

TRAFFIC CONGESTION: HOW PREDICTABLE?
Discovering Volume Trends across Time and
Confirming Fundamental Speed-Flow-Density Relations

Megan Lynn Bernard
Princeton University

May 10, 2005
Professor Alain L. Kornhauser

ABSTRACT:

Though Americans increasingly seek to escape the big cities and enjoy the benefits of suburban life, the major employment bases remain in cities. Because of this fact, millions of Americans experience the daily inconvenience of traffic congestion. From approximately 6:30 am to 10 am, traffic volumes on major roads nearly quadruple as commuters head into work. Though this daily increase is dreadfully predictable, other traffic patterns are entirely less reliable. The purpose of this independent work is to discover trends across weeks, months, and seasons using data from interstates in metropolitan Atlanta, Georgia. Also, this independent work tests that the data provided confirms the volume-speed and density-speed relationships for uninterrupted traffic flow.

TRAFFIC CONGESTION: HOW PREDICTABLE?
Discovering Volume Trends across Time and
Confirming Fundamental Speed-Flow-Density Relations

Table of Contents:

1 Introduction	6
2 Traffic Reporting and Data Collection	11
2.1 Traffic Reporting	12
2.1.1 The Beginning	12
2.1.2 The Traffic Reporting Industry Today	13
2.2 Atlanta Data Collection	15
2.2.1 The Setup of the Atlanta Roadway System	15
2.2.2 Method of Data Collection	17
2.2.3 Good Data, Bad Data, Missing Data	20
3 Volume Trends	23
3.1 Possible Explanatory Variables	24
3.1.1 All Variable Choices	24
3.1.2 The Time Component	25
3.2 Modeling Methods	27
3.3 Atlanta Interstate System	30
4 Speed-Volume-Density Relations	38
4.1 Characteristics of Speed	39
4.2 Fundamental Relationships regarding Speed	43
4.3 Atlanta Interstate System	49
5 Conclusion	54
Bibliography	58

Appendix A: Week-by-Week Volume Data for August	60
Appendix B: Speed-Flow Graphs with Models	62
Appendix C: Speed-Density Graphs with Models	64

1 Introduction

Billions of Americans face the inconvenience of the daily commute to work, making traffic congestion the most widely experienced social problem in the United States. On Monday thru Friday, approximately 252 days a year, commuters driving to and from work experience inevitable delay as huge volumes of drivers navigate the roads in a relatively small time frame. Building in extra time to get to work has become just another regrettable morning routine for most commuters.

The widely varying demand for roadway means finding solutions to congestion is a continual challenge. Because of largely standardized working hours, there is a sharply

peaked demand at times associated with the trip to and from work. For about four hours a day, between 7:30 am and 9:30 am and again between 4:30 pm and 6:30 pm, traffic congestion causes physical and mental stress on commuters. However, this level of traffic demand drops drastically during other parts of the day, meaning that providing efficient yet affordable public transportation is extremely difficult. Also planners must try to balance the demand for expanded infrastructures to manage those few hours a day of heavy traffic with the wasted space on roadways for most other parts of the day.

The prospect of a slow ride to work is all too common for commuters in Atlanta, Georgia. As a sprawling city, the actual population within the city limits is relatively small compared to the population of the wider area of suburbs that contribute to the Atlanta workforce. With only 416,474 people living in the city and a much larger population of 2,604,348 living in and around the city (in Clayton, Cobb, DeKalb, Douglas, Fayette, and Fulton counties), Atlanta is a prime example of urban sprawl.¹ Americans have increasingly sought to escape the city for the greener suburbs; unfortunately this decision by so many working Americans costs a fortune not only in commuting costs, but more importantly in wasted time. In 1999 alone, in 68 urban areas, traffic congestion caused 6.8 billion gallons of wasted fuels, 4.5 billion hours of delay, and 78 billion dollars in total cost.²

For Georgia, a state that had a population growth of 26.4%³ between 1990 and 2000, traffic congestion can only get worse. Expanding the infrastructure of the roadways is evermore a project for the Georgia Department of Transportation (GDOT);

¹ "Georgia Quick Facts from the U.S. Census Bureau." 1 Feb. 2005. U.S. Census Bureau. <<http://quickfacts.census.gov/qfd/states/13000.html>>. (20 March 2005).

² Dr. Peter T. Martin, Yugi Feng, Xiaodong Wang. "Detector Technology Evaluation." Nov. 2003. <http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/pdf/MPC03-154.pdf>. (22 March 2005), 1.

³ "Georgia Quick Facts from the U.S. Census Bureau."

however these new surface roads, too, will eventually reach capacity at peak times of the day. There is an even smaller glimmer of hope for delayed commuters using Atlanta's interstates, unless the Department of Transportation decides to double-decker the interstates, an utter long-shot possibility if even considered. Congestion seems to be a problem that is here to stay, at least for the immediate future.

Since there is no way to prevent congestion, is there a way to predict or possibly beat this congestion? Though there are obvious peak travel times such as the morning rush hour, other less obvious trends surface in the interstate data. Across weeks, months, and seasons, trends in traffic volume emerge, some predictable and some more subtle.

Because the problem of traffic congestion at peak hours is so utterly predictable, this independent work will take the volume spikes for rush hour traffic in the morning and in the evening as the norm and make comparisons based on this assumption. Though this assumption of increases in volume is predictable, the effect that this increase in volume has on speed and travel times is less easily determined.

Obviously as the volume on the road approaches capacity, speeds slow; however many components influence the point at which a road approaches that capacity limit. The Highway Capacity Manual defines capacity as "the maximum hourly rate at which persons on vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions."⁴ Roadway conditions refer to the geometric characteristics of the roadway such as the number of lanes and grade of the road. In contrast, traffic conditions refer to the characteristics of the traffic stream such as the type of vehicle traveling along the road

⁴ Highway Capacity Manual (Special Report 209). (Washington D.C.: Library of Congress, 1985) 1-3.

and the distribution of the vehicles among the lanes. Control conditions refer to the regulatory devices used along the roadway. These roadway, traffic, and control conditions vary between locations however remain static for single locations.

Because this work uses the same fixed points for the entire analysis of Atlanta interstates and make comparisons only between similar elements, these components of capacity have little effect. However, variable components such as weather and traffic accidents do influence volume measures. For the purposes of this analysis, it is assumed that these elements only produce small perturbations in the analysis of the data.

Also, since this analysis uses volume measurements only from highway data, the data considered is only uninterrupted flow data. Uninterrupted flow facilities do not have any fixed elements, such as stop signs, outside the traffic stream which can cause traffic interruptions.⁵ Because of this lack of interruption, there is no time limitation on the use of the roadway space, meaning that the roadway can operate at capacity for indefinite periods of the time without external influence. The implication of this uninterrupted flow trait for the analysis means that, given constant conditions, capacity for the roadways will always be the same.

Using the interstate system has allowed for many external variables that would otherwise bias the data to be controlled. Hence, it is with some certainty that this work can model patterns in the data without serious doubts as to the validity of such analysis. Using this information, this independent work focuses on just such a task.

Throughout the next four chapters, this work attempts to discover trends in traffic volumes across weeks, months, and seasons and calculate how these volumes affect the speed, and ultimately the travel times, for uninterrupted traffic flow on metro-Atlanta

⁵ Highway Capacity Manual (Special Report 209). 1-2.

interstates. Finally, this analysis can be expanded to predict other cities with similar characteristics that might follow the models created for Atlanta.

Chapter 2 will discuss the set-up of the interstate system in Atlanta, its methods for monitoring traffic, and the difficulties discovered when using the data provided by the Georgia Department of Transportation. Chapter 3 begins modeling the traffic flow volumes at various locations along the Atlanta Highway System, thus discovering trends beyond the obvious volume spikes due to rush hour traffic. The volume measures and trends from Chapter 3 are used in Chapter 4 when discussing the implications that volume and density have on the speed of cars traveling on these roads. The fundamental relation between volume and speed is developed, and a discussion of its implications on travel times for not only commuters but all metro-Atlanta drivers follows. Chapter 5 acts as a conclusion to this independent work, recapping what was done, suggesting the application of these trends to other cities, and finally noting the limitations of this work.

2 Traffic Reporting and Data Collection

With the advent of programs like Yahoo Maps, MapBlast, MapQuest, and more, travelers are able to find the statically-weighted shortest route from a Point A origin address to a Point B destination address. However, this static data does not help travelers avoid trouble times during the day or avoid traffic accidents. Without traffic reporting and forecasting, travelers would only be able to plan a route then hope that there is minimal traffic congestion on their chosen path. Both the public and private sector have combined in efforts to gather traffic information that drivers might use in navigating trips. This chapter describes one such traffic monitoring system, the system employed for the

metro-Atlanta interstates. Section 2.1 describes the history of the traffic reporting sector. Section 2.2 looks at how the Georgia Department of Transportation collects data for its roads. And finally Section 2.3 discusses the difficulties present in this system of collection.

2.1 Traffic Reporting

2.1.1 The Beginning

Traffic congestion is a serious social problem and has been for some time. Carole Sauve writes that “[d]uring 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.”⁶ Though the problem then was the horse-and-cart, the problem now is the automobile. With the advent of the automobile in 1885 by Gottfried Daimler and Karl Benz in Germany, society started down the road to becoming what is now a car-dependent culture interconnected by innumerable roads and highways.⁷

The blame for congestion is cyclic; “the car gave us the suburb and the suburb gave us the car.”⁸ Urban sprawl is largely blamed for the problem of traffic congestion. The suburb was created with innocent intentions along rail and trolley lines, but when the electric rail was eliminated, the freedom provided by the automobile attracted attention as a means of escaping the city. Suburban life attracted people out of the city, and the automobile provided a means for travel to jobs, food, and recreation. Now, the American

⁶ Carole Sauve. “Exploring the Auto Culture.” 15 April 1996. <<http://www.flora.org/afo/autocult.html>>. (22 March 2005).

⁷ Sauve.

⁸ Sauve.

society has become so dependent on personal automobiles for transportation that it seems traffic congestion is just a necessary evil.

When radios were first introduced into vehicles, they provided entertainment, news, and weather information. Though the first traffic conditions report is hard to pinpoint, the first documented report occurred in San Francisco in 1957. A private pilot for KSFO-AM radio 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.”⁹ Listeners responded enthusiastically to this reporting and requested more traffic reports. Now there are radio stations devoted solely to providing traffic information for drivers in its region.

2.1.2 The Traffic Reporting Industry Today

Traffic reporting has grown significantly since that first mild observation of the Bay Bridge. Now information is collected by aircraft, cell phone users who report accidents or conditions, police and highway patrol radio frequencies, video detection cameras along the roadways, and sensors built into the pavement on highways. From the 1950’s to the 1990’s, traffic reports were only available by radio or television.

However, with the now widespread use of the Internet and cell phone, new options are available for both gathering and proliferating traffic information. These technologies have advantages over the broadcast technique of radio and television. Cell phones and internet sites are available at the convenience of the user. Because traffic information is not the sole purpose of a radio or television broadcast, the traffic segment of a broadcast is usually constrained to a short segment at a predictable interval in the

⁹ Christopher Schrader. “Reacting in Real Time: Using Historical and Real-Time Information in Forecasting Link Travel Times.” Senior Thesis submitted April 14, 2003. 17.

show, maybe every 15 minutes or so. In these cases, the traveler must wait for the appropriate broadcasts. Also, these broadcasts cover a large area, so they might only hit the traffic information “highlights” and not necessarily the information a specific traveler needs. Internet, cell phone, and other types of non-broadcast traffic information have advantages in that they are generally more specific and always available.

Internet sites provide a variety of information for travelers. Site information generally includes both text and maps. The maps most often display locations of incidents and construction zones. Also, major roads often have data associated specifically with each segment. Examples of provided information include speeds, volume counts, and travel time. Unfortunately there is no standard and each site generally has a different way of conveying data. Some common traffic websites include metrocommute.com, smartraveler.com, etaktraffic.com, accuttraffic.com, traffic.com, traffic411.com, trafficcast.com, trafficonline.com, as well as sites specific to certain urban areas. A drawback of internet site information is that it cannot readily be accessed while a traveler is en route.

Cellular phone services are much more convenient to travelers already on the roads. Travelers can usually dial into a local traffic provider, enter a certain amount of information, and hear traffic conditions for roads they are interested in. This sort of information is useful not just for checking upcoming congestion but also for checking the severity of a traffic situation once experiencing a delay.

In addition to private companies that provide traffic information, the government has taken an increasingly active role in reporting conditions on its roadways. While studies were first conducted manually, noting traffic flow and travel time information,

advances in technology have led to automatically collected information. Certain government agencies have begun to share this information with the public via websites and roadside message boards. In his thesis, Christopher Schrader collects information regarding the development of such technology for various state Departments of Transportation. He notes four distinct stages in the development of public information at the state level: no data collected; data collected but not shared; data collected in real-time and developing ability to share information; and data collected in real-time and shared in real-time with public.¹⁰ Georgia is among the few states collecting data and making that data readily available to the public in a website.

2.2 Atlanta Data Collection

2.2.1 The Setup of the Atlanta Roadway System

The Georgia Department of Transportation attempts to efficiently connect travelers to their destinations using a combination of interstates, county roads, city streets, and state highways. In 2003, the Office of Transportation Data reported that there are 114,862 miles of public roads in Georgia.¹¹ Only 1,244¹² miles are interstates, making only 1.08% of the roadways in Georgia interstates. Not surprisingly, country roads make up a considerable amount of the mileage, approximately seventy-two percent. In contrast, by daily vehicle miles traveled, the interstates have a much more significant

¹⁰ Schrader 21.

¹¹ Vicki Gavalas, ed. "Fact Book 2004." Georgia Department of Transportation. <<http://www.dot.state.ga.us/dot/communications/publications/documents/pdf/factbook/factbook-2004>>. (20 March 2005), 19.

¹² Gavalas 19.

role. Of the 340,276,904¹³ miles traveled daily, approximately 25% are along interstate routes. Clearly the interstates have a vital role in facilitating the movement of travelers on a daily basis.

There are four interstates in and around Atlanta that serve travelers in the area as well as travelers passing through the state. These interstates include Interstate 20 which runs east to west through downtown, Interstate 85 which runs southwest to northeast, Interstate 75 which runs southeast to northwest, and Interstate 285 which encircles the city. Two interstates, I-85 and I-75 merge into one wider interstate for approximately 8 miles through downtown Atlanta. A sample of the internet information provided for these roadways is shown in Figure 2.1.

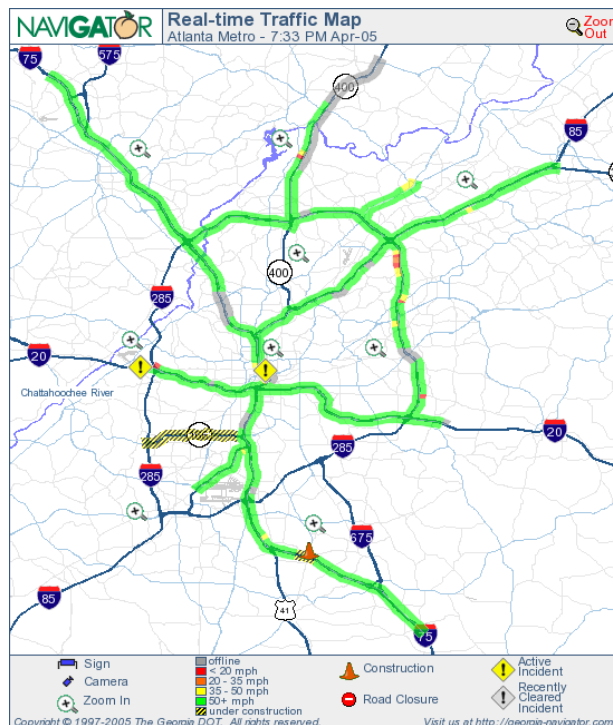


Figure 2.1: Example of information provided by GDOT for Atlanta interstates¹⁴

¹³ Gavalas 19.

¹⁴ “Georgia Quick Facts from the U.S. Census Bureau.”

As previously stated, the GDOT maintains a website which provides real-time traffic information for metro-Atlanta. This website includes information regarding speeds, construction, road closures, accidents, and slow spots.

As far as traffic characteristics, the population density on the north side of Atlanta is much higher than on the south; therefore traffic congestion seems to be most common on Interstate 20 and those portions of the other interstates that lie to the north. Also, Georgia Highway 400 is a toll road through Alpharetta, Roswell, and Buckhead, Georgia, that is known for nearly stand-still traffic at certain times each day. Though there are a significant number of arterial streets that might seem to be alternatives to interstate travel, the population in and around Atlanta has simply grown to the point where congestion is a given on nearly all roads.

2.2.2 Method of Data Collection

In an effort to minimize congestion of the freeway and arterial roadways and to improve traveler safety in the metro-Atlanta area, Georgia developed a traveler information system called NaviGator. NaviGator, Georgia's Intelligent Transportation System (ITS), is a joint effort of the Georgia Department of Transportation (GDOT), the Federal Highway Administration (FHWA), the Metropolitan Atlanta Rapid Transit Authority (MARTA), and the Atlanta Regional Commission.¹⁵ Using telecommunications, video monitoring and detection systems, Geographic Information Systems (GIS) and data management technologies¹⁶, the NaviGator system seeks to provide real-time information about transportation options.

¹⁵ Gavalas 32.

¹⁶ Gavalas 32.

NaviGator uses a video monitoring and detection system to identify congestion, road incidents, and road conditions. The GDOT uses two brands of detectors: Autoscope and Traficon. They mount black and white cameras approximately 80 feet in the air to observe the roadways. These cameras have no pan, tilt, or zoom as they remain fixed to a pole along the side of the interstate. The video signal is used as input for the detection. A typical Traficon installation includes a number of Video Image Processor (VIP) boards integrated into a standard 19" rack together with 1 communication board. Because this method of data collection is non-intrusive, it can be installed and maintained without diverting traffic.¹⁷

There are three types of video processing systems on the market: tripline, closed-loop tracking, and data association tracking. The Atlanta detectors use the tripline system which measures changes in pixels caused by a vehicle relative to the empty road.¹⁸ Upon installation, the processor "learns" what empty road looks like by recognizing when the image changes and thus a vehicle is passing. Also, as part of the installation on the processing end, detection zones are established in the video image.¹⁹ For Atlanta, these companies created zones using the painted white lines which are always 10 feet long with 30 feet in between sets. When a vehicle enters one of these zones, the volume count is increased by one. Also, as the vehicle passes through the zone, the video processor uses the information regarding the vehicle type and amount of time in the fixed zone to calculate the speed. The data collected from these cameras can be recorded in 10, 20, or 30 second periods or also in longer time intervals including 1, 10, 15, 30, or 60 minute

¹⁷ Mark Demidovich. "RE: More info." Email to Megan Bernard. (21 March 2005).

¹⁸ Martin 40.

¹⁹ Traficon Technologies. "Traficon-Principles."
<<http://www.traficon.com/solutions/technology/principles.jsp>>. (25 March 2005).

periods.²⁰ These video processing units can collect several types of data; however, the GDOT uses them only to collect data regarding speed, volume/number of vehicles according to vehicle type, and occupancy. Occupancy refers to the average length of time that the detection zone was occupied and is derived from speed and count.²¹

There are several advantages to using video image processing to collect data. This technology allows for a variety of data material to be collected. With only one camera, video detection can monitor multiple lanes and multiple zones concurrently.²² And finally, because this method of data collection is non-intrusive, detection devices can be added to the system or modified with relative ease.

Currently the GDOT uses approximately 1,300 detectors on mainline, ramp, and arterial roadways²³. Data is aggregated by lane and by roadway location, not only by roadway location. For the purposes of real-time data collection and reporting like the website image shown in Figure 2.1, data is aggregated and sent to the processors every 20 seconds. This data can then be used to update travel situations as conditions change. An example a transmission sent is shown in Figure 2.2.

DetectorID	Speed	Counts (by type)	Occupancy
285001	62.4	13 auto, 3 light truck, 1 long truck	6.5%

Figure 2.2 Twenty Second Data Transmission²⁴

One of the ways GDOT uses this data in real-time is by updating its changeable message signs. These billboard-like signs above the interstates display three types of

²⁰ Peek Traffic. "VideoTrak – 905." <<http://www.its-traffic.com/videotrak905.htm>>. (25 March 2005).

²¹ Demidovich.

²² Martin 108.

²³ Brad Mann. "RE: Data Streams for Metro-Atlanta Interstates." Email to Megan Bernard. (31 Jan. 2005).

²⁴ Demidovich.

messages: travel times messages, incident messages, and abduction messages. The congestion messages relay information about the speed of traffic and the volume of traffic on a specific highway. Incident messages notify travelers of accidents, stalls, or construction on any of the interstates. Finally, child abduction messages are displayed if a child has been abducted in or near Georgia. There are 97 such display boards located along all four interstates and Georgia Highway 400²⁵.

In addition to the real-time data aspect, the information collected by these detectors is archived for historical data analysis. The data is aggregated by lane in 15 minute intervals to be stored. Figure 2.3 shows archived data entries. Historical data is available by year for use in analysis. This independent work uses historical data from the year 2003 for analysis.

```

287|04/07/2003 00:00|96|61.70|3.28
288|04/07/2003 00:00|133|59.04|4.72
289|04/07/2003 00:00|59|52.22|2.21
290|04/07/2003 00:00|37|48.54|1.23
301|04/07/2003 00:00|59|51.14|3.42
302|04/07/2003 00:00|145|57.99|8.08
303|04/07/2003 00:00|123|61.13|5.61
304|04/07/2003 00:00|52|58.96|2.70
347|04/07/2003 00:00|37|66.35|1.99
348|04/07/2003 00:00|115|61.30|6.44

```

Figure 2.3 Archived data as received from GDOT²⁶

2.2.3 Good Data, Bad Data, Missing Data

Supposedly GDOT has 1300 detectors on its Atlanta interstates. Though there are no GPS locations for these detectors, they do supposedly exist throughout Atlanta, and they provide data to their collection facilities at regular intervals. However, not every detector works all the time, and for this analysis, it was imperative to work with detectors

²⁵ Gavalas 32.

²⁶ Mann.

that worked regularly throughout the year in order to discover trends. This independent work uses data from the leftmost lane of four such detectors scattered throughout Atlanta.

The four detectors are located on different interstates at arguably the most congested locations around the city. Detector 285592 is located at the New Northside Drive exit on Interstate 285 East.²⁷ This location is just east of where Interstate 75 intersects the perimeter in a densely populated area, so it has high traffic flows for the morning and evening commute as drivers use the perimeter driving into and out of the city. Detector 75585 is located just north of Delk Road on Interstate 75.²⁸ The detector is outside the perimeter on the southbound left lane; therefore there is much more traffic volume detected in the morning as commuters use the interstate to enter the city for work. Detector 714 is located in downtown Atlanta on the portion of interstate where 75 and 85 join through the middle of the city.²⁹ Located close to the International Boulevard exit on the northbound side, this detector generally would detect a higher traffic volume in the morning as most of the employment is closer to the north side. Finally, Detector 201581 is located on westbound Interstate 20 at the Wesley Chapel Road exit.³⁰ This detector location is inside the perimeter but on the east side, so this location has high traffic volumes in the morning as people go into the city from the eastern suburbs along Interstate 20.

One difficulty for this data set is that there is no concrete way of telling if the detector was malfunctioning at the time of reporting. The only way to know the detector is malfunctioning is if there was no data being reported, and this work avoids using those

²⁷ Mann.

²⁸ Mann.

²⁹ Mann.

³⁰ Mann.

detectors because of a need for regular data anyway. However, if the detection device is under-reporting or over-reporting data values, there is no method of discerning such an error. The only way to notice an error such as this would be to notice uncharacteristically low or uncharacteristically high data points for the volume data. This observation in itself does not mean that the data is erroneous but could rather mean an exceptional circumstance such as a road closure or a cultural event occurring nearby. There are a few outliers that appeared in the data when experimentally graphing different segments. However, because of the method used to create models for the data, these outliers become insignificant. Also, the outliers were not associated with only one detector nor were they reported in a relatively short time frame but rather were spread throughout the year in different detectors.

One example of missing data that impacts this study significantly is the lack of data for all of February and March of 2003. While most days throughout the year reported data, possibly for only a portion of the day for a few select detectors, data is absolutely nonexistent for any detector for any time during those two months. The data for January 2003 is scant, at best, and then completely nonexistent for the next two months. When asked about the lack of data for two months of the year, the GDOT did not have an explanation and could only reiterate that sometimes the detectors malfunctioned. Because there is no data from these months, the analysis of yearly trends suffers. This work can only analyze data from April through December and must assume that January through March data is somehow similar to the other data. Therefore, analysis proceeds with the data that is available for four detectors that were consistent in reporting data.

3 Volume Trends

One of the goals of this independent work is to use historical data to discover trends in traffic volume data. Using the available data from the year 2003, this chapter attempts to model trends in the data using smoothing regression functions across weeks, months, and the entire data set. Also, it makes comparisons among the four different detectors chosen for this analysis. This chapter will first look at possible explanatory variables, then how one might model volume with respect to time for the various intervals, and finally it will focus on the trends in the Atlanta interstate data.

Of the data sent by the GDOT, the four detector locations used had an average of 3205 data points for the time between April and December. Again, only the months from April through December are used in this analysis because data points are unavailable for all of February and March and are available but unreliable for January.

Upon receiving the data from the GDOT, it was sorted using an Opttech Sort program to organize the data by detector, day, and time rather than by day (as it was sent). For each of the four detectors, the data was then separated into the following format in Microsoft Excel:

Detector	Date	Time (24 hour)	Volume	Speed	Occupancy
714	4/7/2003	00:00	121	65.38	3.82
714	4/7/2003	00:15	114	67.42	3.55
714	4/7/2003	00:30	97	66.93	2.98
714	4/7/2003	00:45	62	67.94	1.91
714	4/7/2003	01:00	53	65.23	1.65

Once in this format, the data could be analyzed further.

3.1 Possible Explanatory Variables

3.1.1 All Variable Choices

Under ideal circumstances, one would be able to know a historical volume value as a result of a number of explanatory variables. Explanatory variables could include weather conditions, reports of incidents, special events, reports of construction, and time elements. Thus, one would have:

$$\text{Traffic Volume} \sim f(\text{weather, incidents, special events, construction, time})$$

Equation 3.1: Explanatory variables of traffic volume

All of these variables absolutely have a direct effect on the volume of traffic on the interstate. Unfortunately these variables are also extremely difficult to quantify

exactly. For instance, a concert might increase the volume of cars on the road but by how much depends on the size of the concert. Also, a snow storm might affect the number of vehicles out on the road, but exactly how many travelers will be deterred by the inclement weather? Because these variables are too difficult to quantify within the scope of this work, this analysis focuses only on the concrete time variables when predicting traffic volume.

If it were possible, a study should have been conducted when these other factors had only minimal influence. However, for this study, this sort of data separation was impossible due to a total lack of information about such factors. A more comprehensive study might consider taking such factors into consideration or eliminating them completely to obtain volume estimates under normal conditions only.

3.1.2 The Time Component

Though the time element might seem to be a fairly one-dimensional variable, there are a few classifications to be considered. For travel time, the *Travel Time Data Collection Handbook* suggests four time elements for consideration: month, day of week, day type, and time of day.³¹ Day type refers to whether or not the day is a holiday. Schrader makes the suggestion in his thesis that the year-to-year trend might be another element to consider as roads become more congested with time.³² These five elements can be applied, not only to travel time estimates, but also to volume estimates because travel time is a function of the volume of vehicles on the roadway because of the relations

³¹ Shawn M. Turner, et al. Travel Time Data Collection Handbook. Report No. FHWA-PL-98-035. March 1998. Federal Highway Administration. Texas Transportation Institute. National Technical Information Service. <<http://www.fhwa.dot.gov/ohim/tvtw/natmec/00020.pdf>>. 2-11.

³² Schrader 40.

that both volume and travel time have with speed.

Time Elements for Consideration:

- Time of Day
- Day of Week
- Type of Day (Holiday)
- Month
- Year (year-by-year trend?)

Due to the limitations of both this study and the data from GDOT, this analysis only accounts for a few of these time elements. However, in an expanded study, they could all be incorporated more extensively.

This study is actually greatly simplified by these limitations. For instance, detector data for the Atlanta interstates is not necessarily continuous from year to year. Because managing these detectors is quite a task, detectors malfunction and go offline while new detectors are being added to the system. For this reason, it may be difficult to find year-to-year patterns in the traffic volumes plainly due to lack of available data. Assuming that a detector that works regularly for a period of years, the patterns found in this data set may not be representative of yearly patterns for the roadways as a whole but rather present a trend exclusive to that small data set. Also, the models in this work are analyzing monthly trends, so it is not necessary to use the months component as an explanatory variable. Therefore, this study focuses only on day of the week and time of day as explanatory variables. This study deals only with Mondays throughout the year, thus eliminating the day of the week variable from further calculations. There is only one holiday that falls on a Monday in 2003, Labor Day, and rather inconveniently, none of

the detectors reported data on this day. Subsequently, this work focuses only on time of day in its analysis.

EXPLANATORY VARIABLE (for modeling volume trends): Time of Day
--

3.2 Modeling Methods

Because the data involves time elements, one might be tempted to think that time series is the appropriate model for this data. With enough data and the right tools, one might be able to accurately model yearly trends, seasonality throughout the year, seasonality throughout the week, and seasonality throughout the day. However, the data collection for such a model would have to include a set of measurements taken at regular time intervals.³³ While this data from the Atlanta interstates is aggregated into regular 15 minute increments, there are gaps and irregularities in the data when the detector malfunctioned and did not produce data. It might be possible to extrapolate a regular time series from this irregular time series, but because the volume measurements do vary so widely, such extrapolation gives no guarantee of accuracy and includes a large subjective component based on user input.

Another complication arises when attempting to make the data stationary. In his paper forecasting travel times, Sen, et al. notes that travel times are not stationary due to a daily pattern in which morning and evening peaks occur repeatedly.³⁴ By this same reasoning, volume data is not stationary because of a morning peak and a smaller evening

³³ Rene Carmona. Statistical Analysis of Financial Data in S-Plus. (New York: Springer-Verlag, 2004), 240.

³⁴ Ashish Sen, et al. "Short-term Forecasting of Link Travel Times: A Preliminary Proposal." URBAN TRANSPORTATION CENTER, University of Illinois at Chicago. ADVANCE WORKING PAPER SERIES, Number 7. November 1991. <<http://ais.its-program.anl.gov/advance/reports/REPORTS.HTML/wp7/wp7.html>>.

peak. Adding weekly seasonality, yearly seasonality, and holidays make a time series of volume measures extremely problematic. This model can be done, but again, involves a large subjective component by the user and can be difficult for a large number of locations.

Because time series modeling presents too arduous a task for such a study as this one, a simpler and less subjective method of modeling that more appropriately fits the data for the Atlanta interstates and that can be used in a further study of the roadways must be found. Given the shape and variability of the volume measures, a polynomial regression does not seem appropriate either. Also, one might be tempted to model the morning and evening peaks using a multi-parameter function of normal curves as Schrader did in his thesis with travel times.³⁵ However, the volume data for the entire year for each detector, as shown in Figure 3.1, does not have a simple two peak shape as might have been expected. The location and number of normal curves needed to map the area in between the morning and evening peak is not obvious and might even change from month to month as the volume measures change. Because Schrader's other method of modeling proved to be nearly as good, it is considered next.

³⁵ Schrader 44-47.

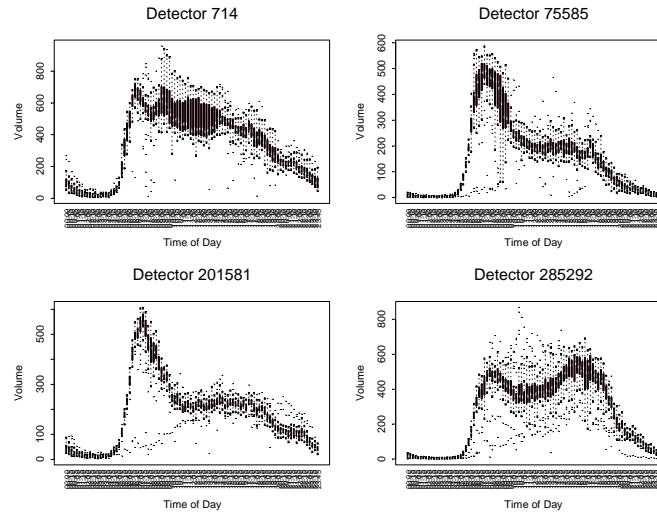


Figure 3.1: All volume data for each detector for year

Nonlinear parametric and nonparametric regressions are considered as a final option. Nonparametric regressions such as the kernel regression prove to be useful measures for modeling the data. These scatterplot smoothers for univariate explanatory variables aim to represent a data set of points $(x_1, y_1), \dots, (x_n, y_n)$ by the graph of a function $y = \varphi(x)$.³⁶ Nonparametric means that the function φ is not expected to be determined by a small number of parameters and the regression function will not be restricted to any specific function class.

The kernel smoother, like other smoothers, relies on observed values of neighboring points to predict the response $\varphi(x)$. However, instead of relying on a certain limited number of points y_i to characterize the smoothing, the local averaging is determined by a weighted average of all the observed values y_i with the weights decreased at increasing distances between x and the corresponding x_i value. The weights are computed with a kernel function $K(x)$ and a smoothing parameter. There are four

³⁶ Carmona 178.

types of kernel functions: box, triangle, parzen, and normal.³⁷ This study uses the normal which uses the Gaussian density function. The equations are shown below. The smoothing parameter is called bandwidth; a number greater than zero, this value is a measure of how closely the smoother will model the data. Choice of this variable can make or break the analysis in most cases; however, for this case, the choice is a little less important. Because most of the trend analysis is comparative, as long as one chooses the same bandwidth to model the data each time, the bandwidth does not have as large an effect on the efficiency of the modeling process.

$$\varphi(x) = \varphi_{b,K}(x) = \frac{\sum_{i=1} K[(x-x_i/b)]}{\sum_{j=1} K[(x-x_j/b)]}$$

$$K_{\text{normal}}(x) = (\sqrt{2\pi})^{-1} * e^{-x^2/2}$$

Equation 3.2: Kernel Scatterplot smoother function, Kernel normal function³⁸

3.3 Atlanta Interstate System

This study evaluates the Atlanta interstates data using the methods described above. For the months April through December of 2003, it evaluates the data for Mondays in search of trends. As stated in Chapter 1, an assumption is made that there will be spikes in volume measures during peak rush hour periods and this work looks beyond these observations for other trends. This model uses the normal kernel regression model to create functions that approximate the data so that one might compare the data across different time intervals, but first the raw data shapes are characterized.

³⁷ Carmona 185.

³⁸ Carmona 184-185.

This work originally makes a blanket assumption that there will be a distinct spike in the volume data associated with the morning and the evening peak. However, upon reviewing the specific data chosen for the Atlanta interstates, this assumption is not supported. There is usually a morning peak, but the evening peak is often not robustly present in the daily data. Upon further consideration, the reasoning for this missing/low evening peak is quite simple. The specific detector locations throughout Atlanta were chosen because of their high traffic volume at some time during the day. If one were looking at volume measurements for both directions at these locations, then the data would probably show the expected morning and evening peaks, but since this analysis is only concerned with one detector measuring volume in one lane going one direction at each point, it loses the data from those same commuters traveling at another time of the day. The home-work trip is actually a round-trip each day including a home-to-work component and a work-to-home component, and this work only accounts for one of those trips as it detects volume on only the southbound traffic on Interstate 75 at a specific location, for example.

Some discoveries worth noting are the different shapes of the detector data. Figure 3.1 above shows whole data set for each detector, and even though the data has a wide range of values for each time of day, there are certain obvious trends present for each detector. It was mentioned above that the locations of these detectors give a reason for their possibly uncharacteristic shapes; however, it is worth a bit deeper look to discover commuter behavior for residents in and around the city. As one might have expected, Detector 201581 and Detector 75585 have high spikes in volume associated with the morning rush hour because they are both located along interstates leading into

the city. Because so much of the workforce lives in the suburbs and commutes to the city, these inbound interstates are densely packed in the mornings while the outbound sides on these interstates have its highest spike in traffic flow in the evening.

Perhaps more interesting are the other two detectors. Detector 285292 has two almost symmetric humps associated with the morning peak and the evening peak with the evening peak being only slightly larger. This detector is along Interstate 285 which encircles the city. One might expect that if people use this road rather than fighting traffic through the city to get to work, then later in the day they would be going the opposite direction and not contribute to the volume count. However, since this detector is located almost exactly north of the city, it might indicate that people on both the east and west side use Interstate 285 to avoid using Interstate 75 and Interstate 85 which go through the city. This interstate is used by residents in general on the north side rather than by just a specific geographic subsection. Detector 714 is a bit different. It has a peak around the morning rush hour, but this peak only gradually drops off during the day then drops significantly in the late evening. The spike in the morning is most likely explained because residents from the south side of the city or the southern suburbs are heading to work on the north side. The reason for the gradual tapering is a bit less obvious. Since this detector is downtown, this shape might indicate that people downtown use the interstate all day long to move throughout the city rather than only relying on the surface streets to provide mobility. Nonetheless all of these detector shapes were not necessarily the shape that might have been expected at first, but upon further consideration, all have explanations for their unique shapes.

A month-by-month analysis using the kernel regression method discussed earlier reveals that each month does not necessarily have the same characteristic shape as the general trend for the year. The Highway Capacity Manual makes some generalizations regarding monthly trends that one might consider for when looking at the data:

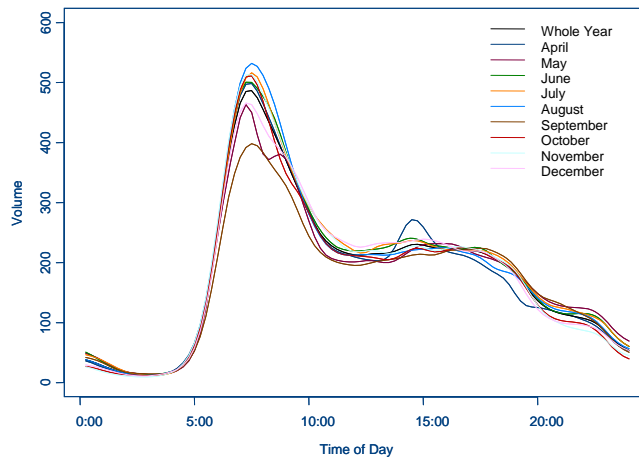
1. Monthly variations are more severe on rural routes than on urban routes.
2. Monthly variations are more severe on rural routes serving primarily recreational traffic than on rural routes serving primarily business routes.
3. Daily traffic patterns vary by month of year most severely for recreational routes.³⁹

Keeping these three things in mind, it might seem as if there would be little or no difference between the months on urban, business routes, but this analysis discovers otherwise. By looking at Figure 3.2, one can observe that most of the months generally do follow the yearly trend (in black). However, there are some outlier months. For instance, for Detector 285292, Detector 714, and Detector 75585, the month of December is different from the other months. One might attribute this difference to holiday preparations including shopping and visiting relatives. The two detectors where the volume measures are much higher than the yearly average are the two detectors that did not have the high volume spike for the commuter traffic but rather the detectors with more continuous traffic volume data, Detector 284292 and Detector 714. Both of the smoothing regressions for these detectors for the month indicate elevated midday traffic volumes with a large spike around the noontime hour. This spike is probably attributed to workers going out to get some extra shopping in during their lunch hour. The two detectors that are associated most with rush hour traffic trends are generally the most

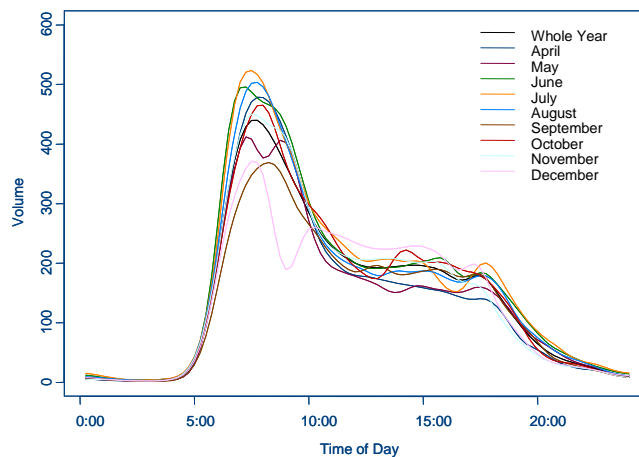
³⁹ Highway Capacity Manual (Special Report 209), 2-6.

continuous from month to month. This is to be expected since most people work year-round. For April, the 285292 detector shows a regression with volumes approximately 100 counts less than the yearly average; however, April is fairly consistent with the yearly trend for all other detectors. One might attribute this lower flow with a road closure or other short-term non-normality in the data. Also, for the two detectors deemed more associated with travel outside the daily commuter peak (Detector 285292 and Detector 714), the November midday traffic volumes are a bit higher than the yearly trend. This volume increase might also be attributed to traveling associated with Thanksgiving and pre-December shopping.

Detector 201581 Month By Month Smoother



Detector 75585 Month By Month Smoother



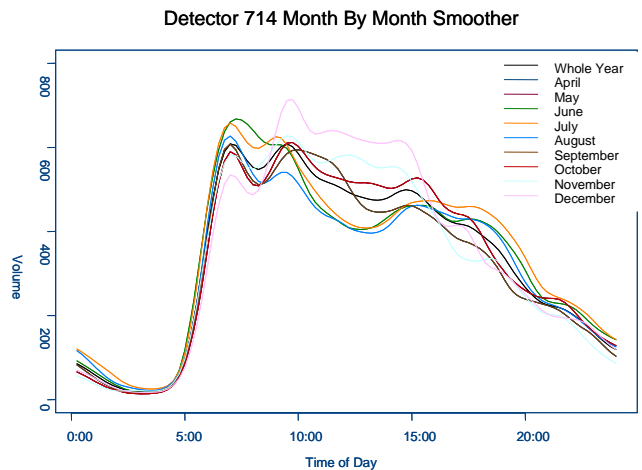
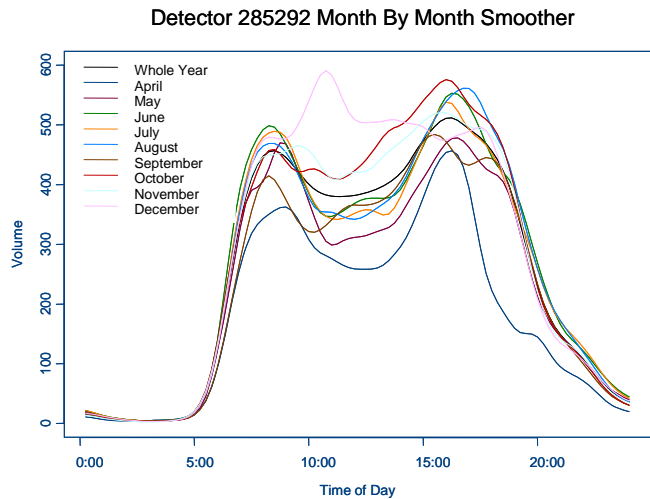


Figure 3.2: Month-by-Month Smoothing Regressions for each Detector

Because the data has gaps where data was not reported, the comparative analysis is more difficult across single days than it is with months. However, even with limited data, one can notice that the volume trend each Monday in a month is fairly similar.

Figure 3.3 graphs traffic volume against time of day and date for the detector located on I-75 (ID#75585) for the month of August. Clearly there is a spike in volume associated with the morning peak, and a smaller more spread-out increase for the afternoon peak.

Also, one notices that volumes on the road are significantly higher during the daylight hours, another sensible trend. Appendix A gives diagrams of the other detector locations

for August. The month of August is chosen only because the data had no significant gaps so trends across the weeks could be more accurately seen. From different angles, it is easier to see the variation among days. One might assume from these graphs and from the monthly smoothing graphs for each detector that generally traffic volume patterns are fairly similar throughout the year. Though this data presents a similar traffic flow for all Mondays in a particular month, one must remember that this data is only for Mondays, and one should not attempt to project a larger assumption that volume flow across all days is similar. In fact, such an assumption would prove utterly wrong if the weekend data had been considered since most of the workforce does not commute to work on the weekend; rather, most traffic is for recreational purposes.

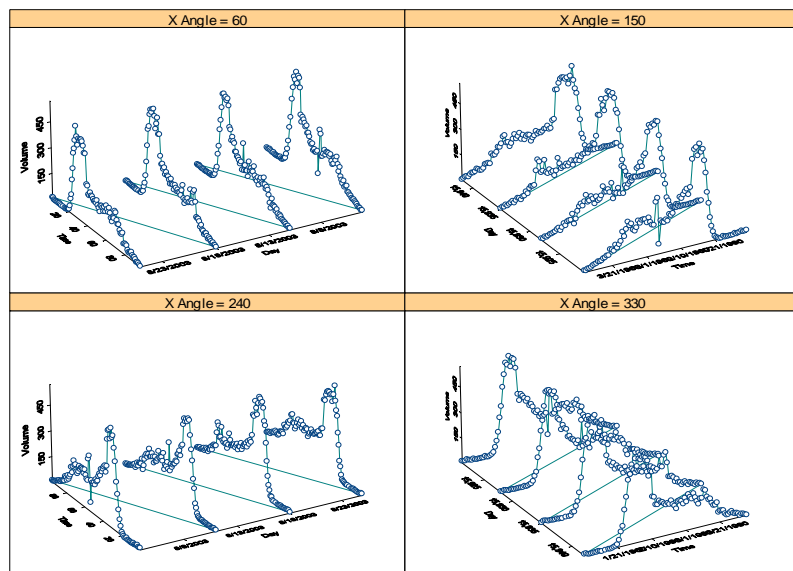


Figure 3.3: Traffic volume as a function of time of day and date for Detector 75585

Though data is very similar throughout the year, different trends in the data do emerge. For instance, the December midday increase for the detectors downtown and on the perimeter are fairly robust outliers from the year-long trend. Generally, the two detectors associated most with the morning commuter traffic, Detectors 201581 and

75585, have monthly models most similar while the other two detectors have a midday variation of between 200 and 300 vehicles during different months. Although there are many minor divergences from the yearly trends present in the monthly data, because of the nature of how few data points are used, one cannot make any concrete statements regarding these differences. This work only observes larger differences. The kernel regression method does prove to be a very effective method of modeling for this case as this analysis is able to smooth these nonparametric functions with relative ease. Also the kernel regression automatically eliminates the rogue points that occur every so often in the data without having to individually go through accounting for those points. Because of the averaging effect of the kernel smoother, what matters most is the trend rather than the individual points. Some trends were revealed by this method of comparative analysis, though not as many as originally expected. Due to the scale of this project, it was not possible to make any discoveries about nuances present in the data but rather it relies only on large trends that are likely to be repeated in other, larger works. A study with more continuous data would be able to verify and elaborate on the monthly trends discovered here.

4 Speed-Volume-Density Relations

While the previous chapter dealt with trends in volume data according to time, this chapter works with the relation between the volume of vehicles on the road and the speed at which they are traveling and also the relation between traffic density and speed. One could simply assume that as traffic volume and density increase, speed decreases; however though this simple logic may be correct, the development of a specific model using historical data is a bit less intuitive. The ultimate goal of Chapter 4 is to develop such models. Section 4.1 discusses aspects of speed for an interstate and how speed varies according to certain parameters. The following section, 4.2, discusses the

fundamental relation between speed and volume and speed and density. It explains the various models which seek to explain these relations. Finally these models are applied to the data obtained from the GDOT to see what the data reveals about speed relations and roadway capacity in section 4.3.

4.1 Characteristics of Speed

Speed is one of the most important measures of the quality of traffic service for a driver. However, this measure is most often dictated as a function of road type and conditions. Just as it would be unreasonable for a traveler on a surface street in the city to drive much faster than about 35 miles per hour, it is unreasonable for a traveler on a rural unpaved road to drive as fast as might be afforded by the openness. As another example, lower speeds are tolerated on steeper slopes because of the comfort factor of the drivers maneuvering these roadways. Specific conditions such as lane width, presence of a median, number of lanes of traffic, and shoulder space also have an effect on speeds. Generally, the more space a vehicle has, the more comfortable the driver will be and the faster he will travel. Speed is a definite function of roadway condition and roadway type but also of vehicle type. Cars obviously travel faster than semi-trucks which is why trucks are constrained to the right lanes for slower traffic. However, possibly surprisingly, national speed trends for all vehicle types show a general increase in average speed on interstate highways. The speed-volume, or speed-flow, relationship has never been definitely characterized (possibly because of the largely qualitative nature of some of its variables). Using the example of I-35W in Minneapolis, the Highway Capacity Manual notes that that speed remains relatively constant despite significant

changes in volume.⁴⁰ If this observation were true for all roads then the analysis of Atlanta interstates would be rather unenlightening. In addition, speed in general would not be nearly as effective as an indicator of the level of service to drivers. However, other experiments have been able to develop general models for the speed relationship; this work examines these models in an effort to fit the data.

Just as this work analyzed trends in volume in the previous chapter, researchers have analyzed variations in vehicular speed for time intervals including years, seasons, months, days, and hours. From 1942 to 1966, the rate of yearly increase in spot speeds on rural highways was approximately 1.0 mph per year.⁴¹ Also, measures of average speeds at different times of the year indicate that speeds are highest in fall and winter, intermediate in spring, and lowest in summer.⁴² These sorts of trends would make sense when considering that summer is the most likely time for travel and recreational driving. Drivers tend to go slower when unfamiliar with an area, so summer tourists would lower these average speeds. There have been statistically significant studies regarding differences in mean monthly speeds as well. As far as variations within the week, only Sunday has been robustly proven to have lower speeds than other days of the week.⁴³ Again thinking about the culture of American society, Sunday is a day of worship for many Americans, and people might be more sensitive to the dangers inherent to speeding on a day that they devote at least part of the day to reverence. As far as how the hour of the day influences speed, this factor is heavily influenced by the amount of light provided for the roadway during the night when natural light is not available. Drivers tend to

⁴⁰ Highway Capacity Manual (Special Report 209), 2-20.

⁴¹ J. C. Oppenlander. "Variables Influencing Spot-Speed Characteristics." Highway Research Board. (1966) 26.

⁴² Oppenlander 26.

⁴³ Oppenlander 26.

travel much faster when they have a better line of sight of the roadway ahead. While there may tend to be trends associated with time of day/month/season/year, there are also other factors influencing speed.

Weather conditions tend to play a large role in vehicle speed, even more so than on volume count. Inclement weather can cause reduction in visibility and impairment of the road surfaces. In general, weather only causes slower speeds; rarely, if ever, does weather cause an increase in speed. The amount of speed reduction depends on severity of the weather. Obviously blizzard conditions would be different than just a light rain storm. The reason volume measures might not be as effected is because most commuters still have to get to work whether it is rainy or sunny; however the speed at which those same commuters move in the two different conditions is necessarily effected by the circumstances.

An interesting characteristic of speed is that it is the reciprocal of travel time. Thus it is valid to think about slower speeds as directly correlating to higher travel times. One might expect that speed data would change drastically throughout the day. After all, anyone who has ever been stuck in rush hour traffic will attest to the fact that it sure seems to take forever to move anywhere. In actuality, the changes in speed for most of the day are not that drastic. Figure 4.1 below shows the speed data for each detector for the year.

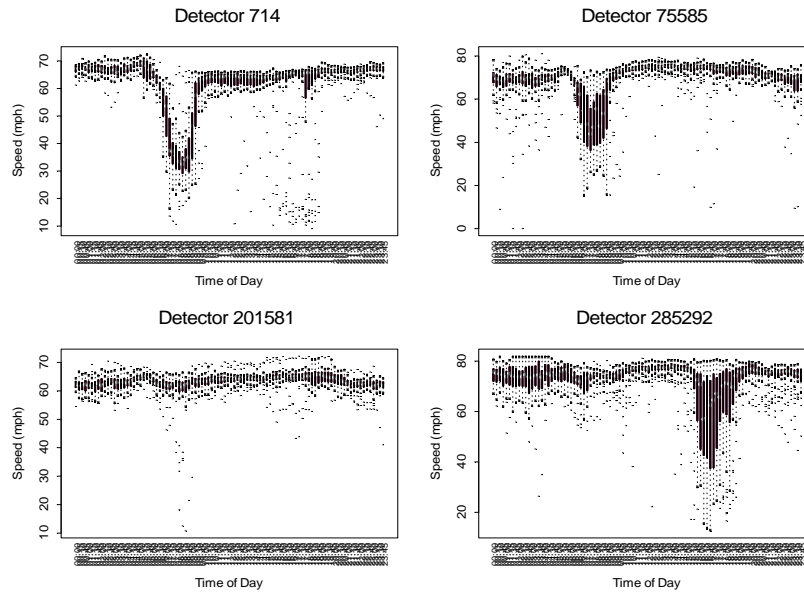


Figure 4.1: Speed Data for each Detector for the Year

Of course, for the two detectors associate with a huge volume of inbound morning traffic, the speeds slow considerably in the morning. However, for Detector 201581, the speeds change only minimally for the entire day. For Detector 285292, the big drop in speeds is associated with the evening rush hour, which is surprising since the volume data for the detector (shown in Figure 3.1) does not show a huge spike in volume at that time but rather shows an increase fairly comparable with the morning peak.

When graphing the inverse of the speed data, one really gets a sense of why it does seem to take forever when getting stuck in rush hour traffic. Figure 4.2 shows a graphical representation of travel time. Though the little humps might seem insignificant, consider the units. A travel time increase of 0.02 on the graph would mean a 0.02 hour increase in travel time per mile. This equates to a one minute and twelve second increase in travel time per mile or an extra 12 minutes to go ten miles. If work were only about 20 miles away, a distance that would take about 20 minutes on the interstate at normal

speeds, it would take an extra 24 minutes on top of those 20 to get there during rush hour. As insignificant as these humps might seem, their effect is actually quite substantial.

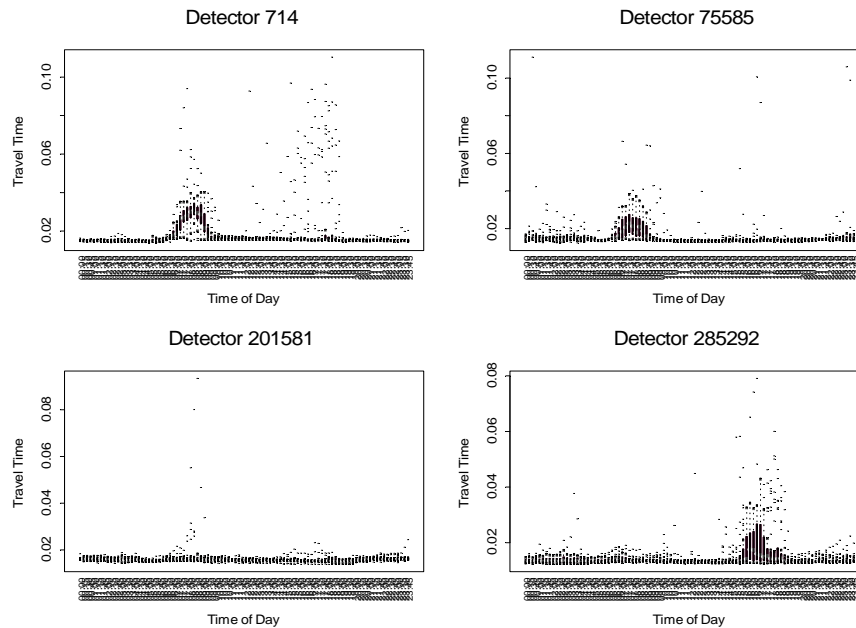


Figure 4.2: Travel Time vs. Time of Day for the Year

The next section explores other speed relations that might not be as well known or obvious.

4.2 Fundamental Relationships regarding Speed

As much as the above factors such as season or weather conditions affect speed, a fundamental relationship has been discovered that reveals the definite, quantifiable relationship between volume of vehicles on the road and speed of the vehicles. Also, considerable research has been done that attempts to estimate the most accurate speed/density relationship. Before describing these relationships, one must first characterize the traffic situation.

Freely flowing traffic is traffic flow when each vehicle can travel at the desired speed of the driver, without being affected by other drivers but rather only constrained by roadway and vehicle constraints. This is only the case if there are few cars on the road and there are multiple lanes for overtaking if necessary. A driver in free flow traffic is subject only to the constraints of his vehicle and the road and thus travels at a speed deemed the desired speed.⁴⁴ The desired speed of each driver is often a function of distance traveled by the driver; however it can also be a function of time of day or some other measure.

When traffic becomes heavier, drivers will have less of an opportunity to maintain their desired speed. More often, drivers will have to reduce their speed to that of a slower vehicle until an opportunity to overtake the slower vehicle emerges. These opportunities to overtake appear less and less often as traffic flow increases. This kind of traffic flow, when some drivers are not free to travel at their desired speed, is called partly constrained traffic.⁴⁵

Traffic can also become completely constrained when it is not possible for drivers to carry out their desired overtaking maneuvers. In this type of traffic congestion, all drivers travel in one or more platoons. A platoon is a line of vehicles in which each vehicle's speed is constrained by the vehicle ahead except for the first vehicle that sets the slow speed.⁴⁶ The decrease in average speed begins slowly in partly constrained traffic but can drop drastically as traffic moves toward completely constrained traffic.

⁴⁴ Wilhelm Leutzbach. Introduction to the Theory of Traffic Flow. (New York: Springer-Verlag, 1988), 93.

⁴⁵ Leutzbach 93.

⁴⁶ Leutzbach 93.

In an effort to avoid having the data affected by platoons and constrained traffic except when necessary, this study uses data from the leftmost lane (first non-HOV lane) on each interstate. Assuming that the left lane is generally used as a lane for faster traffic or for overtaking vehicles in other lanes, speeds will not be as affected by exceptional cars traveling slower than the general body of traffic. Also, by using the left lane, this analysis avoids possible deviations in the data because of vehicles entering or exiting the highway at exits. With few exceptions, vehicles enter and exit the highway from the right; therefore when traveling in the left lane, they are not still accelerating from entry to the highway or yet starting to decelerate in anticipation of exiting the highway.

Traffic flow can increase until a certain rate of flow defined as capacity. A direct measure of absolute capacity is hard to obtain for several reasons. The observation of a high volume or flow rate does not guarantee that a higher flow could not be accommodated at another time. Also, it is not a stable operating condition. Finally, capacity for a given location can change due to variable conditions because the determination of capacity incorporates various considerations including weather conditions, road type, etc. For example, the capacity for a highway is much lower during a blizzard/ice storm than on a sunny, summer day. Capacity is most often calculated by using a density-flow curve for a given location on a highway. Capacity is often deemed as the point at which flow reaches a sort of barrier and speed drops off significantly as the number of vehicles increases.⁴⁷

According to the Highway Capacity Manual, the peak capacity per hour per lane for a multilane highway is 2,000 vehicles under ideal conditions.⁴⁸ The peak hour of

⁴⁷ Highway Capacity Manual (Special Report 209). 2-23.

⁴⁸ Highway Capacity Manual (Special Report 209). 2-2.

traffic volume, or rush hour, is the most critical period for operations and has the highest requirements of capacity. Traffic engineers face the dilemma of providing adequate road space for peak hour capacities with the underutilization of capacity for the rest of the hours in the day. The decision for a compromise might be easier to rationalize if that did not necessarily mean standstill congestion during peak periods.

Incorporating considerations from the free flow versus constrained traffic situations, there is a relation between vehicle speed and density of vehicles on the road. Even if a vehicle ahead of a specific driver is traveling at the same speed as the desired speed of the following driver, the second driver will be more likely to slow down until sufficient headway is available between his car and the car ahead. As speed increases, the acceptable headway necessary in front of a driver's car increases. Conversely, as speed decreases, the required headway decreases. Slower vehicles cause other vehicles to slow down, creating denser traffic as cars leave less headway. Therefore the effect of slower vehicles is twofold: slowing traffic behind and causing the slower traffic to become more densely packed as the vehicles slow.

With regards to the speed-density relationship, as the number of cars on the roadway increases, the speed of these vehicles decreases. Greenshields postulated a linear relationship between speed and density in his 1934 study of capacity. The model is advantageous because of its simplicity and does provide a good fit for some observed cases. A mathematical expression of Greenshield's finding looks like this:

$$S = S_f (1 - D/D_j)$$

where S = speed (mph);
 D = density (vphpl);
 S_f = free-flow speed (mph); and
 D_j = jam density (vphpl).

The linear model is simple and useful; however other nonlinear models have been created to model traffic flow with perhaps better results. Greenberg developed a model based on a “one-dimensional” fluid state which takes the following form:

$$S = S_c * \ln (D_j/D)$$

with S_c equal to the critical speed at capacity (mph). The model describes the behavior of congested flow well but does not work for low densities because the theoretical speed approaches infinity as density approaches zero.⁴⁹

There have been other hypotheses about measuring speed-density models. These other models include a two-part linear model, a three-part linear model, the Underwood model, the Edie model, and a bell curve model. The two-part and three-part linear models are fairly self-explanatory, basically a compilation of linear models. The Underwood model is an exponential decay curve expressed as follows:

$$S = S_f * e^{(-D/D_m)}$$

With D_m being the optimum density. This equation was developed as a steady-state theory for non-congested traffic. The Edie model is a combination of two equations, an exponential for densities less than the optimum and a logarithmic for densities greater than the optimum density. The equations are listed below:

$$S = S_f * e^{(-D/D_m)}, \text{ for } D < D_m;$$

$$S = S_f * \ln (D_j/D), \text{ for } D > D_m.$$

In a statistical analysis by J. Drake, J. Schofer, and A. May, the Edie hypothesis was found to be the best of all models chosen to model modern freeway speeds. However, all hypotheses except the two-part linear model performed well enough to warrant continued

⁴⁹ Highway Capacity Manual (Special Report 209). 2-22.

use for modeling purposes.⁵⁰ The data does not provide a direct density measure, but from the data provided, it can be calculated.

Alternatively, the speed-flow relation can be modeled using the data directly. Flow rate, speed, and density are related by a simple formula, $v = S * D$, known as the fundamental relation; some general observations can be made from this direct relation first before testing it empirically. If there is zero density, then there is zero flow, and if the roadway is at jam density (speed equal to zero), then the flow is also zero. There must be one or more maximum flow values between the point of zero density and the jam density. Generally data shows one such maximum flow rate with one upper curve showing stable flow points and a lower curve showing unstable, forced flow points. All tested models indicate that the stable flow curve is much higher than the unstable flow curve, possibly as much as 200 vehicles per hour higher.⁵¹ Since speed-flow is most readily measured from traffic stream parameters, this type of model is most often formed from the observed data. Most prior models have indicated that multilane flow capacity occurs at a critical speed around 30 miles per hour. This data's level will most likely be different because it evaluates only one lane of a multi-lane highway, the lane that is most likely to be maintaining a higher speed at volumes near capacity. The capacity level is the approximate level that they data points approach but do not cross. A simplified model is shown in Figure 4.3 below. This is the speed-flow model this work will attempt to fit to the data from the Atlanta interstates in the next section.

⁵⁰ J. Drake, J. Schofer, and A. May. "A Statistical Analysis of Speed-Density Hypotheses." Vehicular Traffic Science. (New York: American Elsevier Publishing Co., Inc., 1967), 113-115.

⁵¹ Highway Capacity Manual (Special Report 209). 2-23.

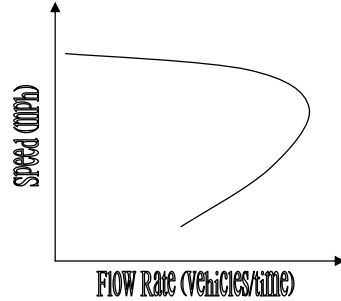


Figure 4.3 Simplified Model for Flow/Speed Data (Fundamental Diagram)

4.3 Atlanta Intestate System

Using the data provided from the four detectors, the first attempt will be to fit a speed-flow relation from the raw data. Then this analysis will use the fundamental relation to calculate density and attempt to fit one of the speed-density models to the data. By using both the speed and the volume data provided by the DOT, this work accepts that the method they use to calculate speed (using an artificial zone and measuring entry to exit time for the zone) is accurate and consistent. If this assumption is not the case then no sort of robust relationship between speed and volume or speed and density can be developed.

The data provided for the Atlanta interstates gives both speed and volume data. By plotting the raw data for the four detectors, one sees that the empirical representation of the speed-flow relation is not as clearly defined as theory might indicate. Unlike the clear shape of Figure 4.3 above, the empirical data points of Figure 4.4 do not easily lend to an obvious regression. Also, one notices that there are several errant points in the data that would probably need to be cleaned from the data in a more sophisticated analysis. For instance, in Detector 75585, there are several flow rates of zero indicating errors in

reporting by the detector device. These points must be errors because there are speeds reported for these flow rates of zero, an impossibility in real life.

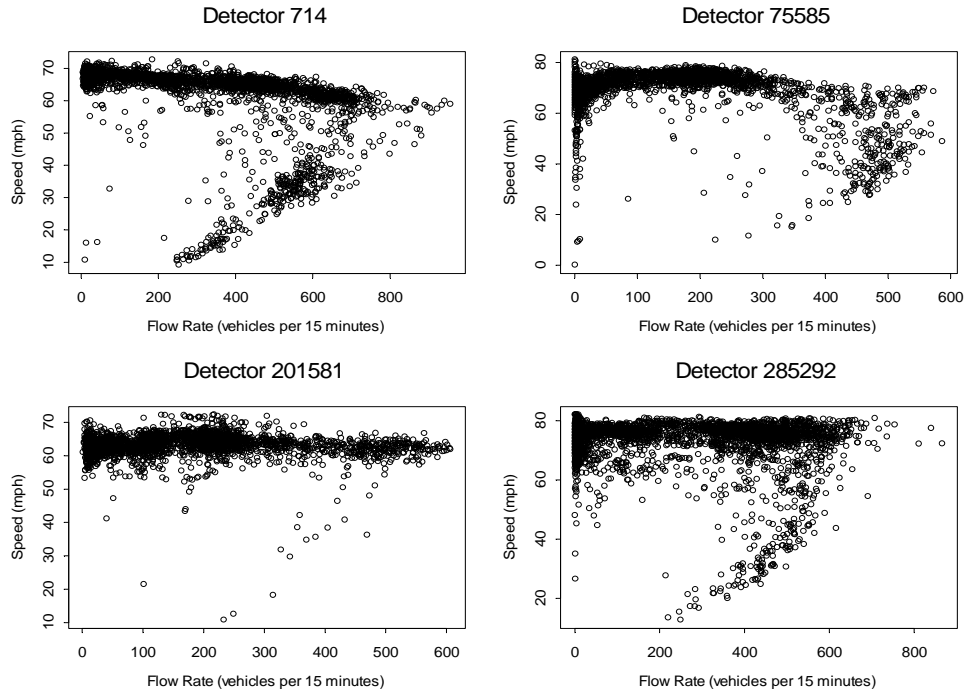


Figure 4.4: Empirical Data of Fundamental Diagram

One can make generalizations about these graphs from the figure above, but it is only by scrutinizing each graph individually that one can create models for the speed-flow relation. Generally, it looks as though vehicles can travel at approximately the same speed until a breakdown capacity at which travel becomes forced and unstable (and proceeds to follow the lower curve).

When creating a general model for this graph, the shape of Figure 4.3 above does emerge. However, the curves for these detectors are obviously not robust and merely approximations of a general shape. One result is found below in Figure 4.5; the rest are in Appendix B.

Empirical Speed-Flow Plot for Detector 714

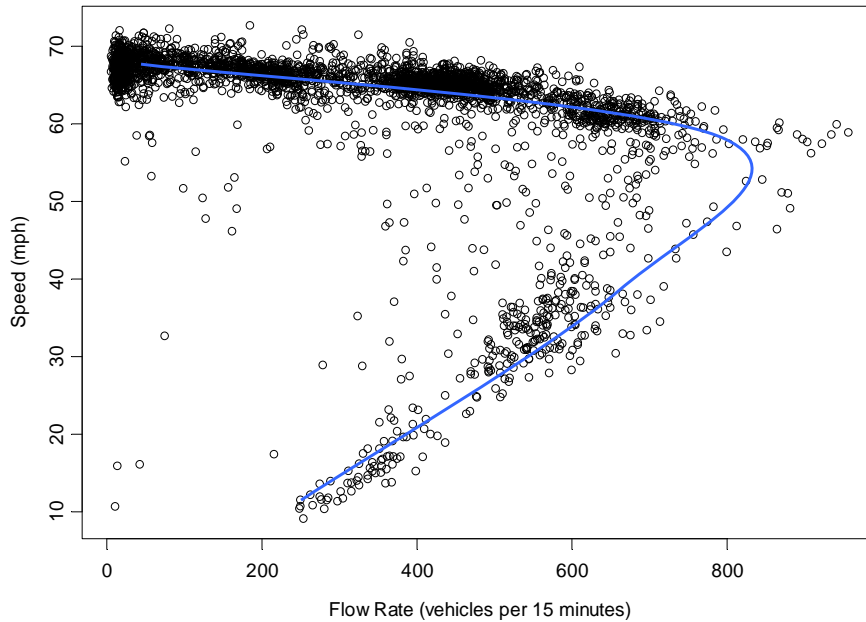


Figure 4.5: Speed-Flow Data and Model

This model shows a capacity, or breakdown point, at around 800 vehicles per 15 minutes. The capacities for each of the four detector points are different, from only around 475 vehicles for detector 201581 to approximately 650 vehicles for detector 285292 to about 525 vehicles for detector 75585. Data has shown that capacity is the point where the speed-flow model breaks down and speeds drop significantly. In this model, there are a few points with flow rates higher than the capacity, but this analysis tried to create a model of best fit rather than a model that took into account the actual values of every single point. Because of the range of variables that could not be accounted for that might have influenced speed, one might agree that this data does indeed support the fundamental speed-flow relationship found in research.

Next, the density is calculated to define the speed-density relationship for the data. Of the models described in the earlier section, the three-part linear model was chosen as the best. An example of one of the models is shown below in Figure 4.6.

Empirical Speed-Density Plot for Detector 714

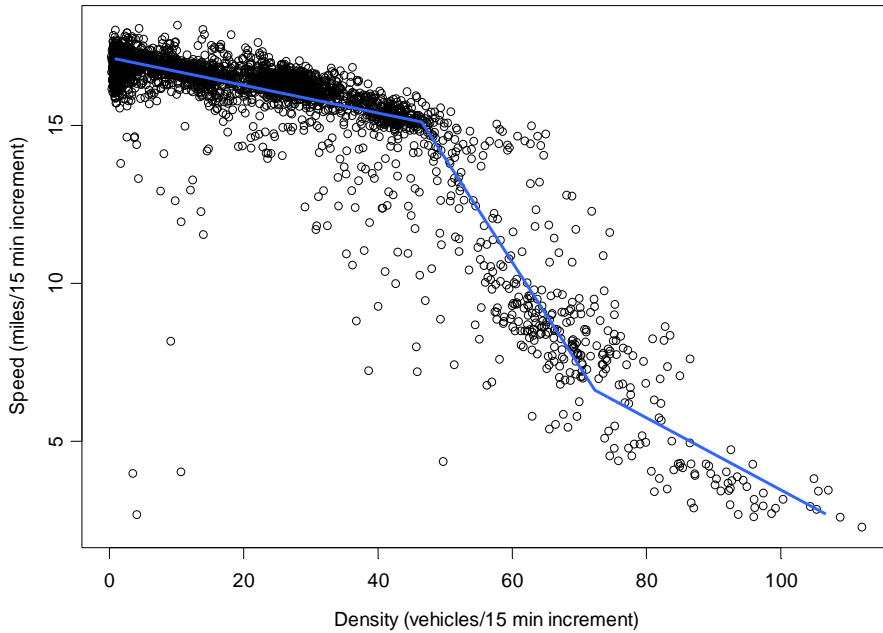


Figure 4.6 Speed-Density Model for one Detector

This three-part linear model was chosen because of the general shape of the raw data points. Each detector has a cluster of dense points approximating a downward sloping line that should be the first segment of the model. Then the graphs become a bit more subjective. Because the data varies so much for the middle segment of the graph, the slope and length of the second line for each graph is an approximation of best fit. However, a different line may be more appropriate given more data points or a different evaluator. The rest of the figures for this model can be found in Appendix C.

Detector 201581 has a strong line of points to model for the first segment of the line but a complete lack of data for a second or third line. This lack of points might mean that this location has the least congestion. Certainly the smaller volume measures support this hypothesis; however it still seems interesting that there are virtually no higher density points to model. Looking back at the time of day versus travel time and time of day

versus speed figures, it is apparent that Detector 201581 does have the least variation in speed and the smallest increases in travel time throughout the day of all of the detectors.

Detector 75585 also has a shape other than what might be expected in the three-part linear model. The first segment of the model is completely flat (slope = 0). This would mean that speed remains constant at low densities. While this idea might seem logical, it is a theory neither supported by the other detectors nor by previous research on density data. Though the segment is flat, it is also a much shorter segment than in the other graphs, so one might just assume that if a larger segment of the graph were taken into account for the first line, the new line would be downward sloping.

Upon further scrutiny of these graphs, it is not robustly clear that the three-part linear model is indeed best. However, the looseness of the data makes it challenging to determine the exact relationship for these models. A three-part model looks appropriate with this data, but as stated earlier, the data for the middle segment of the model is so widely spread that the modeling becomes quite subjective. Nonetheless, this study was able to, at least in part, validate previous research for both the speed-flow relationship and the speed-density relationship. A more sophisticated analysis might be able to further examine the data so as to develop a tighter relationship for these models.

5 Conclusion

It would be nice to know exactly when and where traffic congestion will pop up, before one gets trapped in the middle of miles of backup. New technologies that utilize real-time as well as historical data are making such up-to-the-minute information possible for all drivers, not just commuters during rush hour. However, though rush hour might seem to be the only predictable congestion, there are other trends in traffic data that emerge upon a closer scrutiny.

Using a smoothing regression to account for the main trends in the data while ignoring the infrequent outliers, this study was able to model and discover monthly trends

for the Atlanta detectors. For instance, this study discovered that the volume trend for December detector data shows that much more traffic volume is associated with the midday lunch hour than the morning or evening rush hour. Also, the holiday season in general causes more traffic. The data did not show a bimodal spike in volumes as was originally expected but rather an asymmetric increase during the daylight hours. Upon further inspection, the shapes of the volume measures for all the detectors could be verified given the location of detection.

Possibly the most important thing discovered from the attempted modeling of the data was the impracticality of modeling with such little data. It is nearly impossible to create a model for data week by week with only one data transmission every 15 minutes each day. Only by combining data for each week into monthly segments was it possible to create a kernel smoothing regression that could, with some certainty, be considered a valid and supported model for the month. However, more data would definitely have been beneficial for modeling and credibility purposes.

Lack of data was not a problem for the speed-flow data or the speed-density data; rather it was a lack of concrete, discernable trends. Because this work included research from prior experiments and discovered the relationships to look for in these graphs, it was not hard to find the fundamental relationships. However, had one looked at the graphs first, the relationships might not have been all that clear. By no means are the models developed here robust; rather they are meant to estimate the data and prove that, indeed, the relationships developed by prior researchers are supported by this data. It is only here, in these relationships, that one is easily able to pick out some errors in the data reporting. For instance, a flow rate of zero with a speed of 50 miles per hour is not

possible. Because one sees these errors in data reporting, he must assume that there is a potential for errors elsewhere in the data.

Nevertheless, this work was able to see the relationships for the Atlanta data. A stable upper flow curve and a lower, unstable lower flow curve were modeled for the speed-flow relationship for the detectors. Supposedly, the point that these curves approach but do not pass is the capacity level. However this analysis discovered widely varying capacity measures for the different roadways, an observation that may or may not be realistic given the consistency of lane use and road type. For the speed-density data, it was decided that the three-part linear model suited the data best. Though the first segment of the model fits the linear model well, the next two segments are not as tightly reported. Under ideal circumstances, one would have wanted the data to take on a more defining shape. However, once again, either more data or more accurately reported data might have helped define these models.

If one were to apply these techniques to another city or perform this analysis again with more data, he would first recognize that the choice of the detector location is very influential on the type of volume measure obtained. Because of the way the GDOT collects data, lane-by-lane for each direction, the choice of detector lane and direction greatly changed the shape of the graph. Had detectors been chosen at the same locations but going in the opposite direction, one might have similarly shaped graphs but just reversed (with the morning peak at the evening peak and vice versa). Also, the choice of the city would greatly influence the results. By choosing another city subject to much urban sprawl, the results would be fairly comparable. However, a city without a

concentrated area of employers and a wider suburban community might not experience quite the congestion.

Though this study had its limitations, it was a valuable endeavor to compile information regarding the predictable nature of volume data and also reveal that empirical data for the fundamental relationships of speed, flow, and density are not nearly as obvious as theory might indicate. A further study of volume and congestion trends using more data would advance the analysis started in this independent work and solidify the initial trends, meaning a better estimate of traffic congestion trends and more comprehensive, robust models.

Bibliography

Carmona, Rene. Statistical Analysis of Financial Data in S-Plus. New York: Springer-Verlag, 2004.

Demidovich, Mark. "RE: More info." Email to Megan Bernard. 21 March 2005.

Drake, J., J. Schofer, and A. May. "A Statistical Analysis of Speed-Density Hypotheses." Vehicular Traffic Science. New York: American Elsevier Publishing Co., Inc., 1967.

Gavalas, Vicki, ed. "Fact Book 2004." Georgia Department of Transportation. 20 March 2005 <<http://www.dot.state.ga.us/dot/communications/publications/documents/pdf/factbook/factbook-2004>>.

Georgia Quick Facts from the U.S. Census Bureau. 1 Feb. 2005. U.S. Census Bureau. 20 March 2005 <<http://quickfacts.census.gov/qfd/states/13000.html>>.

Highway Capacity Manual (Special Report 209). Washington D.C.: Library of Congress, 1985.

Leutzbach, Wilhelm. Introduction to the Theory of Traffic Flow. New York: Springer-Verlag, 1988.

Mann, Brad. "RE: Data Streams for Metro-Atlanta Interstates." Email to Megan Bernard. 31 Jan. 2005.

Martin, Dr. Peter T., Yugi Feng, Xiaodong Wang. "Detector Technology Evaluation." Nov. 2003. University of Utah Traffic Lab. 22 March 2005 <http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/pdf/MPC03-154.pdf>.

"NaviGator Real-time Traffic Map." 5 April 2005. Online image. Georgia Navigator Traffic Maps. 5 April 2005. <<http://www.georgia-navigator.com/maps/atlanta>>.

Oppenlander, J. C. "Variables Influencing Spot-Speed Characteristics." Highway Research Board. 1966.

Peek Traffic. "VideoTrak – 905." 25 March 2005 <<http://www.its-traffic.com/videotrak905.htm>>.

Sauve, Carole. "Exploring the Auto Culture." 15 April 1996. <<http://www.flora.org/afo/autocult.html>>.

Schrader, Christopher. "Reacting in Real Time: Using Historical and Real-Time Information in Forecasting Link Travel Times." Senior Thesis submitted April 14, 2003.

Sen, Ashish. et al. "Short-term Forecasting of Link Travel Times: A Preliminary Proposal." URBAN TRANSPORTATION CENTER, University of Illinois at Chicago. ADVANCE WORKING PAPER SERIES, Number 7. November 1991. <<http://ais.its-program.anl.gov/advance/reports/REPORTS.HTML/wp7/wp7.html>>.

Trafigon Technologies. "Trafigon – Principles." 25 March 2005 <<http://www.trafigon.com/solutions/technology/principles.jsp>>.

Turner, Shawn M., et al. Travel Time Data Collection Handbook. Report No. FHWA-PL-98-035. March 1998. Federal Highway Administration. Texas Transportation Institute. National Technical Information Service. <<http://www.fhwa.dot.gov/ohim/tvtw/natmec/00020.pdf>>.

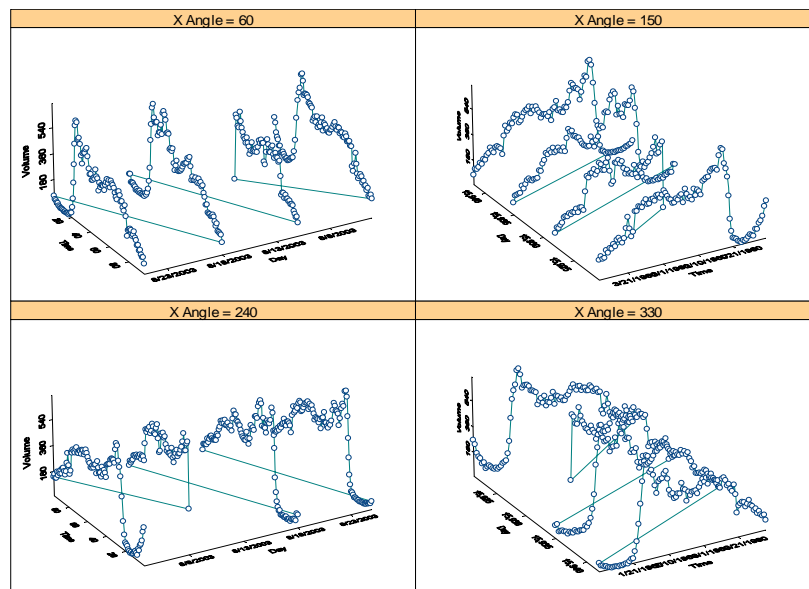
Appendix A

These graphs along with Figure 3.3 were made using the 3-D graphing function in SPlus. They are rotational views of the data for day versus time versus volume for the four Atlanta detectors for the month of August.

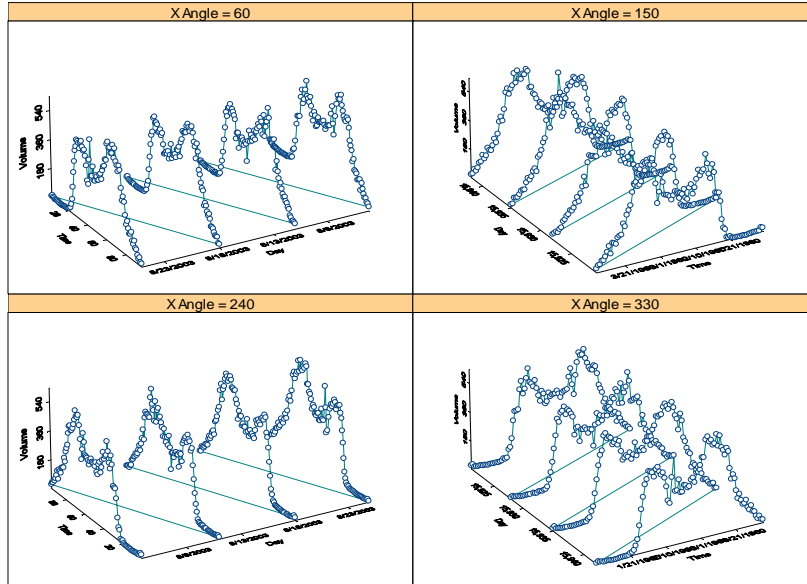
The code used to create the graphs:

```
plot(D714.2[,4],D714.2[,5],main="Empirical Speed-Flow Plot for Detector  
714",xlab="Flow Rate (vehicles per 15 minutes)",ylab="Speed (mph)")  
plot(D75585.2[,4],D75585.2[,5],main="Empirical Speed-Flow Plot for Detector  
75585",xlab="Flow Rate (vehicles per 15 minutes)",ylab="Speed (mph)")  
plot(D201581.2[,4],D201581.2[,5],main="Empirical Speed-Flow Plot for Detector  
201581",xlab="Flow Rate (vehicles per 15 minutes)",ylab="Speed (mph)")  
plot(D285292.2[,4],D285292.2[,5],main="Empirical Speed-Flow Plot for Detector  
285292",xlab="Flow Rate (vehicles per 15 minutes)",ylab="Speed (mph)")
```

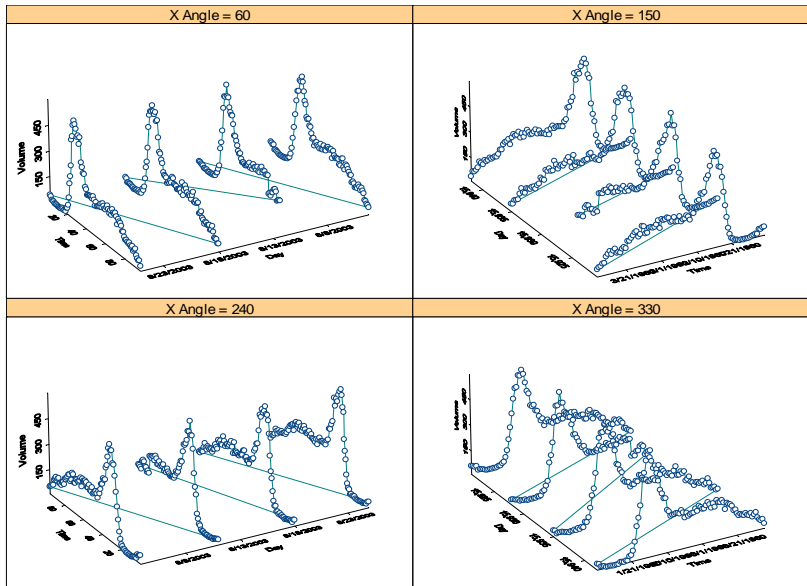
Detector 714



Detector 285292



Detector 201581



Appendix B

These graphs along with Figure 4.4 and Figure 4.5 were made using the plot function in SPlus. They represent speed-flow data for the entire year for each of the four detectors. One must pay special attention to the scale of the graphs as the scale changes for each graph.

The code used to create such graphs (the models are just approximations):

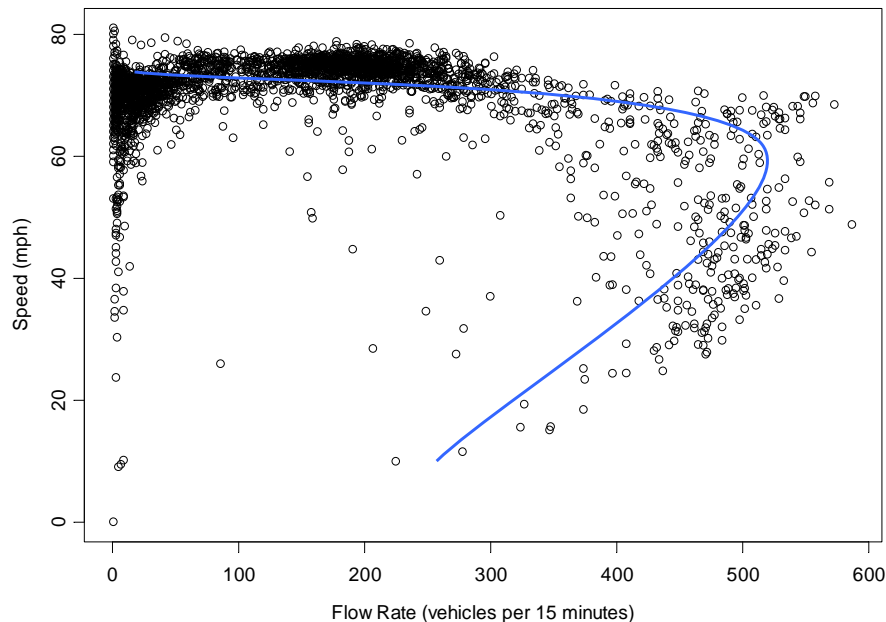
```
Density= D714.2[,4]/((D714.2[,5])/4)
plot (Density, D714.2[,5]/4, xlab="Density (vehicles/15 min increment)",
      ylab="Speed (miles/15 min increment)",
      title(main="Empirical Speed-Density Plot for Detector 714"))

Density2= D285292.2[,4]/((D285292.2[,5])/4)
plot (Density2, D285292.2[,5]/4, xlab="Density (vehicles/15 min increment)",
      ylab="Speed (miles/15 min increment)",
      title(main="Empirical Speed-Density Plot for Detector 285292"))

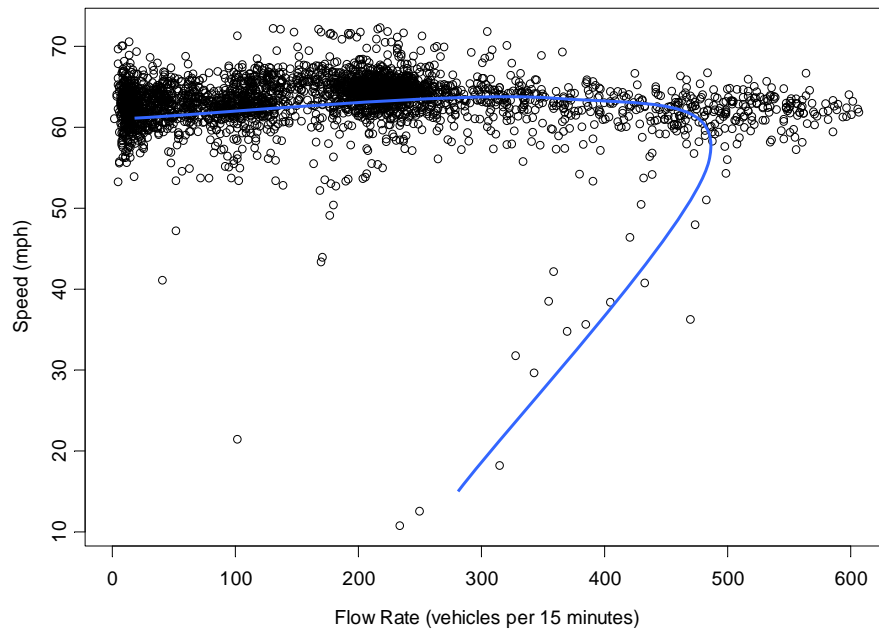
Density3= D201581.2[,4]/((D201581.2[,5])/4)
plot (Density3, D201581.2[,5]/4, xlab="Density (vehicles/15 min increment)",
      ylab="Speed (miles/15 min increment)",
      title(main="Empirical Speed-Density Plot for Detector 201581"))

Density4= D75585.3[,4]/((D75585.3[,5])/4)
plot (Density4, D75585.3[,5]/4, xlab="Density (vehicles/15 min increment)",
      ylab="Speed (miles/15 min increment)",
      title(main="Empirical Speed-Density Plot for Detector 75585"))
```

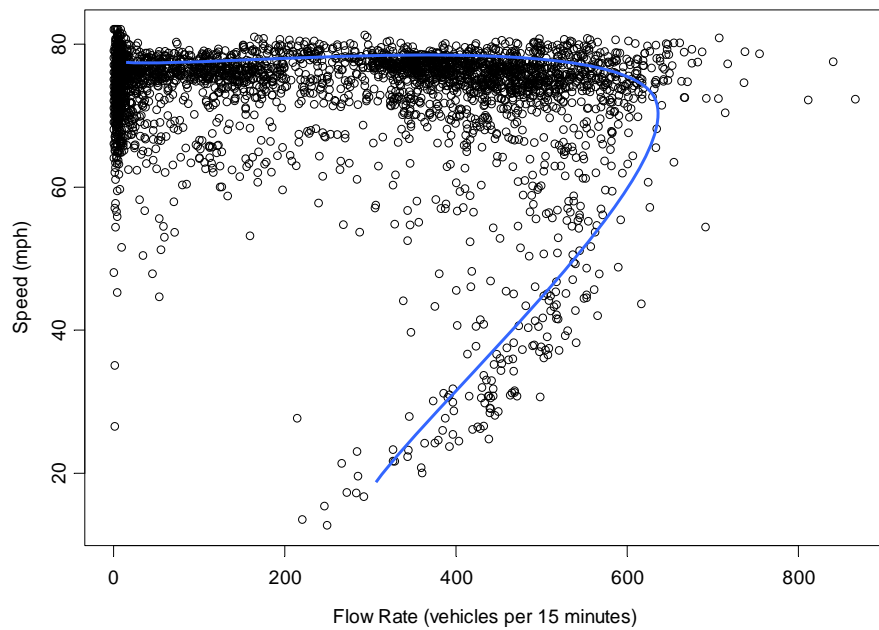
Empirical Speed-Flow Plot for Detector 75585



Empirical Speed-Flow Plot for Detector 201581



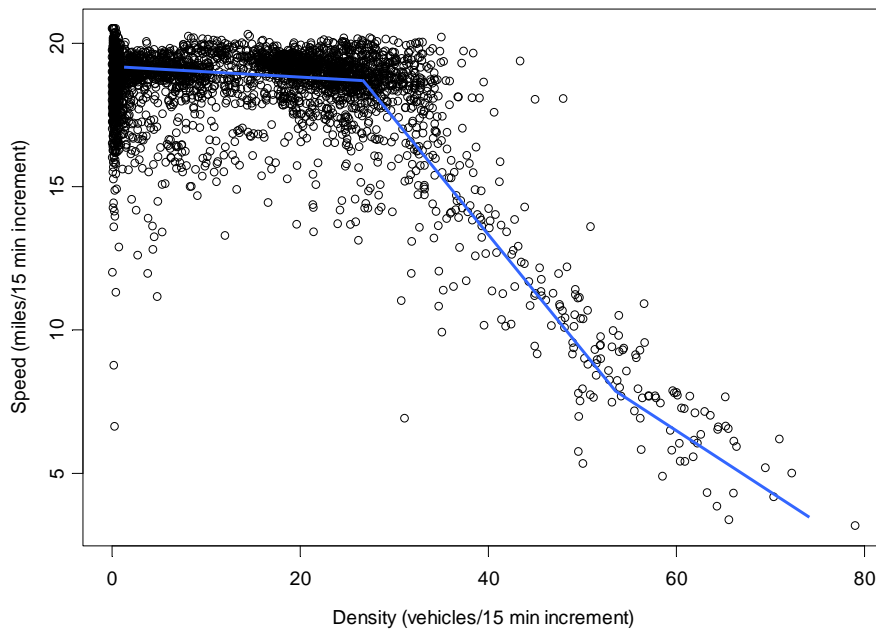
Empirical Speed-Flow Plot for Detector 285292



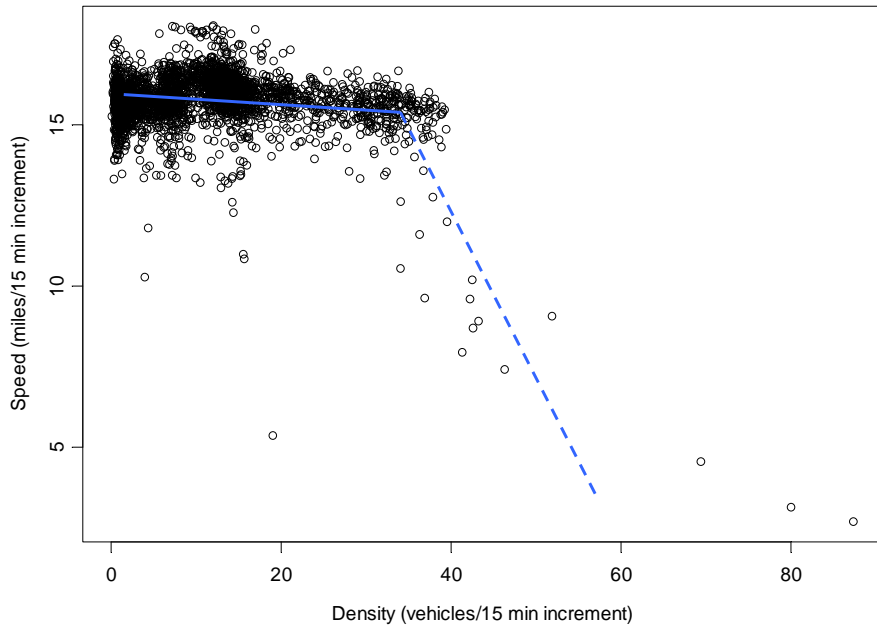
Appendix C

These graphs along with Figure 4.6 were made using the plot function in SPlus. They represent speed-flow data for the entire year for each of the four detectors. The three-part linear models are subjective based upon the writer's judgment; however the choice of a three-linear model over other options is fairly clear.

Empirical Speed-Density Plot for Detector 285292



Empirical Speed-Density Plot for Detector 201581



Empirical Speed-Density Plot for Detector 75585

