5-1-2016

Assessing the Utility of Nighttime Light Satellite Imagery for Adjusting Cost Estimate by Project Location

Su Zhang

Su Zhang
Candidate

Geography and Environmental Studies
Department

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

Approved by the Thesis Committee:

Christopher D. Lippitt, Chairperson

Danqing Xiao

Susan Bogus Halter

ASSESSING THE UTILITY OF NIGHTTIME LIGHT SATELLITE IMAGERY FOR ADJUSTING COST ESTIMATE BY PROJECT LOCATION

by

SU ZHANG

BACHELOR OF CONSTRUCTION MANAGEMENT
AGRICULTURAL UNIVERSITY OF HEBEI
JUNE, 2006
MASTER OF CONSTRUCTION MANAGEMENT
THE UNIVERSITY OF NEW MEXICO
AUGUST, 2010

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Geography

The University of New Mexico
Albuquerque, New Mexico

May, 2016

Dedication

This thesis is dedicated to my family, especially my wife and my grandfather. I will never be able to complete it without their support and encouragement. This thesis is also dedicated to my advisors and colleagues in the Department of Geography and Environmental Studies and Department of Civil Engineering.

Acknowledgements

I would like to express my great gratitude to Dr. Christopher D. Lippitt, my advisor and thesis committee chair, for providing me with the opportunity to pursue my Master's degree in GIScience and work with him as his research assistant. Dr. Lippitt is not only my academic advisor, but also my mentor and dear friend. Dr. Lippitt's enormous support has encouraged me to do well in school and pursue my dreams.

I would also like to thank Dr. Susan Bogus Halter and Dr. Danqing Xiao for serving on my thesis committee and guiding me in the research. I also appreciate their help for both my study at UNM and my life in the United States.

Thanks also go to my amazing colleagues in the Department of Geography and Environmental Studies at UNM for giving me advice and keeping me happy during my study. Without these colleagues, my thesis could not have been completed.

ASSESSING THE UTILITY OF NIGHTTIME LIGHT SATELLITE IMAGERY FOR ADJUSTING COST ESTIMATE BY PROJECT LOCATION

by

Su Zhang

Bachelor of Construction Management, Agricultural University of Hebei, 2006

Master of Construction Management, University of New Mexico, 2010

Master of Science in Geography, University of New Mexico, 2016

ABSTRACT

A fundamental process in construction cost estimation is the appropriate adjustment of costs to reflect project location. Unfortunately, location adjustment factors are not available for all locations. To overcome this lack of data, cost estimators in the United States often use adjustment factors from adjacent locations, referred to as the nearest neighbor (NN) method. However, these adjacent locations may not have similar economic conditions which limit the accuracy of the NN method. This research proposes a new method of using nighttime light satellite imagery (NLSI) to estimate location adjustment factors where they do not exist. The NLSI method for estimating location adjustment factors was evaluated against an established cost index database and results show that NLSI can be used to effectively estimate location adjustment factors. When compared with NN and other alternative location adjustment methods, the proposed NLSI method leads to a 25-40% reduction of the median absolute error. This work contributes to the body of knowledge by introducing a more accurate method for estimating location adjustment factors which can improve cost estimates for construction projects where location adjustment factors do not currently exist.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

The construction industry in the United States is one of the largest industry sectors delivering projects with substantial budgets. Therefore, cost estimating is a one of the most critical processes in a construction project development (Gould and Joyce 2009). Because project costs vary by location, an essential process in cost estimation is the appropriate cost adjustment to reflect project location influence. Unfortunately, location adjustment factors are not available for all locations (i.e. populated areas) across the United States.

Subsequently, in order to overcome this lack of data, cost estimators often use adjustment factors from adjacent locations, referred to as the nearest neighbor (NN) method. A key limitation of the NN method is the assumption that nearby locations share similar economic conditions, such as median household income, unemployment rate, inflation rate, and interest rate. Other proximity-based interpolation methods (Migliaccio et al. 2012; Migliaccio et al. 2013; Zhang et al. 2014) suffer from this same limitation. Existing location adjustment factors are surveyed, compound indices of material, labor, and equipment costs for a specific location, which consider local economic conditions. Hence, it is important to incorporate local economic conditions when estimating location adjustment factors for locations that are not surveyed.

Research has shown that nighttime light satellite imagery (NLSI) can be effectively used as a proxy measure of economic activities (Sutton et al. 2006). In an attempt to improve the process of developing location adjustment factors for locations that have not been surveyed, this research integrates luminance values extracted from NLSI, as a proxy for economic condition, into the cost estimation process. The NLSI-based location adjustment factor estimating method was evaluated using an established and widely adopted cost index database for location adjustment – RSMeans City Cost Index (CCI). Empirical comparison was performed to compare the NLSI-based method with alternative interpolation-based methods using established metrics, including comparison of patterns of overestimation and underestimation, comparison of absolute errors, and formal hypothesis testing of the error differences (Zhang et al. 2014).

Chapter 2: Background

2.1 Cost Estimates and Location Adjustment Factors

Multiple cost estimates are required for various purposes throughout the lifecycle of a construction project. These estimates can be classified into two groups: 1) the conceptual or preliminary cost estimate, which is the basis of successive cost estimates, is used for programming and budgeting and is based on minimal amounts of formal project design, and 2) the detailed or definitive cost estimate, which is used for bidding and is based on nearly complete formal project design. The accuracy of construction project cost estimates is fundamental to its success (Gould and Joyce 2009). From an owners' perspective, both overestimates and underestimates are detrimental. Overestimates push owners to allocate more funding than is actually needed for a specific construction project, which limits the number of projects that owners can pursue. In contrast, underestimates put an owner in the awkward position of having to seek additional funding, decrease the project scope, or even terminate the project which results in lost capital and, frequently, litigation.

Construction project owners often deal with the expected inaccuracy of cost estimates by including contingencies to alleviate the risk of a budget bust. However, using contingencies will reduce financial efficiency through inefficient funding allocations, which is a major issue for project owners, especially public owners or governmental agencies. While inaccurate estimates may be more tolerable during periods of stable economic growth, most governmental agencies struggle to meet capital requirements for new construction and/or renovation of buildings and infrastructures while being subjected to continuous budget cuts. This economic environment makes the accuracy of cost estimates critical to project success, making even slight improvements in the accuracy of location adjustment factors a substantial contribution to the construction industry at large (Layer et al. 2002) .

Three factors greatly affect the accuracy of cost estimates (Gould 1997): 1) Construction cost databases; 2) definition of project scope; and 3) cost adjustment methods. Owners often lack enough completed construction projects or manpower to develop in-house cost databases, so oftentimes they use a published construction cost index from commercial suppliers such as RSMeans. In contrast, large nationwide agencies or international companies have extensive facilities, and thus, enough past construction projects to develop a complete in-house cost database. It is a widespread belief that the cost estimates developed from in-house databases are more accurate than those developed from independent supplier's databases. For most owners, it is impossible to develop precise project scope at the pre-design phase since they only have general ideas about the intended projects. The final critical factor is the method used for adjusting cost estimates. Cost estimates are developed based on historical cost data and adjustment factors, which include location, time, size, and complexity, and thus the accuracy of those adjustments factors directly affect the accuracy of cost estimates.

The most common practice for estimating project costs for various locations is to adjust standardized (e.g., national) costs by applying location factors. The concept of "location adjustment factors" as an input decision for location adjustment was introduced by Johannes et al. (1985) as the construction cost in an area relative to the cost in another area. However, the reality is that only a few location adjustment factor datasets are available. In the United States, the most widely used location adjustment factor dataset is the RSMeans City Cost Index (CCI).

None of the commercial or governmental location adjustment factor datasets include values for all locations in the United States. Even larger commercial suppliers such as RSMeans cannot perform surveys for all cities and towns. In the conterminous United States (excluding the States of Hawaii and Alaska), RSMeans routinely surveys 649 cities out of the over 19,000 incorporated municipalities listed by the U.S. Census Bureau; meaning that less than four percent of municipalities are surveyed. Cost information regarding material, labor, and equipment is collected quarterly in the United States and Canada through a telephone survey and then composed into an index for each of the surveyed cities and then this information is published by RSMeans annually (Waier 2006).

Currently, construction practitioners use the NN method to estimate location adjustment factors for locations where adjustment factors do not exist. NN is a simple, geographical proximity-based interpolation method that adopts the geographically nearest city's location adjustment factor for an unknown city. When compared with alternative interpolation-based methods, however, the NN method has been shown to be less accurate (Zhang et al. 2014). Interpolation-based methods such as conditional nearest neighbor (CNN) and inverse distance weighted (IDW) outperformed the NN method in side-by-side comparisons. Of these two alternative methods, CNN is listed as the most simple to apply (Zhang et al. 2014). CNN is similar to NN, but uses state boundaries in conjunction with geographic proximity to select the nearest neighbor, and thus, selects the nearest neighbor within the same state as the unknown location. IDW is an interpolation method that is commonly used in spatial interpolation. IDW assumes that the value at an unknown location is the weighted average of known location factors within the neighborhood. The weights are inversely related to the distances between the known and unknown sample locations (Lu and Wong 2008).

All of the aforementioned methods rely on the assumption that geographic proximity provides a reasonable proxy for similarity in economic conditions and are ignorant of actual economic factors such as median household income, interest rates, and labor rates. Because the published location adjustment factors are influenced by local economic conditions, it would be reasonable to assume that a method for developing location adjustment factors that includes information on economic conditions would provide a better estimate than a method purely based on proximity. In order to incorporate the influence of local economic conditions, this study proposes and examines a NLSI-based method to improve the process of developing location adjustment factors for locations that have not been surveyed.

2.2 Nighttime Light Satellite Imagery (NLSI)

NLSI is a class of satellite nighttime observations and derived products through the detection of anthropogenic lighting presented on the Earth's surface (Elvidge et al. 2013). It is produced by spaceborne sensors that can collect low light imaging data in spectral bands covering emissions generated by electric lights (Elvidge et al. 2013). The majority of the nighttime light images are derived from nighttime satellite imagery provided by the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) and have become a recognizable spatially explicit global icon of human presence on the planet (Ghosh et al. 2013). Although in 2011 the National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) launched the Suomi National Polar Partnership (SNPP) satellite carrying the first Visible Infrared Imaging Radiometer Suite (VIIRS) instrument to collect low light imaging data, the DMSP-OLS is still widely used because, for more than 40 years, the DMSP-OLS has been the only system collecting global low light imaging data.

NLSI has been used in studies in various fields, such as to estimate global population (e.g., Sutton et al. 2001), to estimate the pattern and impact of urbanization (e.g., Zhang and Seto 2011; Ma et al. 2012), and as a surrogate to estimate economic activities such as gross domestic product (GDP) and income per capita (Ebener et al. 2005; Levin and Duke 2012). NLSI has also been used to investigate and explain biological patterns. For example, Bharti et al. (2011) used NLSI to explain the seasonal fluctuations of measles in Niger.

Evidenced by various studies, NLSI can be effectively used as a proxy measure of economic activities and development, such as GDP or distribution of income in society (Doll et al. 2006; Chen and Nordhaus 2011; Ghosh et al. 2013). As mentioned previously, CCI is a compound index of material, labor, and equipment costs for a location which reflects the relative relationship of construction costs at that location to the national level average. Because the labor, material, and equipment costs are highly correlated to economic activities (Arora and Blackley 1996; Adriaanse et al. 1997; De Long and Summers 1990) and NLSI is a proxy measure of economic activities, we postulate that the NLSI is also a proxy measure of CCI. Specifically, the research proposed here is focused on analyzing the luminance values of NLSI to examine if CCI could be estimated for locations where they have not been surveyed, and if so, how well the estimation is when compared with currently adopted estimation methods.

Chapter 3: Methodology

The research methodology for this study included identification of an appropriate study area and dataset, image processing of the NLSI datasets, regression analysis, and performance evaluation.

3.1 Study Area and Dataset

RSMeans CCI was selected for this study because it is the most widely used location adjustment factor dataset, especially for commercial building construction projects. The 2009, 2011, and 2013 CCI datasets were obtained from RSMeans while the 2009, 2011, and2013 city boundary datasets were obtained from the U.S. Census Bureau or a city's GIS data clearinghouse website to match the CCI datasets. In order to make the proposed method comparable to the NN, CNN, and IDW estimation approaches studied by Zhang et al. (2014), the study area for this research is also the conterminous United States.

Among the 649 cities in the conterminous United States that have CCI values, we excluded 236 cities because there is no distinct boundary (cities interconnect with each other) on the NLSI. Therefore, only 413 cities were selected for analysis, and assigned an exclusive identification number (EID). These cities are shown in Figure 1.

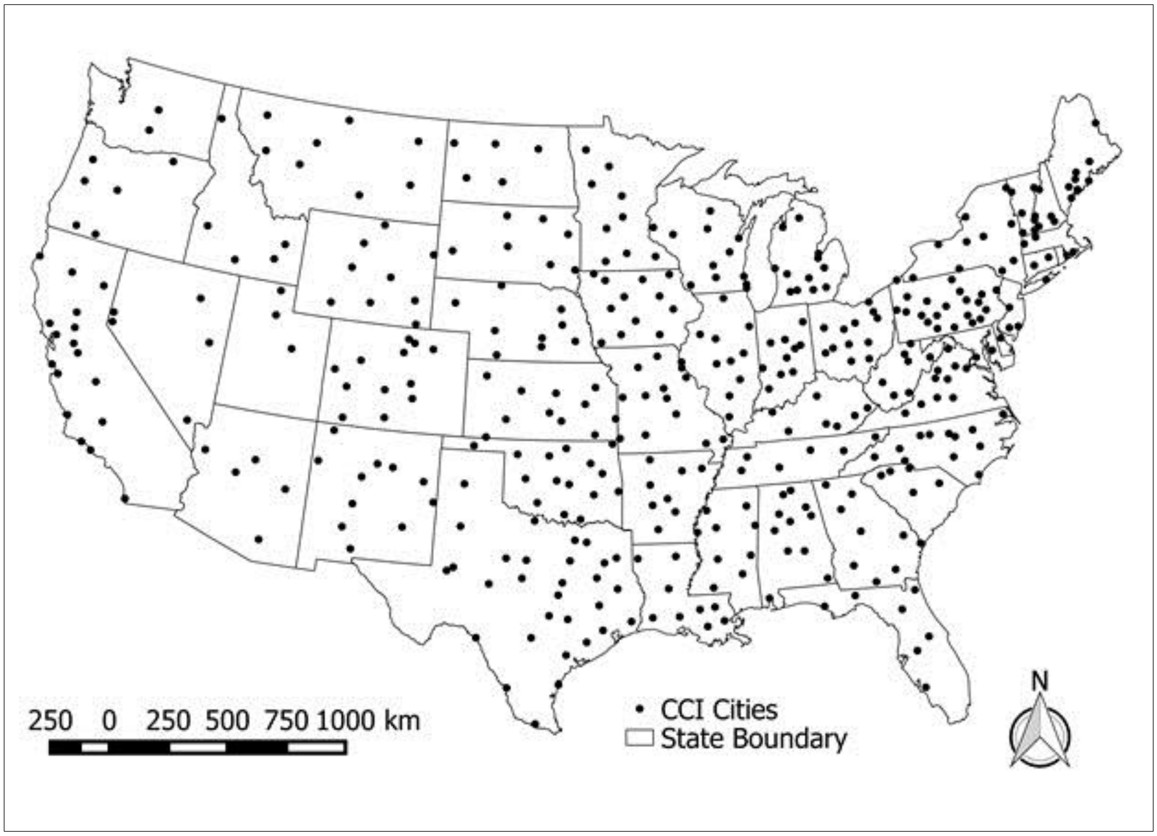

Figure 1. Conterminous Untied States and CCI Cities 2009

The 2009, 2011, and 2013 global NLSI datasets were acquired from the NOAA National Geophysical Data Center. The 2013 global NLSI dataset is the latest publically-available dataset that has been released by NOAA, and therefore, 2014 or 2015 datasets were not used in this study. Global NLSI is collected by the DMSP-OLS sensor and it provides average visible, stable lights, and cloud free coverages. The digital number (DN) values for this imagery range from 0 to 63. The lower the DN value is, the lower luminance value is. The spatial resolution for NLSI is approximately 1 km (0.62 miles).

3.2 Image Processing

The 2009, 2011, and 2013 global NLSI datasets were clipped by the boundary of the conterminous United States and then projected to the USA Contiguous Lambert Conformal Conic coordinate system (Figure 2). The boundaries of the 413 CCI cities were projected to the same coordinate system as the clipped nighttime light satellite imagery and then overlaid with it.

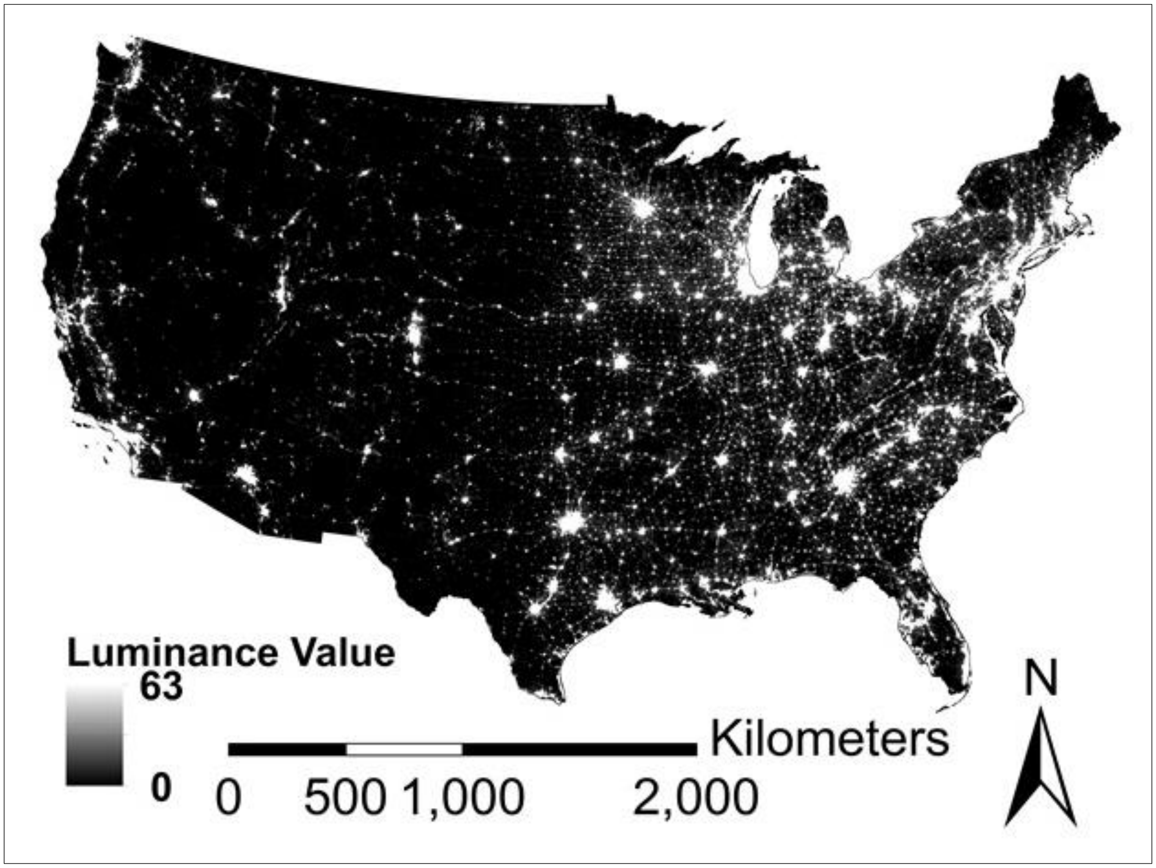

Figure 2. Conterminous United States and Nighttime Light Satellite Imagery 2009

Only the luminance values within the city boundary are useful for this research. This is because the areas outside of these boundaries are not included in the RSMeans survey. Therefore, summary statistics (mean, median, standard deviation, variety [number of unique values for all pixel cells within the boundary of a city], majority, minority, maximum, minimum, range, and sum) of the nighttime light satellite luminance values were extracted for each city and limited to its boundaries as shown in Figure 3, which uses the City of Santa Fe as an example. As shown in Figure 3, varied luminance values exhibit across the boundary, and similar pattern can be found within other cities' limits.

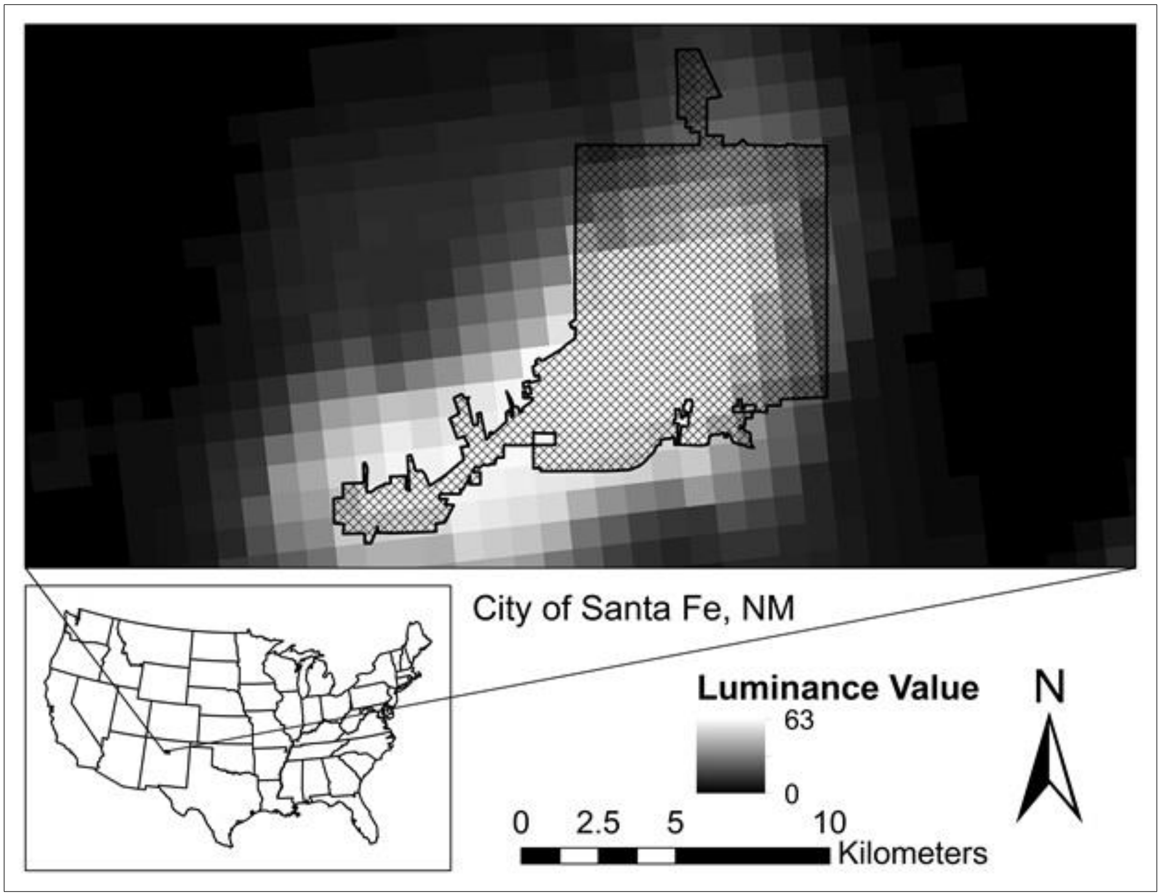

Figure 3. Conterminous United States and Example City Boundary

3.3 Regression Analysis

The dependent variable (i.e., response variable) used in this study is the RS Means CCI values. Choosing the most appropriate independent variables from the statistics available from the nighttime light luminance values is necessary as a precursor to developing the regression model. According to Ghosh et al. (2013), mean brightness of NLSI pixels, which is a measure of the overall brightness, has a significant positive relationship with economic activities such as GDP. Therefore, we postulate that mean brightness of nightlight light pixels could be used as a significant predictor for CCI. In this research context, image texture metrics may also provide significant predictors of CCI. Theoretically, a city with a lower CCI value will have lower economic activity that may lead to heterogeneity in illumination condition throughout its administration boundary. Pearson's correlation analyses were performed and are shown in Table 1.

Table 1. Pearson's Correlation Results between CCI and Luminance Summary Statistics

Year	Pearson's Correlation	CCI	Mean	STD	Range	Variety
2009	CCI	1.0000				
	Mean	0.8970 (0.0001*)	1.0000			
	STD	0.0128 (0.7949)	-0.0315 (0.5227)	1.0000		
	Range	0.0952 (0.0533)	0.0477 (0.3332)	0.8211 (0.0001*)	1.0000	
	Variety	0.1283 (0.0091)	0.0523 (0.2889)	0.6460 (0.0001*)	0.8177 (0.0001*)	1.0000
2011	CCI	1.0000				
	Mean	0.8812 (0.0001*)	1.0000			
	STD	0.0468 (0.3432)	0.0017 (0.9721)	1.0000		
	Range	0.0664 (0.1782)	0.0032 (0.9490)	0.7618 (0.0001*)	1.0000	
	Variety	0.1240 (0.0117)	0.0172 (0.7277)	0.4839 (0.0001*)	0.7685 (0.0001*)	1.0000
2013	CCI	1.0000				
	Mean	0.8704 (0.0001*)	1.0000			
	STD	0.0857 (0.0819)	0.0369 (0.4551)	1.0000		
	Range	0.0844 (0.0868)	0.0160 (0.7453)	0.7427 (0.0001*)	1.0000	
	Variety	0.1454 (0.0031)	0.0371 (0.4516)	0.4819 (0.0001*)	0.7432 (0.0001*)	1.0000

Note: * indicates Pearson correlation coefficient is significant at $p = 0.05$ level.

Results revealed that texture measures, including standard deviation, range, and variety, are significantly correlated with each other, but are not significantly correlated with CCI. Therefore, standard deviation, range, and variety were not included as independent variables in the subsequent analyses. Mean brightness of NLSI pixels instead was significantly correlated with CCI, and therefore, was the only independent variable included in subsequent analyses.

Plotting results revealed that the relationship between CCI and the mean brightness of NLSI pixels is not linear. Therefore, several regression models were investigated to identify the best-fit model. R-squared (R^2) values together with the root mean squared error (RMSE) were used to choose the best-fit regression model.

3.4 Evaluation of Performance of Each Method

To ensure the comparability against previous work (Zhang et al 2014), the performance of each method was assessed in the form of an "error" value, which is the difference between the estimated value and the exact value (Ito 1987). In this study, error is defined as the difference between predicted and actual location factor value. The following equations were used to calculate relative errors and absolute errors for each method.

$$E_i = P_i - A_i \quad (1)$$

$$EA_i = |E_i| \quad (2)$$

Where i denotes the EID for CCI cities from 1 to 413. P_i denotes the predicted value for location i while A_i denotes the actual value for location i. E_i denotes the relative error for location i, while EA_i denotes the absolute error for location i.

3.5 Empirical Comparison

The empirical performance assessment of the NLSI-based method for estimating location adjustment factors was performed against NN, CNN, and IDW. These three methods were selected due to the following reasons. NN is the most widely used method, CNN is the best rough surface interpolation method, and IDW is the best smooth surface interpolation method (Zhang et al. 2014). Again, to ensure the comparability of previous work (Zhang et al 2014), comparison of the performance of the four interpolation methods was based on the following three established metrics:

Comparison of patterns of overestimation and underestimation: overestimation occurs when the difference between estimated CCI value and actual CCI value is positive ($E_i > 0$), whereas underestimation occurs when such difference is negative ($E_i < 0$). An estimate is considered accurate when the corresponding relative error is within a +/-1% range. Pearson chi-squared (X^2) test was used to examine if the observed underestimation, overestimation, and accurate estimation pattern for each method are not significantly different over time.

Pairwise comparison of absolute errors: a method is considered to be better than another one if it has a lower absolute error (EA_i). For each study site, *EA*s were calculated for each of these four methods (NN, CNN, IDW, and NLSI). A pairwise comparison was performed between NLSI and the other three methods (NLSI vs. NN, NLSI vs. NN, and NLSI vs. IDW). Pearson chi-squared (X^2) test was used to examine if the patterns of the observed method comparison results for each pairwise comparison are not significantly different over time.

Formal hypothesis testing of the error differences: for each method, the distribution of the absolute errors (*EA*s) for all study cities was summarized by their means, medians and standard deviations. A pairwise Wilcoxon Signed Rank Test was performed between NLSI-based method and the other three methods (NLSI vs. NN, NLSI vs. CNN, and NLSI vs. IDW) to examine if NLSI-based method yielded absolute errors that were significantly different from the other three methods, and if so, their means were compared to determine which method had the lowest mean, and therefore was considered to be outperforming the other.

3.6 Error Causation Analysis

Environmental factors, including trees canopy and waterbodies, could affect the brightness of NLSI. Tree canopy will obstruct the nighttime light reflection which eventually reduces the luminance values that can be detected by the DMSP-OLS sensor. Water bodies, on the other hand, will increase the luminance values that can be detected by the DMSP-OLS sensor because of ambient light reelection (waterbodies work like mirrors).

The 2011 Tree Canopy dataset was obtained from U.S. Forest Service. This dataset presents a percent tree canopy cover image layer for the conterminous United States, with each pixel's size being 30 x 30 m. Each pixel's value represents the percentage that a pixel is covered by tree canopy. With the help of this dataset, each city's tree coverage percentage was calculated. Pearson's correlation analysis was performed between absolute errors and tree coverage percentage to investigate if the errors of NLSI-based method are caused by tree coverage.

The 2011 National Land Cover Dataset (NLCD) was obtained from U.S. Geological Survey. This dataset presents a land cover type image layer for the conterminous Untied States, with each pixel's size also being 30 x 30 m. Each pixel's value represents the type land cover type that a pixel is covered. Waterbody is a specific land cover type and therefore, each city's waterbody percentage was calculated. Pearson's correlation analysis was performed between absolute errors and waterbody percentage to examine if the errors of NLSI-based method are caused by water reelection.

Chapter 4: Results and Discussion

4.1 Regression Model Selection

The candidate regression models included linear, exponential, logarithmic, polynomial, and power models (Table 2). Linear regression model was selected for comparison due to its fundament ability and generality, although the plotting results revealed that the relationship between CCI and the mean brightness of NLSI pixels is not linear.

Table 2. Regression Models and Corresponding R-squared Value and RMSE

Year	Regression Model	Equation	R^2	RMSE
2009	Linear Least	$Y = 1.11X + 32.06$	0.8046	3.73
	Exponential	$Y = 46.292e^{0.0126X}$	0.8393	3.67
	Power	$Y = 9.2524X^{0.5745}$	0.7971	4.07
	Logarithmic	$Y = 50.118\ln(X) – 108.08$	0.7554	4.32
	Quadratic	$Y = 0.0291X^2 – 1.6976X + 97.898$	<u>0.8536</u>	<u>3.34</u>
	Cubic Polynomial	$Y = 0.0007X^3 – 0.0631X^2 + 2.526X + 35.125$	0.8560	7.18
	Quartic Polynomial	$Y = -0.000008X^4 + 0.0021X^3 – 0.1612X^2 + 5.4156X + 3.9774$	0.8561	3.74
2011	Linear Least	$Y = 1.2583X + 23.205$	0.7764	4.49
	Exponential	$Y = 41.924e^{0.0143X}$	0.8123	4.25
	Power	$Y = 6.4378X^{0.6637}$	0.7661	4.68
	Logarithmic	$Y = 58.118\ln(X) – 140.5$	0.7242	4.99
	Quadratic	$Y = 0.0376X^2 – 2.4357X + 112.03$	<u>0.8374</u>	<u>3.83</u>
	Cubic Polynomial	$Y = 0.0001X^3 + 0.018X^2 – 1.15216X + 98.117$	0.8375	7.64
	Quartic Polynomial	$Y = -0.0002X^4 + 0.0334X^3 – 2.3128X^2 + 69.622X – 698.34$	0.8470	4.13
2013	Linear Least	$Y = 1.1099X + 31.215$	0.7577	4.45
	Exponential	$Y = 46.5e^{0.0126X}$	0.7933	4.35
	Power	$Y = 9.3757X^{0.5696}$	0.7378	4.70
	Logarithmic	$Y = 49.969\ln(X) – 107.97$	0.6965	4.32
	Quadratic	$Y = 0.0291X^2 – 1.7203X + 98.3$	<u>0.8157</u>	<u>3.89</u>
	Cubic Polynomial	$Y = -0.0004X^3 + 0.0927X^2 – 4.6436X + 141.58$	0.8175	8.05
	Quartic Polynomial	$Y = -0.0001X^4 + 0.0229X^3 – 1.4948X^2 + 41.929X – 354.13$	0.8457	4.03

Note: Y indicates CCI values; X indicates mean brightness of nightlight pixels.

For 2009, 2011, and 2013, as shown in Table 2, the linear, exponential, power, logarithmic, cubic, and quartic polynomial models were not selected for regression analysis because they all have high RMSE values despite high R^2 values. The quadratic model was selected because it shows the second highest R^2 value (good model fit) as well as the lowest RMSE value (low error value).

4.2 Empirical Comparison

Overestimations are more frequent than underestimations in each method, particularly for the NLSI-based method. The comparison of the frequency of approximately accurate estimates (within a +/-1% range) suggests the NLSI-based method is more reliable when compared with other methods, as it more frequently produces approximately accurate results (Table 3). Pearson chi-squared (X^2) test revealed that the observed underestimation, overestimation, and accurate estimation pattern for each method are not significantly different over time.

Table 3. Frequency of overestimations and underestimations in the 2009, 2011, and 2013 CCI

Method	Error Classification									Pearson's X^2 Test (DF =4)	
	Underestimates			Overestimates			Accurate Estimates			X^2 Statistics	p-value
	2009	2011	2013	2009	2011	2013	2009	2011	2013		
CNN	139	133	131	162	161	168	112	119	114	0.66	0.9563
NN	150	144	146	160	165	159	103	104	108	0.39	0.9834
IDW	127	121	125	185	187	179	101	105	109	0.64	0.9581
NLSI	99	102	103	199	188	191	115	123	119	0.69	0.9525

Note: DF indicates degree of freedom. The null hypothesis for this Pearson chi-squared test is that the observed underestimation, overestimation, and accurate estimation pattern for each method are not significantly different over time.

Absolute errors of the NLSI-based method were compared pairwise to those of the three established methods (NN, CNN, and IDW). Results are shown in Table 4. The two competing methods were considered to be equal if their absolute errors were within a +/-1% range. A better method has a lower absolute error. Pearson chi-squared (X^2) test revealed that the patterns of the observed results for each pairwise comparison are not significantly different over time.

Table 4. National Level Pairwise Comparison of Absolute Errors for 2009, 2011, and 2013 CCI

Method Comparison	Error Classification									Pearson's X^2 Test (DF =4)	
	NLSI Weaker			NLSI Better			Equal (+/-1%)			X^2 Statistics	p-value
	2009	2011	2013	2009	2011	2013	2009	2011	2013		
CNN vs. NLSI	107	105	109	177	172	165	129	136	139	0.89	0.9260
NN vs. NLSI	95	91	98	199	205	196	119	117	119	0.49	0.9742
IDW vs. NLSI	112	119	116	178	163	167	123	131	130	1.22	0.8743
CNN vs. NLSI	107	105	109	177	172	165	129	136	139	0.89	0.9260

Note: DF indicates degree of freedom. The null hypothesis for this Pearson chi-squared test is that the patterns of the observed method comparison results for each pairwise comparison are not significantly different over time.

A pairwise Wilcoxon Signed Rank Test, which can be used as a robust alternative to the parametric t-test (Wilcoxon 1945), was used to examine if the median absolute errors of NLSI-based were significantly different from the other three methods. As shown in Table 5, at the 5% significance level, there are statistically significant differences between all pairs. Further inspection of the mean absolute error shows that NLSI-based method always has a lower mean absolute error than CNN, NN, and IDW (Table 6). Because the mean absolute errors associated with NLSI were significantly less than those of CNN, NN, and IDW, it can be concluded that NLSI-based method outperformed the other three proximity-based methods to estimate CCI location adjustment factors.

Table 5. National Level Pairwise Wilcoxon Signed Rank Test for Each Method

Year	Pairwise Comparison	P-Value
2009	CNN vs. NLSI	<0.0001*
	NN vs. NLSI	<0.0001*
	IDW vs. NLSI	<0.0001*
2011	CNN vs. NLSI	<0.0001*
	NN vs. NLSI	<0.0001*
	IDW vs. NLSI	<0.0001*
2013	CNN vs. NLSI	<0.0001*
	NN vs. NLSI	<0.0001*
	IDW vs. NLSI	<0.0001*

Note: * indicates pairwise comparison result is significant at $p = 0.05$ level.

Table 6. Descriptive Statistics for Each Method

Year	Methods	Mean	Median	Standard Deviation
2009	CNN	3.01	2.00	3.16
	NN	3.66	2.50	3.92
	IDW	2.89	2.14	2.89
	NLSI	2.42	1.50	2.84
2011	CNN	3.42	2.34	3.61
	NN	3.99	3.10	3.91
	IDW	2.74	1.96	2.72
	NLSI	2.38	1.87	2.68
2013	CNN	3.48	2.45	3.67
	NN	3.89	2.90	3.77
	IDW	2.84	2.00	2.83
	NLSI	2.29	1.71	2.38

Error causation analysis showed that tree coverage is not significantly correlated with absolute errors of the NLSI-based method (correlation coefficient = 0.0983; *p-value* = 0.0459). In addition, waterbody is not correlated with absolute errors of the NLSI-based method (correlation coefficient = -0.0145; *p-value* = 0.7686). These results revealed that s environmental factors, including coverage of trees and water reflection, did not affect the performance of the NLSI-based method. One explanation is that the NLSI used in this study is an annual average daily stable nighttime light image which has already reduced the effects of tree coverage (less tree leaves in fall and no tree leaves in winter) and water reflection. That being said, the errors of NLSI-based method might be caused by other factors such as the sensors used to collect NLSI or the inherent errors of CCI produced during survey.

The direct users of the proposed method are commercial and governmental suppliers of location adjustment factors, such as RSMeans and the U.S. Department of Defense. Since there are only a few major suppliers for location adjustment factors, running a survey would not be very effective practice to collect location adjustment factors. This proposed method has the potential to reduce the number of survey cities required to characterize location adjustment factors, and then essentially reduce the associated cost and time. However, it should be noted that the proposed NLSI-based method is not intended for direct use by construction practitioners such as designers, general contractors, or individual project owners. These construction practitioners will benefit from the ease of computing location adjustment factors from NLSI, which can be automatized instead of undergoing annual time-consuming and labor-intensive surveys. Construction practitioners will also benefit from the effectiveness of the proposed method in the form of more accurate cost estimate results.

Chapter 5: Case Study Example

To illustrate the application of the research results to support cost estimating in a real world setting, it is useful to validate it through a case study. For this case study, we will assume that a large retail chain plans to build a new supermarket in the City of Hannibal, Missouri and there is no location adjustment factor for the City of Hannibal. In reality, Hannibal has a published CCI location adjustment factor, but for illustration purposes we will assume that it does not. The retailer initially develops a preliminary cost estimate based on historical costs for their supermarkets in other U.S. locations and would like to use the CCI dataset to adjust the preliminary estimate for Hannibal, Missouri. The estimated construction cost is $100 million prior to adjusting for project location. Since we are assuming there is no CCI adjustment factor for Hannibal, the company must estimate an adjustment factor. For this case study, we compare the results of using the NN, CNN, IDW, and NLSI methods to estimate the location adjustment factor for Hannibal, Missouri. Using the CNN method, the City of Hannibal's nearest neighbor within Missouri is Bowling Green, Missouri with a CCI of 94.2. Using the NN method, the City of Hannibal's nearest neighbor within conterminous U.S. is Quincy, Illinois with a CCI of 94.9. Using the IDW interpolation tool in ArcMap, the estimated CCI factor for Hannibal, Missouri is 94.5. Using the NLSI approach with the help of the quadratic polynomial regression function, the estimated CCI factor for Hannibal is 91.6. To analyze the accuracy of these methods, we compare the estimated location adjustment factors to the published CCI value for Hannibal, Missouri.

Different interpolation methods will predict different location factors. All of the estimation methods overestimate the adjustment factor, but the NLSI-based method produces a more accurate adjustment factor than the other three methods. Considering the estimated $100 million construction cost (non-adjusted for location), Table 7 shows that the NLSI-based method has the smallest estimate error, overestimating the cost by $2.5 million when compared to RSMeans CCI values collected by direct survey of market conditions. It should be noted that the improved accuracy in this cost estimate is only an improvement in the location adjustment and is relative to the RSMeans CCI estimate. It does not address any uncertainties in the original cost estimate or consider the realized cost of the hypothetical construction project, but does produce significantly improved location adjustment estimates of CCI and, subsequently, construction cost estimates when compared to existing methods.

Table 7. Case Study Results Summary

Location Adjustment Factor Type	Hannibal Location Adjustment Factor Value in 2009	Error	Error Percentage	Estimate Error (millions)
Actual	89.3			
CNN	94.2	4.9	5.5%	$5.5
NN	94.9	5.6	6.3%	$6.3
IDW	94.5	5.2	5.8%	$5.8
NLSI	91.6	2.2	2.5 %	$2.5

Chapter 6: Conclusions

The results of this research support the use of NLSI as a data source for developing location adjustment factors for locations where no adjustment factors currently exist. The analysis shows that the NLSI-based method outperforms proximity-based interpolation methods including NN, CNN, and IDW. One key advantage of the NLSI-based method over purely proximity-based interpolation methods is that it indirectly incorporates local economic conditions, as previous research has shown NLSI to be a proxy measure of economic activities.

A practical advantage of the NLSI-based method for estimating location adjustment factors is that it potentially reduces the amount of costly data collection activities required to develop CCI. A routine annual data collection for a minimum number of survey cities could be used to produce CCI for additional cities. The level of accuracy demonstrated by the NLSI-based method should be sufficient to produce realistic cost estimates for a variety of project locations, especially since it results in more accurate estimates when compared to existing proximity-based interpolation methods, including NN, CNN, and IDW.

Tree coverage and water reflection do not correlate with the errors of the NLSI-based method, and therefore, further studies should be performed to investigate the causation of errors. Other factors that warrant further study include sensitivity analysis and time series analysis to ensure that changes in brightness of NLSI corresponds to the change in CCI values through time.

The overall contribution of this study to the body of knowledge is a preliminary understanding of the relationship between NLSI and CCI location adjustment factors. Although the NN proximity-based location adjustment method is widely used in the construction industry, when compared to alternative proximity-based interpolation methods, the NN method does not perform well. A previous study suggests that CNN and IDW should be used to predict location factors for unknown locations (Zhang et al. 2014). This work found that the NLSI-based location adjustment factor estimation method outperforms all other proximity-based interpolation methods, including NN, CN, and IDW.

The uniqueness of this research is proposing a NLSI-based method which can improve the process of developing location adjustment factors for locations that have not been surveyed. This proposed method can rapidly and accurately predict location adjustment factors for unknown locations. When prediction models are established, cost estimators can estimate the location factors for any inhabited location in the U.S. These findings are highly relevant to construction industry location adjustment factor commercial suppliers because it allows them to improve the process of developing location adjustment factors and ultimately improve the accuracy of cost estimates. As a more accurate method for inferring location adjustment factors, the NLSI-based method has the potential to improve the performance of the construction industry through more accurate and therefore efficient, cost estimates.

References

Adriaanse, A., Bringezu, S., Hammond, A., Moriguchi, Y., Rodenburg, E., Rogich, D., and Schutz, H. (1997). "Resource flows: the material basis of industrial economies." World Resources Institute, Washington, D. C.

Amaral, S., Gamara, G., Miguel, A., Monteiro, V., Quintanilha, A., and Elvidge, C. D. (2005).

"Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data." *Journal of Computers, Environment and Urban Systems* 25， 179 – 195.

Arora, H. K., and Blackley, P. R. (1996). "An empirical analysis of the unit labor cost-product price relation in the U. S. Economy." *Atlantic Economic Journal*, 24(4), 321 – 335.

Bharti, N., Tatem, A. J., Ferrari, M. J., Grais, R. F., Djibo, A., and Grenfell, B. T. (2011). "Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery." *Science*, 334(6061), 1424 – 1427.

DMSP-OLS Nighttime Lights Time Series. (2009). [Online] Available http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

De Long, J.B., and Summers, L.H. (1990). "Equipment Investment and Economic Growth." *Quarterly Journal of Economics*, 106(2), 445 – 502.

Doll, C. N. H., Muller, J, and Elvidge, C. D. (2000). "Nighttime imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions." *AMBIO: A Journal of the Human Environment*, 29(3), 157 – 162.

Doll, C. N. H., Muller, J., and Morley, J. G. (2006). “Mapping regional economic activity from night-time light satellite imagery.” *Journal of Ecological Economics*, 57, 75 – 92.

Doll, C. N. H., and Pachauri, S. (2010). “Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery.” *Journal of Energy Policy*, 38(10), 5661 – 5670.

Elvidge, C. D., Baugh, K., Zhizhin, M., and Hsu, F. C. (2013). “Why VIIRS data are superior to DMSP for Mapping Nighttime Lights”. *Proceedings of the 35th Meeting of the Asia-Pacific Advanced Network*, APAN, Honolulu, HI.

Ebener, S., Murray, C., Tandon, A., and Elvidge, C. C. (2005). “From wealth to health: modeling the distribution of income per capita at the sub-national level using night-time light imagery.” *International Journal of Health Geographics*, 4(5), 1 – 17.

Ghosh, T., Anderson, S. J., Elvidge, C. D., and Sutton, P. C. (2013). “Using nighttime satellite imagery as a proxy measure of human well-being.” *Journal of Sustainability*, 5, 4988 – 5019.

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

Gould, F. E. and Joyce, N. (2009). *Construction project management*, 3rd Ed., Columbus, OH: Prentice Hall.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., and Baugh, K. (2003). “Validation of urban boundaries derived from global night-time satellite imagery.” *International Journal of Remote Sensing*, 24(3), 595 – 609.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., and Elvidge, C. D. (1997a). “A technique for using composite DMSP/OLS ‘City Lights’ Satellite Data to Map Urban Area.” *Journal of Remote Sensing of Environment*, 61(3), 361 – 370.

Imhoff, M. L., Lawrence, W. T., Elvidge, C. D., Paul, T., Levine, E., Privalsky, M. V., and Brown, V. (1997b). “Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States.” *Journal of Remote Sensing of Environment*, 59(1), 105 – 107.

Ito, K. (1987). *Encyclopedic dictionary of mathematics*, 2nd Ed., MIT Press, Cambridge, MA.

Johannes, J. M., Koch, P. D., and Rasche, R. H. (1985). “Estimating regional construction cost differences: theory and evidence.” *Journal of Managerial and Decision Economic*, 6(2), 70 – 79.

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

Levin, N., and Duke, Y. (2012). “High spatial resolution night-time light images for demographics and socio-economic studies.” *Journal of Remote Sensing and Environment*, 119, 1 – 10.

Lu, G., and Wong, D. (2008). “An adaptive inverse-distance weighting spatial interpolation technique.” *Journal of Computers and Geosciences*, 24(9), 1044 – 1055.

Ma, T., Zhou, C., Pei, T., Haynie, S., and Fan, J. (2012). “Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: a comparative case study from China’s cities.” *Journal of Remote Sensing of Environment*, 124, 99 – 197.

Migliaccio, M. C., Guindani, M., D’Incognito, M., and Zhang, L. (2012). “Empirical Assessment of Spatial Prediction Methods for Location Cost-Adjustment Factors.” *Journal of Construction Engineering and Management*, 139(7), 858 – 869.

Migliaccio, M. C., Zandbergen, P., and Martinez, A. A. (2013). “Empirical Comparison of Methods for Estimating Location Cost Adjustments Factors.” *Journal of Management in Engineering*, 31(2), 04014037.

Pozzi, F., Small, C., Yetman, G. (2003). “Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world.” *Earth Observation Magazine*, 12(4), 24 – 30.

RSMeans. (2009). *Building construction cost data*, 67th Ed., Kingston, MA.

RSMeans. (2011). *Building construction cost data*, 69th Ed., Kingston, MA.

RSMeans. (2013). *Building construction cost data*, 71st Ed., Kingston, MA.

Small, C., Pozzi, F., and Elvidge, C. D. (2005). “Spatial analysis of global urban extent from DMSP-OLS night lights.” *Journal of Remote Sensing of Environment*, 96, 277 – 291.

Sutton. P. (1997). “Modeling population density with night-time satellite imagery and GIS.” *Journal of Computers, Environment and Urban Systems*, 21(3/4), 227 – 244.

Sutton, P., Roberts, D., Elvidge, C., and Meij, H. (1997). “A comparison of Nighttime Satellite Imagery and Population Density for the Continental United States.” *Photogrammetric Engineering and Remote Sensing*, 63(11), 1303 – 1313.

Sutton, P., Roberts, D., Elvidge, C., and Baugh, K. (2001). “Census form heaven: an estimate of the global human population using night-time satellite imagery.” *International Journal of Remote Sensing*, 22(16), 3061 – 3076.

Sutton. P. C., and Costanza, R. (2002). “Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation.” *Journal of Ecological Economics*, 41, 509 – 527.

Sutton, P. C. (2003a). “A scale-adjusted measure of ‘urban sprawl’ using nighttime satellite imagery.” *Journal of Remote Sensing of Environment*, 86, 353 – 369.

Sutton, P. C. (2003b). “An empirical environmental sustainability index derived solely from nighttime satellite imagery and ecosystem service valuation.” *Journal of Population and Environment*, 24(4), 293 – 311.

Sutton, P. C., Cova, T. J., and Elvidge, C. D. (2006). “Mapping ‘Exurbia’ in the conterminous United States using nighttime satellite imagery.” *Journal of Geocarto International*, 21(2), 39 – 45.

Townsend, A. C. and Bruce, D. A. (2010). “The use of night-time lights satellite imagery as a measure of Australia’s regional electricity consumption and population distribution.” *International Journal of Remote Sensing*, 31(16), 4459 – 4480.

U.S. Census Bureau. (2014). [Online] Available http://www.census.gov/content/dam/Census/data/developers/understandingplace.pdf, October 15, 2014.

Waier, R. H. (2006). *Building Construction Cost Data 2006.* Robert S Means Co.

Wilcoxon, F. (1945). "Individual comparisons by ranking methods." *Biometrics Bulletin*, 1(6), 80 – 83.

Zhang, S. (2010). "Validation of geographically based surface interpolation methods for adjusting construction cost estimates by project location." Master's thesis, University of New Mexico, Albuquerque, NM.

Zhang, S., Migliaccio, G. C., Zandbergen, P. A., and Guindani, M. (2014). "Empirical assessment of geographically based surface interpolation methods for adjusting construction cost estimates by project location." *Journal of Construction Engineering and Management*, 140(6), 04014015.

Zhang, Q., and Seto, K. C. (2011). "Mapping urbanization dynamics at regional and global scales using multi-temporal DMPS/OLS nighttime light data." *Journal of Remote Sensing of Environment*, 115, 2320 – 2329.