Assessment of a Spline-Based Spatio-Temporal Model for Use with Educational Datasets

BY

SHANNON MILLIGAN
B.S., Oklahoma State University, 2004
M.S., Oklahoma State University, 2006
M.S., Oklahoma State University, 2008

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

George Karabatsos, Chair and Advisor
Yue Yin, UIC. Department of Educational Psychology
Joshua Radinsky, UIC. Department of Curriculum and Instruction
Stanley Sclove, UIC. Department of Information and Decision Sciences
Earvin Balderama, Loyola University Chicago

ACKNOWLEDGMENTS

I would like to thank my dissertation committee, Drs. George Karabatsos, Earvin Balderama, Josh Radinsky, Stan Sclove, and Yue Yin, for their support and guidance throughout the writing process. I would particularly like to thank my chair and advisor, Dr. George Karabatsos, for his willingness to continue to serve as my mentor when I said, "I just watched the movie *Contagion* and think I found my dissertation topic."

I would like to thank my parents, Stephen and Claudia Milligan, for their continued support of my education and their encouragement to pursue my dreams. Likewise, I would also like to thank my friend and fellow PhD student, Jennifer Sweet, for the support and inspiration she has provided me throughout the years. I would be remiss if I did not also thank Dr. Ken Fujimoto for his assistance in furthering my understanding of statistical modeling. For his continued support and belief that I could complete a PhD while working full-time, I extend unending gratitude to Dr. Michael Beazley.

Early work funded by NSF-MMS research grant SES-1156372 contributed to the ideas in this dissertation research.

SMM

TABLE OF CONTENTS

CHAPTER PAGE

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AR	Autoregressive Time Series
ARIMA	Autoregressive Integrated Moving Average Model
ARMA	Autoregressive Moving Average Model
BIC	Bayesian Information Criterion
CPS	Chicago Public Schools
DLM	Dynamic Linear Model
DV	Dependent Variable
EKG	Electrocardiogram
ISAT	Illinois Standards Achievement Test
ISBE	Illinois State Board of Education
MCMC	Markov chain Monte Carlo
MML	Marginal Maximum Likelihood
MSE	Mean-Squared Error
OLS	Ordinary Least Squares
PP1SD	Probability that the standardized beta coefficient falls within 1 standard deviation of 0
PSAE	Prairie State Achievement Examination
RR	Bayesian Ridge Regression
SN	Scaled Neighborhood Criterion

LIST OF ABBREVIATIONS (continued)

SPSS	Statistical Package for the Social Sciences
SSM	State-Space Model
SSVS	Stochastic Search Variable Selection
SVD	Singular Value Decomposition

SUMMARY

In this thesis, we argued for an increase in the use of spatio-temporal analysis in educational research. Though educational research frequently addresses questions that lend themselves well to spatial and spatio-temporal analyses (e.g. those involving differences across schools), the use of such analyses is almost non-existent in this field. We argued that the dearth of usage is due to three challenges: complications of large datasets, complex statistical models, and model selection. To overcome these challenges and therefore support this increase in usage of the analyses, we provided a three-part toolkit for educational researchers.

The first part of our toolkit was the introduction of a Bayesian linear spatio-temporal model which addressed the second challenge (complex statistical models), through incorporation of features familiar to educational researchers, such as the use of an ANOVA-based model structure. This first part of our toolkit also addressed the first challenge (complications of large datasets) through use of Bayesian Ridge Regression and thin-plate splines, which we included to better accommodate spatial, temporal, and spatio-temporal variables. The model also addressed this challenge through inclusion of marginal maximum likelihood estimation in our analyses, as the use of marginal maximum likelihood estimation, along with ridge regression, allowed for avoidance of the lengthy computational time often involved with Markov chain Monte Carlo sampling.

The second part of our toolkit addressed the statistical side of spatio-temporal analysis, as well as the second challenge (complex statistical models), through demonstration of the ease of use of our model for data analysis using the *Bayesian Ridge Regression* software program. This

software program utilizes a point-and-click interface, which is similar to statistical analysis programs that are commonly used in educational research, such as SPSS. To demonstrate the ease of use of our model, we analyzed two educational datasets, one relatively large (*n* = 18,506) and one smaller (*n* = 2,729). To further demonstrate the ease of application of the model and interpretation of the results, we also included sample posterior output, as well as a guide for how to run this particular model in the software program. This output included the time needed to complete each analysis; the fast computational times (under 30 seconds) addressed the first challenge (complications of large datasets). The ease of interpreting the results also addressed the third challenge (model selection), in part through demonstration of the ease of determining significant variables through use of the 50% posterior credible intervals, which are provided in the posterior output. Likewise, we also demonstrated the ease of determining the best combination of predictors to include in a model by using model fit indices [R^2, the D(m) statistic, the Akaike information criterion, and the Bayesian information criterion], which are also provided in the posterior output.

The third part of our toolkit addressed the descriptive side of spatio-temporal analysis through demonstration of techniques such as mapping. This more descriptive side of the analysis is important because it emphasizes further exploration of spatial and spatio-temporal relationships. We emphasized this exploration in part to caution researchers against conflating correlation and causation, especially when sensitive variables (e.g. race and ethnicity) are included in the analysis.

1. INTRODUCTION

1.1 **Background**

Spatial analysis of data, or analysis that requires data that is linked to a geographical location or locations (Weeks, 2004), is by no means a recent development; evidence of it has been found in analysis of weather patterns dating back to 1686 (Cressie, 1991). Snow (1855) used spatial analysis to find the root of the 1854 London cholera outbreak, as did Fisher in his agricultural field trials (Diggle, 2010). Spatial data analysis continues to be used extensively in fields that emphasize analyzing differences across locations, such as ecology and geography. In a similar vein, fields that commonly require observations across time, such as medicine, frequently utilize *temporal* analysis of data. We see more and more application of *spatio-temporal analyses*, or analyses that incorporate both spatial and temporal data, as well as increased development of statistical models capable of accurately representing spatio-temporal data. This increased application and development is due in part to a growing recognition of the value of simultaneously modeling both spatial and temporal aspects of a dataset. For example, the field of epidemiology frequently utilizes spatio-temporal analyses and models in applications such as disease mapping, where researchers attempt to predict the spread of disease across both geographical locations and time.

The arguments for including spatial, temporal, or spatio-temporal analyses have both methodological and statistical bases. Cressie (1991) argued that all data inherently have some spatial and temporal aspects to them, and to leave out these variables is to risk drawing false conclusions about the nature of the relationships between independent and dependent variables. For example, rainfall data collected at stations that are close to each other are likely to be similar, as are data collected at close points in time. Failure to factor these into the data is failure to account for these correlations during the analysis, allowing for possible misattribution of results.

As a methodological example of the importance of including spatial and temporal variables, consider that one purpose of social science research is to explain human behavior (Goodchild & Janelle, 2004). To illustrate, consider a study where we analyze public transportation usage among Chicago residents. If we do not account for variables such as place of residence or location of transportation, we risk creating what Robinson (1950) termed an *ecological fallacy*: drawing conclusions about the behavior of individuals based on findings from analyses conducted on populations. We miss the possible discovery of patterns of behavior in certain parts of the city or among certain groups of people. In addition, by not accounting for time we miss the opportunity to study this behavior *longitudinally*, or with repeated observations across time, rather than as a cross-section, or a single observation at a single point in time, thus missing a chance to strengthen our research design (Goodchild and Janelle, 2004). Longitudinal studies allow researchers to look for consistency in findings and, to an extent, to create an argument for the ability to replicate studies with similar results (Haining, 1990).

We can extend the importance of including spatial and temporal variables to educational data. Educational literature makes a strong case for the need to consider a student's home environment and *neighborhood effects* (Wilson, 1987), or neighborhood characteristics (e.g. concentration of poverty and crime rates), in educational planning (Berliner, 2006; Bronfenbrenner, 1976). Building off this concept of neighborhood effects, researchers have explored numerous variables at the neighborhood-level, including the quality of social relationships and crime rates (Sampson, Raudenbush, & Earls, 1997). Savitz and Raudenbush (2009) extend this research further by creating a *Bayesian* model that allows for *spatial* *dependence*, or the tendency for locations that are close to each other to have similar outcome values. This spatial dependence is often inherent in neighborhood-level analyses[1].

In the remainder of this chapter, we present the opportunities and challenges posed by the increasing use of spatio-temporal models in educational research. We will illustrate the benefits through the use of an example from an analysis of Chicago Public Schools' (CPS) data. This is followed by a literature review covering current spatial and spatio-temporal statistical models. We also include a brief overview of *ridge regression*, a *shrinkage estimation* approach, which will eventually serve as an important part of our proposed spatio-temporal model for educational data. We conclude with a brief look forward to the remainder of this thesis.

1.1.1 **An Illustration of the Benefits of Spatial Analysis**

In May 2013, the Chicago Board of Education decided to close 45 schools in the Chicago Public Schools (CPS) system, effective at the start of the 2013-2014 school year. Per communications from CPS concerning the closings, factors considered in determining which schools to close included school enrollment size (Chicago Public Schools, n.d.). Schools with enrollments both below and above the district-set standard for "Ideal Performance Enrollment," as set forth in the *Chicago Public Schools' Space Utilization Standards* (Chicago Public Schools,

---

[1] However, caution must be taken when considering neighborhoods as a unit of spatial analysis. For one, there is not consensus on how to define the spatial boundaries of neighborhoods (Roberto, 2015), as well as on appropriate ways to measure neighborhoods, such as the use of city blocks (Sampson, 2008). Sampson (2008) also references the ecological fallacy in questioning when (if ever) it is appropriate to draw conclusions about individual residents based on neighborhood-level findings. Per Sampson (2008), the non-experimental nature of much research on "neighborhood effects" casts serious doubts on whether causality can be inferred from research findings. In the same vein, since much research on neighborhoods focuses on racial segregation and poverty, researchers must also consider the caution of Holland (2008) against treating race as a causal variable, in no small part because of the failure to consider other unobserved covariates that may actually account for observed effects.

2011), were targeted for closing. Using school enrollment size as a main criterion may make less intuitive sense than using academic performance, particularly in the cases of schools with large enrollment, but the literature abounds with studies exploring the relationship between school enrollment size and student academic performance. The majority of these studies report a negative relationship between large school enrollment and student performance (Leithwood and Jantzi, 2009). This holds true in elementary schools (Kuziemko, 2006) and secondary schools (Lee and Smith, 1993; Bloom, Thompson, and Unterman, 2010).

Though only 45 of the over 600 CPS schools were closed, Radinsky and Waitoller (2013) estimated that upwards of 133 schools were impacted. This increase was largely due to consolidations and the designation of some schools as "receivers" for displaced students. As a result of their increased estimate of the number of schools impacted by the closings, the number of students impacted also increased substantially, from the roughly 30,000 cited by CPS to almost 47,500 (Radinsky & Waitoller, 2013).

Radinsky and Waitoller (2013) then delved deeper into the data to gain further insight into not only the schools that were closed, but also the students affected by the closings. Through their research, Radinsky and Waitoller determined that the overwhelming majority of the schools that were closed were located in neighborhoods in the West and South sides of the city of Chicago ("Looking Closer at Closings," 2013). In general, these West and South side neighborhoods tend to have disproportionally higher rates of both poverty and minority residents when compared with the rest of Chicago. Not surprisingly, then, the authors found that 81% of the CPS students impacted by the school closings were African-American (Radinsky & Waitoller, 2013).

It was not necessary to use spatial analysis to determine that the majority of the CPS students impacted by the school closings were African-American. What spatial analysis did allow for, however, was a demonstration that the closings were largely concentrated in certain West- and South-side neighborhoods in Chicago. This demonstration opens the door for important policy-related conversations around segregation, both in the city and in CPS, and the ramifications of this segregation for educational access ("Looking Closer at Closings," 2013).

The CPS example illustrates the impact of including spatial elements in research and invites conversations around not only the impact of city or neighborhood-level segregation in education, but also the spatial nature of segregation. We can extend this conversation on segregation into the field of sociology, where recent work on the spatial nature of *residential segregation* (i.e. segregation in housing) is reflected in the research of Roberto (2015). Roberto (2015) examined traditional sociological measures of segregation and determined that many either fail to account for spatial considerations such as geographic scale (i.e. the size of a segregated area or areas relative to a larger area, such as a city) and spatial proximity (i.e. what neighborhoods are next to each other, as well as measures of distance between neighborhoods or certain points within neighborhoods) or fail to do so adequately (e.g. do not consider the possible impact of rivers and other natural boundaries on segregation).

Roberto (2015) addressed this inadequacy through development of a method that utilizes roads in a city and distance along these roads, since roads reflect the presence of boundaries (e.g. if a river is present, not all roads contain bridges over the river, thus creating a boundary), rather than Euclidean distance between points. The use of roads, and acknowledgement of these *spatial boundaries*, also shifts the conceptualization of residential space away from census tracts, reflecting a more realistic view of the spatial boundaries observed in cities and experienced by residents. Roberto built on how segregation is experienced by residents through emphasis on the local effects of segregation; that is, demonstration that segregation is not an index that can be uniformly calculated across a city, but a construct that needs to be considered in terms of scale. For example, does segregation exist on a smaller geographic scale, as in neighborhoods such as Chinatown, or on a larger geographic scale, such as the West and South sides of Chicago?

This illustration from sociological research supports the argument for the need to increase the inclusion of spatial analysis in educational research. Educational research often pursues research questions that lend well to spatial analysis, yet is largely devoid of such analyses. Use cases from other social science disciplines provide a more accessible framework and starting point for extension into education, when compared with those from the natural sciences. The work done by Roberto (2015) on defining spatial boundaries, for example, resonates in educational research on schools, where researchers are concerned with boundaries such as neighborhoods, aldermanic wards, and those used to determine enrollment for schools. For example, an educational researcher might ask if and how school boundaries shifted with the introduction of new charter schools, and what the resident demographics look like within these boundaries. This work on defining spatial boundaries also extends naturally into the city of Chicago, where the borders of the 77 neighborhoods are formed largely by roads (Sampson, 2008).

1.2 **Problem Statement**

Educational research focused on physical locations, such as school locations, is well-suited to spatial analysis. Of the different types of spatial data, which we will cover in more detail later in this thesis, these school locations are classified as "*point*" data, which can be easily characterized using coordinates such as latitude and longitude (Banerjee & Fuentes, 2011). Even research focused on broader locations, such as neighborhoods or cities, has the advantage of looking at areas that are likely to remain geographically fixed across time and have relatively stable boundaries. Contrast this with environmental research studying rainfall amounts across places, where boundaries are more fluid and the locations of interest are likely to change over time. It is worth noting that areas remaining geographically fixed across time does not imply *spatial stationarity*, wherein the mean and *variance-covariance* of the distribution are assumed to be invariant across all spatial locations (henceforth denoted by ***S***) in the domain of interest (Cressie, 1991).

An advantage to incorporating spatio-temporal variables into educational research can be increased validity and accuracy of both findings and interpretation of results (Cressie, 1991). As an example, Savitz and Raudenbush (2009) found that their Bayesian statistical model, which incorporated spatial dependence, was better able to predict future Chicago homicide rates from measures of *collective efficacy* [i.e. "social cohesion among neighbors" (p. 167)] than models that failed to account for this dependence. As another example, think again of the concept of the ecological fallacy (Robinson, 1950), wherein findings based on aggregated data are assumed to be true for individuals and vice versa. Now think of this in terms of research findings across a city-wide school system. If school location-and perhaps *covariates* (predictor variables) related to neighborhood characteristics-are not included, we risk making generalizations that may be true in the bigger picture, but may also only be true for schools in certain parts of the city. In addition, we miss exploring relationships between proximal and distal schools, and what neighborhood characteristics similarly performing schools (e.g. schools with similar average standardized test scores) may have in common (Alperin, 2008). Likewise, if we do not look at findings across multiple time points, we may assume that a one-time drop in performance is a sign of ongoing decline, rather than a fluke. However, it is important to note that including spatial and temporal variables does not guarantee improved results; significance testing of these predictors is important to ensure that variables are intentionally included in models. In addition, it is important to note that these models are largely intended to look at relationships between all predictors, rather than to establish causal relationships. It would be irresponsible to draw causal inference about, for example, increased school enrollments in a particular neighborhood without exploring other relationships and contextual factors.

There are also challenges with using these models in educational research. For one, while educational datasets vary widely in size, some can be large, both in terms of number of cases (subjects) and number of predictors (covariates), denoted *p*, often due to government mandates around data collection for accountability measures. This issue of large datasets is compounded with spatio-temporal models, where each time period *t* of data collection adds another set of data points (cases by variables by time). For a fully Bayesian approach to a *Gaussian process* (i.e. normally distributed) *model*, which is a standard model for spatial data, these large datasets can be computationally unwieldy due to the large-scale *matrix inversions* (i.e. specific algebraic computations) that need to be performed for each *Markov chain Monte Carlo* (MCMC) iteration, or sampling from a given probability distribution (Gelfand, Banerjee, & Finley, 2011; Hancock & Hutchinson, 2006). These MCMC iterations are necessary in order to estimate the covariance function parameter for the model. In the case of the Gaussian process model, the covariance function is of dimension n x n, where n is the sample size of the dataset. Given that the MCMC matrix inversions iterations occur on the order of n^3 (Gelfand, Banerjee, & Finley, 2011; Hancock & Hutchinson, 2006), it is clear to see that parameter estimation using a Gaussian process model would be extremely computationally intensive when used on an educational dataset with a sample size in the hundreds or thousands.

However, with the growing availability of massive amounts of data, from a multitude of sources and with varying degrees of structure, has come increased interest in creating models capable of handling large datasets. Hensman, Fusi, and Lawrence (2013) developed a model that employs *stochastic variational inference* (SVI), where a smaller number of variables are used as a proxy for the whole, and applies to both Gaussian (i.e. normal) and non-Gaussian models. Eidsvik, Finley, Banerjee, and Rue (2012) described a *dimension-reducing*, or *low-rank*, predictive process model that utilizes fixed, non-grid based representations of all spatial locations in ***S***, known as "*knots*." This model also applies to both Gaussian and non-Gaussian models.

Another challenge is the level of technical and statistical knowledge needed to perform these analyses. Many spatial and spatio-temporal models are very complex and difficult to both understand and to use (Stroud, Müller, & Sansó, 2001). For example, a number of these models utilize Bayesian inference; in education graduate programs, training on Bayesian inference is not nearly as common as training in *frequentist* methods (for a detailed description of Bayesian and frequentist inference, please see Section 1.3.1). Finally, running these models often requires knowledge of coding-based statistical programs such as *R* or *MATLAB*.

A third challenge is that of model selection, particularly in large datasets. That is, when there are many variables of potential interest, how does a researcher decide which variables to include in the final model? Even in a study that does not incorporate spatial analysis, it is easy to imagine a large number of background variables, such as student gender, age, and test scores. The addition of spatial locations creates more potential variables, especially when data is collected over time. As an illustration, consider a study where the researcher is interested in looking at the relationship between crime rates and the amount of green space in a neighborhood. In order to control for possible *confounding variables*, a large number of neighborhood-related variables, such as poverty rate and educational attainment, must be included. In addition, a number of spatially relevant variables might be included. For example, Tobler's (1970) first law of geography states that, "everything is related to everything else, but near things are more related than distant things" (p. 236).[2] Therefore, we might include a variable such as distance between neighborhoods, in order to examine whether high correlations exist among our variables between nearby neighborhoods. This phenomenon is also known as *spatial autocorrelation* (Moran, 1950; Cliff & Ord, 1973). Inclusion of distance between neighborhoods, though, also necessitates determination of how to measure distance [i.e. by using Euclidean distance or road distance, as done by Roberto (2015) in a study of segregation and spatial boundaries]. As one might imagine, it could take a great deal of time and iterations of analysis to determine the best combination of variables to model the relationship between crime and green space.

We acknowledge that there are many methods for selecting which variables to include in a model. Some of the more manual methods involve adding and removing variables based on statistical criteria. In the case of the crime rate example, this approach would be particularly cumbersome and time-consuming. Given the propensity for spatio-temporal datasets to contain large numbers of variables, we suggest using methods that allow for automatic variable selection and the use of multiple *model fit indices*.

---

[2] However, researchers must be careful to not assume that close proximity guarantees a higher degree of relationship. For example, while two schools may be physically close to each other, one may be a neighborhood school (students from the surrounding neighborhood) and the other selective enrollment (students from across the city). Therefore, even though the schools are located near each other, they may have very few similarities in terms of attributes such as student body composition.

1.3 **Motivation for the Study**

In thinking about the motivation for the need for spatio-temporal analyses in educational research, we consider the following sample research questions, which incorporate spatial or temporal elements:

1. Does proximity to a charter school have an impact on Illinois Standards Achievement Test (ISAT)/Prairie State Achievement Examination (PSAE) test scores in public schools?
2. Is there a higher percentage of students with autism in schools in certain parts of the city of Chicago?
3. For those students who were displaced by the 2013 school closings, is there a relationship between distance from their place of residence to their new school and the students' reading and math achievement?
4. Does residential proximity to pet coke storage piles in Chicago have an effect on student reading and math achievement?
5. What factors had an impact on the likelihood of a CPS school being closed between 2001 and 2006? Do these factors differ for CPS schools closed in 2013? Do these factors vary across the city of Chicago?

As evidenced by these sample research questions, inclusion of spatial and temporal variables can bring richness to educational studies, and providing an approach that alleviates the current obstacles to their use in education research may result in the greater adoption of these models. Therefore, we seek to alleviate these obstacles through two approaches:

1. Presentation of a *linear* spatio-temporal model that is well-suited to analysis of large datasets.
2. Presentation of a Bayesian approach to analyzing spatio-temporal data that is more accessible to the broader community of educational researchers.

In their presentation of research into the Chicago Public Schools (CPS) school closings, Radinsky and Waitoller highlighted the advantages of both having large datasets in educational research and having the ability to quickly analyze and to report out on this data. Radinsky and Waitoller stressed the latter in particular due to the importance of policymakers having access to research findings when making decisions ("Looking Closer at Closings," 2013). Based on this concern, among others, we focus on the presentation of a new model, rather than on spatial reasoning or spatial analysis tools such as geographic information systems (GIS) [for more coverage of these topics, see Goodchild (2008), Goodchild and Janelle (2010), and the National Research Council (2006)]. This focus on a new model, as well as a suggested new ridge regression-based approach for analyzing spatio-temporal data, ensure that this thesis both contributes to the educational research literature and presents clear and immediate practical applications.

1.4 **Research Goals**

Based on our problem statement and, in particular, the challenges we noted therein, we propose the development of a Bayesian *spline*-based ridge regression model that can account for space and time effects, as well as the effects of other predictors. Our proposed model addresses common problems with the analysis of large datasets, in particular those around the speed and feasibility of the analysis. In addition, our proposed model also provides a useful tool for allowing educational researchers to explore research questions with spatial dimensions (e.g. the relationship between whether a school was closed and its proximity to charter schools). This model is meant to be used with spatial locations that are classified as "point" data, such as schools. In addition, this model provides researchers with a method to analyze spatio-temporal data that is accurate and easy to use, in order to answer the question of whether spatial and spatio-temporal variables help to explain the dependent variable in the regression analysis.

As a supplement, we will also provide a guide for applying our model in Appendix B, thus making it more accessible for educational researchers. This guide will provide both a brief overview of our model as well as steps for using the model in the *Bayesian Ridge Regression* software developed by Dr. George Karabatsos (2016).

1.5 Review of Literature

The classical *geostatistical*, or spatial statistics, model is given by:

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s}) \tag{1.1}$$

where $\mu(s)$ is the mean function and $e(s)$ is the error (Zimmerman and Stein, 2010). This model seems simple when compared with other statistical models, spatial or otherwise, but for early applications in mining, it was sufficient (Diggle, 2010). This model would not be sufficient, however for an application such as disease mapping, for a number of reasons. As both Diggle (2010) and Cressie (1991) note, new research questions and new discoveries necessitate development of better- suited spatial models. These newer, more sophisticated models are better equipped to detect spatial dependence and to allow for *hierarchical structures* (i.e. breaking the model into linked sub-models to allow for greater flexibility in parameter estimation; Wikle, 2010), among many others. In a nutshell, these more sophisticated models lend for confidence in the results of the analyses.

The literature on spatial, temporal, and spatio-temporal models is vast and covers a wide array of model approaches. For the sake of this thesis, we will present a brief overview of model approaches while focusing more on linear regression models (i.e. models that aim to predict or to explain relationships between variables and that assume a linear relationship between the dependent variable[s] and predictor variable[s]), since this is the class of models likely to be most familiar to educational researchers and practitioners. This overview draws extensively on existing guides on spatial modeling and statistics [see Cressie's (1993) seminal book, *Statistics for Spatial Data*], temporal modeling and statistics [see Box and Jenkins' (1976) seminal book, *Time Series Analysis: Forecasting and Control*, and Prado and West's (2010) *Time Series: Modeling, Computation, and Inference*], spatio-temporal modeling and statistics [see *The Handbook of Spatial Statistics* (2010), edited by Gelfand, Diggle, Fuentes, and Guttorp] and the intersection of spatial and spatio-temporal data and the social sciences [see Haining (1990) and Goodchild and Janelle (2004)].

1.5.1 **Spatial Data Analysis**

As a broad, introductory overview, it is worth noting that there exist an abundance of spatial models utilizing both frequentist and Bayesian frameworks. By a frequentist framework, we mean an approach that assumes that model parameters for a dataset (e.g. mean and variance-covariance matrix) would be fixed upon repeated analysis of the data. Prior information about the data is not taken into consideration, and inferences about the data are made based on rules for rejecting or failing to reject the *null hypothesis* (i.e. no relationship between variables). In contrast, a Bayesian framework assumes that model parameters are determined probabilistically through sampling, and therefore not fixed upon repeated analysis. Prior information is taken into consideration, and inferences about the data are based on updating this prior distribution with the data from the sample, then using the resulting posterior distributions (Casella, 2008). Of the references listed above, Cressie (1991) focused more on examples of frequentist models, while Gelfand, Diggle, Fuentes, and Guttorp (2010) also included Bayesian models. Additional examples of frequentist models include the hierarchical spatial model of Royle and Berliner (1999). Likewise, additional examples of Bayesian models include the spatial models of Wikle (2003); Banerjee, Gelfand, Finley, and Sang (2008); and Higdon, Swall, and Kern (1999).

We mentioned earlier that educational research data involving schools would be classified as “point” data, which can be characterized using latitude and longitude (Banerjee & Fuentes, 2011). It could be argued that this represents a best case scenario for spatial data analysis: by definition, spatial data analysis requires data that is linked to a geographical location or locations (Weeks, 2004), and with “point” data the geographic locations are included in the dataset in the form of latitude and longitude coordinates. However, it is important to note that spatial analysis can extend beyond “point” data to larger geographical regions. For example, a researcher interested in looking at academic performance of students from across the city of Chicago may use neighborhoods as the units of spatial analysis. While a neighborhood cannot easily be characterized using a single latitude and longitude coordinate, it is still a spatial location capable of analysis. Indeed, the book *Spatially Integrated Social Science*, edited by Goodchild and Janelle (2004), is divided into sections that recommend spatial analyses for different levels of spatial locations: “point” data, neighborhoods, and larger regions such as countries.

Thus, planning for research involving spatial data includes careful consideration of the spatial locations and type or types of spatial data that will be collected. Haining (1990, p. 39) offered the following considerations of spatial locations:

1. Are the spatial locations of interest a series of points or a larger area?
2. Do the spatial locations cover an entire area or are they clustered in certain parts? (e.g. if one wished to study rainfall amounts in a region, would the amounts be collected across the region or in certain locations within the region?)
3. What boundaries, whether physical or symbolic, exist between spatial locations and how might these boundaries impact the findings? (e.g. boundaries between urban and rural areas)
4. What relationships might exist between spatial locations? What external factors may play into these relationships?

Cressie (1991) urged consideration of whether the spatial process should be classified as *continuous* (or geostatistics), *discrete*, or as part of a *point process*. The importance of identifying the type of spatial process rests largely in the question to be answered. For example, continuous spatial processes are ideal when one seeks to predict or to infer values, while discrete spatial processes are better suited for explanation of phenomena (Gelfand, Diggle, Fuentes, & Guttorp, 2010). Likewise, discrete spatial processes and spatial point processes are more analogous with temporal (time series) analysis due to the countable nature of discrete data and the ability to detect patterns (Cressie, 1991). We now offer brief overviews of each of the three aforementioned types of spatial data processes.

Continuous spatial processes, or geostatistics, are those that allow for spatial locations that vary continuously across the larger, typically two- or three-dimensional spatial domain of interest, denoted as *D*. Cressie (1991) noted that the term “continuous spatial processes” is often used interchangeably with the term “geostatistics.” The aims of these processes are prediction, often referred to as *kriging* (Matheron, 1963) in geostatistics, and inference across *D* (Gelfand, Diggle, Fuentes, & Guttorp, 2010). Because it is unlikely that we sampled all spatial locations across *D*, we seek to predict or to infer values of the dependent variable(s) for unsampled locations. An example of a continuous spatial process would be the monitoring of rainfall across the state of Oklahoma. Here *D* is the state of Oklahoma and we can assume that rainfall data is collected at certain sites across the state. Though we would not have rainfall data for all sites within *D*, we could use known values to make predictions for the unsampled sites. An example from educational research would be looking at test scores from the Chicago Public Schools (CPS) district. If a researcher treated the city of Chicago as the spatial domain of interest and had test scores from at least some CPS schools, the researcher could look at the spatial distribution of test score values across the city and then examine the characteristics of parts of the city with expected higher and lower test scores.

The domain of interest with discrete spatial processes, on the other hand, is only specific spatial locations. Cressie (1991) referred to discrete spatial processes as *lattice*-based because of the finite and therefore easily countable number of spatial locations. In addition, the use of specific spatial locations allows for representation of the sites on a map or grid, and thus also for relatively easy calculation of quantities such as distance between locations. The aims of these processes are to detect *spatial patterns* (e.g. similar outcome values in neighboring sites) in order to facilitate an explanation for phenomenon and to smooth these discrete points to enhance visualization of the data (Gelfand, Diggle, Fuentes, & Guttorp, 2010). An example of a discrete spatial process would be disease mapping in epidemiology. One illustration of disease mapping would be to analyze relative risk of brain cancer across the Chicago neighborhoods, given known numbers of occurrences of brain cancer and potentially background information about those with the disease. This information could then be used to look for patterns across neighborhoods and possible explanations for these patterns. An example from educational research would again be looking at schools with high percentages of students with autism, with an emphasis on looking at the number of these schools in each neighborhood. The researcher could then look for patterns between the neighborhoods, such as whether close-by neighborhoods tend to have similar counts of schools with high percentages.

The final type, spatial point processes, is slightly more difficult to conceptualize than continuous and discrete spatial processes. In spatial point processes, the focus is more on random events in a set, again occurring in a two- or three-dimensional spatial domain of interest, *D*. Here the “point” refers to spatial locations, “process” refers to the model, and the “pattern” or “event” refers to the realization (Cressie, 1991; Gelfand, Diggle, Fuentes, & Guttorp, 2010). Spatial point processes can be defined in two ways: by the spatial locations of the realizations or mathematically based on counts of the realizations, which could then be represented by a *Poisson point process*, among others (Isham, 2010). The aim of spatial point processes is to determine whether the events demonstrate complete spatial randomness or if there is evidence of clustering (i.e. spatial dependence). Examples of spatial point processes include analyzing the locations of trees in a forest, earthquakes, and even lightning strikes. As might be expected, these processes are particularly well-suited to spatio-temporal adaptation due to the relevance of knowing both the time and location of an event (Cressie, 1991). An example from educational research would be looking at the CPS school closings. In this example, the spatial locations would still be the schools, but the event would be the school closings. The researcher could then determine whether evidence exists of clustering within the closed schools, or whether the schools exhibit random distribution across the city.

1.5.2 **Temporal (time series) Data Analysis**

One related concept with temporal, or time series, data analysis is a longitudinal research design. In the longitudinal design instance, we think of measuring variables at different points in time, as the research is being conducted. In a similar vein, we could also think of an electrocardiogram (EKG), measuring a person's heart rate across a period of time. However, time series analysis can also be used to look at future time periods, as with economic or meteorological forecasting.

Time series analysis, like spatial analysis, can also be continuous or discrete. We can use the example of the EKG as an illustration of the difference between continuous and discrete time series analysis. By nature, an EKG is a continuous measure: it measures a person's heart rate across all time points within a given interval. However, if a person's pulse is read at certain time points within that interval, such as every minute, then it is a discrete measure. It is worth noting that while time is often measured at equal intervals, especially in educational research, equal intervals are not required to perform the analyses. The notation of the time series differs, however, depending on whether there are fixed intervals between time points, with:

$$y_t,\ t = 0, 1, \ldots, \mathrm{T} \text{ representing equally spaced intervals and} \tag{1.2}$$

$$y_{ti},\ i = 1, 2, 3, \ldots \text{ representing unequally spaced intervals (Prado \& West, 2010)} \tag{1.3}$$

One important determination in selecting an appropriate model for analyzing time series data is whether the data can be classified as stationary or non-stationary. For time series data to be classified as "stationary," we assume that the probability distribution of the dependent variable y_t, which could be illustrated through use of the mean and variance (or variance-covariance matrix) of y_t, is the same across all time points (Prado & West, 2010). If the distributions vary across temporal observations, the data is classified as "non-stationary" (Box & Jenkins, 1976). Due to our use of a linear spatio-temporal model, we will focus on linear time series models for both stationary and non-stationary data.

For stationary time series data, a common set of linear models is the *autoregressive* time series models (AR). The AR accounts for expected correlations between the data at various time points, also known as *temporal autocorrelation*. The AR is given by:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \epsilon_t \tag{1.4}$$

where:

-y_t is a function of time points,

-φ is a constant parameter,

-p is the order of the autoregression, and

-$\epsilon_t$represents error terms (Prado & West, 2010)

The parameters φ_i can be used to determine the degree of temporal autocorrelation between these included time points by use of Yule-Walker equations (Box & Jenkins, 1976). A description of these equations is outside of the scope of this thesis; therefore, we direct interested readers to Box and Jenkins (1976) for a more extensive overview. The number of included time points (or temporal lag terms) included in the model is given by the order of the autoregression (AR), given by p [i.e. AR(p)], as noted above. For example, a model that only includes t and one lag term, t–1, would be classified as a first-order autoregressive model, or AR(1). In the case of AR(1), the model reduces to:

$$y_t = \varphi_1 y_{t-1} + \epsilon_t \tag{1.5}$$

reflecting that the model only includes *t* and the one lag term, *t*-1.

Another common set of linear models is the *autoregressive moving average models* (ARMA), popularized by Box and Jenkins (1976). The ARMA accounts for temporal autocorrelation in addition to correlations between the lag error terms (Prado & West, 2010). The ARMA is given by:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{1.6}$$

where:

-y_t is a function of time points,

-φ is a constant parameter,

-p is the order of the autoregression,

-q is the order of the correlated lag error terms, and

-$\epsilon_t$represents error terms (Prado & West, 2010)

It is worth nothing that an ARMA that includes zero correlated error terms would reduce to an AR model, which assumes that error terms are uncorrelated (Prado & West, 2010).

However, it is likely in many research scenarios, especially in education, that the mean and variance parameters do vary across time. As such, there has been a rise in the development of time series models equipped to fit non-stationary data, where the probability distribution of the dependent variable y_t varies across all time points (Box & Jenkins, 1976; Prado & West, 2010). Keeping with our focus on linear models, we will briefly highlight two models for non-stationary data: the *autoregressive integrated moving average* (ARIMA) model (Box & Jenkins, 1976) and *dynamic linear models* (DLMs) (West & Harrison, 1997). The ARIMA model incorporates an integrating parameter, *d*, which is given by the number of parameter differences across temporal observations. Per Box and Jenkins (1976), if the time series is integrated *d* times, it will then be transformed into a stationary and therefore more easily interpretable process. This is a complex process, though, and we primarily describe the ARIMA because the previously described ARMA model is a derivation of the ARIMA model when $d = 0$; that is, when the data are stationary. We will focus more on the second models, dynamic linear models (DLMs), due to their additional application in spatio-temporal modeling. A DLM is an extension of the abovementioned AR and ARMA models that incorporates the parameter θ_t, which is a state vector of parameters at a given time t (Prado & West, 2010). The inclusion of this state vector allows for observations of temporal variations for $t = 0, 1, \ldots, T$. In addition, due to the inclusion of this state vector, DLMs are also referred to as *state-space models* (SSMs). We present a description of spatio-temporal DLMs, as well as illustration of a particular model, in the following section.

1.5.3 **Spatio-temporal Data Analysis**

To mirror the overview to the spatial data analysis section, there also exist an abundance of spatio-temporal models utilizing both frequentist and Bayesian frameworks. Additional examples of frequentist models include the spatio-temporal model of Huang and Cressie (1996). Likewise, additional examples of Bayesian models include the spatio-temporal models of Wikle, Berliner, and Cressie (1998) and Savitz and Raudenbush (2009).

Spatio-temporal models are frequently developed through adapting or combining spatial analysis models and time series models. For example, Gelfand, Kim, Sirmans, and Banerjee (2003) proposed a spatio-temporal analog to the temporal autoregressive moving average models (ARMA) via spatio-temporal models that allow for coefficients to vary across space or both space and time. In one instance, Gelfand et al. (2003) modeled the time component by an autoregressive model. These models can also be developed through extensions of existing non-spatial and non-temporal models, however. For example, Karabatsos and Walker (2012) demonstrated how their Gaussian-process based, random partition Bayesian non-parametric regression model can easily be extended to account for spatio-temporal data.

The dynamic linear models (DLMs) mentioned in the previous section are now frequently used in spatio-temporal analysis, including the approach of Stroud, Müller, and Sansó (2001). DLMs are also used in the approach of Gelfand, Banerjee, and Gamerman (2005), which adapted dynamic models for both univariate and multivariate data and was updated by Finley, Banerjee, and Gelfand (2012) to utilize a predictive process approach, thus allowing the model to be better suited for use with large datasets.

In the context of spatio-temporal analyses, DLMs are models that allow for non-stationary spatio-temporal data, where parameters change over time and across spatial locations (Prado & West, 2010). In general, these models assume continuous spatial data ($\boldsymbol{S} = \boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_n}$) and discrete time data (t = 0, 1, …, T). As mentioned previously, DLMs are also referred to as state-space models due to the inclusion of a vector of state parameters, or model parameters that capture the "state" of the outcome at an observation point. In the spatial case, the state could be an observation at one particular spatial location $\mathbf{s}$, while in the spatio-temporal case the state could be an observation at the same particular spatial location at point in time t (Gamerman, 2010). Spatio-temporal DLMs are also highly flexible and adaptable to fit the needs of the data, while running the gamut in terms of complexity. For example, Stroud, Müller, and Sansó (2001) proposed a more complex linear model, assigning *local-mixture weights*, given by Gaussian and non-Gaussian *weighting kernels* (i.e. functions that assign a weight to a given data point based on its distance from a pre-determined point or parameter, such as the mean), to the spatial mean function to allow for inclusion of prior information about these spatial functions. The Stroud, Müller, and Sansó (2001) model is given by:

$$\mathbf{Y_t} = \mathbf{X_t}\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(\mathbf{0}, \mathbf{V_t}), \tag{1.7}$$

$$\beta_t = \mathbf{G}_t\beta_{t-1} + \omega_t, \qquad \omega_t \sim N(\mathbf{0}, \mathbf{W_t}),$$

where:

-$\mathbf{Y_t} = (Y(\mathbf{x}_{t1}), \dots, Y(\mathbf{x}_{tn}))'$ are observations,

-$\mathbf{X_t} = [\mathbf{f}_j(\mathbf{x}_1), \dots, \mathbf{f}_j(\mathbf{x}_n)]'$ is a time-dependent design matrix,

-$\boldsymbol{\epsilon}_t$ is a Gaussian noise process with variance-covariance matrix $\mathbf{V_t}$,

-$\beta_t = (\beta'_{t1}, \dots, \beta'_{tJ})'$ is a vector at *t*, for all *t* = 1, … , *T*,

-$\mathbf{G}_t$ is an evolution matrix, with specifications dependent on the number of temporal components, and

-$\boldsymbol{\omega_t}$ is a Gaussian noise process with variance-covariance matrix $\mathbf{W_t}$,

and the local-weighted mixture given by non-stationary spatial mean function $S(\mathbf{x};\beta_1), \dots, S(\mathbf{x};\beta_T)$ given by:

$$\sum_{j=1}^{J} \pi_j(\mathbf{x})\mathbf{f'}_{\boldsymbol{j}}(\mathbf{x})\beta_j\,, \tag{1.8}$$

where:

-J is the number of mixture components,

-$\pi_j(\mathbf{x})$ is the weighting kernel, centered at the mean of the kernel,

-$\mathbf{f}_{\boldsymbol{j}}$(x) is a vector of known functions, and

-β_j is a vector of random parameters

1.5.4 **Assumptions of Spatial and Spatio-temporal Analyses**

Before moving into the next section, we formally introduce two common assumptions made of data in spatial and spatio-temporal analyses (Cressie, 1991). The first is that the data is stationary, which, as described earlier, is where model parameters (mean and variance-covariance structures) are invariant across spatial locations (Cressie, 1991; Gneiting, 2002). However, models accounting for non-stationary data are becoming more common, though these models commonly assume a Gaussian process (normally distributed mean and variance-covariance parameters) (Damian, Sampson, & Guttorp, 2001; Fuentes, 2002). One such model for non-stationary data, developed by Gelfand, Kottas, and MacEachern (2005), utilizes a Bayesian nonparametric framework, meaning that the model accommodates mean and variance-covariance parameters that are not normally distributed. The Gelfand, Kottas, and MacEachern (2005) model allows for non-stationary data by incorporating *Dirichlet process* mixing, which means that the model generates a random distribution that allows for varying model parameters (mean and variance-covariance structures) across spatial locations.

The second is that the data demonstrates *isotropy*, or that it is "direction invariant" (Haining, 1990, p. 33). In other words, an assumption of isotropy means that relationships between spatial locations are the result of the distance between the locations, not the general direction of the locations in the larger spatial domain (Guan, Sherman, & Calvin, 2004). This intuitively makes sense, as one would expect to see more of a certain variable in spatial locations that are closer to each other.

In educational research, however, it is difficult to automatically assume that data meets these assumptions. Given the wide range of background variables at play with spatial locations, it is unlikely that most educational data is stationary. It is even more unlikely that the data is isotropic. For example, schools that are geographically close to each other would intuitively show similar levels of performance on the dependent variable(s) in consideration. However, these schools may be clustered in certain parts of the city, thus negating directional invariance. Therefore, we prefer spatial and spatio-temporal models that allow for non-stationary, anisotropic data. We focus on those in the following sections on *smoothing terms* and knot selection.

1.5.5 **Smoothing Terms**

As mentioned earlier, a data analysis issue may arise when a dataset used in educational research is large. This size issue is compounded when spatial locations are included and/or if observations are repeated at multiple time points. To elaborate on a point made earlier in this thesis, large datasets are problematic in the matrix inversions performed during Markov chain Monte Carlo (MCMC) iterations (Gelfand, Banerjee, & Finley, 2011; Hancock & Hutchinson, 2006). More specifically, matrix inversions occur on the order of n^3, where n is the number of data points. With large datasets, the matrix inversions become computationally inefficient and the parameter estimates become unstable. If extra dimensions are then added through spatial locations and observations across time, analysis becomes extremely cumbersome (Guhaniyogi, Finley, Banerjee, & Gelfand, 2011). As noted earlier in this chapter, this problem is exacerbated in Gaussian process models.

One solution is to use low-rank or reduced-rank models (Wood, 2003), wherein the information for a particular variable is reduced to a smaller set. One form of low-rank or reduced-rank modeling is the use of splines. The basic model for splines is given by:

$$Y_{\mathrm{i}} = f(\boldsymbol{x_i}) + \varepsilon_{\mathrm{i}} \tag{1.9}$$

where:

- Y_{i} is the dependent variable, observed at location *i*,
- *f* represents the smoothing function,
- $\boldsymbol{x_i}$ represents the predictors and covariates, including spatial and temporal data, at location *i*, and
- ε_{i} represents the measurement error (Nychka, 2000; Wahba, 1990)

Splines are essentially used to “smooth” the data that is included, with locations that may not be included or may be unobserved. Nychka (2000) recommended using splines as a bridge of sorts between kernel-based approaches, such as that employed by Stroud, Müller, and Sansó (2001), and the more classical, lattice-based approaches of kriging (i.e. spatial interpolation; Matheron, 1963).

As a derivation to this basic model, Nychka (2000) encouraged exploration of local smoothing, or smoothing functions that take into consideration that different areas of the spatial domain may require different smoothing functions (e.g. areas that are less densely populated). This concept of local smoothing recognizes growing research into non-stationary spatial processes, where model parameters are assumed to vary across locations. This concept of utilizing different smoothing functions for different spatial areas was supported by Paciorek and Schervish (2004).

The need for splines is clearly illustrated when we think about research involving monitoring rainfall amounts or plant growth across a region, where inferences about the entire region are drawn from data collected at observational sites that almost certainly do not cover the entire region. The connection with educational research may not be as readily apparent, as we think of schools and other "point" sites as being observed. However, in a dataset with many spatial locations, it may be computationally necessary to not include all locations.

Thin-plate splines, introduced in Duchon (1977), among others, are commonly cited in the literature as being applicable for spatial data due to their low-dimensional (*d*; typically *d* = 2) nature; this echoes the often two-dimensional nature of spatial data, especially when represented by latitude and longitude. We remind the reader that splines are smoothing *functions*, and therefore a thin-plate spline is a smoothing function that minimizes the following:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_1(i), \ldots, x_d(i)))^2 + \lambda \boldsymbol{J}_m^d(f) \tag{1.10}$$

where:

-*n* is the sample size,

-y_i is the dependent variable,

-f represents the smoothing function

-x_1 through x_D are the *d* predictor variables and covariates,

-λ is the smoothing parameter, and

-$\boldsymbol{J}_m^d(f)$ represents the penalty function (Hancock & Hutchinson, 2006; Wahba, 1990)

Thin-plate splines are particularly useful due to the inclusion of this penalty function, which penalizes for roughness or "wiggliness" of data, or where curvatures in the function lead to poor fit of the data. This penalization capitalizes on the smoothing naturally afforded by the linear transformations and combinations inherent in this class of splines to smooth out these curvatures (Nychka, 2000; Wood, 2003). In the case of two-dimensional data, such as spatial data, this penalty function is given by:

$$J_2(f) = \iint_{-\infty}^{\infty} (f_{x1x1}^2 + 2f_{x1x2}^2 + f_{x2x2}^2)\, dx_1\, dx_2 \text{ (Wahba, 1990)} \qquad (1.11)$$

Another advantage to using thin-plate splines, is that, unlike *cubic splines*, thin-plate splines are "rotationally invariant," a quality that is particularly appropriate for spatial data (Holmes and Mallick, 2003, p. 355). Finally, these thin-plate splines can be univariate or multivariate; use of multivariate splines alleviates the "piecewise" nature of univariate splines (Hancock & Hutchinson, 2006, p. 1685). Since the spatial data we are using in this research is typically represented by latitude and longitude, we could create interaction terms from univariate splines or use multivariate thin-plate splines. With multivariate splines, our smoothing spline function would adjust to *f*(x,y).

However, Wood (2003) countered that thin-plate splines are computationally inefficient due to the number of parameters estimated being equal to *n*. In addition, per the smoothing function, thin-plate splines are not particularly compatible with linear and generalized linear models. As an alternative, Wood (2003) proposed a class of splines called *thin-plate regression splines*, which are also well-suited for use with latitude and longitude data. Thin-plate regression splines differ from thin-plate splines by relying on *eigenvectors* and *eigenvalues* (essentially, characteristics of matrices and of algebraic transformations of matrices), rather than knots (i.e. a set of spatial locations, which could be all locations included in a dataset or a smaller subset of locations), to determine the flexibility of the model. This removes the issues of how to select knots and how to ensure they are relatively evenly spread across $\boldsymbol{S}$. Wood (2003) noted that thin-plate regression splines require less computational time than thin-plate splines, thus making them better suited to large datasets. This decrease in computational time is due to a reduction of parameters, particularly those not included in the "wiggliness" penalizing function. Thin-plate regression splines are also more compatible with linear and generalized linear models.

Another alternative method is low-rank *Penalized Basis (B) splines*, otherwise known as *P*-splines, which can also accommodate large datasets (Lee & Durbán, 2011). The authors develop this model to address a common lack of acknowledgement for the interaction of spatial and temporal variables in spatio-temporal smoothing models. The penalization, which creates the smoothness of the spline, also addresses the important conceptual issue of anisotropy across spatial locations. Lee and Durbán (2011) do this by developing a *P*-spline based model that allows for smoothing to vary for latitude, longitude, and time. The model follows the general form:

$$\hat{\boldsymbol{y}} = f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) \tag{1.12}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are latitude and longitude, respectively, and $\boldsymbol{x}_t$ is time. Smoothing is provided in general by the use of a penalty term, $\boldsymbol{P}$. The model can be rewritten in a more familiar analysis of variance (ANOVA) form:

$$\hat{\boldsymbol{y}} = f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_3(\boldsymbol{x}_3) + f_{1,2}(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_{1,2,3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) \tag{1.13}$$

where time is now represented by $\boldsymbol{x}_3$ and f_1, f_2, and f_3 are the respective smoothing functions for latitude, longitude, and time. Per the ANOVA model, analysis can be performed on the main effects of latitude, longitude, and time independently, as well as the two-dimensional interaction of latitude and longitude and the three-dimensional interaction of latitude, longitude, and time.

The model also provides the flexibility to reduce the dimensions in the analysis, e.g. to look at all three main effects and the interaction of latitude and longitude while not exploring the interaction of latitude, longitude, and time.

1.5.6 **Knots and Knot Selection**

In low-rank or reduced-rank spatial models (Crainiceanu, Diggle, & Rowlingson, 2008; Stein, 2007; Wikle & Cressie, 1999) the information from all spatial locations is assumed to be captured in a smaller set, either of actual locations or of values derived from the actual locations. This smaller set of variables is more commonly called "knots," which we will cover in this section. As previously mentioned, knots are fixed, in the sense that the locations of the knots remain stationary during the analysis, and non-grid based representations of all spatial locations in $\boldsymbol{S}$ (Eidsvik, Finley, Banerjee, & Rue, 2012). We start with a note that individual knots are frequently denoted by s^*_j, as part of collection of $\boldsymbol{S}^*$ total knots (Gelfand, Banerjee, & Finley, 2011).

A common issue with knot selection is determining the necessary value of $\boldsymbol{S}^*$. Wood (2003) noted that one must take care with knot selection to ensure that selection does not result in completely uneven knot coverage across the spatial domain, whether that selection is fixed prior to running the analysis or automatically done through the analysis. In terms of methods that involve fixing the number of knots prior to the analysis, one suggested method is to have a 1:1 ratio of knots to spatial locations (Holmes & Mallick, 1998; Ramsay & Silverman, 2005). As can be imagined, however, this is not feasible for large datasets. Another suggested method, based on kriging (Matheron, 1963), is to use lattices, or evenly spaced knots on a n x n grid (Cressie, 1991), with potential modifications such as constructing this grid and then randomly placing new knots near existing knots (Diggle & Lophaven, 2006). A solid overview of lattice and modified-lattice approaches is available in Gelfand, Banerjee, and Finley (2011).

Another method is hierarchical Gaussian predictive process models, introduced in Banerjee, Gelfand, Finley, and Sang (2008) and updated in Finley, Sang, Banerjee, and Gelfand (2009) and Banerjee and Fuentes (2011), among others. Hierarchical Gaussian predictive process models, which are meant specifically to address large datasets, utilize a fixed set of knots that may either be drawn directly from ***S*** or may be representative of actual spatial locations. These models, which can be used with both Gaussian and non-Gaussian data, allow for non-stationarity of parameters through the predictive process parameter, which is derived through a linear transformation across all knots. This method also avoids projecting the spatial location data onto a grid, as commonly occurs with unevenly spaced spatial location data (Finley, Sang, Banerjee, & Gelfand, 2009).

In terms of methods that allow for automatic selection through the analysis, one suggested method utilizes stochastic search (i.e. selection based on probability), as done in modified form in Xia, Miranda, and Gelfand (2006). This stochastic search process could also be updated through use of stochastic search variable selection (SSVS), as demonstrated by George and McCulloch (1997). Likewise, Guhaniyogi, Finley, Banerjee, and Gelfand (2011) updated the hierarchical Gaussian predictive process models to allow for “adaptive knots,” which involves an initially fixed number of knots and a subsequent “averaging” across these knots. Through this method, knots are not fixed initially but instead allowed to move on the x-axis during the analysis, essentially allowing the data to tell the knots where they are best suited.

With continuous independent variables, Smith and Kohn (1997) recommended sorting the variable data and placing a knot at every *n*th value to ensure that the knots capture the local flavor of the different areas in the spatial domain. However, as can be imagined knot selection is particularly difficult with bivariate variables such as latitude and longitude. One possible approach is to use clustering in knot selection and placement (Banerjee and Fuentes, 2011; Tibshirani, Walther, and Hastie, 2001). With bivariate variables, Kohn, Smith, and Chan (2001) suggested forming clusters of the bivariate variables and using the cluster centroids, or centers, as the knot locations.

1.5.7 **Shrinkage Estimation and Ridge Regression**

Shrinkage estimation is, essentially, a technique for improving model fit with a focus on the covariance matrix of the model. Shrinkage estimation is particularly useful in cases where the covariance matrix is likely to be unstable, such as when model fit decreases when the model is applied to new datasets and when the sample size of the dataset is less than the number of predictors (Kwan, 2011). In shrinkage estimation, we "shrink" the regression coefficients, or beta (β), or other model parameters toward a pre-defined value. Possible pre-defined values can include zero as well as the mean value of each coefficient or parameter, since this value represents a best estimate of sorts as to the true parameter value (Hoff, 2013).

One such shrinkage estimation methodology is ridge regression. Hoerl (1962) suggested the use of ridge estimation as a means to account for this lack of covariance matrix stability sometimes seen in least squares estimation, a common regression methodology for fitting observed data to a proposed regression line (in the case of linear regression). In a 1970 article, Hoerl and Kennard introduced the concept of ridge regression, wherein a shrinkage parameter, referred to as the *ridge parameter*, or λ, is used to shrink estimates of the regression, or beta (β), coefficients toward zero. When the aforementioned pre-defined value is set to zero, ridge regression aligns more with the Bayesian framework ("Ridge Regression," 2015). In ridge regression, this shrinkage parameter may be estimated using the *Bayesian information criterion* (BIC) (Fan & Tang, 2013), though Karabatsos (2015a) suggested maximization of the *marginal likelihood* instead to boost computational efficiency. Karabatsos (2015a) demonstrated this efficiency through analysis of datasets ranging from three to over 15,000 predictors. The computation time for the posterior estimates ranged from two-thousandths of a second in the case of p = 3 to roughly two minutes for p = 15,154.

Ridge regression addresses three main concerns involving model variables: highly correlated predictors, classification of the dependent variable(s), and selection of predictors. In terms of highly correlated predictors, Hoerl and Kennard (1970) extended the application of ridge estimation to regression in the case of datasets with non-orthogonal predictors, or predictors that are not independent of each other. The property of ridge regression to shrink estimates of the β coefficients toward zero assists with correcting for highly correlated predictors, or *multicollinearity*. This correction of multicollinearity also helps to reduce variance (Myers, 1990). This ability to handle datasets that feature highly correlated predictors is important in educational research, since educational datasets can be large and contain many predictor variables. This issue is also particularly noteworthy with spatial models, where spatial autocorrelation (Moran, 1950; Cliff & Ord, 1973) may lead to highly correlated predictors between nearby spatial locations.

In terms of classification of the dependent variable(s), it should be noted that ridge regression is useful in linear regression, most commonly with continuous dependent variables. Karabatsos (2015a), however, argued that application of ridge regression can be extended to models with binary dependent variables; that is, variables that are limited to two values, often “present” or “not present,” which are typically coded as “0” or “1” in a dataset. Drawing from the school closings research, an example of a binary dependent variable in an educational dataset is whether the school was closed (where “0” represents “not closed” and “1” represents “closed”).

Finally, ridge regression is also effective in use with datasets with large numbers of predictors due to ridge regression allowing for automatic variable selection, by use of the *scaled neighborhood criterion* (SN) or the *interquartile* (50%) or *95% posterior credible intervals* of the slope coefficients of the model (Li & Lin, 2010). Per the SN, significant predictors are those with a marginal posterior probability of less than 0.5. Per use of the 50% or 95% posterior credible intervals of the slope coefficients of the model, non-significant predictors are those with an interval of values close to or including zero. Both methods are welcome news for use with educational datasets which, as previously mentioned, can contain large numbers of predictors. The issue of large numbers of predictors is also of concern in the case of spatio-temporal models, where variables are measured repeatedly over time. In the case of spatial and spatio-temporal models, this automatic variable selection can also be extended to automatic selection of knots for the splines.

In the final section of this chapter, we present an argument for the need for an expansion of spatio-temporal analysis in educational research. We present advantages to including spatial and temporal variables, while also noting particular challenges with implementation of spatio-temporal models. We also lay the groundwork for presentation of our linear, ridge regression-based spatio-temporal model, which we believe assists with alleviation of the noted challenges to implementation of other spatio-temporal models. At the conclusion of this chapter, we present a framework for the remaining chapters of this thesis.

1.6 **Problem Statement Revisited**

As mentioned previously, there are clear advantages in incorporating spatial and temporal variables into educational research. These advantages fall into three primary domains: statistically, methodologically, and ease of application. In terms of statistical advantage, one could argue that all data inherently have some spatial and temporal aspects to them, and excluding space and time-related variables can lead to false conclusions about the nature of the relationships between predictors and the dependent variable(s) (Cressie, 1991). This argument is supported by the research of Savitz and Raudenbush (2009), who reported decreased expected mean squared errors of neighborhood means in models that incorporated spatial dependence in comparison to other models that did not account for spatial dependence.

In terms of methodological advantage, accounting for spatial and temporal variables allows for greater consideration of context in educational research, where here “context” means latent, or not directly observable, factors that may also influence our findings. In terms of space this context might include factors such as locations of schools, characteristics of the neighborhoods these schools are in, and possible patterns of findings between schools that are spatially near to each other (Alperin, 2008). We ground this need for spatial context in the work of Wilson (1987) on neighborhood effects and the work done by researchers such as Bronfenbrenner (1976) and Berliner (2006) to push consideration of the influence of the student’s home environment on educational performance, among others. In terms of time this context primarily refers to the ability to track results longitudinally, thus allowing the researcher to look for consistency or inconsistency of results over time (Haining, 1990) and ultimately strengthening the design of the study (Goodchild and Janelle, 2004).

Finally, in terms of an advantage in the ease of application, a fair amount of educational research focuses on locations such as schools as units of analyses. A school can be represented spatially using latitude and longitude coordinate points, which lends to more straightforward spatial and spatio-temporal analyses over more broad locations such as neighborhoods and parks. As mentioned earlier, however, there is even an advantage when using broader areas such as neighborhoods as the spatial locations of interest, as these areas are less likely to change much over time in terms of location and geographical boundaries. It is certainly easier to conceptualize spatial analysis utilizing neighborhoods as spatial locations rather than analysis using difficult to predict locations, such as hurricane trajectories.

However, there are also challenges that serve as obstacles to using spatio-temporal analyses in educational research. These challenges fall into three primary domains: complications of large datasets, complex statistical models, and model selection. In terms of the challenge of large datasets, we again note that educational datasets can be large, with large numbers of cases (subjects) and predictors. This issue is compounded when one includes spatial and temporal variables, where each time period means another set of data points. We recall that parameter estimation in large datasets can be computationally intensive due to the matrix inversions performed during MCMC iterations (Gelfand, Banerjee, & Finley, 2011; Hancock & Hutchinson, 2006), and that this issue is worse with Gaussian models due to the n x n covariance matrix. The challenge of working with large datasets is daunting enough, especially given a tendency in educational research to prefer models utilizing the normal distribution. However, it becomes intimidating to think of then adding spatial location data, especially if data is collected over time, because each new temporal observation multiplies the number of data points.

The second challenge, complex statistical models, refers both to the level of skill and knowledge necessary to perform these analyses as well as ease of interpretation of the findings. As Stroud, Müller, and Sansó (2001) noted, spatial and spatio-temporal models tend to be very complex, often assuming a level of statistical and mathematical training far beyond that obtained by many educational researchers. As such, these models are conceptually difficult, in terms of understanding how the models function, how to implement the analyses, and how to interpret the results. The difficulty in implementing the analyses is compounded by the frequent need to utilize coding-based statistical programs, such as *R*, which can entail the need to write or to adapt programming code. Many educational graduate programs emphasize more standard statistical programs that utilize "point and click" menu interfaces, such as Statistical Package for the Social Sciences (SPSS), leaving educational researchers with little to no formal training on more advanced programs.

The final challenge, model selection, is particularly pronounced in large datasets due to the large number of variables of potential interest for the model. Earlier in this chapter we provided an example of a study where a researcher is interested in looking at the relationship between crime rates and the amount of green space in a neighborhood. To control for possible confounding variables, a large number of background variables, such as poverty rate and educational attainment, must be included. However, given that there may be high correlations among the variables between nearby neighborhoods, the researcher should also include a number of spatially relevant variables. Given the large number of variables, it could take a great deal of time, whether on the part of a computer or the researcher, to determine the best combination of variables to model the relationship between crime and green space. We acknowledge that there are many methods for selecting which variables to include in a model. As noted, some of these methods can be more manual, and can involve a trial and error approach to fitting the model. This approach would be particularly cumbersome and time-consuming with a large dataset, such as that in the crime rate example. Given the propensity for spatio-temporal datasets to contain large numbers of variables, we suggest the employment of methods that allow for automatic variable selection and believe that our model fits this criterion. We also suggest the use of multiple model fit indices in order to determine the best combination of variables to include in our model.

1.7 **An Overview of the Rest of the Thesis**

In Chapter 2 we will present our Bayesian linear spatio-temporal model, developed for use with educational datasets, particularly the type of datasets developed by schools for annual reporting purposes (e.g. containing information about student demographics and student performance, among other variables). The model is also developed for use with locations that can be characterized using latitude and longitude, such as schools. We opt to use the Bayesian statistical framework, in part because of the utility of Bayesian statistics with complex statistical analyses, which we argue include spatio-temporal analyses (Cowles, Kass, & O’Hagan, 2009). Our linear model is an extension of the low-rank *P*-spline model introduced in Section 1.5.5. Our model could also be considered a type of a dynamic linear model (DLM) (introduced in Section 1.5.3) because it allows for model parameters to change over time and across spatial locations. We capture the temporal elements through inclusion of a smoothing function *f* of autoregressive terms. Because we utilize a discrete spatial process model (introduced in Section 1.5.1), we focus on “point-based” spatial data, or spatial data represented by latitude and longitude (Banerjee & Fuentes, 2011); latitude and longitude are each represented by a smoothing function *f*. Other relevant predictors, such as test scores, are included and represented with smoothing function $f_z(\mathbf{z})$. Through our linear model, we can identify the independent effects of each predictor (space, time, and other relevant predictors) as well as model the interactions of the predictors. Finally, we use ridge regression, introduced in Section 1.6, to estimate model parameters, including the coefficients of our spline (smoothing) functions.

In Chapter 3 we will demonstrate the utility of our model on two educational datasets, which we describe in Chapter 2. We do this by focusing on computational efficiency, or the time required to complete the analyses, and also the ease of implementing the model. Finally, we analyze these datasets both with and without spatial and temporal variables, in order to determine whether model fit is improved through inclusion of these variables.

In Chapter 3 we will also make a case for inclusion of spatio-temporal analysis in educational research. We do so by framing use of our Bayesian model as a tool for educational research, as well as by presenting examples of educational research questions that lend to spatio-temporal analysis. Part of this framework will include a guide (in Appendix B) on how to run the analyses in *Bayesian Ridge Regression* software, developed by Dr. George Karabatsos (2016). It is our hope that this thesis will inspire readers who may be new to spatio-temporal analysis and/or Bayesian modeling to incorporate both into their research and data analysis. We also present considerations to be mindful of when conducting spatial and spatio-temporal research, including taking caution against conflating correlation and causation.

Finally, in Chapter 4 we will conclude with an assessment of whether our model addresses the three noted challenges to using spatio-temporal analysis in educational research (complications of large datasets, complex statistical models, and model selection). We also discuss any limitations to our model and possible next steps for future research.

2. METHODOLOGY

2.1 Presentation of the Bayesian Spatio-Temporal Model

As mentioned in Chapter 1, one challenge to incorporating spatio-temporal models into educational research is the complexity of many spatio-temporal models. Given that most educational researchers are likely familiar with linear models, we propose a dynamic linear model (DLM) that adapts the spline-based analysis of variance (ANOVA) model developed by Lee and Durbán (2011). The model follows the general form:

$$y = f_{st}(s_1, s_2, \mathbf{x}_t) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2) \tag{2.1}$$

where s_1 and s_2 are latitude and longitude, $\mathbf{x}_t$ represents a vector of predictors at time t, and f_{st} is the smoothing function for spatio-temporal interaction. We represent the joint modeling of latitude (s_1) and longitude (s_2) via the vector $\mathbf{s}$, as reflected in our expanded model (Equation 2.2). The use of this ANOVA-based model allows researchers to capture the interaction between the spatial and temporal variables. In addition, Lee and Durbán (2011) noted that this model type, along with the use of smoothing functions, addresses the issue of anisotropy across spatial locations by allowing for smoothing to vary for latitude, longitude, and time. In addition, the use of smoothing functions accounts for any predictor variables that have a non-linear relationship with the dependent variable (Karabatsos, 2015b).

The model can be rewritten in a more familiar ANOVA form. We update the model developed by Lee and Durbán (2011) (Equation 1.13) to include additional predictors and a specified error term:

$$y = f_t(\boldsymbol{x}_t) + f_z(\mathbf{z}) + f_s(\mathbf{s}) + f_{st}(\mathbf{s}, \boldsymbol{x}_t) + f_{stz}(\mathbf{s}, \boldsymbol{x}_t, \mathbf{z}) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2) \tag{2.2}$$

where f_t, f_s, and f_{st} are the respective smoothing functions for time, space, and the spatio-temporal interaction. The parameter $\boldsymbol{z}$ represents a vector of non-spatial and non-temporal predictors, and therefore $f_z(\boldsymbol{z})$ is the respective smoothing function. We model the interaction of the spatio-temporal variables and other predictors in the parameter $f_{stz}(\mathbf{s}, \boldsymbol{x}_t, \boldsymbol{z})$. The error term, ε, is assumed to be normally distributed, with a mean of zero and variance σ^2. As previously noted, the use of an ANOVA-based model allows for analysis on the main effects of time and non-spatio-temporal predictors, the two-dimensional interaction of latitude and longitude, the three-dimensional interaction of latitude, longitude, and time, and a final interaction between latitude, longitude, time, and additional model predictors.

Since we use a linear model, we include splines as a way to account for any non-linear relationships between predictor variables and the dependent variable. The use of splines also transforms these non-linear relationships into the linear relationships expected in a linear model. We elect to use splines over the perhaps more familiar polynomial functions (i.e. $\beta_1 x + \beta_2 x^2 + \cdots + \beta_q x^q$) because the use of local knots in the spline terms provides greater stability and less sensitivity to outliers than does the polynomial function (Magee, 1998).

Based on the work of Nychka (2000) and Wood (2003) in support of thin-plate splines with spatial data, we will use thin-plate splines as the smoothing functions for our ANOVA-based model. As noted, thin-plate splines can be univariate or multivariate, with multivariate splines alleviating the "piecewise" nature of univariate splines (Hancock & Hutchinson, 2006, p. 1685). We will utilize univariate splines in this research; however, due to the bivariate nature of spatial data represented by latitude and longitude, we will create interaction terms from the univariate latitude and longitude thin-plate splines.

This use of thin-plate splines is, in part, due to the accommodation for the effect of each observed covariate value, including spatial location, which helps to prevent poor fit of the data. The use of thin-plate splines also allows for transformation of each linear regression coefficient,

β_k, into a univariate or multivariate spline function, noted by f_k. Per Green and Silverman (1994), this univariate thin-plate spline function is given by:

$$f_k(x_k) = \beta_{k0}x_k + \sum_{l=1}^{L_k} \beta_{kl} \{(x_k - t_l)^2 \log|x_k - t_l|\} \quad (2.3)$$

where t_l represents a knot of the spline and $\{(x_k - t_l)^2 \log|x_k - t_l|\}$ represents a term that "transforms" non-linear relationships between the predictor variables and the dependent variable into linear relationships. Since we will use univariate thin-plate splines for the temporal vector $\mathbf{x}_t$ and the predictor vector $\mathbf{z}$, we update the function given in Equation 2.3 for each parameter:

$$f_{x_t}(\boldsymbol{x_t}) = \beta_{x_t}x_t + \sum_{l=1}^{L_k} \beta_{\boldsymbol{x}_t l} \{(\boldsymbol{x}_t - t_l)^2 \log|\boldsymbol{x}_t - t_l|\} \quad (2.4)$$

$$f_z(\boldsymbol{z}) = \beta_{\boldsymbol{z}}\mathbf{z} + \sum_{l=1}^{L_k} \beta_{\boldsymbol{z}l} \{(\boldsymbol{z} - t_l)^2 \log|\boldsymbol{z} - t_l|\} \quad (2.5)$$

One option for smoothing the spatial terms is to use a bivariate thin-plate spline function. The bivariate thin-plate spline function is given by:

$$f_s(\boldsymbol{s}) = \sum_{l=1}^{L_k} \beta_{\boldsymbol{s}l} \{|\boldsymbol{s} - \boldsymbol{t}_l|^2 \log|\boldsymbol{s} - \boldsymbol{t}_l|\} \quad (2.6)$$

where **s** again represents a vector of latitude and longitude, and $|\boldsymbol{s} - \boldsymbol{t}_l|$ represents the Euclidean distance between **s** and the $\boldsymbol{t}_l$ knot points. Another option is to create an interaction term from the univariate latitude and longitude splines. Therefore, it follows that the interaction terms in our model will represent the interactions of the univariate splines, as well as the possible interaction of the univariate splines with the one bivariate spline. In a dataset with many predictors, especially with these predictors observed at multiple points in time, the large number of spline terms may result in multicollinearity. We will further discuss this issue later in this section.

Given that linear models can be represented by $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$, we can represent all of the thin-plate spline function terms, f_k, by a super vector $\boldsymbol{\beta}$ and all of the predictor variables by a super vector $\mathbf{x}_i^T$. This super vector $\mathbf{x}_i^T$ also includes the $\{(x_k - t_l)^2 \log|x_k - t_l|\}$ terms of the univariate thin-plate spline functions and the $\{|\boldsymbol{s} - \boldsymbol{t}_l|^2 \log|\boldsymbol{s} - \boldsymbol{t}_l|\}$ term of the bivariate thin-plate spline function. The error term, ε_i, is assumed to be normally distributed, with a mean of zero and variance σ^2. Therefore, our model can be seen as a very large version of a traditional linear regression model. In addition, since our model uses a Bayesian statistical framework, the linear model can be rewritten as $\mathrm{n}(y_i|\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2)$. Per Bayes' theorem, a *prior probability distribution* is assigned for a given model parameter, such as $\boldsymbol{\beta}$. This prior probability distribution is updated with the observed dataset, resulting in a *posterior probability distribution*. In Bayesian statistics, all statistical inference is based on this resulting posterior distribution. For our Bayesian linear model, the prior distribution for our regression parameters, $\boldsymbol{\beta}$, is given by $\boldsymbol{\beta}|\sigma^2 \sim \mathrm{n}_p\left(0, \sigma^2\lambda^{-1}\mathbf{I}_{\boldsymbol{p}}\right)$ and the prior distribution for σ^2 is given by $\sigma^2 \sim \mathrm{IG}(a, b)$ (where IG represents the *inverse-gamma* distribution). Taken jointly, the prior distribution $\pi(\boldsymbol{\beta}, \sigma^2)$ becomes $\mathrm{n}_p\left(\boldsymbol{\beta}|\mathbf{0}, \sigma^2\lambda^{-1}\mathbf{I}_{\boldsymbol{p}}\right)\mathrm{ig}(\sigma^2|a, b)$ or, combined further as *normal-inverse gamma*, $\mathrm{nig}(\boldsymbol{\beta}, \sigma^2|\mathbf{0}, \lambda^{-1}\mathbf{I}_{\boldsymbol{p}}, a, b)$. Finally, for the purpose of statistical inference, we obtain the posterior parameter values through the following:

$$\pi(\boldsymbol{\beta}, \sigma^2|\mathcal{D}_n) = \frac{\prod_{i=1}^{n} n\left(y_i\middle|\mathbf{X}\boldsymbol{\beta}, \sigma^2\right)\mathrm{nig}\left(\boldsymbol{\beta}, \sigma^2\middle|\mathbf{m}, \lambda^{-1}\mathbf{I}_{\boldsymbol{p}}, a, b\right)}{\iint \prod_{i=1}^{n} \mathrm{n}\left(\mathbf{y}\middle|\mathbf{X}\boldsymbol{\beta}, \sigma^2\right)\mathrm{nig}\left(\boldsymbol{\beta}, \sigma^2\middle|\mathbf{m}, \lambda^{-1}\mathbf{I}_{\boldsymbol{p}}, a, b\right)\mathrm{d}\boldsymbol{\beta}\mathrm{d}\sigma^2} \tag{2.7}$$

The marginal mean of $\boldsymbol{\beta}$ in the above posterior distribution is given by:

$$\bar{\beta}_\lambda = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}_p\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y} \tag{2.8}$$

where $\bar{\beta}_\lambda$ is the ridge estimator used in ridge regression, further discussed later in this section, and $\left(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}_p\right)$ represents the posterior covariance matrix of $\boldsymbol{\beta}$ (Karabatsos, 2015a).

As mentioned earlier, a consequence of having a large number of spline terms is an increased risk of multicollinearity. Under ordinary least-squares regression (OLS), this multicollinearity would result in a non-positive definite matrix of $\hat{\beta}$ variables. These OLS estimates are thus poorly defined and unstable, with very large errors that approach infinity. A perfect correlation between predictors would create a singular design matrix, and therefore a non-existent OLS estimate of $\hat{\beta}$ (Karabatsos, 2015a). In order to account for this multicollinearity, we elect to use the *regularization method* (Yeomans, 2015) of shrinkage estimation. In shrinkage estimation, the regression coefficients or other model parameters "shrink" toward a pre-defined value; in the case of the particular shrinkage estimation technique we will employ, ridge regression, a prior distribution involving a penalty parameter is used to "shrink" insignificant parameters toward zero, though without equaling zero (Karabatsos, 2015a).

We select ridge regression over other shrinkage estimation techniques because of its utility with the following: highly correlated predictors (multicollinearity), linear regression models, and selection of predictors. In addition, Nychka (2000) supports the use of ridge regression with spatial data, especially alongside thin-plate splines. Highly correlated predictors are of particular concern in the case of datasets with spatial variables, where nearby locations can have similar values when compared to more distant locations (Tobler's first law of geography). This concern is compounded when observations are repeated across time, as in spatio-temporal datasets. As noted, if this multicollinearity is not accounted for, estimation of the β parameters can be inaccurate due to inflated mean-squared error (MSE). Ridge regression addresses multicollinearity through use of a ridge or shrinkage parameter, λ. In general, larger values of λ result in increased shrinkage of the estimated regression coefficients. In addition, when $\lambda = 0$, the ridge regression estimate of the regression coefficients is equal to the OLS estimate, as OLS is a special case of ridge regression where $\lambda \to 0$ (Karabatsos, 2015a)

We will apply the Bayesian Ridge Regression (RR) model using marginal maximum likelihood (MML) to estimate the ridge parameter λ, as suggested by Karabatsos (2015a). The use of MML allows for faster computation of the ridge parameter λ and faster parameter estimation through use of a *singular value decomposition* (s.v.d) of **X**. Determination of the MML ridge estimate $\hat{\lambda}$ will be accomplished by finding the positive-valued estimate that maximizes the following equation:

$$\log\pi(\mathcal{D}_n|\boldsymbol{\lambda}) = \log|\overline{\mathbf{V}}_{\boldsymbol{\lambda}}|^{\frac{1}{2}} - \log|\mathbf{V}_{\boldsymbol{\lambda}}|^{\frac{1}{2}} + a\log b - \bar{a}\log\bar{b}_{\lambda} + \log\Gamma(\bar{a}) - \log\Gamma(a) - \frac{n}{2}\log\pi$$

$$= \frac{q}{2}\log\lambda - \frac{1}{2}\sum_{k=1}^{q}\log(\lambda + d_k^2) + a\log b - \bar{a}\log\left\{b + \frac{1}{2}(\mathbf{y}^{\mathbf{T}}\mathbf{y} - \sum_{\boldsymbol{k=1}}^{\boldsymbol{q}}\frac{\hat{\alpha}_k^2 d_k^4}{\lambda + d_k^2})\right\} +$$

$$\log\Gamma(\bar{a}) - \log\Gamma(a) - \frac{n}{2}\log\pi \qquad (2.9)$$

which simplifies to:

$$q\log\lambda - \sum_{k=1}^{q}\log(\lambda + d_k^2) - n\log\left(\mathbf{y}^{\mathrm{T}}\mathbf{y} - \sum_{k=1}^{q}\frac{\hat{\alpha}_k^2 d_k^4}{\lambda + d_k^2}\right) \qquad (2.10)$$

Because the purpose of this paper is to develop linear models that are easier to understand and to utilize, we assign a normal inverse gamma prior [nig(β, $\sigma^2|\mathbf{0}$, $\lambda^{-1}\mathbf{I}_p$, a, b)] to (β, σ^2), which aligns the Bayesian RR model with the Bayesian normal linear model. Once the s.v.d, including *orthogonal* matrices **U** and **W**, is incorporated, the prior distribution for the normal inverse gamma prior updates to nig(β, $\sigma^2|\mathbf{0}$, $\mathbf{V}_{\lambda}$, a, b). As mentioned earlier, inclusion of the s.v.d allows for faster computational speeds due to its ability to avoid matrix inversion. Karabatsos (2015a) demonstrated the faster computational time in determining the ridge estimate for the MML RR when compared to both of the other ridge regression models, as well as other models including *LASSO* and *Elastic Net* models, using datasets ranging from n (sample size) = 24 to 52,397 and p (number of predictors) = 3 to 15,154. The MML RR model consistently outperformed all other models, regardless of dataset size and number of predictors, including in cases where $p > n$.

In this study, we will analyze datasets, described in the next section, using the *Bayesian Ridge Regression* software (Karabatsos, 2016). Using our spatio-temporal model, we can accommodate predictor variables related to spatial location and time as well as non-spatio-temporal predictor variables of interest. It is worth noting that the temporal variables in the model can include autoregressive lag terms; that is, we can include observations from a single time point *t* or from a given number of time points prior to *t*. For example, if we wished to use trend data from a three year period, we would use *t*, *t*-1 (first lag term), and *t*-2 (second lag term). If we had not previously constructed these lag terms, we could do so using the "Modify Data Set" menu in the *Bayesian Ridge Regression* software (Karabatsos, 2016).

To begin the analysis, we will first select our model, in this case the "Fast Ridge Regression" model. After indicating our dependent variable, we will determine which predictor variable(s) to include in our model. We will then use the software to generate thin-plate spline terms for the included predictor variables; interaction terms can then be created from the univariate spline terms as needed.

In addition to applying the Bayesian Ridge Regression (RR) model using *marginal maximum likelihood* (MML) to estimate the ridge parameter λ, we will also use this model to obtain posterior parameter estimates. For most of the model parameters, interpretation of the *beta coefficients* (slope parameters) is similar to that done in ordinary least-squares (OLS) regression, including inclusion of both standardized and unstandardized coefficients. This includes the slope parameters for the constructed thin-plate spline terms, as we only interpret the linear regression term, $\beta_{k0}x_k$, from the function in Equation 2.3. The remainder of the function represents the control for non-linear effects and will not be immediately interpretable.

We will determine significant predictors using either the interquartile (50%) or 95% posterior credible interval. We determine the 95% posterior credible interval based on an α (alpha) value of 0.05 and credible interval (1 – α)% (Levy, 2012). In this case of the former, the posterior credible interval establishes 50% probability that the parameter value is included in a percentile range, here between the 25% and 75% percentiles. In the case of the latter, the interval establishes 95% probability that the parameter value is included in a percentile range, here between the 2.5% and 97.5% percentiles. A predictor is determined to be significant when the posterior credible interval does not include zero (Karabatsos, 2015a). Because we are focusing particularly on large datasets, we apply the central limit theorem to establish that as the sample size approaches infinity, the probability distribution becomes the normal distribution. Thus, with large datasets we can assume an underlying normal distribution, which makes statistical inference more familiar and accessible. Finally, due to the facility of selecting significant predictors; the ability of the model to accommodate large datasets and number of predictors; and the fast computational speeds, all spatial knots and knots used in the model smoothing function(s) (if they differ) will be included along with all other covariates. This contributes to ease of spatio-temporal analysis by alleviating the need to select appropriate knots prior to the statistical analysis.

We will also identify any outliers in the data using the *standardized residual*, which should be familiar to most educational researchers. The standardized residuals focus on the observed values of the dependent variable and are given by:

$$z_i = \frac{y_i - \mathrm{E}_n(Y_i|\mathrm{x}_i)}{\sqrt{\left(\mathrm{Var}_n\left|(Y_i|\mathrm{x}_i)\right.\right)}} \qquad (2.11)$$

where $\mathrm{E}_n(Y_i|\mathrm{x}_i)$ and $\mathrm{Var}_n|(Y_i|\mathrm{x}_i)$ are the predicted mean and the predicted variance of the dependent variable, respectively. Analyzing standardized residuals allows the researcher to compare the observed dependent variable values to those predicted by the regression model. Residual values greater than three or less than negative three indicate that the observed value may be an outlier (Karabatsos, 2015b). In this event that an observed value takes on a residual value greater than three or less than negative three, we will address the outlier(s) by adding predictors with "0" and "1" values into the model. Here, a predictor with a "1" value indicates the presence of an outlier and serves to transform the outlying data point into a non-outlier (Xiong & Joseph, 2013).

Finally, we will assess model fit using four methods: R^2, the D(m) statistic, the *Akaike information criterion* (AIC), and the Bayesian information criterion (BIC). The first method is more relevant when looking at a single statistical model, while the last three are more relevant for model selection between several models. The R^2 statistic, which should also be familiar to most educational researchers, represents the amount of variance in the dependent variable(s) that is accounted for by the predictor variable(s) in the model. A larger R^2 value represents a greater amount of variance accounted for by the predictor variable(s).

The D(m) statistic (Gelfand & Ghosh, 1998) uses a Bayesian framework and is more relevant when comparing two or more models. The D(m) statistic is given by:

$$D(\mathrm{m}) = \sum_{i=1}^{n}(y_i - \mathrm{E}_n[Y|x_i, m])^2 + \sum_{i=1}^{n} \mathrm{Var}_n[Y|\mathrm{x}_i, m] \qquad (2.11)$$

where the term $\sum_{i=1}^{n}(y_i - \mathrm{E}_n[Y|x_i, m])^2$ assesses the fit of the model to the data and the term $\sum_{i=1}^{n} \mathrm{Var}_n[Y|\mathrm{x}_i, m]$ promotes parsimony by penalizing for complexity. This penalty term characteristic makes this statistic well-suited for use with spatio-temporal models, as illustrated in Gelfand and Ghosh (1998). When ranking models, the model with the best fit to the data is that with the lowest D(m) value.

The AIC statistic (Akaike, 1973) is given by:

$$-2\log L + 2p \tag{2.12}$$

where -2log$L$ is the information criterion (given as the log-likelihood) and 2p is the penalty term, with p representing the number of parameters in the model. When ranking models, the model with the lowest AIC value is that considered to be closest to the true distribution, particularly for predicted samples from the distribution (Dziak, Coffman, Lanza, & Li, 2012).

The BIC statistic (Schwarz, 1978) is given by:

$$-2\log L + \log(n)\mathrm{p} \tag{2.13}$$

where -2log$L$ is again the information criterion, log(n)p is the penalty term, with p again representing the number of parameters in the model. When ranking models, the model with the lowest BIC value is that considered to be the true model (Dziak, Coffman, Lanza, & Li, 2012).

In the context of our spatio-temporal model, we rank models using different combinations of predictor variables, including spatial and temporal variables, to assess which produces the best model fit. For example, it may be that inclusion of spatial location only adds to model complexity, rather than the fit of the model to the data or to the ability of the model to best reflect the true distribution of the data. We utilize multiple fit criteria when ranking models due to the affordances and limitations afforded by each; for example, the BIC penalizes more for model complexity than the AIC, leading to a possibility of underfitting with BIC and overfitting with AIC. Therefore, using both AIC and BIC provides more information to consider when selecting the model.

2.2 **Description of the Datasets**

We will analyze two education-related datasets to both illustrate how our model functions, as well as to provide examples of spatio-temporal analysis in educational research. The first dataset (referred to as "CPS") contains data on 554 schools in the Chicago Public Schools (CPS) District across eight years. This set represents information aggregated from data available on the Illinois State Board of Education (ISBE) website. Since the CPS dataset contains many rows of data (18,506) and a large number of potential predictor variables (20 +), we can use this dataset to test our model's capacity to accurately estimate posterior parameters from large datasets. CPS (the district) includes both charter and public schools; however, the ISBE website, where we obtained information for this dataset, reports data for charter networks (e.g. Noble Street Charter) in aggregate (that is, results for all schools within the network are combined into a single value). For this reason, the schools in these charter networks are excluded from analysis.

"CPS" contains the following school-level variables for grades 3, 4, 5, 6, 7, 8, and 11: percentage of students meeting or exceeding standards on the Illinois Standards Achievement Test (ISAT) and Prairie State Achievement Examination (PSAE) Math and Reading subject tests and enrollment. The descriptive statistics for the ISAT/PSAE Math scores, ISAT/PSAE Reading scores, and enrollment are included in Table I. With respect to the inclusion of examination results, we opt to include the subject-level examinations rather than the composite scores in order to provide higher granularity in the analysis. The numbers for the ISAT/PSAE Reading and Math exams are the percentage of students meeting or exceeding set state standards on the ISAT or PSAE. We will use the Math and Reading examination results, rather than the Writing and Science scores, because they are more consistently measured across grades each year. We also include the 2011-2012 probation status for each school, which is determined in part by the number of students meeting or exceeding said set state standards on the ISAT or PSAE, as well as 2012-2013 space utilization status (whether a school was underutilized, overcrowded, or determined to be efficiently enrolled, per CPS standards) (Chicago Public Schools, 2011).

Table I

DESCRIPTIVE STATISTICS FOR THE CPS DATASET

Variable	Mean	Standard Deviation
ISAT/PSAE Reading	63.47	19.55
ISAT/PSAE Math	70.38	19.47
Enrollment	629.35	366.43

In addition, we added the latitude and longitude of each school as variables to capture the spatial component of the analysis. This spatial component allows for the potential to explore prediction of performance based on the location of the school, as well as the potential to explore possible patterns or clustering in the data based on distance between schools (e.g. we would expect to find similar performance levels in schools that are located near each other). The data also includes school year as a variable and includes the following eight academic years: 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012, and 2012-2013. All school-level variables are distinguished by year, allowing us to test the potential to capture temporal components of the model in analysis, including the temporal lag terms described in Section 2.1. This temporal component provides us with the flexibility to make predictions and to explore trends in performance and enrollment over time, whether across all schools, across all schools in a neighborhood, or for a single school. This temporal component also allows us to look for trends with schools that were closed by CPS for the 2013-2014 school year.

The second dataset (referred to as “CPS Demo”) contains a cross-section of data from the CPS dataset (n = 2,729, thus allowing us to use this dataset to test our model's capacity to accurately estimate posterior parameters from a smaller dataset). This dataset contains the following data from the CPS dataset: 2011-2012 probation status, 2012-2013 space utilization status, and spatial variables (latitude, longitude, and latitude by longitude). The CPS Demo dataset also contains enrollment numbers, Reading scores, Math scores, and school-level race/ethnicity data (from the *CPS 2011-2012 Racial/Ethnic Report*) for the 2011-2012 academic year. The *Racial/Ethnic Report* includes counts and percentages of students identifying in the following categories: White, African-American, Native American, Asian/Pacific Islander (retired), Hispanic, Multi-Racial, Asian, Hawaiian/Pacific Islander, and Not Available (Chicago Public Schools, 2016). In order to create a variable to represent the percentage of minority students in a particular school, we summed all of the categories except for "White" and divided by the total number of students in that school. The descriptive statistics for the ISAT/PSAE Math scores, ISAT/PSAE Reading scores, enrollment numbers, and the racial composition variable are presented in Table II.

Table II

DESCRIPTIVE STATISTICS FOR THE CPS DEMO DATASET

Variable	Mean	Standard Deviation
ISAT/PSAE Reading (2011-2012)	67.85	18.74
ISAT/PSAE Math (2011-2012)	76.10	17.56
Enrollment (2011-2012)	589.07	362.05
Percentage Minority (2011-2012)	91.13	16.70

2.3 <u>**Spatio-temporal Analysis for Educational Researchers**</u>

Finally, we will make a case for increased use of spatio-temporal analysis in educational research. We present our ANOVA-based spatio-temporal model as a tool for educational research that uses “point” data [location data characterized by using coordinates such as latitude and longitude (Banerjee & Fuentes, 2011)], such as school location, as well as a means for introducing educational researchers to research questions that lend well to spatio-temporal analysis. We conclude with considerations to be mindful of when including spatio-temporal variables, particularly spatial variables.

In Chapter 1, we introduced three challenges that serve as obstacles to using spatio-temporal analyses in educational research. These challenges are: complications of large datasets, complex statistical models, and model selection. In Chapter 3 of this thesis, we will demonstrate how this thesis research addresses these challenges. In particular, we will demonstrate this through analysis of the CPS and CPS Demo datasets; these analyses will also serve as case studies for how spatio-temporal analysis can be utilized in educational research. As a reflection of the wide range in size of datasets used in educational research, we are careful to include both a large dataset (CPS) and a smaller dataset (CPS Demo).

We address the complex statistical models challenge through use of an ANOVA-based model, since most educational researchers will be familiar with ANOVA models and analyses. We also address this challenge by using the *Bayesian Ridge Regression* software program (Karabatsos, 2016), which uses a menu-drive, point-and-click interface, to analyze our datasets. To illustrate use of our model in this software program, we will include a step-by-step instructional guide in Appendix B of this thesis.

We address the complications of large datasets and model selection challenges through analysis of these datasets. In terms of complications of large datasets, we will demonstrate that by using Bayesian Ridge Regression and marginal maximum likelihood (MML), we avoid the lengthy computational time that can be involved with the matrix inversions performed during Markov chain Monte Carlo (MCMC) sampling. In terms of model selection, we use either the interquartile (50%) or 95% posterior credible interval to determine significant variables. We will also provide posterior output to illustrate how easy it is to interpret the analysis results, both in terms of variable slope parameters and in terms of determining significant predictors. In addition, we will also use multiple model fit indices to determine the best combination of predictors to include in the model, based on different model fit criteria (e.g. parsimony).

As part of seeing this thesis as a toolkit for educational researchers looking to begin incorporating spatio-temporal analysis into their research, we close by presenting considerations to be mindful of when including spatial and temporal variables. In particular, we focus on the need to remember the distinction between correlation and causation, and how this distinction becomes particularly important in the context of spatio-temporal analysis. To illustrate this, consider the research by Radinsky and Waitoller (2013) on the CPS school closings. Radinsky and Waitoller discovered that African-American students were impacted by the school closings at much higher rates than other racial groups. In addition, the majority of the schools that closed were located in the West and South sides of Chicago. There are relationships between race, spatial location, and schools closing, but could a researcher attribute race and location as causes of the closings? We draw from Holland (2008) and Sampson (2008) in arguing that taking a more descriptive approach is preferable to establishing causality, particularly with sensitive variables such as race. The researcher needs to systematically consider the potential influence of other effects and interactions, particularly those that may not have been included in the model, when interpreting any relationships discovered through statistical analysis.

3. RESULTS

3.1 **An Overview of the Toolkit and Chapter**

We structure Chapter 3 around the components of our toolkit for incorporation of spatio-temporal analysis into educational research. In this overview, we present the three components of our toolkit and what each corresponding section will entail. In addition, we also refer back to the three challenges to using spatio-temporal analyses in educational research, which were introduced in Chapter 1: complications of large datasets, complex statistical models, and model selection. We then map which section of this chapter addresses which of the three challenges.

The first component of our toolkit is the Bayesian linear spatio-temporal model introduced in Chapter 2; we will provide a brief overview of the model to frame our subsequent analyses. This model, in part, addresses the challenge of complex statistical models since it is an ANOVA-based linear model, which will be familiar to most educational researchers. The second component of our toolkit, and therefore the second section of this chapter, is a demonstration of this model using an education-related dataset. We will provide an overview of the datasets (introduced in Chapter 2), including descriptive statistics, as well as the results of our statistical analyses using our Bayesian linear spatio-temporal model. In terms of our three challenges, the first and third challenges will be directly addressed through the results of our analyses. To show how our model addresses the issue of lengthy computational times, we will present the time needed to complete each analysis. Likewise, to address model selection issues, our results will include posterior output as well as output from multiple model fit indices: R^2, the D(m) statistic, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). To further address the challenge of complex statistical models, we will include an instructional guide for running our model in Appendix B.

The final component and section pertains to considerations to keep in mind when including spatio-temporal variables, particularly spatial variables, in analyses. An especially important consideration is that of taking care to not conflate causation and correlation. As established in the previous chapter, there is much to be mindful of when implying causation, particularly in terms of variables such as race and ethnicity (including racial composition of a school's student body, as well as racial composition of the area or neighborhood surrounding a school). We explore non-causal methods for providing spatial analysis and insight into relationships.

3.2 Overview of the Bayesian Spatio-Temporal Model

We introduced our thin-plate-spline-based Bayesian linear model in Section 2.1 of Chapter 2. In this section we offer an overview of the model in order to frame our subsequent analyses and reporting of results. Our model is notated as:

$$y = f_t(\boldsymbol{x}_t) + f_z(\mathbf{z}) + f_s(\mathbf{s}) + f_{st}(\mathbf{s}, \boldsymbol{x}_t) + f_{stz}(\mathbf{s}, \boldsymbol{x}_t, \mathbf{z}) + \varepsilon,\ \varepsilon \sim \mathrm{N}(0, \sigma^2) \quad (3.1)$$

We represent the joint modeling of latitude (s_1) and longitude (s_2) via the vector $\mathbf{s}$, the modeling of predictors at time t via the vector $\mathbf{x}_t$, and the modeling of non-spatial and non-temporal predictors via the vector $\boldsymbol{z}$. We model the interaction of the spatio-temporal variables in the parameter ($\mathbf{s}$, $\boldsymbol{x}_t$), and the interaction of the spatio-temporal variables and other predictors in the parameter ($\mathbf{s}$, $\boldsymbol{x}_t$, $\boldsymbol{z}$). The error term, ε, is assumed to be normally distributed, with a mean of zero and variance σ^2. Our model can also be written in the more general linear representation:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \text{ with } \varepsilon_i \sim \mathrm{N}(0, \sigma^2) \quad (3.2)$$

and the Bayesian linear representation:

$$\mathrm{n}(y_i | \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2) \quad (3.3)$$

where the super vector $\boldsymbol{\beta}$ represents the thin-plate spline function terms, f_k, and super vector $\mathbf{x}_i^T$ represents the predictor variables.

The smoothing (spline) functions are defined as the following:

f_t = vector of variables at time t

f_s = vector of spatial variables

f_z = vector of non-spatial and non-temporal variables

f_{st} = interaction of spatial and temporal variables

f_{stz} = interaction of spatial, temporal, and non-spatial, non-temporal variables

The use of spline terms transforms non-linear relationships between predictor variables and the dependent variable into linear relationships (Karabatsos, 2015b), with the aim of improving model fit. We use univariate thin-plate splines for the vector of temporal variables and the vector of non-spatial, non-temporal variables; we use the interactions of univariate splines to represent the spatial variables, as well as other interaction terms in our model.

As mentioned in Chapter 2, we employ the shrinkage estimation technique of ridge regression due to its facility with multicollinearity. Multicollinearity becomes a concern in this research due to the number of data points and predictors in the dataset, as well as the repeated observations over time. In addition, multicollinearity is a concern because of the large number of spline terms that can be created when applying our Bayesian linear model. The use of ordinary least-squares regression (OLS) is not recommended in the presence of multicollinearity due to unstable or non-existent parameter estimates, along with errors approaching infinity (Karabatsos, 2015a). To avoid this issue, we use a particular ridge regression technique known as Bayesian Ridge Regression (RR) using marginal maximum likelihood (MML) in our statistical analyses (Karabatsos, 2015a). The use of RR provides the added advantage of decreasing the computational time needed to complete the analyses.

3.3 **Demonstration of the Model**

In this section, we begin with an overview of our first dataset. We then demonstrate use of our thin-plate-spline-based Bayesian linear model through a series of statistical analyses, comparing the results from Bayesian Ridge Regression and Bayesian regression. We present the results of these analyses, with each analysis utilizing a different combination of predictor variables from the CPS dataset.

For this research, we will analyze data on schools in the city of Chicago, particularly the Chicago Public Schools (CPS) district. We offer more extensive detail on the datasets in Section 2.2 of Chapter 2, but provide an overview to frame the analyses we report on in the next section. The particular focus of our first dataset, which we refer to as CPS, is on the 2013-2014 school closings. CPS contains data on 554 schools, which were selected based on the following criteria:

1. The school must have been open during the 2012-2013 school year.
2. The school must not be a charter school [the Illinois State Board of Education (ISBE) website reports data for charter school networks in aggregate. Therefore, no charter school data is included].

The CPS dataset contains the following data on each of the 554 schools: percentage of students meeting or exceeding standards on the ISAT and PSAE Math and Reading subject tests (broken down by grade level), as well as enrollment. The subject test scores and enrollment numbers for each school are captured for the 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011, and 2011-2012 academic years. The descriptive statistics for the ISAT/PSAE Math scores, ISAT/PSAE Reading scores, and enrollment are presented in Table III.

Table III

DESCRIPTIVE STATISTICS FOR THE CPS DATASET

Variable	Mean	Standard Deviation
ISAT/PSAE Reading (across all years)	63.47	19.55
ISAT/PSAE Math (across all years)	70.38	19.47
Enrollment (across all years)	629.35	366.43

The CPS dataset also includes variables that are CPS policy-related; we include these variables due to their stated relevance to the school closings. The first variable, which we use as the dependent variable in our analyses, is whether the school was closed for the 2013-2014 academic year (coded “0” for “not closed” and “1” for “closed”). We also include whether the school was on probation during the 2011-2012 academic year (coded “0” for “not on probation” and “1” for “on probation”), as well as the space utilization status at the beginning of the 2012-2013 academic year (coded “0” for “efficient use” and “1” for “underutilized or overcrowded”) (Chicago Public Schools, 2011). The frequency data for each of the aforementioned variables is included in Table IV.

Table IV

FREQUENCIES FOR PROBATION STATUS, SPACE UTILIZATION STATUS, AND SCHOOL CLOSURES

School Status	Number of Schools
On probation during 2011-2012 academic year	245/554 (44.22%)
Either underutilized or overcrowded during 2012-2013 academic year	359/554 (64.80%)
Closed for 2013-2014 academic year	45/554 (8.12%)

Finally, since the schools represent finite, countable spatial data points, we incorporate a spatial element into our model through inclusion of the latitude and longitude points for each of the 554 schools. In addition, temporal elements are represented through the inclusion of ISAT/PSAE Math scores, ISAT/PSAE Reading scores, and enrollment numbers for seven academic years, as well as the use of six temporal lag terms (2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, and 2010-2011). We can model the spatio-temporal elements through creation of interaction terms between the spatial variable and a given temporal variable (e.g. the interaction of school location and enrollment numbers for a temporal lag period of one year).

3.3.1 **Baseline CPS Analysis (No Spatial Variables)**

We framed our first analysis around the following question: was school enrollment a significant predictor of which schools were closed for the 2013-2014 year? We focused on school enrollment due to the claim by the CPS district that space utilization (determined by enrollment) was the main criterion used to determine which schools to close. We used whether the school was closed for the 2013-2014 academic year (binary coded “0” for “not closed” and “1” for “closed”) as our dependent variable (DV). To reflect the CPS criterion, we included the space utilization status at the beginning of the 2012-2013 academic year (binary coded “0” for “efficient use” and “1” for “underutilized or overcrowded”) and enrollment numbers as predictor variables. We also included whether the school was on probation during the 2011-2012 academic year (binary coded “0” for “not on probation” and “1” for “on probation”), Reading scores, and Math scores as predictor variables in order to account for school-level academic performance.

Prior to running the analysis, we constructed univariate thin-plate splines for each of the continuous predictor variables (i.e. Reading scores, Math scores, and enrollment numbers). In constructing our thin-plate splines, we placed knots at 40 evenly spaced quantiles of the variable data. The use of 40 knots is suggested as a default value by Ruppert (2002), regardless of sample size (though particularly for datasets with $n > 150$). Given the size of our dataset ($n = 18{,}506$), we opted to use this default value.

We ran this analysis, which intentionally does not incorporate spatial location or temporal lags, first in part to orient the reader to our model and to use of the Bayesian Ridge Regression software (Karabatsos, 2016). In the remainder of this section, we present the results of our analysis. We include posterior estimates of our predictors; since $p = 121$, however, we will not include all posterior output in this section. The full posterior output, as generated by the Bayesian Ridge Regression software (Karabatsos, 2016), is included in Table XV, Appendix A, along with a guide for how to run the analysis in Appendix B. As indicated in Chapter 2, we include the following model fit statistics: R^2, the D(m) statistic, the Akaike information criterion

(AIC), and the Bayesian information criterion (BIC), as well as the number of outliers in the data. Finally, we also include the time needed to complete the computations.

The posterior estimates are provided in Table V. We include the 2011-2012 probation and 2012-2013 space use status variables, as well as selected Reading, Math, and enrollment knot estimates. Per our analysis, a school being on probation in 2011-2012 was a significant predictor of the school closing in 2013-2014, when controlling for all other variables in the analysis. We determined this because the PP1SD (the probability that the standardized beta coefficient falls within 1 standard deviation of 0) is less than .5 (.00) and the 50% posterior interval (shown in Table V between the 25% percentile and the 75% percentile) does not contain zero. Therefore, based on the PP1SD and the 50% posterior interval, we conclude that 2011-2012 probation status is a particularly strong predictor of a school closing. Similarly, a school being classified as “underutilized” or “overcrowded” during the 2012-2013 school year was a significant predictor of the school closing, again when controlling for all other variables [PP1SD = .00; 50% posterior interval (0.027, 0.030)]. Based on these values, space use status is also a particularly strong predictor of a school closing.

The percentage of students meeting or exceeding standards on the ISAT and PSAE Reading subject test was a significant predictor of a school closing in 2013-2014, though only for a particularly small range of percentages (roughly between 23% and 30.3% of students meeting or exceeding standards) [PP1SD = .372; 50% posterior interval (-0.136, -0.043)]. Likewise, the percentage of students meeting or exceeding standards on the ISAT and PSAE Math subject test was a significant predictor of a school closing in 2013-2014, though also only for a particularly small range of percentages (roughly between 25.6% and 35% and then 40.5% and 44% of

Table V

POSTERIOR ESTIMATES FOR THE FIRST CPS ANALYSIS

Predictor	E[β|Data]	SE[β|Data]	PP1SD	2.5% percentile	25% percentile	75% percentile	97.5% percentile
Intercept	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2011-2012 Probation	0.051	0.002	.000	0.047	0.050	0.053	0.056
2012-2013 Space Use Status	0.029	0.002	.000	0.025	0.027	0.030	0.033
Reading (Knot 1)	-0.028	0.090	.660	-0.204	-0.088	0.033	0.149
Reading (Knot 2)	-0.089	0.069	.372	-0.224	-0.136	-0.043	0.046
Reading (Knot 3)	-0.004	0.089	.682	-0.179	-0.064	0.056	0.171
Math (Knot 1)	0.026	0.087	.661	-0.144	-0.032	0.085	0.196
Math (Knot 2)	-0.158	0.057	.036	-0.269	-0.196	-0.120	-0.047
Math (Knot 3)	-0.027	0.078	.654	-0.181	-0.080	0.026	0.127
Enrollment (Knot 1)	-0.393	0.088	.000	-0.565	-0.452	-0.333	-0.221
Enrollment (Knot 2)	0.320	0.097	.011	0.129	0.255	0.386	0.512
Enrollment (Knot 3)	0.228	0.105	.118	0.023	0.158	0.299	0.433
Enrollment (Knot 39)	-0.017	0.026	.583	-0.068	-0.034	0.000	0.033
Enrollment (Knot 40)	0.297	0.081	.004	0.138	0.242	0.351	0.455

students meeting or exceeding standards) [[PP1SD = .036; 50% posterior interval (-0.196, -0.120)][PP1SD = .491; 50% posterior interval (0.026, 0.150)]]. More research is needed on any practical implications of the significance of the Reading and Math scores. Finally, enrollment number was also a significant predictor, though only for certain ranges (below 320 and above 4278) (all 50% posterior intervals do not include zero). The practical implications of this significance seem more apparent, as the significance of both low and very high enrollment numbers seems to confirm the policy statement by CPS that enrollment was the main factor used to determine which schools to close.

The model fit statistics, as well as computational time, are included in Table VI.

Table VI

MODEL FIT STATISTICS AND COMPUTATIONAL TIME FOR FIRST CPS ANALYSIS

Model Fit	**Value**
R^2	.15
D(m)	2697.95
AIC	4094.09
BIC	4413.65
Number of Outliers ($z > 3.00$)	514 (2.78%)
Computational Time	0.54 seconds

Though the D(m), AIC, and BIC model fit statistics are presented here, the values will be interpreted in the next section as we compare our first and second models. We can, however, interpret the R^2 statistic, number of outliers, and computational time for this model. Our R^2 value of .15 indicates that only 15% of the variance in our school closing dependent variable is accounted for by the predictor variables we included in this first model. Therefore, we conclude that more variables, or a different combination of variables, should be included in a subsequent analysis. Conversely, it is relatively impressive that we only observed 514 outliers (as determined by any *Y* value with a standardized residual greater than 3.00), or extreme data observations, out of 18,506 cases. This lends support for the validity of our model. Likewise, the computational time of 0.54 seconds for analysis of a dataset with n = 18,506 and p = 121 lends great support for the use of ridge regression in our Bayesian linear model.

3.3.2 **CPS Analysis with Spatial Variables**

For our second analysis, we utilized the same CPS dataset and the same dependent variable (whether the school was closed for the 2013-2014 year) as in our first analysis. Since our goals are to both demonstrate the functionality of the model and to compare model performance, we also utilized the same base predictor variables (2012-2013 space utilization status, 2011-2012 probation status, enrollment numbers, Reading scores, and Math scores). However, we extended our model to include spatial locations (point data) in order to both demonstrate the utility of the model and to answer the following question: was the spatial location of the school a significant predictor of whether the school closed in 2013-2014?

We again constructed univariate thin-plate splines for each of the continuous predictor variables (i.e. Reading scores, Math scores, and enrollment numbers), with knots at 40 evenly spaced quantiles. To capture the spatial locations, we included the latitude and longitude coordinates (point data) of each school in our dataset. We then created separate univariate thin-plate splines for latitude and longitude, with knots at 10 evenly spaced quantiles. We selected a smaller number of initial splines since we then created latitude by longitude interaction terms from these splines, resulting in 100 latitude by longitude variables.

We now present the results of our analysis, including posterior estimates of our predictors. Since $p = 241$, we again only include part of our posterior output. We also include the R^2, D(m), AIC, and BIC model fit statistics, with a comparison between the performance of our first model and that of the spatial model used in this analysis. Finally, we again include the number of outliers in the data and the time needed to complete the computations.

The posterior estimates are provided in Table VII. We include the 2011-2012 probation and 2012-2013 space use status variables, as well as selected Reading, Math, enrollment, and spatial knot estimates. As in our first analysis, a school being on probation in 2011-2012 was a significant predictor of the school closing in 2013-2014, when controlling for all other variables in the analysis [PP1SD = .000; 50% posterior interval (0.065, 0.069)]. In addition, space use status during 2012-2013 was again a significant predictor of the school closing [PP1SD = .000; 50% posterior interval (0.025, 0.028)]. The strength of both probation status and space use status as predictors remained the same from the first analysis to the second, particularly in terms of PP1SD (.000 for both variables, across analyses).

The percentage of students meeting or exceeding standards on the ISAT and PSAE Reading subject test continued to be a significant predictor of a school closing in 2013-2014, though only for a few small ranges (including between 23% and 30.3%, and then between 56.6% and 58% of students meeting or exceeding standards) [[PP1SD = .506; 50% posterior interval (-0.197, -0.030)][PP1SD = .394; 50% posterior interval (0.283, 0.963)]]. Likewise, the

Table VII

POSTERIOR ESTIMATES FOR THE SECOND CPS ANALYSIS

Predictor	E[β|Data]	SE[β|Data]	PP1SD	2.5% percentile	25% percentile	75% percentile	97.5% percentile
Intercept	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2011-2012 Probation	0.067	0.003	.000	0.062	0.065	0.069	0.072
2012-2013 Space Use Status	0.026	0.002	.000	0.022	0.025	0.028	0.031
Reading (Knot 2)	-0.113	0.124	.506	-0.356	-0.197	-0.030	0.129
Math (Knot 2)	-0.130	0.095	.348	-0.315	-0.193	-0.066	0.056
Math (Knot 3)	-0.169	0.159	.457	-0.481	-0.276	-0.061	0.144
Enrollment (Knot 1)	-1.006	0.190	.000	-1.379	-1.134	-0.877	-0.632
Enrollment (Knot 2)	0.747	0.260	.030	0.238	0.572	0.923	1.257
Enrollment (Knot 3)	0.936	0.423	.112	0.107	0.651	1.221	1.765
Enrollment (Knot 4)	-1.259	0.471	.047	-2.181	-1.576	-0.941	-0.336
Enrollment (Knot 39)	0.028	0.034	.536	-0.039	0.005	0.051	0.095
Latitude (Knot 2)	-1.185	0.482	.072	-2.130	-1.510	-0.860	-0.241
Longitude (Knot 2)	-0.454	0.530	.525	-1.492	-0.811	-0.097	0.584
Lat*Long (Knot 4)	-0.299	0.271	.441	-0.831	-0.482	-0.116	0.233
Lat*Long (Knot 6)	0.409	0.235	.227	-0.052	0.250	0.568	0.870

percentage of students meeting or exceeding standards on the ISAT and PSAE Math subject test was a significant predictor of a school closing in 2013-2014, with a slightly more expanded range (primarily between 25.6% and 47.5% of students meeting or exceeding standards) (50% posterior intervals do not include zero). Again, we encourage more research on the practical implications of the significance of the Reading and Math scores. Enrollment number also remained a significant predictor for values below 336.30 (excluding a small range between 285 and 303) and between 1352 and 4278 (50% posterior intervals do not include zero). The continued significance of both low and very high enrollment numbers again seems to confirm the CPS statement about the use of enrollment data in determining which schools to close.

Finally, spatial location was a significant predictor of a school closing, in terms of the main effects of both latitude and longitude as well as the interaction of latitude and longitude. Six of the 10 latitude terms were significant, based on 50% posterior intervals that did not include zero. Likewise, six of the 10 longitude terms were significant, again based on the same criterion. Of the latitude by longitude interaction terms, 76 of the 100 knots were significant, based on either a PP1SD < .50 or 50% posterior intervals that did not include zero. We provide more detail on the practical implications of this significance in Section 3.4.

The model fit statistics, as well as computational time, are included in Table VIII. We also include the results from the first analysis for comparison. In terms of our current analysis, the R^2 value of .27 indicates that 27% of the variance in our school closing dependent variable is accounted for by the predictor variables we included in this second model, which is a 12% increase over our first model. This lends strong support for including the spatial location variables in our analysis. Likewise, the lower *D*(m) (2697.95 vs. 2330.03), BIC (4413.65 vs.

2776.17), and AIC (4094.01 vs. 1502.69) values for our second model also lend support for including the spatial variables.

Table VIII

COMPARISON OF MODEL FIT STATISTICS AND COMPUTATIONAL TIME FOR THE FIRST AND SECOND CPS ANALYSES

Model Fit	**Value-Analysis #1**	**Value-Analysis #2**
R^2	.15	.27
D(m)	2697.95	2330.03
AIC	4094.01	1502.69
BIC	4413.65	2776.17
Number of Outliers ($z > 3.00$)	514 (2.78%)	407 (2.20%)
Computational Time	0.54 seconds	1.33 seconds

In terms of overall model performance, we only observed 407 outliers, a decrease from our first model and again impressive for a dataset with n = 18,506. This continues to lend support for the validity of our model, even with the added predictor variables. Likewise, the computational time of 1.33 seconds for analysis with n = 18,506 and p = 241 lends great support for the use of ridge regression in our Bayesian linear model, including when the model is expanded to include spatial variables.

3.3.3 <u>**CPS Analysis with Spatial/Spatio-Temporal Variables and Discussion**</u>

For our third and final analysis, we utilize the same CPS dataset and dependent variable (whether the school was closed for the 2013-2014 academic year) as in our previous two analyses. In order to further demonstrate the functionality of the model and to compare model performance, we use the same base predictor variables as well as the spatial location variables from the second analysis. For this analysis, however, we extend our model to include both a temporal variable as well as a spatio-temporal variable. For the temporal variable, we include the enrollment data from the 2010-2011 academic year (first temporal lag term, or t-1). We selected enrollment data since this variable has shown the most significant relationship with the dependent variable across the previous analyses. For the spatio-temporal variable, we created an interaction between the latitude by longitude interaction terms from the previous analysis and the 2010-2011 enrollment variable. The addition of these variables allows us to answer the question of whether time, as well as spatial location at a particular point in time, were significant predictors of whether a school was closed in 2013-2014.

We again constructed univariate thin-plate splines for each of the continuous predictor variables (i.e. Reading scores, Math scores, and enrollment numbers), with knots at 40 evenly spaced quantiles. We also included the separate univariate thin-plate splines for latitude and longitude, with knots at 10 evenly spaced quantiles, as well as the 100 latitude by longitude interaction variables. For the new temporal variable, we constructed a univariate thin-plate spline for the 2010-2011 enrollment variable, with knots at seven evenly spaced quantiles. We then constructed the interaction terms for latitude, longitude, and time, resulting in 700 latitude by longitude by time interaction variables.

We now present the results of our analysis, including posterior estimates of our predictors. Since $p = 948$, we again only include part of our posterior output. We also include the R^2, D(m), AIC, and BIC model fit statistics, with a comparison between the performance of the current spatio-temporal model and the models used in the two previous analyses. Finally, we again include the number of outliers in the data and the time needed to complete the computations.

The posterior estimates are provided in Table IX. We include the 2011-2012 probation and 2012-2013 space use status variables, as well as selected Reading, Math, enrollment, spatial, temporal, and spatio-temporal knot estimates. As in our two previous analyses, a school being on probation in 2011-2012 was a significant predictor of the school closing in 2013-2014, when controlling for all other variables in the analysis [PP1SD = .000; 50% posterior interval (0.065, 0.068)]. In addition, space use status during 2012-2013 was again a significant predictor of the school closing [PP1SD = .000; 50% posterior interval (0.023, 0.026)]. The strength of both the probation and space use status variables as a predictor of a school closing continued to remain the same as in the previous two analyses, particularly in terms of PP1SD (.000 across all analyses).

The percentage of students meeting or exceeding standards on the ISAT and PSAE Reading subject test continued to be a significant predictor of a school closing, though again only for a few small ranges (including between 23% and 30.3%, and then between 56.6% and 58% of students meeting or exceeding standards) [[PP1SD = .258; 50% posterior interval (-0.157, -0.065)][PP1SD = .519; 50% posterior interval (0.022, 0.167)]]. Likewise, the percentage of students meeting or exceeding standards on the ISAT and PSAE Math subject test was a significant predictor of a school closing in 2013-2014, though, as in the first analysis, only for a particular few small ranges (between 25.6% and 35% and then between 40.5% and 44% of students meeting or exceeding standards) [[PP1SD = .044; 50% posterior interval (-0.187, -0.113)][PP1SD = .449; 50% posterior interval (0.038, 0.162). As with the previous findings, we encourage more research on the practical implications of the continued significance of the Reading and Math scores. Enrollment number also remained a significant predictor for values below 320, between 1352 and 1560, and above 4278 (50% posterior intervals do not include

Table IX

POSTERIOR ESTIMATES FOR THE THIRD CPS ANALYSIS

Predictor	E[β|Data]	SE[β|Data]	PP1SD	2.5% percentile	25% percentile	75% percentile	97.5% percentile
Intercept	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2011-2012 Probation	0.066	0.003	.000	0.061	0.065	0.068	0.071
2012-2013 Space Use Status	0.024	0.002	.000	0.019	0.023	0.026	0.029
Reading (Knot 2)	-0.111	0.068	.258	-0.244	-0.157	-0.065	0.022
Math (Knot 2)	-0.150	0.055	.044	-0.258	-0.187	-0.113	-0.041
Math (Knot 4)	0.100	0.092	.449	-0.081	0.038	0.162	0.281
Enrollment (Knot 1)	-0.395	0.090	.000	-0.572	-0.456	-0.335	-0.219
Enrollment (Knot 2)	0.272	0.098	.039	0.079	0.205	0.338	0.465
Enrollment (Knot 3)	0.254	0.107	.083	0.045	0.182	0.326	0.463
Enrollment (Knot 38)	-0.044	0.059	.559	-0.160	-0.084	-0.004	0.071
Latitude (Knot 1)	-0.088	0.110	.542	-0.303	-0.162	-0.014	0.127
Longitude (Knot 1)	0.087	0.102	.527	-0.113	0.018	0.156	0.287
Lat*Long (Knot 8)	0.169	0.109	.284	-0.044	0.096	0.242	0.382
Lat*Long (Knot 9)	-0.137	0.109	.386	-0.351	-0.211	-0.064	0.077
Enrollment 2010 (Knot 1)	-0.002	0.119	.683	-0.236	-0.082	0.079	0.232
Lat*Long* Time (Knot 23)	-0.095	0.104	.508	-0.299	-0.165	-0.024	0.110

zero). The significance of both low and very high enrollment numbers across all three analyses provides strong support for the use of enrollment data in determining which schools to close.

Finally, spatial location was a significant predictor of a school closing, in terms of the main effects of both latitude and longitude as well as the interaction of latitude and longitude. Six of the 10 latitude terms were again significant, based on 50% posterior intervals that did not include zero. Three of the 10 longitude terms were significant (down from six in the previous analysis), again based on 50% posterior intervals that did not include zero. Of the latitude by longitude interaction terms, 25 of the 100 knots were significant (down from 76 in the previous analysis). The temporal term (2010-2011 enrollment) was not itself a significant predictor; however, 165 of the 700 spatio-temporal terms created by the interaction of latitude, longitude, and 2010-2011 enrollment were significant. As with the previous analysis, we will provide more detail on the practical implications of this significance in Section 3.4.

The model fit statistics, as well as computational time, are included in Table X. We also include the results from the first two analyses for comparison.

Table X

COMPARISON OF MODEL FIT STATISTICS AND COMPUTATIONAL TIME FOR THE FIRST, SECOND, AND THIRD CPS ANALYSES

Model Fit	**Value-Analysis #1**	**Value-Analysis #2**	**Value-Analysis #3**
R^2	.15	.27	.27
D(m)	2697.95	2330.03	2341.56
AIC	4094.01	1502.69	1700.88
BIC	4413.65	2776.17	3810.46
Number of Outliers	514 (2.78%)	407 (2.20%)	420 (2.27%)
Computational Time	0.54 seconds	1.33 seconds	12.04 seconds

Our R^2 value of .27 indicates that 27% of the variance in our dependent variable is accounted for by the predictor variables we included in this third model. Therefore, unlike between the first and second analyses, there was no change in R^2 value between the second and third analyses. This indicates a need to consider additional predictor variables beyond the standardized test scores, enrollment, and the spatial, temporal, and spatio-temporal variables. In addition, both the AIC and BIC values for our third model are higher than those for the second model (1700.88 vs. 1502.69 and 3810.46 vs. 2776.17, respectively), providing strong support for including the spatial variables but less support for the increased model complexity from adding the temporal and spatio-temporal variables (or, perhaps, the particular temporal period selected for this analysis). We note a considerable divergence between the AIC and BIC values, particularly in the third model, which we believe is due to the BIC penalizing more for the additional model complexity. Finally, we note that the D(m) statistic is slightly higher for the third model than the second model (2341.56 vs. 2330.03), which we suggest is due to number of covariates in the model (p = 948) and the nature of the D(m) statistic to penalize for complexity.

In terms of overall model performance, we observed 420 outliers, a decrease from the 514 detected in our first model but a slight increase from the 407 detected in our second model. However, this is still impressive given the size of the dataset and the 707 additional covariates in the third model. We conclude that this lends support for the validity of our spatio-temporal model. Likewise, the computational time of 12.04 seconds for an analysis with n = 18,506 and p = 948 lends great support for the use of ridge regression in our Bayesian linear model, including when the model is expanded to include temporal and spatio-temporal variables.

3.3.4 **Addressing Outliers**

As noted previously, our analyses revealed 514, 407, and 420 outliers for our first, second, and third models, respectively. We determined which values were outliers using a threshold of a standardized residual (*z*) lower than negative three or greater than positive three, or equivalent to three standard deviations above or below the mean and therefore an extreme data observation. It is perhaps not surprising that most of the outliers were cases within schools that were closed prior to the 2013-2014 school year. Per the criteria used by CPS, these schools largely represented the most extreme cases across the district, whether in terms of academic performance, enrollment numbers, or both.

In order to address these outliers in our dataset, we applied a technique developed by Xiong & Joseph (2013) to dummy code the outlier cases. Per this technique, each case in the dataset was dummy coded as either "0" (not an outlier) or "1" (outlier), with a new variable created for each case. We then ran the analysis again, but this time we included the newly created "outlier" (value of 1) dummy coded variables in our set of covariates. We applied this technique to all three models and present the updated results in Table XI.

Table XI

COMPARISON OF MODEL FIT STATISTICS AND COMPUTATIONAL TIME FOR ALL THREE CPS ANALYSES (AFTER CORRECTION FOR OUTLIERS)

Model Fit	**Value-Analysis #1**	**Value-Analysis #2**	**Value-Analysis #3**
R^2	.39	.39	.41
D(m)	1985.75	1975.43	1938.03
AIC	-1115.46	-1266.87	-1559.04
BIC	2824.02	2242.25	2431.95
Computational Time	5.97 seconds	6.05 seconds	24.37 seconds

As shown in Table XI, accounting for the outliers greatly improved the model performance across all models. The R^2 value improved by more than twofold for the first model (from 0.15 to 0.39) and the D(m), AIC, and BIC values improved (decreased) for all models. These results support utilizing the Xiong & Joseph (2013) technique to account for outliers, particularly given that computation time remains below 30 seconds, even with p = 1368.

The decrease in D(m), AIC, and BIC values from the first model to the second and third supports the inclusion of spatial, temporal, and spatio-temporal variables in the model. Per the D(m) and AIC values, the outlier-adjusted third (spatio-temporal) model provides the best fit to the dataset. The BIC values support the outlier-adjusted second (spatial) model, most likely due to the added complexity in the third model. Overall, the model performance and fast computational speeds continue to support the use of ridge regression in our Bayesian model.

3.4 Considerations When Interpreting Spatial and Spatio-Temporal Analyses

The final part of our toolkit pertains to the interpretation of statistical analyses conducted using spatial and spatio-temporal variables. How should an educational researcher interpret findings of statistical significance and evidence to support inclusion of spatial and spatio-temporal variables, such as model fit indices (e.g. AIC and BIC)? In terms of statistical analysis, the significant latitude by longitude interaction knots form a “smoothing” grid for discrete spatial data (introduced in Section 1.5.1). Though our datasets utilize point data (represented by latitude and longitude), the grid serves to form a more condensed representation of these spread out points, thus better facilitating the regression analysis.

In general, though, we argue that a researcher should avoid making causal claims about these spatial and spatio-temporal variables and should instead incorporate more work on exploring correlations (relationships) between variables, including those established through statistical analyses and those discovered through digging deeper into the findings. This exploration of relationships can include elements such as visualization of the data through maps, descriptive elements such as counts, and a more qualitative investigation of the spatial locations, which serves to uncover elements that may contribute to the research findings (e.g. a school may be in a neighborhood with low median resident income, which may impact resources available to the schools as well as to the students outside of school) (Sampson, 2008). These elements could then be incorporated into subsequent statistical analyses.

We illustrate these considerations with interpretations of our own analyses conducted in Sections 3.3.3 and 3.3.4 of this chapter. In Section 3.3.3, we analyzed the academic performance and enrollment-related variables for 554 schools, while adding in spatial location as a predictor variable. The results of this analysis, as shown in part in Table VII, indicate both that spatial location is a significant predictor of whether a school closed in 2013-2014 and that the inclusion of spatial location improves model fit (across multiple fit indices). Likewise, in Section 3.3.4, we analyzed the same academic performance, enrollment, and spatial variables, while adding temporal and spatio-temporal elements. The results of this final analysis, as shown in part in Table IX, indicate that spatial location and the spatio-temporal variable were significant predictors and that the inclusion of temporal and spatio-temporal variables improved model fit (per most fit indices).

Since spatial location was a significant predictor of a school closing, we begin with a visualization of the 45 CPS schools that were closed for the 2013-2014 school year. In Figure 1, we present a map of the closed schools, created using the free and publicly available Google Fusion Tables program (2016).

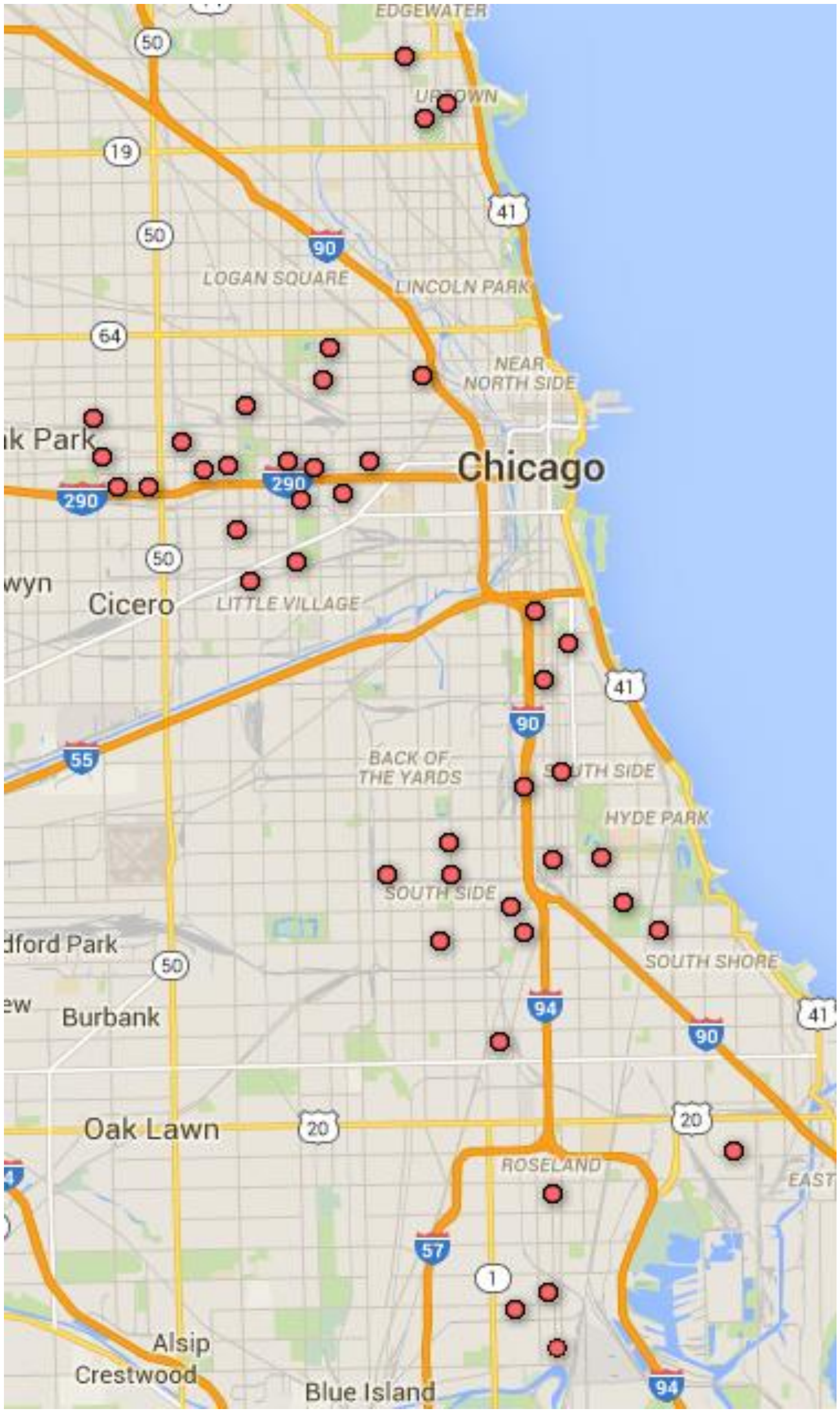

Figure 1. The 45 Chicago Public Schools schools that were closed prior to the start of the 2013-2014 school year.[3]

[3] There are 43 schools shown on the map because Roque du Duprey Elementary School and Von Humboldt Elementary School shared an address. Likewise, Williams Middle Prep Academy and Williams Multiplex Elementary School also shared an address.

Examining the map reveals that most of the schools that closed were located in Chicago's West and South sides. On the surface, we could then infer a relationship between a school's location in the West or South side of the city and likelihood of closing. However, what additional information could we provide about these areas of Chicago and the various neighborhoods located within? With this question, we are able to begin a more in-depth exploration of additional factors that may have contributed to a school being selected for closure.

As a first step, we recall the West and South sides of Chicago tend to have a higher percentage of minority residents than other areas of the city. Per Chicago census data, the primary minority group in the West and South sides is African-Americans, who comprise as much as 85.8% to 100% of the population in some areas of the West and South sides (Frankel, 2013). Likewise, we revisit the research by Radinsky and Waitoller (2013) on the CPS school closings, which found that 81% of the students impacted by the school closings were African-American. Thus, we further explore our initial findings through an additional statistical analysis that incorporates racial composition of the student body as a covariate.

For our follow-up analysis, we utilize the CPS Demo dataset described in Chapter 2. We elected to only use data from 2011-2012 since the focus of this analysis is exploration of additional factors related to spatial location. The CPS Demo dataset contains the following data from the CPS dataset: 2011-2012 probation status, 2012-2013 space utilization status, spatial variables (latitude, longitude, and latitude by longitude interaction terms), and 2011-2012 enrollment numbers, Reading scores, and Math scores. We also added race/ethnicity data from the *CPS 2011-2012 Racial/Ethnic Report* (Chicago Public Schools, 2016). Though the *Racial/Ethnic Report* includes counts and percentages of students identifying in nine categories (listed in Chapter 2), for the sake of this analysis we summed the number of students identifying in any category except for “White” and divided by the total number of students enrolled in order to create a 2011-2012 Percentage Minority variable. The descriptive statistics for the ISAT/PSAE Math scores, ISAT/PSAE Reading scores, enrollment numbers, and percentage minority variable are presented in Table XII.

Table XII

DESCRIPTIVE STATISTICS FOR THE CPS DEMO DATASET

Variable	Mean	Standard Deviation
ISAT/PSAE Reading (2011-2012)	67.85	18.74
ISAT/PSAE Math (2011-2012)	76.10	17.56
Enrollment (2011-2012)	589.07	362.05
Percentage Minority (2011-2012)	91.13	16.70

3.4.1 **Analysis of CPS Demo Dataset and Discussion**

Prior to running the analysis, we again created univariate thin-plate splines for the continuous predictor variables (i.e. Reading scores, Math scores, enrollment numbers, and racial composition), using knots at 40 evenly spaced quantiles of the data. We again also created univariate thin-plate splines for latitude and longitude, and then created a latitude by longitude interaction term from the two univariate splines.

We now present the results of our analysis, including posterior estimates of our predictors. The results shown are post-application of the Xiong & Joseph (2013) technique to account for 71 outliers ($z > 3$). Since $p = 336$, we again only include part of our posterior output, along with the R^2, D(m), AIC, and BIC model fit statistics and the time needed to complete the computations.

The posterior estimates are provided in Table XIII. We include the 2011-2012 probation and 2012-2013 space use status variables, as well as selected Reading, Math, enrollment, spatial, and racial composition knot estimates. Though we only considered data from the 2011-2012 academic year, the results for this analysis are similar to those from the three CPS dataset analyses. In this analysis, probation status, space use status, and very low and very high enrollment numbers were significant predictors of a school closing, which is consistent with the CPS analyses. In addition, 28 of the 100 latitude by longitude interaction terms were significant predictors, which continues to support the relationship between spatial location and school closings.

However, unlike in the CPS dataset analyses, the Reading and Math score variables were not significant predictors. This indicates support for a more longitudinal analysis of the school closings, rather than the cross-sectional approach taken in this analysis (wherein we only analyzed scores from 2011-2012). The racial composition variable was also not a significant predictor, an initially rather unexpected finding given the spatial locations of the closed schools. We will explore this finding further in the Section 3.4.2.

Table XIII

POSTERIOR ESTIMATES FOR THE CPS DEMO ANALYSIS

Predictor	E[β|Data]	SE[β|Data]	PP1SD	2.5% percentile	25% percentile	75% percentile	97.5% percentile
Intercept	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2011-2012 Probation	0.049	0.006	.000	0.038	0.045	0.053	0.060
2012-2013 Space Use Status	0.012	0.005	.065	0.003	0.009	0.016	0.022
2011-2012 Minority (Knot 1)	-0.005	0.021	.669	-0.047	-0.019	0.009	0.036
Reading (Knot 1)	0.003	0.020	.678	-0.037	-0.011	0.016	0.042
Math (Knot 1)	-0.000	0.020	.682	-0.040	-0.014	0.013	0.039
Enrollment (Knot 1)	-0.066	0.020	.011	-0.106	-0.080	-0.053	-0.027
Enrollment (Knot 2)	-0.035	0.021	.239	-0.076	-0.049	-0.021	0.005
Enrollment (Knot 3)	-0.020	0.021	.494	-0.061	-0.034	-0.006	0.021
Enrollment (Knot 40)	0.068	0.020	.008	0.029	0.055	0.082	0.107
Latitude (Knot 1)	-0.009	0.020	.640	-0.048	-0.022	0.005	0.031
Longitude (Knot 1)	0.004	0.020	.671	-0.035	-0.009	0.018	0.043
Lat*Long (Knot 3)	0.016	0.018	.517	-0.020	0.004	0.028	0.052

Finally, in Table XIV we present model fit indices and computational time for both the second CPS dataset analysis (with outliers accounted for) and for the current analysis (with and without the percentage minority variable).

Table XIV
COMPARISON OF MODEL FIT STATISTICS AND COMPUTATIONAL TIME FOR THE SECOND CPS DATASET ANALYSIS AND CPS DEMO ANALYSIS

Model Fit	**Initial Spatial Analysis**	**Updated Spatial Analysis (With Race)**	**Updated Spatial Analysis (Without Race)**
R^2	0.39	0.46	0.45
D(m)	1975.43	257.85	261.80
AIC	-1266.87	-472.29	-436.14
BIC	2242.25	209.17	213.91
Computational Time	6.05 seconds	7.29 seconds	7.91 seconds

It is important to note that the second CPS dataset analysis contained n = 18,506 and p = 648 (p = 241 when excluding the outlier dummy coded variables), while the current analysis contained n = 2,729 and p = 336. The difference in n is due to the longitudinal nature of the CPS dataset, as compared to the cross-section analyzed in the CPS Demo dataset. Given these, the *D*(m) and BIC model fit indices improve (decrease) considerably for the current analysis. In addition, the amount of variance in our dependent variable accounted for by our model increased from 39% in the second CPS analysis to 46% in the current analysis.

We also compared the model fit indices for two analyses run on the CPS Demo dataset: first with the racial composition variable (n = 2,729 and p = 336), and then with that variable excluded (n = 2,729 and p = 307). As shown in Table XIV, the *D*(m) and BIC model fit indices improve (decrease) when race is included. In addition, R^2 value increases slightly, from 45% to 46%. Therefore, even though our percentage minority variable was not a significant predictor of whether a school closed, the model fit indices and R^2 value indicate that, at least when considering the 2011-2012 school year, there is definite value in including this variable in the model.

3.4.2 **Final Considerations**

The question of the relationship between a school's racial composition and whether the school closed provides an excellent case for Sampson's (2008) argument that researchers should utilize descriptive approaches as much as, if not more than, statistical analyses (such as regression), where results that indicate a relationship between variables could potentially be misinterpreted to imply causation. In the same vein, relying on results of statistical analyses without incorporation of additional descriptive approaches, such as maps, can result in a failure to uncover unobserved variables that actually account for what we observe in our data.

To this point, the results of the regression analysis indicate that while inclusion of the percentage minority variable improves model fit and the R^2 value, the variable itself is not a significant predictor of whether a school was closed. However, it is important to remember that the average percentage of minority students in the observed schools was 91.13%. Indeed, 95.8% of the schools were "majority minority;" that is, more than 50% of the students in these schools identify with a racial/ethnic group other than White. Furthermore, the racial composition of the student population at all of the 45 schools closed for 2013-2014 was at least 87% minority, and 38/45 (84%) were at least 99% minority. Therefore, it is possible that the skew of the data toward higher values may complicate the regression analysis. It may be advisable to compare analysis results with a combined racial variable, as done in this thesis, with an analysis that looks at individual racial/ethnic groups.

It could also be, however, an illustration of Holland's (2008) point that race itself is not often the actual variable of interest, and therefore should not be used as a causal variable. Rather, what we attribute to race is often instead a product of other unobserved predictors that we have not included in our analysis and will not uncover without more (often qualitative) investigation. After all, though all of the schools that were closed had student populations composed of at least 87% minority students, there were 400 schools (72.3% of the total) that matched this percentage and yet were not closed.

Given this, we conclude with three maps, again created using the Google Fusion Tables program (2016), that hopefully inspire more questions and more exploration of the school closing data. First, in Figures 2 and 3 we capture visualizations of spatio-temporal data through maps of schools that had enrollments lower than 350[4] (in 2005-2006 and 2011-2012, respectively). The juxtaposition of these maps demonstrates not only the increase in the number of schools with lower enrollments, but also where these schools tend to be located in Chicago (West and South sides). Finally, in Figure 4 we present a map of those schools with 2011-2012 racial/ethnic student body compositions of at least 50% White. We present Figure 4 as a contrast to Figures 1, 2, and 3, as the schools with lower minority enrollments tend to be located in the North and East sides of the city. It is our hope that the maps inspire researchers to explore these trends across different areas of Chicago and to delve deeper into what characteristics of these varying spatial locations may have also served as factors influencing which schools were closed.

---

[4] We select 350 as our cut point due to its frequency as the upper bound of the significant enrollment knot points (for lower enrollment values) across many of the analyses performed in this thesis.

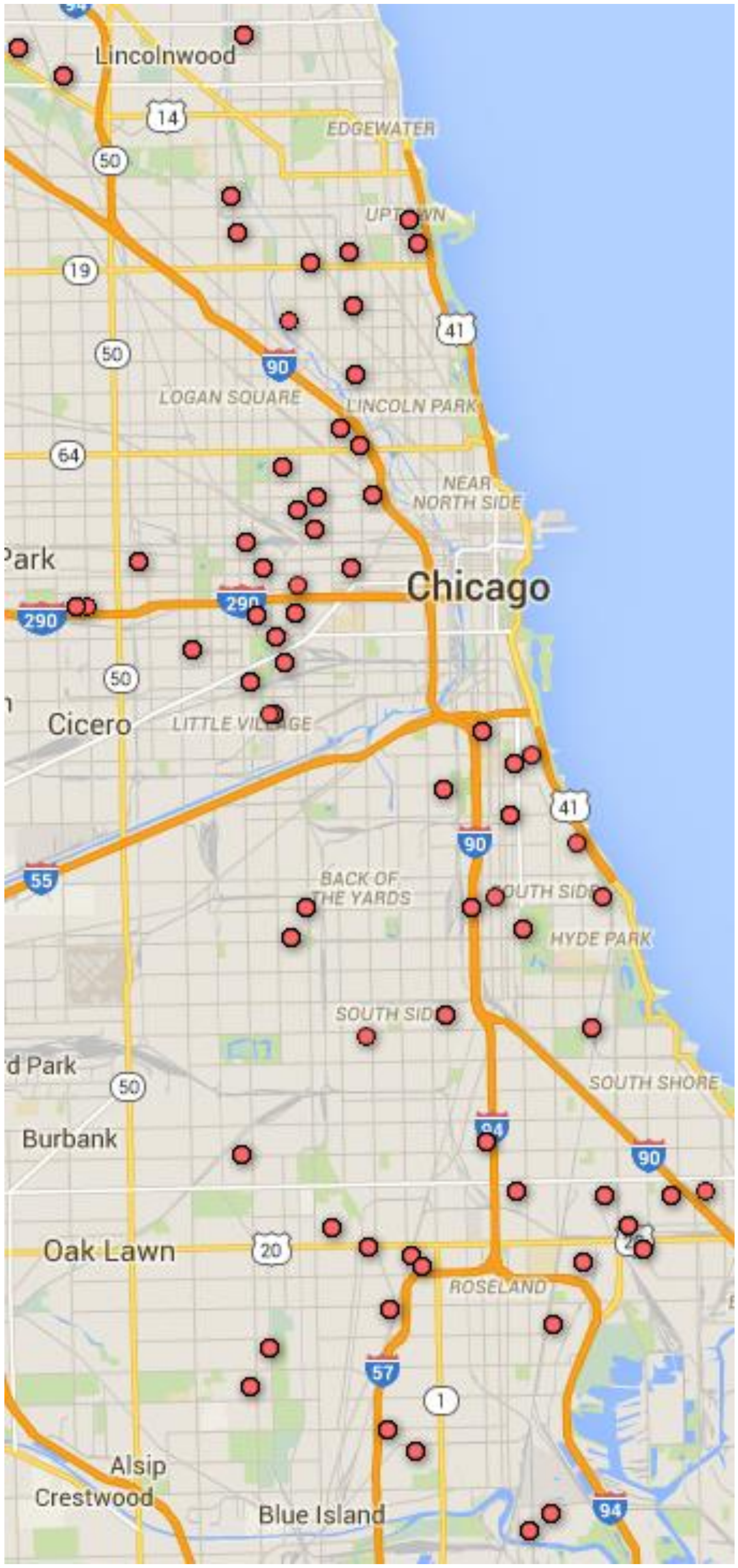

Figure 2. CPS schools with enrollments under 350 students (2005-2006)

Figure 3. CPS schools with enrollments under 350 students (2011-2012)

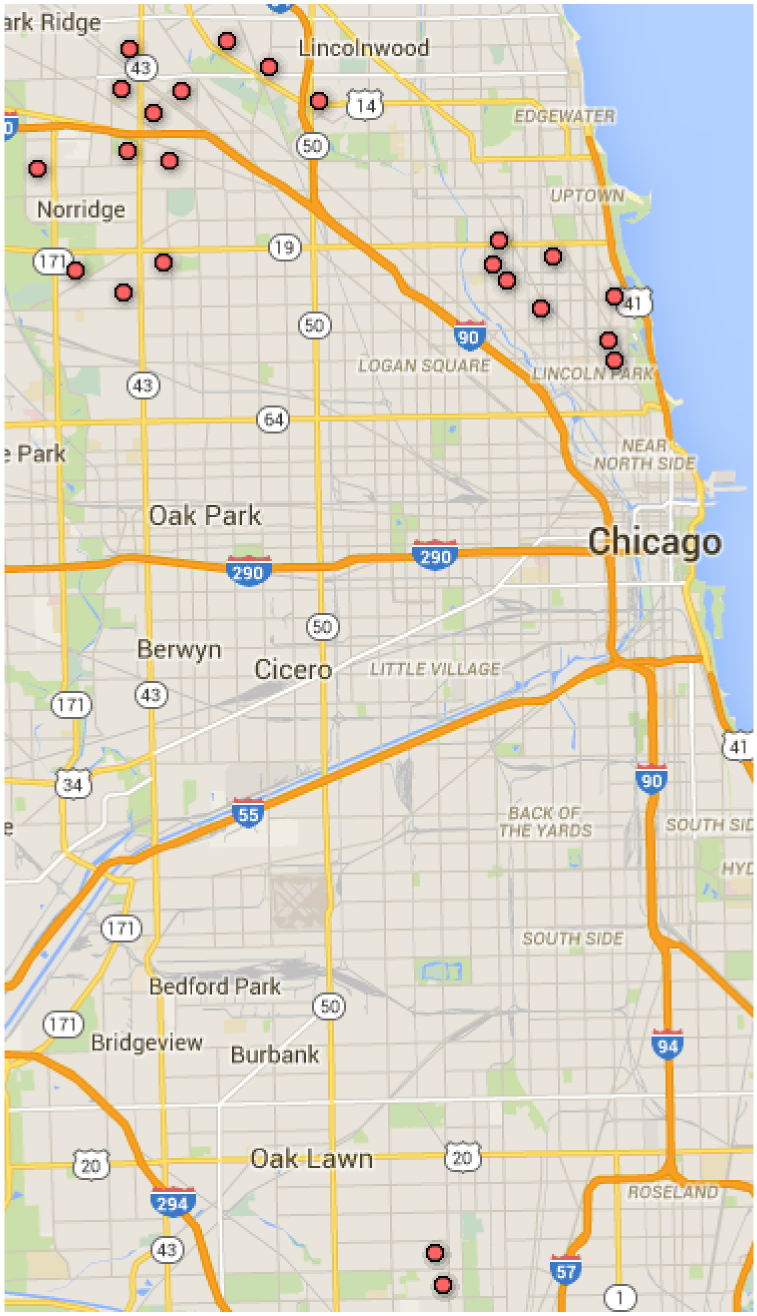

Figure 4. CPS schools with 2011-2012 student body racial/ethnic composition of at least 50% White

4. CONCLUSIONS

4.1 **Summary and Contributions**

In this thesis, we argue for an increase in the use of spatio-temporal analysis in educational research and provide a toolkit for encouraging this increased use, particularly in research looking at fixed spatial locations (e.g. schools) that can be characterized by latitude and longitude coordinates (referred to as point data). We argue that this analysis should be both statistical and more descriptive. The first part of our toolkit addresses the statistical side of spatio-temporal analysis through introduction of our Bayesian linear spatio-temporal model. Our Bayesian linear model incorporates thin-plate splines and ridge regression to better accommodate spatio-temporal variables. Our model also uses a structure akin to that used in analysis of variance (ANOVA) models, in terms of inclusion of main effects and interaction terms. This model is facilitated by use of the *Bayesian Ridge Regression* software program, developed by Dr. George Karabatsos (2016), which is the second part of our toolkit. The third part of our toolkit addresses the more descriptive side of spatio-temporal analysis through identification and demonstration of techniques suggested by Sampson (2008), such as counts and visualization through mapping. This part of our toolkit also emphasizes the use of these descriptive techniques to further explore spatial and spatio-temporal relationships, often through use of more qualitative methods. This exploration is emphasized in part due to a need to caution researchers to avoid conflating correlation and causation, especially when sensitive variables such as race are implicated. This third part of our toolkit is facilitated in part by the free and publicly available Google Fusion Tables program (2016), which allows users to generate maps from Excel spreadsheets.

However, we also note three challenges that serve as obstacles to using spatio-temporal analyses in educational research: complications of large datasets, complex statistical models, and model selection. In this section, we note how we addressed each challenge individually. In the next section, we will assess how well we addressed each challenge.

In Chapter 3 of this thesis, we addressed these challenges through analysis of two datasets, one relatively large (n = 18,506) and one smaller (n = 2,729). We addressed the first challenge, complications of large datasets, through use of Bayesian Ridge Regression (RR) and marginal maximum likelihood (MML) in our statistical analyses, as the use of MML with RR allows researchers to avoid the lengthy computational time often involved with more traditional Markov chain Monte Carlo (MCMC) sampling. We addressed the second challenge, complex statistical models, through use of an ANOVA-based model, which is familiar to most educational researchers, as well as through use of the menu-driven, point-and-click *Bayesian Ridge Regression* software program (Karabatsos, 2016) for our analyses. Finally, we addressed the third challenge, model selection, through demonstration of the ease of determining significant variables by using the interquartile (50%) posterior credible intervals provided in the output produced by the *Bayesian Ridge Regression* software program (Karabatsos, 2016). Likewise, we also demonstrated the ease of determining the best combination of predictors to include in a model by using the model fit indices [R^2, the D(m) statistic, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC)] also provided in the posterior output.

The primary contribution of this thesis is addressing the dearth of spatial (in terms of geographic location) and spatio-temporal analyses in educational research. Despite the fact that educational research explores research questions that lend well to these types of analyses, such as those using schools and neighborhoods as units of analyses, the inclusion of spatial and spatio-temporal variables is virtually non-existent in the literature. In this thesis, we not only identify this need in educational research, but also identify potential roadblocks for researchers: the three challenges noted above. Through creation of this Bayesian linear model, and demonstration of how it can be implemented in a more user-friendly software program, we hope to not only inspire researchers to learn more about spatial and spatio-temporal analysis, but to begin to incorporate it into their own research.

4.2 **Assessment of Contributions**

In this section, we assess how well we addressed each of the three challenges to incorporating spatial and spatio-temporal analyses. We start with the first challenge, complications of large datasets. As noted in the previous section, one issue with using MCMC sampling in analysis of large datasets is the lengthy amount of computational time needed. We proposed to address this by using Bayesian Ridge Regression (RR) and marginal maximum likelihood (MML). After running the analyses documented in Chapter 3, we conclude that the use of RR and MML adequately addresses this challenge. We support this through the observed computational times (shown in Tables X, XI, and XIV in Chapter 3), including 0.54 seconds to analyze a dataset with $p = 121$ and $n = 18{,}506$. The average computational time, including for analyses with over 500 covariates, was between six and seven seconds. For the analysis with the highest number of covariates ($p = 1368$) and $n = 18{,}506$, the computational time was still under 30 seconds (24.37 seconds). In contrast, the computational time to run this same analysis using a binary probit linear model with MCMC sampling (20,000 iterations) is upwards of 34 hours.

We assess the second challenge, complex statistical models, in two parts. The first part is the model itself, which we conclude addresses the challenge due to it ultimately being a familiar

ANOVA-based linear model. In ANOVA terms, the model parameters are a mix of main effects (e.g for predictors such as Reading scores) and interaction terms (e.g. for the latitude by longitude interaction term predictor variable). Even the thin-plate spline terms, which may be a new concept for many educational researchers, are made more conceptually accessible while conducting the analysis due to how the spline terms are created in the software program. The second part is, then, using the model for analysis. We conclude that this addresses the challenge due to the use of the *Bayesian Ridge Regression* software (Karabatsos, 2016), which allows for a point-and-click experience similar to using SPSS (complete with a guide in the "Help" menu that provides step-by-step instructions on how to set up and run analyses). This is in stark contrast to programs commonly used for spatial analysis, such as *R*, which are primarily coding-based. The menu-driven set-up also simplifies creating the model by including options to create the thin-plate spline terms and interaction terms. In addition, the posterior output is presented in a format that is both easy to read and to interpret, including descriptions of how to determine significance. We model this by including partial output from our results in Tables V, VII, IX, and XIII in Chapter 3, both to provide readers with an idea of what the output looks like as well as how to interpret it. For full illustration, we also include the posterior output from our first CPS analysis (Section 3.3.2) in Table XV, Appendix A.

We assess the final challenge, model selection, similarly to the challenge of complex statistical models. We conclude that we adequately address this challenge in part based on the consistency in parameter estimates across the three CPS analyses in Chapter 3 (i.e. significant predictors tended to largely remain the same across analyses), as well as that the ridge parameter, λ, is greater than zero for all analyses, indicating that the ridge regression is addressing multicollinearity in our dataset. We also conclude that we adequately address model selection based on the model fit indices provided in our analyses. As with the previous challenge, the clear layout of the model fit indices [R^2, the D(m) statistic, AIC, and BIC] in the posterior output greatly facilitates our ability to address this challenge (the model fit indices are included in the output provided in Table XV, Appendix A). We model the availability of the model fit indices in Tables VI, VIII, X, XI, and XIV in Chapter 3. Moreover, we are also able to model the utility of the various fit indices through our analyses, including comparing the results between models and the impact of the inclusion of various predictor variables on the model. For example, though we did not find 2011-2012 school racial composition (“percentage minority”) to be a significant predictor of whether a school closed, the model fit indices improved when this variable was included in the analysis. Therefore, we conclude that the model with the best fit to our dataset, and therefore the best to use for analysis, is the one including the percentage minority variable.

Moreover, we conclude that this thesis serves as strong evidence for the inclusion of spatial and spatio-temporal analyses in educational research. We demonstrate that use of our model mitigates the primary challenges to incorporating these analyses. The model provides for a conceptually easy way to include spatial location as a variable, while also correcting for any multicollinearity between the variables, spatial or otherwise. The model also allows for very fast analyses of datasets with many predictor variables and a large n, as is often the case with datasets that include spatial and temporal elements.

Perhaps the strongest evidence is the analyses themselves. We conducted three analyses on the same dataset in Chapter 3: one with no spatial or spatial-temporal variables, one with spatial variables, and one with spatial and spatio-temporal variables. Spatial location was a significant predictor in both the second and third analyses, and the spatio-temporal terms were significant in the third analysis. More compelling still is that despite the increase in model complexity with the addition of these variables, almost all model fit indices indicated that the spatio-temporal model was the best fit for the dataset. In terms of real-life application, this is particularly powerful in that it implies that one should not have a discussion about the CPS school closings without acknowledging the role of spatial location and trends across time.

4.3 **Recommendations for Future Work**

Since we used the CPS school closings as our case study for this thesis, we hope researchers will continue to ask questions that dig deeper into the school closings and the impact felt by CPS students and their families. For our primary analyses, we intentionally included only variables that were cited by the district as factors considered when deciding which schools to close. This was, in part, to explore whether statistics would back stated policy. This was also to give a relatively cut and dry baseline for starting to explore what relationships may exist between spatial location and school performance overall. The statistical analyses, as well as the maps generated using Google Fusion Tables (2016), provided tremendous insight into overall trends with schools in various parts of Chicago.

The final analysis, which included a variable based on the racial/ethnic composition of a school's student body, only scratches the surface of variables to include when further investigating the school closings (and CPS schools as a whole). The model fit indices support the inclusion of a race-based variable, though we strongly support Holland's (2008) argument that we must drill down to find other variables that may better explain our findings. We also encourage researchers to look beyond demographic variables and to explore questions related to a school's surroundings, such as whether proximity to a charter school or ample green space impacted the likelihood of a school closing. Regardless of what variables a researcher elects to include, we hope that this thesis will inspire researchers to explore the vast amounts of existing data available from CPS and ISBE, among others, and to explore this data using our model and free tools such as the *Bayesian Ridge Regression* software (Karabatsos, 2016) and Google Fusion Tables (2016).

We hope educational researchers will be inspired to ask what research questions they might have that would benefit from spatial and spatio-temporal analysis, and that they will feel confident to start incorporating these analyses into their research. This thesis is meant to serve as a toolkit for introducing educational researchers to these types of analyses; this model makes spatial analysis (particularly of discrete spatial processes utilizing locations identified using latitude and longitude) accessible and hopefully instills a curiosity into other types of spatial analysis. For example, there would be great merit in analyzing the spatial patterns of the closed CPS schools to determine whether spatial clustering exists. Likewise, there would be merit in breaking down the CPS analysis by neighborhood and analyzing differences in counts of closed schools across neighborhoods. This would lend well to further descriptive analyses as well. Both are outside of the capacity of this current model, but hopefully educational researchers will employ additional existing spatial analysis models or, better yet, continue to create new (and more accessible) models.

In terms of the model and analysis, it would be interesting for a spatial statistician to compare the multicollinearity corrected for by ridge regression with techniques commonly used in spatial statistics to account for spatial autocorrelation, such as inclusion of a spatial covariance matrix. Given that the ridge parameter values indicated correction for multicollinearity, this was outside of the scope of this dissertation. However, it may be worthwhile to see if there is an advantage to using ridge regression in conjunction with a technique specifically meant to account for spatial autocorrelation.

In addition, more research could be done on knot selection using ridge regression, especially knot selection for spatial locations. We opted to use 40 evenly spaced quantiles for many of our spline knots, as per Ruppert (2002). In the case of spatial location, we used 10 for latitude and 10 for longitude, in order to have a more manageable number of latitude by longitude interaction terms to analyze. The ridge regression analysis easily identified significant spatial locations and a map of the significant knots revealed a grid that quite adequately covered all spatial locations included in the dataset. However, more investigation could be done on the smoothing done by the thin-plate splines for the spatial terms in particular. While the literature strongly suggests the use of thin-plate splines for spatial data, it would be interesting to explore use of ridge regression with other types of splines. In addition, it would also be interesting to redo the analyses conducted in this thesis using different numbers of spatial knots, in order to gauge the impact on both smoothing and model fit.

Finally, an overarching motivation for this thesis was the desire to make a traditionally complex statistical analysis more broadly accessible. As demonstrated in this thesis, there can be tremendous value in including spatial and spatio-temporal analyses in research, especially for educational researchers. However, educational researchers are rarely trained in spatial or spatio-temporal analysis. The literature abounds with innovative and broadly useful statistical models; unfortunately, it often takes a statistician or extensive training in statistics to be able to fully comprehend the articles, let alone to use the models in their research. We hope that additional work will be done to make learning about advanced statistics more accessible to researchers who are not statisticians, in order that these models may be more broadly understood and applied. To go along with this, we also hope for increased development of software programs such as *Bayesian Ridge Regression* (Karabatsos, 2016), which aim to make advanced statistical analyses accessible through use of a familiar point-and-click interface. This increase in training on use of advanced statistics, coupled with more accessible platforms with which to implement these analyses, only stands to strengthen educational research.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267- 281). Budapest, Hungary: Akademai Kiado.

Alperin, J. P. (2008). *A spatio-temporal model for the evaluation of education quality in Peru* (Master's thesis). Retrieved from ProQuest Digital Dissertations and Theses. (MR 43577)

Banerjee, S., & Fuentes, M. (2011). Bayesian modeling for large spatial datasets. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(1), 59-66. doi:10.1002/wics.187

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 825-848. doi: 10.1111/j.1467-9868.2008.00663.x

Berliner, D. (2006). Our impoverished view of educational reform. *The Teachers College Record*, *108*(6), 949-995. doi: 10.1111/j.1467-9620.2006.00682.x

Bloom, H. S., Thompson, S. L., & Unterman, R. (2010, June). Transforming the high school experience: How New York City's new small schools are boosting student achievement and graduation rates. *MDRC*. Retrieved from http://files.eric.ed.gov/fulltext/ED511106.pdf

Box, G. E. P, & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control (revised ed*.). San Francisco: Holden-Day.

Bronfenbrenner, U. (1976). The experimental ecology of education. *Educational Researcher*, *5*(9), 5-15. doi: 10.3102/0013189X005009005

Casella, G. (2008). *Bayesians and frequentists: Models, assumptions, and inference* [PDF document]. Retrieved from http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf

Chicago Public Schools (n.d.) *Summary of 2012-2013 Guidelines*. Retrieved from http://cps.edu/About_CPS/Policies_and_guidelines/Documents/2012-2013SummaryofGuidelines.pdf

Chicago Public Schools (2011, December 28). *Chicago Public Schools Space Utilization Standards*. Retrieved from http://cps.edu/About_CPS/Policies_and_guidelines/Documents/SpaceUtilizationStandards.pdf

Chicago Public Schools (2016). *Racial/Ethnic Report: School Year 2011-2012.* Retrieved from http://cps.edu/SchoolData/Pages/SchoolData.aspx

Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. London, UK: Pion.

Crainiceanu, C. M., Diggle, P. J., & Rowlingson, B. (2008). Bivariate binomial spatial modeling of Loa loa prevalence in tropical Africa. *Journal of the American Statistical Association*, *103*(481), 21-37. doi: 10.1198/016214507000001409

Cressie, N. A. C. (1991). *Statistics for spatial data*. New York, NY: John Wiley & Sons, Inc.

Damian, D., Sampson, P. D., & Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, *12*(2), 161-178. doi: 10.1002/1099-095X(200103)12:2<161::AID-ENV452>3.0.CO;2-G

Diggle, P. J. (2010). Historical introduction. In A. E. Gelfand, P. J. Diggle, P. Guttorp, & M. Fuentes (Eds.), *Handbook of spatial statistics* (pp. 3-14). Boca Raton, FL: CRC Press.

Diggle, P., & Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, *33*(1), 53-64.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive theory of functions of several variables* (pp. 85-100). Berlin, Germany: Springer.

Dziak, J.J., Coffman, D.L., Lanza, S.T., & Li, R. (2012). *Sensitivity and specificity of information criteria* (No. 12-119). University Park, PA: The Methodology Center, The Pennsylvania State University. Retrieved from http://methodology.psu.edu/media/techreports/12-119.pdf

Eidsvik, J., Finley, A. O., Banerjee, S., & Rue, H. (2012). Approximate Bayesian inference for large spatial datasets using predictive process models. *Computational Statistics & Data Analysis*, *56*(6), 1362-1380. doi: 10.1016/j.csda.2011.10.022

Fan, Y., & Tang, C. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B, 75*, 531-552. doi: 10.1111/rssb.12001

Finley, A. O., Banerjee, S., & Gelfand, A. E. (2012). Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *Journal of Geographical Systems*, *14*(1), 29-47. doi: 10.1007/s10109-011-0154-8

Finley, A. O., Sang, H., Banerjee, S., & Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational Statistics & Data Analysis*, *53*(8), 2873-2884. doi: 10.1016/j.csda.2008.09.008

Frankel, S. (Cartographer). (2013). African-American population by census tract in Chicago, IL (2011) [Demographic map]. Retrieved from https://commons.wikimedia.org/wiki/File:African_American_Population_by_Census_Tract_in_Chicago,_IL_(2011).svg#file

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika, 89*(1), 197–210. doi: 10.1093/biomet/89.1.197

Gamerman, D. (2010). Dynamic spatial models including spatial time series. In A. E. Gelfand, P. J. Diggle, P. Guttorp, & M. Fuentes (Eds.), *Handbook of spatial statistics* (pp. 437-448). Boca Raton, FL: CRC Press.

Gelfand, A. E., Banerjee, S., & Finley, A. O. (2011). Spatial design for knot selection in knot-based dimension reduction models. In J. Mateu & W. Muller (Eds.), *Spatio-temporal design: Advances in efficient data acquisition* (pp. 142-169). Chichester, UK: John Wiley & Sons, Inc.

Gelfand, A. E., Banerjee, S., & Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics, 16*(5), 465-479. doi: 10.1002/env.715

Gelfand, A. E., Diggle, P., Fuentes, M., & Guttorp, P. (Eds.). (2010). *Handbook of spatial statistics*. Boca Raton, FL: CRC Press.

Gelfand, A. E., & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika, 85*(1), 1-11. doi: 10.1093/biomet/85.1.1

Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*(462), 387-396. doi: 10.1198/016214503000170

Gelfand, A. E., Kottas, A., & MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, *100*(471), 1021-1035. doi: 10.1198/016214504000002078

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*(2), 339-373. Retrieved from http://www-stat.wharton.upenn.edu/~edgeorge/Research_papers/GeorgeMcCulloch97.pdf

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, *97*(458), 590-600. doi: 10.1198/016214502760047113

Goodchild, M. F. (2008). Combining space and time: new potential for temporal GIS. In A.K. Knowles (Ed.), *Placing history: How maps, spatial data, and GIS are changing historical scholarship*, (pp. 179-198). Redlands, CA: ESRI Press.

Goodchild, M. F., & Janelle, D. G. (Eds.). (2004). *Spatially integrated social science*. New York, NY: Oxford University Press.

Goodchild, M. F., & Janelle, D. G. (2010). Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, *75*(1), 3-13. doi: 10.1007/s10708-010-9340-3

Google Fusion Tables [Computer program]. (2016). Retrieved from https://fusiontables.google.com/DataSource?dsrcid=implicit

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. Boca Raton, FL: CRC Press.

Guan, Y., Sherman, M., & Calvin, J. A. (2004). A nonparametric test for spatial isotropy using subsampling. *Journal of the American Statistical Association*, *99*(467), 810-821. doi: 10.1198/016214504000001150

Guhaniyogi, R., Finley, A. O., Banerjee, S., & Gelfand, A. E. (2011). Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics*, *22*(8), 997-1007. doi: 10.1002/env.1131

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. New York, NY: Cambridge University Press.

Hancock, P. A., & Hutchinson, M. F. (2006). Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling & Software*, *21*(12), 1684-1694. doi: 10.1016/j.envsoft.2005.08.005

Hensman, J., Fusi, N., & Lawrence, N. D. (2013). Gaussian processes for big data. *arXiv preprint arXiv:1309.6835,* 1-9. Retrieved from http://arxiv.org/abs/1309.6835

Higdon, D., Swall, J., & Kern, J. (1999). Non-stationary spatial modeling. In J.M. Bernardo & J.O. Berger (Eds.), *Bayesian Statistics 6,* (pp. 761-768). Oxford, UK: Oxford University Press.

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, *58,* 54-59. Retrieved from http://www.aiche.org/resources/publications/cep

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12,* 55-67. Retrieved from http://math.arizona.edu/~hzhang/math574m/Read/Ridge.pdf

Hoff, P. (2013, October 31). *Shrinkage estimators* [PDF document]. Retrieved from http://www.stat.washington.edu/people/pdhoff/courses/581/LectureNotes/shrinkage.pdf

Holland, P. W. (2008). Causation and race. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, white methods: Racism and methodology*, (pp. 93-109). Lanham: Rowman & Littlefield.

Holmes, C. C., & Mallick, B. K. (1998). Bayesian radial basis functions of variable dimension. *Neural Computation*, *10*(5), 1217-1233. doi: http://dx.doi.org/10.1162/089976698300017421

Holmes, C. C., & Mallick, B. K. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, *98*(462), 352-368. doi: 10.1198/016214503000143

Huang, H. C., & Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics & Data Analysis*, *22*(2), 159-175. Retrieved from https://www.researchgate.net/profile/Hsin-Cheng_Huang/publication/223074020_Spatio-Temporal_Prediction_of_Snow_Water_Equivalent_Using_the_Kalman_Filter/links/54aca93c0cf21c47713a4c06.pdf

Isham, V. (2010). Spatial point process models. In A. E. Gelfand, P. J. Diggle, P. Guttorp, & M. Fuentes (Eds.), *Handbook of spatial statistics* (pp. 283-298). Boca Raton, FL: CRC Press.

Karabatsos, G., & Walker, S. G. (2012). A latent-variable Bayesian nonparametric regression model. *arXiv:1212.3712*, 1-23. Retrieved from http://arxiv.org/abs/1212.3712

Karabatsos, G. (2015a). Fast marginal likelihood estimation of the ridge parameter in ridge regression and generalized ridge regression for big data. *arXiv:1409.2437*, 1-44. Retrieved from http://arxiv.org/abs/1409.2437

Karabatsos, G. (2015b, October 13). *Regression and HLM* [PDF document].

Karabatsos, G. (2016, in press). A menu-driven software package of Bayesian nonparametric (and parametric) mixed models for regression analysis and density estimation. *Behavior Research Methods*. doi:10.3758/s13428-016-0711-7

Kohn, R., Smith, M., & Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, *11*(4), 313-322. doi: 10.1023/A:1011916902934

Kuziemko, I. (2006). Using shocks to school enrollment to estimate the effect of school size on student achievement. *Economics of Education Review*, *25*(1), 63-75. doi: 10.1016/j.econedurev.2004.10.003

Kwan, C. C. Y. (2011). An introduction to shrinkage estimation of the covariance matrix: A pedagogic illustration. *Spreadsheets in Education (eJSiE)*, *4*(3), 1-32. Retrieved from http://epublications.bond.edu.au/ejsie/vol4/iss3/6/

Lee, D-J., & Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling, 11*(1), 49-69. doi: 10.1177/1471082X1001100104

Lee, V. E., & Smith, J. B. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education*, 164-187. Retrieved from http://files.eric.ed.gov/fulltext/ED346617.pdf

Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research*, *79*(1), 464-490. doi: 10.3102/0034654308326158

Levy, R. (2012). *Probabilistic models in the study of language*. Retrieved from http://idiom.ucsd.edu/~rlevy/pmsl_textbook/book_draft.pdf

Li, Q., & Lin, N. (2010). The Bayesian elastic net. *Bayesian Analysis, 5*, 151-170. doi: 10.1214/10-BA506

Looking closer at closings. (2013, December 20). Retrieved from http://education.uic.edu/news-events/news/looking-closer-closings#.VitrOHvGpBo

Magee, L. (1998). Nonlocal behavior in polynomial regressions. *American Statistician, 52*, 20-22. doi: 10.1080/00031305.1998.10480531

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246-1266. doi: 10.2113/gsecongeo.58.8.1246

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*, 17-23. doi: 10.2307/2332142

Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). Boston, MA: PWS-Kent.

National Research Council (2006). *Learning to think spatially: GIS as a support system in K-12 education*. Washington, DC: National Academies Press.

Nychka, D. W. (2000). Spatial process estimates as smoothers. In M. G. Schimek (Ed.), *Smoothing and regression: Approaches, computation, and application* (pp. 393-424). New York, NY: John Wiley & Sons, Inc.

Paciorek, C., & Schervish, M. (2004). Nonstationary covariance functions for Gaussian process regression. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in Neural Information Processing Systems 16* (pp. 273–280). Cambridge, MA: MIT Press.

Prado, R., & West, M. (2010). *Time series: Modeling, computation, and inference*. Boca Raton, FL: CRC Press.

Radinsky, J., & Waitoller, F. R. (2013). *Chicago Public Schools actions: Impact on students.* Chicagoland Researchers and Advocates for Transformative Education, Chicago.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York, NY: Springer.

Ridge regression [PDF document]. (2015). Retrieved from http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf

Roberto, E. (2015, September 15). The spatial context of residential segregation. *arXiv:1509.03678v1*, 1-27. Retrieved from https://arxiv.org/abs/1509.03678

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357. Retrieved from https://www.jstor.org/stable/2087176

Royle, J. A., & Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics, 4*(1), 29-56. Retrieved from http://www.jstor.org/stable/1400420

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*(5328), 918-924. doi: 10.1126/science.277.5328.918

Sampson, R. J. (2008). Moving to inequality: Neighborhood effects and experiments meet social structure. *American Journal of Sociology, 114*(1), 189-231. doi: 10.1086/589843

Savitz, N. V., & Raudenbush, S. W. (2009). Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology*, *39*(1), 151-183. doi: 10.1111/j.1467-9531.2009.01221.x

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464. Retrieved from http://www.jstor.org/stable/2958889

Smith, M., & Kohn, R. (1997). A Bayesian approach to nonparametric bivariate regression. *Journal of the American Statistical Association*, *92*(440), 1522-1535. doi: 10.1080/01621459.1997.10473674

Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). London, England: John Churchill.

Stein, M. L. (2007). Spatial variation of total column ozone on a global scale. *The Annals of Applied Statistics*, 191-210. Retrieved from http://www.jstor.org/stable/4537428

Stroud, J. R., Müller, P., & Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B*, *63*(4), 673-689. doi: 10.1111/1467-9868.00305

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423. doi: 10.1111/1467-9868.00293

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(2), 234-240. doi: 10.2307/143141

Wahba, G. (1990). *Spline models for observational data*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Weeks, J. R. (2004). The role of spatial analysis in demographic research. In M. F. Goodchild, & D. G. Janelle, (Eds.), *Spatially integrated social science* (pp. 381-399). New York, NY: Oxford University Press.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). New York, NY: Springer.

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, *84*(6), 1382-1394. doi: 10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2

Wikle, C. K. (2010). Hierarchical modeling with spatial data. In A. E. Gelfand, P. J. Diggle, P. Guttorp, & M. Fuentes (Eds.), *Handbook of spatial statistics* (pp. 89-106). Boca Raton, FL: CRC Press.

Wikle, C. K., Berliner, L. M., & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, *5*(2), 117-154. doi: 10.1023/A:1009662704779

Wikle, C. K., & Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, *86*(4), 815-829. Retrieved from http://www.jstor.org/stable/2673587

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago, IL: University of Chicago Press.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95-114. doi: 10.1111/1467-9868.00374

Xia, G., Miranda, M. L., & Gelfand, A. E. (2006). Approximately optimal spatial design approaches for environmental health data. *Environmetrics*, *17*(4), 363-385. doi: 10.1002/env.775

Xiong, S. & Joseph, V. R. (2013). Regression with outlier shrinkage. *Journal of Statistical Planning and Inference, 143*(11), 1988-2001. doi: 10.1016/j.jspi.2013.06.007

Yeomans, M. (2015, July 7). What every manager should know about machine learning. *Harvard Business Review*. Retrieved from https://hbr.org/2015/07/what-every-manager-should-know-about-machine-learning

Zimmerman, D. L., & Stein, M. (2010). Classical geostatistical models. In A. E. Gelfand, P. J. Diggle, P. Guttorp, & M. Fuentes (Eds.), *Handbook of spatial statistics* (pp. 29-44). Boca Raton, FL: CRC Press.

APPENDICES

APPENDIX A

BAYESIAN RIDGE REGRESSION (VERSION 20-JUNE-2015) (Copyright 2014, George Karabatsos)

POSTERIOR DISTRIBUTION: DESCRIPTIVE STATISTICS

Fast ridge regression

y_i | x_i ~ f(y|x_i), i = 1,...,n
f(y|x) = normal(y | x'beta, sigma^2)
beta | sigma^2 ~ Normal(0, sigma^2*(1/lambdahat)*I_p)
sigma^2 ~ InverseGamma(eps, eps)
lambdahat: Marginal Maximum Likelihood Estimate (MMLE) of lambda (penalty).
Automatically, y is centered to have mean 0, and
each predictor (x) variable is rescaled to have mean 0 and variance 1.

Dependent Variable (Y):
Closed for 2013-2014

Sample Size: n = 18506

The p = 121 covariates are listed in Posterior Summaries table.

Censor indicators of Y (lower- & upper-bounds, resp.)
< None >

Observation Weight Variable:
Each observation has weight 1

Working Directory:
Data File: CPS ISBE Data-Final 4-8-16-No Charter-Open in 2012 (No 2012 or 2013)-Demo.DAT
Model File: Frr 6-19-2016 23h42m18s CPS ISBE.model
Text Output File: POSTERIOR SUMMARY Frr 6-19-2016 23h42m18s CPS ISBE.txt
Coefficients File: COEFFICIENTS Frr 6-19-2016 23h42m18s CPS ISBE.txt
Coeff. Covariances File: COV COEFF Frr 6-19-2016 23h42m18s CPS ISBE.txt
Residual File: RESIDUALS Frr 6-19-2016 23h42m18s CPS ISBE.txt

---

APPENDIX A (continued)

=============================
RESULTS OF THE DATA ANALYSIS
=============================

Posterior Predictive Model Fit Statistics and Penalty Estimates

---

Stat.
Model posterior predictive SSE D(m) = 2697.950
Model SSE fit to data Gof(m) = 1345.852
Penalty (predictive variance) P(m) = 1352.098
Proportion of variance explained, R squared = 0.150
Model log marginal likelihood = 4251.434
Model loglikelihood = -2006.159
Model Deviance = -2*loglikelihood = 4012.318
Model d.f. (effective # of parameters) = 40.845
BIC = Deviance + log(n)*d.f. = 4413.651
AIC = Deviance + 2*d.f. = 4094.008
GCV = mean[{(y - E[y|X])./(1-df/n)}.^2] = 0.073
Lambda (shrinkage penalty) estimate = 4.648
Coefficient prior variance v = 1/lambda = 0.215
Error variance sigma^2 (posterior mean) = 0.073
sigma^2 (posterior s.d.) = 0.001

Standardized Residuals (z_i) of Dependent Variable Responses:

Min	5%	10%	25%	50%	75%	90%	95%	Max
-2.254	-0.920	-0.810	-0.549	-0.186	0.122	0.335	2.873	3.521

1687 (9.1%) of all n = 18506 observations are outliers (with |z_i| > 2).

Computation Times:
Initialization: 0.16117 seconds.
Estimation algorithm: 0.30516 seconds.
Total computation time: 0.54396 seconds.

Column labels for the output table below:

beta: Standardized Coefficients (posterior mean)
(based on y zero-mean centered, and X variables rescaled to mean 0 and variance 1).
If y is a vector of z-scores (with mean zero and variance 1), then beta gives the coefficients

APPENDIX A (continued)

on a correlation scale. Furthermore, the beta coefficients strictly range from -1 to +1 when all of the X variables (columns) are uncorrelated.

SD: Posterior standard deviation of standardized coefficient.

PP1SD: Posterior probability that the standardized coefficient is within 1 standard deviation of 0.

Intervals: The 50% posterior (interquartile) interval of beta is given by the 25% and 75% percentiles. The 95% posterior interval of beta is given by the 2.5% and 97.5% percentiles.

Significance: A covariate is a "significant" predictor when: zero lies outside the 50% interval, or when PP1SD < .50.

Click the Posterior Summaries button in order to generate additional Love plots and/or box plots.

TABLE XV

MARGINAL POSTERIOR SUMMARY ESTIMATES

Covariate	ID	beta	SD	PP1SD	25%	75%	2.50%	97.50%
Intercept	0	0	0	1	0	0	0	0
2011-2012 Probation	1	0.051	0.002	0	0.05	0.053	0.047	0.056
2012-2013 Space Use Status	2	0.029	0.002	0	0.027	0.03	0.025	0.033
TP:Reading0	3	-0.028	0.09	0.66	-0.088	0.033	-0.204	0.149
TP:Reading23	4	-0.089	0.069	0.372	-0.136	-0.043	-0.224	0.046
TP:Reading30.2577	5	-0.004	0.089	0.682	-0.064	0.056	-0.179	0.171
TP:Reading34.3115	6	0.042	0.097	0.639	-0.023	0.108	-0.148	0.233
TP:Reading37.5	7	0.016	0.101	0.677	-0.052	0.084	-0.182	0.213
TP:Reading40	8	-0.016	0.102	0.677	-0.084	0.053	-0.216	0.185
TP:Reading42.5	9	0.036	0.102	0.653	-0.033	0.105	-0.164	0.236
TP:Reading45	10	0.044	0.104	0.64	-0.026	0.114	-0.159	0.248
TP:Reading47	11	-0.005	0.106	0.682	-0.076	0.067	-0.212	0.203
TP:Reading48.7	12	-0.048	0.107	0.636	-0.12	0.024	-0.257	0.161
TP:Reading50	13	-0.075	0.107	0.574	-0.147	-0.002	-0.285	0.135
TP:Reading52	14	-0.058	0.106	0.613	-0.13	0.013	-0.266	0.15
TP:Reading53.7	15	0.009	0.106	0.681	-0.063	0.081	-0.2	0.217
TP:Reading55	16	0.053	0.107	0.625	-0.019	0.126	-0.156	0.263

TP:Reading56.6038	17	0.096	0.106	0.512	0.024	0.167	-0.113	0.304
TP:Reading58	18	0.039	0.106	0.651	-0.033	0.11	-0.17	0.247
TP:Reading60	19	-0.028	0.107	0.666	-0.1	0.044	-0.238	0.182
TP:Reading61	20	-0.053	0.108	0.627	-0.125	0.02	-0.264	0.158
TP:Reading62.2	21	-0.044	0.108	0.644	-0.116	0.029	-0.255	0.167
TP:Reading64	22	0.006	0.108	0.682	-0.067	0.078	-0.205	0.217
TP:Reading65.1	23	0.035	0.107	0.658	-0.038	0.107	-0.176	0.245
TP:Reading66.7	24	0.026	0.108	0.669	-0.047	0.099	-0.186	0.237
TP:Reading68	25	-0.011	0.108	0.68	-0.084	0.062	-0.223	0.201
TP:Reading69.2	26	-0.013	0.108	0.679	-0.085	0.06	-0.224	0.199
TP:Reading71	27	-0.013	0.108	0.679	-0.086	0.061	-0.225	0.2
TP:Reading72	28	0.015	0.109	0.678	-0.059	0.088	-0.198	0.227
TP:Reading73.7	29	0.003	0.108	0.683	-0.07	0.076	-0.21	0.215
TP:Reading75	30	-0.002	0.109	0.683	-0.075	0.071	-0.215	0.211
TP:Reading76.5	31	-0.013	0.109	0.679	-0.086	0.06	-0.226	0.2
TP:Reading78	32	-0.01	0.109	0.681	-0.084	0.063	-0.224	0.203
TP:Reading79.5654	33	-0.001	0.109	0.683	-0.074	0.072	-0.214	0.212
TP:Reading81	34	0.016	0.109	0.678	-0.058	0.089	-0.198	0.229
TP:Reading82.8	35	0.023	0.109	0.671	-0.05	0.097	-0.19	0.237
TP:Reading84.1	36	0.015	0.108	0.678	-0.058	0.088	-0.197	0.228
TP:Reading86	37	0.003	0.107	0.683	-0.07	0.075	-0.207	0.213
TP:Reading88	38	-0.003	0.105	0.682	-0.074	0.068	-0.21	0.204
TP:Reading90.6885	39	-0.008	0.103	0.681	-0.078	0.061	-0.21	0.193
TP:Reading93.5	40	-0.034	0.102	0.657	-0.102	0.035	-0.234	0.166
TP:Reading97.3962	41	-0.046	0.107	0.64	-0.118	0.027	-0.255	0.164
TP:Reading100	42	-0.015	0.104	0.678	-0.085	0.056	-0.219	0.19
TP:Math0	43	0.026	0.087	0.661	-0.032	0.085	-0.144	0.196
TP:Math25.6	44	-0.158	0.057	0.036	-0.196	-0.12	-0.269	-0.047
TP:Math35	45	-0.027	0.078	0.654	-0.08	0.026	-0.181	0.127
TP:Math40.5	46	0.088	0.092	0.491	0.026	0.15	-0.092	0.268
TP:Math44	47	-0.007	0.096	0.681	-0.072	0.057	-0.195	0.18
TP:Math47.5	48	-0.04	0.098	0.643	-0.106	0.026	-0.232	0.152
TP:Math50	49	-0.022	0.1	0.671	-0.09	0.045	-0.219	0.174
TP:Math52.6	50	0.006	0.1	0.682	-0.061	0.074	-0.191	0.203
TP:Math55	51	0.02	0.102	0.673	-0.048	0.089	-0.179	0.22
TP:Math57	52	0.01	0.103	0.681	-0.06	0.079	-0.192	0.211
TP:Math59	53	0.002	0.103	0.683	-0.068	0.071	-0.2	0.204
TP:Math61	54	-0.03	0.104	0.663	-0.1	0.04	-0.234	0.174
TP:Math62.7	55	0.001	0.105	0.683	-0.069	0.072	-0.204	0.207
TP:Math64.2	56	0.057	0.105	0.614	-0.014	0.128	-0.149	0.263
TP:Math66	57	0.039	0.106	0.65	-0.032	0.11	-0.168	0.246

TP:Math67.7	58	-0.036	0.106	0.655	-0.108	0.035	-0.244	0.172
TP:Math69	59	-0.048	0.107	0.635	-0.121	0.024	-0.259	0.162
TP:Math70.5	60	-0.023	0.108	0.671	-0.096	0.049	-0.235	0.188
TP:Math72	61	0.013	0.109	0.679	-0.061	0.087	-0.201	0.227
TP:Math73	62	0.018	0.11	0.676	-0.056	0.092	-0.198	0.233
TP:Math74.2	63	0.02	0.11	0.675	-0.054	0.095	-0.196	0.236
TP:Math75.6	64	0.017	0.11	0.677	-0.058	0.091	-0.2	0.233
TP:Math76.6	65	0.011	0.111	0.68	-0.063	0.086	-0.206	0.228
TP:Math77.9	66	0.001	0.112	0.683	-0.075	0.076	-0.218	0.219
TP:Math79	67	-0.006	0.112	0.682	-0.082	0.069	-0.226	0.214
TP:Math80	68	-0.023	0.112	0.673	-0.099	0.053	-0.243	0.198
TP:Math81	69	-0.009	0.112	0.681	-0.085	0.067	-0.229	0.212
TP:Math82	70	-0.01	0.112	0.681	-0.085	0.066	-0.229	0.21
TP:Math83.1	71	0.003	0.112	0.682	-0.072	0.079	-0.215	0.222
TP:Math84.6	72	0.015	0.111	0.678	-0.06	0.09	-0.203	0.233
TP:Math85.8654	73	0.019	0.111	0.676	-0.056	0.094	-0.199	0.237
TP:Math87	74	0.011	0.111	0.68	-0.064	0.086	-0.207	0.229
TP:Math88.2	75	0.001	0.111	0.683	-0.073	0.076	-0.216	0.219
TP:Math90	76	-0.009	0.11	0.681	-0.083	0.066	-0.225	0.208
TP:Math91	77	-0.012	0.11	0.68	-0.086	0.062	-0.227	0.203
TP:Math93	78	-0.005	0.109	0.682	-0.078	0.068	-0.218	0.208
TP:Math95	79	-0.019	0.108	0.675	-0.092	0.054	-0.231	0.193
TP:Math97	80	-0.043	0.109	0.645	-0.116	0.03	-0.256	0.17
TP:Math100	81	-0.04	0.103	0.648	-0.109	0.03	-0.241	0.162
TP: Enrollment 34	82	-0.393	0.088	0	-0.452	-0.333	-0.565	-0.221
TP: Enrollment 211	83	0.32	0.097	0.011	0.255	0.386	0.129	0.512
TP: Enrollment 244	84	0.228	0.105	0.118	0.158	0.299	0.023	0.433
TP: Enrollment 265	85	0.075	0.107	0.573	0.003	0.148	-0.135	0.286
TP: Enrollment 285	86	0.101	0.108	0.501	0.028	0.174	-0.112	0.313
TP: Enrollment 303	87	0.076	0.109	0.575	0.002	0.149	-0.138	0.29
TP: Enrollment 320	88	0.045	0.109	0.643	-0.029	0.119	-0.169	0.259
TP: Enrollment 336.2692	89	-0.033	0.109	0.66	-0.107	0.04	-0.248	0.181
TP: Enrollment 355	90	-0.096	0.109	0.519	-0.169	-0.022	-0.31	0.118
TP: Enrollment 369	91	-0.086	0.109	0.548	-0.159	-0.012	-0.299	0.128
TP: Enrollment 386	92	-0.023	0.109	0.672	-0.096	0.051	-0.236	0.191
TP: Enrollment 404	93	0.065	0.109	0.602	-0.009	0.138	-0.148	0.278
TP: Enrollment 421	94	0	0.109	0.683	-0.074	0.073	-0.214	0.214
TP: Enrollment 437	95	-0.067	0.109	0.598	-0.14	0.007	-0.28	0.147
TP: Enrollment 452	96	-0.024	0.109	0.671	-0.098	0.05	-0.238	0.19
TP: Enrollment 467	97	0.059	0.109	0.614	-0.014	0.133	-0.154	0.272
TP: Enrollment 484	98	0.073	0.108	0.579	0.001	0.146	-0.138	0.285

TP: Enrollment 501	99	-0.031	0.108	0.663	-0.103	0.042	-0.242	0.181
TP: Enrollment 519	100	-0.169	0.108	0.279	-0.242	-0.096	-0.38	0.042
TP: Enrollment 536	101	-0.103	0.108	0.491	-0.176	-0.031	-0.315	0.108
TP: Enrollment 551	102	0.011	0.107	0.68	-0.061	0.084	-0.199	0.222
TP: Enrollment 568	103	0.109	0.107	0.472	0.036	0.181	-0.101	0.319
TP: Enrollment 586	104	0.078	0.106	0.564	0.006	0.15	-0.131	0.287
TP: Enrollment 603	105	0.016	0.106	0.677	-0.055	0.087	-0.191	0.223
TP: Enrollment 624	106	-0.001	0.104	0.683	-0.072	0.069	-0.206	0.203
TP: Enrollment 645	107	0.018	0.103	0.675	-0.051	0.088	-0.184	0.22
TP: Enrollment 669	108	-0.003	0.103	0.682	-0.072	0.066	-0.204	0.198
TP: Enrollment 690	109	-0.001	0.101	0.683	-0.069	0.068	-0.199	0.198
TP: Enrollment 714	110	-0.047	0.099	0.63	-0.114	0.02	-0.24	0.146
TP: Enrollment 748	111	0.023	0.093	0.668	-0.04	0.086	-0.16	0.206
TP: Enrollment 787.6538	112	0.084	0.093	0.509	0.021	0.147	-0.098	0.266
TP: Enrollment 824.1923	113	-0.055	0.088	0.595	-0.115	0.005	-0.228	0.118
TP: Enrollment 877	114	0.003	0.086	0.682	-0.054	0.061	-0.165	0.171
TP: Enrollment 924	115	-0.036	0.079	0.633	-0.09	0.017	-0.192	0.119
TP: Enrollment 1004	116	0.015	0.069	0.671	-0.031	0.062	-0.119	0.15
TP: Enrollment 1096.0385	117	0.022	0.059	0.65	-0.018	0.061	-0.093	0.137
TP: Enrollment 1245	118	-0.02	0.061	0.657	-0.062	0.021	-0.141	0.1
TP: Enrollment 1352	119	-0.009	0.058	0.676	-0.048	0.03	-0.123	0.104
TP: Enrollment 1560	120	-0.017	0.026	0.583	-0.034	0	-0.068	0.033
TP: Enrollment 4278	121	0.297	0.081	0.004	0.242	0.351	0.138	0.455

APPENDIX B

Guide to Running the Bayesian Linear Spatio-Temporal Model in the Bayesian Ridge Regression Software (Karabatsos, 2016)

Note: There are excellent instructions for general data analysis in the "Help" menu in the Bayesian Ridge Regression Software (Karabatsos, 2016). The following instructions pertain to the specific model presented in this thesis.

1. <u>To import data into the software</u>:
    a. Under the "File" menu, select "Import, save, and open data file," to import the comma-delimited data file (.csv) into the software
    b. Once the data file is imported, the software will prompt the user to save the file as a data (.dat) file. This new file can be opened directly into the software using the "Open data (.dat) file" option under the "File" menu

2. <u>To create univariate thin-plate splines</u>:
    a. Under the "Modify Data Set" menu, select "Construct spline or weighted covariates"
    b. Then select "Univariate thin-plate (TP) splines"
    c. Click on the variable of interest and click "OK"
    d. Select the number of knots for the spline, based on a number of equally-spaced quantiles (for many of the variables in this thesis, we selected 40 quantiles). Click "OK"
    e. The indicated number of spline variables are now added to the dataset, using the naming convention "TP: <variable name>"

3. <u>To create interaction terms (for all variables, including spline terms)</u>:
    a. Under the "Modify Data Set" menu, select "Construct interaction and/or polynomial covariates"
    b. Then select "Construct interaction terms"
    c. Click on the variables of interest (up to 50 at a time) and click "OK"
    d. The new interaction term(s) are now added to the dataset, using the naming convention "<variable 1>*<variable 2>" (e.g. latitude*longitude)
    e. To create a 3-way interaction (e.g. latitude*longitude*time), follow the above instructions to create the 2-way interaction (e.g. latitude*longitude)
        i. Then repeat to create interaction terms between the 2-way interaction and the third variable of interest (e.g. time)

4. <u>To run the analysis:</u>
    a. Click on "Specify New Model"
    b. Select "Fast ridge regression" and click "OK"
    c. Select the dependent variable (the software allows for a single dependent variable) and click "OK"
    d. Select the independent variable(s) (referred to as covariates/predictor variables) and click "OK"

APPENDIX B (continued)

e. The selected Dependent Variable and Covariate(s) are displayed on the right-hand side of the screen. This is a quick way to double-check that the variables are correct prior to running the analysis
f. After verifying that everything looks accurate, click on "Run Posterior Analysis"
g. Depending on the number of variables included (and your computer), a small window may pop up with a notification that the analysis is running
h. Once the analysis is complete, a Posterior Summary (as shown in Appendix A) pops up

5. <u>To detect outliers:</u>
   a. The Posterior Summary provides information on how many outliers were detected, using the threshold of a standardized residual (z) greater than 2. This information is located under "Standardized Residuals (z_i) of Dependent Variable Responses"
   b. An additional output file is generated titled "RESIDUALS Frr <name of dataset>." This file indicates which of the individual cases were outliers

6. <u>To correct for outliers</u>:
   a. Under the "Modify Data Set" menu, select "Basic Data Editing"
   b. Then select "Add Case (row) ID variable"
   c. To generate the dummy codes [per the Xiong & Joseph (2013) outlier correction method], select the "Modify Data Set" menu and then click on "Dummy/binary code variables"
   d. Then select "Dummy/binary code a categorical variable (positive integer valued)
   e. In the screen that says "What type of binary coding?" select "1 versus 0"
   f. Select "Case ID" as the categorical variable
   g. A dummy coded variable for each case is now added to the dataset
   h. Repeat the steps for running the analysis, but this time include the dummy coded variables *for those cases that were determined to be outliers*
   i. The newly generated Posterior Summary provides the results for the analysis, excluding the outlier variables

VITA

SHANNON MILLIGAN
1748 N. Washtenaw Ave., Chicago, IL
918-688-6921 · Shannon.milligan@gmail.com

EDUCATION

Doctor of Philosophy in Educational Psychology **2016**
Focus Area: Measurement, Evaluation, Statistics, and Assessment
University of Illinois at Chicago, Chicago, Illinois
Dissertation title: "Assessment of a Spline-Based Spatio-Temporal Model for Use With Educational Datasets"
Advisor: George Karabatsos, PhD

Master of Science in Educational Leadership **2008**
Focus Area: College Student Development
Oklahoma State University, Stillwater, Oklahoma

Master of Science in Research, Evaluation, Measurement, and Statistics **2006**
Oklahoma State University, Stillwater, Oklahoma

Bachelor of Science in Psychology **2004**
Oklahoma State University, Stillwater, Oklahoma

RESEARCH EXPERIENCE

Undergraduate Research Assistant **2003**
Department of Psychology
Oklahoma State University

- Studied the effect of price of cigarettes on nicotine consumption and the interactions of various forms of chewing gum with nicotine deprivation
- Conducted library research; actively recruited subjects; and administered both surveys and carbon monoxide detector tests

Undergraduate Research Assistant **Fall 2003**
Department of Psychology
Oklahoma State University

- Studied the ideal, actual, and ought selves with respect to the Big Five personality traits
- Administered computerized and paper and pencil measures; checked data for accuracy; entered the data into the Idiogrid software; and analyzed the data

PROFESSIONAL EXPERIENCE

Coordinator of Assessment, Faculty Center for Ignatian Pedagogy **2010-present**
Loyola University Chicago

- Coordinate annual assessment reporting process for all academic programs at Loyola University Chicago
- Advise academic programs on the creation of assessment plans
- Advise co-curricular programs on assessment processes
- Collect and analyze data from FCIP assessments
- Create surveys for use in FCIP assessment
- Co-coordinate the Assessment Certificate Program professional development program
- Serve on the Loyola Experience Steering Group
- Serve on the Division of Student Development Assessment Working Group
- Co-teach the Online Teaching Course
- Supervise one graduate intern

Resident Director for Administration and Assessment **2008-2010**
University of Illinois at Chicago

- Developed and implement surveys and other assessment tools for the Housing department
- Created assessment reports by the Housing department for use in program improvement
- Collected and analyze data from Housing assessments, including annual Housing Survey with over 500 respondents
- Supervised one Graduate Assistant, one Student Staff Supervisor, and office student workers
- Coordinated administrative processes including contract cancellation appeals and housing assignments
- Served on crisis duty rotation for campus
- Coordinated the student employee hiring process and payroll for Housing
- Served as Technical Coordinator of MAP-Works assessment with Educational Benchmarking, Inc.

TEACHING EXPERIENCE

Online Teaching Course **2011-present**
Loyola University Chicago

- Work in a teaching team of staff members to deliver curriculum on online teaching and general pedagogy to Loyola faculty, staff, and graduate students
- Develop and teach instructional units on assessment and building assessment strategies into the classroom

Student Development Training for Resident Assistants **Fall 2007**
Oklahoma State University

- Worked in a teaching team consisting of another graduate student and an undergraduate student to teach Resident Assistants a curriculum based on student development theory in relation to the RA position

Teaching Assistant **Spring 2004**
Department of Psychology
Oklahoma State University

- Held office hours; created and maintained grade and attendance records; and assisted with lecture portion of class

Teaching Assistant **Fall 2002**
Department of Psychology
Oklahoma State University

- Assisted graduate TA in lab portion of class, which focused on the use of SPSS

PUBLICATIONS

Hidding, G. J., Scheidenhelm, C. L., & Milligan, S. M. (2014). A generalized rubric for reflective assignments with the Ignatian Pedagogy Paradigm. *Journal of Jesuit Business Education*, *5*(1), 89-104.

PROFESSIONAL PRESENTATIONS

Milligan, S. (2016, April). Writing and revising program learning outcomes. Invited workshop facilitator. American Islamic College, Chicago, IL.

Milligan, S. & Sweet, J. (2016, January). Faculty development week assessment workshop. Invited workshop facilitators. Kennedy-King College, Chicago, IL.

Sweet, J. & Milligan, S. (2016, June). Responding constructively to criticism in assessment. Association for the Assessment of Learning in Higher Education Conference, Milwaukee, WI.

Sweet, J. & Milligan, S. (2015, June). Building agency in assessment through an inter-institutional program. Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Milligan, S. & Sweet, J. (2014, September). How do we know what students are learning? Invited workshop facilitators. Chicago State University, Chicago, IL.

Hidding, G., Scheidenhelm, C. & Milligan, S. (2013, July). A generalized rubric for reflective assignments with the Ignatian Pedagogy Paradigm. Colleagues in Jesuit Business Education Conference, St. Louis, MO.

Sweet, J. & Milligan, S. (2013, June). Integrating assessment across the curricular and co-curricular domains. Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Kehoe, A., Milligan, S., Stark, C. & Sweet, J. (2013, March). ePortfolios and assessment. Invited panelist. Chicago State University, Chicago, IL.

Arkalgund, R., Sweet, J. & Milligan, S. (2012, June). An ontological framework for designing the assessment of learning outcomes. Association for the Assessment of Learning in Higher Education Conference, Albuquerque, NM.

Milligan, S. & Kehoe, A. (2012, March). How to implement ePortfolios on your campus. Higher Learning Commission Conference, Chicago, IL.

Scheidenhelm, C. & Milligan, S. (2012, March). Introduction to assessment. Invited workshop facilitators. Chicago State University, Chicago, IL.

Green, P., Kehoe, A., Milligan, S. & Scheidenhelm, C. (2012, February). Assessment, ePortfolios, and experiential learning: Expanding the model. Association of American Colleges and Universities General Education and Assessment Meeting, New Orleans, LA.

Milligan, S. & Sweet, J. (2011, April). Improving survey design. Invited Lecturer, ELPS 431, Evaluation in Higher Education, Loyola University Chicago, Chicago, IL.

Sweet, J. & Milligan, S. (2010, October). Improving assessment through communities of practice. IUPUI Assessment Institute, Indianapolis, IN.

Kessler, D. & Milligan, S. (2008, March). Guiding those in the middle: Meeting the needs of bisexual students. National Association of Student Personnel Administrators Conference, Boston, MA.

Barraclough, J., Milligan, S., Smith, E. & Swigert, L. (2007, May). Hazing: Damned if you do, damned if you do nothing. Southwest Association of College and University Housing Officers Conference, Oklahoma City, OK.

AWARDS AND HONORS

2004-2005 Student Staff Member of the Year **May 2005**
Oklahoma State University

2003-2004 Outstanding Senior in Psychology **May 2004**
Oklahoma State University