REACTING IN REAL TIME:
Using Historical and Real-Time Information in Forecasting Link Travel Times

Christopher Corbett Schrader
April 14, 2003
Professor Alain L. Kornhauser

Submitted in partial fulfillment
Of the requirements for the degree of
Bachelor of Science in Engineering
Department of Operations Research & Financial Engineering
Princeton University
I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Christopher Schrader

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other instructors or individuals for the purpose of scholarly research.

Christopher Schrader
ABSTRACT:

One of the most common decisions an individual makes is how to get from Point A to Point B. When the individual makes his choice, he receives some value for his action. The better the choice, the more value to the decision maker. In route choice, the most distinct measure of value is travel time. Current systems optimize based on hypothetical travel times calculated from distance and assumed speed. The focus of this thesis is to improve this estimate of travel time by developing a methodology for forecasting that combines historical and real-time observations of travel time. A process is described by which a function for individual link travel time may be developed using historical data. A forecasting method is then demonstrated using historical data and real-time information. This process is then applied to road segments in the Milwaukee Highway System.

ACKNOWLEDGEMENTS:

I am indebted to Professor Alain Kornhauser for his guidance and technical support during this project. A special thanks also goes out to Koray Simsek, Dries Darius, and Rong He for their help and support, and Ashirul Amin, Ryan Goldenberg, Katy Milkman, John Cranston, Charles Perkins, Jordan Rapp, Bill Tamashunas, Laurel Harvey, and Mike Bodden for their help with the graduate student shuttle. A personal thanks to my parents, Tom and Kathy Schrader, for their advice and encouragement throughout this year. Thanks to James, Melissa, Aaron, Jackie, Meryl, Ryan, Ethan, the entire third helping, and everyone else who made these past four years so fun.
# Table of Contents

1 Introduction

2 A Survey of Traffic Information
   2.1 Academic Interpretations ........................................ 8
      2.1.1 A Traffic Flow Model ........................................ 8
      2.1.2 The Travel Time Data Collection Handbook ................. 12
   2.2 Information for Use by Real People in Real-Time ............ 16
      2.2.1 The Beginnings of Traffic Reporting ....................... 16
      2.2.2 Development of the Traffic Reporting Industry ............ 18
      2.2.3 The Public Sector and Data Collection ..................... 21
   2.3 Attempts at Combining Real-Time and Historical Information .... 23

3 Including Historical Information ................................. 29
   3.1 Defining Historical Travel Times ............................... 30
   3.2 Examining Travel Time Data ...................................... 34
      3.2.1 Good Data, Bad Data ....................................... 34
      3.2.2 APPLICATION: Milwaukee Highway System .................. 36
   3.3 Representing Travel Time ......................................... 39
      3.3.1 Time Elements ............................................. 39
      3.3.2 Creating an Expectation of Travel Times ................... 41
      3.3.3 APPLICATION: Milwaukee Highway System .................. 47
   3.4 Missing Information ............................................. 54
   3.5 Using Predicted Travel Times .................................... 57

4 Including Real-Time Information ................................ 61
   4.1 Defining Real-Time Travel Times ............................... 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>Real-Time Travel Times in the Real World</td>
<td>63</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Exponential Smoothing</td>
<td>64</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Recurring Traffic Patterns</td>
<td>69</td>
</tr>
<tr>
<td>4.2.3</td>
<td>APPLICATION: Milwaukee Highway System</td>
<td>72</td>
</tr>
<tr>
<td>4.3</td>
<td>Additional Issues</td>
<td>81</td>
</tr>
<tr>
<td>4.4</td>
<td>Using Forecasted Travel Times</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>Summary &amp; Conclusion</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>90</td>
</tr>
<tr>
<td>Appendix A:</td>
<td>Description of the Milwaukee Highway System</td>
<td>94</td>
</tr>
<tr>
<td>Appendix B:</td>
<td>Development of Historical Travel Time</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Function: Milwaukee Highway System</td>
<td></td>
</tr>
<tr>
<td>Appendix C:</td>
<td>Implementation and Results of Real-Time Application: Milwaukee Highway System</td>
<td>111</td>
</tr>
<tr>
<td>Appendix D:</td>
<td>Princeton University Graduate Shuttle Web Application</td>
<td>122</td>
</tr>
</tbody>
</table>
1 Introduction

Every day, people make decisions based upon tasks they must complete and information they are presented with. A decision maker chooses an action such that it maximizes the perceived value to him. The better the information he has, the better his decision.

One of the simplest and most common decisions people make is how to get from point A to point B. This thesis addresses the topic of providing the decision maker with better information to help him make informed choices when taking any trip.

When setting out from A to B, the decision maker must choose a route and mode of transportation. Different choices have different costs to the decision maker and affect
the perceived value of his decision. When measuring the total perceived value of any given route, several attributes that measure value must be considered.\(^1\) Objective, easily measured attributes include travel time, cost of tolls, and cost of gasoline. The decision maker also considers less measurable, subjective attributes that contribute to the value of his decision. Among these are scenic appeal of the route, convenience of a mode, drivability of roads, annoyance factor (for congestion, etc.), and so on. In the end, the value of any route is some function of time, money, and several less measurable attributes.

Certainly, all of these attributes directly affect the value of any trip to the decision maker. However, we consider travel time to be the most important attribute measuring the value of any trip. In doing so, we hope to present one piece of information that will help drivers to make more informed decisions. Thus, for the purposes of this thesis, a decision maker must minimize his travel time between point A and point B in order to maximize his value in traveling between these two points.

\[
\max\limits_{\text{route choice}} (\text{Value to Decision Maker}) \equiv \min\limits_{\text{route choice}} (\text{Travel Time})
\]

\text{Equation 1.1: Travel time as value to the decision maker}

When the decision maker initially chooses how to get from A to B, he begins to gather and evaluate information so that he may make the best decision possible. In the most basic case, when a traveler begins a trip he may consult a map or rely on personal experience to select the best route for his travel. The decision maker must decide which path to take given only a graphic representation of links in a network, his own knowledge

\(^1\) Keeney, Ralph L. and Howard Raiffa. \textit{Decisions with Multiple Objectives: Preferences and Value Tradeoffs}. New York: Cambridge University Press, 1993. 34.
about those roads, and the area and his own knowledge of peak travel times in which he is traveling. We will call this the 0th generation of vehicle routing systems.²

More recently, computer based options have become available that aid travelers by providing a route based upon user inputs and map databases. Staying with the convention of labeling the use of paper maps and knowledge as the 0th generation, the 1st generation of vehicle routing systems includes electronic maps and methods for generating directions (routes) of how to “optimally” get from A to B.³ These routes are fixed and the tools that generate them cannot be used to regenerate a route after travel has begun. Examples of 1st generation systems that produce directions based purely on a Point A origin address and a Point B destination address include Yahoo Maps and MapBlast. Systems such as MSN Maps & Directions and RandMcNally.com allow the user to choose between “shortest” and “fastest” routes, adding a more subjective aspect to the value created by any one route. Some websites, including MapQuest, take this one step further by allowing the user to decide to avoid toll roads, avoid ferries, and several other options.

Second generation vehicle routing systems are mobile devices that travel along the route with the decision maker.⁴ Some systems, such as DeLORME’s Xmap Handheld Street Atlas, simply allow the decision maker to bring the 1st generation systems with them as digital directions or databases. Still other systems equip travelers with GPS (Global Positioning System) based mobile systems. GPS technology allows a device to calculate the position of the traveler at any moment in time. The system consists of 24 satellites orbiting the earth. A GPS receiver needs to have line of sight

---

³ Erera 5.
⁴ Erera 8.
contact with four satellites to determine its position. The receiver obtains the location of
satellites via a one-way broadcast and then computes its location in XYZ coordinates that
can then be translated into latitude and longitude. Most 2\textsuperscript{nd} generation systems use GPS
and only track location of the traveler. The GUI (Graphic User Interface) for such a
system might include a list of directions and a map with an indicator for current location
of the traveler. An example of a 2\textsuperscript{nd} generation system is Garmin’s GPS receiver with
Rand McNally’s MapSource Software. This package allows the decision maker to get
location-to-location directions and observe his progress along his route in map form via
real-time GPS coordinates.

Third generation vehicle routing systems take the latter type of 2\textsuperscript{nd} generation
systems one step further by allowing for real-time route updating based on current GPS
position and a destination.\textsuperscript{5} Really, the 3\textsuperscript{rd} generation system solves the A to B problem
with A as the decision maker’s current location and B as the final destination. Examples
of 3\textsuperscript{rd} generation systems include Pocket CoPilot, CoPilot Truck, and Garmin’s StreetPilot
III.

But how can the decision maker be sure that the vehicle routing systems are
providing him with the best possible route to get to his destination? The simple answer is
that he cannot. He can only expect the best solution based on the information that the
routing systems use to create his “optimal” route. Each of the systems in the 1\textsuperscript{st}, 2\textsuperscript{nd}, and
3\textsuperscript{rd} generations suggest routes based on calculating value using time independent and
condition independent costs for traveling along any given roadway. What this means is
that the cost, or value to the decision maker, is determined using a calculated estimate of
travel time that is not based on actual travel times. This can be accomplished by

\textsuperscript{5} Erera 10.
associating an average speed with each type of road and then calculating the sum of
distances on different types of roads divided by the speed on each, by associating a single
average speed for an entire trip, or by being very meticulous and gathering the speed
limits on each link in a network of roads and using that measure (or some function of it)
along with the length of the link to calculate the total travel time.

These methods clearly lack one thing: they do not use real measures of travel
time. Instead, they use theoretical measures to calculate something very real. The travel
times they create are not based upon any actual measure of time. The next step, the step
to get to the 4th generation of vehicle guidance systems, involves using actual travel times
as a measure of value for traveling from point A to point B. Obtaining and using that
information is the crux of this thesis.

Throughout the next four chapters, we attempt to include measured travel times as
an aid the decision maker both in theory and in practice. This will be done in two parts.
First, historical travel times will be discussed. An approach involving a continuous
parameterized function will be introduced and implemented. Then, real-time travel times
will be introduced, and issues involving their use in combination with historical
information for travel time forecasting will be examined. Throughout, the Milwaukee
Highway System will be examined to illustrate the application of discussed concepts for
predicting travel times on roadways.

Chapter 2 will discuss a history of traffic information, reporting, and forecasting,
what current work is being done in the field, and what technological drivers are pushing it
into the future. This discussion leads directly into Chapter 3, a description of historical
travel times, why they are important, and how they are to be used. In this chapter, we
present our own method for creating an estimate of travel time based on historical data. Chapter 4 examines real-time sources of travel time information and how to include them in travel time prediction. Our approach uses exponential smoothing techniques to create a smoothed estimate of actual travel time and then forecasts future travel times from that smoothed estimate and knowledge of historical traffic patterns. Chapter 5 acts as a conclusion to this thesis, recapping what was done, noting limitations of the work within, and suggesting topics for further investigation. Additionally, an appendix is included discussing implementation of historical and real-time information in examining the development and use of an online tracking system for Princeton University’s graduate student shuttle.
Without traffic reporting and forecasting, a driver would only learn about traffic conditions when he is engulfed in them and it is too late to change his route. For this reason, a great deal of effort, both public and private, has been put into gathering traffic information for drivers to use. This chapter describes the different facets of the evolution of traffic forecasting. Section 2.1 looks at academic interpretations of traffic forecasting. Section 2.2 describes how traffic information is distributed to the public and the current information collection environment. And, section 2.3 discusses attempts to combine historical and real-time information to produce traffic forecasts.
2.1 Academic Interpretations

2.1.1 A Traffic Flow Model

A significant amount of work has been done with traffic information to alleviate congestion on roadways in order to minimize the travel time of individual travelers. When constructing roads and making changes to the current network of roads, civil engineers require information about how drivers will behave. They must know how many people will be traveling on each road, old and new, and must decide if the proposed change is effective based upon their project goals.

Modeling traffic patterns and making estimates for road volumes through a traffic flow model is one possible way to obtain this information. This concept involves trip start points, endpoints, and routes, all of which will become very important when we begin to deal with measured travel times. A traffic flow model is essentially a calculated projection of traffic flows through a network and can be summarized by the following four steps:6

- Trip Generation
- Trip Distribution
- Modal Split
- Traffic Assignment

In trip generation, the input is demographic information about the land area the trips are being generated for, a division of the land area into zones, information about the people who live there, expected growth factors, and any other variables the model can feasibly include. The outputs of this step are *productions* *(P)_i* and *attractions* *(A)_j* for

---

Each zone. Productions and attractions are different from origins and destinations because productions and attractions describe where all trips start and end, a concept similar to sources and sinks. For example, trips are produced in the home and are attracted to work space and shopping areas. Therefore, residential areas tend to have large numbers of productions and commercial zones tend to have large numbers of attractions. Origins and destinations on the other hand describe the path of any one individual trip. If the system being modeled is self-contained, the sum of all productions should equal the sum of all attractions.

After productions and attractions are generated, one must model which productions interact with which attractions. This is trip distribution. The inputs are $P_i$ and $A_j$ for all zones, and the outputs are a matrix $T_{ij}$ of trips. Each element of this matrix represents the total number of trips taken from zone $i$ to zone $j$. A time dimension should be added to this for traffic forecasting. Contained in this three dimensional matrix are trips from $i$ to $j$ in time interval $t$.

Trip distribution is traditionally done by using a “gravity model.” The gravity model assigns trips based upon the relative attractiveness of each zone. Obviously, closer zones are easier to get to and require less travel time, thus they are more attractive. The model is called a gravity model because closer attractions pull harder on productions just as one physical object pulls harder on another if they are closer together. Equation 2.1 describes one gravity model.
\[ T_{ij} = \frac{P_i A_j F(t)_{ij} K_{ij}}{\sum_{j=1}^{n} A_j F(t)_{ij} K_{ij}} \]

where:
- \( T_{ij} \) = trips between \( i \) and \( j \)
- \( F(t)_{ij} \) = accessibility of \( j \) from \( i \) at time \( t \)
- \( A_j \) = attractiveness of \( j \) (# of attractions)
- \( P_i \) = productions from \( I \)
- \( K_{ij} \) = socio-economic factor.

**Equation 2.1: Trip Distribution.**

Within this \( T_{ij} \) matrix, we want the rows to sum up to the number of productions at \( i \) and the columns to sum up to the number of attractions at \( j \). In the formulation above, the elements of the rows of \( T_{ij} \) do sum to \( P_i \). However, the sum of the elements of the columns does not equal the number of the attractions. Conceptually, many possible reasons exist why this might occur. For example, people may use one trip to go to many attractions, such as a person running errands. Kornhauser suggests we use an adjustment factor \( A_{jk} \) to modify the attractions of each zone:

\[ A_{jk} = \frac{A_j A_{j,(k-1)}}{C_{j,(k-1)}} \]

Where:
- \( A_{jk} \) = adjusted attraction factor for attraction zone (col) \( j \), iteration \( k \)
- \( A_{j,0} = A_j \) (note: \( k = 0 \))
- \( C_{j,k} \) = actual attraction total for zone \( j \), iteration \( k \)
- \( A_j \) = desired attractions for zone \( j \)
- \( k \) = iteration number

**Equation: 2.2: Adjustment of Trip Distribution.**

---

We then use this adjustment factor to iterate until $C_{j,k} \sim A_j$. At this point, all of the columns and rows in the trip matrix will add up to the correct number of attractions and productions.

The third step in creating a traffic flow model is modal split. The first two steps modeled how many trips occur between different zones at what times and with what probabilities. This third step, modal split, models which mode of transportation these trips use and in what proportions. The trips can be divided up into modes in any number of ways, but the mode with the highest overall value to the traveler will always get the highest proportion of the trips. One popular way to do this is the Logit Model in which the probability of using any mode is proportional to the exponential of the utility for using that mode.

The final step of the traffic flow model is actually assigning traffic to links in the network. This step will predict the volume on any road for time intervals throughout a day, week, month, or year. There are a number of ways in which to assign traffic to each trip simply because there are multiple paths in most road networks to get from any production to any attraction. Costs are assigned to each link in the network. These costs can be constant, or can vary with some external factor, such as link volume. Once costs are assigned, traffic can be allocated by an all or nothing method in which the path from production to attraction with the least cost is always used by the traveler. This does not seem realistic, and consequently many multi-path assignment techniques are used in formulation of traffic flow models. For example, the model may be designed to find the $k$ least cost paths from $i$ to $j$, $k$ being any integer. The traffic then could be divided up such that the best path gets the most traffic, the second best path gets the second most
traffic, and so on until the $k^{th}$ best path gets the least traffic. Additionally, an “essentially equal” paths method is often implemented such that paths whose total travel time is less than some measure epsilon away from that of the “best” path are considered equal and traffic is divided among these paths.

After traffic assignment, the model is complete. If done properly, the model can give information about expectation of flow on any link in a network at any time and recalculate with changes to the network. With this information, a civil engineer or city planner will be able to predict traffic flows along any link in his network and thereby create some measure of travel time.

Unfortunately, while this model can be used to compute travel times, it can only create theoretical travel times. Additionally, in using this model, a civil engineer would most likely be looking for a system-optimal solution, while the individuals in real life attempt to solve their own problem in a user-optimal way. In this thesis, we look for a more concrete measure of travel times that will be able to help a single user maximize his utility. We look to use measured travel times to get better estimates of actual travel times

2.1.2 The Travel Time Data Collection Handbook

The end goal of the traffic flow model is to be able to create measures of predicted traffic flow such that roadway infrastructure decisions may be made based upon these predictions. The Travel Time Data Collection Handbook, published online by the FHWA, provides several frameworks for measuring actual traffic conditions in the form
of travel times.\textsuperscript{9} Put together with renewed interest after ISTEA Legislation (1991),\textsuperscript{10} the *Travel Data Time Collection Handbook* also draws from a motivation to improve the infrastructure and manage congestion on the roadways. The motivations of this thesis, real-time and case specific improvement of individual driver travel times does not involve any change in infrastructure. However, the frameworks provided for collecting travel times in the handbook are important in describing the concepts of travel time data throughout the entirety of this thesis. In fact, the methods in the handbook can be used for studies where the motivation is to provide information to the public, which is parallel with the motivation of this thesis.

The *Travel Time Data Collection Handbook* defines travel time as “the time required to traverse a route between any two points of interest.”\textsuperscript{11} Travel times consist of both a running time and stopped time.\textsuperscript{12} If a direct measure of time is not available, the handbook recommends that the user employ some other measure from which time can be extrapolated, such as speed or volume.

As its overarching goal, the handbook discusses how to properly design and implement a travel time study by creating the following nine-item list of steps and suggesting several alternatives for most of the steps. I include the entire list and expand upon several of the elements that are the most relevant to this thesis:

- Establish Study Purpose and Objectives:

  The study purpose and objective define what it hopes to accomplish. For example, a study could hope to calculate travel times on Chicago’s main arteries


\textsuperscript{10} Turner 1-1.

\textsuperscript{11} Turner 1-1.

\textsuperscript{12} Turner 1-5.
so that infrastructure decisions may be made to decrease congestion and lower travel times. For this thesis, the purpose is to collect travel times, use them to calculate a useful forecast of future travel times, and provide these forecasts to any decision maker in a form such that he can make better value maximizing choices.

- Understand Uses and Users:
  In our study, the users are individuals. We intend to create travel time forecasts for their benefit.

- Define Study Scope:
  The scope of the study includes geographic areas that control where the study is to take place, facility types that determine specific routes, road segments, rates, and sample sizes to use, and time that determines the length of the study and the time intervals in which to collect information. This thesis.

Figure 2.1: Travel Time Study Procedure.
examines one data set with scope including the arterial highways in Milwaukee during the month of June 2002.

- Select Data Collection Technique:
  The Travel Time Data Collection Handbook mentions four different categories of travel time collection techniques: test vehicle, license plate matching, ITS Probe Vehicle, and emerging and non-traditional. Test vehicles are dispatched specifically for picking up travel times. They can drive as an average car, floating car, or max car and collect data throughout their trip. License plate matching can be done manually or by video character recognition. License plates are matched passing consecutive checkpoints and the time taken between those points is recorded as a travel time. Probe vehicle techniques involve “passive instrumented vehicles and remote sensing devices.”

  13 The devices in these vehicles record a variety of statistics that can be used to calculate travel time. For example, common devices collect latitude and longitude coordinates with a time stamp. From this information, one can determine how long it took the vehicle to go from one coordinate to another. Finally, the handbook mentions emerging technologies. This obviously covers a broad range of topics, some of which include inductance loops, aerial video, weigh in motion stations, and detecting electronic toll collection.

  The Travel Time Data Collection Handbook suggests a check for existing data sources that match the purpose and objectives of any study. For the purposes of this thesis, we will be exclusively using a preexisting data set. This does limit

---

13 Turner 1-4.
us because concepts can only be demonstrated if the data obtained is of proper scope. Because of the existence of this data source, the final steps of this guide to travel time collection do not relate directly to this thesis. I list the steps below for reference.

- Develop Data Collection Schedule & Equipment
- Conduct Training
- Perform Pilot Studies or Trial Runs
- Collect Data
- Reduce Data & Quality Control

We will return to the concepts of the *Travel Time Data Collection Handbook* when discussing the data used in this thesis and as we discuss the calculation and use of travel times. Now, we move from the academic concepts behind traffic modeling and travel time estimation to discuss the reporting and estimation of traffic and travel times for use by people in real-time.

2.2 Information for Use by Real People in Real-time

2.2.1 The Beginnings of Traffic Reporting

Traffic and the problems it causes are certainly not new. Carole Sauve writes, “During the Roman Civilization, Julius Caesar became so frustrated by traffic congestion
that he banned the movement of carts during daylight hours…this stands as the world’s first traffic report.”

In 1864, the first gasoline powered automobile was produced in Germany. The first radios began being installed in commercial vehicles in the 1930’s by Motorola. This allowed for broadcast of information to people while in transit. At first, news, entertainment programs, and weather information were broadcast. Although it is difficult to tell exactly when radio stations began to broadcast traffic information, the first documented radio traffic report occurred in San Francisco in 1957. Private pilots often reported on the weather from the sky. One morning, a pilot working for KSFO-AM reported, “A stalled car on the upper deck of the Bay Bridge…and commented that as a result, traffic was backed up to the toll plaza.” Listeners responded with enthusiasm, requesting more traffic reports, and soon the pilot was making daily reports about traffic in the Bay Area. Popularity soon spread, as KMPC in Los Angeles reported traffic conditions from two aircraft the very next year. It is impossible to tell whether the information being provided was of any use to drivers, whether it affected any travel decisions they had to make, or whether it was simply of entertainment value to listeners. What we can take away is that people consciously began to think about reporting traffic conditions.

---

18 Chan 8.
19 Chan 8.
2.2.2 Development of the Traffic Reporting Industry

Today, traffic reporting is much more common than even ten years ago. Information is collected by aircraft, police and highway patrol radio frequencies, cell phone users who report accidents or conditions, closed circuit television cameras, and sensors built into the highway infrastructure. The traffic information collection and reporting industry, if we may call it an industry, began in the private sector. The largest private traffic reporting service is Metro Networks, based out of Houston, TX. Metro Networks currently collects information in any of the ways listed above for almost all of the major metropolitan areas in the United States and provided reports in the form of broadcasts to 1275 radio stations and 110 television stations nationwide as of 1999.\textsuperscript{20} Until 2000, the second largest service was Shadow Broadcast Services, based out of Philadelphia, PA. Shadow provided a similar service to that of Metro Networks, but to a smaller number of markets. Shadow also provided news, sports, and other broadcasts and had gone through several reorganizations in the past ten years.\textsuperscript{21} In the past three years, Westwood One, the parent company of Metro Networks, acquired Shadow Broadcast Services, and the two companies now operate as one source of traffic information.\textsuperscript{22} The existence and actions of these companies and their survival is a testimony to the value of traffic information. If no value was being provided by reports, the companies would not be able to make money from them and cut the programs. Again, it is difficult to tell if the value provided is entertainment, or if people use traffic information to help make driving decisions.

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{20} Chan 10.
\item \textsuperscript{21} Chan 11.
\item \textsuperscript{22} Company Profile. 2001. Westwood One. <http://westwoodone.com/aboutus_co_profile.asp>.
\end{itemize}
\end{footnotesize}
From the 1950’s to the 1990’s, traffic reports were constrained to television and radio. The development of the cell phone and Internet led to new possibilities for the transmission of traffic information. These new technologies have many advantages over a broadcast technique for transmitting. First, they are available anytime. Radio station traffic reports usually come “at the ten’s,” or at differently spaced intervals throughout a stations programming. This does not make it easy for a driver to get traffic information when he wants it. The same is true of television. Additionally, traffic reports generally last less than a minute and a half, and therefore can only provide a limited amount of information. Because of this, the driver may not be provided with information that is relevant to his trip because a station deems that not enough listeners will be affected by that specific information. Internet, cell phone, and other types of non-broadcast traffic reports have advantages over broadcast reports in that the reports are generally customize-able and available anytime.

Internet sites generally contain both text and maps. Maps may display speeds numerically or as color coded roads. Incidents are often displayed on maps as icons. Additionally, users can often choose a road or area and get a traffic report specific to this request. One drawback of Internet sites is that there is no standard for providing information. Some services provide maps with speeds; some provide maps with a congestion measure. Some sites provide travel times, and some provide volume counts. Notable traffic websites include metrocommute.com, smartraveler.com, etaktraffic.com, accutraffic.com, traffic.com, traffic411.com, trafficcast.com, and trafficonline.com.

Each of these websites acts as a provider to multiple cities, and contains a variety of the information described above. Often, the site will provide graphics or information

---

23 Chan 14.
to local information providers such as newspaper web pages or TV channel web pages. Metro Networks runs the etaktraffic.com site and uses its established information collection methods to provide information for both the Internet as well as radio and television broadcasts. These sites stay in business by providing billable personalized services, allowing advertisement on their site, selling information to news agencies, or partnering with governmental agencies. The sites that survive will be the sites that provide enough value to decision makers to make them willing to pay for traffic services.

Cellular phone services offer many of the same advantages as Internet websites but add one key feature: people can access the information from their vehicles. There are services set up such that any cell phone user can dial a specific number, enter a small amount of information, and get traffic reports for any number of specific road segments. This can be useful to check upcoming roads, or while stuck in traffic, to learn the severity of the traffic situation and what could be causing such a delay. Drivers could then possibly choose another road they know to be an alternate route and check the traffic information for that location. A prime example of such a service is offered by Smart Route Systems.\textsuperscript{24} SmartRoute provides phone numbers for users to call and then gives an audio menu for which a user then must choose a route. This is very similar to the act of clicking a link on the Internet except that the driver can obtain the information while in transit.

\textsuperscript{24} Chan 18.
2.2.3 The Public Sector and Data Collection

More recently than the development of private sector traffic information, the government has become involved in reporting the conditions on the roadways it owns and maintains. Governments began to collect traffic flow and travel time information for the purpose of maintaining roadways. At first studies were done manually and as technologies became available, states and local governments began to develop ways of automatically collecting information about their roads. With the development of data collection, we begin to see the implementation of road message signs and we see the development of websites.

Through my work on this thesis, I had the opportunity to contact members of each of the 48 state Departments of Transportation in the continental United States. While states now provide some information about their roads to the public, four distinct stages of the development of public road information are clear at the state level: 1) No data collected, 2) Data collected, but not shared, 3) Data collected in real-time and ability to share information is in development, and 4) Data is collected in real-time and shared with the public.

States that do not collect data tend not to because of one of two reasons: they do not see a need or they do not have a sufficient budget. For example, the Vermont DOT notes that it does have a monitoring system, but that it does not collect any information, as it does not believe the information would be of any use.25 Additionally, the Rhode Island DOT notes that it does have the infrastructure to collect speed data, but due to loss of staff and budget, they do not collect any data at current.26

---

26 Bucci, Joe. “Real Time Traffic Information.” E-mail to Christopher Schrader. 9 July 2002.
States that collect information but do not provide it to the public are generally collecting information to analyze their roadways, and the information is not collected in real-time. Arkansas\textsuperscript{27}, South Carolina\textsuperscript{28}, Michigan\textsuperscript{29}, and Nevada\textsuperscript{30} are examples of states that fall into this category. According to the Arkansas DOT, this is a very developmental place to be as most states realize the need for real-time information and are slowly headed in that direction.\textsuperscript{31}

Among states that are currently on the path to displaying roadway information to the public are Colorado\textsuperscript{32}, Oregon\textsuperscript{33}, Connecticut\textsuperscript{34}, and Massachusetts\textsuperscript{35}. Connecticut is currently installing infrastructure near Hartford with the intent of displaying traffic information to the public. The other three states given as examples have infrastructure, and receive data in real-time. They are all currently working to display that information to the public via the Internet.

Finally, several states do have functioning Internet sites that provide traffic information to the public. Wisconsin\textsuperscript{36}, Ohio\textsuperscript{37}, Georgia\textsuperscript{38}, Minnesota\textsuperscript{39} all maintain websites with some form of roadway information. Among the problems with current information sources is a lack of standards for displaying information. The \textit{Travel Time Data Collection Handbook} suggests several different formats.\textsuperscript{40} However, formats are

\textsuperscript{27} Flanagan, Ed E. “Real-Time Traffic.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{28} Beck, William B. “RE: Real-Time Traffic Information (Email Generated by Comments Page).” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{29} Parsons, Bob. “Re: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{30} McCurdy, Bryan. “Nevada Speed Monitoring.” E-mail to Christopher Schrader. 8 July 2002.
\textsuperscript{31} Flanagan, Ed E. “Real-Time Traffic.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{32} Thomas, Scott. “RE: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{33} Marchant, Jack. “RE: Traffic question.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{34} Healy, Michael. “Real time Traffic information.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{35} Bond, Russ. “Info. Request.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{36} DeCabooter, Phil. “FW: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{37} Saylor, George. “Traffic Data.” E-mail to Christopher Schrader. 15 July 2002.
\textsuperscript{38} Waters, Marion. “Real time traffic information for speeds.” E-mail to Christopher Schrader. 2 July 2002.
\textsuperscript{39} Osborn, Patrick. “re:fwd Traffic Data.” Email to Christopher Schrader. 25 July 2002.
\textsuperscript{40} Turner.
certainly not identical across states. Some states, such as Wisconsin, provide travel times. Others provide maps with congestion levels or volume occupancies. Still others provide information about speed at any number of sensors.

This paper involves the development and use of travel times. From the previous discussion, it is clear that travel time information, while available in certain instances, does not exist in bulk for the majority of the nation’s roadways or even highways. We will use one public data source from the Wisconsin DOT that is formatted into travel times. As we will see in the next section, previous attempts to use real-time travel times in practice have required the creation of a specific set of data created and monitored by the people administering each specific project.

2.3 Attempts at Combining Real-Time and Historical Information

The previous sections deal mainly with ways in which traffic information is collected and reported. This section is devoted to introducing several applications in which the collection of information was used to predict future travel times. This, of course, is also the main goal of this thesis. Applications described herein use many of the concepts that we use in chapters three and four, but also have some limitations that will be evident in our brief discussion of them.

ADVANCE

While many public and private entities display real-time information, very few attempts to combine real-time information and historical information to forecast travel
times exist. Of these, the ADVANCE Project in Chicago, IL is the largest functional test. The description below highlights important facets relevant to this thesis and is an amalgamation of concepts discussed in Bowcott, ADVANCE Information Source, and ADVANCE Project Formal Evaluation.

ADVANCE stands for Advanced Driver and Vehicle Advisory Navigation Concept and involves the combined efforts of the Federal Highway Administration, the Illinois Universities Transportation Research Consortium of University of Illinois at Chicago and Northwestern University, the Illinois Department of Transportation, the American Automobile Association, and Motorola, Inc. ADVANCE’s goal was to collect and use traffic information to affect the route choice of drivers by providing them with least cost solutions. By doing this, ADVANCE hoped to show that providing the right type of information allows drivers to avoid traffic hazards and reduce trip travel time.

Four systems make up the organizational structure of ADVANCE. The first of these is the Mobile Navigation Assistant. This system is an in-vehicle device that has both hardware and software components. The hardware includes a memory card, speaker, compass, antennas, sensors, RF modem, GPS receiver, navigational computer, CD-ROM drive, and display unit. The software component both calculates a route for the driver and provides him with information. First, destination is requested via the display unit and then current location is calculated from a GPS signal. The Mobile Navigation Assistant uses travel time information received from the other three components of ADVANCE to calculate and optimal route for the driver. The driver is given visual and

---

audio cues as to how to follow the route. If he deviates from the route, the device will query as to whether it should recalculate the route from the current location.

A second system, the Traffic Information Center (TIC), collects and distributes traffic information. Real-time information is collected via loop detector readings, probe vehicle observations, reports of traffic incidents, and weather reports. Observations are sent to the TIC remotely by radio frequency. Once received, a third system, Traffic Related Functions, processes the information into a useable format. The TIC then relays this information to vehicles containing Mobile Navigation Assistants in a format that provides a travel time and link ID number for each road segment in the ADVANCE study area.

Traffic Related Functions, the third of ADVANCE’s four systems, are a set of algorithms that operate within the Traffic Information Center. These algorithms create usable travel time estimates by combining the information collected by the TIC. Original estimates of travel time are created from a network equilibrium flow model similar to that described in Section 2.2.1 that generates historical profiles based upon theoretical measures. We will look to improve upon that in Chapter 3 by creating historical link profiles based previously recorded travel times. The Traffic Related Functions then create profiles for all links based on received information. This information, in combination with the original estimates, is used to forecast travel times 5, 10, 15, and 20 minutes into the future. If traffic conditions were forecasted to be abnormal, an updated travel time was sent to all Mobile Navigation Assistants and used in route guidance.
Messages to and from Mobile Navigation Assistants and the Traffic Information center were sent by Radio Frequency (RF) communications, the final part of ADVANCE’s operating system.

During development in the early 1990s and deployment in 1994 and 1995, many questions arose with the development of such a system and a body of papers developed to explain concerns and implementation. Topics ranged from how to combine information, to how to predict travel times, to how often travel time observations were required, to how to establish initial estimates of travel time, to eliminating bad data, to dealing with accuracy and reliability of information. The papers in the collection served as reference point for much of this thesis and as such are referred to often throughout the next two chapters. An alphabetical listing of the papers produced from the ADVANCE project can be found at <http://advance.dis.anl.gov/advance/reports/alpha.listing.html>.

Throughout this thesis, we focus on concepts similar to those of the Traffic Related Functions. We look to generalize many of the concepts discussed in ADVANCE papers, focusing on the forecasting of travel times. As in ADVANCE, our efforts are aimed at developing both an initial estimate of travel times with historical information and forecasts with real-time information. In the advance project, original link travel times were created from a flow model, still a theoretical measure. We look to improve on this by creating a historical profile based upon previously observed travel times. When creating this profile, methods were implemented in advance that generally kept profiles for discrete time intervals throughout each day. We look to improve on this by developing a continuous parameterized function to describe the historical profile of a link.

---

at any time of day. Finally, smoothing methods were used in the ADVANCE project to predict future travel times. In Chapter 4, we take a slightly different approach by including trend factors in smoothing based on patterns observed in historical information.

**Navigation Technologies & TrafficCast**

More recently, private companies have begun to see the use of data to forecast travel times as a reasonable source of income. On October 15, 2002, two companies, Navigation Technologies and TrafficCast, announced their joint intention to provide in-vehicle route guidance and trip advice by combining historical roadway profiles with real-time information and using concepts similar to ADVANCE.\(^{45}\)

Essentially, this partnership is an extension of the ADVANCE Project. Navigation Technologies was involved with funding and scenario tree development for ADVANCE’s route choice software.\(^{46}\) Now, they look to combine most necessary elements of the ADVANCE project to forecast travel times for commercial use. Working with information from state DOTs, police, emergency dispatchers, and other governmental agencies, TrafficCast and Navigation Technologies hope to combine historical traffic information with weather, incident, and construction information.\(^{47}\) The end goal of this project is to create travel time forecasts that a vendor may provide wirelessly to any user in his home or vehicle to help him make informed decisions. Notice that the listed information does not contain direct measures of travel time, but rather quantities that may predict it. Currently, the project is in development, and was

---


\(^{46}\) ADVANCE Project Formal Evaluation.

\(^{47}\) Navigation Technologies.
demonstrated on the small scale at the Intelligent Transportation Systems World Congress this past October in Chicago.\textsuperscript{48}

Little focus appears to be given to real-time travel time observations. Instead, efforts seem to be placed on obtaining information about factors that may influence or predict real-time observations. Navigation Technologies’ approach claims that future travel time is a function of current predictors of travel time and historical profiles. Our approach focuses on travel time observations themselves. Therefore, we use real-time measures of travel time along with a historical profile to produce a forecast of future travel time.

\textsuperscript{48} Navigation Technologies.
Our goal in this thesis is to create forecasts of travel times that will aid decision makers. We have identified two pools of information for this task: historical travel times and real-time travel times. This chapter looks individually at historical travel times, and breaks that problem down into several parts. First, we take a closer look at the definition of travel times and create our own definition. Next, we establish a methodology for developing an estimate of travel times from historical data. This step is the bulk of the work accomplished in this chapter and involves examining applicable data, investigating time categories, and developing a parameterized function as an estimate of travel time.
The steps needed to create this function are applied to road segments in the Milwaukee Highway System.

3.1 Defining Historical Travel Times

The Federal Highway Administration defines travel time as:

“The time necessary to traverse a route between any two points of interest.”

The Lake County Transportation Improvement Project board defines travel time as:

“The time it takes a vehicle to traverse a section of roadway.”

The University of Minnesota ITS Institute defines travel time as:

“The amount of time required to travel from one point to another on a given route.”

These definitions highlight three important characteristics which will help make up our definition of historical travel times: *sets of beginning and end points, a route between those points, and a time it takes to traverse that route*, the travel time.

Beginning (A) and end (B) points set specific boundaries on where time can be measured. It is important to have them well defined otherwise the time calculation is meaningless. Additionally, if one is to calculate a route travel time that is between two

49 Turner 1-5.
non-adjacent points of interest, he may have to add the travel times of intermediate sets of beginning and end points. *

It is also important to define a route between each set of A and B points. In any road network, many routes between any two points are possible, each with a different travel time at any instant. Defining a route clearly identifies the entire segment of road for which a travel time is associated.

The concept of the time it takes to go from point A to point B on a specified route is then self-explanatory. However, the above definitions are missing one important concept that will allow us to define a historical travel time: they do not include any notion of observation of travel time or when that observation takes place. To say that we know what a travel time was in the past we must both have observed that travel time and know at what time it was observed.

Note that by definition, we cannot assign a travel time to a single instant. For the remainder of this thesis, we follow the convention of assigning a reference time to each observation representative of when it ended. Only at this time is all the necessary information in describing a travel time known. Thus, if we say an observation of 37 seconds for the travel time from A to B was recorded at 12:00:00, we mean that some measure was taken that indicates someone passing A at 11:59:23 would then pass B at 12:00:00. Therefore, we can define a historical travel time as a data point that can be received or constructed, that reflects the time it takes to traverse a specific route from one location to another ending at a given time in the past.

* Other considerations do come into play when adding travel times. For example, there may be an intersection that is represented by an endpoint. In order to add travel times on links adjacent to this point, one ending with this point and one beginning with it, we must take into consideration the time it takes to go through the intersection. We do not consider this problem here, and leave it as an area for further work.
Inherent in our definition is the fact that a travel time can be directly measured or constructed from some other measure. As described in Chapter 2, the agencies that collect traffic information do so in a number of different formats. They therefore must be able to arrive at travel times in different ways. Actual time measurements are ideal, but not extremely common in current practice. More common are calculations that are extrapolated from other data, often speed or volume measurements. Speed measurements are straightforward in the fact that the units always involve time. The assumption that a measured speed is an average over a certain length of roadway must be made in order to calculate travel time. With this assumption, travel time is simply the distance through which the speed is assumed an average for divided by the measured speed. Equation 3.2 explains this calculation. Travel times obtained from Milwaukee were calculated in this fashion by the DOT.

As seen in Equation 3.3, calculations based upon volume are a bit more subtle but can be conceptually simplified to be the number of vehicles on a roadway during a given time period multiplied by the length of that time period.\textsuperscript{52}

\textbf{DEFINITION:} A historical travel time is a data point that can be received or constructed and reflects the time it takes to traverse a specific route from one location to another location ending at a given time in the past.

**Calculation for Travel Time Based on Speed Measurement:**

\[
\text{Estimated Travel Time (sec) = } \frac{\text{Segment Length (km)}}{\text{Time Mean Speed (km/h)}} \times (3,600 \text{ sec/hour})
\]

\[
\text{Time Mean Speed} = \text{avg speed} = \frac{\sum v_i}{n}
\]

\[
v_i = \text{speed of } i^{th} \text{ vehicle}
\]

\[
n = \text{number of observations}
\]

**Equation 3.1: Travel time calculation based on speed.**

---

**Calculation for Travel Time Based on Volume Measurement:**

\[
TIT = \frac{\sum_{m=1}^{m} \sum_{n=1}^{n} k_{it} (L_i N_i)}{m} \cdot T
\]

Where:

- \( TIT \) = total travel time expended in the system during the selected time period (vehicle - hours)
- \( n \) = number of subsections in the system
- \( m \) = number of observations during the time period selected
- \( T \) = time period of observations (hours)
- \( k_{it} \) = density in subsection \( i \) at time \( t \) (vehicles per lane - mile)
- \( L_i \) = length of subsection \( i \) (miles)
- \( N_i \) = number of lanes in subsection \( i \)

**Equation 3.2: Travel time calculation based on roadway volume.**

We concede that there are subtle differences within each of these methods, and subtle differences in their accuracy and variability. Nevertheless, once travel times are calculated, then the methodology described later in this chapter can be applied regardless of the way the travel times were obtained.

---

53 Turner 1-7.
54 May 213-214.
3.2 Examining Travel Time data

Now, we move on to describe the process by which data measurements can become useful. This process involves two steps. First, we clean any data that is to be used. Then, we use a cleaned data source to develop a representation of historical travel times. At the end of this section we apply these concepts to the Milwaukee Highway System.

3.2.1 Good Data, Bad Data

The process by which we obtain data and clean it so that it is usable in developing representations of historical travel times is described in this brief section. The process involves obtaining a data set and eliminating bad data.

In obtaining a set of historical travel times, it is important to keep in mind the different parts of our definition from Section 3.1. Any obtained data must have a time stamp to tell us when (in history) the travel time will be for. The data must also contain location information referencing start and endpoints of the measurement. Note that we skipped over a fairly large step here. We have assumed that data is received with a travel time indicated. As mentioned in 3.1 this is often not the case. If it is not, the data must be in a format such that beginning and endpoints can be inferred. For example, if a probe vehicle is reporting data, its reports must contain a timestamp as well as location data. From any two consecutive points, the user may infer a start point, endpoint, and time in-between. Additionally, if a sensor is reporting speeds, travel times must be calculated and include reference to specific endpoints for which he assumed the speed to be an
average. We suggest the creation of travel times in the following format and continue our
discussion assuming use of a format of similar nature:

<table>
<thead>
<tr>
<th>Travel Time Format:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Point (A), Endpoint (B), Time Stamp (t), Travel Time (TT)</td>
</tr>
</tbody>
</table>

This format contains all the information relevant to describing travel times
prescribed by our definition. Note that travel time information must indicate a route,
which the above formulation may not. We assume this route implicit in the Start Point
and Endpoint as the optimal path between the two. Without this assumption, we are
forced to include a large chunk of information in the above that will vary in length based
upon the complexity of each individual route chosen.

Bad data can be defined as any data that will not accurately represent the travel
time over a given segment of roadway. Bad data can be recognized and thrown out if it
does not provide relevant or correct information within the confines of our definition of
historical travel time. Conditions for throwing out data include the following:

• If any one of Start Point, Endpoint, Start Time, or Travel Time does not exist
• If any of our four items does not “make sense”
• If no route can be implied from the original data source

If one of the above quantities does not exist, the entire observation will not
provide us with any useful information. Defining whether the quantities “make sense” is
more difficult and mainly a matter of judgment. Obviously if the time is outside of the
definable range of times, it does not make sense. For example, if a travel time is zero or
negative, it does not make sense. Beyond that, the determination of what makes sense is
application specific. This concept will become more apparent when we deal with the
Milwaukee Highway System at the end of this section.
Above, we have assumed the path from A to B to be the optimal path given a network flow problem. A quick example of how a path cannot be inferred involves the dynamic routing software for ALK’s Copilot. Copilot provides a driver with an optimal path from his current location to a point B and collects location information with a timestamp. It is possible to construct travel times based upon this data, and a route can be implied if the driver follows copilot’s instructions. Copilot will recalculate a route if a driver deviates from the calculated path. Therefore, if between any two recorded data points Copilot recalculates, then the path between those two points cannot be the optimal path. Consequently, the travel time calculated between them is identified as invalid, and the point is thrown out.

3.2.2 APPLICATION: Milwaukee Highway System

In doing this study, we looked for examples of travel time data sets with which to apply the concepts being presented in hopes of showing that one could predict useful travel time information. One such data set was received from John Mishefske at the Milwaukee Department of Transportation via email.\textsuperscript{55} We were fortunate to receive a data set with many necessary elements and calculations completed for us.

Data was for seventeen A to B paths collected at inconsistent time intervals every several minutes for the entire month of June 2002 and July 6, 2002. Links included segments from intersection to intersection of major highways as well as from major exits to intersections. A complete description of these links may be found in Appendix A.

\textsuperscript{55} Mishefske, John. “RE: Traffic Information (Summer Contact).” E-mail to Christopher Schrader. 17 February 2003.
In all, 224,621 data points were received in the format shown in Figure 3.1. The data elements in our definition are evident in this format. A timestamp including time of day and date is included twice, endpoints and path are noted directly after the timestamp, and travel time is indicated at the end of each observation.

![Figure 3.1: Travel time data for Milwaukee Highway System in received format.](image)

Travel times for this data set were calculated from speeds collected by a series of “Trap” detectors along each of the seventeen paths. Each travel time was then calculated by dividing the length of a road segment by the recoded speed on that segment. Several segments represented each path in the data set. Relevant travel times were added together to get a total path travel time. Times at which an insufficient sample size of speed measurements occurred due to equipment malfunction or lack of traffic were reported as null values and speeds were capped at 60 mph by the Milwaukee DOT to reduce the likelihood of unrealistic travel times.\(^{56}\)

We then parsed this data into a more usable format in Microsoft Excel. Path was implied from start points and endpoints and a single timestamp was allocated to each set of seventeen data points as they were always received simultaneously. The calculated time stamp contained two parts, date and time in seconds from midnight. Data was then in the following format:

\(^{56}\) Mishefske, John. “RE: Traffic Information (Summer Contact).” E-mail to Christopher Schrader. 24 February 2003.
Upon initial inspection of data, two major problems arose that would demand removing some data points from our analysis. First, there were often reportings of travel times equal to zero. With one exception, every A and B combination showed at least some observations of zero, while two links showed virtually all zeros for the month of June. As stated previously, travel times of zero are infeasible and these data points were therefore removed. Additionally, on one of the links there appeared to be a “cloud” of travel times that were substantially higher than any other measures of travel time. This link was also the link that did not show any zero values for travel time. As we would expect travel times to grow in a continuous fashion from free flow to very congested, this cloud seemed very unusual. It was hypothesized that for this link, extremely high travel times were reported instead of zero values when equipment was malfunctioning. These points were subsequently removed.

After removing bad data from the received database of Milwaukee, there were a different number of observations for each link. Number of observations ranged from 95 to 13,211. Because of this discrepancy, an array was created for each A, B pair containing timestamp information and corresponding travel time information for all non-
removed values. Two endpoint sets showed fewer than 100 observations remaining during the one-month period. Of these 100 observations, most occurred at free flow conditions, and as such, no relevant information could be obtained by studying them further. They were discarded as unusable.

3.3 Representing Travel Time

Once we had an acceptable data set of historical travel times, we set out to model this information to create an estimate of travel time that will be useful in predicting future travel times. Our approach, developed through empirical observation and analysis of the Milwaukee Highway System, produces a continuous parameterized function of travel time as a function of time of day.

3.3.1 Time Elements

Under ideal circumstances, we would want to be able to know a historical expectation of travel time along any route given a number of explanatory variables. Possible explanatory variables include weather conditions, reports of incidents, reports of construction, special events, and time elements.\(^57\) Thus, we have:

\[
\text{Historical Travel Time} \sim f(\text{weather, incidents, construction, events, time})
\]

Equation 3.3: Elements of historical travel time.\(^58\)

We concede that all of these explanatory variables \textit{absolutely} have a direct impact on travel time. We would like to be able to input the fact that there is a concert tonight at


a specific location and that it is raining heavily. However, these variables tend to be highly volatile. The amount rain affects traffic depends heavily on the severity of the rain. The amount a concert affects travel time depends heavily on the level of popularity of the performer. We focus on the most concrete of these explanatory variables, time elements, as our main predictor of travel time.

The *Travel Time Data Collection Handbook* suggests the organizer of any historical travel time study should conduct the study when other factors are minimal.\(^{59}\) As such, we suggest that, if possible, times when there are special events or conditions be cleaned from the data before creating travel time estimates for normal conditions.

That being said, the *Travel Time Data Collection Handbook* mentions the following time elements for consideration: month, day of week, day type, and time of day.\(^{60}\) By day type, we mean to include a category for holidays. We add another category not mentioned in the *Travel Time Data Collection Handbook*, year. We hypothesize that travel times may show a year-to-year trend, becoming more congested with time.

<table>
<thead>
<tr>
<th>Time Elements for Consideration:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Month</td>
</tr>
<tr>
<td>• Day of Week (Weekend, Weekday, or specific day)</td>
</tr>
<tr>
<td>• Day Type (Holiday)</td>
</tr>
<tr>
<td>• Time of Day (Actual Measure or discrete segments)</td>
</tr>
<tr>
<td>• Year (Is there any year-to-year trend?)</td>
</tr>
</tbody>
</table>

\(^{59}\) Turner 2-11, 2-13, 2-25.  
\(^{60}\) Turner 2-11.
3.3.2 Creating an Expectation of Travel Times

Because of the time elements involved, we may immediately think of modeling historical travel times with some sort of time series. This is certainly a possibility. In papers describing processes for the ADVANCE project in Illinois, Ashish Sen, et al. mention ARIMA time series as one possibility for using historical information to predict future travel times. With enough historical data and the right computational tools, one can expect to get an accurate representation of year-to-year trends in travel times, seasonality throughout the year (possibly even including recurring holidays), seasonality throughout the week, and seasonality throughout the day.

However, implementing an ARIMA model necessitates that observations are received at regularly spaced intervals. In the current data collection environment mentioned in Chapter 2, data is not generally collected at regularly spaced intervals. One certainly can extrapolate a regular time series from an irregular time series, especially with travel times since measurements of travel time are not likely to vary much over a small time span. However, the complexity of this increases computational effort by both the computer and an individual, as it may involve a subjective component.

Additionally, in order to model a time series, the data must first be made stationary. Sen, et al. notes that recorded travel times are not stationary due to a diurnal pattern in which morning and afternoon peak periods occur repeatedly. Add weekly seasonality, yearly seasonality, recurring holidays, and year-to-year trends, and making a

---


63 Sen.
time series of travel times stationary becomes a problematic task. Again, this can certainly be done, but requires a subjective user judgment that cannot be done efficiently for a large number of different A, B pairs.

Also, given the current state of data collection, historical data is not as available from three years ago as it is now, and not as available now as it will be three years from now. As such, it may be difficult to find year-to-year patterns or yearly seasonality simply due to a lack of availability of data. If patterns do exist, they may not be representative of what is actually going on due to small sample size. For this reason, we will focus in on day of week and time of day as explanatory variables describing travel time.

**SUGGESTION:** Consider Time of Day and Day of Week as the most important explanatory variables describing travel time.

Even with this in mind, the complexities of making a time series stationary still exist. This simplification does however make it easier to discretize travel time observations into time categories. This leads to simpler ARIMA approach hinted at but not explicitly stated by Shbaklo, et al.\(^6\) Any day can be broken down into discrete time sections. Within that section, an ARIMA \((0,1,1)\), or exponential smoothing, model can be applied. If an assumption of stationarity through days in any given time section is correct, the series will converge to a single value that can be used as an expectation for that time segment.

This solution, while feasible and somewhat simple, presents one of two problems. Either it breaks our historical expectations into segments that do not accurately represent

---

what is happening continuously throughout the day, or it creates a representation that has too many parameters, three for each time segment (endpoints and an expectation).

We do not see these as acceptable solutions, and create a solution which will more holistically provide an expectation of travel times given past data and which will not require as many parameters to describe it.

We first look for some other predictor that will describe the data set and then look to minimize the data requirements to describe that expectation. We expect this predictor to be able to describe trends throughout single days as well as different days of the week. Furthermore, empirically we expect the travel times to be greater during morning and afternoon rush hours. Figure 3.3 shows the characteristics we wish a predictor to describe graphically.

![Figure 3.3: Travel time as a function of time of day and day of week.][1]

---

65 Sen Figure 4.
Given the shape of the expected travel times shown in Figure 3.3, polynomial regression does not seem adequate. This leads us to consider nonlinear parametric and nonparametric regressions.

Nonparametric regressions such as kernel regression and the projection pursuit algorithm will provide very good expectations of travel time. However, these methods create estimates for a finite number of discrete times. The predictors are almost continuous if large amounts of data is supplied. However, because each individual estimate is given as a value, the number of parameters required to represent this method of estimation is equal to two multiplied by the number of times travel times estimated. This is entirely too large, and we search for a better solution.

With days of the week as an explanatory variable, nonlinear parametric regressions appear difficult. Days are a discrete quantity, and we must provide a function for which parameters can be estimated in order to perform this type of regression. However, if we execute a regression for each day of the week, we find that we are looking to estimate parameters of a continuous function with a single independent variable, time of day. This is very possible, and we are given the freedom to select the number of parameters we wish to use.

The question remains, then, what type of function to use to estimate travel time as a function of time of day. Our solution is a ten-parameter function that is the sum of three normal curves and a constant. We develop the conceptual background for this function both by empirical observation and by analysis of Milwaukee travel times. Our empirical reasoning is developed below, and our work with the Milwaukee data is described in Section 3.3.3.
Clearly, there is some non-zero minimum travel time for traversing any path. This is described by the line intersection the origin in the Figure 3.3. To get our composite function, we must add functions representing the morning and afternoon peak periods to this constant. We know that travel time is a direct function of volume from Section 3.1. Let us assume these peaks are in major part due to trips to and from work. Let us also assume that each individual travels to and from work at some average time with some variance. In his book *Stuck in Traffic*, Anthony Downs notes that work trips are concentrated during these peak times because most workplaces begin and end their day at the same time in order to make interaction between businesses possible.\(^\text{66}\) We then infer that any road segment will have some time at which it is likely to have the most volume, again with some variance. This directly leads to the notion that there is some static time of day at which travel time will peak on any road segment, again with some variance.

We notice from Milwaukee data that the morning bump tends to have a symmetric bell shape. This, along with the aforementioned logic, leads us to suppose that travel times during the morning peak period may be normally distributed around some mean. In evaluation of the Milwaukee Highway System travel times, afternoon peak travel times appear to show the same bell shaped curve, but with an extra hump before travel time reaches its maximum expected value. Evidence of this characteristic will be presented in the analysis section to follow. Empirically, we conjecture that this is an additional set of trips dependent on afternoon activities such as the recess of schools and

---

errands and can be represented by a third bell curve. Downs also mentions these sources of trips and contends that they often occur during a peak period.\textsuperscript{67}

Thus, we may conjecture that there are three bell curves, each representing one of the sets of trips described above, which correspond to travel congestion throughout each day. Each curve is multiplied by a coefficient and added together with a constant for a minimum free flow travel time, leaving a ten-parameter function. This ten-parameter function is a continuous estimate of travel time as a function of time of day. This function is presented in Equation 3.4.

\begin{equation}
\text{Weekday Travel Time}
\end{equation}

\[ TT = f(t) = K + C_1 \cdot \eta(\mu_1, \sigma_1) + C_2 \cdot \eta(\mu_2, \sigma_2) + C_3 \cdot \eta(\mu_3, \sigma_3) \]

Where:

\[ \eta(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(t-\mu)^2/2\sigma^2} \]

and \( K, C_i, \mu_i, \sigma_i \) are parameters to be estimated

\text{Equation 3.4: Continuous function estimating link travel time as a function of time of day.}

In sum, we have defined explanatory variables for travel time, suggesting day of week and time of day as the most important for travel time estimation. We have examined methods for modeling travel time based upon these variables and developed a ten-parameter function including the sum of a constant and three normal curves as a solution likely to represent travel time as function of time of day. The next section

\textsuperscript{67} Downs 15.
analyzes travel times on the Milwaukee Highway System to evaluate the feasibility of such a solution and show its efficacy.

### 3.3.3 APPLICATION: Milwaukee Highway System

In this section, we analyze the travel time data collected for road segments in the Milwaukee Highway System and fit the ten parameter function described in the previous section to each link, modeling travel time as a function of time of day. We perform this analysis of historical data to see if our solution will work with existent data, and evaluate the robustness of our solution.

After cleaning the Milwaukee data as described in section 3.2.2, we examine the remaining data to evaluate which time categories would be relevant to our analysis. Data was for the month of June 2002. As such, finding year-to-year trends as well as seasonality throughout a single year was irrelevant. We are left then with day of week, time of day, and holidays as explanatory variables. Figure 3.4 graphs travel time against time of day and day of week. It is clear from this diagram that there is a definite bimodal pattern occurring within time of day and that there are dissimilarities between days, especially between weekdays and weekends.
We further examine days of the week as a possible descriptor of travel time. Data was assigned integer values ranging from one for Monday to seven for Sunday. A typical plot for one of the fifteen paths is shown in Figure 3.5(a). Indeed, there is a noticeable difference between days. As mentioned above, this difference is especially large between weekdays and weekends. We ran linear regressions on both weekdays and weekends to see if there was a noticeable difference in the magnitude of travel times each day. Weekdays tended to show little difference across days. Weekends also were likely to show little difference between Saturday and Sunday, coefficients tending to be near zero. There was generally a significant difference in the intercepts of these two regressions for each link, indicating a difference in traffic patterns from weekdays to weekends, but not between different weekdays or different weekend days. This difference is highlighted in

Figure 3.4: Travel time as a function of day of week and time of day.
Figure 3.5(b). Weekdays were assigned a binary indicator of one and weekends were assigned a binary indicator of zero.

We come up with a similar pattern when examining travel times as a function of time of day. Weekdays produce significantly higher travel times than do weekends. Figure 3.5(c) depicts this and gives a representation of what typical travel time distributions will look like for weekdays and weekends. Notice that the bimodal pattern is evident, although the magnitudes of the two bumps will vary from link to link. Also notice that for this particular link, the afternoon peak period appears to have a second bump on the earlier side of peak afternoon rush. This pattern occurs for nearly every link and led to the addition of a third bell curve in the formulation of the equation for travel times as a function of time of day.

We have identified time of day and a binary weekday or weekend indicator as explanatory variables, and now simplify the process by one further step. We note that increased travel times on weekends were not due to a repeated pattern, yet single isolated incidents that did not recur.

Furthermore, it was noted that many increased travel times occurred in the hours surrounding midnight, both on weekdays and weekends. Further investigation revealed that a festival in the Milwaukee area ending at midnight, Summerfest, coincided directly with the dates for which travel times were highest around midnight. This falls under the category of a recurring holiday and as such should not affect normal travel times, which we are trying to predict. Inclusion of this information is absolutely necessary to predict travel time in the most robust manner. We suggest this as an area for further research with this data set and continue to try to eliminate such data so that we can model travel
time patterns under “normal” conditions. As such, we note that weekend travel times are generally free flow, and move on to computationally model weekday travel times.

Our first step in doing this is to run a kernel regression on each of the A, B pairs to confirm our hypothesis for a generalized ten-parameter function to estimate weekday travel time as a function of time of day. Figure 3.6 shows several plots of kernel
regressions superimposed over data points. All kernels show at least one of the peak periods while most show both. Afternoon peak periods do show a higher travel time to the left of the maximum peak travel time, indicating the presence of a third normal curve.

Notice that some links have moderate travel times in the hours around midnight. These A, B pairs tend to originate from Downtown Milwaukee, and as such may be attributable to Summerfest. As we are looking to model normal travel time conditions, we are hesitant to include measurements due to Summerfest in any representation of normal conditions.

Figure 3.6: Kernel smoother estimates of travel time as a function of time of day for all weekdays.
We must then decide how to represent these distributions in terms of the ten-parameter function described in the first half of this section. Two options present themselves: we can try to fit this function to the data points themselves or we can try to fit them to the kernel estimator. Both procedures were performed on several links. In each case, Microsoft Excel was programmed to calculate the ten-parameter function for any time indicator while allowing freedom to choose the ten parameters. Excel Solver was run, allowing each of the parameters to vary, and minimizing the Sum of Squared Errors between the ten-parameter function and either the kernel or the observed data. Parameters were constrained to values that did not represent “festival travel times.” This indicates that the means of the bell curves were constrained to peak times. Figure 3.7 displays one of the solutions we found. In this figure each of the three normal curves are shown originating at zero. Above them is their sum plus the free flow value compared against the data to which it was fit.

![Figure 3.7: Ten-parameter function fit to data from Milwaukee Highway System. Function is summation of three bell curves and a constant.](image-url)
There was little difference in the results between parameters estimated with kernel predictors and those estimated with actual data. Figure 3.8 depicts one example of the two ten-parameter functions. No two functions differed much more than this one. The main difference between the two methods appeared to be that functions produced with the actual data tended to be smoother. The subjective component of picking a bandwidth for the kernel estimator may have something to do with this result. Higher bandwidths produce smoother results and it is the task of the user to choose an acceptable bandwidth for each regression.

![Figure 3.8: Differences between ten parameter functions fit to data points and a kernel smoother.](image)

To eliminate this subjective choice of bandwidth, we perform the parameter estimation with the actual travel time observations. This was performed for each of the fifteen links in the Milwaukee dataset that remained after eliminating bad data. Results were consistent with our assumptions. In general, afternoon peak periods tended to last a longer period of time and in many cases had a higher estimation of travel time. Morning peak periods were shorter in length and peaked more quickly. Results are displayed
graphically in Figure 3.9. Estimated parameters and further graphs are contained in Appendix B.

![10 Parameter Function Fit to Weekday Data](image)

Figure 3.9: Ten parameter functions estimating travel time as a function of time of day fit to individual observations.

### 3.4 Missing Information

Although it is not the focus of this thesis, some mention must be given to road segments for which there is no available travel time information. This section is devoted to a shortened conceptual look at this topic. Four possibilities are introduced and suggested as areas for further investigation.
The first possibility is one that has already been suggested in the introduction: continue to use the information that produces 3rd generation route guidance travel time estimates. Recall that these estimates are generally created by associating a speed with any given roadway and then dividing the roadway length by that speed. Most often, that assigned speed is some sort of free flow measure or guesstimate of an average speed. This may seem like reverting to a previous state that we have tried to surpass, but in the end, it may be the only option available. It is obvious that observed travel times must be at or above free flow travel times. By associating a cost equivalent to a free flow quantity, we will be undervaluing the cost of traveling on any given link at some times throughout every day. This may make road segments with unknown travel times appear more desirable to a driver when in fact the expectation should be that the roadway is congested. Note that the opposite could also be true. We could assume that because no travel time observations exist, the road in question is not traveled as much and less important; therefore, free flow conditions might exist at most times. This is obviously case specific. For this reason, further investigation is needed into this area.

Second, a look at overall local road network trends may provide insight into travel times for additional road segments. The most straightforward example of this is taking an expected value of the local system dependent upon time, normalizing it by distance or a percentage of free flow travel time, and then applying this value to any road segment for which there is no information. To increase accuracy, this solution should then be broken down by road type. This approach assumes much. In the Milwaukee Highway System, typical peak periods do exist, but magnitudes of travel times varied greatly and depended on roadway direction. Using network trends may provide a better estimate than
free flow times alone, but still assumes too much to be applied in general cases, and must be looked at individually.

A third possible area of research that could lead to better estimates of travel time on road segments for which there are no recorded travel times is work with traffic flow models. By combining known information into the framework of this type of model, better theoretical measures of travel time on unknown road segments maybe developed. Examples of traffic flow models are many. We introduced a generalized model in Section 2.1.1. Adolph May’s *Traffic Flow Fundamentals* devotes an entire chapter to presenting different models of this type.⁶⁸ Shbaklo, et al also describe several different ways to model upstream and downstream traffic flows in order to predict travel time.⁶⁹ Perhaps with some of these models and pieces of travel time information, estimates of travel time for non-observed links may be established. However, with many examples to choose from, no single correct way of modeling traffic flow to predict travel times has been accepted. Thus, the procedure of modeling traffic flow to predict travel times requires further research.

A final area in which additional road segments could be assigned updated travel time estimates is an analysis of similar road segments. For example, if we do not have any travel time information for a given road segment, we could look at all road segments of a similar type within a certain distance of this first roadway. If a noticeable pattern occurred, we may be able to infer travel time estimates for the segment in question. Additionally, if we notice relationships between nearby parallel roadway travel times, we may be able to create a travel time estimate for a given no-observation road segment

---

⁶⁸ May 283 – 320.
⁶⁹ Shbaklo.
based upon a link parallel to it. This is all speculation and makes assumptions that may be proven incorrect. Hopefully, with enough work, detailed statistical analysis in the future will lead to conclusions that can produce travel time estimates in this way.

We might expect that a combination of effort in all of these areas will lead to the best estimate of travel times for roads with no existing data. In rural settings, fewer trip productions and attractions exist, causing less incentive to exit roadways and less volume in general. We may expect that a flow model would work well in this case without the complexities of urban traffic. In urban settings, we may expect that a comparison of similar road types to yield better results. We will not know without deeper investigation, and no conclusions can be drawn from this brief discussion. The topic of estimating travel times for links on which no travel time observations have been recorded is an area in need of further work.

3.5 Using Predicted Travel Times

Our goal in this thesis was to create a forecast of travel time that has the capability to aid decision makers by reducing their trip travel times. This chapter suggests a solution by creating that prediction with historical information. This section now briefly describes the use of this information and leaves the main application of it as an area for further work.

In Chapter 1, we discussed route guidance as the main use of travel time information. In the current, 3rd generation of route guidance, travel time is a measure based on speed and distance. In any method of route guidance, these travel time
measures are assigned as costs to the traveler. A computer algorithm then aids the
decision maker by solving a least cost network flow problem and provides the traveler
with the route that has the least cost.

With the information from a historical estimate, the costs on any network may be
updated. However, we must remember that each historical estimate corresponds to only
one instant in time, and that time corresponds to the end of the travel time observation.
Therefore, costs must be assigned accordingly. For example, if we are at a point A at
time \( t \), we can find the expected travel time to B by solving the simple equation:

\[
\Delta t = \tau_{AB}(t + \Delta t)
\]

*where:*  
\( \Delta t \) = current estimate of travel time  
\( t \) = current time  
\( \tau_{AB}(\cdot) \) = historical estimate function of travel time from A to B

*Equation 3.5: Estimated travel time from A to B beginning at time \( t \).*

This estimate of travel time, \( \Delta t \), may then be assigned as the cost on a link only at time \( t \),
and refer to the time span \( t + \Delta t \).

When the desired path of the decision maker is longer than any A,B pair for
which there exists a historical estimate, it may be necessary to combine historical
estimates from more than one pair of endpoints. The obvious way to do this is by
addition of adjacent segments. However, in each application, care must be taken to
ensure that all travel time has been accounted for. For example if a driver must go from
A to B and one possible path is A to C followed by C to B, there may be an intersection
at point C, and it may take time to go through this intersection. Additionally, adjacent
road segment travel times may not be independent, causing some discrepancies in
estimation of travel times over multiple road segments. These subjects are left as areas for further research.

As an explanatory example, we perform this operation on one path in the Milwaukee Highway System and display the results in Figure 3.10. We measure the estimated travel time from Moorland Road to the Mitchell Interchange beginning 50,000 seconds from midnight on any given weekday, or at 1:53:20 PM. We measure this time by way of the Zoo and Hale interchanges.

As shown in the table section of Figure 3.10, 220.1 seconds solves Equation 3.5 from Moorland Road to the Zoo Interchange beginning at 50,000 seconds. For a highway system, we expect the time to change between road segments to be negligible, making addition of segment travel times possible in this case. We continue as such. Beginning at 50,220.1 seconds, 317.03 seconds solves Equation 3.5 from the Zoo to Hale interchanges. And, beginning at 50,537.13 seconds, 351.7 seconds solves Equation 3.5 from the Hale to Mitchell interchanges. Thus, the estimated total travel time from Moorland Road to the Mitchell interchange is 888.83 seconds, or 14 minutes 48.83 seconds.

If costs are available for multiple paths between a start point and endpoint, route guidance may be used to select a route. By some least cost algorithm, a computer will be able to compare path travel times and provide the route to the decision maker that minimizes his travel time. In this way, the historical estimates we created can be used by a decision maker.
In this chapter, we have used historical data to create an expectation of travel time. Our solution was the development of a ten-parameter function with time of day as the independent variable. This function, the sum of three normal curves and a constant, matched our expectations when analyzed with data from the Milwaukee Highway System. However, this solution is only an expectation of travel time. In the next chapter, we add real-time information to this function and forecast travel times based on the combination of both historical and real-time information sources.
In discussing historical travel times, we have attempted to develop a strategy whereby an individual can know some expectation of the travel time along any link in a network of roadways and therefore know an expectation of the total travel time from any location to any destination using a given path. The goal in developing historical travel times is to replace a theoretical travel time with some actual predicted measure. However, all historical travel times can be is an expectation of what is to come. They are constructed from many individual observations, each of which will lie above or below our historical expectation. If we receive these observations in real-time, the accuracy of predictions for future travel times should go up, especially in the short term. It is the goal
of this chapter to describe a process to forecast travel times, combining real-time information with historical.

This chapter will be organized much as Chapter 3 was. Section 4.1 will give a brief definition of real-time travel times. Section 4.2 introduces the concepts necessary to short-term travel time forecasting and applies those concepts to the Milwaukee Highway System. Section 4.3 will discuss issues for further investigation. And, section 4.4 will relate the results of this brief study to the real world.

4.1 Defining Real-Time Travel Times

Recall our definition of a historical travel time: A historical travel time is a data point that can be received or constructed and reflects the time it takes to traverse a specific route from one location to another ending at a given time in the past.

A real-time travel time must contain all the same information. Two locations must be specified, a start point (A), and an endpoint (B). A route must be indicated between A & B. As mentioned in Chapter 3, it may be inferred that this path is the optimal path between the two points. The travel time is then the amount of time taken to go from A to B along the given path. Again, observations might not be received as measurements of time. As such, it may be necessary to extrapolate a time measurement out of an observed set of volumes or speeds.

Our definition of historical travel times also included a time stamp indicating when the observation took place. We adopted the convention that this time stamp referred to the time at which a vehicle traversing the path would pass point B, as that is
the only time when all necessary components of our definition would be known. For real-time travel times we again adopt this convention. By real-time, we mean now. Thus, for a travel time observation of \( t \) seconds from A to B received now to be a real-time observation, it must indicate that a vehicle passing A \( t \) seconds ago would currently be passing B. It is obvious that this concept mirrors that of historical travel times. Accordingly, our definition does:

```
DEFINITION: A real-time travel time is a data point that can be received or constructed and measures the time it takes to traverse a specific route from one location to another location ending now.
```

Thus, instants after a real-time travel time observation is received, it becomes a historical travel time. Often, we will refer to the “most recent real-time observation.” We will generally treat these observations as if they are in real-time even if they are received a short while ago because they provide the most relevant information to the actual conditions of a roadway. In the following section, we describe one possible process through which real-time travel times can be used to forecast travel times into the short-term future.

4.2 Real-Time Travel Times in the Real World

In order to produce usable information for a decision maker, we must produce an estimate of travel time for the seconds during which a driver will be actually passing from A to B. Ashish Sen, et al. write:
Estimates of travel time need to be for the time period when the vehicle actually traverses the link. Since a desirable route needs to be given when the driver asks for it, but the computation of such a route requires travel times which occur later, we need to be able to forecast such travel times.

It follows that our goal is to produce these forecasts. In Chapter 3, we came up with a solution to estimate of travel time as a function of time of day. This approach is static in nature and based strictly on observations of travel time in the past. In order to be more accurate in our prediction, we may look to include more information. If a real-time information source is available and collected, we will have a real-time measurement as well as recent observations and a historical estimate upon which to base forecast of travel time.

In this section, we take the combination of these information types on any path and look to predict future travel times along that same path. Sections 4.2.1 and 4.2.2 present two concepts and how they are necessary to this process, and section 4.2.3 describes their use in forecasting travel times for the Milwaukee Highway System.

### 4.2.1 Exponential Smoothing

Exponential smoothing is a method of “smoothing” a time series of observations. In this method, the most recent observations are given a high weight and previous observations are given lower weights that decrease exponentially with the age of the observation. This is where the name exponential smoothing comes from. Nearly all time series have an error or variance associated with their observation. By weighting past

---

Sen.
events in this way, this method creates a smoother picture of the time series without the noise associated with observation.

Three types of exponential smoothing are commonly mentioned in textbooks: single, double, and triple exponential smoothing. Single exponential smoothing weights an observation at time $t-1$ with the current smoothed estimate to get an updated estimate for the value of the time series at time $t$. This method is equivalent to an ARIMA (0,1,1) time series model and is given by Equation 4.1.

\[
S_t = \alpha y_{t-1} + (1 - \alpha) S_{t-1} \quad 0 \leq \alpha \leq 1 \quad t \geq 3
\]

Equation 4.1: Single Exponential Smoothing

Here, $S_t$ is the smoothed estimate and $y_t$ is an observed value of a given time series. A smoothing factor $\alpha$ is included to weight previous observations. By this formula, $y_{t-1}$ is weighted by $\alpha$, $y_{t-2}$ is weighted by $\alpha(1-\alpha)$, $y_{t-3}$ is weighted by $\alpha(1-\alpha)^2$, and so on. The smoothing factor $\alpha$ can be a constant or a direct function of $t$ such as $1/t$ so long as it is between zero and one and should be estimated because different values for $\alpha$ will produce different smoothing effects.

Double exponential smoothing adds a trend component to the single exponential smoothing concept. In this approach, an observation at $t-1$ is weighted with a previous estimate plus a trend factor, or estimated slope. This slope is also generally a smoothed estimate of the overall trend of the time series. As such, double exponential smoothing requires two smoothing parameters, given in Equation 4.2 as $\alpha$ and $\gamma$. Here, $S_t$ and $y_t$ again refer to the smoothed estimate and observation, respectively, and $b_t$ is the smoothed estimate of the trend of the time series at time $t$. By adding an estimate of the trend to the

---

previous estimate of the value of a time series observation, double exponential smoothing is able to fit a time series that has a trend better than single exponential smoothing, which will lag behind if the value of time series observations continually increases or decreases.

\[
S_t = \alpha y_{t-1} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1
\]

\[
b_t = \gamma(S_t + S_{t-1}) + (1-\gamma)b_{t-1} \quad 0 \leq \gamma \leq 1
\]

**Equation 4.2: Double Exponential Smoothing.**

The third common type of exponential smoothing is triple exponential smoothing. This smoothing method incorporates both a trend and a seasonal component. The seasonal component allows the smoothing estimate to better follow seasonal trends such as the oscillation of a wave or yearly temperature patterns. Equation 4.3 describes a specific instance of triple exponential smoothing named the Holt-Winters method. In this method, \(I_t\) indicates a seasonal index, \(L\) is a lag measurement of the number of periods in each season, and \(\beta\) is a third smoothing factor.

\[
S_t = \alpha \frac{y_{t-1}}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1
\]

\[
b_t = \gamma(S_t + S_{t-1}) + (1-\gamma)b_{t-1} \quad 0 \leq \gamma \leq 1
\]

\[
I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad 0 \leq \beta \leq 1
\]

**Equation 4.3: Triple Exponential Smoothing.**

Exponential smoothing provides us with a good way of following time series observations. It can also enable us to predict future events. If no observation is available at time \(t\), the only information that can be known is the smoothed estimate at time \(t\), and

---


trend and seasonality estimates if they exist. Because of this, the forecast of single exponential smoothing will always converge to a single value. Any forecast using double exponential smoothing will be the smoothed estimate at the time of the last observation plus a time increment to be forecasted for. Likewise, the forecast for triple exponential smoothing is just a double exponential smoothing forecast divided by the seasonal index for the period of the forecast. Figure 4.1 shows a time series with both a pattern and a trend. The time series ends at time 25 and future values are forecast. Notice that the blue single exponential smoothing converges to a single value, green double exponential smoothing continues to increase with a slight constant trend, and red triple exponential smoothing increases and follows the seasonality of the actual data.

![Figure 4.1: Single, double, and triple exponential smoothing compared to a seasonal time series with a trend.](image)

---

Exponential smoothing and adaptations thereof have been used in modeling traffic events. Shbaklo et al. introduce exponential smoothing as an excellent model for traffic quantities.\textsuperscript{75} They note a study performed by Levin in which volume data was collected and represented using several different ARIMA models and several different time lags between smoothing. Levin concluded that exponential smoothing, or ARIMA (0,1,1), produced the most statistically significant volume forecasts of the ARIMA models examined.\textsuperscript{76} We know that travel time can be modeled as a direct function of the volume of a roadway. Therefore, we can conclude that exponential smoothing will also provide the most statistically significant results when forecasting travel times into the short-term future.

Additionally, Shbaklo et al. mention an approach to travel time forecasting that involves both the concepts of exponential smoothing and a historical profile of travel time. This approach was developed by Hoffman and Janko and produces a forecast of future travel time not by weighting the most recent observations overall, but by weighting a recent observation with a historical smoothed profile for that time of day.\textsuperscript{77} That is, for a set number of time intervals throughout the day, a historical profile exists that is a smoothing of previous day’s observations. To create a forecast, Hoffman and Janko weight this historical profile against new information to create an updated forecast, essentially performing single exponential smoothing for each time interval throughout the day. Their method is given by Equation 4.4.
\[
\bar{t}_{l,n}^{(new)} = \theta \bar{t}_{l,n} + (1 - \theta) \bar{t}_{l,n}^{(old)}
\]

where:

\(\bar{t}_{l,n}^{(new)}\) = new travel time value of the std. profile for link \(l\) during time interval \(n\)

\(\bar{t}_{l,n}^{(old)}\) = previous existing travel time for link \(l\) during time interval \(n\)

\(\bar{t}_{l,n}\) = corresponding travel time during most recent observation

\(\theta\) = Weighting factor for the most recent observation

Equation 4.4: Hoffman and Janko historical profile travel time prediction method.\(^{78}\)

### 4.2.2 Recurring Traffic Patterns

Also important to our goal of forecasting travel times in the short term is an understanding of how travel times are likely to behave throughout any given day. We present several concepts here so that they may be incorporated into our process of prediction.

As we noted in Chapter 3, traffic patterns will have peak periods where the observed travel times are more likely to be above the free flow travel time. We go one step further here by noting that on a given day, current travel times are highly correlated to recent travel times; that’s why Shbaklo can say that ARIMA (0,1,1) works. Because of this, on any given day, if travel times are high during a peak period, they will remain high. If travel times are low during a peak period, they will remain low.

However, because of the natural cause of travel times, we can expect that there will still be a mean time at which people use the roadway with some variance. As such, we expect that a time series of travel times will exhibit the overall pattern of the historical model of peak periods. That is, we can expect travel times to increase to a maximum

---

\(^{78}\) Shbaklo.
point and then decrease after the maximum volume has occurred on the roadway. Overall, we expect travel times to mirror the general trend of the historical function developed in Chapter 3, and to mirror that trend either above or below the function, not reverting back to it during peak hours. Figure 4.2 shows this empirically.

This brings up another important dynamic. We can expect that there will be a time at which enough vehicles traveling during peak travel periods will reach their destinations to sufficiently lower traffic volume to that of non-peak periods. With the above description, we would expect this to occur after each peak period. Realistically, in many local road networks, there will often be a daytime travel volume that is above the free flow limit. In this case, we can still expect travel times to drop off significantly after the morning peak. We would very rarely expect nighttime travel times not to return to free flow in the absence of an event.

Fortunately, in the Milwaukee Highway System, kernels smoothed to free flow or near free flow travel times between peak periods as well as in the night and early morning hours. Therefore, this “return to normalcy” is inherent in our ten-parameter function modeling historical travel time. In Figure 4.2, the pink line represents this function for one particular link. Overlaid onto this function are the actual recorded travel times for one day in June 2002. The return to normalcy is clear in the actual data. After each peak period ended, the travel times along this road segment quickly returned to free flow. Data showed similar trends during almost all days on all segments for which data was obtained.

Additionally, note the pattern of travel times around peak periods in Figure 4.2. In each case, data mirrors the pink historical travel time function, residing at a relatively
constant proximity to it. In other words, if travel times are high for a given peak period, they will remain high throughout that peak period. Data was similar for nearly every day on every link examined.

Figure 4.2: Traffic patterns relative to a historical estimation of travel time for the Milwaukee Highway System.

Thus, we know that travel times will mirror our historical perception of travel times and that travel times will return to free flow conditions outside of peak hours. We now look to model these realities and predict them in the Milwaukee Highway System using a variation of exponential smoothing.
4.2.3 APPLICATION: Milwaukee Highway System

As we have already noted, the Milwaukee Highway system exhibits the properties described in section 4.2.2. These properties of travel times are inherently different for different times of the day, and our model must be able to include these dissimilarities.

During peak hours, we expect travel times to mirror the shape of the historical distribution. For this to happen, any smoothing and forecast should reflect the slope of the historical distribution as a trend. The next smoothing and subsequent forecasts should also rely upon the most recent real-time observation and other recent observations in decreasing order of importance. This leads us to use a concept similar to double exponential smoothing. Double exponential smoothing allows for a trend factor to be added on to any smoothed estimate and smooths that trend factor. In this case, we already know what the trend will be. It is the slope of historical distribution function.

Therefore, during peak periods we create a smoothed estimate of travel time by weighting the most recent real-time observation with the most recent smoothed estimate plus the difference in the historical times corresponding to the time of the previous smoothed estimate and the time associated with the current estimate. *

A forecast of the next period’s travel time would then just be the previous smoothed estimate plus the difference between the historical travel time function at the time of the smoothed estimate and forecast. This was implemented for the Milwaukee data in Microsoft Excel using Equation 4.5.

* Note that there are other possibilities. For example we could multiply (Tau(t)-Tau(t-1))*S_{n-1}/Tau(t-1) to keep the proportions the same, where Tau is our historical function. However, this has some problems. For example, if rush hour volumes occur early in the day and travel times are way above the historical estimate, the prediction for the middle of the peak period may be astronomical.
\[
S_n = \theta X_{n-1} + (1-\theta)[\tau(t_n) - \tau(t_{n-1}) + S_{n-1}]
\]

Forecast:
\[
S_n = \tau(t_n) - \tau(t_{n-1}) + S_{n-1}
\]

where:
\[
\tau(t_n) = 10 \text{ parameter function estimated in Chapter 3}
\]
\[
\{\theta\} \sim \text{smoothing parameters} \in \{0,1\}
\]

Equation 4.5: Smoothing and forecasting travel times given a historical estimate and real-time information during peak periods.

This concept will hold true during morning and afternoon peak travel periods. However, we noted a return to free flow travel times at non-peak periods. Thus, these times must be treated differently. We want our smoothed estimates of travel time to both decay to free flow travel times and to reflect the most recent real-time observation. For this reason, when working with the Milwaukee data, we smoothed the most recent real-time observation not directly with the previous smoothed estimate, but with a weighted average of the previous smoothed estimate and the free flow travel time. This is again essentially equivalent to double exponential smoothing in that the weighted average of a previous smoothed estimate and free flow travel time can be written as the smoothing term minus a trend term.

In forecasting during non-peak periods, we simply remove the dependence on a real-time observation and weight the most recent smoothing with the free flow travel time for each link. By the iterative nature of this process, an exponential decay to the free flow travel times is observed, as weights on previous smoothing decay exponentially. Equation 4.6 describes this process. Note that different weighting parameters were used in \(c\) and \(\zeta\). This was done because we might expect the time intervals of real-time
observations and forecasts to be necessarily different. This extra parameter allows flexibility in choosing a time increment for the iterative process of forecasting travel times.

\[
\text{Smoothing:} \\
S_n = \phi X_{n-1} + (1-\phi)\left[\varsigma S_{n-1} + (1-\varsigma)\tau(0)\right]
\]

\[
\text{Forecast:} \\
S_n = c S_{n-1} + (1-c)\tau(0)
\]

where:
\[
\tau(t_n) = 10 \text{ parameter function estimated in Chapter 3}
\]
\[
\{\phi, \varsigma, c\} \sim \text{smoothing parameters} \in \{0,1\}
\]

Equation 4.6: Smoothing and forecasting travel times given a historical estimate and real-time information during off-peak periods.

These two circumstances were combined and implemented in Microsoft Excel in an iterative process. Real-time information could be received from Milwaukee via the internet or simulated by selecting a random day of existing historical data and progressing through that day observation by observation. In either case, data was parsed into 17 columns each corresponding to a beginning and endpoint pair. A timestamp was then associated with this vector noting the time at which it was received. This information was then placed into a “Real-Time Information Sheet” in Microsoft Excel. The worksheet could handle as many recent real-time observations as necessary, so long as they were placed in successive rows and the row following the most recent real-time observation was left blank. A section of this worksheet is included in Figure 4.4. This figure diagrams the process by which smoothed estimates of travel time and forecasts were made.
In another worksheet, parameters were input. \( \theta, \phi, \zeta, \) and \( c \) were included and allowed to range between 0 and 1. An additional parameter was included for the time increment at which forecasts were to be produced. In other words, if the increment were set to 30 seconds, a forecast would be created for every thirty second interval beginning with the interval immediately following the most recent real-time observation.

Historical estimates of travel times were then created on each link for every value of \( t \) that a real-time observation was received as well as each increment past the most recent real-time observation over a relatively short time horizon. These estimates were created based upon the ten-parameter function described in Chapter 3 and the actual parameters estimated for each link of the highway network.

An iterative process was then begun in which smoothing and forecasting were performed by the conditions described above in the following way. A Visual Basic macro written for Excel stepped through a series of increasing times, beginning with the time at which an initial real-time observation was received. This macro is included in Appendix C. Times steps were associated directly with the time of observations, generally occurring about every three minutes, until the most recent real-time observation. Beginning with the first time, if a real-time observation was available, a smoothed estimate was created differing if \( t \) lay within a peak travel time period. If no real-time information was available at \( t \), a forecast was made based upon whether \( t \) lay within a peak period. For the Milwaukee Highway System, a peak period was characterized by a historical estimate of travel time that exceeded 105\% of free flow travel time. This process is described in Figure 4.3.
Process repeated iteratively from first real time observation until some predetermined period after the most recent real-time observation.

\[
\begin{align*}
    \text{if} \ [\text{Real Time Information Received at } t] \\
    \quad \text{if} \ [t \text{ During Peak Hour}] \\
    \quad \quad S_n &= \theta X_{n-1} + (1 - \theta) \left[ \tau(t_n) - \tau(t_{n-1}) + S_{n-1} \right] \\
    \quad \text{else} \\
    \quad \quad S_n &= \phi X_{n-1} + (1 - \phi) \left[ \varsigma S_{n-1} + (1 - \varsigma) \tau(0) \right] \\
    \quad \text{end if} \\
    \text{else} \\
    \quad \text{if} \ [t \text{ During Peak Hour}] \\
    \quad \quad S_n &= \tau(t_n) - \tau(t_{n-1}) + S_{n-1} \\
    \quad \text{else} \\
    \quad \quad S_n &= c S_{n-1} + (1 - c) \tau(0) \\
    \quad \text{end if} \\
\end{align*}
\]

where:
\[
\begin{align*}
    \tau(t_n) &= 10 \text{ parameter function estimated in Chapter 3} \\
    \{\theta, \phi, \varsigma, c\} &\sim \text{smoothing parameters } \in \{0,1\}
\end{align*}
\]

Figure 4.3: Algorithm describing the smoothing and prediction of travel time observations

A final worksheet was added to calculate the amount recent observations were weighted at the moment of making a first forecast. Weights were assigned by the process described in 4.2.1 and differed according to whether or not an observation occurred during a peak period.

Note that while all data was stored in the spreadsheet used for this process, the actual data requirements of this process are fairly small. The only information that needs to be stored is the most recent travel time observation for each link, the most recent
smoothing for each link, a timestamp for each, and parameters for historical travel time functions.

Figure 4.3 describes the process of travel time prediction. Figure 4.4 describes the functional implementation of this process in Microsoft Excel. Arrows note flow of information. Times corresponding to real-time observations are used to calculate historical function values. Historical Function Values, Real-time values, and Smoothing Parameters are used to calculate Smoothed Estimates and Forecasts. The number of real-time observations and smoothing parameters are used to calculate weights on previous real-time observations.
Figure 4.4: Implementation of process of predicting travel times based on a smoothed estimates of real-time observations and a historical estimate of travel times. Arrows denote information flow.

The results of this method were first tested empirically. Simulations were run by stepping through individual days and graphically displaying the results. The forecasts
appeared to predict future travel times with some degree of accuracy. An example of this empirical testing is shown in Figure 4.5. In this figure, the pink line represents the historical estimate of travel time. The yellow points represent actual observations of travel time, and the blue points indicate the smoothed estimate of travel times and that estimate forecasted into the future. Looking at this figure, the blue forecast is able to closely predict actual travel time.

Empirical testing provided reasonably encouraging support in our belief that this method will in fact work to predict travel times. As further evidence, a quantitative measure of performance was sought out. A least absolute deviation measure over some time horizon into the future was constructed and implemented.
The Visual Basic macro used to step through any individual day was altered to collect a running total of the least absolute deviation between the predicted and actual travel times for 6, 15, 30, and 60 minute time horizons after all observations. Smoothing parameters were chosen empirically for this testing. A total least absolute deviation was kept for each link in the Milwaukee Highway System. These totals were then divided by the number of observations to get an average least absolute deviation for each link. As a final step, these totals were expressed as a percentage of free flow travel time so that they may be compared directly and an overall measure of accuracy may be constructed for the entire system.

As a measure for comparison, average least absolute deviations were calculated for each time observation using the historical estimates of travel time as a predictor and using the free-flow travel time as a predictor.

This procedure was performed for each of the twenty weekdays in the month of June 2002 on all links in the Milwaukee Highway System. In general, the real-time prediction method outperformed the historical estimate and free-flow estimate of travel time. Shorter time horizons produced better results. These results are summarized in Figure 4.6 below.

<table>
<thead>
<tr>
<th>Time Horizon of Prediction</th>
<th>Percent of Observations Better Than Historical Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 min</td>
<td>95.9%</td>
</tr>
<tr>
<td>15 min</td>
<td>89.2%</td>
</tr>
<tr>
<td>30 min</td>
<td>78.7%</td>
</tr>
<tr>
<td>60 min</td>
<td>63.2%</td>
</tr>
</tbody>
</table>

Figure 4.6: Performance of real-time prediction method.

Additionally, a system wide measure of accuracy was established for each method. This measure was an average least absolute deviation expressed as a percentage
of free flow travel time. Again, the real-time prediction method outperformed the historical and free flow methods for all time horizons. As expected, shorter time horizons created more accurate predictions. This information is summarized in Figure 4.7.

![Figure 4.7: Average least absolute deviations of prediction methods expressed as a percentage of free flow travel time.](image)

Thus, we can conclude that given the correct smoothing parameters, our method of predicting travel time based on real-time information and a historical estimate will outperform the historical estimate and outperform a constant measure for travel time.

### 4.3 Additional Issues

As in Chapter 3, the issue of non-existent data requires further research and deserves mention here. There are almost certainly going to be cases in any application when real-time information will not be available. This was evident in the Milwaukee
data by reportings of zero when equipment malfunctioned or no observation was recorded. We dealt with this by not including this value in any prediction, but relying on a forecast from the most recent real-time observation.

Oftentimes, when a zero value was received, many other road segments showed observed travel time values. In this case, a better estimate of travel time on the unknown road segment may be available if correlations exist between it and a known value. A simple example might be if we knew that travel times on one road 1 averaged 56% of those on road 2 with some variance. If travel times were reported for road 2, but not for road 1, we could create a confidence interval for travel times on road 1. We might be able to update this confidence interval further with the most recent real-time observations and obtain an expectation of travel time more accurate than one relying only on previous observations. This is obviously a very simple explanatory example. The mathematics behind this concept become much more complicated in cases where multiple road segments are all correlated. Shbaklo, et al. discuss this topic briefly, but do not get into the mathematics or apply the concept to ADVANCE.\textsuperscript{79} In a paper presented this year at the Transportation Research Board’s 82\textsuperscript{nd} Annual Meeting, Gajewski and Rilett develop the mathematics for correlation between travel times in great detail.\textsuperscript{80} In this paper, they develop a method in which natural cubic splines produce an expectation of travel time. They then applied Bayesian statistical methods to model the correlations between road segment travel times performed data analysis on a corridor in Houston, Texas. One interesting result was that correlation of travel times between road segments dropped significantly in congested traffic conditions. This result may indicate that the use of

\textsuperscript{79} Shbaklo.
correlations might not be as helpful as first thought, particularly because we expect high correlations during non-congested hours as most roads will be at free flow traffic volumes. We search for correlation relationships that can help predict congestion. As such, further research into this topic is required to come to any applicable conclusions.

4.4 Using Predicted Travel Times in the Real World

As stated at the conclusion of Chapter 3, the value of predicting travel time is to create a forecast of travel time that will be directly relevant to a decision maker and influence his decisions. A total travel time along a path from any start point to any endpoint is simply the sum of a continuous set of segments that connects the two points (plus time spent at intersections of road segments and any other considerations). Multiple paths generally exist between points of interest and the optimal path is considered to be that which minimizes the travel time of the decision maker. With real-time travel time information, costs assigned to roadways are more accurate and choices can be made that better optimize the value to the decision maker.

The key to forecasting link travel times is that the forecasts need to be for the time period when the user is expected to traverse the given link. Recall that we associated a travel time measurement with the end of an interval. Given this assumption, we may use an approach similar to Equation 3.5 to calculate travel time forecasts. However, Equation 3.5 is specific to a continuous estimation of historical travel time as a function of time of day. Our forecasts, because they are calculated iteratively, are for discrete times, and we must modify our approach as such.
As a solution to this problem, we may either associate each moment we want to predict with one of our discrete forecasts or we may look to somehow predict between the forecasts. An easy way to do this is simply by connecting all discrete points with a line. This line will represent any trend inherent in the discrete forecasts and produce a more accurate prediction of future travel time. The point where this continuous function of joined line segments satisfies Equation 3.5 is the predicted travel time we are looking for. Equation 4.7 solves one instance of Equation 3.5 when adjacent discrete travel time forecasts are connected with a straight line.

\[
\Delta t = \frac{T(t_1) - T(t_2)}{(T(t_1) + t_2) - (T(t_2) + t_1)}
\]

solves
\[
\Delta t = \left(\frac{T(t_2) - T(t_1)}{t_2 - t_1}\right) \Delta t + \left(\frac{T(t_1) - T(t_2) - T(t_1)}{t_2 - t_1}ight)
\]

where:
\(\Delta t\) = forecasted travel time on a link beginning at an start - time
\(T(t)\) = forecast of travel time \(t\) seconds from link start - time
\(t_1\) = closest time observation less than where equation 3.5 is satisfied
\(t_2\) = closest time observation greater than where equation 3.5 is satisfied

**Equation 4.7: Extension of Equation 3.5 to case of line segment between two discrete points**

As an example, we applied this process to the same path examined in Chapter 3, Moorland Road to the Mitchell Interchange. We forecast beginning at 50,000 seconds from midnight on Wednesday July 12, 2002, knowing reported real-time travel time observations to that point. We forecast for the entire route from that point, recording an actual travel time from Moorland Road to the Zoo Interchange, and updating the forecast at that point. We repeat this process upon actual arrival at the Hale interchange as if
simulating a driver actually traversing this route. Figure 4.8 displays the results of these calculations graphically. Actual travel times are recorded in black. Forecasts beginning at Moorland, the Zoo, and the Hale are red, green, and blue respectively. Times of arrival at endpoints are indicated below the points, and predicted and actual travel times are recorded above the link they refer to. As expected, the closer a driver gets to reaching the Mitchell Interchange, the more accurate the predictions of travel time become.
Imagine yourself, in the not so distant future, leaving your home in Rockford, IL for a long weekend getaway at a lake cottage in northern Indiana. You have directions prepared for you on a dashboard display unit, with indicated travel times. You notice that your trip is planned to avoid Chicago, even though you normally drive through the city, because you would arrive there during rush hour on Friday afternoon. As you drive to your destination, following the given instructions, you receive an auditory cue to drive through Chicago instead of around. A map on your in-dash screen changes to reflect this command. The device tells you that your estimated time of arrival has dropped from 7:10 PM to 6:53 PM. From what you know about how the device in your dashboard works,
you understand that it has received information that the traffic situation has changed and that the fastest route to your destination is now through Chicago.

This reality is not as far off as we might think, and the work presented in this thesis has provided some very important building blocks to help us get there. We have developed a methodology for forecasting the travel times that a device like the one described above would use to calculate the fastest route to any destination by combining historical and real-time observations of travel time.

We have developed the concept for, and demonstrated the feasibility of describing travel time as a continuous function of time of day. Our solution, a ten-parameter function consisting of the sum of a constant and three normal curves, was developed through logical reasoning and empirical observation of link travel times from the Milwaukee Highway System. The function was then fit to travel time data for links in Milwaukee by minimizing the total least absolute deviation. Resulting functions showed that two normal curves matched morning and afternoon rush hours, with the third fitting to the noon side of the afternoon rush. The historically based forecast of travel time ending at any point in time is equal to the value of this function evaluated at that time. This approach improves upon existing methods in two ways. First, the function is continuous, allowing expected travel time to be calculated at any instant throughout the day. Most current attempts at solving this problem create travel time estimates for a number of discretized time intervals. Second, because the function is limited to a relatively small number of parameters, the data requirements for producing a predicted travel time are smaller than using discretized methods where at least two parameters are required for each time interval throughout the day.
To produce a more precise forecast of travel time, we developed a method to include a real-time information source with the historical function when predicting travel time. Our method created a smoothed estimate of current travel time based upon recent observations and then forecasted travel times into the future based upon this smoothed estimate and observable traffic patterns. Forecasts increased and decreased with the slope of the historical function during peak travel periods, and decayed back to free flow travel time between peak periods. As expected, the forecast with real-time information was able to predict travel time better than the historical function and the constant measure of free flow travel time, especially for short time horizons. As the time horizon of the forecasts approached one hour, their accuracy dropped to near that of the historical estimate. Both continued to outperform the free flow travel time as a predictive measure at this time interval.

While we believe this model to be a feasible method of forecasting travel times, the work and discussion in this thesis certainly has limitations. As mentioned throughout, further research is needed in cases where there is no available information. Often times, there are no available observations of travel times for a given link, making the creation of a historical estimate impossible. This problem also occurs with real-time information. There were often cases in the Milwaukee Highway System when some links reported travel times and others did not. Null values were not included in forecasts and it was speculated that through further research, relationships could be developed between travel times on different links such that if one road segment did not report a travel time, it could be accurately predicted from reported observations on other road segments.
This leads directly to the question of availability of data. Presently, both historical and real-time travel times are not widely available. While some states collect and distribute traffic information, most states are just beginning to make this information available. And, when traffic information is available, it is not often recorded as a travel time, but as a speed or road volume, necessitating additional computation to arrive at travel time observations. For the use of travel time information to be possible, information needs to become more readily available. Otherwise, we will have a solution only applicable to certain areas or specific roadways.

That being said, we have shown that application is possible. We have described two very simple examples in which travel time was forecast over a path. These examples, however rudimentary, demonstrate the applicability of forecasts using historical and real-time information. With added work and combination of the concepts in this thesis with third generation in-vehicle guidance systems, the use of forecasted link travel times will enable travelers to make better, value maximizing travel decisions.
Bibliography


Beck, William B. “RE: Real-Time Traffic Information (Email Generated by Comments Page).” E-mail to Christopher Schrader. 3 July 2002.

Bond, Russ. “Info. Request.” E-mail to Christopher Schrader. 10 Jul 2002.


Bucci, Joe. “Real Time Traffic Information.” E-mail to Christopher Schrader. 9 July 2002.


<http://westwoodone.com/aboutus_co_profile.asp>.

DeCabooter, Phil. “FW: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 Jul 2002.


Flanagan, Ed E. “Real-Time Traffic.” E-mail to Christopher Schrader. 2 July 2002.


Glossary. Jan 1999. Lake County Transportation Improvement Project.  

Glossary of Terms. ITS Institute, University of Minnesota.  

Gamble, Amy. “RE: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 July 2002.

Harvey, Laurel. “Re: question.” E-mail to Christopher Schrader. 7 Apr 2003.

Healy, Michael. “Real time Traffic information.” E-mail to Christopher Schrader. 2 July 2002.


McCurdy, Bryan. “Nevada Speed Monitoring.” E-mail to Christopher Schrader. 8 Jul 2002.
Mishefske, John. “RE: Traffic Information (Summer Contact).” E-mail to Christopher Schrader. 24 February 2003.

*Motorola History.* 2003. Motorola, Inc. &lt;http://www.motorola.com/content/0,1037,115-110,00.html&gt;.


Parsons, Bob. “Re: Real-Time Traffic Information.” E-mail to Christopher Schrader. 8 Jul 2002.


Thomas, Scott. “RE: Real-Time Traffic Information.” E-mail to Christopher Schrader. 2 Jul 2002.


Waters, Marion. “Real time traffic information for speeds.” E-mail to Christopher Schrader. 2 Jul 2002.
Appendix A:
Description of the Milwaukee Highway System

Data Summary:

Collected From: 17 sets of Start and End Points in the Milwaukee Highway System

Collected: June 2002 and July 6, 2002 at inconsistent intervals averaging about 3 minutes

224,621 Total Observations

Format:

[Time] Date – Time Date, Roadway, Direction, Starting Point - Ending Point, Travel Time


Milwaukee Highway System Information

Travel times were obtained on the pictured road segments approximately every three minutes for the month of June 2002

Segment Endpoints
1. Brown Deer Road & I-43
2. Good Hope Road & US 45
3. Capitol Drive & I-43
4. Burleigh Drive & US 45
5. Moorland Road & I-94
6. Layton Road & I-43
7. Zoo Interchange (US 45, I-94, & I-94E)
9. Hale Interchange (I-43 & I-94E)

Complete Segments
5-3 I-43 NB Downtown - Capitol Dr.
5-1 I-43 NB Downtown - Brown Deer Rd.
3-5 SB Capitol Dr. - Downtown
1-8 I-43 SB Brown Deer Rd. to Downtown
7-9 I-94 EB Zoo - Hale
9-10 I-94 EB Hale - Mitchell
9-7 I-94 WB Hale - Zoo
10-9 I-94 WB Mitchell - Hale
5-8 I-94 EB Moorland - Downtown
5-7 I-94 EB Moorland - Zoo
6-8 I-94 NB Layton - Downtown
5-6 I-94 SB Downtown - Layton
3-1 I-94 WB Downtown - Moorland
8-7 I-94 WB Downtown - Zoo
7-2 US45 NB Zoo - Good Hope Rd.
4-7 US45 SB Burleigh - Zoo
2-7 US45 SD Good Hope - Zoo
Summary Statistics for Link Travel Times in the Milwaukee Highway System

<table>
<thead>
<tr>
<th>Path</th>
<th>7-9</th>
<th>7-2</th>
<th>5-7</th>
<th>5-8</th>
<th>10-9</th>
<th>6-8</th>
<th>9-7</th>
<th>9-10</th>
<th>2-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1st Qu:</td>
<td>311</td>
<td>586</td>
<td>220</td>
<td>606</td>
<td>367</td>
<td>0</td>
<td>275</td>
<td>329</td>
<td>600</td>
</tr>
<tr>
<td>Mean:</td>
<td>299.6718</td>
<td>587.7964</td>
<td>232.8395</td>
<td>638.2641</td>
<td>374.0905</td>
<td>313.3813</td>
<td>278.0318</td>
<td>337.7492</td>
<td>636.0039</td>
</tr>
<tr>
<td>Median:</td>
<td>311</td>
<td>586</td>
<td>220</td>
<td>606</td>
<td>367</td>
<td>391</td>
<td>275</td>
<td>329</td>
<td>600</td>
</tr>
<tr>
<td>3rd Qu:</td>
<td>311</td>
<td>586</td>
<td>223</td>
<td>621</td>
<td>377</td>
<td>391</td>
<td>275</td>
<td>329</td>
<td>600</td>
</tr>
<tr>
<td>Max:</td>
<td>833</td>
<td>1173</td>
<td>1567</td>
<td>2130</td>
<td>1607</td>
<td>2254</td>
<td>1581</td>
<td>1982</td>
<td>2194</td>
</tr>
<tr>
<td>Total N:</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
</tr>
<tr>
<td>Std Dev:</td>
<td>82.40777</td>
<td>23.40633</td>
<td>56.03363</td>
<td>101.6986</td>
<td>76.40421</td>
<td>198.5435</td>
<td>104.1454</td>
<td>85.95502</td>
<td>191.3375</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Path</th>
<th>8-7</th>
<th>8-5</th>
<th>8-6</th>
<th>8-3</th>
<th>8-1</th>
<th>3-8</th>
<th>4-7</th>
<th>1-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>216</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1st Qu:</td>
<td>387</td>
<td>629</td>
<td>393</td>
<td>216</td>
<td>0</td>
<td>0</td>
<td>270</td>
<td>0</td>
</tr>
<tr>
<td>Mean:</td>
<td>404.2969</td>
<td>664.1825</td>
<td>377.4395</td>
<td>240.1492</td>
<td>428.8971</td>
<td>1.853478</td>
<td>279.3293</td>
<td>5.352002</td>
</tr>
<tr>
<td>Median:</td>
<td>388</td>
<td>629</td>
<td>393</td>
<td>216</td>
<td>660</td>
<td>0</td>
<td>270</td>
<td>0</td>
</tr>
<tr>
<td>3rd Qu:</td>
<td>399</td>
<td>632</td>
<td>417</td>
<td>216</td>
<td>660</td>
<td>0</td>
<td>270</td>
<td>0</td>
</tr>
<tr>
<td>Max:</td>
<td>1759</td>
<td>2443</td>
<td>2120</td>
<td>1244</td>
<td>4856</td>
<td>425</td>
<td>1390</td>
<td>914</td>
</tr>
<tr>
<td>Total N:</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
<td>13213</td>
</tr>
<tr>
<td>Std Dev:</td>
<td>137.7085</td>
<td>130.6001</td>
<td>201.6633</td>
<td>68.71451</td>
<td>346.0225</td>
<td>22.02327</td>
<td>130.8834</td>
<td>61.19272</td>
</tr>
</tbody>
</table>

Frequency of Observations is constant throughout the day. This is true of all links as data was received in packets with one observation for every link.
Plot of Original Data – Time of Day v. Travel Time for 16 of 17 links

S-Plus code for input of Milwaukee Highway Data

"MKE.TS" <- matrix(c(6,0,0:00:18,6/1/2002,311,586,224,629,374,391,275,329,600,387,629,2120,216,0,0,270,0,6,0,0:03:26,6/1/2002,311,586,220,606,373,391,275,329,600,387,629,1362,216,0,0,270,0,6,0,0:06:34,6/1/2002,314,586,220,606,374,391,275,329,600,387,629,1186,216,0,0,270,0,6,0,0:09:41,6/1/2002,313,586,220,606,367,391,275,329,600,387,629,1804,216,0,0,270,0,6,0,0:12:48,6/1/2002,311,586,220,606,379,391,275,337,600,387,629,1853,216,0,0,270,0,6,0,0:15:55,6/1/2002,318,586,220,606,324,500,391,275,336,600,387,629,1356,216,0,0,270,0,
6,0,23:50:53,7/6/2002,311,586,220,606,367,391,275,329,600,0,629,393,216,60,0,270,0,6,0,23:53:53,7/6/2002,311,586,220,606,367,391,275,329,600,0,629,393,216,660,0,270,0,6,0,23:57:10,7/6/2002,311,586,220,606,367,391,275,336,600,0,629,393,216,660,0,271,0), nrow = 13213, ncol = 21, dimnames = list(character(0),c("Day_of_Week","Weekday","Time","Date","Zoo_Hal","Zoo_GHp","Moo_Zoo","Moo_Dtw","Mit_Hal","Lay_Dtw","Hal_Zoo","Hal_Mit","GHp_Zoo","Dtw_Zoo","Dtw_Moo","Dtw_Lay","Dtw_Cap","Dtw_BrD","Cap_Dtw","Bur_Zoo","BrD_Dtw")))
Appendix B:
Development of Historical Travel Time Function: Milwaukee Highway System

Cleaning Data – Removing Null Values and Time Elements

S-Plus code for Removing Null Values from Milwaukee Highway Data

```r
for (i in 1:length(MKE[,1])) {
  if (MKE$Zoo.Hal[i] != 0) Zoo.Hal <- rbind(Zoo.Hal, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Zoo.Hal[i]))
  if (MKE$Zoo.GHp[i] != 0) Zoo.GHp <- rbind(Zoo.GHp, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Zoo.GHp[i]))
  if (MKE$Zoo.Zoo[i] != 0) Zoo.Zoo <- rbind(Zoo.Zoo, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Zoo.Zoo[i]))
  if (MKE$Zoo.Dtw[i] != 0) Zoo.Dtw <- rbind(Zoo.Dtw, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Zoo.Dtw[i]))
  if (MKE$Zoo.Mit[i] != 0) Zoo.Mit <- rbind(Zoo.Mit, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Zoo.Mit[i]))
  if (MKE$Zoo.Dtw[i] != 0) Dtw.Dtw <- rbind(Dtw.Dtw, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.Dtw[i]))
  if (MKE$Moo.Hal[i] != 0) Moo.Hal <- rbind(Moo.Hal, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Moo.Hal[i]))
  if (MKE$Moo.GHp[i] != 0) Moo.GHp <- rbind(Moo.GHp, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Moo.GHp[i]))
  if (MKE$Moo.Zoo[i] != 0) Moo.Zoo <- rbind(Moo.Zoo, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Moo.Zoo[i]))
  if (MKE$Moo.Dtw[i] != 0) Moo.Dtw <- rbind(Moo.Dtw, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Moo.Dtw[i]))
  if (MKE$Moo.Mit[i] != 0) Moo.Mit <- rbind(Moo.Mit, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Moo.Mit[i]))
  if (MKE$Moo.Dtw[i] != 0) Dtw.Moo <- rbind(Dtw.Moo, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.Moo[i]))
  if (MKE$Moo.Lay[i] != 0) Dtw.Lay <- rbind(Dtw.Lay, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.Lay[i]))
  if (MKE$Moo.Cap[i] != 0) Dtw.Cap <- rbind(Dtw.Cap, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.Cap[i]))
  if (MKE$Moo.BrD[i] != 0) Dtw.BrD <- rbind(Dtw.BrD, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.BrD[i]))
  if (MKE$BrD.Dtw[i] != 0) Dtw.BrD <- rbind(Dtw.BrD, cbind(MKE$Day[i], MKE$Weekend[i], MKE$Secs[i], MKE$Dtw.BrD[i]))
}
```
for(i in 1:length(MKE[,1]))
{ if(MKE$Cap[i] != 0) Dtw.Cap <-
  rbind(Dtw.Cap,cbind(MKE$Day[i],MKE$Weekend[i],MKE$Secs[i],MKE$Dtw.Cap[i]))
}
for(i in 1:length(MKE[,1]))
{ if(MKE$BrD[i] != 0) Dtw.BrD <-
  rbind(Dtw.BrD,cbind(MKE$Day[i],MKE$Weekend[i],MKE$Secs[i],MKE$Dtw.BrD[i]))
}
for(i in 1:length(MKE[,1]))
{ if(MKE$Cap.Dtw[i] != 0) Cap.Dtw <-
  rbind(Cap.Dtw,cbind(MKE$Day[i],MKE$Weekend[i],MKE$Secs[i],MKE$Cap.Dtw[i]))
}
for(i in 1:length(MKE[,1]))
{ if(MKE$BrD.Dtw[i] != 0) Bur.Zoo <-
  rbind(Bur.Zoo,cbind(MKE$Day[i],MKE$Weekend[i],MKE$Secs[i],MKE$BrD.Zoo[i]))
}

S-Plus Code for Separating Weekdays from Weekends within the Milwaukee Highway Data

```
Zoo.Hal.Wk <- NULL
Zoo.GHp.Wk <- NULL
Moo.Zoo.Wk <- NULL
Moo.Dtw.Wk <- NULL
Mit.Hal.Wk <- NULL
Lay.Dtw.Wk <- NULL
Hal.Zoo.Wk <- NULL
Hal.Mit.Wk <- NULL
GHP.Zoo.Wk <- NULL
Dtw.Zoo.Wk <- NULL
Dtw.Moo.Wk <- NULL
Dtw.Lay.Wk <- NULL
Dtw.Cap.Wk <- NULL
Dtw.BrD.Wk <- NULL
Cap.Dtw.Wk <- NULL
Bur.Zoo.Wk <- NULL
BrD.Dtw.Wk <- NULL

Zoo.Hal.End <- NULL
Zoo.GHp.End <- NULL
Moo.Zoo.End <- NULL
Moo.Dtw.End <- NULL
Mit.Hal.End <- NULL
Lay.Dtw.End <- NULL
Hal.Zoo.End <- NULL
Hal.Mit.End <- NULL
GHP.Zoo.End <- NULL
Dtw.Zoo.End <- NULL
Dtw.Moo.End <- NULL
Dtw.Lay.End <- NULL
Dtw.Cap.End <- NULL
Dtw.BrD.End <- NULL
Cap.Dtw.End <- NULL
Bur.Zoo.End <- NULL
BrD.Dtw.End <- NULL

for(i in 1:length(Zoo.Hal[,1]))
}
for(i in 1:length(Zoo.GHp[,1]))
{ if(Zoo.GHp[i,2] != 1) Zoo.GHp.End <- rbind(Zoo.GHp.End,Zoo.GHp[i,])
  else Zoo.GHp.Wk <- rbind(Zoo.GHp.Wk,Zoo.GHp[i,])
}
```
for (i in 1:length(Moo.Zoo[,1]))
  else Moo.Zoo.Wk <- rbind(Moo.Zoo.Wk, Moo.Zoo[i,])
for (i in 1:length(Moo.Dtw[,1]))
  else Moo.Dtw.Wk <- rbind(Moo.Dtw.Wk, Moo.Dtw[i,])
for (i in 1:length(Mit.Hal[,1]))
  else Mit.Hal.Wk <- rbind(Mit.Hal.Wk, Mit.Hal[i,])
for (i in 1:length(Lay.Dtw[,1]))
for (i in 1:length(Hal.Zoo[,1]))
  else Hal.Zoo.Wk <- rbind(Hal.Zoo.Wk, Hal.Zoo[i,])
for (i in 1:length(Hal.Mit[,1]))
  if (Hal.Mit[i,2] != 1) Hal.Mit.End <- rbind(Hal.Mit.End, Hal.Mit[i,])
  else Hal.Mit.Wk <- rbind(Hal.Mit.Wk, Hal.Mit[i,])
for (i in 1:length(GHp.Zoo[,1]))
for (i in 1:length(Dtw.Zoo[,1]))
  else Dtw.Zoo.Wk <- rbind(Dtw.Zoo.Wk, Dtw.Zoo[i,])
for (i in 1:length(Dtw.Moo[,1]))
  if (Dtw.Moo[i,2] != 1) Dtw.Moo.End <- rbind(Dtw.Moo.End, Dtw.Moo[i,])
  else Dtw.Moo.Wk <- rbind(Dtw.Moo.Wk, Dtw.Moo[i,])
for (i in 1:length(Dtw.Lay[,1]))
  else Dtw.Lay.Wk <- rbind(Dtw.Lay.Wk, Dtw.Lay[i,])
for (i in 1:length(Dtw.Cap[,1]))
  else Dtw.Cap.Wk <- rbind(Dtw.Cap.Wk, Dtw.Cap[i,])
for (i in 1:length(Dtw.BrD[,1]))
  else Dtw.BrD.Wk <- rbind(Dtw.BrD.Wk, Dtw.BrD[i,])
for (i in 1:length(Cap.Dtw[,1]))
for (i in 1:length(Bur.Zoo[,1]))
for (i in 1:length(BrD.Dtw[,1]))
  if (BrD.Dtw[i,2] != 1) BrD.Dtw.End <- rbind(BrD.Dtw.End, BrD.Dtw[i,])
  else BrD.Dtw.Wk <- rbind(BrD.Dtw.Wk, BrD.Dtw[i,])
Creating Smoothed Kernel Estimates of Travel Time as Function of Time of Day

Choosing Bandwidth for Kernel Estimators

Example with Hale Interchange to Zoo Interchange

MKEker1 <- ksmooth(Hal.Zoo[,3],Hal.Zoo[,4],kernel="normal",bandwidth = 100)
MKEker2 <- ksmooth(Hal.Zoo[,3],Hal.Zoo[,4],kernel="normal",bandwidth = 1000)
MKEker3 <- ksmooth(Hal.Zoo[,3],Hal.Zoo[,4],kernel="normal",bandwidth = 5000)
MKEker4 <- ksmooth(Hal.Zoo[,3],Hal.Zoo[,4],kernel="normal",bandwidth = 3000)
plot(Hal.Zoo[,3],Hal.Zoo[,4])
lines(MKEker1,col=2)
lines(MKEker2,col=3)
lines(MKEker3,col=4)
lines(MKEker4,col=5)
Differences Between Kernels for Weekdays and All Combined Data
Example with Moorland Road to Downtown Interchange

plot(Moo.Dtw.Wk[,3],Moo.Dtw.Wk[,4],xlab="seconds from midnight")
lines(Moo.Dtw.Wk.ker,col=5)
lines(Moo.Dtw.ker,col=6)
S-Plus Code for Running Kernels
For Weekdays on All Links in Milwaukee Highway System

```r
### KERNELS ###
Zoo.GHp.Wk.ker <- ksmooth(Zoo.GHp.Wk[3],Zoo.GHp.Wk[4],kernel="normal",bandwidth = 3000)
Moo.Zoo.Wk.ker <- ksmooth(Moo.Zoo.Wk[3],Moo.Zoo.Wk[4],kernel="normal",bandwidth = 3000)
Moo.Dtw.Wk.ker <- ksmooth(Moo.Dtw.Wk[3],Moo.Dtw.Wk[4],kernel="normal",bandwidth = 3000)
Mit.Hal.Wk.ker <- ksmooth(Mit.Hal.Wk[3],Mit.Hal.Wk[4],kernel="normal",bandwidth = 3000)
Hal.Mit.Wk.ker <- ksmooth(Hal.Mit.Wk[3],Hal.Mit.Wk[4],kernel="normal",bandwidth = 3000)
GHP.Zoo.Wk.ker <- ksmooth(GHP.Zoo.Wk[3],GHP.Zoo.Wk[4],kernel="normal",bandwidth = 3000)
Dtw.Moo.Wk.ker <- ksmooth(Dtw.Moo.Wk[3],Dtw.Moo.Wk[4],kernel="normal",bandwidth = 3000)
Dtw.Lay.Wk.ker <- ksmooth(Dtw.Lay.Wk[3],Dtw.Lay.Wk[4],kernel="normal",bandwidth = 3000)
```

Fitting the 10-Parameter Function
Differences Between Fitting to Kernel and Data
Zoo Interchange – Hale Interchange
Plots of Curves Fit To Data
Travel Time v. Time of Day

Zoo – Hale

Zoo – Good Hope

Moorland – Zoo

Moorland - Downtown

Downtown – Zoo

Good Hope – Zoo
Estimated Parameters for 10-Parameter Function fit to Links in Milwaukee Highway System

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dtw.Cap</td>
<td>216</td>
</tr>
<tr>
<td>Dtw.Brd</td>
<td>666</td>
</tr>
<tr>
<td>Cap.Dtw</td>
<td>311</td>
</tr>
<tr>
<td>Brd.Dtw</td>
<td>329</td>
</tr>
<tr>
<td>Zoo.Hal</td>
<td>275</td>
</tr>
<tr>
<td>Hal.Mit</td>
<td>367</td>
</tr>
<tr>
<td>Hal.Zoo</td>
<td>606</td>
</tr>
<tr>
<td>Mit.Hal</td>
<td>220</td>
</tr>
<tr>
<td>Moo.Dtw</td>
<td>391</td>
</tr>
<tr>
<td>Moo.Zoo</td>
<td>393</td>
</tr>
<tr>
<td>Lay.Dtw</td>
<td>629</td>
</tr>
<tr>
<td>Dtw.Lay</td>
<td>387</td>
</tr>
<tr>
<td>Dtw.Moo</td>
<td>586</td>
</tr>
<tr>
<td>Dtw.Zoo</td>
<td>270</td>
</tr>
<tr>
<td>Zoo.GHp</td>
<td>600</td>
</tr>
<tr>
<td>Bur.Zoo</td>
<td>600</td>
</tr>
<tr>
<td>GHp.Zoo</td>
<td>600</td>
</tr>
</tbody>
</table>

=PARAMETERS!H$5+PARAMETERS!H$9/(SQRT(2*PI())*PARAMETERS!H$11)*EXP(-(SB9-PARAMETERS!H$10)/2*PARAMETERS!H$11))+PARAMETERS!H15/(SQRT(2*PI())*PARAMETERS!H$14)*EXP(-(SB9-PARAMETERS!H$13)/2*PARAMETERS!H$14))

=PARAMETERS!H$6/(SQRT(2*PI())*PARAMETERS!H$8)*EXP(-(SB9-PARAMETERS!H$7)/2*PARAMETERS!H$8))

=PARAMETERS!H$9/(SQRT(2*PI())*PARAMETERS!H$11)*EXP(-(SB9-PARAMETERS!H$10)/2*PARAMETERS!H$11))

=PARAMETERS!H15/(SQRT(2*PI())*PARAMETERS!H$14)*EXP(-(SB9-PARAMETERS!H$13)/2*PARAMETERS!H$14))
Appendix C:
Implementation & Results of Real-Time Application: Milwaukee Highway System

Empirical Results

Code for Microsoft Excel Macro to Simulate Real-Time Information

Starts at random point on a random day, and progressively add data points to real-time data through the end of the day.

```vba
Sub copypastemacro()
    ' copypastemacro Macro
    ' Macro recorded 3/19/2003 by Koray D. Simsek
    ' Macro adapted by Chris Schrader

    Sheets("OUTPUT").Select
    Range("C8:T1208").Select
    Range("T1208").Activate
    Selection.ClearContents

    Dim i As Integer
    Dim j As Integer

    j = 0

    Do
        Randomize
        i = Int((8750 * Rnd) + 7) ' Generate random value between 7 and 8757.
        Loop Until Range("E7").Offset(i, 0) < 86400

    Sheets("DATA").Select

    Do While Range("E7").Offset(i, 0) > Range("E7").Offset(i - 1, 0)
        Range("E7:V7").Offset(i, 0).Select
        Selection.Copy
        Sheets("OUTPUT").Select
        Range("C8").Offset(j, 0).Select
        Selection.PasteSpecial Paste:=xlPasteValues
        i = i + 1
        j = j + 1

    Sheets("DATA").Select
    Loop

    End Sub
```
Progression Through Sample Day:
Moorland – Downtown       June 14, 2002

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>THETA</td>
<td>PHE</td>
<td>CAI</td>
<td>C</td>
</tr>
<tr>
<td>0.3000</td>
<td>0.8000</td>
<td>0.5000</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Time Step: 0:03:00
Quantitative Measure of Accuracy

Code for Microsoft Excel Macro to Evaluate Real-Time Information: Adapted from Simulation Code

Starts at beginning of a given day, and progressively add data points to real-time data through the end of the day, calculating Least Absolute Deviations for 6, 15, 30, and 60 minute time horizons.

Code shortened to include process for only one link, Capitol Drive – Downtown

Sub copypastemacro()
  ' copypastemacro Macro
  ' Macro recorded 3/19/2003 by Koray D. Simsek
  ' Macro adapted by Chris Schrader
Sheets("OUTPUT").Select
Range("C8:T1208").Select
Range("T1208").Activate
Selection.ClearContents

Dim i As Integer, j As Integer
Dim k As Integer, c As Integer, kk As Integer, cc As Integer
Dim LAD_60 As Double
Dim LAD_6 As Double
Dim LAD_15 As Double
Dim LAD_30 As Double

' CAP-DTW
Dim pred_6_out As Double, tt_6_out As Double, hist_6 As Double
Dim pred_60_out As Double, tt_60_out As Double, hist_60 As Double
Dim pred_15_out As Double, tt_15_out As Double, hist_15 As Double
Dim pred_30_out As Double, tt_30_out As Double, hist_30 As Double
k = 0
     c = 0
     kk = 0
     cc = 0
LAD_60 = 0
LAD_6 = 0
LAD_15 = 0
LAD_hist_6 = 0
LAD_hist_15 = 0
LAD_hist_60 = 0

j = 0
i = 2470 ' set to beginning of a day
Sheets("DATA").Select
Do While Range("E6").Offset(i, 0) > Range("E6").Offset(i - 1, 0)

    Range("E7:V7").Offset(i, 0).Select
    Selection.Copy
    Sheets("OUTPUT").Select
    Range("C8").Offset(j, 0).Select
    Selection.PasteSpecial Paste:=xlPasteValues

' CAP - DTW -----------------------------------------------
Sheets("OUTPUT").Select
    pred_6_out = Range("BP8").Offset(j + 2, 0).Value ' sets variable to pred travel time 6 mins out
    pred_60_out = Range("BP8").Offset(j + 20, 0).Value
    pred_15_out = Range("BP8").Offset(j + 5, 0).Value
    pred_30_out = Range("BP8").Offset(j + 10, 0).Value
Sheets("DATA").Select
    tt_6_out = Range("F7").Offset(i + 2, 0).Value ' sets variable to actual travel time 6 mins out
    tt_60_out = Range("F7").Offset(i + 20, 0).Value
    tt_15_out = Range("F7").Offset(i + 5, 0).Value
    tt_30_out = Range("F7").Offset(i + 10, 0).Value

If Range("E7").Offset(i + 2, 0).Value < Range("E7").Offset(i + 3, 0).Value And Range("E7").Offset(i + 2, 1).Value > 0 Then
    LAD_6 = LAD_6 + Abs(tt_6_out - pred_6_out)
    k = k + 1
End If

If Range("E7").Offset(i + 20, 0).Value < Range("E7").Offset(i + 21, 0).Value And Range("E7").Offset(i + 20, 1).Value > 0 Then
    LAD_60 = LAD_60 + Abs(tt_60_out - pred_60_out)
    c = c + 1
End If

If Range("E7").Offset(i + 5, 0).Value < Range("E7").Offset(i + 6, 0).Value And Range("E7").Offset(i + 5, 1).Value > 0 Then
    LAD_15 = LAD_15 + Abs(tt_15_out - pred_15_out)
    kk = kk + 1
End If

If Range("E7").Offset(i + 10, 0).Value < Range("E7").Offset(i + 11, 0).Value And Range("E7").Offset(i + 10, 1).Value > 0 Then
    LAD_30 = LAD_30 + Abs(tt_30_out - pred_30_out)
    cc = cc + 1
End If

' ------------------------------
Sheets("Sheet1").Select
' CAP - DTW
    Range("C7") = LAD_6 / k
    Range("C8") = LAD_15 / kk
    Range("C9") = LAD_30 / cc
    Range("C10") = LAD_60 / c

End Sub
Average Least Absolute Deviation expressed as a percentage of free flow travel time for road segments (Top) and the entire system (Bottom) over the entire month of June 2002.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Free Flow</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.2%</td>
<td>6.9%</td>
<td>5.3%</td>
<td>3.7%</td>
<td>5.7%</td>
<td>11.2%</td>
<td>5.6%</td>
<td>8.1%</td>
<td>9.3%</td>
<td>9.5%</td>
<td>18.3%</td>
<td>8.6%</td>
<td>13.4%</td>
<td>0.5%</td>
<td>19.2%</td>
<td>13.2%</td>
</tr>
<tr>
<td><strong>Historical Information</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.5%</td>
<td>7.2%</td>
<td>5.6%</td>
<td>3.5%</td>
<td>7.6%</td>
<td>11.4%</td>
<td>6.7%</td>
<td>6.2%</td>
<td>8.2%</td>
<td>10.3%</td>
<td>12.2%</td>
<td>5.8%</td>
<td>8.6%</td>
<td>0.9%</td>
<td>11.2%</td>
<td>8.5%</td>
</tr>
<tr>
<td><strong>Historical and Real Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>7.1%</td>
<td>6.7%</td>
<td>4.4%</td>
<td>3.5%</td>
<td>7.2%</td>
<td>8.9%</td>
<td>5.7%</td>
<td>5.7%</td>
<td>8.0%</td>
<td>11.6%</td>
<td>12.3%</td>
<td>5.4%</td>
<td>8.3%</td>
<td>0.9%</td>
<td>10.4%</td>
<td>7.7%</td>
</tr>
<tr>
<td>30 min</td>
<td>6.3%</td>
<td>6.2%</td>
<td>4.1%</td>
<td>3.2%</td>
<td>5.7%</td>
<td>7.4%</td>
<td>5.1%</td>
<td>5.1%</td>
<td>7.5%</td>
<td>9.9%</td>
<td>11.3%</td>
<td>4.7%</td>
<td>7.5%</td>
<td>0.8%</td>
<td>10.0%</td>
<td>6.8%</td>
</tr>
<tr>
<td>15 min</td>
<td>5.4%</td>
<td>5.1%</td>
<td>3.6%</td>
<td>2.8%</td>
<td>4.5%</td>
<td>6.2%</td>
<td>4.5%</td>
<td>4.2%</td>
<td>6.2%</td>
<td>8.0%</td>
<td>9.8%</td>
<td>4.1%</td>
<td>6.6%</td>
<td>0.7%</td>
<td>9.1%</td>
<td>5.9%</td>
</tr>
<tr>
<td>6 min</td>
<td>4.3%</td>
<td>3.7%</td>
<td>3.1%</td>
<td>2.4%</td>
<td>3.5%</td>
<td>5.1%</td>
<td>4.0%</td>
<td>3.3%</td>
<td>4.8%</td>
<td>6.0%</td>
<td>8.1%</td>
<td>3.3%</td>
<td>5.5%</td>
<td>0.5%</td>
<td>7.7%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Free Flow</th>
<th>Historical Information</th>
<th>Historical and Real Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.2%</td>
<td>7.5%</td>
<td>60 min</td>
</tr>
<tr>
<td></td>
<td>6.3%</td>
<td>6.0%</td>
<td>30 min</td>
</tr>
<tr>
<td></td>
<td>5.4%</td>
<td>5.4%</td>
<td>15 min</td>
</tr>
<tr>
<td></td>
<td>4.3%</td>
<td>4.3%</td>
<td>6 min</td>
</tr>
</tbody>
</table>
Daily Average Least Absolute Deviations for road segments in the Milwaukee Highway System expressed in seconds.

<table>
<thead>
<tr>
<th>THETA</th>
<th>0.3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHE</td>
<td>0.8000</td>
</tr>
<tr>
<td>CAI</td>
<td>0.5000</td>
</tr>
<tr>
<td>C</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Time Step</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0:03:00</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6/3/2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>49.83</td>
<td>10.40</td>
<td>10.79</td>
<td>17.74</td>
<td>18.34</td>
<td>22.13</td>
<td>34.83</td>
<td>14.54</td>
<td>49.84</td>
<td>34.08</td>
<td>22.56</td>
<td>26.00</td>
<td>5.79</td>
<td>23.41</td>
</tr>
<tr>
<td></td>
<td>30 min</td>
<td>65.25</td>
<td>11.65</td>
<td>11.18</td>
<td>24.99</td>
<td>24.15</td>
<td>25.58</td>
<td>44.34</td>
<td>18.61</td>
<td>61.75</td>
<td>40.55</td>
<td>30.00</td>
<td>32.78</td>
<td>7.15</td>
<td>26.36</td>
</tr>
<tr>
<td></td>
<td>60 min</td>
<td>70.79</td>
<td>12.57</td>
<td>11.87</td>
<td>41.09</td>
<td>38.59</td>
<td>31.90</td>
<td>62.96</td>
<td>24.33</td>
<td>86.85</td>
<td>43.69</td>
<td>45.48</td>
<td>43.56</td>
<td>8.31</td>
<td>28.70</td>
</tr>
<tr>
<td>6/4/2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-Time</td>
<td>6 min</td>
<td>0.47</td>
<td>10.52</td>
<td>10.01</td>
<td>78.41</td>
<td>10.05</td>
<td>29.58</td>
<td>27.22</td>
<td>11.51</td>
<td>6.40</td>
<td>60.58</td>
<td>37.18</td>
<td>33.13</td>
<td>8.84</td>
<td>17.72</td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>0.71</td>
<td>11.54</td>
<td>10.67</td>
<td>101.79</td>
<td>12.57</td>
<td>35.77</td>
<td>35.99</td>
<td>14.63</td>
<td>8.18</td>
<td>73.35</td>
<td>45.64</td>
<td>38.56</td>
<td>10.63</td>
<td>22.00</td>
</tr>
<tr>
<td></td>
<td>30 min</td>
<td>0.93</td>
<td>12.63</td>
<td>10.00</td>
<td>125.12</td>
<td>15.60</td>
<td>47.42</td>
<td>45.80</td>
<td>17.05</td>
<td>10.83</td>
<td>87.07</td>
<td>51.67</td>
<td>40.35</td>
<td>10.85</td>
<td>26.51</td>
</tr>
<tr>
<td></td>
<td>60 min</td>
<td>1.07</td>
<td>13.83</td>
<td>9.35</td>
<td>165.24</td>
<td>19.76</td>
<td>57.76</td>
<td>53.01</td>
<td>20.03</td>
<td>12.08</td>
<td>105.08</td>
<td>60.22</td>
<td>46.81</td>
<td>9.99</td>
<td>32.02</td>
</tr>
<tr>
<td>6/5/2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-Time</td>
<td>6 min</td>
<td>9.49</td>
<td>19.19</td>
<td>3.32</td>
<td>5.25</td>
<td>5.67</td>
<td>8.38</td>
<td>11.88</td>
<td>5.53</td>
<td>11.59</td>
<td>52.85</td>
<td>11.64</td>
<td>13.53</td>
<td>1.56</td>
<td>12.20</td>
</tr>
<tr>
<td></td>
<td>30 min</td>
<td>11.93</td>
<td>14.03</td>
<td>4.15</td>
<td>8.33</td>
<td>8.91</td>
<td>10.29</td>
<td>19.77</td>
<td>10.51</td>
<td>17.67</td>
<td>86.16</td>
<td>16.19</td>
<td>17.72</td>
<td>2.80</td>
<td>15.53</td>
</tr>
<tr>
<td>6/6/2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>6.73</td>
<td>5.24</td>
<td>9.74</td>
<td>11.90</td>
<td>12.05</td>
<td>17.88</td>
<td>11.23</td>
<td>64.11</td>
<td>16.94</td>
<td>15.71</td>
<td>18.75</td>
<td>1.78</td>
<td>42.30</td>
<td>42.91</td>
</tr>
<tr>
<td></td>
<td>30 min</td>
<td>7.25</td>
<td>6.04</td>
<td>11.11</td>
<td>12.38</td>
<td>13.71</td>
<td>18.58</td>
<td>12.62</td>
<td>69.61</td>
<td>18.55</td>
<td>14.62</td>
<td>17.67</td>
<td>2.41</td>
<td>45.75</td>
<td>45.01</td>
</tr>
<tr>
<td></td>
<td>60 min</td>
<td>6.94</td>
<td>6.27</td>
<td>12.29</td>
<td>12.51</td>
<td>12.96</td>
<td>17.12</td>
<td>13.37</td>
<td>81.82</td>
<td>18.70</td>
<td>16.72</td>
<td>16.89</td>
<td>2.81</td>
<td>37.29</td>
<td>51.24</td>
</tr>
<tr>
<td>6/7/2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-Time</td>
<td>6 min</td>
<td>1.93</td>
<td>9.56</td>
<td>4.59</td>
<td>20.12</td>
<td>16.54</td>
<td>13.20</td>
<td>8.21</td>
<td>44.37</td>
<td>26.60</td>
<td>16.64</td>
<td>18.05</td>
<td>1.13</td>
<td>21.36</td>
<td>27.84</td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>1.92</td>
<td>10.82</td>
<td>5.47</td>
<td>23.42</td>
<td>18.42</td>
<td>15.87</td>
<td>10.23</td>
<td>61.27</td>
<td>26.77</td>
<td>17.25</td>
<td>20.26</td>
<td>1.90</td>
<td>22.34</td>
<td>30.27</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Real-Time</td>
<td></td>
<td>Real-Time</td>
<td></td>
<td>Real-Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.03</td>
<td>10.85</td>
<td>10.63</td>
<td>6.06</td>
<td>15.96</td>
<td>14.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>2.01</td>
<td>10.05</td>
<td>10.15</td>
<td>5.20</td>
<td>16.05</td>
<td>13.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>12.59</td>
<td>11.56</td>
<td>9.65</td>
<td>21.43</td>
<td>14.17</td>
<td>27.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>14.88</td>
<td>14.25</td>
<td>11.27</td>
<td>26.78</td>
<td>15.33</td>
<td>32.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>14.92</td>
<td>15.27</td>
<td>12.57</td>
<td>30.97</td>
<td>14.53</td>
<td>37.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.97</td>
<td>2.69</td>
<td>4.56</td>
<td>8.18</td>
<td>9.05</td>
<td>14.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 min</td>
<td>5.77</td>
<td>3.74</td>
<td>4.75</td>
<td>10.61</td>
<td>15.16</td>
<td>11.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>6.97</td>
<td>5.07</td>
<td>4.97</td>
<td>14.13</td>
<td>16.34</td>
<td>17.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>7.34</td>
<td>6.64</td>
<td>5.65</td>
<td>17.55</td>
<td>18.77</td>
<td>21.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>8.45</td>
<td>16.90</td>
<td>7.08</td>
<td>32.71</td>
<td>11.66</td>
<td>14.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.02</td>
<td>11.44</td>
<td>5.05</td>
<td>14.79</td>
<td>10.74</td>
<td>15.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 min</td>
<td>6.92</td>
<td>13.82</td>
<td>5.96</td>
<td>20.85</td>
<td>9.64</td>
<td>12.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>8.04</td>
<td>15.92</td>
<td>7.11</td>
<td>26.60</td>
<td>11.78</td>
<td>13.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>8.45</td>
<td>16.90</td>
<td>7.08</td>
<td>32.71</td>
<td>11.66</td>
<td>14.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>18.15</td>
<td>34.81</td>
<td>7.50</td>
<td>8.64</td>
<td>25.37</td>
<td>12.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.85</td>
<td>23.52</td>
<td>6.23</td>
<td>3.83</td>
<td>13.73</td>
<td>9.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 min</td>
<td>13.48</td>
<td>27.32</td>
<td>6.64</td>
<td>4.76</td>
<td>16.92</td>
<td>9.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>16.38</td>
<td>31.97</td>
<td>7.04</td>
<td>5.96</td>
<td>21.68</td>
<td>10.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>18.15</td>
<td>34.81</td>
<td>7.50</td>
<td>8.64</td>
<td>25.37</td>
<td>12.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>20.59</td>
<td>37.17</td>
<td>27.62</td>
<td>15.35</td>
<td>75.46</td>
<td>59.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.73</td>
<td>22.32</td>
<td>13.15</td>
<td>7.55</td>
<td>40.44</td>
<td>37.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 min</td>
<td>15.53</td>
<td>26.78</td>
<td>16.57</td>
<td>9.60</td>
<td>49.76</td>
<td>44.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>18.96</td>
<td>32.39</td>
<td>21.46</td>
<td>12.16</td>
<td>58.58</td>
<td>50.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>20.59</td>
<td>37.17</td>
<td>27.62</td>
<td>15.35</td>
<td>75.46</td>
<td>59.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>4.12</td>
<td>16.68</td>
<td>5.99</td>
<td>8.49</td>
<td>11.04</td>
<td>9.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>5.45</td>
<td>19.39</td>
<td>6.86</td>
<td>11.41</td>
<td>13.61</td>
<td>10.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

119
<table>
<thead>
<tr>
<th></th>
<th>30 min</th>
<th>60 min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.01</td>
<td>28.21</td>
</tr>
<tr>
<td></td>
<td>15.40</td>
<td>29.52</td>
</tr>
<tr>
<td>Real-Time</td>
<td>15.68</td>
<td>60.36</td>
</tr>
<tr>
<td></td>
<td>22.00</td>
<td>72.19</td>
</tr>
<tr>
<td></td>
<td>27.57</td>
<td>83.95</td>
</tr>
<tr>
<td></td>
<td>30.33</td>
<td>86.66</td>
</tr>
<tr>
<td>Real-Time</td>
<td>12.44</td>
<td>114.20</td>
</tr>
<tr>
<td></td>
<td>15.92</td>
<td>144.53</td>
</tr>
<tr>
<td></td>
<td>19.23</td>
<td>174.77</td>
</tr>
<tr>
<td></td>
<td>21.66</td>
<td>193.13</td>
</tr>
<tr>
<td>Real-Time</td>
<td>4.74</td>
<td>35.59</td>
</tr>
<tr>
<td></td>
<td>4.99</td>
<td>34.38</td>
</tr>
<tr>
<td></td>
<td>5.28</td>
<td>32.88</td>
</tr>
<tr>
<td></td>
<td>4.39</td>
<td>32.52</td>
</tr>
<tr>
<td></td>
<td>13.99</td>
<td>19.06</td>
</tr>
<tr>
<td></td>
<td>19.66</td>
<td>24.67</td>
</tr>
<tr>
<td></td>
<td>30 min</td>
<td>60 min</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>6/26/2002 Dtw.Cap</td>
<td>17.70</td>
<td>20.10</td>
</tr>
<tr>
<td>Dtw.Brd</td>
<td>24.37</td>
<td>27.63</td>
</tr>
<tr>
<td>Zoo.Hal</td>
<td>7.27</td>
<td>7.49</td>
</tr>
<tr>
<td>Hal.Mit</td>
<td>13.53</td>
<td>14.82</td>
</tr>
<tr>
<td>Hal.Zoo</td>
<td>12.20</td>
<td>13.48</td>
</tr>
<tr>
<td>Mit.Hal</td>
<td>14.59</td>
<td>14.01</td>
</tr>
<tr>
<td>Moo.Dtw</td>
<td>23.03</td>
<td>26.37</td>
</tr>
<tr>
<td>Moo.Zoo</td>
<td>11.07</td>
<td>13.93</td>
</tr>
<tr>
<td>Lay.Dtw</td>
<td>43.88</td>
<td>50.51</td>
</tr>
<tr>
<td>Dtw.Lay</td>
<td>37.20</td>
<td>35.71</td>
</tr>
<tr>
<td>Dtw.Moo</td>
<td>21.10</td>
<td>24.29</td>
</tr>
<tr>
<td>Dtw.Zoo</td>
<td>20.83</td>
<td>21.74</td>
</tr>
<tr>
<td>Zoo.GHp</td>
<td>2.52</td>
<td>2.90</td>
</tr>
<tr>
<td>GHp.Zoo</td>
<td>15.29</td>
<td>18.31</td>
</tr>
<tr>
<td>6/27/2002 Dtw.Cap</td>
<td>5.07</td>
<td>0.78</td>
</tr>
<tr>
<td>Dtw.Brd</td>
<td>9.06</td>
<td>10.15</td>
</tr>
<tr>
<td>Zoo.Hal</td>
<td>10.79</td>
<td>10.15</td>
</tr>
<tr>
<td>Hal.Mit</td>
<td>15.75</td>
<td>16.15</td>
</tr>
<tr>
<td>Hal.Zoo</td>
<td>17.55</td>
<td>17.15</td>
</tr>
<tr>
<td>Mit.Hal</td>
<td>27.65</td>
<td>27.25</td>
</tr>
<tr>
<td>Moo.Dtw</td>
<td>12.39</td>
<td>12.95</td>
</tr>
<tr>
<td>Moo.Zoo</td>
<td>22.13</td>
<td>22.85</td>
</tr>
<tr>
<td>Lay.Dtw</td>
<td>67.76</td>
<td>71.63</td>
</tr>
<tr>
<td>Dtw.Lay</td>
<td>18.72</td>
<td>20.86</td>
</tr>
<tr>
<td>Dtw.Moo</td>
<td>2.82</td>
<td>2.59</td>
</tr>
<tr>
<td>Dtw.Zoo</td>
<td>48.73</td>
<td>52.12</td>
</tr>
<tr>
<td>Zoo.GHp</td>
<td>24.43</td>
<td>35.17</td>
</tr>
<tr>
<td>Bur.Zoo</td>
<td>61.63</td>
<td>71.63</td>
</tr>
<tr>
<td>GHp.Zoo</td>
<td>48.73</td>
<td>52.12</td>
</tr>
<tr>
<td>6/28/2002 Dtw.Cap</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>Dtw.Brd</td>
<td>7.88</td>
<td>11.70</td>
</tr>
<tr>
<td>Zoo.Hal</td>
<td>25.32</td>
<td>14.32</td>
</tr>
<tr>
<td>Hal.Mit</td>
<td>15.11</td>
<td>9.80</td>
</tr>
<tr>
<td>Hal.Zoo</td>
<td>15.47</td>
<td>12.75</td>
</tr>
<tr>
<td>Mit.Hal</td>
<td>9.34</td>
<td>1.67</td>
</tr>
<tr>
<td>Moo.Dtw</td>
<td>10.19</td>
<td>35.29</td>
</tr>
<tr>
<td>Moo.Zoo</td>
<td>9.46</td>
<td>15.28</td>
</tr>
<tr>
<td>Lay.Dtw</td>
<td>29.89</td>
<td>18.66</td>
</tr>
<tr>
<td>Dtw.Lay</td>
<td>1.10</td>
<td>1.38</td>
</tr>
<tr>
<td>Dtw.Moo</td>
<td>8.68</td>
<td>23.25</td>
</tr>
<tr>
<td>Dtw.Zoo</td>
<td>9.85</td>
<td>30.22</td>
</tr>
<tr>
<td>Zoo.GHp</td>
<td>22.53</td>
<td>18.72</td>
</tr>
<tr>
<td>Bur.Zoo</td>
<td>41.68</td>
<td>65.23</td>
</tr>
<tr>
<td>GHp.Zoo</td>
<td>52.72</td>
<td>71.63</td>
</tr>
</tbody>
</table>
Appendix D:
Princeton University Graduate Shuttle Web Application

Throughout the writing of this thesis, I have contributed a large portion of work to a project that illustrates the relationship of real-time and historical information in travel time forecasting: web display of the Princeton University graduate student shuttle. This project began in the fall of 2002 with conceptual development of the need for a shuttle bus to serve graduate students. The Princeton Borough planning board required Princeton University to make a “good faith” attempt at a shuttle bus in order to allow construction of more graduate student housing. The motivation was two fold: get graduate students to campus more frequently and alleviate the crowded parking situation caused by graduate student cars.

As the project became a reality, a need to accurately display shuttle information to graduate students was apparent. Without appropriate information about when the shuttles (two shuttles were to be deployed) would be at a given graduate student stop, they would be difficult to catch, and graduate students would not find them an acceptable alternative to get to campus. In addition to me, Katy Milkman, Ryan Goldenberg, and John Cranston took on the task of providing useful information via the internet to graduate

---

81 Harvey, Laurel. “Re: question.” E-mail to Christopher Schrader. 7 Apr 2003.
students on Princeton’s Campus. Our suggestions, design, and application were submitted as an end of term project in ORF 467 in January of 2003.\footnote{Schrader, Christopher, Katy Milkman, John Cranston, and Ryan Goldenberg. “Princeton Shuttle Finder.” Term paper submitted January 12, 2003.}

After the end of the fall term, implementation became the task of a private company, ALK, and students working in the Transportation Information and Decision Engineering (TIDE) lab at Princeton University. Currently, Ashirul Amin and I are the main contributors from the TIDE lab, and Jordan Rapp and Charles Perkins are the main contributors from ALK. Alain Kornhauser has overseen the project on both ends. The final product, still under development, displays a combination of real-time information and historical information in the hope that it will influence the graduate student decision-maker to take the shuttle, rather than use a car or stay at home. The project can currently be found on the internet at \url{http://205.246.138.50}.\footnote{\textit{P-Rides Transit}. April 2003. ALK. NJ TIDE. Princeton University. <\url{http://205.246.138.50}>.}

In this appendix section, both the historical and real-time aspects of this project are introduced, the difficulties of providing real-time information are addressed, and implications on this thesis are discussed.

**Concept**

The concept for this application is simple: provide the current location of shuttles and estimated times of arrival (ETA) at all stops so that graduate students may make informed decisions about when to use the shuttle. Shuttles calculate location by GPS technology, send that location to a server where information is retained and necessary information is calculated. The essential function of the server is equivalent to that of the Traffic Information Center and Traffic Related Functions of the ADVANCE project
mentioned in Chapter 2. The server receives data, makes travel time calculations to stops, and sends data to users via the World Wide Web. Figure D.1 displays the relevant components of the Graduate Student Shuttle Web Application. Shuttles receive signals from GPS satellites. They then send their location to a server, which displays that location along with calculated ETAs on the internet.

Figure D.1: Concept of Graduate Student Shuttle Website.

Including Historical Information

Historical information can be used to provide ETAs directly to users in the form of a schedule. A Princeton University committee of graduate students and administration originally designed the shuttle schedule in accordance with the perceived needs of graduate students, basing intervals between stops on estimated travel times. Shuttle drivers, hired by A1 Limousine, were to stick to the schedule through original implementation. If they arrived at a stop early, they were to wait until the scheduled

---

84 ADVANCE Project Formal Evaluation.
85 Schrader, et al. 2.
departure time to leave. If they arrived late, time was to be made up at the next stop to which they arrived early. In this way, a consistent, historical pattern of travel time was created. Historical information was displayed on the website via a tabular schedule. Figure D.2 shows a screenshot of the display of the schedule.

Historical information was also used in calculation of ETAs. Given no other information, the ETA to any shuttle stop should be (Current Time) – (Next Scheduled Arrival Time). If real-time information is received, an ETA to the next stop may be available. We may then be able to assume that the travel time from the next stop to the stop after that is the historical travel time between those stops, or some function of that time. This technique is currently being implemented and will be further discussed below.

Figure D.2: Screenshot of Schedule view of Graduate Student Shuttle Website.
Including Real-Time Information

Devices were installed in each of two shuttles in order to collect and transmit real-time shuttle locations. Included in each device was an iPAQ 3670 personal digital assistant, an iPAQ dual PCMCIA expansion sleeve, a CoPilot CF GPS receiver, a Sierra Wireless Aircard 750, and a Pocket PC 5 Volt / 2 Amp DC Power Adapter. The devices receive signals from GPS satellites to determine their location. Once location is determined, it is sent to the server via a cellular communication and placed in a text file `busloc.txt`. Each shuttle updates this file every ten seconds throughout connection to the server. The `busloc.txt` file contains an ID identifying each shuttle, a location (latitude and longitude), a speed, a heading, and a timestamp.

<table>
<thead>
<tr>
<th>ID</th>
<th>LAT</th>
<th>LON</th>
<th>SPEED</th>
<th>HEADING</th>
<th>TIMESTAMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shuttle1</td>
<td>40329743</td>
<td>-74657372</td>
<td>2556</td>
<td>172</td>
<td>9:06 AM</td>
</tr>
<tr>
<td>Shuttle2</td>
<td>40350165</td>
<td>-74651712</td>
<td>1609</td>
<td>222</td>
<td>9:07 AM</td>
</tr>
</tbody>
</table>

From the information contained in `busloc.txt`, shuttle icons are displayed over a map on the webpage. The shuttle icons are refreshed every five seconds, to give the most accurate information to the user.

In addition to visual information, ETAs should be calculated based upon the real-time information contained in `busloc.txt` and displayed on the website. Schrader, et al. describe an algorithm with which to make the calculation of ETA for all future stops.\(^\text{86}\)

This algorithm is as follows: Given a shuttle location, speed, and heading, determine whether the shuttle is at a stop, or between two stops, and determine those stops. If the shuttle is at a stop, the ETA can be drawn directly from the historical schedule. If the shuttle is between two stops, the ETA to the next stop is calculated to be the distance to

\(^\text{86}\) Schrader.
the next stop as a percentage of total distance between the two stops multiplied by the historical time between the two stops. In both cases, the travel time to subsequent stops is taken to be the historical travel time.

Schrader, et al. implement this algorithm in a format using grids to divide up different locations on the map of the shuttle route.\textsuperscript{87} In this approach, percent distance to any next stop is a direct function of which grid the shuttle is located in and shuttle heading. This information is displayed on the internet both in a table of ETAs and mouse-over boxes that appear whenever a user places his cursor over a stop. Figure D.4 shows the tabular ETA view of the website, and Figure D.6 shows the map view of the website with ETA mouse-over.

**Problems with Implementation**

This technique worked well for demonstrative purposes in presentation of the paper, but failed in the real world. Shuttles began running on February 3, 2003 and throughout the first two months of operation, real-time information was received with little consistency. Rarely did both shuttles report their location to the server, and oftentimes neither shuttle sent information. This makes the algorithm described by Schrader, et al. incomplete as it does not account for the possibility of no information. In Chapter 3 and Chapter 4 of this thesis, major recommendations were made for research into dealing with cases in which no information was available. Here, a scheme to provide some measure of estimated time of arrival, whether or not real time information is available, is needed.

\textsuperscript{87} Schrader.
As a temporary solution, calculation of ETAs was changed to draw directly from the schedule. Based on the current time, the next arrival time for each bus at each stop is looked up in a data file. Scheduled arrival times are displayed in the mouse-over boxes and the table of ETAs. As drivers are instructed to follow the schedule as best as possible, this solution is fairly accurate. Real-time information is displayed graphically through the shuttle icons, and a feature allowing the user to determine the quality of the information has been added: By scrolling-over any shuttle, the user will see a pop-up box with the time stamp of the most recent observation of the shuttle’s location. If either shuttle’s location has not been updated for a certain amount of time, that shuttle’s icon is placed in a “garage” just above the map legend.

![Princeton University Shuttle Expected Arrival Times](image)

**Figure D.4:** ETA table view of website.
Figure D.5: Map view of website containing ETA mouse-over.

Figure D.6: Map view of website containing last update mouse-over.
Implications for this Thesis

The implications of the graduate student shuttle web application on this thesis are two fold: it shows the combination of real-time and historical information in practice, and it shows the implementation and difficulties of a possible method for receiving real-time travel times.

In this project, historical information sets a base expectation of how long it will take to travel between each stop, and ETAs are calculated from this. In much the same way, the historical function developed in Chapter 3 can create an estimated travel time along the path it represents. Additionally, the combination of real-time and historical information is used in calculating ETAs and displaying information to the user. This information helps the user decide when to go out and catch the bus. Similarly, the travel time procedure developed in Chapter 4 uses a combination of historical and real-time travel time observations to forecast future travel times. Indeed, the graduate student shuttle website is a simple application of the concepts developed throughout this thesis.

Also important is the implication this project has on the collection and distribution of real-time information. Throughout this thesis, we have highlighted the importance of further research into how travel times should be predicted when information is not available. The inability of the shuttles to consistently communicate with the server highlights this fact. There are certainly going to be situations when, for whatever reason, real-time observations are not received. Information systems need to strive to receive consistent observations, but they also must be prepared for the contingency in which no information is available.
Next Steps

Reliability of communication with the shuttles needs to be improved before further steps become relevant. Problems currently lie in driver’s misuse of the devices, wireless communication, and hardware malfunction.

Once communication with the shuttles is acceptable, a new algorithm to calculate ETAs must be implemented. This algorithm needs to be capable of handling cases when information is received as well as situations when one or both shuttles have ceased communication with the server. Possibilities including a grid approach and map-matching of shuttles to roadways have been discussed and need to be investigated further. Without implementation of ETA calculation based on real-time observation, the user of this website is forced to extrapolate an expected arrival time from information provided graphically. In an ideal situation, the website will be able to do this for the user.

Finally, with reception and logging of observations from shuttles, there is the opportunity for the updating of historical information. A larger number of observations can lead to more exact expected travel times and ETAs. With this information, forecasts will become more accurate, both with and without real-time observations, and users will be able to make better decisions.