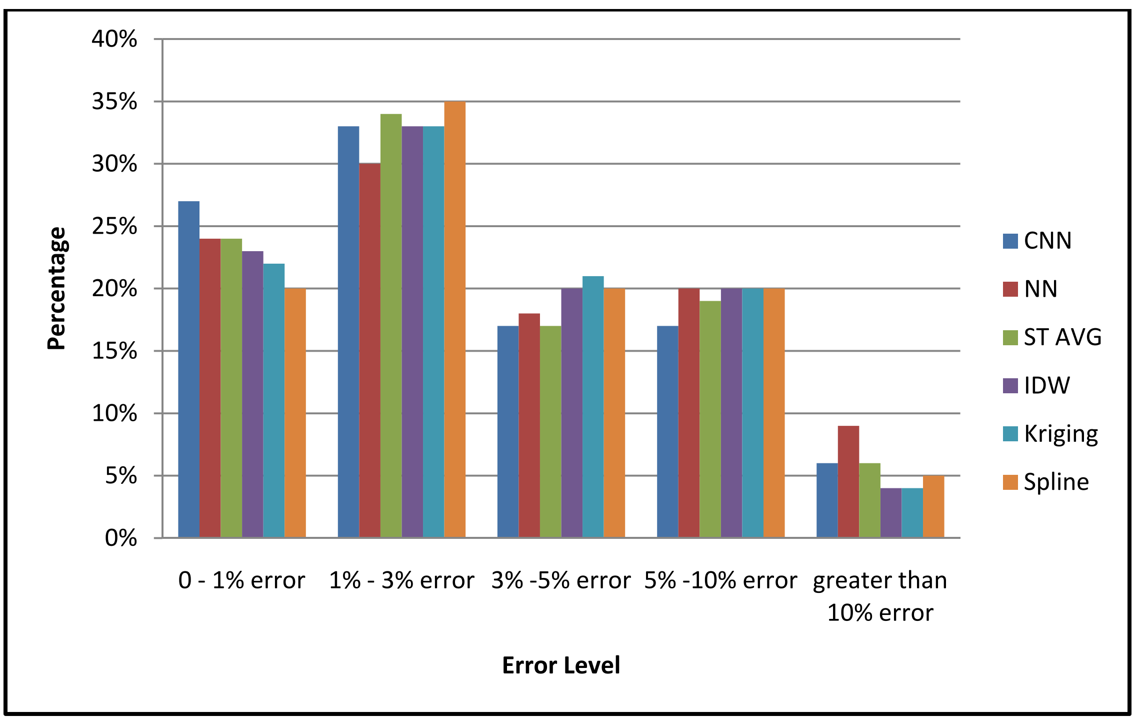

Figure E1.5 National-Level Error Percentage Comparison for 2009 CCI

Exhibit E2. National-Level Error Percentage Comparison for DoD ACF Dataset

Table E2. 1 National-Level Error Percentage Comparison for ACF 2005

Interpolation Methods Error for ACF 2005						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	137	41	37	80	42
	percentage	41%	12%	11%	24%	12%
NN	count	101	40	42	71	83
	percentage	30%	12%	12%	21%	25%
ST AVG	count	49	96	63	101	28
	percentage	15%	29%	18%	30%	8%
IDW	count	119	56	44	84	34
	percentage	35%	17%	13%	25%	10%
Kriging	count	98	72	45	91	31
	percentage	29%	22%	13%	27%	9%
Spline	count	100	77	44	85	31
	percentage	30%	23%	13%	25%	9%

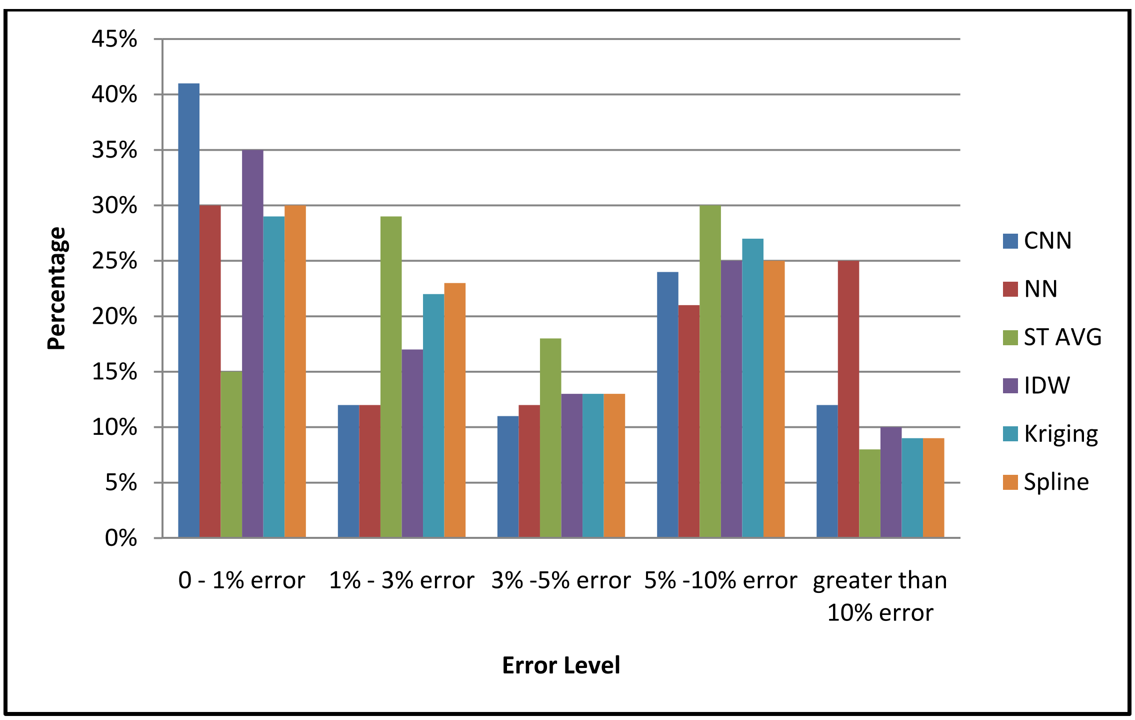

Figure E2.1. National-Level Error Percentage Comparison for ACF 2005

Table E2. 2 National-Level Error Percentage Comparison for 2006 ACF

Interpolation Methods Error for ACF 2006						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	158	52	31	58	38
	percentage	47%	16%	9%	17%	11%
NN	count	102	56	33	72	74
	percentage	30%	17%	10%	21%	22%
ST AVG	count	56	84	66	101	30
	percentage	16%	25%	20%	30%	9%
IDW	count	125	77	43	69	23
	percentage	37%	23%	13%	20%	7%
Kriging	count	97	94	51	64	31
	percentage	29%	28%	15%	19%	9%
Spline	count	104	81	68	59	25
	percentage	31%	24%	20%	18%	7%

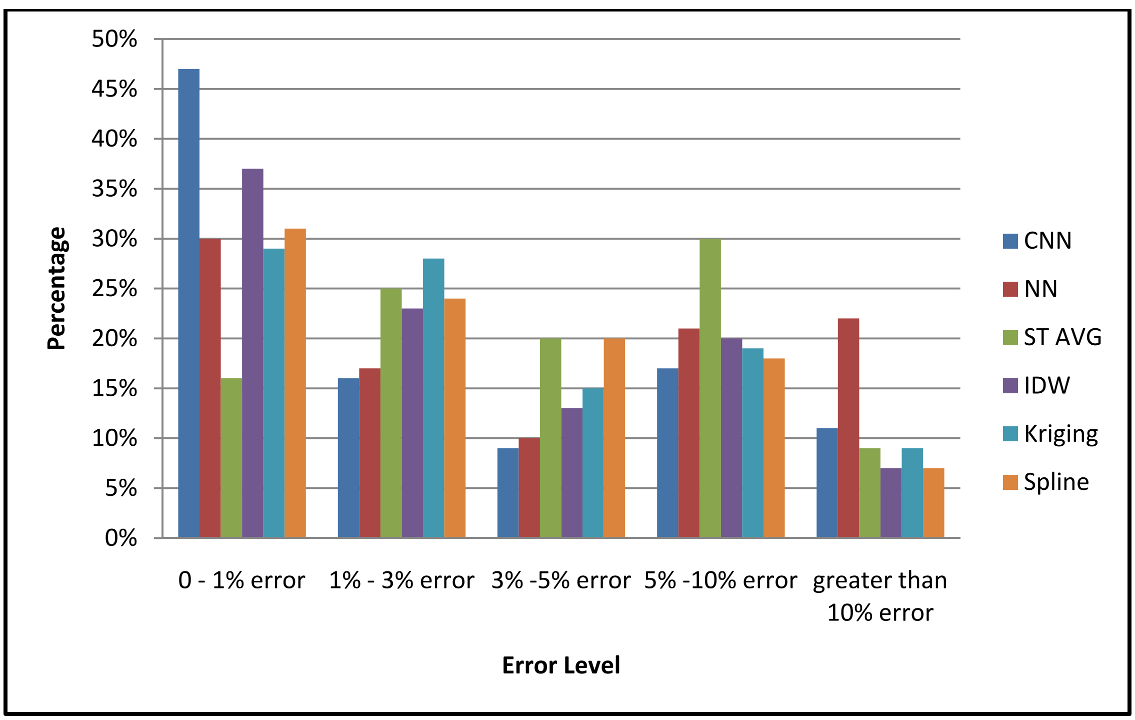

Figure E2.2 National-Level Error Percentage Comparison for 2006 ACF

Table E2. 3 National-Level Error Percentage Comparison for 2007 ACF

Interpolation Methods Error for ACF 2007						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	159	45	34	56	43
	percentage	47%	13%	10%	17%	13%
NN	count	113	45	45	58	76
	percentage	34%	13%	13%	17%	23%
ST AVG	count	48	97	50	107	35
	percentage	14%	29%	15%	32%	10%
IDW	count	114	67	52	77	27
	percentage	34%	20%	15%	23%	8%
Kriging	count	50	90	68	97	32
	percentage	15%	27%	20%	29%	9%
Spline	count	94	86	56	74	27
	percentage	28%	26%	17%	22%	8%

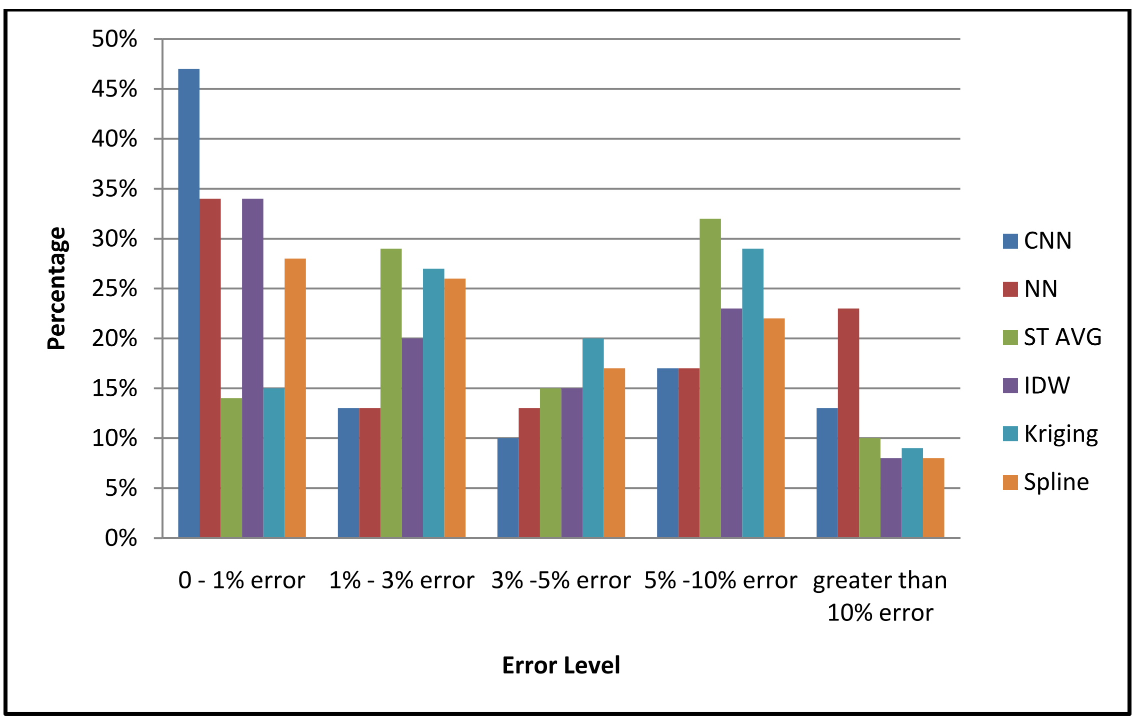

Figure E2.3 National-Level Error Percentage Comparison for 2007 ACF

Table E2. 4 National-Level Error Percentage Comparison for 2008 ACF

Interpolation Methods Error for ACF 2008						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	156	45	39	51	46
	percentage	46%	13%	12%	15%	14%
NN	count	110	47	39	59	82
	percentage	33%	14%	12%	18%	24%
ST AVG	count	44	78	69	110	36
	percentage	13%	23%	20%	33%	11%
IDW	count	110	59	62	77	29
	percentage	33%	18%	18%	23%	9%
Kriging	count	60	83	71	91	32
	percentage	18%	25%	21%	27%	9%
Spline	count	86	79	62	81	29
	percentage	26%	23%	18%	24%	9%

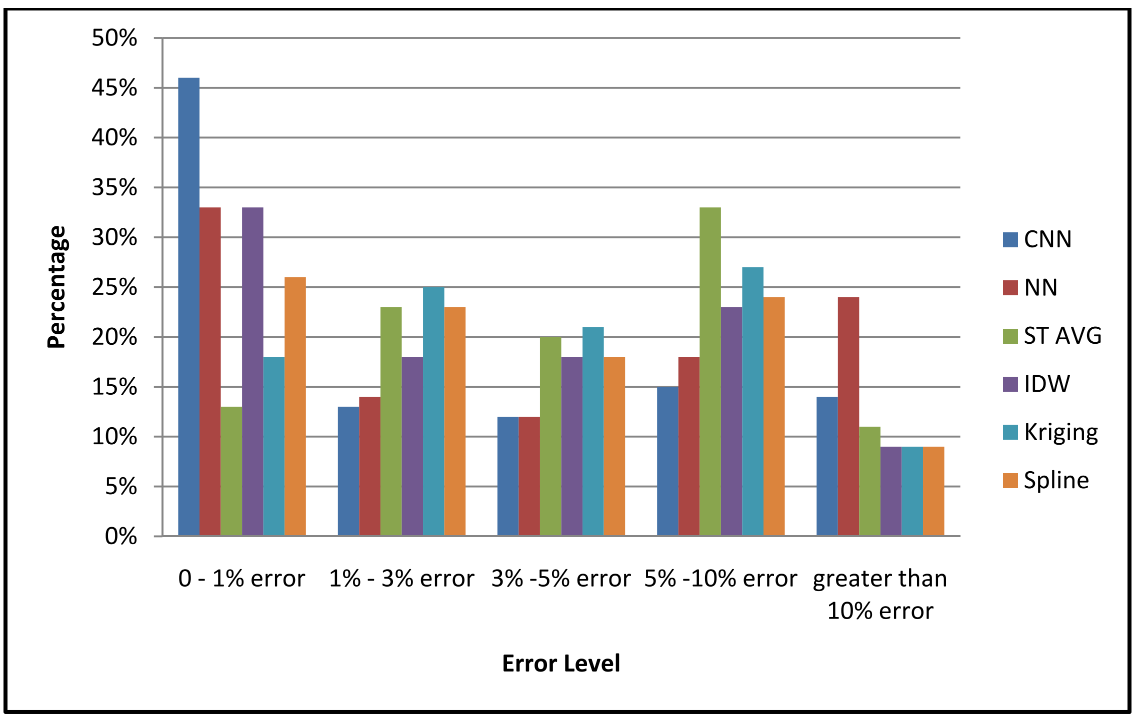

Figure E2.4 National-Level Error Percentage Comparison for 2008 ACF

Table E2. 5 National-Level Error Percentage Comparison for 2009 ACF

Interpolation Methods Error for ACF 2009						
Interpolation Methods	Comparison	Very Low (0-1%)	Low (1%-3%)	Medium (3%-5%)	High (5%-10%)	Very High (>10%)
CNN	count	162	39	35	61	40
	percentage	48%	12%	10%	18%	12%
NN	count	101	50	43	72	71
	percentage	30%	15%	13%	21%	21%
ST AVG	count	57	103	52	85	40
	percentage	17%	31%	15%	25%	12%
IDW	count	127	68	44	67	31
	percentage	38%	20%	13%	20%	9%
Kriging	count	71	103	50	85	28
	percentage	21%	31%	15%	25%	8%
Spline	count	100	86	54	66	31
	percentage	30%	26%	16%	20%	9%

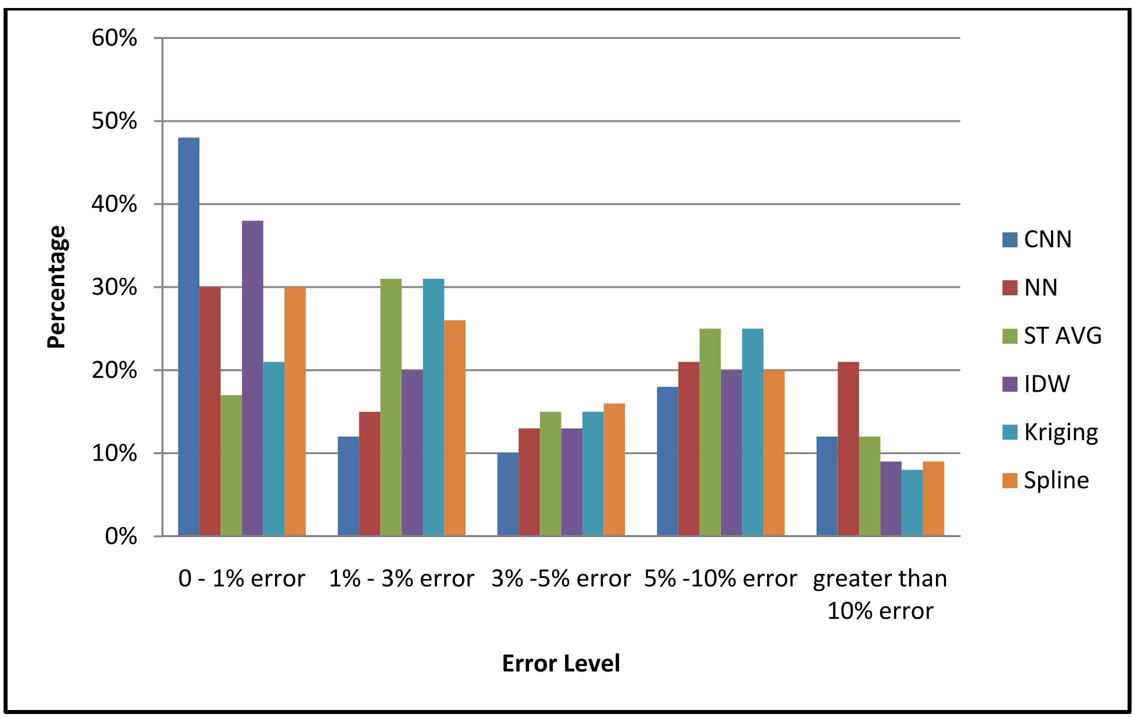

Figure E2.5 National-Level Error Percentage Comparison for 2009 ACF

APPENDIX F. ERROR DESCRIPTIVE STATISTICS SUMMARY

Exhibit F1 Summary of Error Descriptive Statistics for RSMeans CCI Dataset

Table F1.1 Summary of Error Descriptive Statistics for CCI 2005

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.10	2.40	2.51	2.40	2.60	2.50
Mean	3.17	3.79	3.72	3.31	3.46	3.45
Standard Deviation	3.32	4.09	3.76	2.96	2.98	3.09

Table F1.2 Summary of Error Descriptive Statistics for CCI 2006

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.00	2.40	2.59	2.50	2.70	2.50
Mean	3.17	3.79	3.80	3.31	3.47	3.39
Standard Deviation	3.34	4.14	3.76	2.93	2.98	3.10

Table F1.3 Summary of Error Descriptive Statistics for CCI 2007

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.00	2.40	2.30	2.20	2.20	2.30
Mean	2.96	3.55	3.50	3.00	3.08	3.16
Standard Deviation	3.09	3.72	3.63	2.78	2.82	2.89

Table F1.4 Summary of Error Descriptive Statistics for CCI 2008

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	1.90	2.40	2.18	2.20	2.20	2.30
Mean	2.89	3.55	3.41	2.94	3.01	3.14
Standard Deviation	2.93	3.81	3.66	2.80	2.82	3.00

Table F1.5 Summary of Error Descriptive Statistics for CCI 2009

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.02	2.60	2.14	2.30	2.43	2.40
Mean	3.06	3.73	3.39	3.06	3.15	3.26
Standard Deviation	3.09	3.89	3.64	2.82	2.87	3.04

Exhibit F2 Summary of Error Descriptive Statistics for DoD ACF Dataset

Table F2.1 Summary of Error Descriptive Statistics for ACF 2005

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.00	4.00	3.54	2.50	2.90	2.60
Mean	4.29	6.90	4.59	4.16	4.41	4.25
Standard Deviation	5.25	8.21	3.89	4.92	4.52	4.56

Table F2.2 Summary of Error Descriptive Statistics for ACF 2006

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	1.00	3.00	4.00	2.10	2.40	2.50
Mean	3.57	5.76	4.87	3.65	3.98	3.86
Standard Deviation	5.10	6.77	4.57	4.66	4.77	4.61

Table F2.3 Summary of Error Descriptive Statistics for ACF 2007

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.00	3.00	3.87	2.50	3.90	2.70
Mean	3.89	5.97	5.07	4.04	4.91	4.16
Standard Deviation	5.60	7.46	4.70	4.91	4.75	4.89

Table F2.4 Summary of Error Descriptive Statistics for ACF 2008

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	2.00	4.00	4.33	3.00	3.80	3.10
Mean	4.02	6.29	5.36	4.25	4.89	4.44
Standard Deviation	5.71	7.50	4.68	4.85	4.77	4.79

Table F2.5 Summary of Error Descriptive Statistics for ACF 2009

	CNN	NN	ST AVG	IDW	Kriging	Spline
Median	1.00	4.00	3.23	2.10	3.00	2.40
Mean	3.97	5.88	4.93	3.96	4.52	4.15
Standard Deviation	5.90	6.79	4.89	4.92	4.87	4.84

APPENDIX G. 2D DISTRIBUTION VISUALIZATION

Exhibit G1. 2D Distribution Visualization for RSMeans CCI Dataset

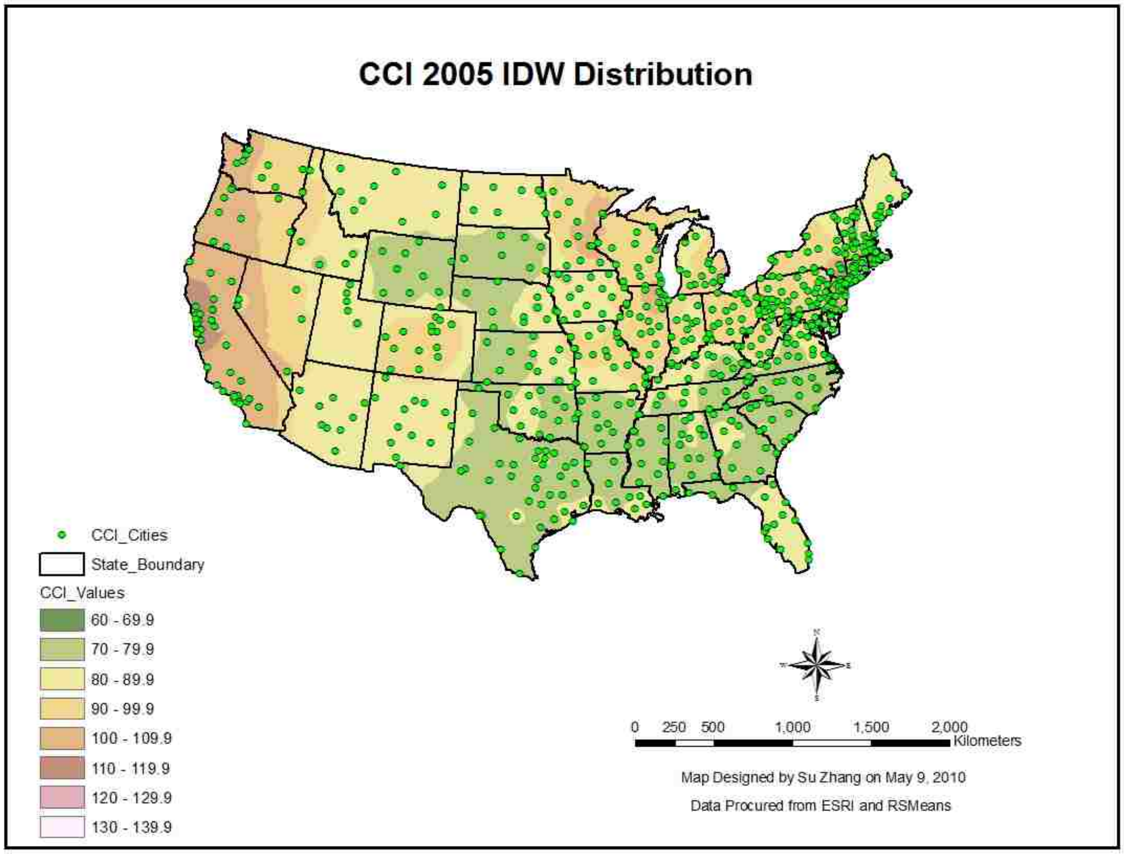

Figure G1.1 2D IDW Distribution for 2005 CCI

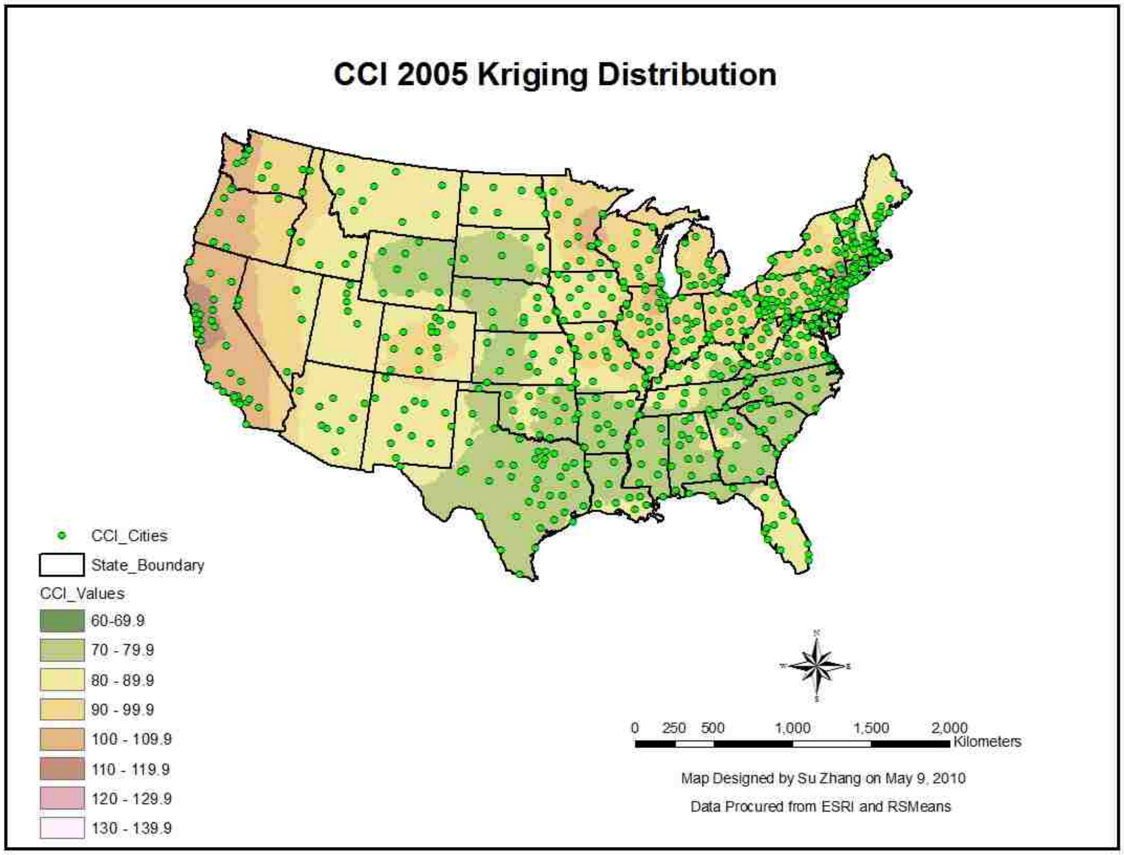

Figure G1.2 2D Kriging Distribution for 2005 CCI

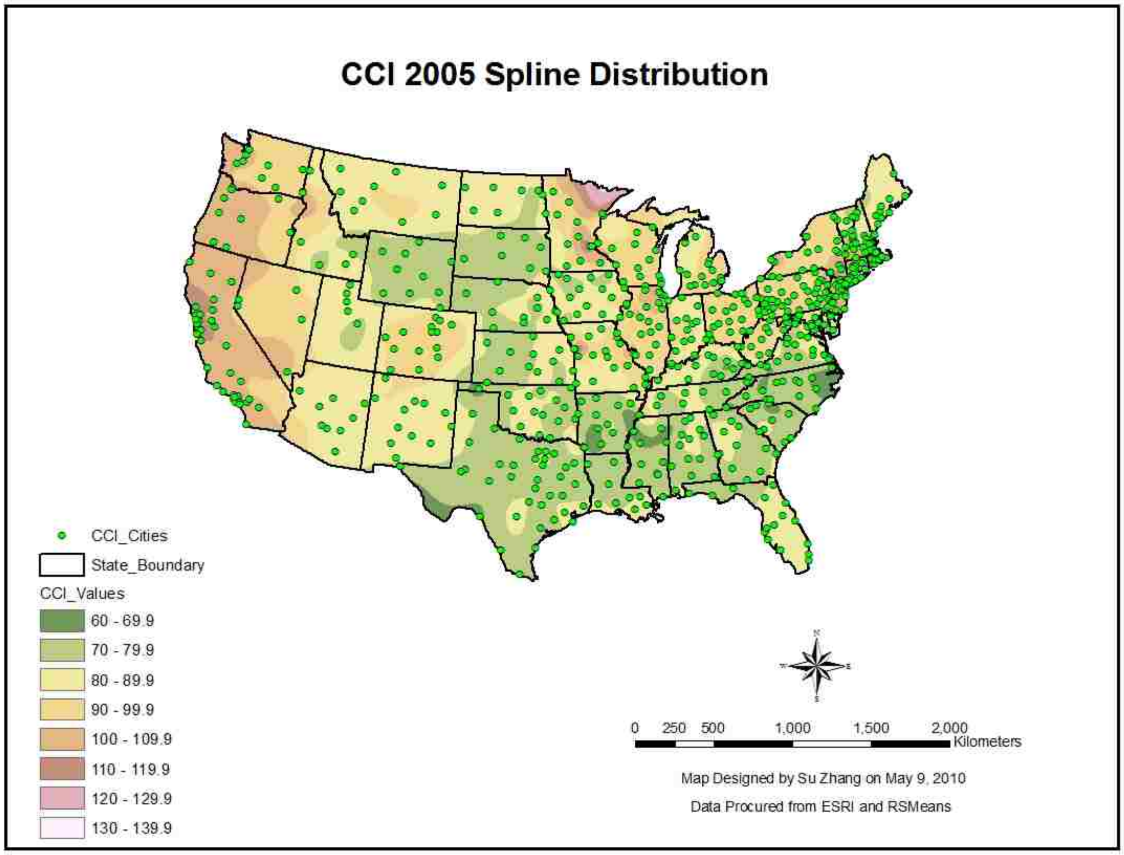

Figure G1.3 2D Spline Distribution for 2005 CCI

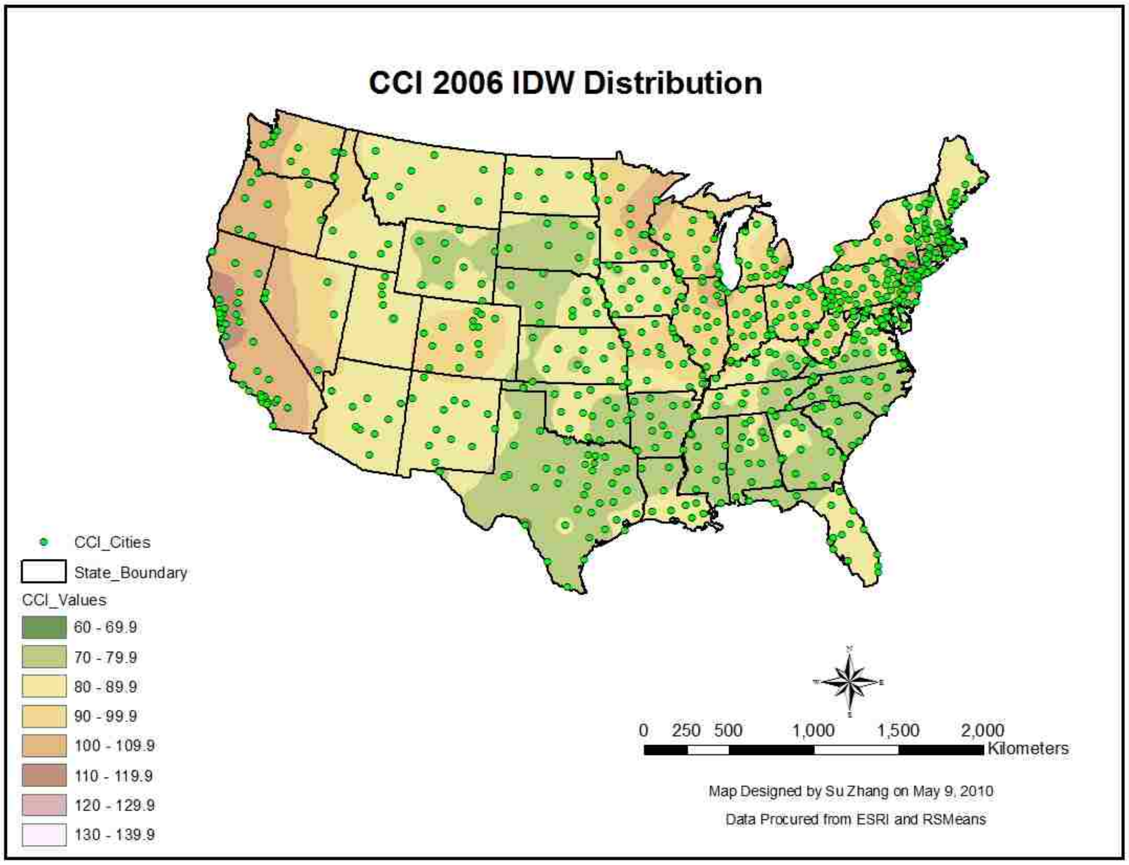

Figure G1.4 2D IDW Distribution for 2006 CCI

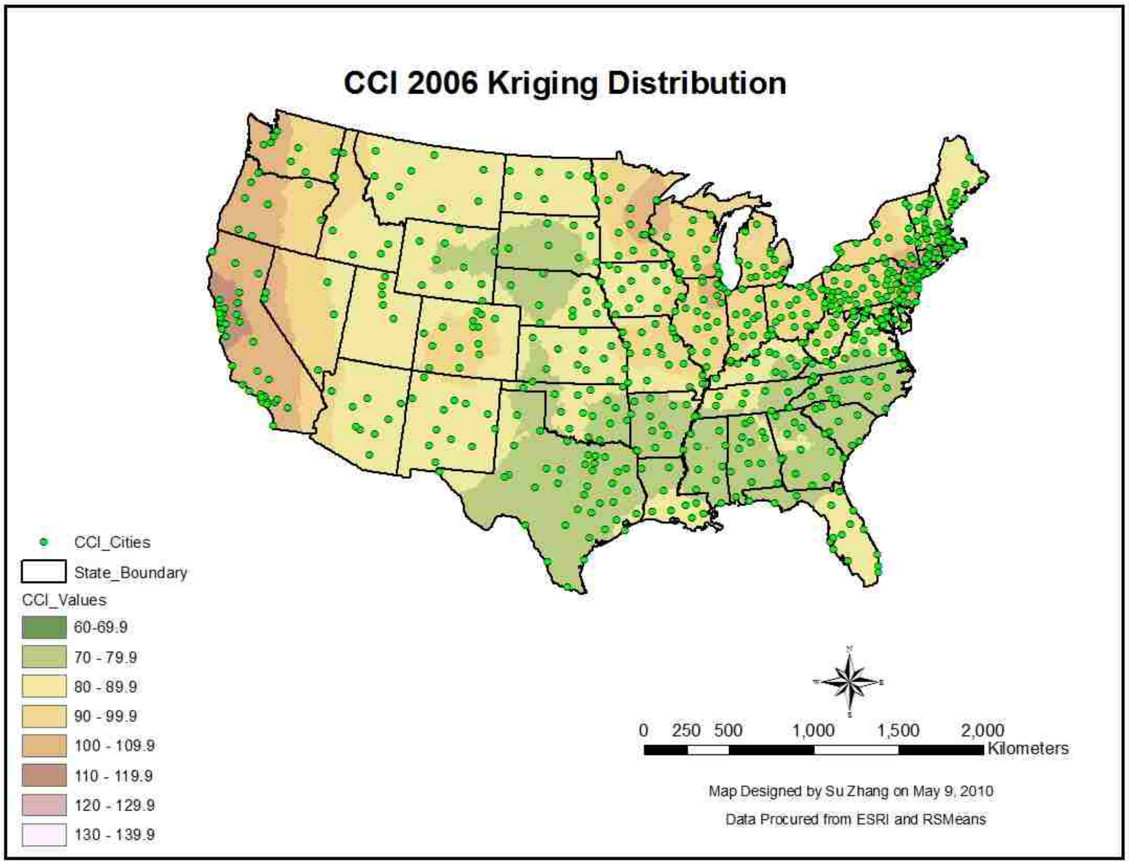

Figure G1.5 2D Kriging Distribution for 2006 CCI

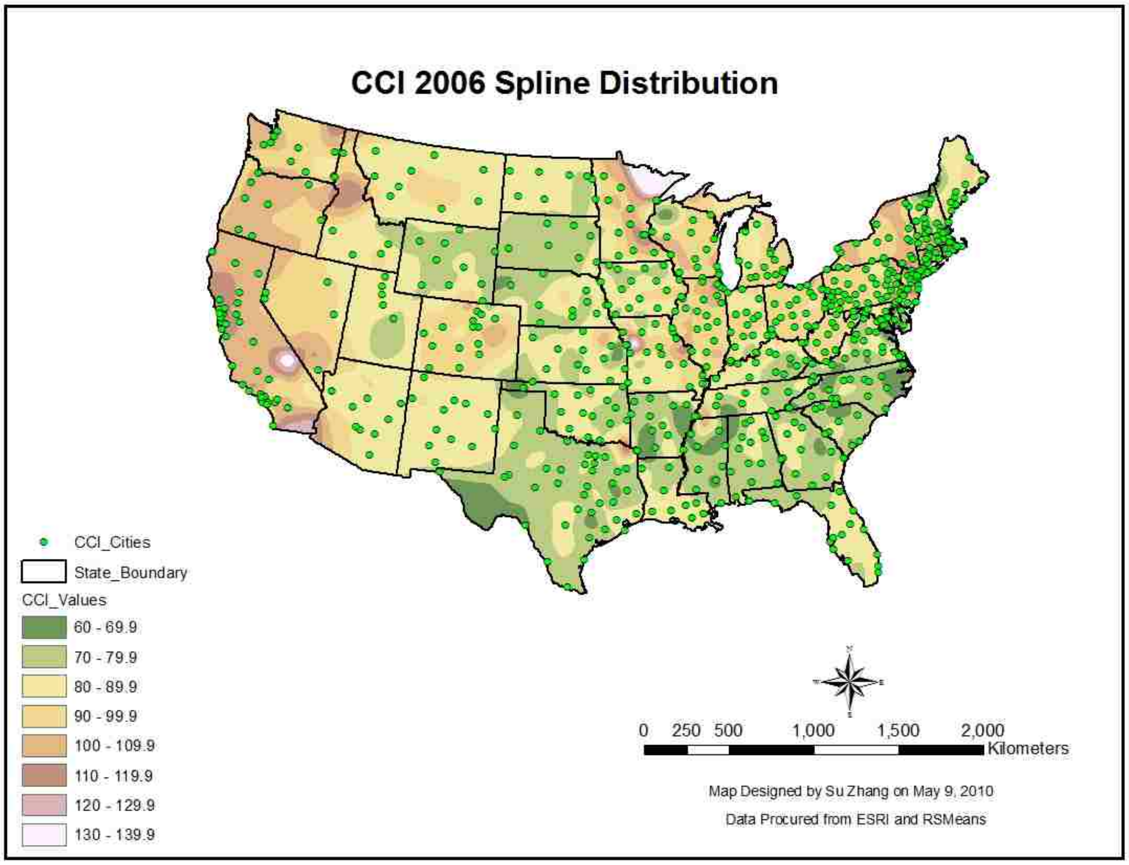

Figure G1.6 2D Spline Distribution for 2006 CCI

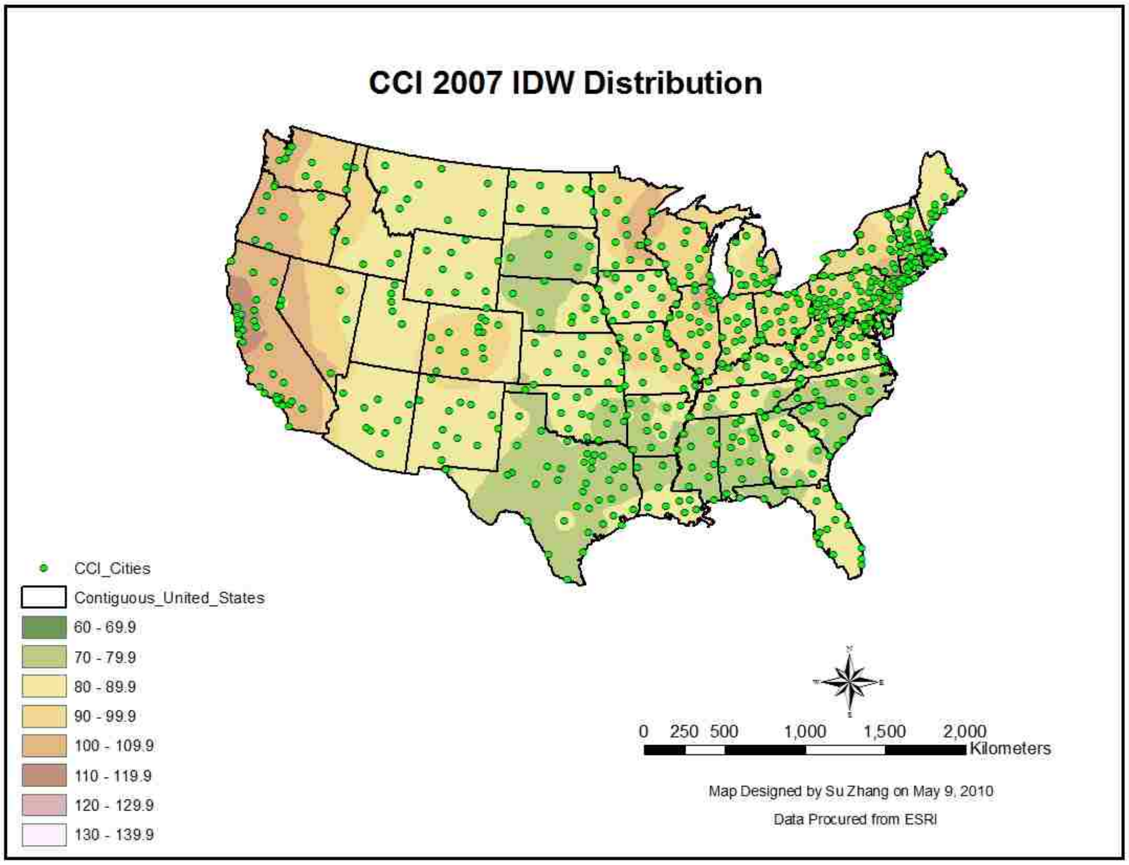

Figure G1.7 2D IDW Distribution for 2007 CCI

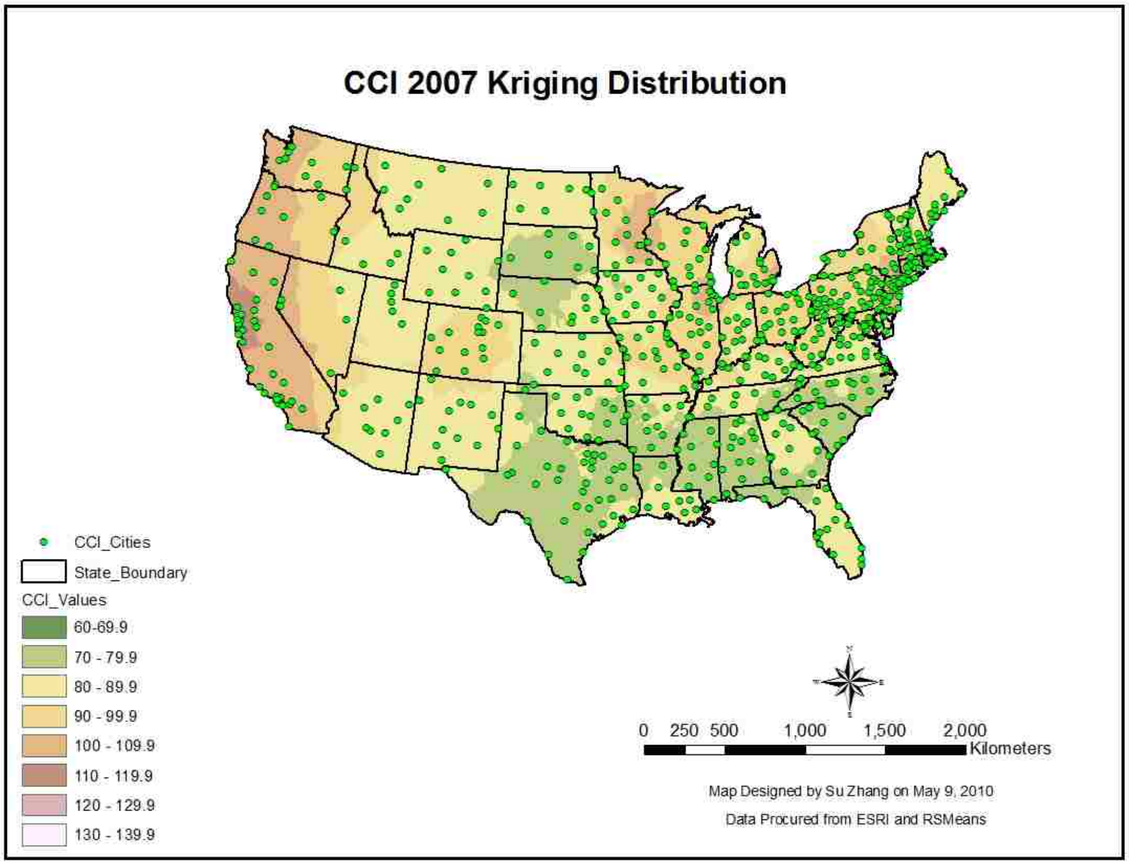

Figure G1.8 2D Kriging Distribution for 2007 CCI

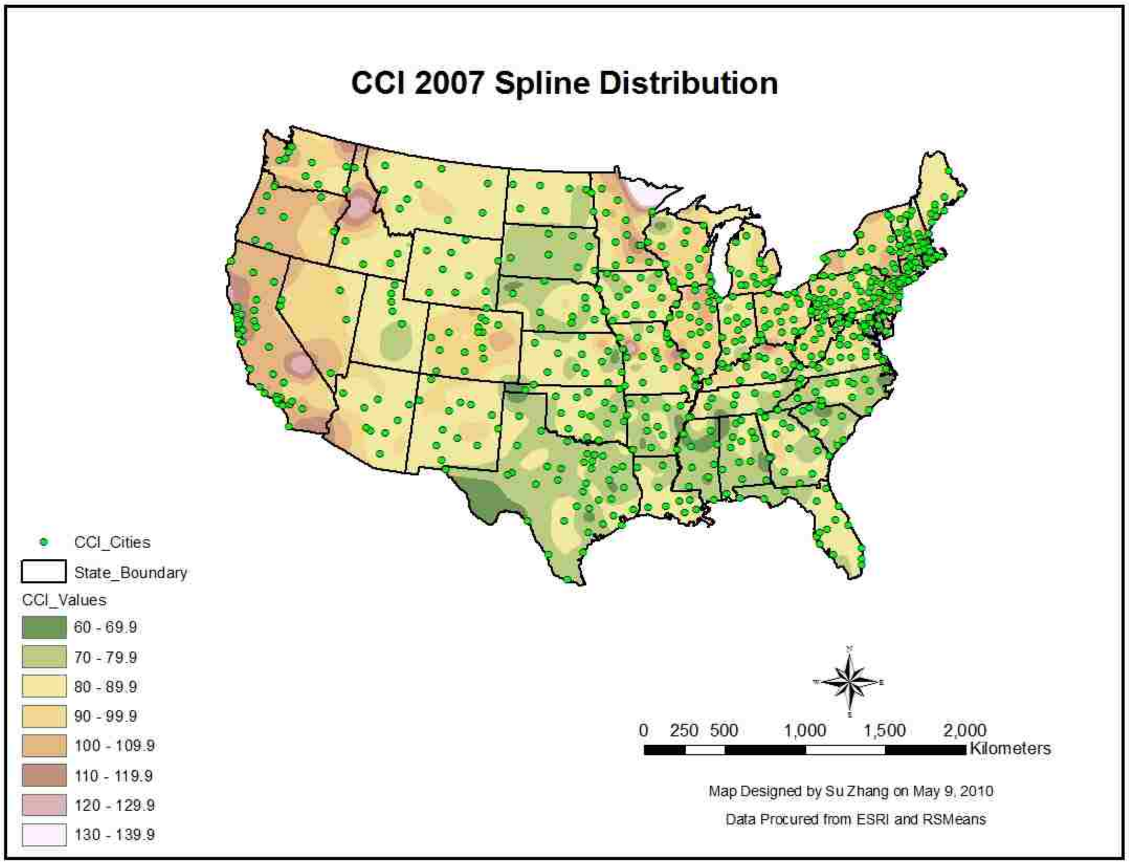

Figure G1.9 2D Spline Distribution for 2007 CCI

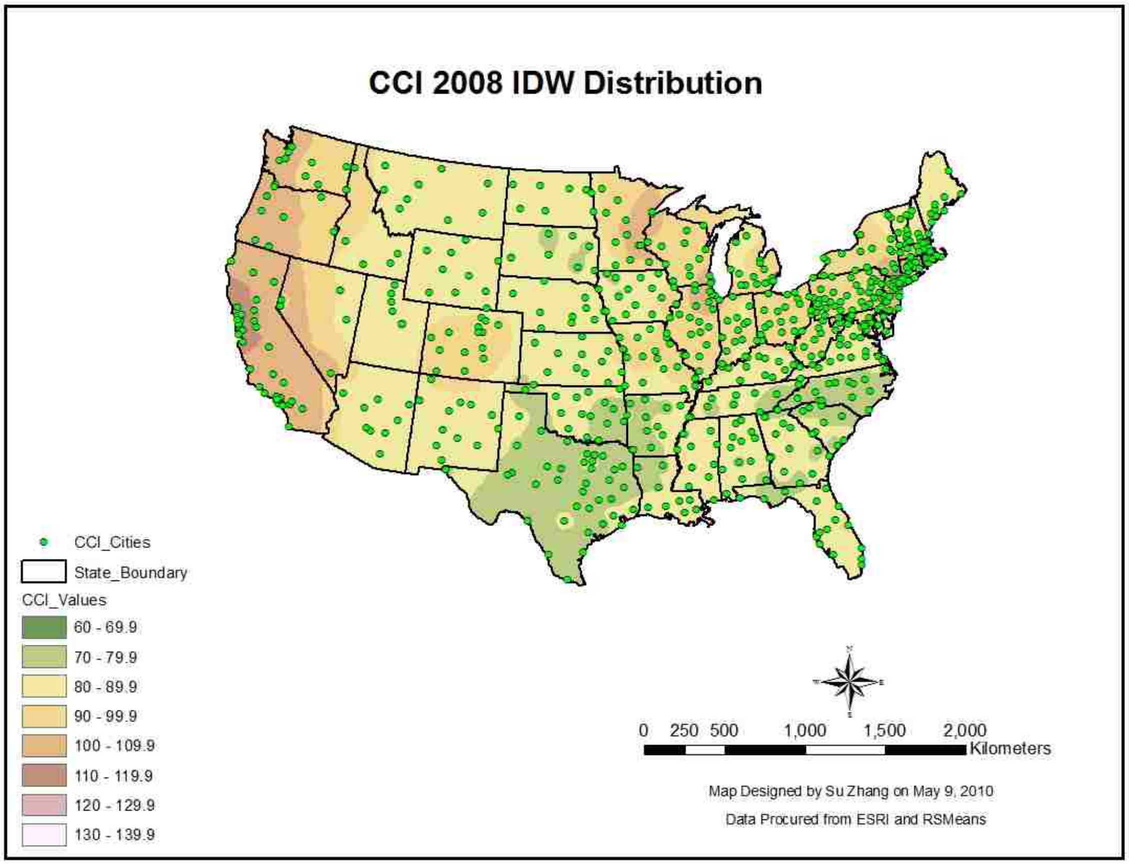

Figure G1.10 2D IDW Distribution for 2008 CCI

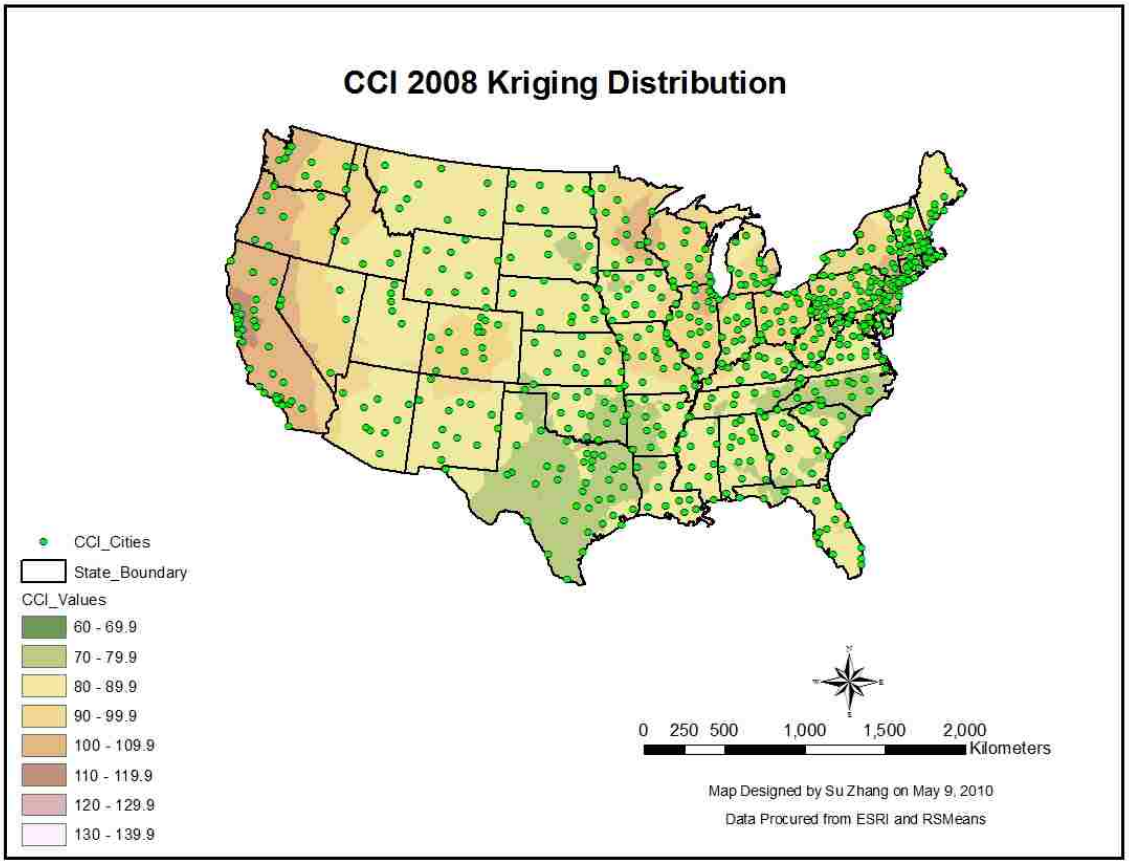

Figure G1.11 2D Kriging Distribution for 2008 CCI

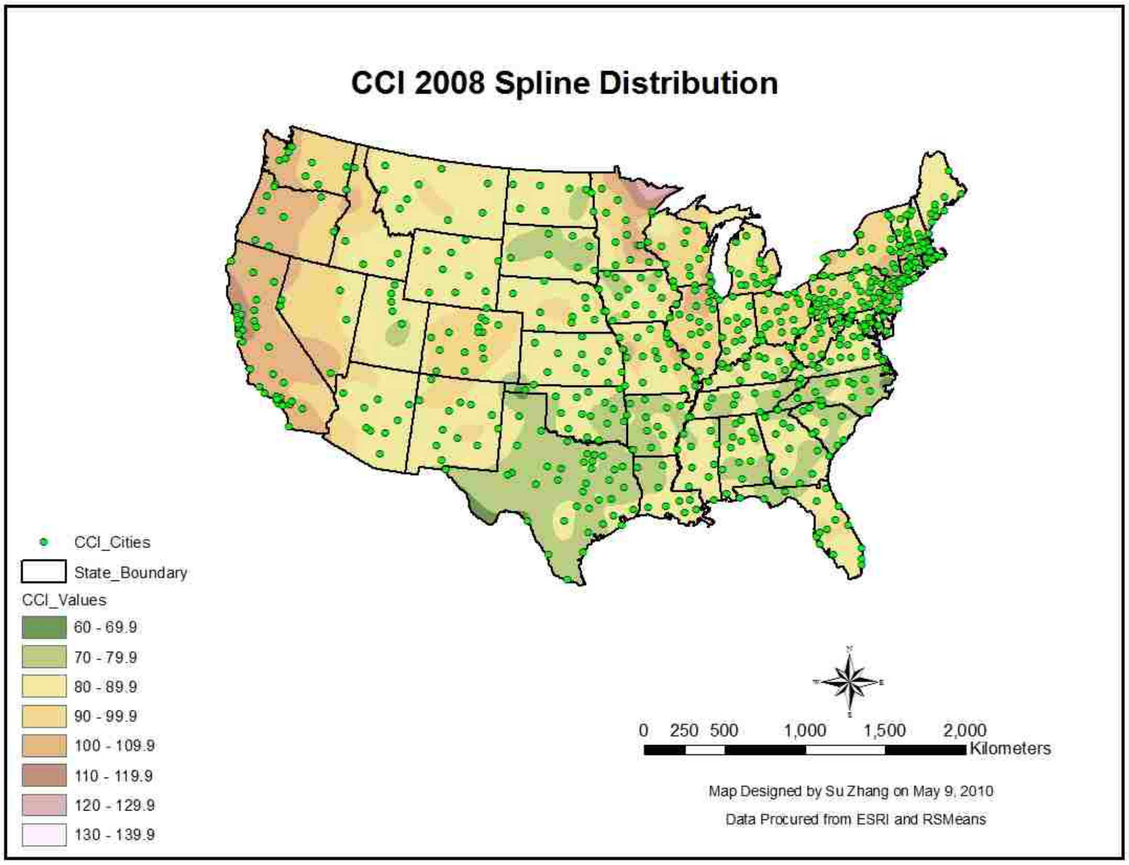

Figure G1.12 2D Spline Distribution for 2008 CCI

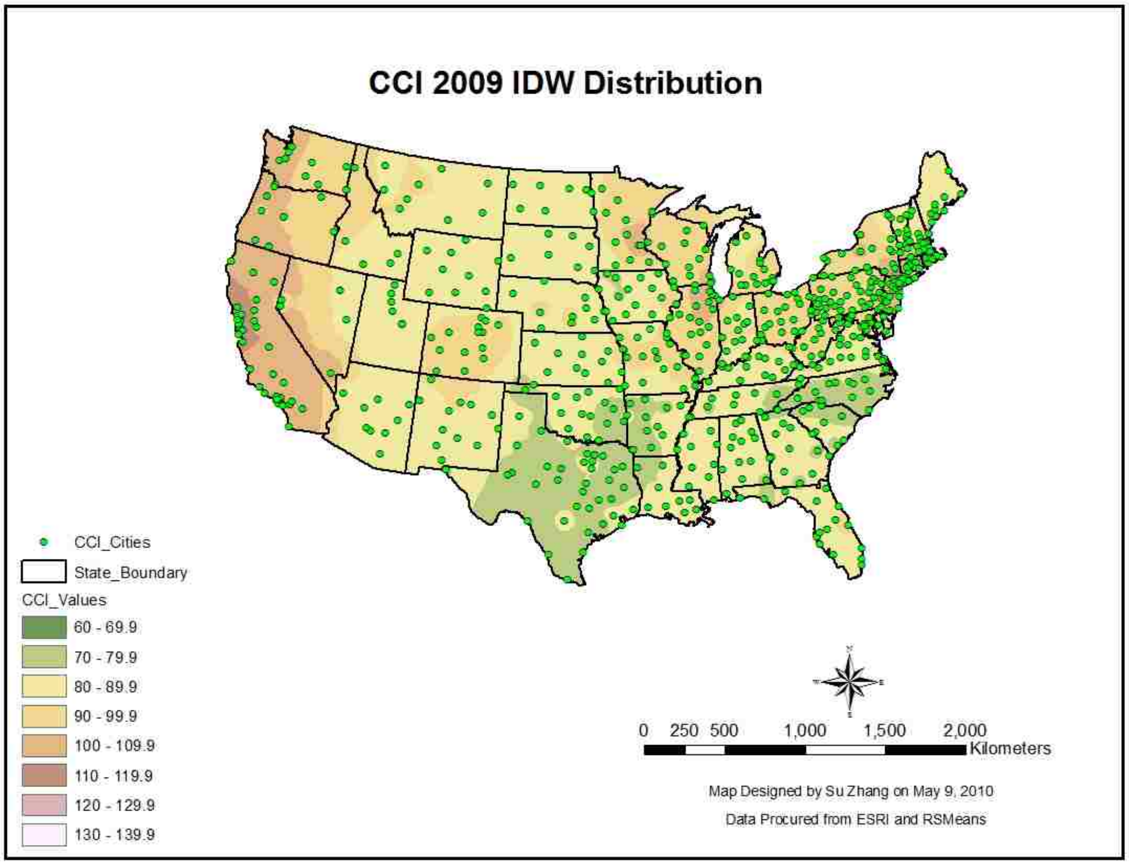

Figure G1.13 2D IDW Distribution for 2009 CCI

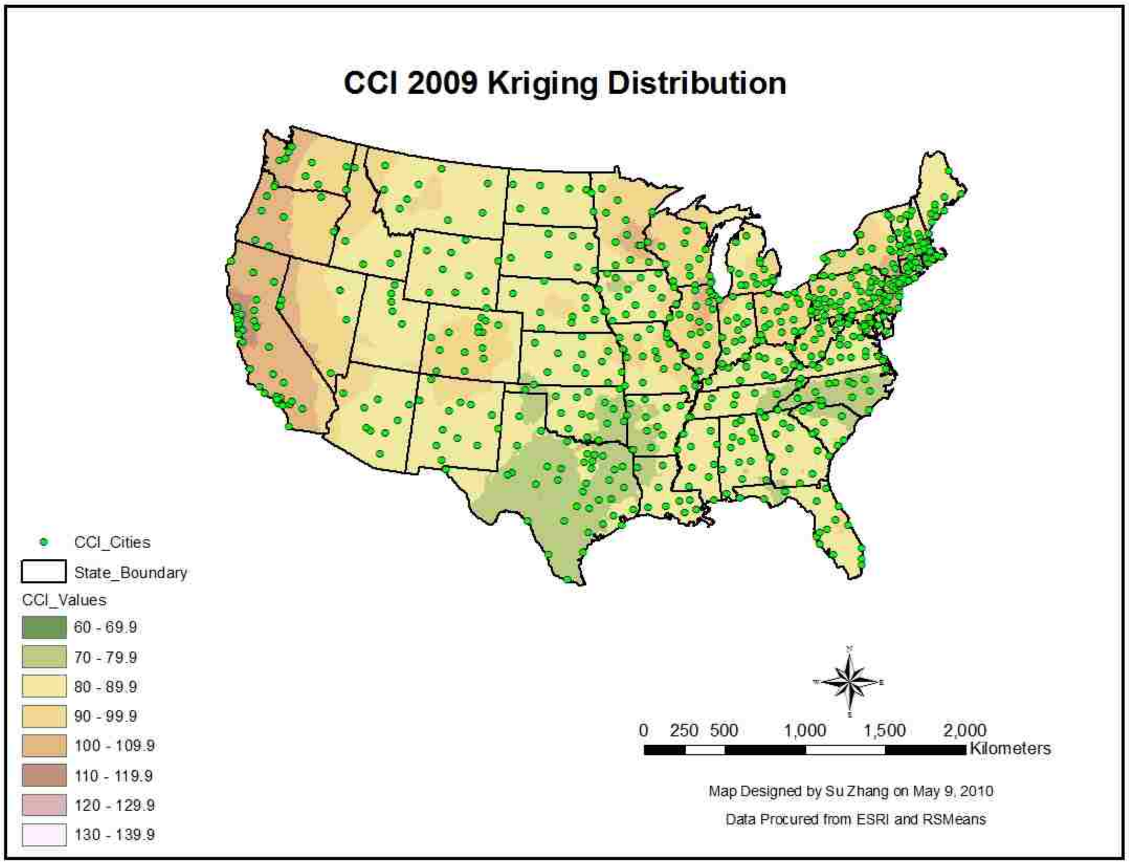

Figure G1.14 2D Kriging Distribution for 2009 CCI

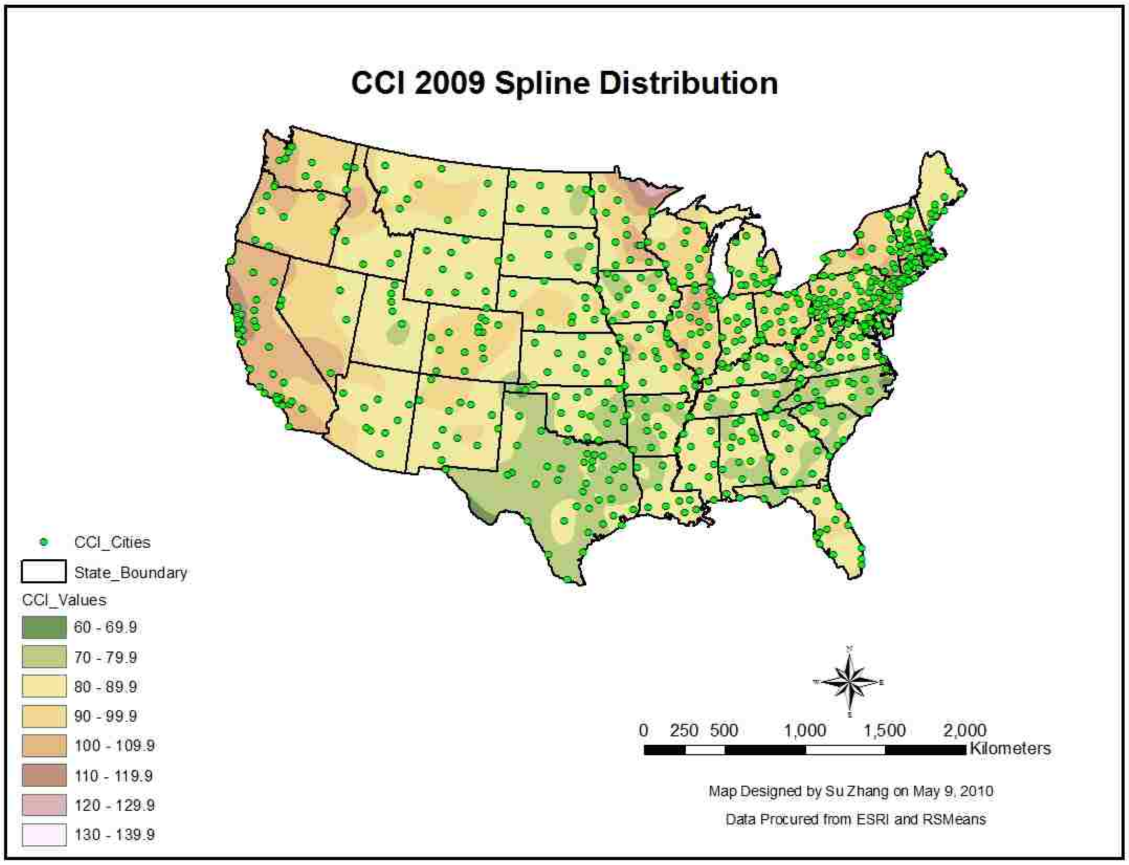

Figure G1.15 2D Spline Distribution for 2009 CCI

Exhibit G2. 2D Distribution Visualization for DoD ACF Dataset

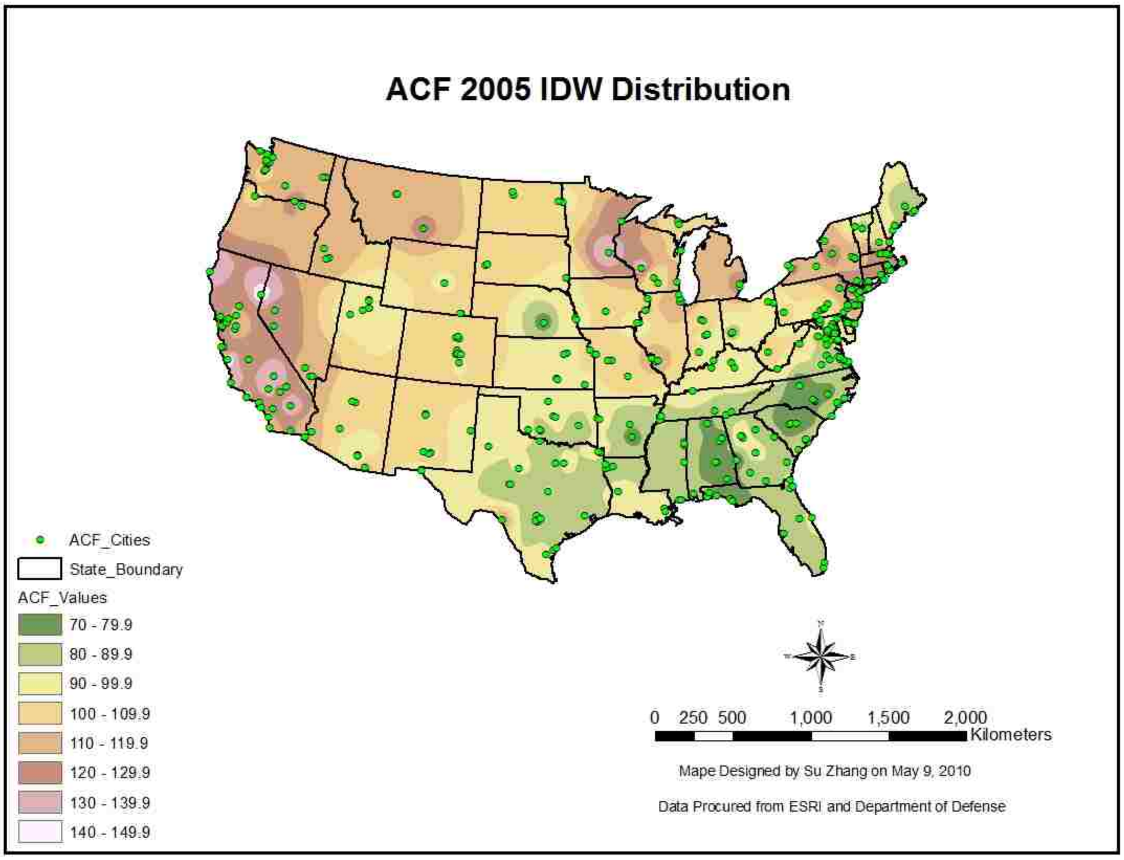

Figure G2.1 2D IDW Distribution for 2005 ACF

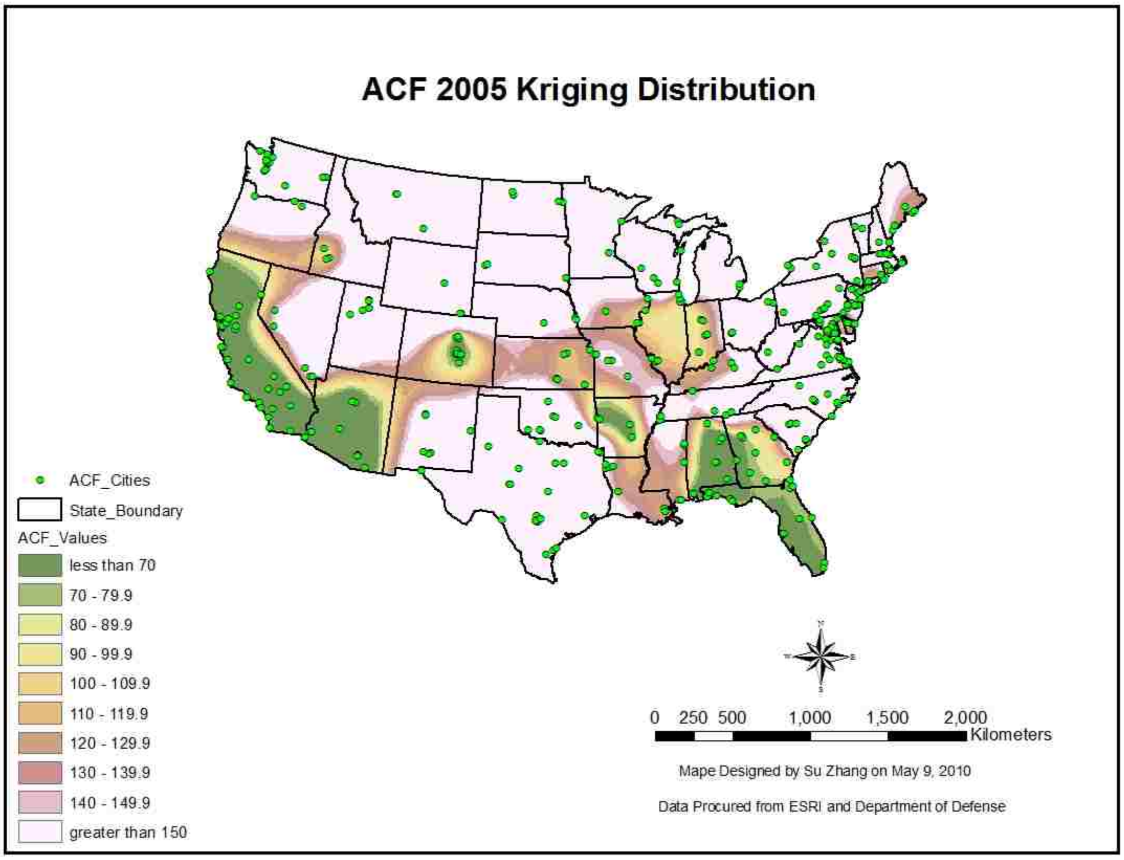

Figure G2.2 2D Kriging Distribution for 2005 ACF

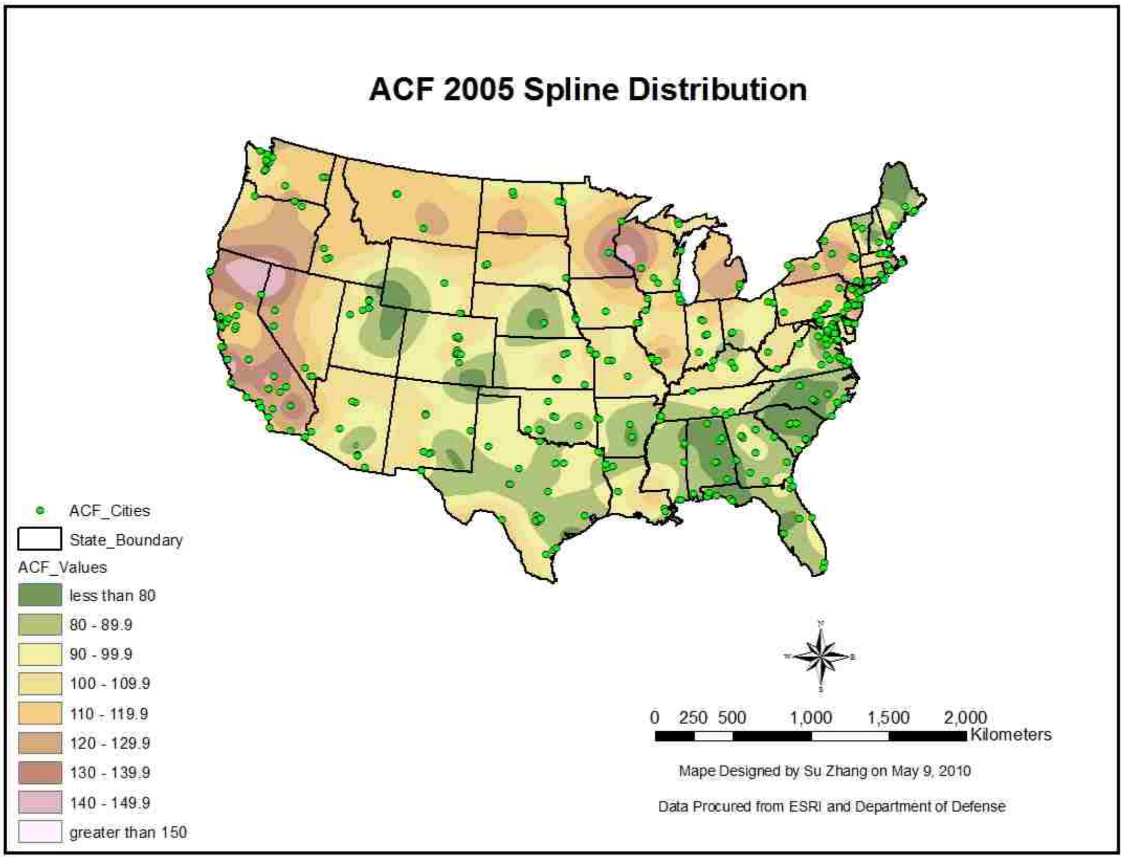

Figure G2.3 2D Spline Distribution for 2005 ACF

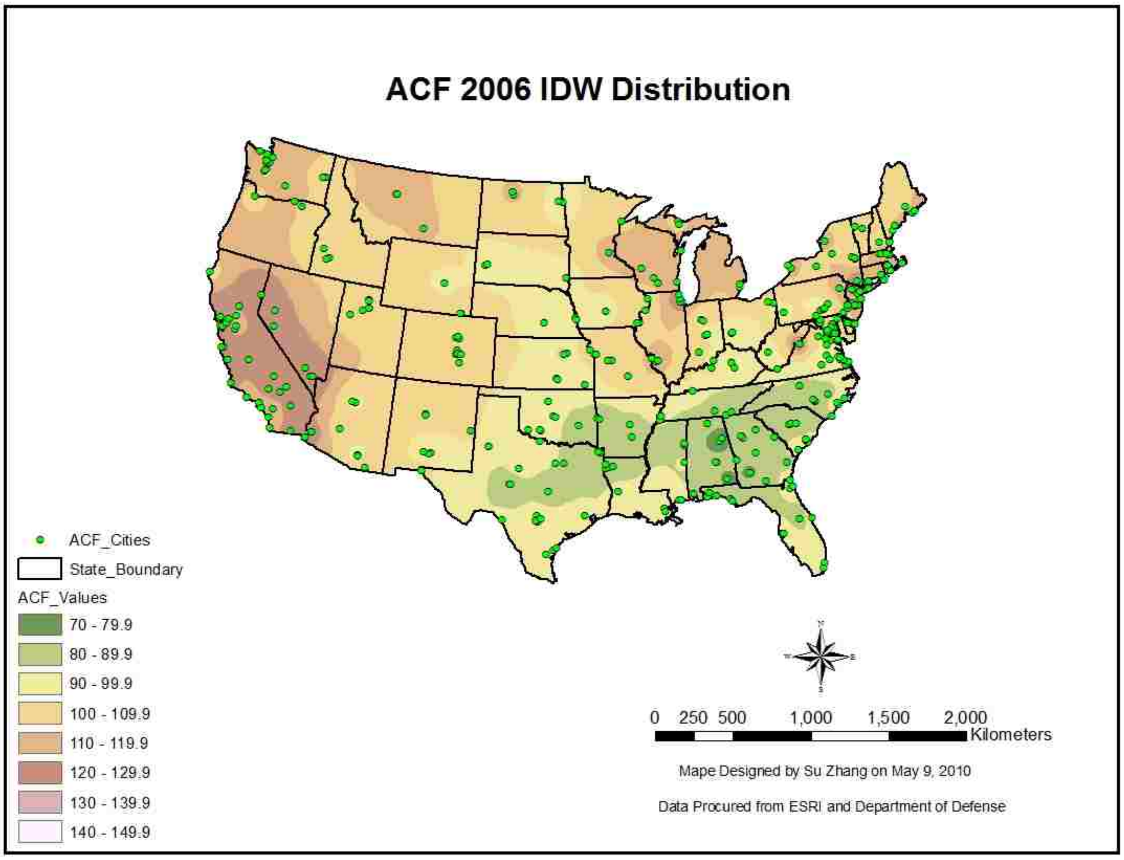

Figure G2.4 2D IDW Distribution for 2006 ACF

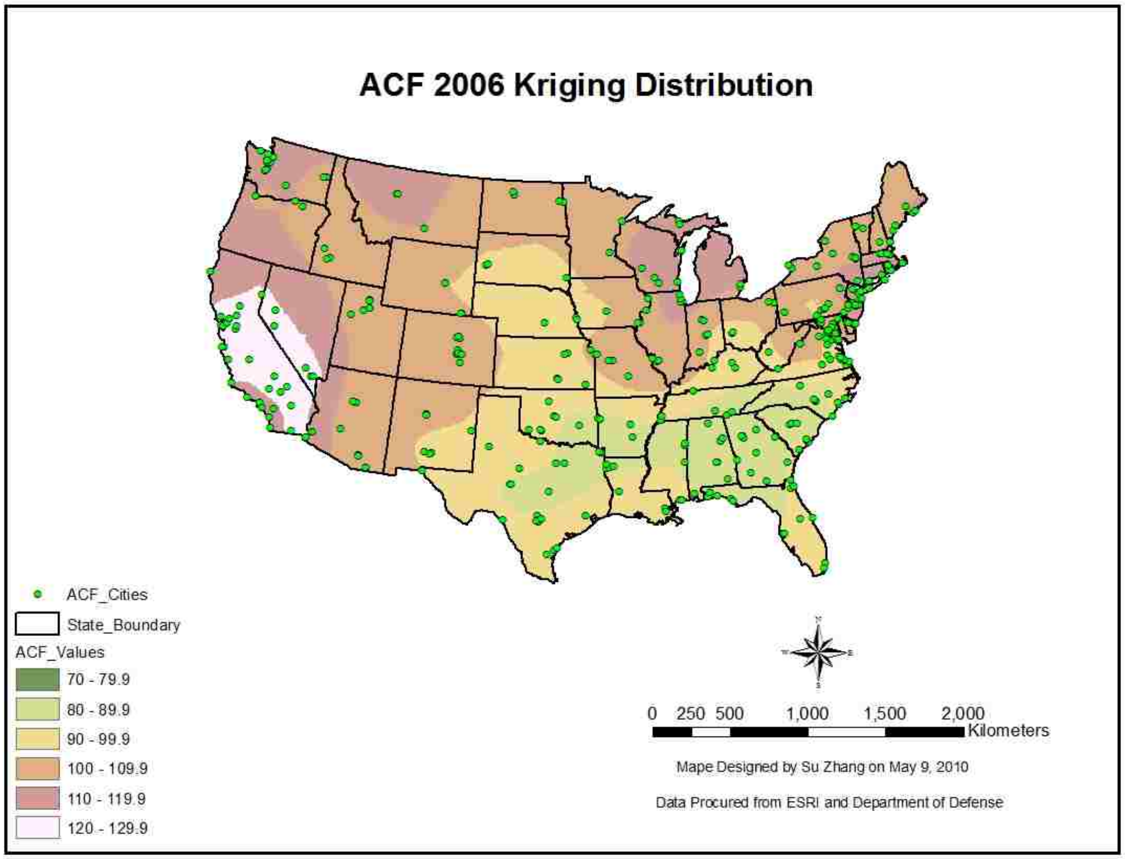

Figure G2.5 2D Kriging Distribution for 2006 ACF

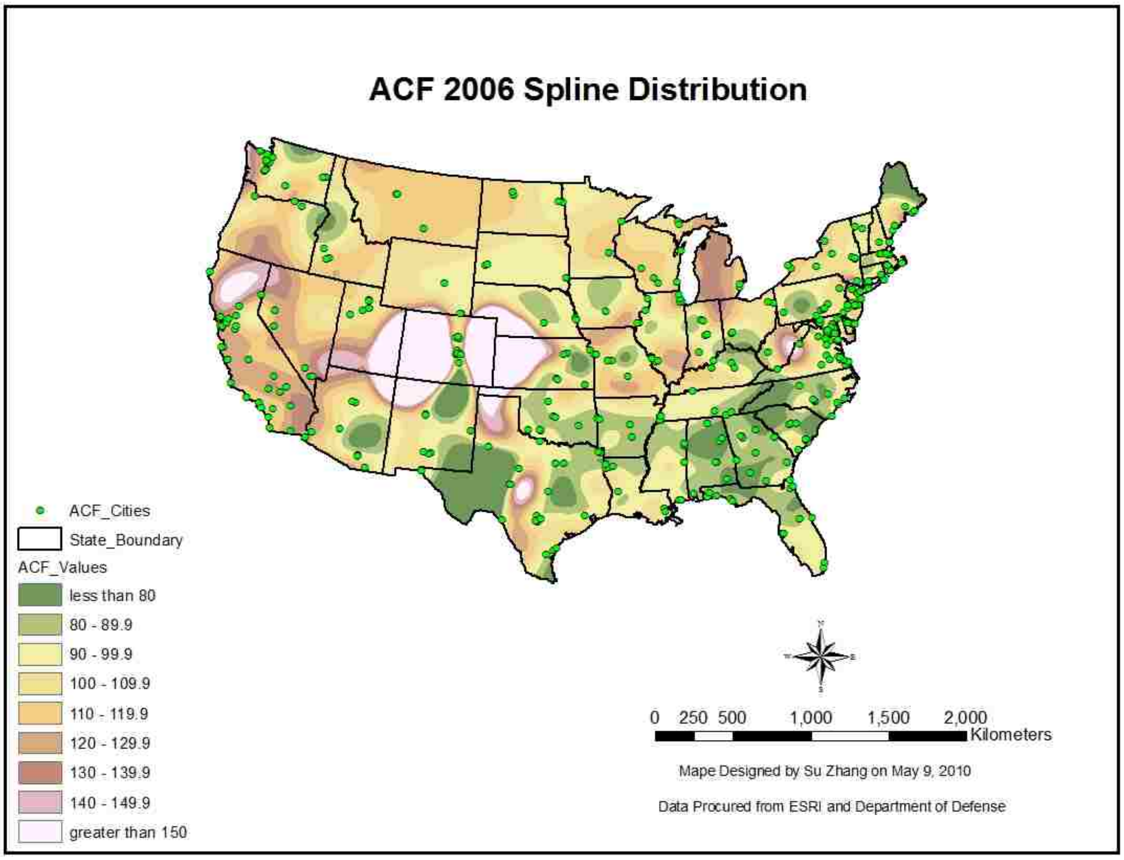

Figure G2.6 2D Spline Distribution for 2006 ACF

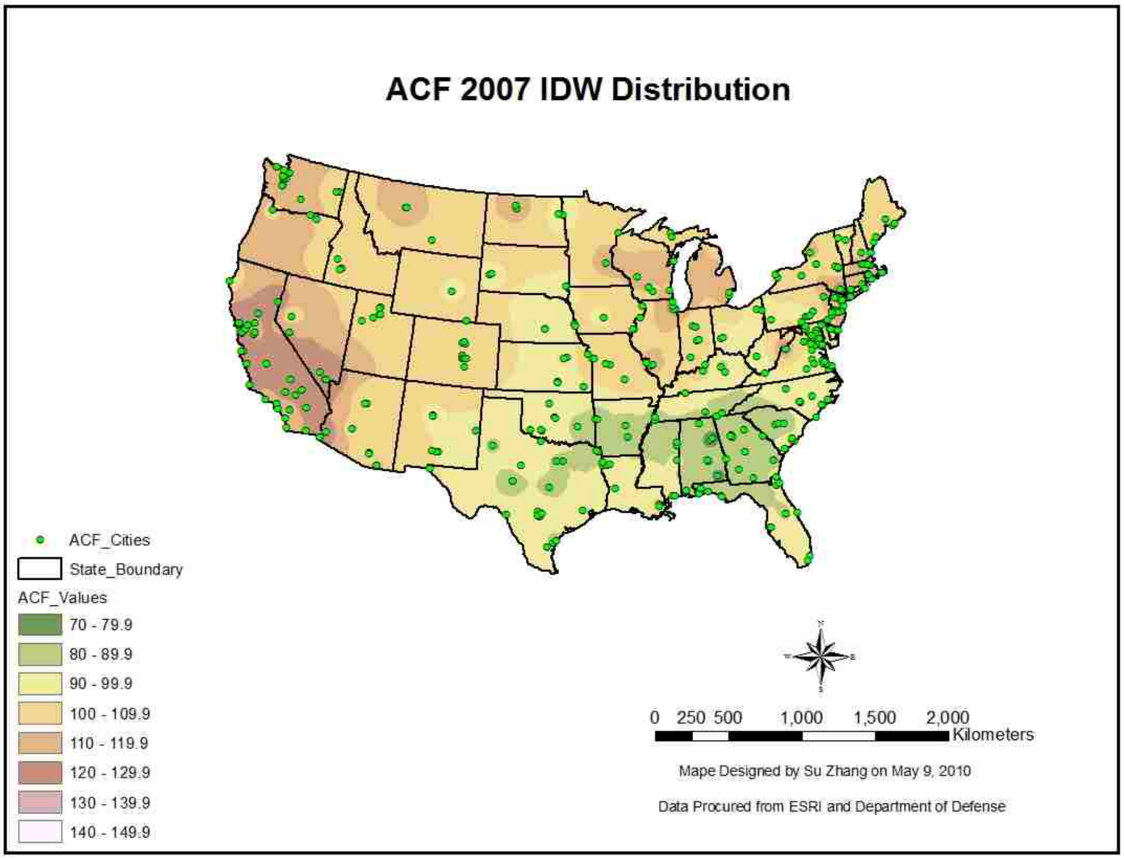

Figure G2.7 2D IDW Distribution for 2007 ACF

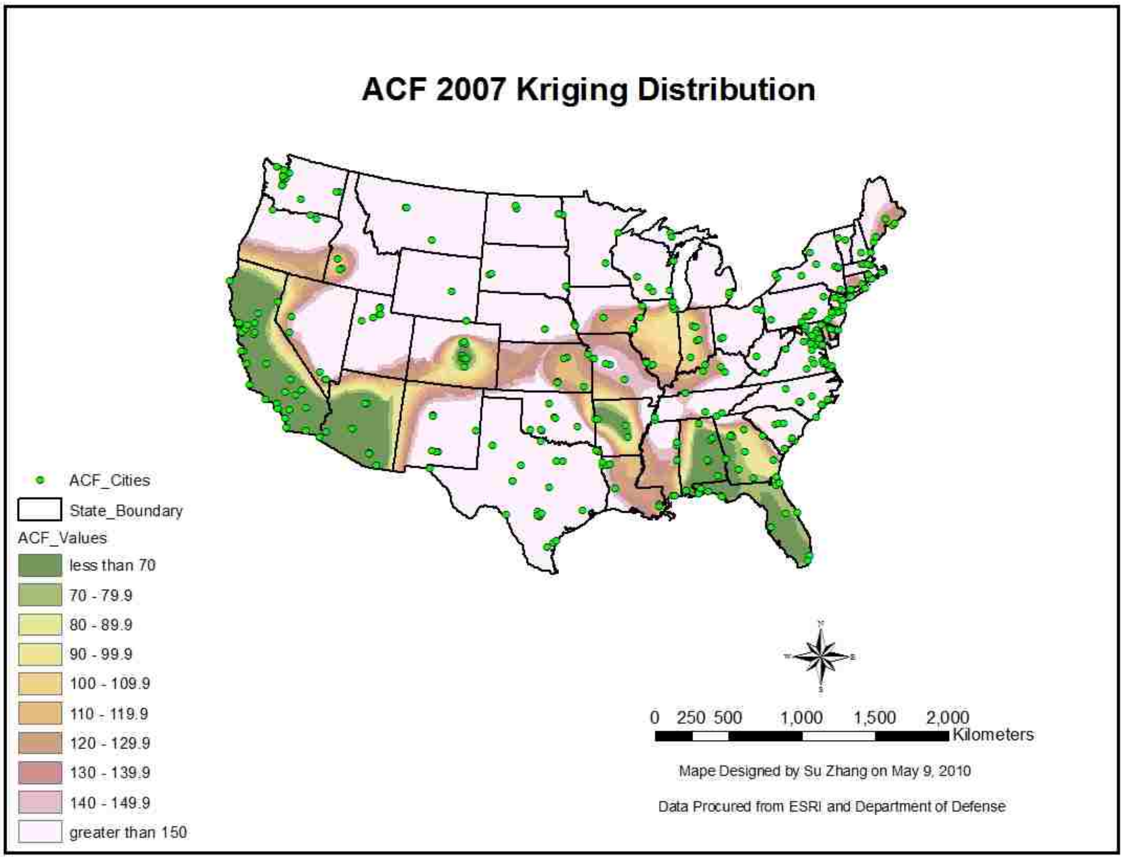

Figure G2.8 2D Kriging Distribution for 2007 ACF

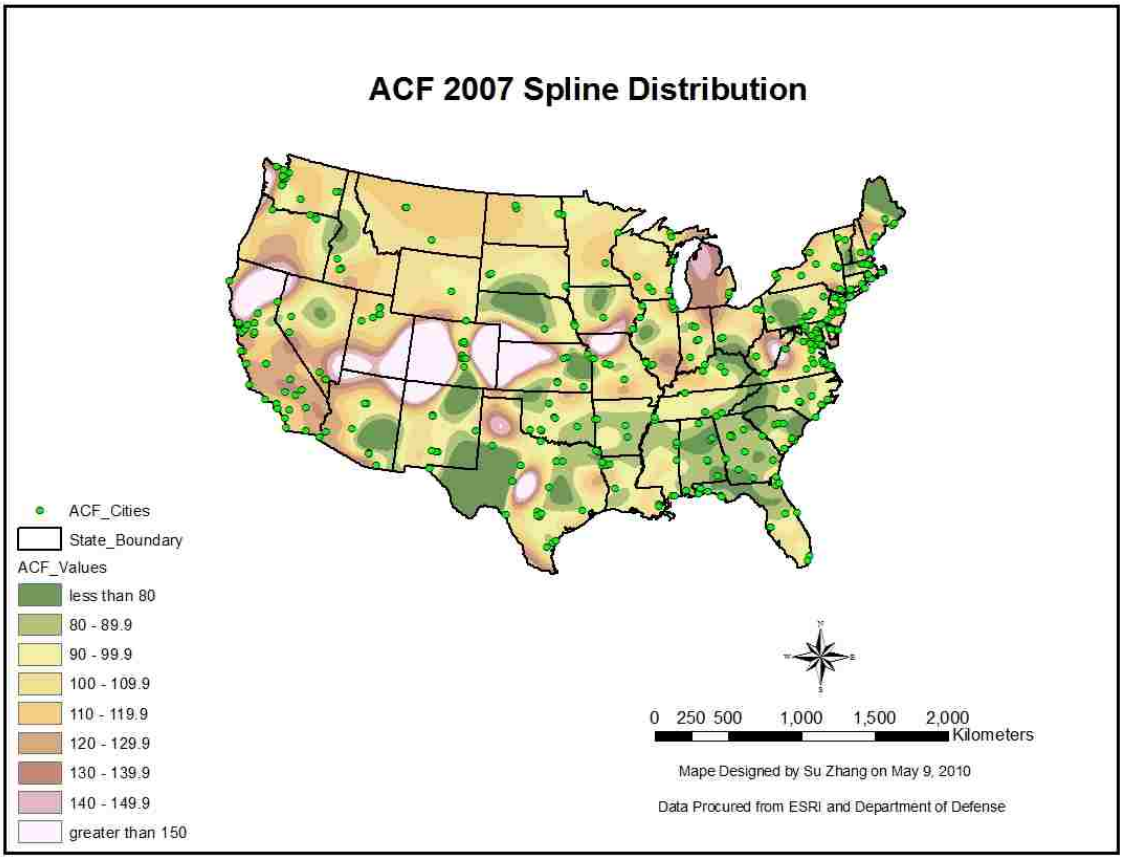

Figure G2.9 2D Spline Distribution for 2007 ACF

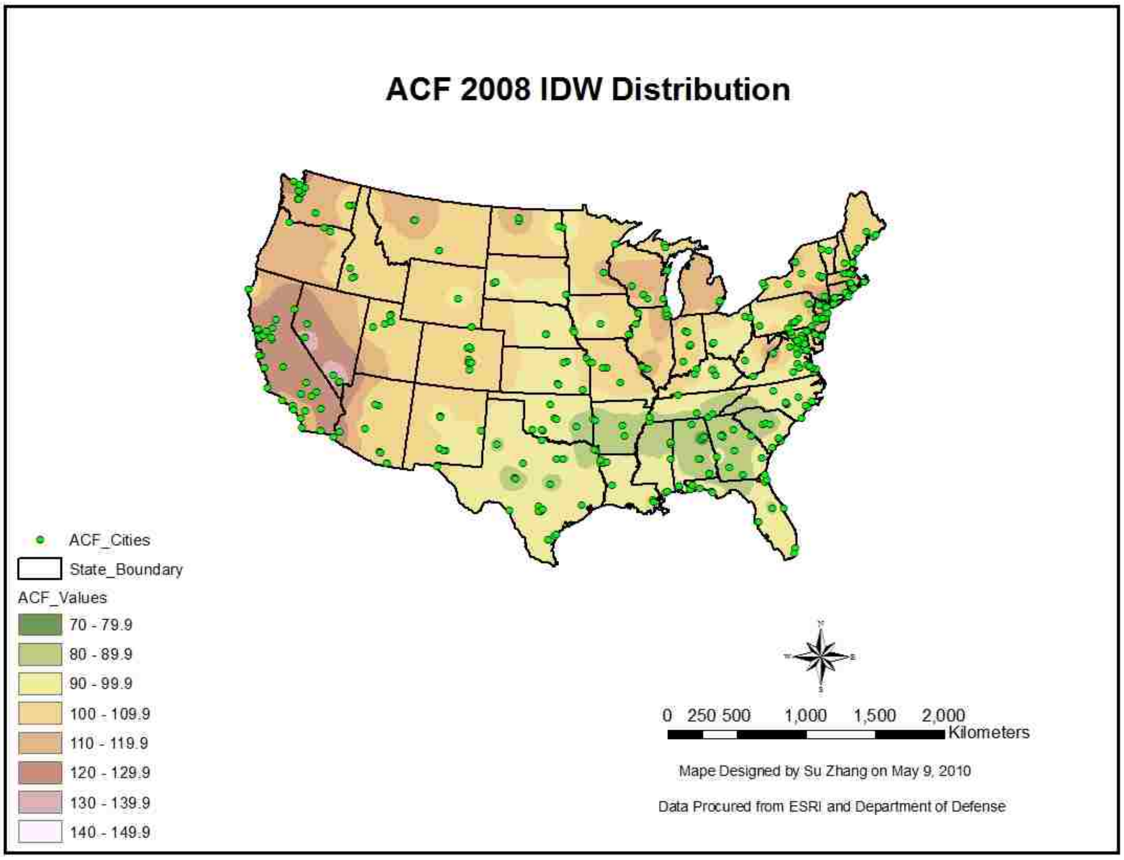

Figure G2.10 2D IDW Distribution for 2008 ACF

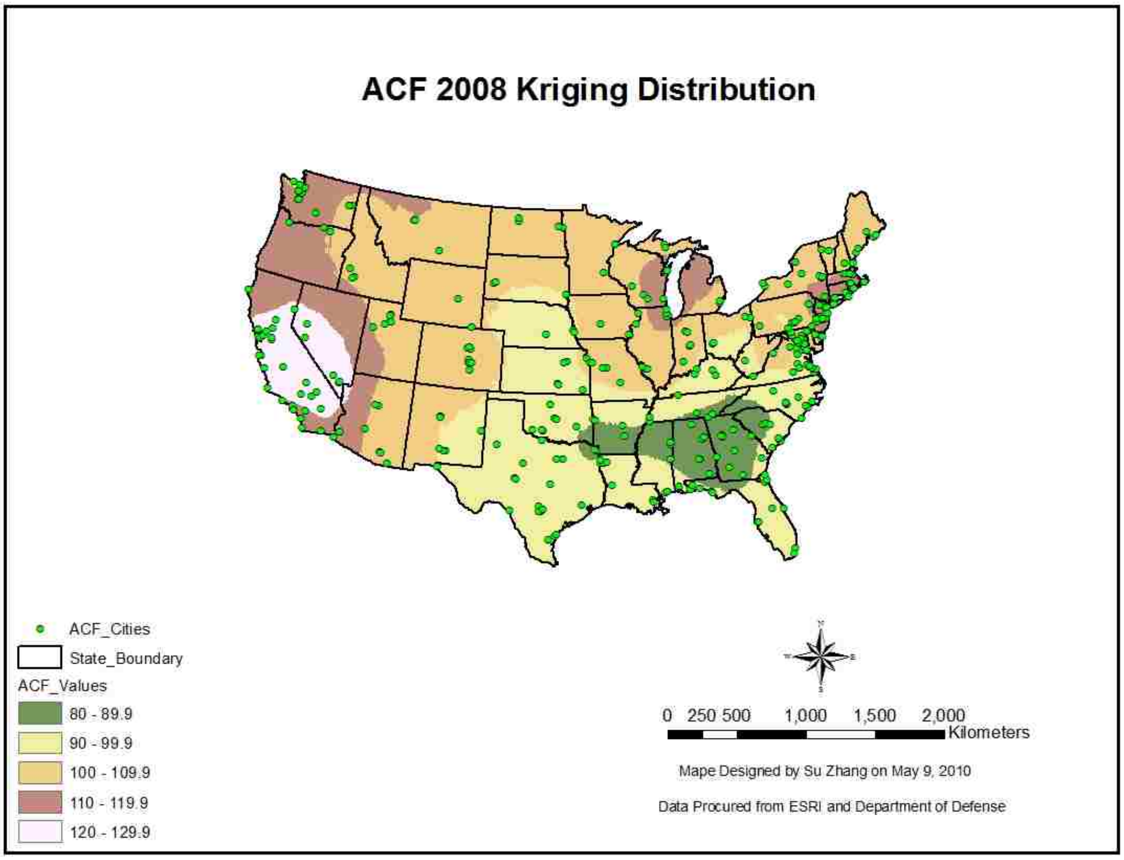

Figure G2.11 2D Kriging Distribution for 2008 ACF

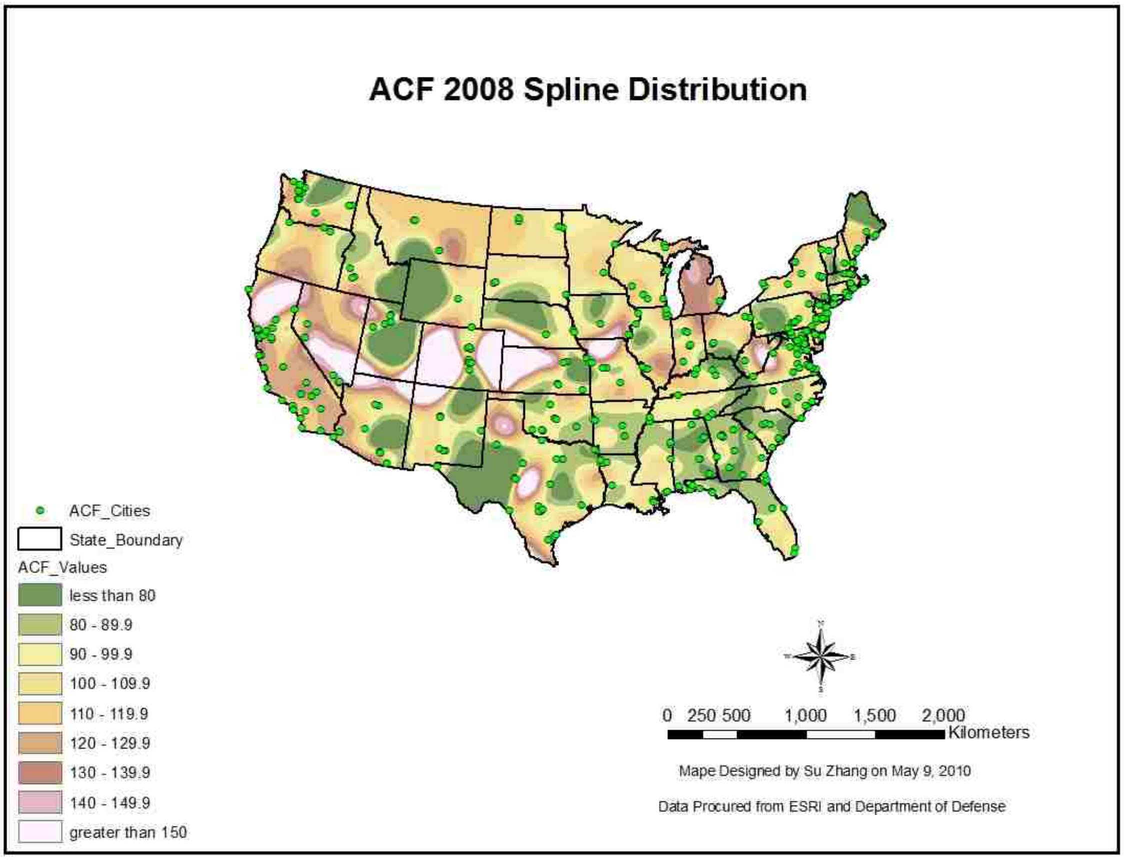

Figure G2.12 2D Spline Distribution for 2008 ACF

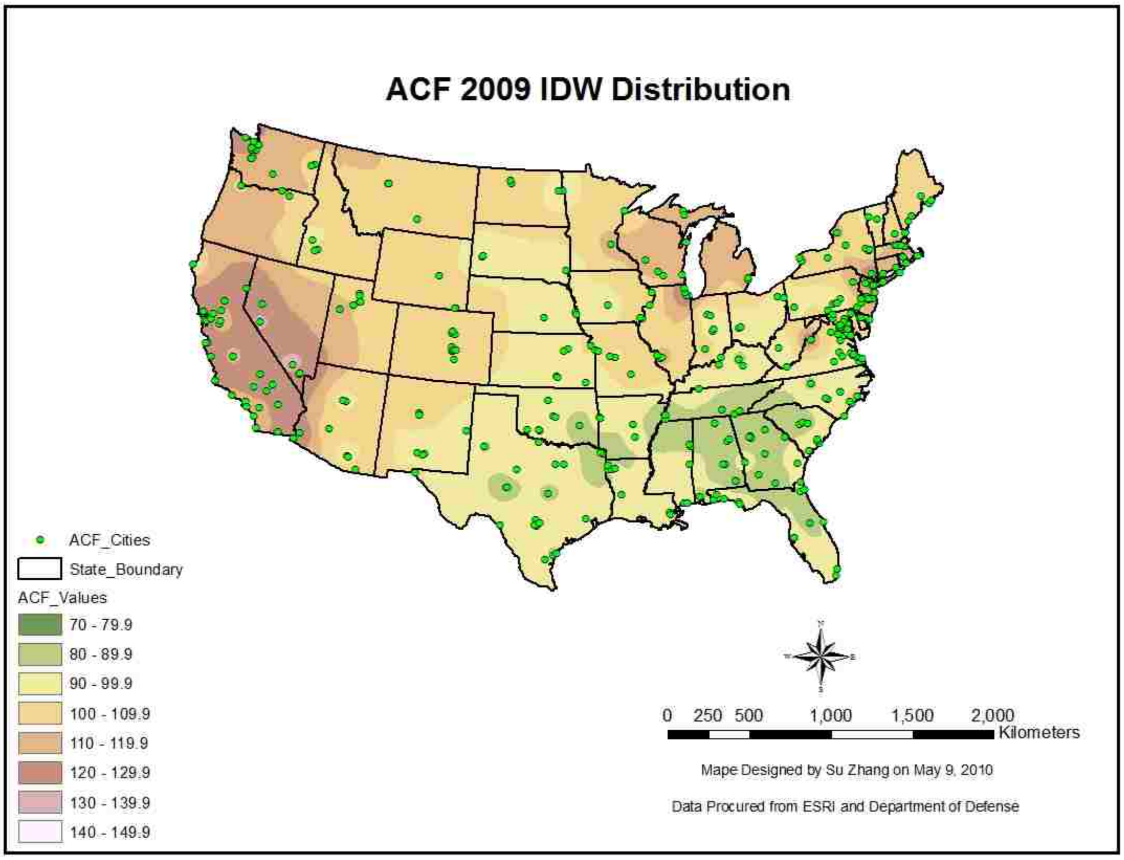

Figure G2.13 2D IDW Distribution for 2009 ACF

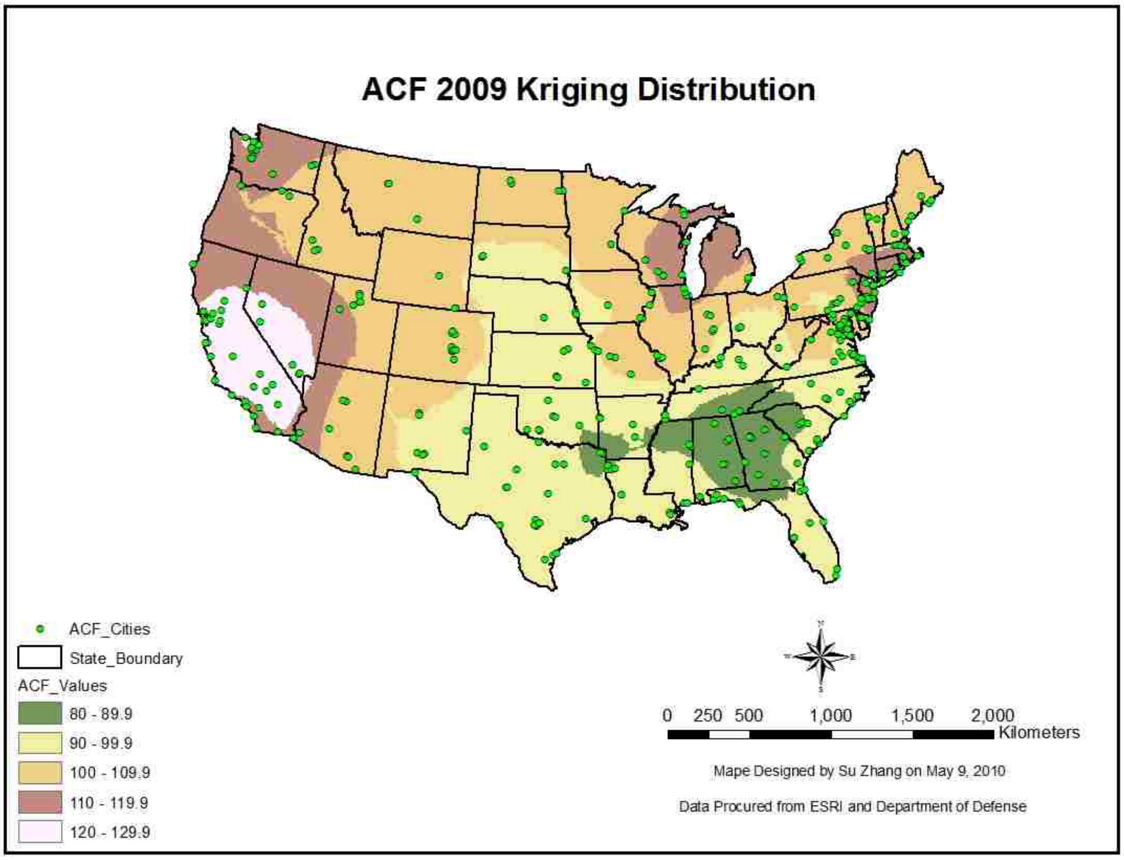

Figure G2.14 2D Kriging Distribution for 2009 ACF

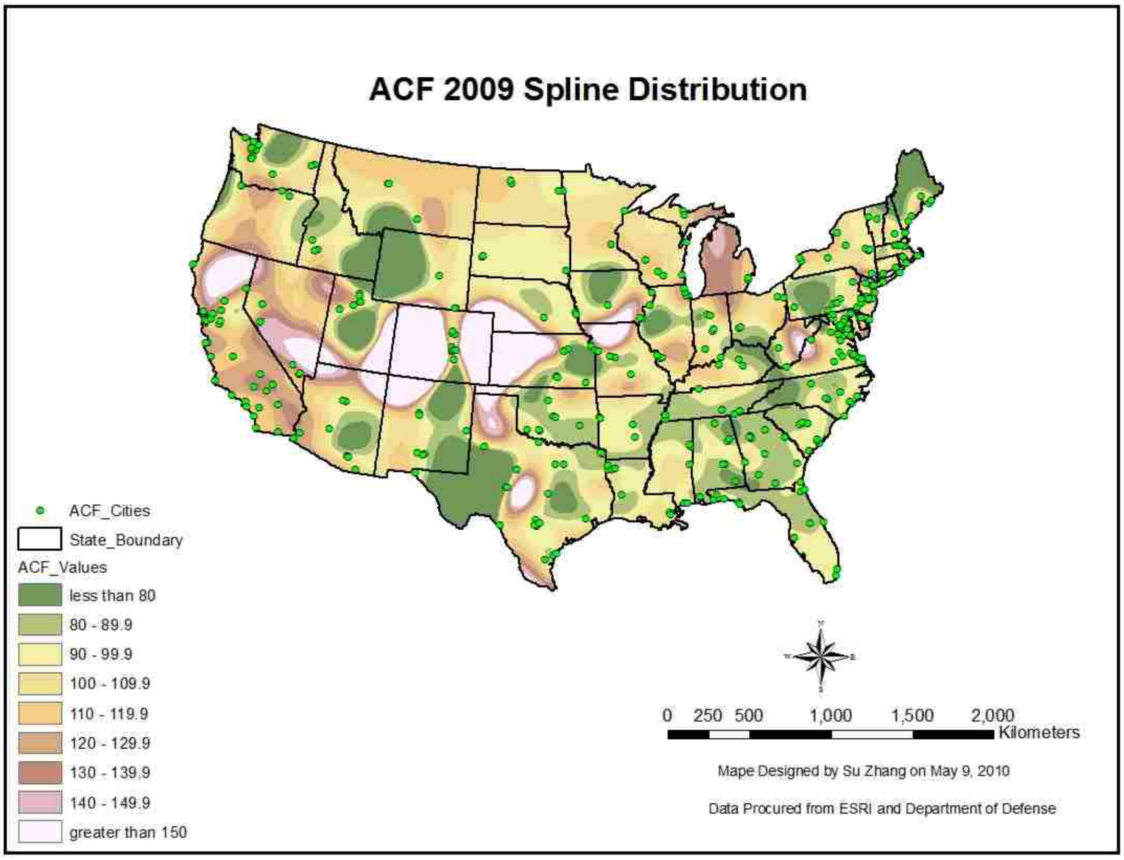

Figure G2.15 2D Spline Distribution for 2009 ACF