Spring 2017

Extending time series forecasting methods using functional principal components analysis

Isaac Michael Wagner-Muns

EXTENDING TIME SERIES FORECASTING METHODS USING FUNCTIONAL PRINCIPAL COMPONENTS ANALYSIS

by

ISAAC MICHAEL WAGNER-MUNS

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

in

SYSTEMS ENGINEERING

2017

Approved by

Dr. Ivan Guardiola
Dr. Ruwen Qin
Dr. Cihan Dagli

ABSTRACT

Traffic volume forecasts are used by many transportation analysis and management systems to better characterize and react to fluctuating traffic patterns. Most current forecasting methods do not take advantage of the underlying functional characteristics of the time series to make predictions. This paper presents a methodology that uses Functional Principal Components Analysis (FPCA) to create smooth and differentiable daily traffic forecasts. The methodology is validated with a data set of 1,813 days of 15 minute aggregated traffic volume time series. Both the FPCA based forecasts and the associated prediction intervals outperform traditional Seasonal Autoregressive Integrated Moving Average (SARIMA) based methods.

ACKNOWLEDGMENTS

Many thanks to Ivan Guardiola for his patience and guidance.

Thanks to my parents Anne Wagner and Raleigh Muns for raising me to be stubborn and inquisitive.

TABLE OF CONTENTS

Page

ABSTRACT iii

ACKNOWLEDGMENTS iv

LIST OF ILLUSTRATIONS vii

LIST OF TABLES viii

SECTION

1. INTRODUCTION 1

2. LITERATURE REVIEW 4

  2.1. TRAFFIC DATA ANALYSIS 4

    2.1.1. Imputation 4

    2.1.2. Pattern Recognition 5

    2.1.3. Data Reduction 6

    2.1.4. Forecasting 6

  2.2. FUNCTIONAL DATA ANALYSIS 7

3. TRAFFIC DATA DISCUSSION 9

  3.1. DATA SOURCE 9

  3.2. EXAMINING THE DATA 9

4. FPCA FORECASTING METHODOLOGY 11

  4.1. FIT CONTINUOUS BASIS FUNCTIONS TO DISCRETE TIME SERIES . 11

4.2. GENERATE FUNCTIONAL PRINCIPAL COMPONENTS AND DAILY SCORES 13

4.3. FIT SARIMA MODEL TO PRINCIPAL COMPONENT SCORES 16

4.4. GENERATE FORECASTED FUNCTION AND PREDICTION INTERVALS 16

5. FPCA FORECASTING RESULTS 18

5.1. FITTING BASIS FUNCTIONS TO TRAFFIC TIME SERIES 18

5.2. GENERATION OF FUNCTIONAL PRINCIPAL COMPONENTS 19

5.3. FITTING SARIMA MODELS TO COMPONENT SCORES 23

5.4. GENERATION OF FORECASTS AND PREDICTION INTERVALS 26

6. SUMMARY AND CONCLUSIONS 31

REFERENCES 35

VITA 40

LIST OF ILLUSTRATIONS

Figure Page

4.1 Outline of FPCA Forecasting Methodology. 12

5.1 Histogram of Fitting Residuals. 20

5.2 A sample day's functional fit plotted with its associated discrete time series. 20

5.3 Scree Plot of Number of Components vs. Quantity of Described Variation. 21

5.4 Demonstration of how FPCA harmonics modify the mean function. 22

5.5 Because accidents are rare, they are not characterized within the first three principal components of the data. 23

5.6 3d scatter plot of FPCA scores colored by day of week (Mon - Blue, Tues - Magenta, Wed - Purple, Thurs - Orange, Fri - Yellow, Sat - Red, Sun - Green). 24

5.7 Partial autocorrelation of FPCA scores. 25

5.8 An example day close to the median forecast RMSE. 27

5.9 An example day from the highest (95% to 100%) quantile of L_2 Distance. This poor forecast was produced due to road construction on the highway impeding normal traffic flow [Star-Tribune, 2004]. 27

5.10 Comparison of forecasting methods. 90% prediction intervals for traditional SARIMA (in orange) and FPCA SARIMA (in blue). 29

5.11 Comparison of forecasting methods. 90% Prediction intervals for traditional SARIMA (in orange) and FPCA SARIMA (in blue). Note that after 45min to 1 hour (3–4 steps) ahead, the prediction intervals for traditional SARIMA are wider than for FPCA SARIMA. 30

LIST OF TABLES

Table Page

5.1 Summary of Residual Fitting Statistics. 19

5.2 Summary of Daily Error Properties. 26

5.3 Comparison of summary statistics from 365 residual samples. 28

SECTION

1. INTRODUCTION

As urban populations continue to grow, traffic problems become increasingly larger and more frequent. Traffic congestion is one of the most visible and frustrating traffic problems that people commonly experience. When the demand for a roadway exceeds its capacity, traffic speeds slow to a crawl. Traffic congestion affects very large numbers of people and can produce a significant "drag" on the economic activity of an urban area [Sweet, 2014]. Intelligent Transportation Systems (ITS) are used to manage the complexity of a traffic network and mitigate the negative impacts of traffic congestion. ITS utilize historical, live, and forecasted data from a wide variety sensors and locations to determine effective traffic management policies. Traffic volume is one the most prevalent and common types of data used by ITS to determine effective traffic control strategies [Zhang et al., 2011, Vlahogianni et al., 2014]. Specifically, traffic volume data is used as the primary input into traffic network simulations. These simulations are used to determine optimal routing, metering, traffic lights, and variable speed limit policies [Taniguchi and Shimamoto, 2004, Papageorgiou et al., 2003, Carlson et al., 2010]. Traffic control policies must be adaptive and flexible to changes in daily traffic volume profiles. Control policies during periods of congestion are very different than policies during free flowing periods. For these reasons, high quality forecasts are an important component of any effective traffic management system. From early applications of traditional time series methodologies by Ahmed and Cook [1979] to modern applications of support vector regression and simulated annealing methods by Hong [2011], the development of predictive analytical techniques and models that seek to increase forecast efficiency and accuracy remains an important area of research among practitioners and academics alike.

Traffic volume data must be measured before it can be quantified and characterized. Measurement of traffic volume is performed using many different methods, each with differing benefits and problems. In the past, traffic volume would be measured manually by having an observer count cars as they passed on a road. Manual measurement methods have fallen out of favor as the cost of automation has decreased. In its discussion of traffic sensing technologies, Chapter 9 of Dalgleish and Hoose [2008] states that inductive loops, axle detectors, "above ground" methods, and image processing methods are commonly used to measure traffic automatically. Inductive loop detectors are buried beneath the road surface. They are used to collect traffic volume measurements over long periods of time. Inductive loops detect vehicles using RF in a way similar to how metal detectors function. While inductive loop sensors use RF to measure traffic, axle detectors measure the pressure exerted by vehicles as they drive over the sensor. Axle detectors are somewhat less permanent than inductive loops, but can be very rapidly placed and removed. While both inductive loop and axle detector sensors are placed near the road surface, "above ground" detectors are placed above the road on bridges or gantries. Above ground detectors use a number of passive and active methods to count vehicles as they pass. Similar to above ground sensors, technologies based on image processing use cameras pointed at the road to count traffic. Unlike above ground sensors, image processing methods are more intelligent and analyze video feeds to count and categorize traffic. These popular methods of traffic measurement supply data that is the foundation of traffic analysis.

The dynamics of traffic volume arises from the interactions of many complex systems [Chowdhury et al., 2000, Helbing, 2001, Mahnke et al., 2005]. These complicated interactions make traffic volume measurements nonstationary, noisy, and chaotic. Despite its variability, traffic behavior is still regular. Much of the regularity of traffic volume comes from seasonal effects. From day to day, traffic volume profiles appear similar [Li et al., 2015, Williams et al., 1998]. Plots of traffic volume over the course of weekdays usually have two distinct peaks of morning and evening rush periods. The next seasonal effect in traffic has a weekly period. Although successive days share many similarities, there are stronger similarities found between days from successive weeks. For example, a traffic profile of a Monday will share similarities with the traffic of the Sunday preceding it [Williams et al., 1998]. However, it will share more similarities with the Monday profile taken a week previously. Finally, traffic volume has very long period yearly seasonal effects. Driver behavior changes with the weather conditions predominant to a particular season [Datla and Sharma, 2008]. The aggregated interactions of many independent drivers using a transportation system is the source of the emergent properties of transportation systems. Despite the complex behavior of traffic volume patterns, regularities in the data can be exploited to produce quality forecasts. Intelligently managing the complexity of a transportation system can significantly increase its efficiency and effectiveness.

The first section of the thesis will discuss the background surrounding traffic applications of time series techniques related to FDA. The next section will present a preliminary analysis of the traffic data used to justify the methodology. Next, it will describe the FDA forecasting methodology itself. The next part of the thesis will discuss the results of each step of the methodology and compare the quality of functional forecasts to forecasts produced by SARIMA. Finally, the thesis will conclude that FDA methods can greatly improve traffic analysis and forecasting.

2. LITERATURE REVIEW

2.1. TRAFFIC DATA ANALYSIS

Traffic data is necessary for making informed policy decisions about road systems. Intelligent transportation systems seek ways to manage and optimize traffic. As such, ITS need timely and accurate information about the state of the road network. The traffic patterns that ITS manage are quite complex. The behavior of traffic volume time series data comes from the complex system that emerges from the aggregated behavior of large numbers of people [Xiao et al., 2016]. Li et al. [2015] states that ITS handle traffic data in four distinct categories: imputation, pattern recognition, data reduction, and forecasting. Imputation is the process of replacing missing data in traffic time series caused by failures in sensors/recorders. Pattern recognition analyzes data to categorize and determine the regularities in observed traffic events. Data reduction is used to reduce the volume of the data while preserving the characteristics important for traffic analysis. Finally, forecasting is the method of making predictions of future traffic volumes based on present and historical observations. Although the primary subject of this work is forecasting, these categories are closely related. This literature review will give a brief overview of each category.

2.1.1. Imputation. Traffic volume data is collected from imperfect and unreliable sensors. Sometimes, sensors can break and output either incorrect measurements or no data at all. ITS make decisions based on the data they receive. Inference based on incorrect data will lead to incorrect management decisions. Many imputation methods have been implemented to address sensor reliability issues. In Chen et al. [2003], data from malfunctioning traffic volume sensors were found by their volume measurements. Imputation is performed by a linear regression using historical data and the data of neighboring valid sensors. Neural networks have also been used for accurate and effective imputation of traffic data. Lint et al. [2005] exploit the properties of a state space neural network to continue to produce useful outputs under the influence of missing data.

In addition to neural network based methods, more traditional methods from multivariate statistics have been used to approach the traffic imputation problem. Principal Components Analysis (PCA) is a statistical method that separates the dominant linearly correlated phenomena in a time series. Qu et al. [2009] uses maximum likelihood estimation to impute missing data based on the dominant modes of traffic variation determined using PCA. Although PCA is an excellent tool for characterizing variation, additional assumptions of continuity can be made to increase imputation quality. Assuming continuity means that nearby data points are assumed to be at least somewhat dependent on one another. Instead of directly analyzing the time series Chiou et al. [2014] assume that the discrete time series data are observations throughout time from a realization of a random continuous and differentiable function and use FPCA to impute the missing data. Chiou et. al. found the functional methodology to be more effective than traditional discrete PCA based methods.

2.1.2. Pattern Recognition. Traffic volume patterns behave differently depending on the season, day, weather, and many other factors. Furthermore, traffic patterns will change as the population and road demand changes. The ability to recognize and differentiate these traffic patterns is important for ITS decision making. Weijermars and Berkum [2005] demonstrate that supervised clustering based on knowledge of working versus non-working days produces useful results. It also demonstrates the existence of weekly seasonal effects in traffic behavior. Because PCA methodologies categorize linearly correlated events over time into separate principal components, it is useful for many pattern recognition tasks. Guardiola et al. [2014] examine control chart plots of traffic data FPCA scores to recognize long-term traffic trends. Shorter term pattern recognition is important for determining when data imputation is necessary. FPCA is used by Chiou et al. [2014] to classify outlier data into sensor failures or abnormal events. Outlier detection can also make use of forecasting methodologies. Guo et al. [2015] compare forecasts of traffic with observed traffic events, classifying data as abnormal if it differs significantly from the forecasts. Furthermore, Guo et al. [2015] demonstrate the close relationship between forecasting, pattern recognition, and imputation. All methodologies require an accurate model of traffic events but use it differently depending on the desired output.

2.1.3. Data Reduction. The volume of data in traffic analysis has massively grown as the cost of sensors and data storage has decreased in recent years. When these data become unwieldy and difficult to work with, methods that reduce the bulk while maintaining the important phenomena become invaluable. As data sets grow, so does the effects of the "curse of dimensionality" [Donoho, 2000, Bellman, 1957]. As dimension grows, differences in volume between relatively similar solids becomes greatly magnified. Methods discussed in Naymat [2015] such as random projection, downsampling, and averaging are all effective in reducing data set size. PCA based methods as described by Tran [2008] and Xing et al. [2015] are a more intelligent way of reducing data size. Unlike random projection, PCA projects the data onto a basis generated from the data itself. PCA methods find an orthonormal basis for the data by categorizing sets of linearly correlated events into independent basis components. These principal components are ordered based on how much of the total variation they characterize within the data. By truncating the basis to only the components that express a large amount of variation, PCA preserves much of the variation in the data while reducing dimensionality. Xing et al. [2015] specifically demonstrates the effectiveness PCA has for reducing traffic data volume. Reducing data volume not only leads to more economical storage of observations, but also enables better classification techniques by avoiding the 'curse' of dimensionality.

2.1.4. Forecasting. The necessity to forecast traffic volume data has existed almost as long as traffic volume time series data. Because of its simplicity and versatility, ARIMA has been applied to forecast many types of time series. Ahmed and Cook [1979] used ARIMA in 1979 to effectively forecast traffic volume. Since then, ARIMA has remained the baseline standard to which other methods are compared [Vlahogianni et al., 2014, Smith and Demetsky, 1997, Williams and Hoel, 2003]. Beginning in the 1990s, neural network (NN) methods have gained popularity for their effectiveness in traffic volume forecasting. Yun et al. [1998] found neural networks to be more effective than traditional time series forecasting methods, while other research, such as that from Smith and Demetsky [1997], found that NN methods were as effective as ARIMA but harder to correctly configure and train. Despite these drawbacks, neural networks have remained a popular foundation for traffic volume time series forecasting [Yin et al., 2002, Dia, 2001, Boto-Giralda et al., 2010]. More recently, methods from statistical learning theory such as support vector machines have been applied to traffic volume. Hong [2011] found SVR methods to be significantly more accurate than ARIMA methods. Although effective, Castro-Neto et al. [2009] showed that different forecasting methods have different levels of effectiveness of forecasting typical vs atypical traffic days. There has been extensive research in traffic volume forecasting methods over the last forty years. Many methods have been shown to be effective with differing costs and benefits.

2.2. FUNCTIONAL DATA ANALYSIS

Functional data analysis (FDA) is a field of statistics that works with functional (rather than multivariate or univariate) data sets [Ramsay and Dalzell, 1991]. Commonly high dimensional multivariate data can be approximated with smooth functions. A number of books and articles discuss the methods commonly used in FDA. Much of the foundational description of FDA was recorded in Ramsay and Dalzell [1991]. Although previous papers had discussed functional techniques, Ramsay and Dalzell produced a solid theoretical basis for FDA. First, Ramsay and Dalzell defined the space where FDA techniques are executed. Rather than $\mathbb{R}^n$ in multivariate analysis, FDA works in the Hilbert space of square integrable absolutely continuous functions defined on a compact domain, called $\mathcal{H}$. Although there are more general function spaces, choosing $\mathcal{H}$ allowed Ramsay and Dalzell to take advantage of many properties of projection and distance that are inherent to Hilbert spaces. From the construction of the space, it follows that a random functional variable X is a function:

$$X : (\Omega, \Sigma, P) \rightarrow \mathcal{H}$$

where (Ω, Σ, P) is a probability space. In more detail, Ω is the sample space, Σ is a sigma algebra on Ω, and P is a probability measure on Σ. Each realization of X produces a function rather than a vector or scalar.

Ramsay and Dalzell use these basic definitions to show a number of techniques to analyze functional data. The most important technique they discuss (for the purposes of this thesis) is functional principal components analysis. FPCA is a data reduction and analysis technique that decomposes the covariance operator into an orthogonal basis of eigenfunctions. Each eigenfunction describes a common mode of variation present in the data. FPCA is the foundation of the forecasting methodology discussed in this thesis. FPCA is discussed further in the methodology section of the thesis as well as in the reviews Tran [2008] and Shang [2013]. In contrast to the works of Ramsay et. al., Hsing and Eubank [2015] gives a different perspective of FPCA and other FDA techniques. While Ramsay exploits the properties of reproducing kernel Hilbert spaces (which $\mathcal{H}$ also happens to be), Hsing uses the properties of the theory of linear operators to achieve the same results. The reader gains a more complete understanding of FDA techniques by understanding these perspectives.

Reading Ramsay et al. [2009] gives a practical demonstration of common FDA methods and software tools. After a working understanding of FDA is gained through direct exercise, works such as Ramsay and Silverman [2002] give the reader more detailed descriptions of FDA practice.

3. TRAFFIC DATA DISCUSSION

3.1. DATA SOURCE

The data used to validate this thesis' methodology comes from a traffic sensor data set obtained from the Minnesota Department of Transportation Traffic Data Archives [Minnesota Traffic Observatory, 2008]. These traffic time series data were collected with induction loop detectors from 1 January, 2004 through 31 December, 2008. In total, 1827 days of traffic measurements were collected. The detectors were located at the S110 measurement station on I-94 near the Minneapolis/Saint Paul area. The detectors measured the eastbound traffic volume on I-94 at 30 second intervals. Each day consisted of 2880 observations. To reduce data size while maintaining important behavioral characteristics, the traffic data are consolidated into time series consisting of 96 fifteen-minute aggregates [Weijermars and Berkum, 2005]. Large numbers of zero counts in the time series indicate detector failure. Data from failing detectors can adversely affect model development, so it is replaced with good data from a different day. Daily traffic volume time series that had more than four counts of zero vehicles (greater than one hour with no traffic) were replaced with the previous week's observed day to preserve the weekly seasonality properties of the data. Out of the set of 1827 daily time series, only 12 days of profiles were replaced this way. Once the general characteristics of the data have been established, more detailed analysis can begin.

3.2. EXAMINING THE DATA

Examining the data can help demonstrate that the use of FDA techniques are appropriate. Traffic volume data is not stationary throughout the day. On any given day, the series starts at low volumes, then rapidly rises to high volumes during the morning rush. As the day progresses, the traffic remains relatively even until the evening rush, where another increase is generally observed. After the evening rush ends, the traffic volume decreases to its early morning levels. This intra-day variability in the time series means that it is nonstationary. FDA is more appropriate for modeling and characterizing nonstationary data than standard time series models [Ramsay and Dalzell, 1991, Ramsay et al., 2009]. Furthermore, the behavior of traffic throughout the entire day is of most interest. FDA is used to analyze curves and thus a daily traffic volume profile is a single datum rather than 96 individual data.

Traffic volume data exhibit seasonal effects on many scales. Traffic volume has strong autocorrelation for both daily and weekly lags [Williams et al., 1998]. Daily autocorrelation can be seen in the daily regularity of rush periods. Traffic volume profiles are also correlated from week to week. The traffic profiles of two Sundays one week apart will be more similar than a Sunday and its following Monday. In Williams et al. [1998], the authors discuss that modeling multiple seasonal periods with ARIMA is difficult. By taking a daily traffic profile as a single datum, FDA can more easily characterize longer period seasonal effects. FPCA in particular will be useful for this data because its extraction of the the average intra-day trend seamlessly incorporates the daily regularity into the representations it creates. This will enable examination and modeling that incorporates longer period weekly seasonal phenomena.

4. FPCA FORECASTING METHODOLOGY

The methodology presented herein combines techniques from functional data analysis and time series forecasting. The method extends traditional time series forecasting methods onto a larger scale. Instead of a fifteen minute prediction, this method predicts an entire day ahead. This method is outlined in Figure 4.1. The time of day t ranges from 1 to T, where T is the number of discrete observations in a daily traffic time series. A particular day where traffic observation occurred is denoted by d. First, functions $f_d(t)$ (abbreviated as f_d or f when appropriate) representing the observed time series $\{a_{t,d}\}_{t=1}^{T}$ are constructed from a basis of k splines, $\boldsymbol{\phi} = \{\phi_i(t)\}_{i=1}^{k}$. Next, FPCA is used to create a truncated orthonormal basis $\{\psi_i\}_{i=1}^{N}$ consisting of N principal component functions. Represented with this new basis, each day $f_d(t)$ consists of coordinates, called FPCA scores, $b_{i,d} = \{< f_d, \psi_i >\}_{i=1}^{N}$. Then, for each sequence of daily FPCA scores a SARIMA model is selected with minimal Akaike information criterion (AIC). This model creates forecasts of day $d + 1$'s FPCA scores $\hat{b}_{i,d+1}$ and then combines them to produce a forecast of the next day's traffic profile. Finally, prediction intervals for the forecasts are computed from the residuals between the modeled days and their respective observed days.

4.1. FIT CONTINUOUS BASIS FUNCTIONS TO DISCRETE TIME SERIES

The first step in the proposed functional time series forecasting method is representing the discrete time series data with a smooth function that is a linear combination of continuous basis functions. Most commonly, basis functions can be polynomials, splines, and sine/cosine functions. As some data can be more efficiently represented with one basis over the other, the underlying form of the functional basis is at the discretion of the modeler [Ramsay et al., 2009]. The problem of fitting a function to a time series is as follows. If $\{a_{t,d}\}_{t=1}^{T}$ is a day's discrete traffic volume time series consisting of T observations, a finite

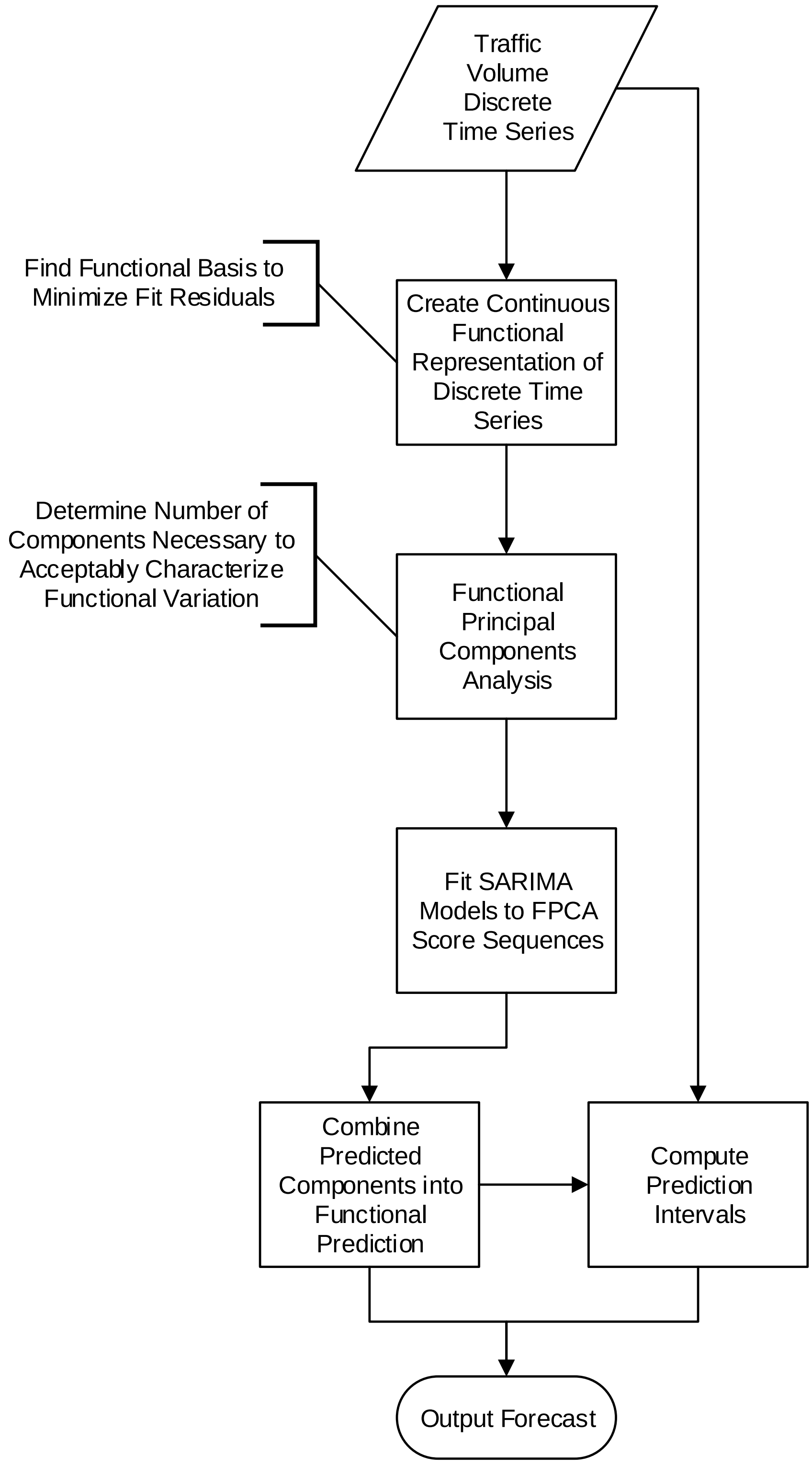

Figure 4.1. Outline of FPCA Forecasting Methodology.

basis of functions $\boldsymbol{\phi} = \{\phi_i(t)\}_{i=1}^{k}$ with domain $[1, T]$ can be created to approximate the discrete series. To create this approximation, a vector $\boldsymbol{c} = \{c_i\}_{i=1}^{k}$ is created. The corresponding basis function ϕ_i is scaled by c_i. The function $f_d(t) = \sum_{i=1}^{k} c_k \phi_k(t)$ will be fitted to the discrete time series. To fit $f_d(t)$ to $a_{t,d}$ the sum of squared errors (SSE) between the functional fit and the time series must be minimized. By modifying the values of $\{c_k\}_{i=1}^{k}$, $\min SSE = \sum_{t=1}^{n} (a_t - f(t))^2$.

This minimization is achieved through the method of linear least squares. The method of linear least squares reduces the above problem to a linear algebra problem that is very easily solved by a computer. Determining the kind and number of basis functions necessary for a quality representation is directly driven by the data that will be represented. A good functional representation of a time series reduces the size of its residuals and eliminates regular patterns from them. The determination of k (the number of functions in the finite basis) is further discussed in the results section of this paper.

4.2. GENERATE FUNCTIONAL PRINCIPAL COMPONENTS AND DAILY SCORES

The next step in the functional time series forecasting method is the application of functional principal components analysis, or FPCA. Whereas principal components analysis (PCA) used in multivariate statistics creates a singular value decomposition of the covariance matrix, FPCA creates an orthonormal decomposition of the *covariance operator*. The primary difference is that PCA is performed on data of finite dimensionality (like a discrete time series) while FPCA is performed on data of infinite dimensionality [Tran, 2008, Shang, 2013]. PCA has been shown to be useful in working with traffic data [Li et al., 2015, Xing et al., 2015, Chen et al., 2012]. FPCA further extends this utility.

Suppose that X is a random variable of functional traffic volume profiles taking events in a probability space (Ω, Σ, P) into the space of L_2 (square integrable) functions with domain $[1, T]$. If $\mu = E[X]$; then the covariance operator on functional data is defined as

$$C_X(f) = E[< (X - \mu), f > (X - \mu)] \tag{4.1}$$

where f is a function in L_2, and $C_X(f)$ is the covariance of f with X. Note that the inner product for infinite dimensional objects is computed as an integral, i.e, $< f, g >= \int_1^T f(t)g(t)dt$. Equation 4.1 shares strong similarities to the finite dimensional covariance matrix. The only major difference is that f and μ are continuous functions rather than discrete time series.

The covariance operator is compact, linear, self-adjoint, and positive semidefinite, so the spectral theorem may be applied (Lemma 6 in Tran [2008]). The spectral theorem for compact self-adjoint linear operators shows the existence of an eigendecomposition of the covariance operator (Theorem 7 in Tran [2008]). This decomposition yields the functional principal components—the eigenfunctions of the covariance operator.

The spectral theorem asserts that there exists a sequence of eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ of C_X such that $|\lambda_i|$ decreases to 0 as $i \to \infty$. The spectral theorem also asserts the existence of eigenfunctions $\{\psi\}_{i=1}^{\infty}$ of C_X corresponding to the eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$. The eigenfunctions form an orthonormal basis for $L^2[1, T]$. In addition, the functions in the orthonormal basis are uncorrelated with one another. Lack of correlation simplifies the creation of SARIMA forecasts of the FPCA scores.

Any functional representation of a time series f_d can also be represented by projection onto the computed orthonormal basis. In other words, $f_d = \mu + \sum\limits_{i=1}^{\infty} \psi_i < f_d, \psi_i >$.

Because the eigenvalues of the covariance operator decrease to zero, this representation in the new orthonormal basis can be truncated into a finite series that approximates f, the functional representation of the time series:

$$f_d \approx \mu + \sum_{i=1}^{N} \psi_i < f_d, \psi_i > \tag{4.2}$$

This truncated representation massively reduces dimensionality but still preserves most of the variation. Truncating the series still accurately approximates f because each each eigenfunction quantifies a progressively smaller amount of the variance than the previous eigenfunction. Determination of N, the number of terms in the series to keep, depends on the required quantity of variance that is needed to be kept by the FPCA representation (Equation 4.3). This determination is data specific and will be further explored in the results section of the paper. The quantity of variation in the truncated representation is computed as the ratio of the partial sum of the eigenvalues of C_X to the total sum of the eigenvalues of C_X:

$$V = \frac{\sum_{i=1}^{N} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \tag{4.3}$$

Equation 4.3 is used to produce a scree plot that shows the quantity of variation contained in the truncated basis representation versus the size of the basis. The scree plot is a convenient visual method to determine the number of principal components to keep. It is used in Section 5.2 to justify how many principal component functions are used.

The coefficients $\{< f_d, \psi_i >\}_{i=1}^{N}$ in Equation 4.2 are called the principal component scores of the function. The principal component scores of a day gives its coordinate representation in the truncated orthonormal basis computed with FPCA. In this way, FPCA generates a daily sequence of principal component scores. These principal component scores are what will be forecasted in the following section.

4.3. FIT SARIMA MODEL TO PRINCIPAL COMPONENT SCORES

After the daily sequences of FPCA scores have been created, a SARIMA$(p, d, q)(P, D, Q)_m$ model is fitted to the M most recent daily scores to produce next-day score forecasts. The nonseasonal and seasonal orders of autoregression (resp. p, P), differencing (resp. d, D), and moving average (resp. q, Q) are selected automatically using the `auto.arima` function from the ‘forecast’ package in R. The `auto.arima` function selects parameters so that the AIC of the model is minimized. For computational efficiency, only the most recent daily scores are used in the model selection process. Finding the quantity of recent scores that are needed is discussed in the results section. The seasonal lag (m) of the SARIMA model is manually determined through examination of autocorrelation plots of the scores and is also discussed in the results section.

4.4. GENERATE FORECASTED FUNCTION AND PREDICTION INTERVALS

After the FPCA score models are created, forecasted scores can be produced. If $\hat{b}_{(i,d+1)}$ is the ith principal component score forecast created by its respective SARIMA model, then the functional forecast is created by replacing $< f_t, \psi_i >$ in Equation 4.2 with $\hat{b}_{(i,d+1)}$. The forecast for the next day (Equation 4.4) is the mean function added to a linear combination of the largest N estimated principal component functions with their respective forecasted next day principal component scores:

$$\hat{f}_{d+1} = \mu + \sum_{i=1}^{N} \hat{b}_{i,d+1}\psi_i \tag{4.4}$$

The SARIMA models of the FPCA scores can produce forecasts for the observed days in the training set. The residual differences between the M modeled days $\{\hat{f}_\tau\}_{\tau=d-M}^{d}$ and the M observed time series $\{a_{t,\tau}\}_{\tau=d-M}^{d}$ are analyzed to produce prediction intervals for the forecast $\hat{f}_{d+1}$. These prediction intervals are difficult to generate through parametric methods. The model used to create the forecasts is based on a relatively small data set.

There is a small number of days where the model can be compared with the observed days. Outliers can significantly distort the statistical properties of the residuals. To reduce the effects of outliers, a nonparametric bootstrap process is used.

If the residual between the functions produced by the FPCA SARIMA model and the observed traffic volume time series is $\epsilon_d(t) = f_d(t) - a_{t,d}$, then the set of residuals to be sampled for the bootstrapping process is $B = \{\epsilon_\tau\}_{\tau=d-M}^{d}$.

A large number of simulated training set residuals, $\mathcal{B}$, is created by sampling with replacement from B. A simulated residual set b found in $\mathcal{B}$ consists of residuals ϵ_{τ_i}, where τ_i is *any* sequence of M integers between $d - M$ and d. For each simulated residual set b in $\mathcal{B}$, the 5% and 95% population quantiles are calculated at every time t from 1 to T. Finally, these 5% and 95% population quantiles for every element of $\mathcal{B}$ are averaged to produce the bootstrapped 5% and 95% population quantiles. These bootstrapped quantiles are added to the forecasted function to create the prediction intervals.

5. FPCA FORECASTING RESULTS

This section demonstrates that the methodology presented above is appropriate for traffic modeling and forecasting. Each step of the methodology is justified. Approximation of traffic time series with a basis of splines will be justified by careful examination of the residuals of the approximations. The use of FPCA and number of components will be justified by examining what and how much variation is explained with each principal component harmonic. Next, the SARIMA models employed to forecast the FPCA scores will be justified by examining the properties of the FPCA score time series. The Akaike Information Criterion (AIC) will be used to justify the choice of the model parameters. Finally, the functional forecasts will be compared with direct SARIMA forecasts by comparing the magnitude of their respective root mean squared error (RMSE). Daily traffic forecasts produced from the forecasted functional principal component scores will be show to perform better than forecasts produced from a SARIMA model directly fitted to the time series. This demonstrates that FPCA increases the forecasting abilities of traditional time series forecasting.

5.1. FITTING BASIS FUNCTIONS TO TRAFFIC TIME SERIES

A basis of continuous functions had to be created before they could fit the discrete time series. Splines were selected over a polynomial or a Fourier basis because the daily traffic time series are nonstationary and nonperiodic on the time scales on which the functional fits are made [Ramsay et al., 2009]. A basis of 32 splines sufficiently described the continuous behavior of the daily traffic series. To verify the quality of the functional fits, the residuals were examined for any major trends or patterns. Table 5.1 shows that on

Table 5.1. Summary of Residual Fitting Statistics.

	Normalized Traffic Count (Divided by 1709)	Actual Traffic Count
Min.	-0.3856000	-658.9904
1st Qu	-0.0110600	-18.90154
Median	-0.0000091	-0.0155519
Mean	0.0000997	0.1703873
3rd Qu.	0.0106000	18.1154
Max	0.9122000	1558.9498

average, the functional representation differs from the observed time series by less than half a car. Figure 5.1 provides an additional illustration that the functional fits to the traffic time series approximate the time series very well.

The noise that is left out of the functional fits lies in a narrow band that surrounds them. For example, Figure 5.2 shows that 95% of the residual noise from a sample day's functional representation lies within a relatively small band about the function. The narrow noise band around the fitted traffic volume function indicates that the function characterizes much of the behavior of the discrete time series. This further demonstrates the existence of an underlying smooth process driving the traffic volume observations.

5.2. GENERATION OF FUNCTIONAL PRINCIPAL COMPONENTS

Although the dimensionality of the data has been reduced by fitting continuous functions to the time series, the components of each vector that scale associated basis functions are highly correlated with their neighbors. This correlation between components is exploited with FPCA to further reduce the dimensionality of the data.

The first step in conducting FPCA is to determine the number of components necessary to characterize the dominant phenomena found in the data. A scree plot is used to show the relationship between the number of principal components and the overall variation that they describe. Scree plots are produced by calculating V (Equation 4.3), the fraction of

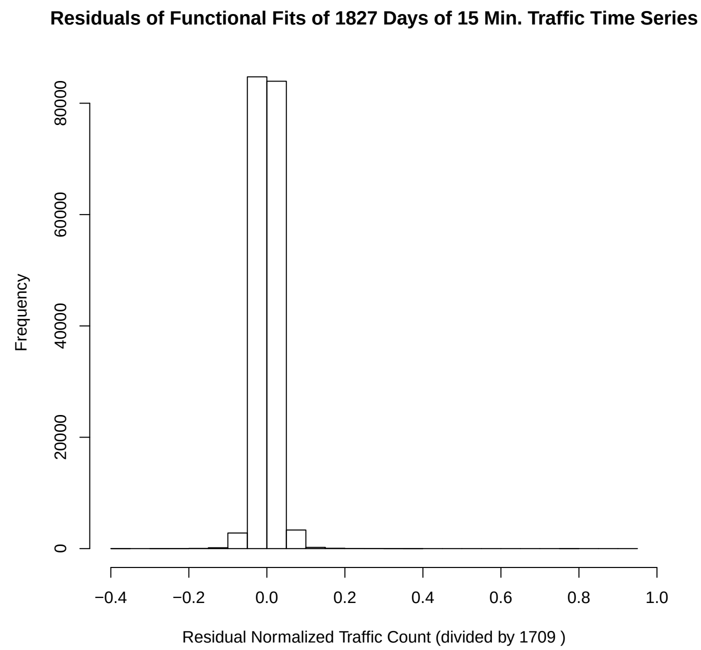

Figure 5.1. Histogram of Fitting Residuals.

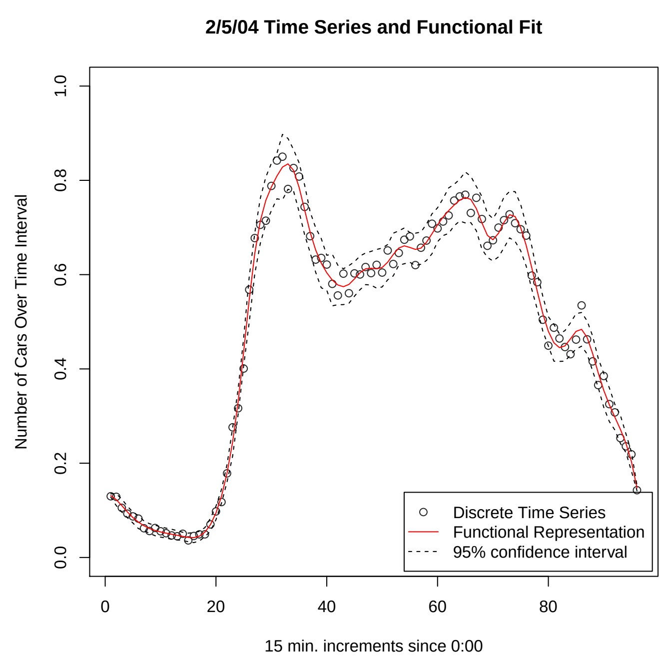

Figure 5.2. A sample day’s functional fit plotted with its associated discrete time series.

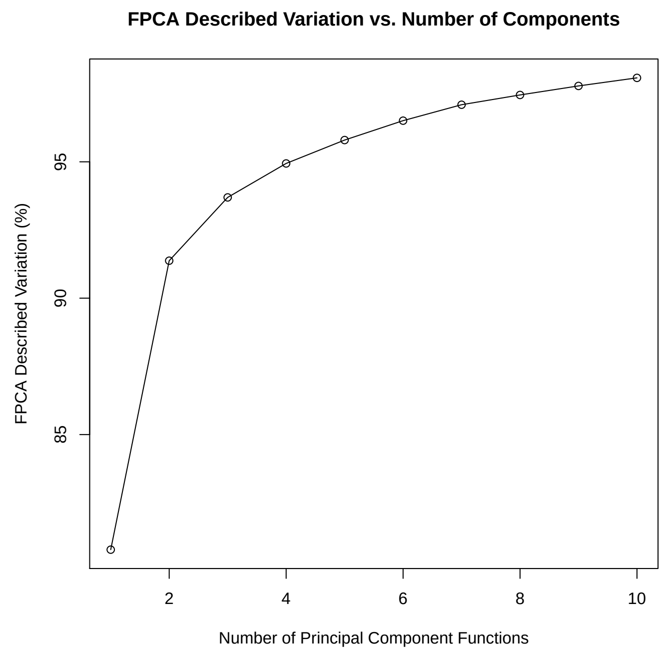

Figure 5.3. Scree Plot of Number of Components vs. Quantity of Described Variation.

variance described by N FPCA harmonics, for a range of values of N. Figure 5.3 is the scree plot corresponding to the first ten functional principal components for the entire functional traffic data set. The first three principal components describe 93.6% of the variation in the functional time series data. The FPCA representation of the data will be truncated at $N = 3$ because the remaining 6.4% of variation is deemed negligible for the purposes of forecasting.

FPCA does not preserve all of the variation of the data. Depending on what one wishes to model, this can be a cost or a benefit. When producing traffic volume forecasts, the reduction in variation of data characterized with FPCA is exploited to make the models more robust against short term transient behaviors such as minor car accidents. FPCA representations keep the *principal* modes of variation found in a time series. Consistent events are very effectively described with FPCA. In many cases, the first principal component function directly corresponds to a clearly observable phenomena in the data [Ramsay et al., 2009]. Higher order principal components do not reflect a singular phenomenon in the

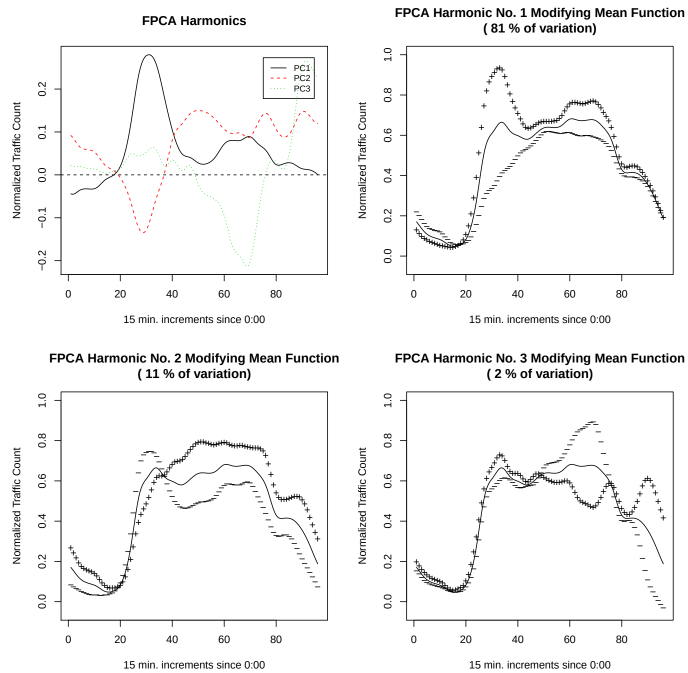

Figure 5.4. Demonstration of how FPCA harmonics modify the mean function.

data insomuch as they provide “contrast” that better separates phenomena occurring at different times in the data. Figure 5.4 shows that the rush hour periods (or the lack thereof on weekends) are so consistent that they are described by the first principal component harmonic. Higher order principal components appear to modify the intensity and duration of morning and evening rush periods in the data. Consistently occurring events are described with relatively few components. However, FPCA does not represent transient events as well. Figure 5.5 shows such a transient event (most likely an accident) that is essentially ‘ignored’ by the FPCA representation of the day. It is the nature of transient events like these to be difficult to predict. Incorporating these transients into a model can skew the forecasts and prediction intervals that the model can produce. FPCA is useful for the creation of generic forecasts because it fails to capture very unusual events that would otherwise reduce the quality of the model.

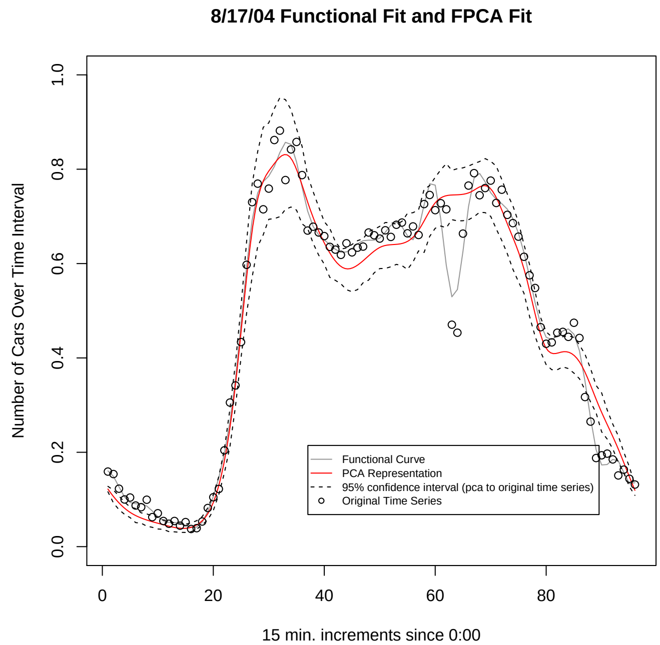

Figure 5.5. Because accidents are rare, they are not characterized within the first three principal components of the data.

FPCA is also useful for forecasting because the FPCA representations have uncorrelated scores and low dimensionality. Where each day in the original data set is described with 96 values, each FPCA representation of a day's traffic events only needs three values to be described. Reducing the dimensionality of the data set reduces the complexity of making forecasts.

5.3. FITTING SARIMA MODELS TO COMPONENT SCORES

Once the data has been represented using FPCA, the next step is to make models of the daily FPCA scores.

First, the scores must be examined manually to determine the seasonality lag of the model. Figure 5.6 shows the distribution of FPCA scores colored by the day of the week. There is strong visual evidence of clustering based on the day of the week, which indicates

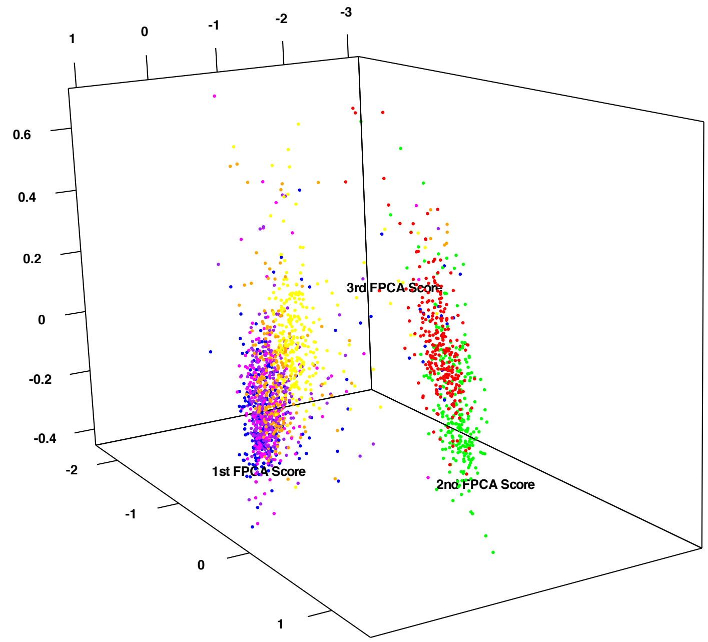

Figure 5.6. 3d scatter plot of FPCA scores colored by day of week (Mon - Blue, Tues - Magenta, Wed - Purple, Thurs - Orange, Fri - Yellow, Sat - Red, Sun - Green).

weekly seasonality. Previously performed traditional SARIMA modeling of traffic time series has also indicated weekly seasonality [Williams et al., 1998]. Examination of the partial autocorrelation plots in Figure 5.7 further confirms weekly seasonality. All three score sequences have strong seven day autocorrelations. Because of this, the SARIMA models for the FPCA scores will incorporate weekly seasonality.

Next, the number of recent daily scores needed to select the model parameters is determined. Comparison of automatically selected models using the 28 most recent days of scores versus the 14 most recent days of scores yielded no difference in the parameters selected by `auto.arima`. However, using the most recent 14 days yielded a large reduction in computation time. Using fewer than 14 days would not ensure sufficient data to select the seasonal parameters (P, D, Q) of the model. Training sets for model parameter selection will utilize only the last 14 days of FPCA scores.

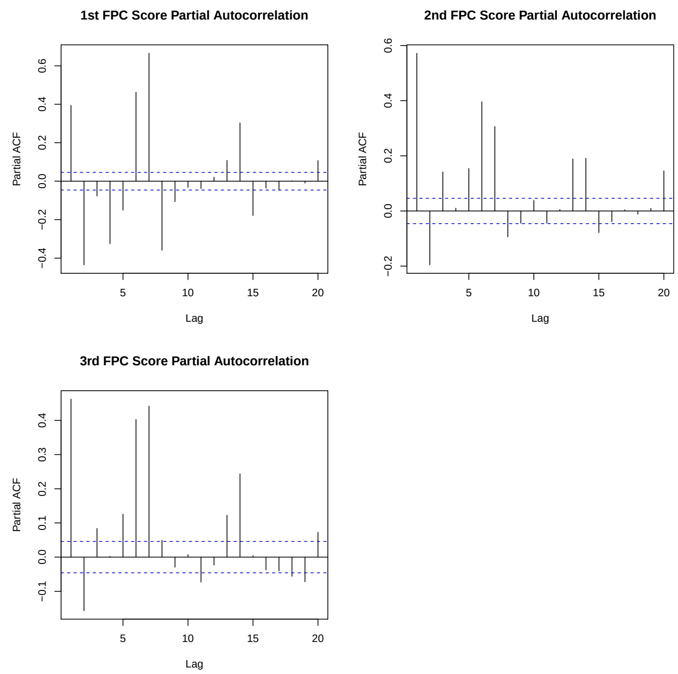

Figure 5.7. Partial autocorrelation of FPCA scores.

Next, SARIMA models are fitted to each sequence of FPCA scores. The rest of the SARIMA model parameters are determined automatically based on the scores in the training set. The orders of autoregressive (p), differencing (d), and moving average (q) parameters of the model are fitted to the training set with the objective of minimization of the AIC. The most commonly selected model is a SARIMA $(0,0,0)(0,1,0)_7$ model. This strongly reiterates the weekly seasonality characteristics of the data.

Once SARIMA models have been fitted to their associated sequence of scores, they can produce next-day score forecasts. To produce a functional forecast, the FPCA score forecasts are linearly combined with their respective FPCA harmonic functions and then added to the mean function. This functional forecast is a prediction of the entire next day's traffic profile.

Table 5.2. Summary of Daily Error Properties.

	MPE	MAPE	RMSE
Min	-750.6	4.1	0.20
5% Pop.	-23.4	5.7	0.28
Median	-1.3	9.0	0.41
Mean	-4.6	15.0	0.58
95% Pop.	+10.3	42.8	1.61
Max	+72.4	761.1	3.03

5.4. GENERATION OF FORECASTS AND PREDICTION INTERVALS

These forecasts are only useful if they manage to describe the next day's events. There are many ways to measure the quality of a forecast. The methods that will be used here are mean percent error (MPE), mean absolute percent error (MAPE) and root mean squared error (RMSE). MPE will demonstrate how the model overpredicts or underpredicts traffic over time while MAPE and RMSE will show how far apart the models are from reality.

The mean and median values of MPE and MAPE are useful for determining if the models are consistently overpredicting or underpredicting traffic volume. MPE and MAPE do poorly when the observed data is zero or near zero. Outliers are best found using the RMSE measure. Since the RMSE is computed without using division, zero/small observed values do not distort the distribution of the population as much.

Figure 5.8 illustrates a forecast with near median RMSE. Some poor forecasts can be traced to phenomena outside the scope of the model. For example, Figure 5.9 illustrates a very poor forecast that was produced on 12 June. This forecasted day is in the 95th percentile of forecast error. The large error occurred because a section of the highway was closed for the entire day [Star-Tribune, 2004].

Although the forecasts are useful, they do not provide information about the accuracy of the predictions. Prediction intervals must be generated for the functional forecasts to better understand their reliability. The relative size of prediction intervals is another important

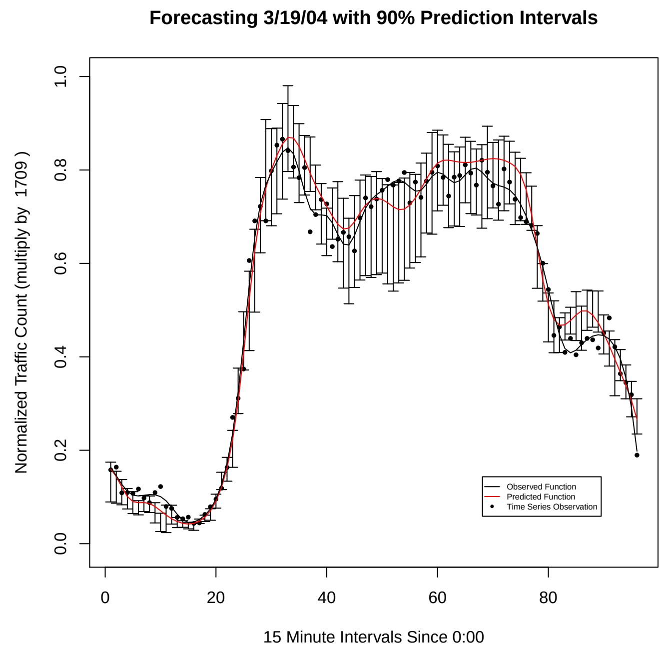

Figure 5.8. An example day close to the median forecast RMSE.

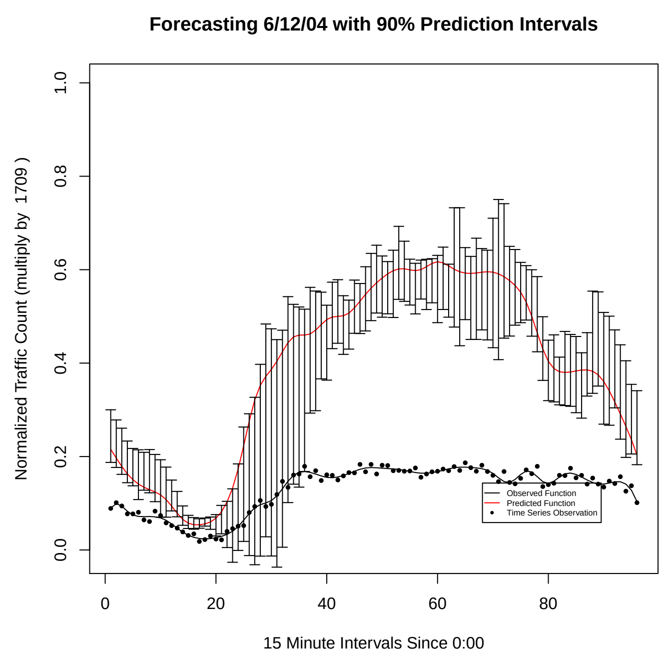

Figure 5.9. An example day from the highest (95% to 100%) quantile of L_2 Distance. This poor forecast was produced due to road construction on the highway impeding normal traffic flow [Star-Tribune, 2004].

Table 5.3. Comparison of summary statistics from 365 residual samples.

	SARIMA RMSE	FPCA-SARIMA RMSE
Min.	0.2504	0.2026
1st Qt.	0.6349	0.3398
Median	1.6833	0.4148
Mean	1.5637	0.5859
3rd Qt.	2.3205	0.6004
Max.	4.1339	2.6181

metric that can be used to compare forecasting models. To demonstrate the improvement FDA makes upon traffic time series forecasting, its predictions are compared with SARIMA. These two models are compared over the same training sets. Comparison of traditional SARIMA forecasting methods to this functional forecasting method is difficult because the two methods operate on different time scales. Traditional time series methods produce forecasts 15 minutes at a time, while functional time series methods produce forecasts one day at a time.

Forecasts for the same days were produced by both methods for comparison. The residuals of the SARIMA model and the functional model were compared. Because finding optimal parameters for a SARIMA model on a large time series is computationally intensive, forecasts from both models were produced for a random sampling of 365 days out of the available data. Note that continuity of the data was preserved. A day's forecast is always produced from the previous 14 days traffic data.

The results of comparing the RMSE of SARIMA vs FPCA-SARIMA can be found in Table 5.3. The FPCA-SARIMA model clearly outperforms the traditional SARIMA model. Though both models use the same data to create forecasts, the mean RMSE of the FPCA-SARIMA is nearly half that of the SARIMA model.

In terms of prediction interval size, traditional SARIMA is outperformed by functional SARIMA. Traditional SARIMA prediction intervals were computed using a similar bootstrapping process that was used for the functional model. However, the traditional

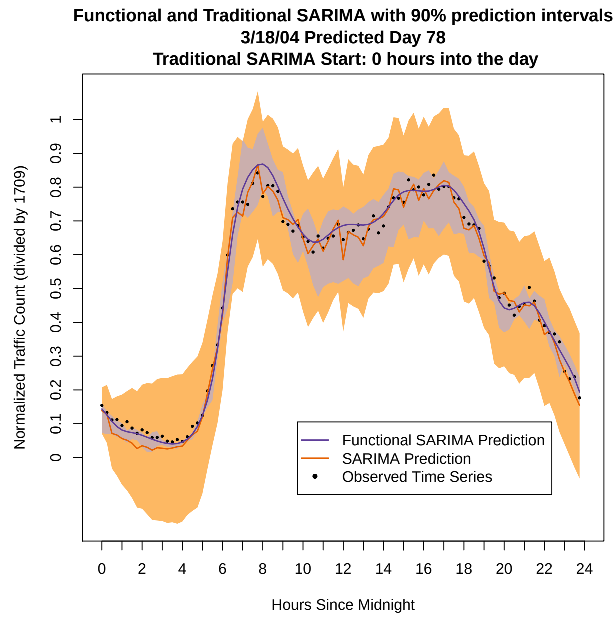

Figure 5.10. Comparison of forecasting methods. 90% prediction intervals for traditional SARIMA (in orange) and FPCA SARIMA (in blue).

SARIMA models make predictions that are many steps farther into the future than the functional forecasts. This naturally causes the prediction intervals to grow in size as time passes. Figure 5.11 and Figure 5.10 demonstrate that while traditional SARIMA predictions become increasingly uncertain with time, functional SARIMA prediction intervals remain relatively narrow. Functional SARIMA prediction intervals are more variable over time, however. Functional SARIMA predicts the predominant phenomena (the functional principal components) that will characterize a day. These predominant phenomena have variable effects throughout the day. Uncertainty in the rush-hour related components does not change the uncertainty (or lack thereof) of the midday related traffic components.

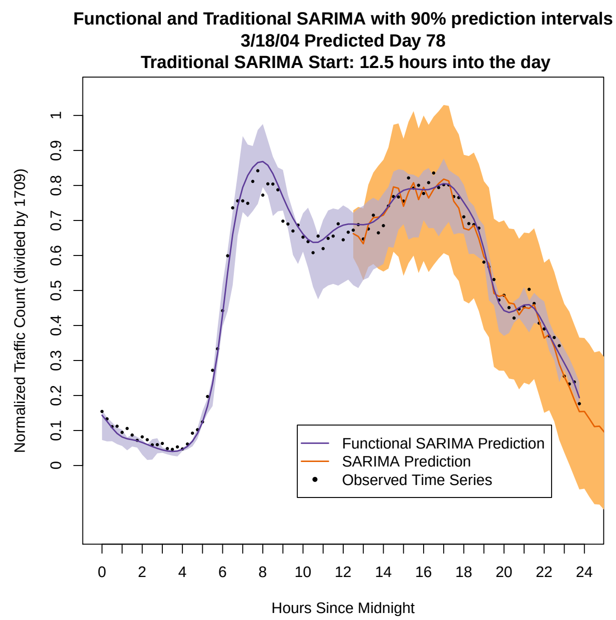

Figure 5.11. Comparison of forecasting methods. 90% Prediction intervals for traditional SARIMA (in orange) and FPCA SARIMA (in blue). Note that after 45min to 1 hour (3–4 steps) ahead, the prediction intervals for traditional SARIMA are wider than for FPCA SARIMA.

6. SUMMARY AND CONCLUSIONS

This thesis demonstrated a novel application of functional data analysis (FDA) and functional principal components analysis (FPCA) to the problem of traffic volume forecasting. Functional statistical methods were employed to extend the forecasting power of traditional time series methods.

To reduce data volume while preserving the temporal features of the traffic data, each day of the time series was represented as a single functional datum. FPCA was then used to determine the principal component functions of the traffic data. The principal component functions represent the dominant modes of variation found in the data. FPCA represented each day of traffic by determining the quantity (FPCA score) of each mode of variation found in the data. While the original daily time series of traffic consisted of 96 observations of traffic volume, the FPCA representation of the same time series consists of only three scores. Finally, two weeks of traffic represented as 14 FPCA score triples were used to train three SARIMA models. In this way, FPCA extended the ability of a traditional time series forecasting methodology to produce functional daily traffic forecasts.

The application of this methodology yielded promising results. The forecasts produced using FDA methods were consistently more accurate than the forecasts produced using traditional SARIMA methods. Furthermore, the functional forecasts prediction intervals remained more narrow over longer periods of time than the equivalent SARIMA forecasts. This shows that the applying FDA methods to time series forecasting allows for more accurate and longer-term forecasts to be produced from the same data set.

The presented methodology is more effective than traditional time series methods for a number of reasons. First, FDA methods exploit the underlying functional characteristics of the time series data to simplify and reduce the size of the data while preserving its important characteristics. Much like continuous functions, traffic time series observations are locally correlated with one another: the traffic volume observed at 9:30 are going to be similar to the traffic at 9:35. Table 5.1 confirmed the continuous functions that fit the discrete data characterized a very large component of overall traffic behavior. Using FDA methods, a day consisting of 96 data points becomes a single point—a functional datum. Time series forecasting methods make their best forecasts a short period of time into the future. Working directly with the time series data, good forecasts can only be made 15–45 minutes into the future. In contrast, by working with the functional representation of the time series, accurate forecasts are made an entire day into the future.

Next, there are analytical tools available for functional data that are not for traditional discrete time series. Many times, differencing discrete time series amplifies the error contained in the data. This larger error can obscure correlations between the differenced time series and other phenomena of interest. FDA represents discrete time series with smooth functions, removing some of the noise present in the time series. The lack of this noise in the functional data allows stronger inferences to be made from the rates of change within the functional data. The smooth and differentiable functional data represents more useful information about traffic flow than the corresponding discrete time series. FPCA is another analytical tool that makes this methodology so effective. By representing functional data with an orthonormal basis tailored to the data itself, a very clear view of the dominant behaviors of the data is created. FPCA extends the utility of PCA methods applied to discrete time series by maintaining the continuity properties inherent to functional data. Principal components analysis on discrete data characterizes less variation in the data per component because a large amount of random noise is still present. The smooth functional data does not have this random noise, so a much larger amount of the variation can be characterized using FPCA. Smooth data is easy to analyze.

Finally, representing each day with FPCA scores enables much easier application of time series methods. Time series methods such as SARIMA struggle to characterize multiple seasonal effects at once. The SARIMA model used for comparison in this thesis performed seasonal differencing at a 96 observation (one day) lag. When used with the FPCA representation of data, the SARIMA model performed seasonal differencing at a 7 observation (one week) lag. Characterizing the data with FPCA simplifies clustering and hence simplifies forecasting. Characterizing weekly lag periods in traffic models is useful because of the seasonal characteristics of the data. For example, there is more similarity between two Monday traffic profiles one week apart than between a Sunday traffic profile and a Monday traffic profile a single day apart. In this way, FPCA enables traditional time series methods to look further within the data to characterize the longer term seasonal effects in the data.

The forecasting method described in this paper leads to many areas of future research. FPCA is used in this paper to summarize a 15-minute sample time series into a 24-hour sample time series. Other commonly used traffic volume models like neural networks and support vector machines could be used in place of SARIMA to forecast the daily FPCA scores of traffic time series.

Currently, the presented method produces forecasts only on a daily basis. It does not take partially observed days into account when forecasting. Relatively anomalous days of traffic volume (such as road closures or holidays) can be difficult to predict, but easier to detect as they occur. The papers Guardiola et al. [2014], Chiou [2012] as well as Figure 5.6 suggest that a clustering analysis of FPCA scores is a promising method to classify traffic volume profiles and detect anomalous ones. Further development of models similar to Chiou [2012] that detect anomalous days as they occur and accordingly change the forecast of the remainder of the day will lead to even better overall forecast quality.

The presented forecasting method could also be combined with existing dynamic traffic routing systems. These systems intelligently avoid areas that will have heavy traffic in the future [Kok et al., 2012]. More intelligent traffic forecasts would produce higher quality routes.

Other functional data analysis tools could also be applied to traffic volume problems. FDA could be used to better understand the relationship between traffic volume and driving conditions. Temperature, precipitation, and visibility conditions all affect driver behavior. A functional response model could be created to better understand how weather affects traffic volume over the course of a day, allowing for ITS to respond to a more nuanced variety of traffic conditions.

FDA methods could be used to create a continuous dynamic model of traffic behavior. There exists a set of governing laws that dictate how a segment of a road behaves. Simple models of these laws have been previously shown to resemble commonly found dynamic systems [Dorogush, 2013]. Derivation of a set of governing equations that accurately respond to the changing travel demand driving function would allow advanced traffic management systems to more intelligently control the traffic patterns of a city. Instead of being applied for forecasting, FDA methods could be used to produce simulated traffic volume profiles to validate the dynamic models. The continuous and differentiable nature of the functional data fit with cubic splines allows direct input into models that utilize differential equations.

Functional data analysis methods provide a number of new perspectives with which to observe and predict traffic phenomena. In particular, functional principal components analysis has the ability to greatly extend the utility of current traffic forecasting methods.

This thesis demonstrated the implementation and verification of a novel forecasting methodology applied to traffic volume time series. Functional data analysis is a relatively new field of statistics. Many of its benefits have not been fully exploited in real world applications. By showing the effectiveness of this functional forecasting methodology, this thesis helps build the bridge between practice and theory. Furthermore, this thesis has demonstrated that despite the large body of research in traffic volume forecasting, new perspectives on the data can bring insights and benefits to both researchers and engineers.

REFERENCES

Star-Tribune. I-94 closed this weekend through minneapolis. *Star-Tribune*, Jun 2004.

Sweet. Traffic congestions economic impacts: Evidence from us metropolitan regions. *Urban Studies*, 51(10):2088–2110, 2014.

Zhang, Wang, Wang, Lin, Xu, and Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), December 2011.

Vlahogianni, Karlaftis, and Golias. Short-term traffic forecasting: Where we are and where were going. *Transportation Research Part C: Emerging Technologies*, 43:319, 2014.

Taniguchi and Shimamoto. Intelligent transportation system based dynamic vehicle routing and scheduling with variable travel times. *Transportation Research Part C: Emerging Technologies*, 12(3-4):235250, 2004.

Papageorgiou, Diakaki, Dinopoulou, Kotsialos, and Wang. Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12):2043–2067, Dec 2003. ISSN 0018-9219.

Carlson, Papamichail, Papageorgiou, and Messmer. Optimal mainstream traffic flow control of large-scale motorway networks. *Transportation Research Part C: Emerging Technologies*, 18(2):193 – 212, 2010. ISSN 0968-090X.

Ahmed and Cook. Analysis of freeway traffic time-series data by using box-jenkins techniques. *Transportation Research Record*, (722):1–9, 1979.

Hong. Traffic flow forecasting by seasonal svr with chaotic simulated annealing algorithm. *Neurocomputing*, 74(12-13):20962107, 2011.

Dalgleish and Hoose. *Highway traffic monitoring and data quality*. Artech House, 2008.

Chowdhury, Santen, and Schadschneider. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, 329(46):199 – 329, 2000. ISSN 0370-1573.

Helbing. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.*, 73: 1067–1141, Dec 2001.

Mahnke, Kaupus, and Lubashevsky. Probabilistic description of traffic flow. *Physics Reports*, 408(12):1 – 130, 2005. ISSN 0370-1573.

L. Li, Su, Zhang, Lin, and Z. Li. Trend modeling for traffic time series analysis: An integrated study. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):3430–3439, 2015.

Williams, Durvasula, and Brown. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board*, 1644: 132141, 1998.

Datla and Sharma. Impact of cold and snow on temporal and spatial variations of highway traffic volumes. *Journal of Transport Geography*, 16(5):358 – 372, 2008. ISSN 0966-6923.

Xiao, Yang, and Ye. Physics of day-to-day network flow dynamics. *Transportation Research Part B: Methodological*, 86:86–103, 2016.

Chen, Kwon, Rice, Skabardonis, and Varaiya. Detecting errors and imputing missing data for single-loop surveillance systems. *Transportation Research Record*, (1855):160–167, 2003.

Van Lint, Hoogendoorn, and Van Zuylen. Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, 13(56):347 – 369, 2005. ISSN 0968-090X.

Qu, Li, Zhang, and Hu. Ppca-based missing data imputation for traffic flow volume: A systematical approach. *IEEE Transactions on Intelligent Transportation Systems*, 10(3): 512–522, Sept 2009. ISSN 1524-9050.

Chiou, Zhang, Chen, and Chang. A functional data approach to missing value imputation and outlier detection for traffic flow data. *Transportmetrica B: Transport Dynamics*, 2 (2):106–129, Feburary 2014.

Weijermars and Van Berkum. Analyzing highway flow patterns using cluster analysis. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 308–313, Sept 2005.

Guardiola, Leon, and Mallor. A functional approach to monitor and recognize patterns of daily traffic profiles. *Transportation Research Part B: Methodological*, 65:119136, 2014.

Guo, Huang, and Williams. Real time traffic flow outlier detection using short-term traffic conditional variance prediction. *Transportation Research Part C: Emerging Technologies*, 50:160 – 172, 2015. ISSN 0968-090X. Special Issue on Road Safety and Simulation.

Donoho. High-dimensional data analysis: The curses and blessings ..., Aug 2000. URL `http://statweb.stanford.edu/ donoho/lectures/cbms/curses.pdf`.

Bellman. *Dynamic programming*. Princeton Univ Pr, 1957. ISBN 9780691079516,069107951X.

Naymat. Mitigating the influence of the curse of dimensionality on time series similarity measures. *IJCAT International Journal of Computer Applications in Technology*, 52(1): 94, 2015.

Tran. An introduction to the theoretical properties of functional principal components analysis. Master's thesis, "The University of Melbourne", 2008.

Xing, Zhou, Hong, Huang, Bian, and Xie. Traffic flow decomposition and prediction based on robust principal component analysis. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, volume 2015-October, pages 2219–2224, 2015.

Smith and Demetsky. Traffic flow forecasting: Comparison of modeling approaches. *Journal of Transportation Engineering J. Transp. Eng.*, 123(4):261266, 1997.

Williams and Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering J. Transp. Eng.*, 129(6):664672, 2003.

Yun, Namkoong, Rho, Shin, and Choi. A performance evaluation of neural network models in traffic volume forecasting. *Mathematical and Computer Modelling*, 27(9):293 – 310, 1998. ISSN 0895-7177.

Yin, S.c. Wong, Xu, and C.k. Wong. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, 10(2):8598, 2002.

Dia. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*, 131(2):253 – 261, 2001. ISSN 0377-2217. Artificial Intelligence on Transportation Systems and Science.

Boto-Giralda, Díaz-Pernas, González-Ortega, J. Díez, Antón-Rodríguez, Zarzuela, and I. Díez. Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 25 (7):530–545, 2010. ISSN 1467-8667.

Castro-Neto, Y. Jeong, M. Jeong, and Han. Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, 36(3 PART 2):6164–6173, 2009.

Ramsay and Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539572, 1991.

Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis AStA Adv Stat Anal*, 98(2):121142, Feb 2013.

Hsing and Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley, 1 edition, 2015. ISBN 0470016914,9780470016916.

Ramsay, Hooker, and Graves. *Functional Data Analysis with R and MATLAB*. Springer New York, 2009.

Ramsay and Silverman. *Applied functional data analysis : methods and case studies*. Springer series in statistics. Springer, 2002. ISBN 0387224653,9780387224657,9780387954141,0387954147,0585472750,9780585472751.

University of Minnesota Minnesota Traffic Observatory. Traffic data archives, minnesota traffic observatory, university of minnesota, 2008. URL `http://tcfreeways.itslab.umn.edu/traffic_data.php`.

Chen, Wang, Li, Hu, and Zhang. The retrieval of intra-day trend and its influence on traffic prediction. *Transportation Research Part C*, 22:103–118, 2012.

Chiou. Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann. Appl. Stat. The Annals of Applied Statistics*, 6(4):15881614, 2012.

Kok, Hans, and Schutten. Vehicle routing under time-dependent travel times: The impact of congestion avoidance. *Computers and Operations Research*, 39(5):910918, 2012.

Dorogush. Dynamic model of traffic flow on a circular highway. *Dokl. Math. Doklady Mathematics*, 88(3):690694, 2013.

VITA

Isaac Wagner-Muns received dual Bachelors of Sciences degrees in Mining Engineering and Applied Mathematics from the Missouri University of Science and Technology in 2015. He continued his studies in Missouri S&T's department of Systems Engineering and recieved his MS in Systems Engineering in May of 2017.