A METHODOLOGY FOR DERIVING PERFORMANCE MEASURES FROM SPATIO-TEMPORAL TRAFFIC CONTOUR MAPS USING DIGITAL IMAGE ANALYSIS PROCEDURES

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
In partial fulfillment of the
Requirements for the degree of
Master of Science in
Civil Engineering

in

The Department of
Civil and Environmental Engineering

by
Prashanth Kotha
B.Tech, Kakatiya University, 2001
December 2003

Acknowledgements

I would like to thank my advisor Dr. Sherif Ishak for his guidance, timely comments, support and patience throughout the course of this work. I thank Dr. Chestor Wilmot and Dr. Brian Wolshon for being on my defense committee. I thank all my friends who have helped me in the successful completion of this work. Last but not least, I thank my parents for their support and encouragement

This research is supported by the National Science Foundation (NSF) through grant No. CMS/0230216 to Louisiana State University. The research award was granted under a joint program between NSF and USDOT on information and communication systems for surface transportation (ICSST). The opinions, findings, and conclusions in this study are those of the authors' and not necessarily those of the NSF or USDOT. This study does not constitute a standard, specification, or regulation.

Table of Contents

List of Tables

List of Figures

Abstract

The main focus of this study is to improve the data analysis tools used in performance monitoring and level of service assessment of freeway systems. The proposed study presents a methodology to develop new second-order statistical measures that are derived from texture characterization techniques in the field of digital image analysis. The new measures are capable of extracting properties such as smoothness, homogeneity, regularity, and randomness in traffic behavior from the spatio-temporal traffic contour maps. To study the new performance measures a total of 14270, 15-min traffic contour maps were generated for a section of 3.4 miles of I-4 in Orlando, Florida for 24 hours over a period of 5 weekdays. A correlation matrix was examined using the obtained measures for all the constructed maps, which is used to check for information redundancy. This resulted in retaining a set of three second-order statistical measures: angular second moment (*ASM*), contrast (*CON*), and entropy (*ENT*). The retained measures were analyzed to examine their sensitivity to various traffic conditions, expressed by the overall mean speed of each contour map. The measures were also used to evaluate level of service for each contour map. The sensitivity analysis and level of service criteria can be implemented in real time using a stand-alone module that was developed in this study. The study also presents a methodology to compare the traffic characteristics of various congested conditions. To examine the congestion characteristics, a total of 10,290 traffic contour maps were generated from a 7.5-mile section of the freeway for a period of 5 weekdays.

1 Introduction

1.1 General

This research study is motivated by the need to improve the existing data analysis and presentation tools used to evaluate the performance of our surface transportation system (McNeil et al. 2000). A growing number of urban freeways are currently instrumented with advanced traffic monitoring systems where large amounts of spatial and temporal data, sometimes in the order of a few gigabytes per day, are collected in real- time and archived into traffic data warehouses. This massive amount of data is yet to be fully utilized primarily because of the lack of appropriate data analysis tools. Despite the tremendous advancements in transportation data collection technologies, no parallel efforts have been made to improve the existing performance measures and maximize the utility of information extraction methods from both archived and real-time transportation data. Current measures in the form of point estimates of average travel time and delay are essential properties derived from first-order statistics and do not adequately reveal certain local and system-wide properties such as smoothness, coarseness, regularity, homogeneity, entropy, and others. Such properties can only be revealed by second-order statistics and via other advanced procedures that are capable of characterizing the quality of traffic operations by exploiting spatiotemporal dependencies of constructed spatiotemporal traffic contour maps, a concept that is similar to texture characterization of digital images.

This study is aimed at addressing the needs of two communities: freeway management agencies and academic researchers. On one hand, traffic management centers seek new opportunities to improve their real-time operation and management functions and to advance the methods used to assess the impact of minor/major capital improvements. On the other hand, researchers seek better capabilities to discover more of the behavior of traffic under non-stationary transient stages, to identify certain factors or conditions that may impact safety, to distinguish between the traffic characteristics during recurrent and non-recurrent conditions, and to develop comprehensive and composite measures of the level of service at both local and global scales. The role of information and communication technologies has been fundamental to this research study due to the extensive data requirements of the approach. In the absence of highway traffic monitoring systems, it would be virtually impossible to consider an approach that is contingent upon the availability of real-time, high resolution data.

1.2 Problem Statement

The concept of Intelligent Transportation Systems (ITS) has introduced an array of technological components into aspects of planning, operation, maintenance, safety, and design of our surface transportation system. As the backbone of ITS, information and communication technologies have substantially overcome the data acquisition obstacles through a wide spectrum of advanced data collection and communication devices. This has tremendously increased our ability to manage and control major transportation system facilities in real-time. Most of the ITS implementation efforts were intended for major urban freeways to provide effective solutions to the perpetually escalating problems of urban congestion. Currently, several hundreds of freeway miles are instrumented with traffic surveillance devices, all of which are primarily installed to achieve following goal; that is to improve the operation, safety, and productivity of our freeway system.

As the ITS instrumentation efforts continue to pervade our urban freeway system nationwide, real-time data from hundreds of miles is simply being either archived or disposed off. Currently, little information has been utilized from the considerable amount of information that can be extracted from this data. This is primarily due to the lack of advanced data mining methods that are specifically developed to maximize the utility of information from existing surveillance systems. Such methods must be capable of manipulating large amounts of data for the purpose of extracting the most useful information and removing information redundancies. Essentially, the need for new methods was not justified in the past when the transportation data acquisition process, in the absence of today's technology, was quite a challenge. Today, we face another challenge; that is how to process the overabundance of data in order to extract the most useful information and maximize the efficiency of Traffic Management Centers (TMC) operations and management functions. This can only be achieved through advanced analytical techniques and performance measures that are capable of extracting the most valuable information from real-time and archived data while reducing the overwhelming level of redundancies therein.

Traffic management and control functions undertaken by TMCs and other related agencies are performed through a series of sequential and parallel tasks that involve processing information from collected data and then making decisions, if necessary. Today, three major challenges continue to face researchers and practitioners in this field. First, how can we extract the most relevant information that supports the operational functionalities of TMCs and other interested transportation agencies, given the vast amount of data and despite the tremendous redundancies therein? Second, how can we utilize the extracted information to advance our level of understanding of the traffic characteristics under a wide spectrum of operating conditions? Third, how can we use the extracted features to advance our level of understanding of traffic behavior in the context of other related applications, such as accident analysis and incident detection? This study addresses the first question using feature extraction techniques and texture characterization methods from the field of digital image analysis to advance the data analysis and presentation tools in the field of traffic operations and performance monitoring systems.

1.3 Study Objectives

The ultimate goal of this study is to improve freeway data analysis and presentation tools by accomplishing the following objectives:

1. Develop a new set of performance measures by quantifying special features of spatiotemporal traffic contour maps using feature extraction and textural characterization techniques
2. Develop an on-line module for performance assessment of extended time periods and long freeway segments.
3. Study and compare the operational characteristics of different congestion periods using newly developed performance measures.

2 Literature Review

The proposed research study explores innovative procedures from the field of digital image analysis to analyze and extract special properties that can be used to measure the operational performance of freeway systems. As the proposed research is multi-disciplinary in nature, this chapter is divided into four sections. The first section gives a detailed description of conventional performance measures that are used to assess the performance of transportation facilities. The second section introduces the concept of traffic contour maps. In the third section, digital images and their applications are discussed followed by an introduction to basic tools and techniques used in digital image analysis in the fourth section.

2.1 Conventional Performance Measures

Conventional performance measures are selected to answer a few basic questions (Roess et al. 1998). For a highway facility, the questions are "how good was my trip?", "how many of us did the system serve?", "how well did the system move us?", "how much of the system resources were used up in the process?", and "are there any deficiencies in the infrastructure?" The common measures used to answer the previous questions include travel time, speed, delay, volume, demand to capacity ratio, congestion extent and duration, and others. These first-order statistical measures are used individually or collectively to evaluate the level of service on highway facilities (Banks 2002; Gaber and Hoel 1999; HCM 2000; Khisty and Lall 1990; May 1990; Papacostas and Prevedouros 2001; Roess et al. 1998 and Wright 1996). While these measures reveal common properties of the system operation, several other properties go undetected, and therefore, are the focus of this research study. A recent study by Dahlgren et al.

(2002) examined the utilization of archived operations data collected by TMCs to measure and improve freeway performance. This study emphasized that better utilization of archived data would take time and experimentation. So, there is an immediate need to find new methods and measures to use the collected data to the fullest extent. This study also concluded that the most of the TMCs are not utilizing the archived data properly to improve their operational efficiency; they, in turn, see a little connection between historical archived data and the crisis they manage on a day-to-day basis. The authors also concluded that the key to effective data archiving and disseminating is responsiveness and adequate funding of the appropriate organization, which is collecting the data.

Another study by Choe et al. (2002) on freeway performance in the state of California, presented a comprehensive system of freeway performance monitoring. The study was one of a few that emphasized the importance of using operational analysis tools in the context of PeMS (Freeway Performance Measurement System) in California. The study presented a system that is capable of providing traffic management centers with uniform and comprehensive assessment of freeway performance in real-time. The authors showed how the system can benefit several entities in transportation, including both transportation system users and providers (planners, operators, etc.). The study emphasized a few applications such as freeway operational analysis, bottleneck identification and analysis, level of service characterization, incident impacts, and assessment of Advanced Traffic Management and Information Systems (ATMIS) strategies.

In another study by Bertini et al. (2002) certain performance measures were generated for a freeway corridor in Portland, Oregon. The study used the archived data to determine the functionality of the facility with respect to measures such as mobility, economic development, quality of life, the environment, resource conservation, and safety. The study showed that by using real time data from loop detectors, it is possible to obtain information about the functionality of a facility by developing performance measures and tracking them over time. This information can in turn, help the transportation agencies to have a better vision of the current performance of the transportation network, its evolution over time, and also help setting objectives to improve the performance of the facility.

Another study by Kwon et al. (2000) attempted to capture special features and trends in travel times from day to day. The study attempted to predict travel times solely based on loop detector data. The study concluded that the current traffic conditions are good in predicting travel times up to 20 minutes in advance, while historical data is more useful in predicting travel times for longer range time periods. The study highlighted some issues related to feature extraction and proposed an unusual measure for each day by comparing it with the average.

A paper by Brydia et al. (1998) dealt with the development of ITS data management system (ITS DataLink) that is used to store, analyze, and present data from the TransGuide center in San Antonio, Texas. A lot of traffic data is being collected through ITS components by TMCs. This data is simply discarded after their use, and nothing is been done to share this data among other transportation groups or agencies within the same jurisdiction. On the other hand, transportation analysts and researchers often struggle to obtain accurate real-time data, which they use for various purposes such as calibration of new transportation models, and decision making etc. This particular paper addresses this issue by studying various data collection, archiving, and sharing techniques. This study also looks into issues associated with storage, aggregation and analysis of the collected data, and computer hardware and software requirements for its management.

In another paper by Banks (1998) studies were conducted regarding the needs, opportunities, and techniques for measuring the performance of the California Department of Transportation (Caltrans) units involved in traffic management for urban freeways. The study concluded that successful performance measurement for traffic management system requires both physical and institutional changes. Case studies on two Caltrans districts showed that these districts are ready to provide a physical infrastructure adequate to quantify major traffic management system performance measures once it is fully deployed. However, neither district has the staffing and organizational structure to perform activities such as data collection, performance monitoring, evaluation studies, and traffic data quality control. Consequently, successful performance measurement requires significant institutional changes.

A study by Papiernik et al. (2000) presented a case study based on the enterprise data warehouse and Programming and Scheduling (P&S) Data Mart that is being developed and implemented for the Virginia Department of Transportation (VDOT). The study explicitly described how in one division of the VDOT, P&S will be benefited by investing in Information Technology (IT) to achieve its strategic goals. The study also details the design approach, methodology, and implementation procedure for the P&S

decision to support data mart. The study also highlights the methodology for capturing the performance measures that have been defined by the P&S division in context of its strategic outcome areas. The study concluded that the system should utilize web-based technologies to store, organize, and share the data so that the end users can utilize this data for decision support and analytical processing.

In a recent study by Shaw et al. (2003) presented a technique for measuring, estimating and reporting reliability performance measures at the highway segment, facility, and system level. The study emphasized that with congestion levels increasing, and limitations and restrictions imposed on new constructions, there is an urgent need to look beyond existing level of service to assess the performance of a transportation system. The study concludes that better understanding of the reliability performance measure can be useful for transportation planners and traffic engineers to provide accurate information to travelers through ITS.

From the literature it is clear that a lot of research has been done in the field of data collection, archiving, and dissemination technologies but very little has been done in utilizing the collected data to fullest extent. Currently, most of the collected data is simply either stored or discarded. This research study presents a new technique to develop performance measures for freeway systems that are instrumented with traffic monitoring devices. Preliminary results have been published in Ishak (2003).

2.2 The Concept of Traffic Contour Maps

The latest edition of the Highway Capacity Manual (2000) describes the use of time-space domain traffic contour maps in performance measures of freeway facilities. Contour maps are used to identify valleys of lower-speed or high-density regions. In essence, contour maps can simply be considered pictorial representations of traffic conditions in time and space. They are suitable for an overall assessment of the entire freeway facility over the study period. This concept was introduced in the early 1990s and was followed by a surge of traffic modeling efforts using pattern recognition tools such as artificial neural networks. In such research efforts, traffic conditions were primarily represented as spatiotemporal patterns. Examples are found in the area of automated incident detection systems (Hsiao et al. 1994, Ishak and Al-Deek 1998, Ishak and Al-Deek 1999, Ritchie and Ruey 1993, Ruey and Ritchie 1995), short-term traffic predictions (Abdulhai et al. 1999, Park and Rilett 1998, Rilett and Park 1999), and many others. With this concept in mind, spatio-temporal traffic contour maps can be analyzed using techniques developed for digital image analysis. However, the feature extraction process in the context of the study is not used for pattern recognition application, but rather for revealing traffic characteristics of the constructed spatio-temporal traffic contour maps. A detailed description of traffic contour maps is given in section 4.2 of the methodology chapter.

2.3 Digital Images and Applications

2.3.1 Digital Image

In this section the definition of a digital image, its fields of application, its representation, and different processes of digital image analysis will be addressed. An image is defined as a two-dimensional function f (x, y), where x and y are spatial co-ordinates. The amplitude of 'f' at any pair of co-ordinates (x, y) is called the intensity or gray level of the image at that point (Gonzales and Woods 2001). When x, y and amplitude values of 'f' are all finite and, discrete quantities, the image is called a digital image. i.e., a digital image is composed of a finite number of elements, each with a particular location and value. These elements of the image are called pixels. In short a computer or digital image is nothing but a matrix of a two-dimensional array of pixels, where the value of each pixel is proportional to the brightness of the corresponding point in the image (Nixon and Aguado 2002).

2.3.2 Applications

The areas of applications of digital image analysis are so broad that their uses are categorized according to the sources of images such as gamma, x-ray, synthetic etc. This section gives different applications of digital image analysis categorized by the image source. Applications of imaging based on gamma rays include nuclear medicine and astronomical observations. X-ray images are used in medical diagnostics and also used extensively in industry and other areas like astronomy. Imaging in the ultraviolet band is used in lithography, industrial inspection, microscopy, lasers, biological imaging, and astronomical observations. Applications in the visual and infrared band out-weight by far all other image sources in terms of scope of application. Their applications include light microscopy, astronomy, remote sensing, industry, and law enforcement. Imaging in the microwave band is used in radar fields. The unique feature about imaging radar is its ability to collect data over virtually any region at any time, regardless of weather or ambient lightning conditions. Radio band image applications include the medical field and astronomy.

Apart from the electromagnetic spectrum image applications that are discussed above, there are a number of other image modalities such as Acoustic imaging, Electron microscopy, and synthetic (computer generated) imaging. Acoustic imaging is applied in geological exploration, industry, and medical fields. Fractals are the striking examples of computer-generated images. A fractal is an iterative reproduction of a basic pattern according to some mathematical rules (Gonzales and Woods 2001). This is one of the methods to analyze the texture of an image.

2.4 Digital Image Analysis Tools

In the field of digital image analysis, feature extraction and textural characterization can be performed using one of the three approaches: structural, statistical, or a combination of both approaches. The Structural approach is the most basic approach to texture description, which generates Fourier transforms from images and then groups the transform data in some way so as to obtain a set of measurements such as entropy, energy, and inertia (Nixon and Aguado 2002). Other structural approaches include Wavelet transform and Gabor Wavelet transform.

Another approach that is common among texture characterization techniques is based on a statistical approach that requires building a co-occurrence matrix from pair-wise pixel information at different distances and relative inclination. The co-occurrence matrix measures spatial relationships, as opposed to frequency contents in structural approaches. Measures such as smoothness, roughness, homogeneity, randomness and others can be derived from the co-occurrence matrices (Nixon and Aguado 2002, and Theodoridis and Koutroumbas 1999). The last approach is a combination of structural and statistical approaches. This approach, often referred to as statistical geometric features (SGF), combines geometrical structures with statistical ones and leads to measures of irregularity (Nixon and Aguado 2002). A detailed description of the measures used in this study is given in section 4.3 of the methodology chapter.

3 Data Collection

The data used in this study is collected from a 40-mile stretch of I-4, in Orlando, Florida for a period of five weekdays from April 2nd, 2001 to April 6th, 2001 for 24 hours in both east bound and west bound directions. Three traffic parameters (speed, volume, and lane occupancy) data are collected from a total of 70 dual detector stations that are spaced approximately 0.5 miles apart and stream data every 30 seconds to the Orlando Regional Traffic Management Center (RTMC). In addition to the loop detectors, the I-4 corridor is instrumented with Closed Circuit TV cameras (CCTV) and Changeable Message Signs (CMS) in each direction along the I-4 corridor. FIGURE 3-1 shows the 40-mile corridor of I-4 that extends from west of US 192 to east of Lake Mary Blvd. TABLE 3-1 shows the description and location of each detector station.

The real-time and archived data is accessible through the Internet at http://trafficinfo.engr.ucf.edu/i4traffic/realtime/speedMap.asp. The loop detector data is collected in real time via a T1 link between the Orlando RTMC and the ITS lab at the University of Central Florida. Speed, volume, and lane occupancy data is downloaded and compiled into an SQL server that supports multiple publicly accessible web applications such as real time and short-term travel time predictions between user-selected on- and off-ramps. FIGURE 3-2 shows a sample of speed data recorded on August 28th, 2002 between 7:00 am and 7:15 am at loop detector station 38 (Church St.) using 1-minute aggregated data.

The I-4 corridor is composed of three lanes in each direction. At each detector station on the mainline, there are two 6'x 6' loops embedded in each lane and connected to a 170-type controller located in a cabinet adjacent to the roadside. FIGURE 3-3 displays the configuration of a typical loop detector station in one direction of travel. The dual loops in each lane are used to measure the vehicle speed as the distance between the two loops divided by the difference between the two detectors' actuation times.

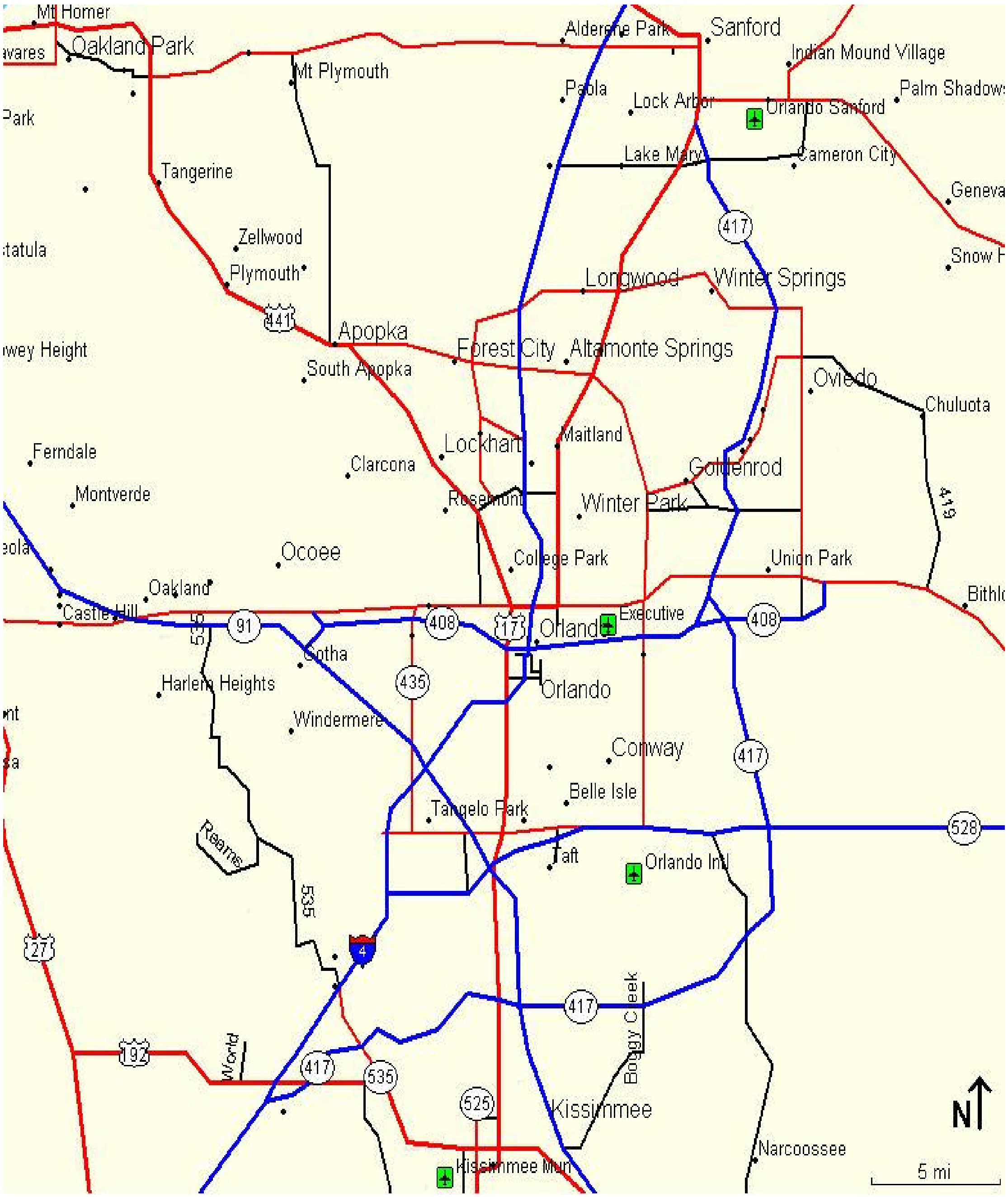


FIGURE 3-1: MAP OF THE I-4 STUDY CORRIDOR IN ORLANDO, FLORIDA.

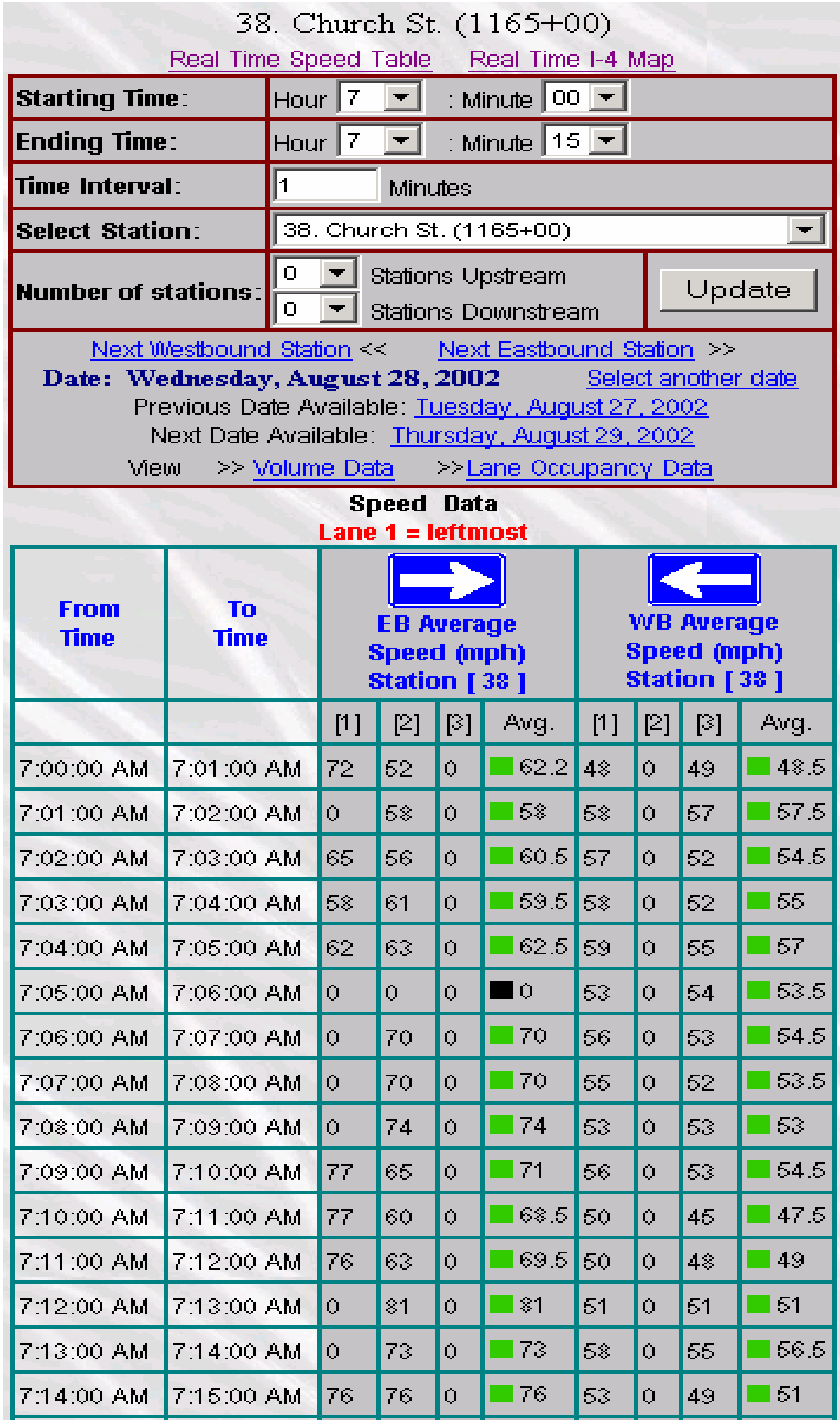
Occupancy and volumes are directly measured with a single loop detector. Therefore, in each direction, each detector station reports three values of speed and six values of

TABLE 3-1: LOCATION OF LOOP DETECTOR STATIONS ON THE 40-MILE CORRIDOR OF I-4 IN ORLANDO, FLORIDA

From Station	To Station	Location	Spacing (feet)
1	2	West of 192	-
2	3	West of 192	2600
3	4	US 192	2470
4	5	West of Osceola	3300
5	6	East of Osceola	3530
6	7	SR 536	33.30
7	8	East of SR 536	3370
8	9	West of SR 535	3360
9	10	West of SR 535	3.40
10	11	SR 535	3000
11	12	West of Rest Area	3200
12	13	Rest Area	4090
13	14	West of Central Florida Pkwy	3020
14	15	Central Florida Pkwy	2980
15	16	528 EB Ramp	2910
16	17	528 WB Ramp	3250
17	18	West of 482	3100
18	19	West of 482	3450
19	20	SR 482	2000
20	21	West of 435	3100
21	22	West of 435	2600
22	23	SR 435	3000
23	24	435 WB Ramp	2900
24	25	Turnpike	2200
25	26	Turnpike WB Ramp	2900
26	27	Camera 21	26.10
27	28	West of John Young Pkwy	28.90
28	29	West of John Young Pkwy	2900
29	30	John Young Pkwy	4100
30	31	East of John Young Pkwy	2400
31	32	Rio Grande	2600
32	33	Orange Blossom Trail	2400
33	34	Michigan	2500
34	35	Kaley	2400

TABLE 3-1: (CONTINUED)

From Station	To Station	Location	Spacing (feet)
35	36	Camera 28	2700
36	37	Camera29	2700
37	38	Church St	1800
38	40	Robinson	3000
39	41	SR 50	2500
40	42	Ivanhoe	2600
41	43	Princeton	2700
42	44	Winter Pk	2600
43	45	Par Ave	2600
44	46	Minnesota	3000
45	47	SR 426	2200
46	48	Site 1393	2300
47	49	Lee Rd	2600
48	50	East of Lee Rd	1700
49	51	Kennedy	2800
50	52	414 EB Ramp	3000
51	53	East of SR 414	1800
52	54	Wymore	3300
53	55	East of Wymore	2700
54	56	West of SR 436	2900
55	57	SR 436	2400
56	58	West of SR 434	3800
57	59	West of SR 434	2900
58	60	SR 434	3500
59	61	434 Ent Ramp	3400
60	62	434 Ext Ramp	1900
61	63	West of EEWill	2800
62	64	East of EEWill	2600
63	65	Rest Area	3000
64	66	East of Rest Area	2700
65	67	West of Lake Mary Blvd	2100
66	68	West of Lake Mary Blvd	2500
67	69	Lake Mary	2800
68	70	Lake Mary	2300
69	71	East of Lake Mary Blvd	3500



38. Church St. (1165+00)

Real Time Speed Table   Real Time I-4 Map

Starting Time:	Hour 7 : Minute 00
Ending Time:	Hour 7 : Minute 15
Time Interval:	1 Minutes
Select Station:	38. Church St. (1165+00)
Number of stations:	0 Stations Upstream 0 Stations Downstream Update

Next Westbound Station <<   Next Eastbound Station >>

Date: Wednesday, August 28, 2002   Select another date

Previous Date Available: Tuesday, August 27, 2002

Next Date Available: Thursday, August 29, 2002

View >> Volume Data >> Lane Occupancy Data

Speed Data

Lane 1 = leftmost

From Time	To Time	EB Average Speed (mph) Station [38]				WB Average Speed (mph) Station [38]			
		[1]	[2]	[3]	Avg.	[1]	[2]	[3]	Avg.
7:00:00 AM	7:01:00 AM	72	52	0	62.2	48	0	49	48.5
7:01:00 AM	7:02:00 AM	0	58	0	58	58	0	57	57.5
7:02:00 AM	7:03:00 AM	65	56	0	60.5	57	0	52	54.5
7:03:00 AM	7:04:00 AM	58	61	0	59.5	58	0	52	55
7:04:00 AM	7:05:00 AM	62	63	0	62.5	59	0	55	57
7:05:00 AM	7:06:00 AM	0	0	0	0	53	0	54	53.5
7:06:00 AM	7:07:00 AM	0	70	0	70	56	0	53	54.5
7:07:00 AM	7:08:00 AM	0	70	0	70	55	0	52	53.5
7:08:00 AM	7:09:00 AM	0	74	0	74	53	0	53	53
7:09:00 AM	7:10:00 AM	77	65	0	71	56	0	53	54.5
7:10:00 AM	7:11:00 AM	77	60	0	68.5	50	0	45	47.5
7:11:00 AM	7:12:00 AM	76	63	0	69.5	50	0	48	49
7:12:00 AM	7:13:00 AM	0	81	0	81	51	0	51	51
7:13:00 AM	7:14:00 AM	0	73	0	73	58	0	55	56.5
7:14:00 AM	7:15:00 AM	76	76	0	76	53	0	49	51

FIGURE 3-2: SAMPLE OF SPEED DATA COLLECTED AT LOOP DETECTOR STATION 38 (CHURCH ST.) ON AUGUST 28TH, 2002 FROM 7:00 AM TO 7:15 AM.

Volume and lane occupancy every 30 seconds. Speeds and lane occupancies are expressed as the average of all vehicles with the 30-second period, while volumes represent the total vehicular count within the period.

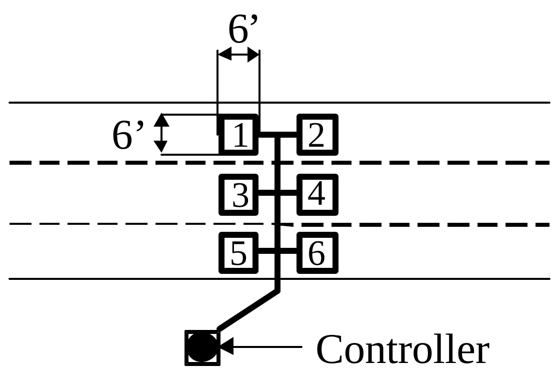


FIGURE 3-3: TYPICAL LOOP DETECTOR STATION.

3.1 Data Filtering

Prior to the analysis stage, the detector data must be filtered to remove abnormalities and smoothed to reduce noise and random fluctuations. In order to construct traffic contour maps we just need one speed value from each detector station. However, from the loop detectors we get three speed values for each lane. Therefore, the three speeds produced at a loop detector station should be reduced to one value. Also we will get an error or negative value if the loop detector station or communication infrastructure between the loops and the TMC is down. To account for above stated reasons, we need to filter the data obtained from the loop detectors. As we are using only speed data in this study, the procedure adopted to filter speed data is given below.

Each loop detector station produces three (one per each lane) speed values per time slice. This data is filtered as follows:

- If all the three values are positive numbers, then average of all the three values is taken to obtain a single value for the station.
- If any two values are positive numbers, then average of these two values is taken to obtain a single value for the station.
- If any one of the values is positive number, then that value is taken as speed value for that station.
- If none of the values is a positive number, then that station is down. If this station is not at the beginning or ending of the study section, then take the average of upstream and downstream values to obtain a single value for the station. If the station is at the beginning or ending of the study section, then use the speed value at the upstream station if the last station is down, and use the downstream station if the first station is down.

3.2 Summary

This chapter presented information on study section used and data parameters collected in this research. The traffic data was collected using a dual loop detectors placed along the study section. This chapter also discussed the data filtering techniques, which make the data usable. The filtered data is used to conduct further analysis as given in the subsequent chapters.

4 Methodology

The proposed approach in this study seeks to identify a viable set of measures that are capable of extracting the most important features from spatiotemporal traffic contour maps in a manner similar to extracting features and characterizing textures of digital images in the field of image analysis and pattern recognition. Several references were cited in the course of implementing this study and were used to provide an overview of the most common techniques in the field of image analysis (Micheli-Tzanakou 2000, Nixon and Aguado 2002, Seul et al. 2000, and Theodoridis and Koutroumbas 1999).

As explained earlier, textural description of digital images can be addressed using one of three approaches: structural, statistical, and a combination of both. The structural measures can only be used for images that have periodic texture patterns, i.e. images with a certain texture repeating, which is not applicable to traffic contour maps. Traffic contour maps do not exhibit periodic textural pattern as the traffic on a freeway is not periodic, i.e. the characteristics of traffic will not be the same at two different time intervals in a given day. Therefore, only the statistical approach was used in this study to analyze the traffic contour maps.

4.1 Representation of a Digital Image

As mentioned earlier a digital image is a matrix of a two-dimensional array of pixels with the value of each pixel proportional to the brightness of corresponding point in the image. Assuming an image $f(x, y)$ is sampled so that the resulting image has M rows and N columns. With this sense a digital image can be represented as follows:

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \ldots\ldots a_{0,N-1} \\ a_{1,0} & a_{1,1} & a_{1,2} & \ldots\ldots a_{1,,N-1} \\ a_{2,0} & a_{2,1} & a_{2,2} & \cdots\cdots a_{2,N-1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M-1,0} & a_{M-1,1} & a_{M-1,2} & \cdots a_{M-1,N-1} \end{bmatrix}$$

Where $a_{i,\,j} = f\,(x=i,\, y=j) = f\,(i,\, j)$, i.e. $a_{i,j}$ is the intensity of the pixel located at point $x=i$ *and* $y=j$.

The image represented above has M rows and N columns, which means it has a spatial resolution of size M X N with intensity of pixels varying from 0 to L levels i.e. the image has a gray level intensity of L levels. Due to computer hardware considerations the number of gray levels is usually an integer power of 2, the most common exponential number is 8 bits. 8 bits means the image has 2^8 (256) gray levels varying from 0 to 255.

4.2 Spatio-temporal Traffic Contour Maps

A spatio-temporal traffic contour map is constructed with point measurements that are observed over time and space. FIGURE 4-1 shows the anatomy of a typical spatio-temporal traffic contour map. The figure shows a grid of pixels, each having an intensity $I\,(t,x)$, where x and t represent the spatial and temporal dimensions, respectively. The intensity level $I\,(t,x)$ is arbitrarily expressed in terms of one of the three traffic parameters: speed, flow, or lane occupancy. An example of a speed contour map is shown in FIGURE 4-2, which is constructed for a 4-mile section of I-4 extended from loop detector station 30 to station 37 over a 2-hour period. The contour map is drawn with speed data which is updated every minute. The spatial resolution of this speed contour map is 120 X 9 with the gray level intensity varying from 0 to 70, because the maximum speed is set to 70 mph and minimum to zero. The traffic contour maps can be constructed for the entire section of freeway under study or a specific section of interest.

Also, maps can be constructed for different time periods such as 15 min, 20 min, 1hour, etc. The traffic contour maps constructed with small spatiotemporal dimensions are more appropriate for revealing local traffic characteristics, while those constructed with large spatiotemporal dimensions are more suitable for an overall system-wide assessment of traffic performance.

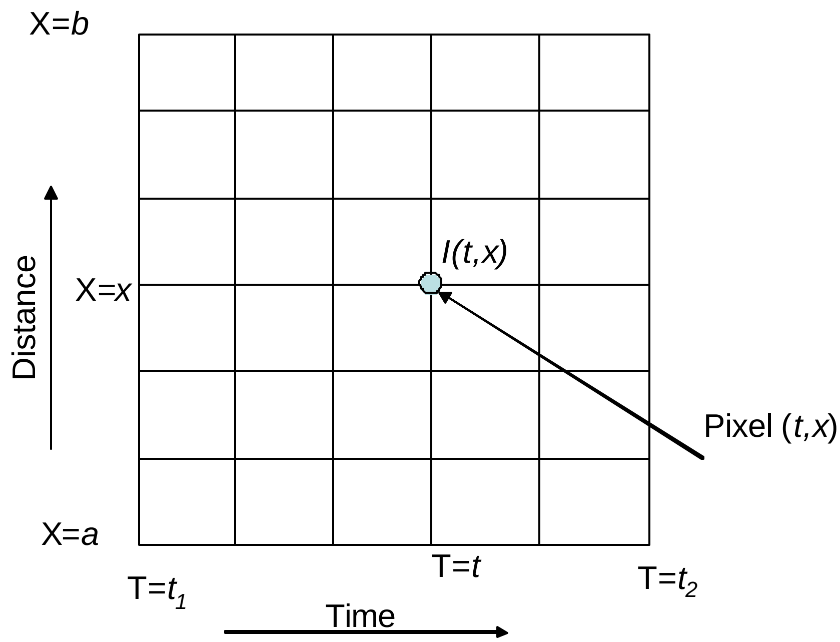


FIGURE 4-1: SPATIOTEMPORAL TRAFFIC CONTOUR LAYOUT REPRESENTED BY TRAFFIC PARAMETER I.

4.3 Statistical Approach

The statistical approach used in the study will depend on the construction of the co-occurrence matrices for relative distances and relative orientations between a point and all of its surroundings in the time-space domain of a traffic contour map as shown in FIGURE 4-3. The figure shows that the relative differences between each point *(t, x)* and its

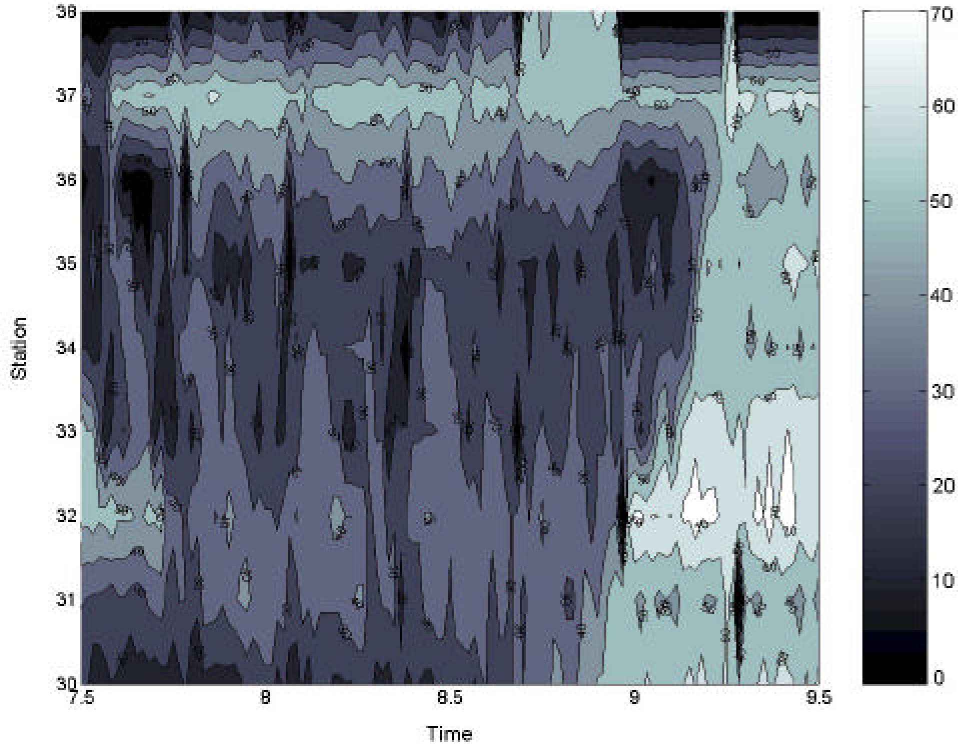


FIGURE 4-2: AN ILLUSTRATION OF TRAFFIC CONTOUR MAP FOR 4-MILE STRETCH OF I-4 TAKEN IN THE MORNING PEAK HOURS (7:30 AM-9:30 AM).

surroundings are taken at angles in the range from 0° to 360°. The angular variation was chosen such that a wide range of shockwave speeds is covered, when emanating from or into (*t*,*x*). Given that the spacing between two consecutive stations is nearly half a mile and that traffic measurements are taken every 30 seconds, it was found that the inclusion of 10 point measurements in the temporal dimension would result in the following shockwave speeds 0, 6, 6.7, 7.5, 8.6, 10, 12, 15, 20, 30, and 60 mph, starting from 0° and moving to 90° anticlockwise. Although the figure shows only one quarter (0° to 90°), the three remaining quarters (from 90° to 360°) can be similarly constructed as mirrors of the

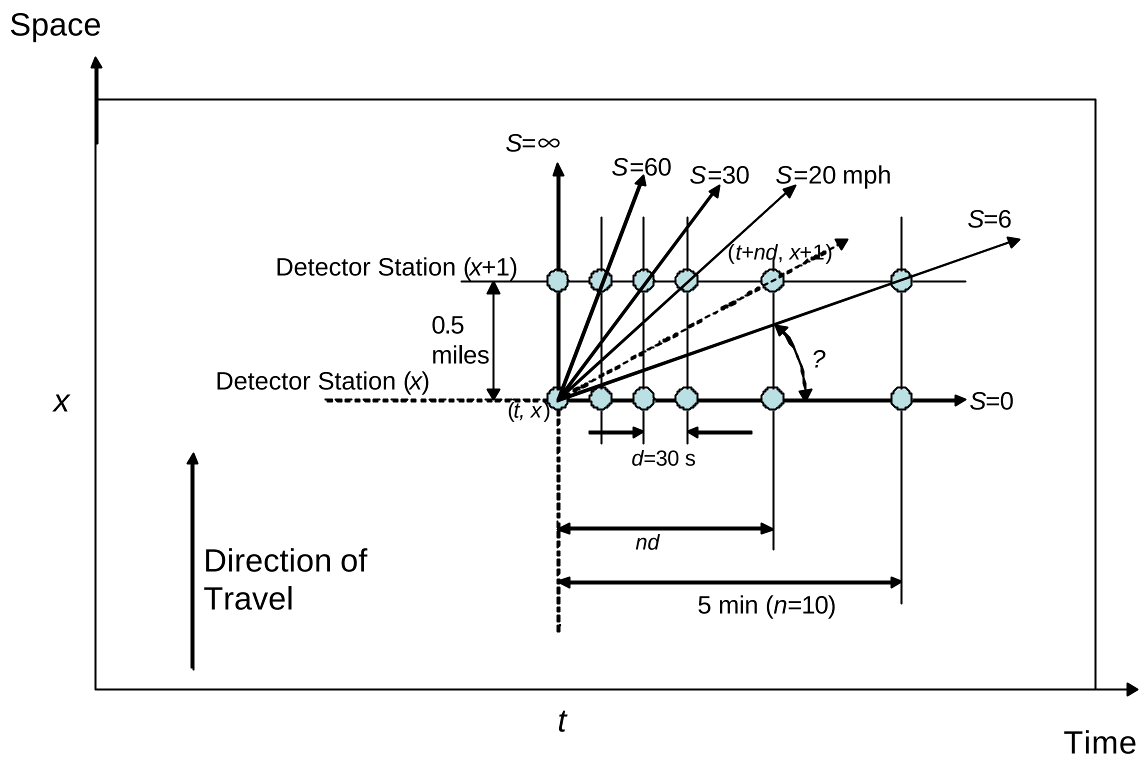


FIGURE 4-3: THE DIFFERENT ORIENTATIONS USED TO CALCULATE *P(I,J).*

first quarter. Each point will be compared to each of the neighboring points as shown in FIGURE 4-3 in all orientations from 0° to 360°. For each angle (?) (possible shockwave speed), a co-occurrence matrix can be constructed as follows:

$$A^q = \frac{1}{R}\begin{bmatrix} h(0,0) & h(0,1) & . & . & h(0,N) \\ h(1,0) & h(1,1) & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ h(N,0) & . & . & . & h(N,N) \end{bmatrix} \qquad [1]$$

Where $h(i,j)$ is the number of point pairs at angle *?* that have values *i* and *j*, respectively. Mathematically,

$$h(i,j) = \sum_t \sum_x f_{i,j}(t,x) \qquad \forall t \in 0,\ldots,N_t; x \in 0,\ldots,N_x \qquad [2]$$

And

$$f_{i,j}(t,x) = \begin{cases} 1 & if\ I(t,x) = i\, and\, I(t \pm nd, x \pm 1) = j \\ 0 & Otherwise \end{cases} \quad [3]$$

Where

$I(t,x)$ = the intensity at time t and location x, measured by one of the three traffic parameters. The value of $I(t,x)$ can be rounded off or discretized to suppress the effect of noise and minor disturbances in traffic conditions.

N_t = Number of observations in the temporal dimension of the contour map

N_x = Number of observations in the spatial dimension of the contour map

n = Number of 30-second observations in reference to time t

d = Temporal resolution of the contour map (30 seconds in this study)

R = the total number of possible pairs

N = the depth of the traffic contour map, measured by the number of discrete levels of the representative traffic parameter

$P(i,j) = h(i,j)/R$ and is defined as the probability of observing adjacent pairs of traffic contour map pixels with values i and j in all possible orientations; i.e. temporally, spatially, and spatio-temporally, as shown in FIGURE 4-3.

Using the probability values, a family of second-order statistical measures can be calculated. The most common statistical measures are: Angular Second Moment (*ASM*), Contrast (*CON*), Inverse Difference Moment (IDF), and Entropy (*ENT*). Each measure is explained in detail next.

4.3.1 Angular Second Moment (*ASM*)

ASM is a measure of texture smoothness of the image. In the context of spatiotemporal traffic contour maps, high *ASM* values indicate uniform transitions or very low disturbances in the traffic conditions, while low *ASM* values reflect that the traffic conditions are non –uniform or exhibit major disturbances in the traffic contour map under study. Mathematically, *ASM* is expressed as

$$ASM = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j)^2 \qquad [4]$$

Where, N is measure of color quality of the digital image. For traffic contour map, it is one of the three traffic parameters.

4.3.2 Contrast (*CON*)

Contrast is a measure typically used to indicate local color variations of a digital map. In the context of traffic contour maps, contrast emphasizes local variations in traffic conditions where, local variations indicate the severity of disturbance in the region considered. Since traffic contour maps are spatiotemporal in nature, the local variations can be temporal (horizontal), spatial (vertical), or spatiotemporal (angular), in reference to FIGURE 4-3. High contrasts exhibit severe local disturbances that could be attributed to heavy weaving or merging maneuvers and abrupt changes in demand or capacity as a result of lane-blocking incidents. Mathematically, Contrast is defined as

$$CON = \sum_{n=0}^{N-1} n^2 \left[\sum_{i=0}^{N-1} \sum_{j=0_{|i-j|}}^{N-1} P(i,j) \right] \qquad [5]$$

4.3.3 Inverse Difference Moment (*IDF*)

IDF is a complimentary measure that relates inversely to the contrast measure. It is also a direct measure of the local homogeneity of a digital image. Low IDF is associated with low homogeneity and vice versa. For a traffic contour map, this property can be similarly measured to quantify the degree of homogeneity in traffic conditions in the time-space domain. In mathematical terms, *IDF* is measured as

$$IDF = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1 + (i - j)^2} \qquad [6]$$

4.3.4 Entropy (*ENT*)

This measure is often used to quantify the expected amount of surprise or uncertainty in a random variable. In information theory, entropy is considered to be the average amount of information received when a random variable is observed. Also, entropy can be used to measure the randomness in a digital map. In traffic contour maps, high Entropy indicates high randomness and vice versa. Mathematically, Entropy is defined as

$$H_{x,t} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \log_2 P(i, j) \qquad [7]$$

Where x and t refer to spatial and temporal dimensions of a traffic contour map.

4.3.5 Other Statistical Measures

Additional second-order statistical measures will be evaluated to quantify other properties such as:

$$\text{CORRELATION} = \frac{\left[\sum_i \sum_j (ij) P(i, j)\right] - m_x m_t}{s_x s_t} \qquad [8]$$

$$\text{SUM ENTROPY} = \sum_{i=0}^{2N-2} P_{x+t}(i) \log P_{x+t}(i) \qquad [9]$$

$$\text{DIFFERENCE ENTROPY} = -\sum_{i=0}^{2N-2} P_{x-t}(i) \log P_{x-t}(i) \qquad [10]$$

$$\text{INFORMATION MEASURE} = \frac{ENT_{tx} - ENT_{tx}^{1}}{\max(ENT_x, ENT_t)} \qquad [11]$$

Where

m_x, m_t = The speed mean observed over x and t, respectively

s_x, s_t = The speed standard deviation observed over x and t, respectively; and

$$ENT_{tx}^{1} = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j) \log[P_x(i) P_t(j)] \qquad [12]$$

$$P_x(i) = \sum_{j=0}^{N-1} P(i,j) \qquad [13]$$

$$P_t(i) = \sum_{i=0}^{N-1} P(i,j) \qquad [14]$$

$$P_{x \pm t}(k) = \sum_{i=0}^{N-1} \sum_{j=0_{|i \pm j|=k}}^{N-1} P(i,j) \qquad [15]$$

$$ENT_x = -\sum_{i=0}^{N-1} P_x(i) \log P_x(i) \qquad [16]$$

$$ENT_t = -\sum_{j=0}^{N-1} P_t(j) \log P_t(j) \qquad [17]$$

Apart from structural and statistical approaches for analyzing image textures, there are other techniques, such as measuring the microstructure of texture patterns, and analyzing textures using fractals to name a few. However, in this study only statistical approach is used to analyze spatio-temporal traffic contour maps.

4.4 Summary

This chapter presented the concept of traffic contour maps and its relevance in digital image analysis. The chapter also discussed the methodology implemented to derive the second order statistical measures from digital images. The measures discussed in this chapter are implemented in case of spatio-temporal traffic contour maps and the results obtained were discussed in the next chapter.

5 Second-Order Statistical Measures

In this chapter, the methodology used to derive the performance measures and analysis conducted using those measures are discussed. In order to develop new performance measures using the collected traffic data, a special module was developed in Visual Basic. With the help of the developed module, performance measures are calculated for the selected contour maps. The module is given in Appendix A. For the visual representation of the traffic contour maps MATLAB is used. An example of such a map can be seen in FIGURE 4-2.

A sample of the data collected from the database will be in the form given in the FIGURE 5-1. The data contains the following fields: hour, minute, second, time, station, east bound speed (ES), west bound speed (WS), east bound volume (EV), west bound volume (WV), east bound occupancy (EO), and west bound occupancy (WO). The module is written such that we can calculate performance measures for the constructed contour map by plugging in the following inputs: direction of travel (eastbound, west bound, or both), the traffic parameter (speed, volume, and occupancy) with which it can construct the traffic contour map, the starting point and ending point of the freeway section, the spatial dimensions (number of loop detector stations), temporal dimensions (minutes) of the contour map, and the level of aggregation of the traffic parameter. In this study we used 5 mph. A sample of module input requirements is given in Appendix A.

Once the required inputs are selected the module is allowed to run, and the output

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Hour	Minute	Second	Time	Station	ES	ES	ES	WS	WS	WS	EV	EV	EV	WV	WV	WV	EO	EO	EO	WO	WO	WO
2	0	0	30	00:30.0	2	68	57	NULL	68	58	NULL	5	3	NULL	4	2	NULL	2	3	NULL	2	1	NULL
3	0	0	30	00:30.0	3	71	64	NULL	0	58	NULL	4	4	NULL	0	2	NULL	1	2	NULL	0	1	NULL
4	0	0	30	00:30.0	4	64	83	NULL	67	0	NULL	4	2	NULL	2	0	NULL	1	0	NULL	1	0	NULL
5	0	0	30	00:30.0	5	57	0	NULL	55	55	NULL	5	0	NULL	5	2	NULL	2	0	NULL	2	0	NULL
6	0	0	30	00:30.0	6	53	46	52	64	0	53	4	2	10	2	0	4	2	1	7	1	0	3
7	0	0	30	00:30.0	7	66	66	61	0	64	58	1	4	3	0	3	1	2	5	2	0	3	1
8	0	0	30	00:30.0	8	66	53	45	61	55	47	1	6	2	1	6	4	0	5	3	0	4	3
9	0	0	30	00:30.0	9	47	0	66	74	70	0	2	9	4	2	5	0	20	3	2	0	2	4
10	0	0	30	00:30.0	10	60	63	68	56	56	64	5	6	1	3	1	2	3	4	0	4	0	3
11	0	0	30	00:30.0	11	53	56	65	54	60	70	6	5	2	7	6	3	2	2	0	2	2	1
12	0	0	30	00:30.0	12	57	56	67	60	59	46	6	6	6	7	7	9	3	4	3	7	4	8
13	0	0	30	00:30.0	13	63	56	53	64	63	56	8	7	6	2	5	7	4	5	4	1	2	4
14	0	0	30	00:30.0	14	53	58	64	64	59	56	7	9	4	2	6	5	5	6	2	1	3	5
15	0	0	30	00:30.0	15	68	57	53	62	73	59	6	11	8	6	2	4	2	7	10	3	1	2
16	0	0	30	00:30.0	16	62	53	45	61	61	59	7	7	2	4	4	5	2	2	0	1	1	2
17	0	0	30	00:30.0	17	62	0	66	0	66	66	4	4	6	0	4	3	2	4	3	0	4	1
18	0	0	30	00:30.0	18	53	57	66	64	59	54	5	4	1	2	7	7	3	3	0	1	4	4

FIGURE 5-1: SAMPLE OF THE DATA EXTRACTED FROM THE DATABASE

is obtained. From the output, we get new performance measures such as Average (*AVG*), Variance (*VAR*), Angular Second Moment (*ASM*), Entropy (*ENT*), Contrast (*CON*), Inverse Difference Moment (*IDF*), Correlation (*CORR*), Sum Entropy (*SUMENT*), Difference Entropy (*DIFFENT*), and Information Measure (*INFO*).

5.1 Information Redundancy Check

In order to quantify the information redundancy among the measures, a correlation matrix is constructed among all the measures. The correlation coefficient between a pair of measures was used to indicate degree of association between each pair of measures and hence can be used to eliminate information redundancies. Measures with high correlation coefficient strongly indicate that they both quantify the same property, and therefore, should not be used together. The threshold for the correlation coefficient between any pair of measures is selected as 0.9. Consequently, if any pair of measures has correlation coefficient value more than 0.9, one of the measures should be eliminated to remove redundancy in information. A redundancy score was used to count the number of times the correlation coefficient exceeded the threshold value 0.9 for each measure. This score is used to eliminate information redundancy. Measures with high scores indicate high contribution to information redundancy, and therefore are eliminated first. TABLE 5-1 shows redundancy scores for different measures.

TABLE 5-1: CORRELATION MATRIX BETWEEN MEASURES

	MEAN	*VAR*	*ASM*	*CON*	*IDF*	*CORR*	*ENT*	*SUMENT*	*DIFFENT*	*INFO*	Overall Redundancy Score
MEAN	1.000	-0.618	0.472	-0.545	0.605	0.937	-0.780	-0.748	-0.741	-0.705	**1**
VAR	-0.618	1.000	-0.475	0.749	-0.500	-0.652	0.608	0.692	0.660	0.643	**0**
ASM	0.472	-0.475	1.000	-0.492	0.907	0.625	-0.820	-0.811	-0.798	-0.858	**1**
CON	-0.545	0.749	-0.492	1.000	-0.560	-0.635	0.689	0.744	0.785	0.692	**0**
IDF	0.605	-0.500	0.907	-0.560	1.000	0.775	-0.890	-0.851	-0.883	-0.922	**2**
CORR	0.937	-0.652	0.625	-0.635	0.775	1.000	-0.847	-0.816	-0.829	-0.813	**1**
ENT	-0.780	0.608	-0.820	0.689	-0.890	-0.847	1.000	0.984	0.974	0.974	**3**
SUMENT	-0.748	0.692	-0.811	0.744	-0.851	-0.816	0.984	1.000	0.972	0.971	**3**
DIFFENT	-0.741	0.660	-0.798	0.785	-0.883	-0.829	0.974	0.972	1.000	0.965	**3**
INFO	-0.705	0.643	-0.858	0.692	-0.922	-0.813	0.974	0.971	0.965	1.000	**4**

As shown in TABLE 5-1, Information Measure (*INFO*) received highest redundancy score (4) and therefore, will be eliminated first. Measures Entropy (*ENT*), Sum Entropy (*SUMENT*), Difference Entropy (*DIFFENT*) have a redundancy score of 3 and since three of these are highly correlated, only one of the three will be retained and the other two will be eliminated. *IDF* measure has a redundancy score of 2, with strong correlation with *ASM* and *INFO*. Since *INFO* has already been eliminated, we eliminate one from *ASM* and *IDF* and we choose *IDF*. Measures *MEAN* and *CORR* have a redundancy score of 2 and are inter-correlated. For simplification purpose we eliminate

CORR measure. Measures *VAR* and *CON* did not exhibit any strong correlation between any measures, and therefore, both measures are retained.

From a total of 10 performance measures only 5 are retained and the others are eliminated owing to information redundancy. The five measures, which we have retained, are MEAN, *VAR*, *ASM*, *CON* and *ENT*. Further analysis is conducted on the retained measures as given in the following sections.

5.2 Sensitivity Analysis

This section provides the sensitivity of the new second-order statistical measures to various traffic conditions that are quantified by the mean speed of each contour map. The sensitivity analysis helps us detect relationships between first-order and second-order properties of traffic conditions. In order to conduct the sensitivity analysis, the second order statistical measures obtained are grouped by *MEAN* over 5-mph intervals for each of the 14,270 contour maps constructed. For each speed interval, descriptive statistics (mean and 95% confidence intervals) of *ASM*, *CON*, and *ENT* were calculated.

Descriptive statistics of *ASM*, *CON*, and *ENT* were plotted against mean speed intervals derived from corresponding traffic contour maps. FIGURE 5-2 shows a plot between descriptive statistics of *ASM* (measure of smoothness) versus mean speed intervals. The figure indicates whether there is any significant differences exist across different mean speed intervals. In the context of spatiotemporal traffic contour maps, high *ASM* values indicate uniform transitions or very low disturbances in the traffic conditions, while low *ASM* values reflect that the traffic conditions are non–uniform or exhibit major disturbances in traffic. From the figure we see that *ASM* values are lowest for the speed interval from 25 to 30 mph, which indicates that this speed interval marks the roughest conditions of traffic contour maps.

From the figure we can also see that the *ASM* confidence bounds do not overlap with other limits, so we can conclude that differences are statistically significant. This indicates that the highest level of non-uniformity in traffic conditions is likely to be observed in that speed range. On the other hand the highest mean *ASM* is observed in the speed range of 55-60 mph, when traffic operates at or near free-flow conditions and speeds are likely to be more uniform. A slight drop in *ASM* is observed in the speed range of 60-65 mph, which might be due to variations in speeds under free-flow conditions where drivers have the freedom of choosing their own desired speeds.

FIGURE 5-3 shows a plot between the descriptive statistics of *CON* (measure of contrast) versus mean speed intervals. In the context of traffic contour maps, high contrast levels indicate high local disturbances in the traffic conditions and vice versa. From the figure we can observe maximum *CON* values in the speed range 35-40 mph, which represents maximum local variations. We can also see that the difference in *CON* values is not statistically significant between 35-40 mph interval and 40-45 mph interval, which represents that in speed range between 35-45 mph, there is lowest level of homogeneity on traffic conditions. This speed range marks the transient states from and into congested conditions, where high levels of local disturbances due to traffic breakdowns or recovery are often occurred. As the speed range increases to 60 mph, we can see a decrease in *CON* value, which represents high level of homogeneity. We can also observe a slight increase in *CON* value in the speed range 60-65 mph, which can be attributed to driver's ability to choose their desired speeds under extremely unconstrained conditions.

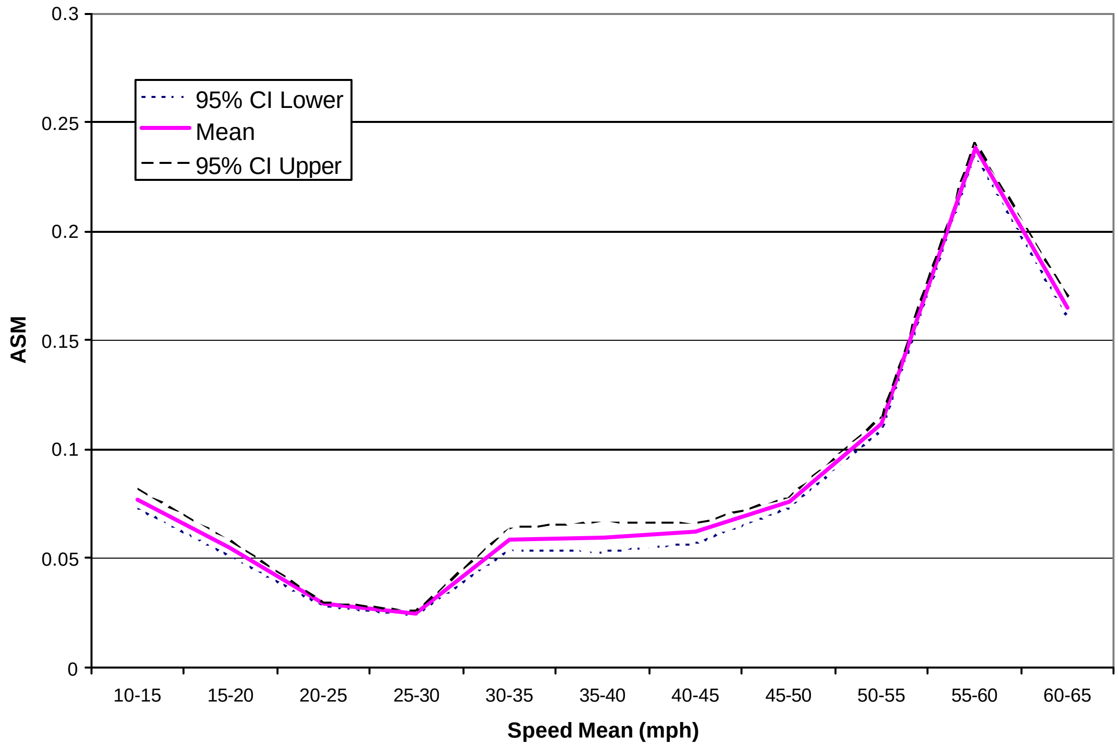


FIGURE 5-2: MEAN *ASM* AND 95% CONFIDENCE INTERVALS FOR DIFFERENT TRAFFIC CONDITIONS

FIGURE 5-4 shows a plot between descriptive statistics of *ENT* (measure of randomness) versus mean speed intervals. In traffic contour maps high Entropy indicates high randomness and vice versa. The figure shows maximum *ENT* (highest randomness) at speed range 25-30 mph. Lower randomness is observed during heavy congestion (20 mph or less) where traffic has already plunged into extreme forced-flow conditions and the lowest levels of randomness are observed during speed ranges 55 mph or higher. Statistically significant differences in the values of *ENT* exist for different speed ranges, which are consistent with the other results from *ASM* and *CON*.

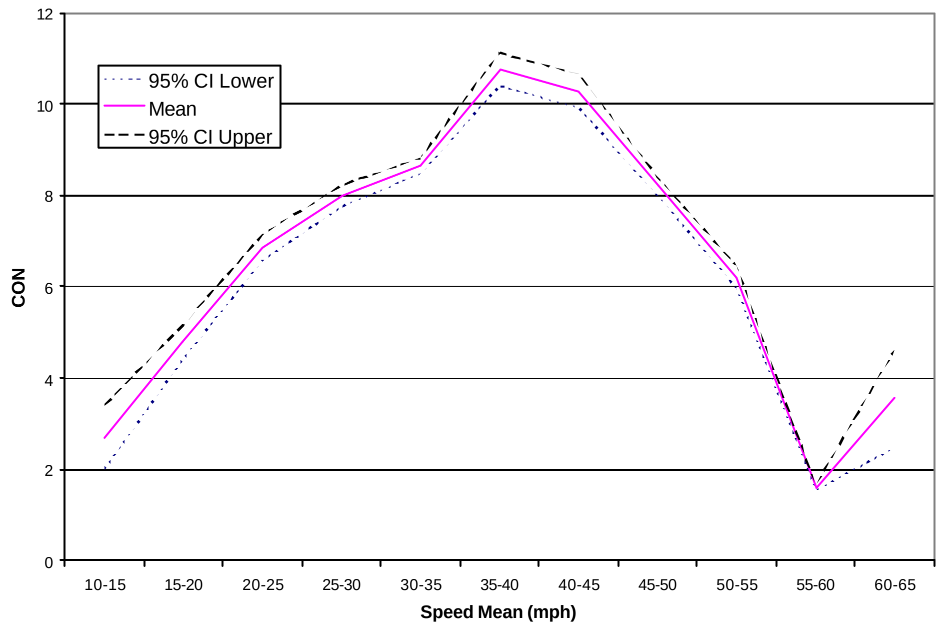


FIGURE 5-3: MEAN *CON* AND 95% CONFIDENE INTERVALS FOR DIFFERENT TRAFFIC CONDITIONS

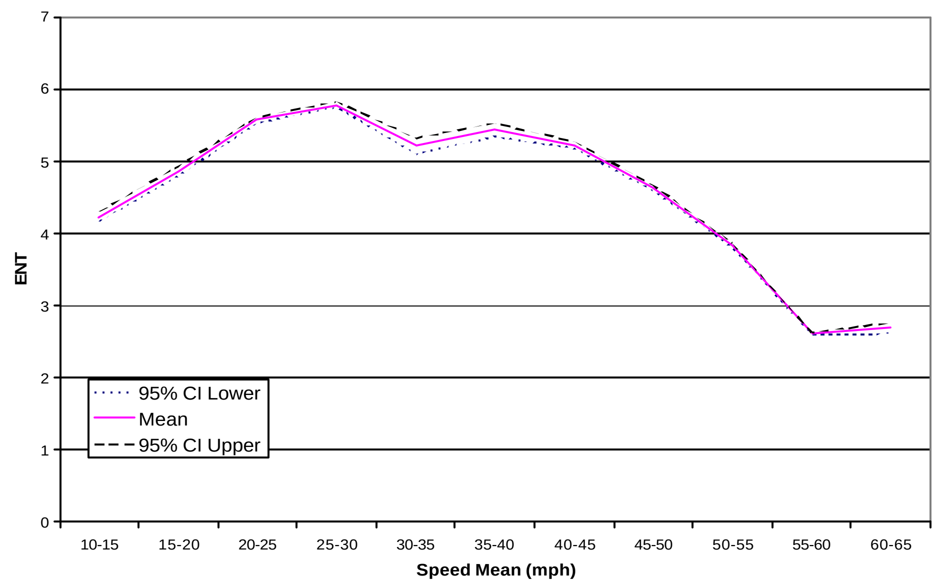


FIGURE 5-4: MEAN *ENT* AND 95% CONFIDENCE INTERVALS FOR DIFFERENT TRAFFIC CONDITIONS

5.3 Level of Service (LOS)

In order to assess the performance of the freeway system, we have established LOS criteria, similar to those used with conventional measures in the highway capacity manual based on travel time, density, etc. As we have three new measures and each represents a different property, the LOS will be established for each measure separately. Following the conventions of the highway capacity manual for estimating the level of service on basic freeway segments, we may divide the plausible range of each measure into 6 distinct categories: A through F.

FIGURE 5-5 shows LOS based on *ASM*. For smoothness, it is divided into 6 arbitrarily selected categories. LOS A indicates very smooth operation of traffic with value greater than or equal to 0.5. On the other side LOS F is characterized with very coarse operation with values less than or equal to 0.1. Intermediate values represent gradual variation from smooth operation to coarse operation.

Similar LOS criteria were established for the measures contrast (homogeneity) and entropy (randomness). For homogeneity, *CON* is also divided into six arbitrarily selected regions A through F. For LOS A, $CON \leq 0.5$, which indicates highly homogenous traffic conditions where the homogenous conditions could be free-flow conditions or heavily congested conditions. For LOS F, $CON \geq 8$ indicates very low homogeneity, which means the traffic is fluctuating from free-flow to intermediate congestion to heavy congested conditions and vice versa. The LOS criterion for *CON* measure is shown in FIGURE 5-6.

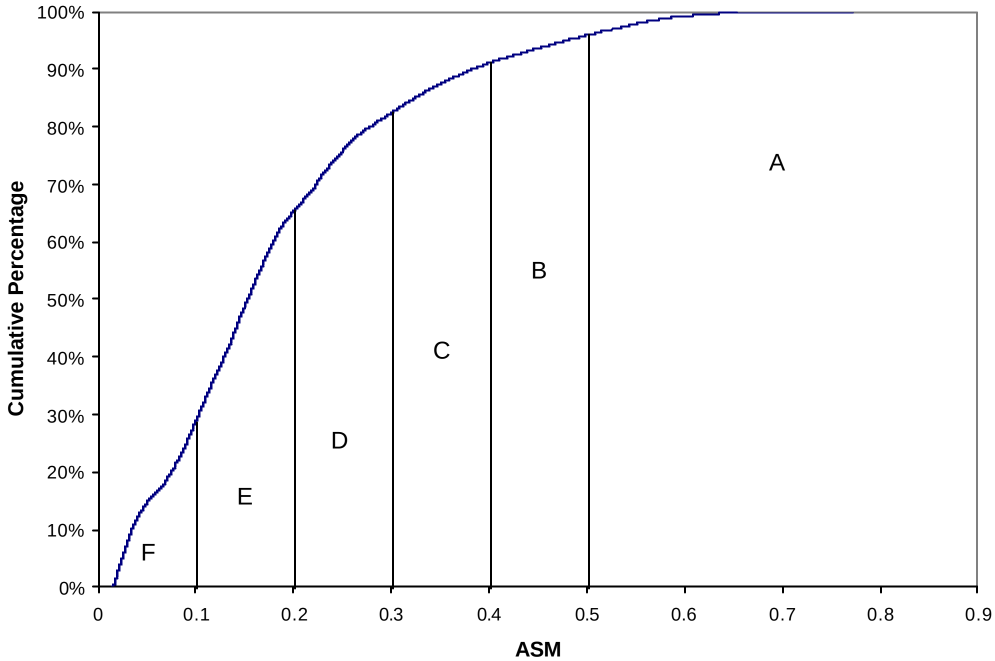


FIGURE 5-5: LEVEL OF SERVICE CRITERIA BASED ON *ASM*

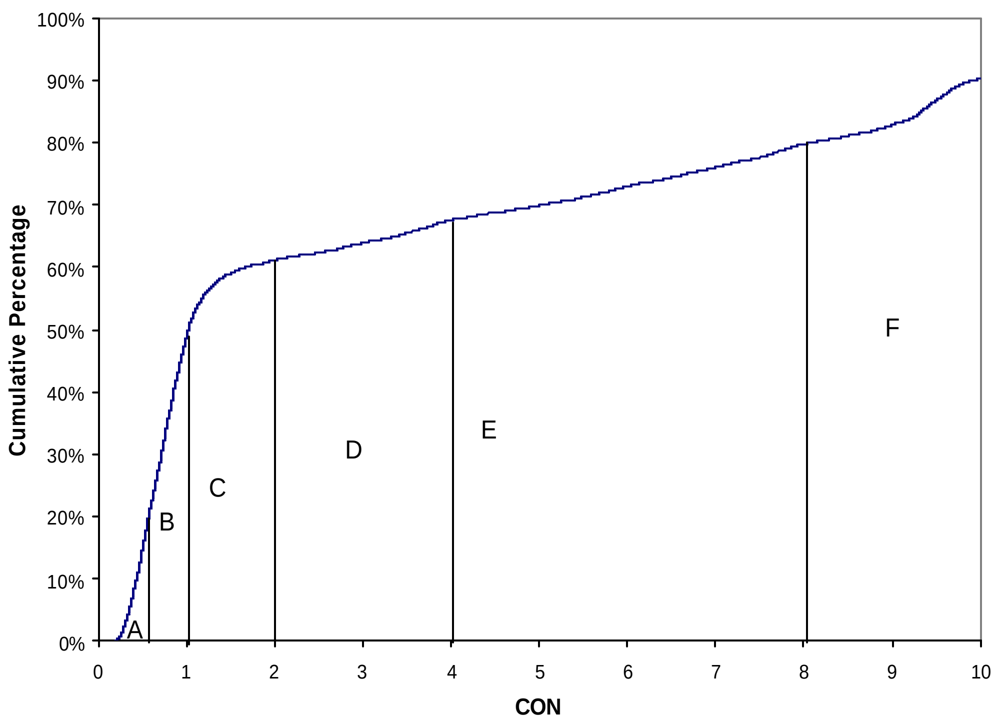


FIGURE 5-6: LEVEL OF SERVICE BASED ON *CON*

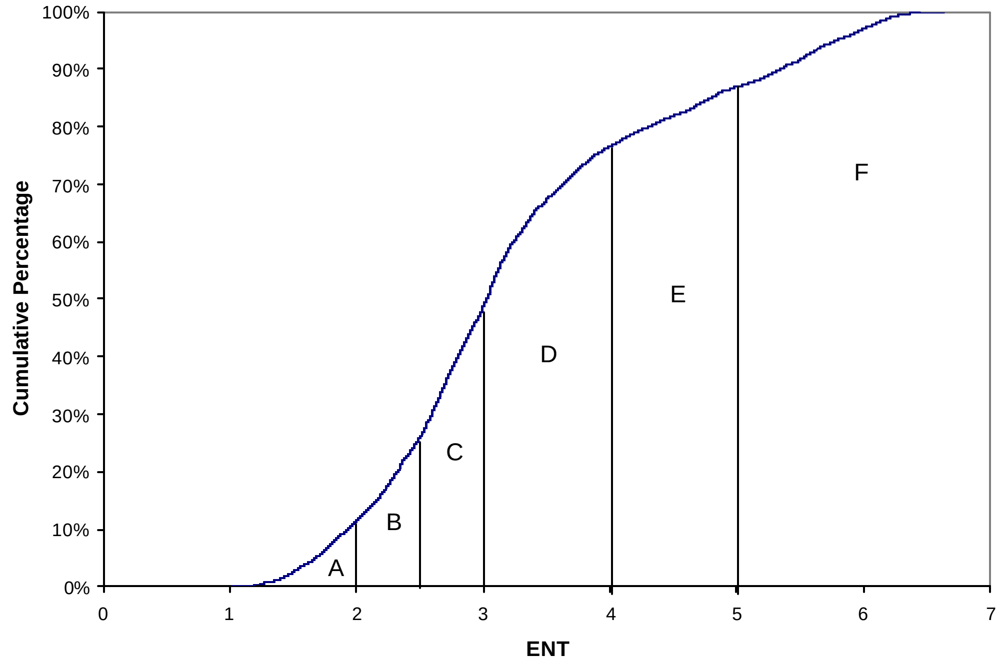


FIGURE 5-7: LEVEL OF SERVICE BASED ON *ENT*

LOS for *ENT* which measures randomness is also divided into six regions: A through F. for LOS A, ENT $\leq$ 2, which represents low randomness in traffic conditions and for LOS F, *ENT* $\geq$ 5, which represents high randomness in traffic conditions. FIGURE 5-7 shows LOS for *ENT*.

5.4 On-Line Performance Assessment of Newly Developed Measures

The newly developed second-order statistical measures can be implemented on-line for extended time periods and long freeway sections. The on-line implementation is achieved by specifying the spatial and temporal dimensions of the traffic contour maps. The spatial dimensions is specified by length of the freeway section for which we need the performance monitoring and temporal dimensions is determined by selecting the time period for which we need monitoring. The spatial and temporal dimensions of the constructed traffic contour maps can be updated in real-time by moving time space window incrementally over time.

To facilitate the on-line implementation, we make use of the module that was developed to conduct the analysis of this study. The module can be executed in real time and the user will be able to specify the following parameters in constructing traffic contour maps: direction of travel (eastbound, west bound, or both), the traffic parameter (speed, volume, and occupancy) with which it can construct the traffic contour map, the starting point and ending point of the freeway section, the spatial dimensions (number of loop detector stations), temporal dimensions (minutes) of the contour map, and the level of aggregation of the traffic parameter. In this study we have used 5 mph.

5.5 Summary

This chapter discussed the analysis conducted on first and second order statistical measures obtained from spatio-temporal traffic contour maps. A correlation coefficient matrix was constructed among the measures to eliminate the information redundancy and measures with low correlation coefficient are retained. With the retained measures, sensitivity of the measures to various traffic conditions is presented. Level of service criteria was established to describe the performance of the freeway system. The sensitivity analysis and level of service criteria established in this study can be implemented in real-time using the developed module. In the next chapter we discuss the different statistical tests used to compare the characteristics of congestion regions for all the weekdays.

6 Statistical Comparisons

In this chapter we use the loop detector data collected from stations 0 to 69 for the same five days to compare the characteristics of congestion regions using the developed performance measures. To conduct this, we identify the congested regions in the east bound direction only. Based on observations, congested regions were identified over a 7.5-mile freeway section that falls between loop detector stations 25 to 40, from 7:30 am to 10:00 am.

For each congestion region a traffic contour map was constructed and the second-order statistical measures were calculated for a time-space window of 5 minutes and 3 stations. As the study area has 16 stations over a period from 7:30 am to 10:00 am, the total size of the study area is 151x16 observations. This results in a total of 14*147=2058 time space windows for each day.

After obtaining the selected second order statistical measures for all the five days, statistical analysis was conducted to check whether the calculated measures for different days follow the same distribution or not, i.e. comparing different congestion regions to check whether they represent same congestion conditions or not. In the next section a brief introduction on the statistical tests used in this study is given.

6.1 Statistical Analysis

In this study, we compare the characteristics of congested regions of different days to check whether their distributions are similar. To achieve this we use a set of statistical tests to compare distributions of measures obtained from different days. As we are not assuming the obtained measures for different days to follow any particular distribution, it is appropriate to use non-parametric tests to compare the distributions of different days.

As the measures of different days are unpaired, the appropriate statistical tests that we can employ are Pearson's Chi-Square test for independence in Two-Way contingency tables, Kuskal-Wallis (K-W) Test and Kolmogorov-Smirnov (K-S) Test. While the first two tests are used to compare the distributions of measures of all the days at a time, K-S test is used to compare distributions of two days at a time. This test is done by forming pairs of two days from all the weekdays. To test for equality of means between all pairs of days, we have used two-sample t-test. All the statistical tests were carried out in SYSTAT statistical software package. In the next few sections, different statistical tests used in this study along with the results are discussed.

6.2 Pearson's Chi-Square Test for Independence in Two-Way Contingency Tables

A two-way contingency table is a table of frequencies obtained when a set of objects is simultaneously classified under two different categorizations. If the first categorization (rows) has r levels and second categorization (columns) has c levels then the data sets of this form are referred to as r x c contingency tables. In order to test for independence in these tables we use Pearson's chi-square test statistic.

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

Where o_{ij} is the observed frequency and e_{ij} is the expected frequency. The expected frequency e_{ij} is calculated using

$$e_{ij} = \frac{N_i O_j}{N}$$

Where N_i is the total number of observations of category r, i.e. row totals, and O_j is the total number of observations of category c, i.e. column totals. N is the total size of the sample. Therefore, the null hypothesis that all the samples follow same frequency distribution is rejected when.

$$X^2 > X^2_{a,df}$$

Where $X^2_{a,\, df}$ is the chi-square percent point function with level of significance *a* and degrees of freedom *df=(r-1)*(c-1).*

6.2.1 Chi-Square Test Results

In order for the Pearson's chi-Square test static to be applied to *ASM*, the obtained measures for different days are divided into 9 bins. Therefore, the *ASM* measure can be represented as a Two-Way contingency table as shown in the TABLE 6-1. The null hypothesis tested is that measures of all the days follow same frequency distribution.

TABLE 6-1: TWO-WAY CONTINGENCY TABLE FOR *ASM*

	0-0.1	0.-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-1.0	Total
day 1	760	247	260	279	74	62	83	68	225	2058
day 2	624	203	201	359	108	78	105	82	298	2058
day 3	623	230	244	302	107	80	101	95	276	2058
day 4	784	407	252	268	64	77	0	49	157	2058
day 5	313	427	437	360	134	62	77	72	176	2058
Total	3104	1514	1394	1568	487	359	366	366	1132	10290

The null hypothesis is rejected as the Pearson's Chi-Square test static ($\boldsymbol{X^2}$ =736.7) obtained was greater than Chi-Square percent point function ($\boldsymbol{X^2_{0.05,32}}$=46.194) with 0.05 level of significance and 32 degrees of freedom. This indicates that there is no statistical evidence or very little probability (<0.001) that the frequency distributions of measures of all the days are same.

Similarly, for the measures *CON* and *ENT* the Two-Way contingency tables are given in TABLE 6-2 and TABLE 6-3, respectively. The measure *CON* is divided into 9 bins and *ENT* is divided into 12 bins. For these measures also the null hypothesis tested was that the frequency distributions of all the weekdays are same. For *CON* measure $X^2 = 656$ and Chi-Square percent point with 0.05 level of significance and 32 degrees of freedom, $X^2_{0.05,32} = 46.194$ and for *ENT* measure $X^2 = 625.37$ and Chi-Square percent point with level of significance 0.05 and 44 degrees of freedom is, $X^2_{0.05,44} = 59.194$. As $X^2 > X^2_{\alpha,df}$ for both measures, the null hypothesis is rejected as there is no statistical evidence or very little probability (<0.001) that the frequency distributions of the measures for all the weekdays are similar. The Pearson's Chi-Square test statistic output as obtained from the SYSTAT software is given in TABLE 6-4.

TABLE 6-2: TWO-WAY CONTINGENCY TABLE FOR *CON*

	0--5	5--10	10--15	15--20	20--25	25--30	30--35	35--40	40--45	Total
day 1	1329	271	166	87	48	28	41	85	3	2058
day 2	1253	225	136	81	69	17	102	106	69	2058
day 3	1291	261	118	102	48	52	50	130	6	2058
day 4	1261	267	158	84	101	48	74	53	12	2058
day 5	1532	131	85	9	7	3	57	218	16	2058
Total	6666	1155	663	363	273	148	324	592	106	10290

TABLE 6-3: TWO-WAY CONTINGENCY TABLE FOR *ENT*

	0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5	4.5-5.0	5.0-5.5	5.5-6.0	Total
day 1	225	83	135	229	261	211	102	192	287	217	101	15	2058
day 2	298	98	195	250	263	173	131	136	182	203	110	19	2058
day 3	276	102	198	230	250	207	105	156	223	205	86	20	2058
day 4	169	106	124	244	218	254	143	174	214	256	137	19	2058
day 5	194	141	202	514	291	259	115	112	93	77	58	2	2058
Total	1162	530	854	1467	1283	1104	596	770	999	958	492	75	10290

6.3 Kruskal-Wallis (K-W) Test

The Kruskal-Wallis test, also called Kruskal-Wallis H test, is a non-parametric test to compare three or more unpaired groups. It is also called Kruskal-Wallis one-way analysis of variance by ranks. If we have K independent samples of sizes $n_1, \ldots\ldots\ldots n_K$, we combine all the samples into one large sample, arrange them in ascending order and

TABLE 6-4: PEARSON'S CHI-SQUARE TEST OUTPUT FROM SYSTAT

ASM

TABULATE DAYS *BINS$
Frequencies
DAYS (rows) by BINS$ (columns)

	0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-1.0	Total
Day 1	760	247	260	279	74	62	83	68	225	2058
Day 2	624	203	201	359	108	78	105	82	298	2058
Day 3	623	230	244	305	107	80	101	95	276	2058
Day 4	784	407	252	268	64	77	0	49	157	2058
Day 5	313	427	437	360	134	62	77	72	176	2058
Total	3104	1514	1394	1568	487	359	366	366	1132	10290

Expected values
DAYS (rows) by BINS$ (columns)

	0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-1.0
Day 1	620.8	302.8	278.8	313.6	97.4	71.8	73.2	73.2	226.4
Day 2	620.8	302.8	278.8	313.6	97.4	71.8	73.2	73.2	226.4
Day 3	620.8	302.8	278.8	313.6	97.4	71.8	73.2	73.2	226.4
Day 4	620.8	302.8	278.8	313.6	97.4	71.8	73.2	73.2	226.4
Day 5	620.8	302.8	278.8	313.6	97.4	71.8	73.2	73.2	226.4

Test Statistic	Value	Df	Probability
Pearson’s Chi-Square	736.7	32	<0.0001

CON

TABULATE DAYS *BINS$
Frequencies
DAYS (rows) by BINS$ (columns)

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	Total
Day 1	1329	271	166	87	48	28	41	85	3	2058
Day 2	1253	225	136	81	69	17	102	106	69	2058
Day 3	1291	261	118	102	48	52	50	130	6	2058
Day 4	1261	267	158	84	101	48	74	53	12	2058
Day 5	1532	131	85	9	7	3	57	218	16	2058
Total	6666	1155	663	363	273	148	324	592	106	10290

TABLE 6-4 (CONTINUED

Expected values

DAYS (rows) by BINS$ (columns)

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Day 1	1333	231	132.6	72.6	54.6	29.6	64.8	118	21.2
Day 2	1333	231	132.6	72.6	54.6	29.6	64.8	118	21.2
Day 3	1333	231	132.6	72.6	54.6	29.6	64.8	118	21.2
Day 4	1333	231	132.6	72.6	54.6	29.6	64.8	118	21.2
Day 5	1333	231	132.6	72.6	54.6	29.6	64.8	118	21.2

Test Statistic	Value	Df	Probability
Pearson's Chi-Square	736.7	32	<0.0001

ENT

TABULATE DAYS *BINS$

Frequencies

DAYS (rows) by BINS$ (columns)

	0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5	4.5-5.0	5.0-5.5	5.5-6.0	Total
Day 1	225	83	135	229	261	211	102	192	287	217	101	15	2058
Day 2	298	98	195	250	263	173	131	136	182	203	110	19	2058
Day 3	276	102	198	230	250	207	105	156	223	205	86	20	2058
Day 4	169	106	124	244	218	254	143	174	214	256	137	19	2058
Day 5	194	141	202	514	291	259	115	112	93	77	58	2	2058
Total	1162	530	854	1467	1283	1104	596	770	999	958	492	75	10290

Expected values

DAYS (rows) by BINS$ (columns)

	0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5	4.5-5.0	5.0-5.5	5.5-6.0
Day 1	232.4	106	170.8	293.4	256.6	220.8	119.2	154	199.8	191.6	98.4	15
Day 2	232.4	106	170.8	293.4	256.6	220.8	119.2	154	199.8	191.6	98.4	15
Day 3	232.4	106	170.8	293.4	256.6	220.8	119.2	154	199.8	191.6	98.4	15
Day 4	232.4	106	170.8	293.4	256.6	220.8	119.2	154	199.8	191.6	98.4	15
Day 5	232.4	106	170.8	293.4	256.6	220.8	119.2	154	199.8	191.6	98.4	15

Test Statistic	Value	Df	Probability
Pearson's Chi-Square	625	44	<0.0001

assign ranks, and then find R_i, the average of the ranks of the observations in i^{th} sample.

The statistic is given as

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{(\sum R_i)^2}{n_i} - 3(N-1)$$

Where N is the total number of cases, n_i is the number of cases in a given group and $(S R_i)^2$ is the sum of ranks squared for a given group of subjects.

We reject the null hypothesis H_o, that the K independent groups follow same distribution if $H > X^2_{a,\ df}$. Where $X^2_{a,\ df}$. is the chi-square percent point with level of significance a and degrees of freedom $df=(K-1)$.

6.3.1 Results of K-W Test

The Kruskal-Wallis test is applied for the measures *ASM*, *CON*, and *ENT* to test the null hypothesis that each measure comes from the same distribution. From the results of the test, the null hypothesis is rejected for all the measures, as the probability of occurrence of the null hypothesis is less than 0.0001. Therefore, this test also concludes that the distributions of measures in congestion regions for all weekdays differ significantly. The results of the test statistic as obtained from the SYSTAT software for all the measure is given in TABLE 6-5.

6.4 Discussion

Based on the results of the above two statistical tests we can conclude that the characteristics of congestion regions for the considered weekdays under study were statistically different. Although the distribution of measures for all weekdays differed significantly, it is possible that distributions of any two different days behave similarly. The same can be observed in FIGURE 6-1, FIGURE 6-2 and FIGURE 6-3, which give

TABLE 6-5: KRUSKAL-WALLIS TEST OUTPUT FROM SYSTAT

ASM

>KRUSKAL VAR00001 * VAR00002
Categorical values encountered during processing are:
VAR00002 (5 levels)
1, 2, 3, 4, 5
Kruskal-Wallis One-Way Analysis of Variance for 10290 cases
Dependent variable is VAR00001
Grouping variable is VAR00002

Group	Count	Rank Sum
1	2058	1.000370E+07
2	2058	1.012258E+07
3	2058	1.112390E+07
4	2058	9.023942E+06
5	2058	1.15698E+07

Kruskal-Wallis Test Statistic = 245.157
Probability is 0.000 assuming Chi-square distribution with 4 df

CON

>KRUSKAL VAR00001 * VAR00002
Categorical values encountered during processing are:
VAR00002 (5 levels)
1, 2, 3, 4, 5
Kruskal-Wallis One-Way Analysis of Variance for 10290 cases
Dependent variable is VAR00001
Grouping variable is VAR00002

Group	Count	Rank Sum
1	2058	1.000370E+07
2	2058	1.012258E+07
3	2058	1.112390E+07
4	2058	9.023942E+06
5	2058	1.15698E+07

Kruskal-Wallis Test Statistic = 245.157
Probability is 0.000 assuming Chi-square distribution with 4 df

TABLE 6-5 (CONTINUED)

ENT

>KRUSKAL VAR00001 * VAR00002
Categorical values encountered during processing are:
VAR00002 (5 levels)
1, 2, 3, 4, 5
Kruskal-Wallis One-Way Analysis of Variance for 10290 cases
Dependent variable is VAR00001
Grouping variable is VAR00002

Group	Count	Rank Sum
Day 1	2058	1.13847E+07
Day 2	2058	1.02974E+07
Day 3	2058	1.04668E+07
Day 4	2058	1.17818E+07
Day 5	2058	9.0164425E+06

Kruskal-Wallis Test Statistic = 255.258
Probability is 0.000 assuming Chi-square distribution with 4 df

the distributions of measures *ASM*, *CON*, and *ENT*, respectively for all the weekdays. To test this possibility we have used K-S test to compare distributions between pairs of two days for all week days. The K-S test along with its results is discussed in the next section.

6.5 Kolmogorov-Smirnov (K-S) Test

The K-S test is a non-parametric test to compare equality of distributions of two groups. Consider two distribution functions $F_A(x)$ and $F_B(x)$, the equivalence of these distribution functions is based on the comparison of the two empirical cumulative distribution functions $\hat{F}_A(x)$ and $\hat{F}_B(x)$. The test statistic is given below.

$$M = \max_x |\hat{F}_A(x) - \hat{F}_B(x)|$$

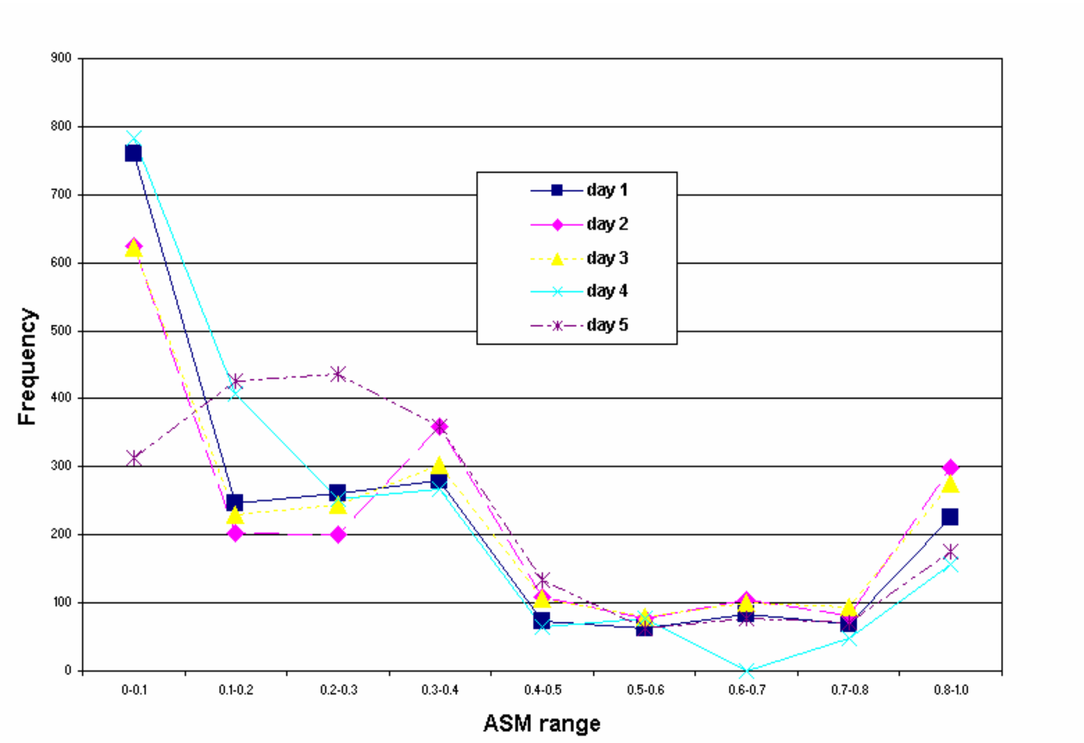


FIGURE 6-1: DISTRIBUTION OF *ASM* FOR ALL WEEKDAYS

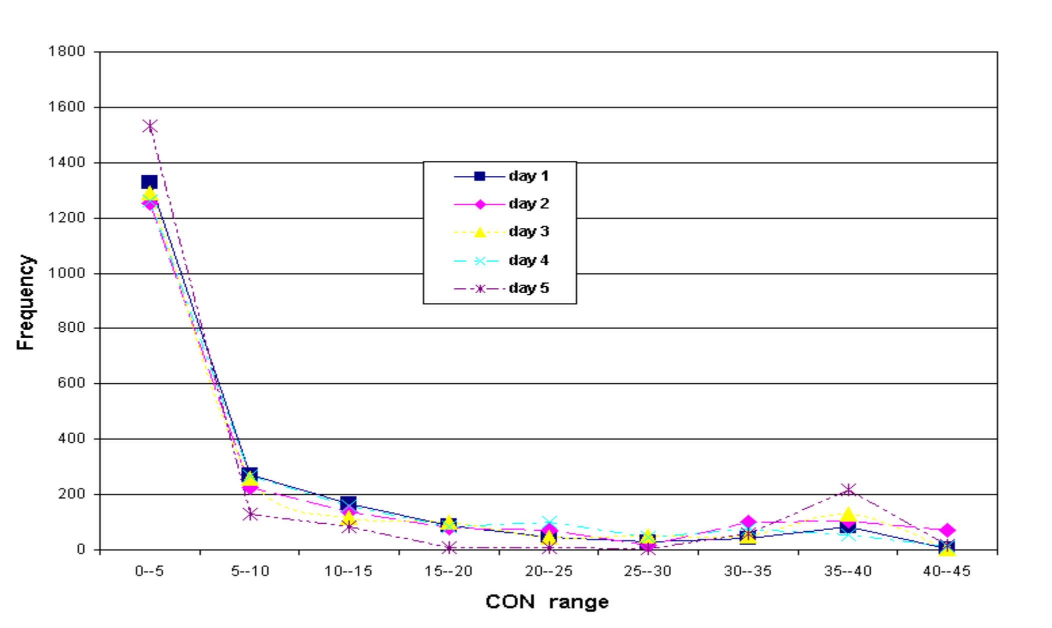


FIGURE 6-2: DISTRIBUTION OF *CON* FOR ALL WEEKDAYS

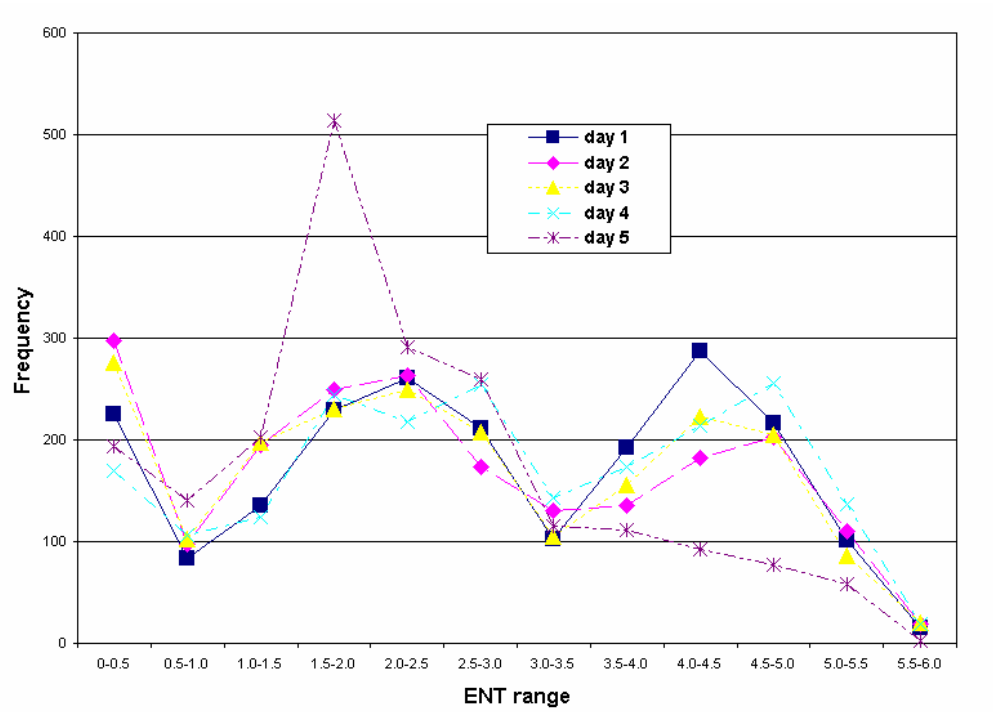


FIGURE 6-3: DISTRIBUTION OF *ENT* FOR ALL WEEKDAYS

Where M is the maximum vertical distance between the two empirical cumulative distribution functions. This measurement is used to test how close together the two distribution functions are. The null hypothesis tested here is that the two groups have same cumulative distribution functions. The M statistic value is compared with the critical point

$$d_a \sqrt{1/n + 1/m}$$

Where n and m are the sizes of the two groups and the value of d_a depends on level of significance a and are given below

a	0.20	0.10	0.05	0.02	0.01
d_a	1.07	1.22	1.36	1.52	1.63

The null hypothesis is rejected when *M* statistic is greater than the critical point and is accepted when *M* statistic is smaller than the critical point.

6.5.1 K-S Test Results

To test the possibility of any two days having same frequency distribution of the measures, we group the weekdays into pairs of days. The null hypothesis is both the groups have the same cumulative frequency distribution of measures. The summary results obtained for all pairs for different measures are given in TABLE 6-6, TABLE 6-7 and TABLE 6-8. In these tables "*N*" represents that there is no statistical evidence or very little probability that both distributions are equal and "*Y*" represents that there is no statistical evidence to reject the null hypothesis, i.e. both distributions have similar frequency distribution. "*NA*" denotes "not applicable". The results of the test statistic obtained for all measures from SYSTAT are given in Appendix B.

From TABLE 6-6, we can see that for *ASM* measure only Day 2 and Day 3 pair have the same frequency distribution and all other pairs have significantly different distributions. For *CON* measure there are no pairs of days for which the frequency distributions are similar. The same can be seen in TABLE 6-7. Similarly, from TABLE 6-8 we can observe that for measure *ENT* there are two pairs of two days (Day 2 & Day 3, and Day 1 & Day 4) for which the frequency distributions are similar.

TABLE 6-6: K-S TEST RESULTS SUMMARY FOR *ASM*

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	*N*	*N*
Day 2	*N*	*NA*	***Y***	*N*	*N*
Day 3	*N*	***Y***	*NA*	*N*	*N*
Day 4	*N*	*N*	*N*	*NA*	*N*
Day 5	*N*	*N*	*N*	*N*	*NA*

TABLE 6-7: K-S TEST RESULTS SUMMARY FOR CON

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	*N*	*N*
Day 2	*N*	*NA*	*N*	*N*	*N*
Day 3	*N*	*N*	*NA*	*N*	*N*
Day 4	*N*	*N*	*N*	*NA*	*N*
Day 5	*N*	*N*	*N*	*N*	*NA*

TABLE 6-8: K-S TEST RESULTS SUMMARY FOR ENT

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	***Y***	*N*
Day 2	*N*	*NA*	***Y***	*N*	*N*
Day 3	*N*	***Y***	*NA*	*N*	*N*
Day 4	***Y***	*N*	*N*	*NA*	*N*
Day 5	*N*	*N*	*N*	*N*	*NA*

Although there are only a couple of pairs of days which have exhibited equality of distributions, it is possible that the means of the distributions may not differ significantly. To check this we use two-sample t-test. The two-sample t-test along with its results is discussed in the following section.

6.6 Two-Sample T-Test

The two-sample t-test is used to test the null hypothesis that the means of two independent groups are equal. The test is implemented with a level of significance *(a)* of 0.05. The null hypothesis is rejected when the test statistic t is greater than the value obtained form t-distribution table for level of significance a and degrees of freedom $df=n_1+n_2-1$, where n_1 is the number of observations in first group and n_2 is the number of observations in second group.

6.6.1 Results of T-Test

The results summary of T-test for measures *ASM*, *CON*, and *ENT* are given in TABLE 6-9, TABLE 6-10, and TABLE 6-11, respectively. In these tables also "*N*" represents that there is statistical evidence that the means of both days differ significantly, "*Y*" represents that there is no statistical evidence that the means of both days differ significantly and *"NA"* represent not applicable. The results obtained from SYSTAT are given in Appendix C.

TABLE 6-9: TWO-SAMPLE T-TEST RESULTS SUMMARY FOR *ASM*

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	*N*	*N*
Day 2	*N*	*NA*	***Y***	*N*	*N*
Day 3	*N*	***Y***	*NA*	*N*	*N*
Day 4	*N*	*N*	*N*	*NA*	*N*
Day 5	*N*	*N*	*N*	*N*	*NA*

TABLE 6-10: TWO-SAMPLE T-TEST RESULTS SUMMARY FOR *CON*

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	*N*	***Y***
Day 2	*N*	*NA*	*N*	*N*	*N*
Day 3	*N*	*N*	*NA*	***Y***	*N*
Day 4	*N*	*N*	***Y***	*NA*	***Y***
Day 5	***Y***	*N*	*N*	***Y***	*NA*

TABLE 6-11: TWO-SAMPLE T-TEST RESULTS SUMMARY FOR *ENT*

	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*NA*	*N*	*N*	*N*	*N*
Day 2	*N*	*NA*	***Y***	*N*	*N*
Day 3	*N*	***Y***	*NA*	*N*	*N*
Day 4	*N*	*N*	*N*	*NA*	*N*
Day 5	*N*	*N*	*N*	*N*	*NA*

From TABLE 6-9 for measure *ASM* we can see that only Day 2 and Day 3 pair have exhibited equality in means. This result is consistent with K-S test results for this measure, which exhibited equality in distributions. For measure *CON* three pairs of days (Day 1 & Day 5, Day 3 & Day 4, and Day 4 & Day 5) exhibited similarity in means, these results are in complete contrast with K-S test results for the same measure, where no pair exhibited any similarity in distributions. The summary of results can be seen in

TABLE 6-10. From TABLE 6-11 for measure *ENT,* we can see that only one pair of days (Day 2 & Day 3) exhibited equality in means. This is also consistent with K-S test results.

6.7 Summary

This chapter discussed various statistical tests used in this study to compare the characteristics of congested regions for 5 selected weekdays. The results of the tests are summarized below:

From the results of Chi-Square test and K-W test it can be said that the distributions of the measures for the week days are significantly different. However, it is possible that distributions of any two different days may behave similarly. To test this we use K-S test by forming pairs of two days from all the weekdays. Also, to check for equality of means between any pairs of days we use two-sample t-test.

From the results of K-S test and two-sample t-test it can be seen that only one pair of days, Day 2 and Day 3 exhibited similarity in distributions and equality in means for all measures except for *CON* measure. Form this we can conclude that the traffic during congestion periods for this pair of days, behaved similarly. We can also notice that, for a couple of pairs of days though the frequency distributions of measures differed significantly, they exhibited equality in means (Examples: Day 1 & Day 5, Day 3 & Day 4, and Day 4 & Day 5 for measure *CON*) and one pair of days exhibited similarity in frequency distributions with significantly different means (Example: Day 1 & Day 4 for measure *ENT*).

7 Summary and Conclusions

7.1 Summary of the Study

The study presented an approach to characterize the spatio-temporal traffic contour maps in a manner similar to feature extraction and textural characterization of digital images. The approach focused on characterizing the spatio-temporal traffic contour maps that quantify characters such as smoothness, regularity, homogeneity, and randomness. There are several techniques available in the field of textural characterization of digital images such as structural, statistical, and combination of structural and statistical measures. However, in this study only the statistical approach was used.

The first order statistical measures and the newly developed second-order statistical measures were used to quantify a total of 2854 contour maps generated per day for five weekdays over a period of 24 hours. Therefore, the total number of contour maps was equal to 14,270. In order to remove the redundancy among the obtained measures, a redundancy check was applied between first order and second order measures and among the second order measures as well. The measures, which were retained after redundancy check, were *AVG, VAR*, *ASM*, *CON*, and *ENT*. Among these *AVG* and *VAR* are first order measures, and *ASM*, *CON*, and *ENT* are second-order measures. Sensitivity of the three second-order measures towards the traffic conditions were studied by dividing speed means incrementally into 5-mph intervals and calculating mean and confidence bounds for each measure. The second order measures were also used in performance assessment and level of service estimation, an approach similar to the one used for level of service on freeway segments in the highway capacity manual.

In order to study the traffic characteristics during congestion on the freeways, we have identified the congestion regions for five weekdays and contour maps which represent the congested conditions were generated. A total of 2058 contour maps were generated each day, totaling 10290 for the five weekdays. These contour maps were studied using second order statistical measures *ASM*, *CON*, and *ENT*, which were retained after redundancy check. Comparisons were made using these second-order statistical measures with statistical tests to check whether distributions of congestion regions of all the weekdays are equal. Two-sample t-test was used to test for equality of means between each pair of two days from all the weekdays.

7.2 Conclusions

This research study presented a new methodology to derive the performance measures from the spatio-temporal traffic contour maps using digital image analysis tools. The new measures are capable of quantifying characters such as smoothness, regularity, homogeneity, and randomness, which is not possible with conventional performance measures. The obtained measures were studied for their sensitivity towards various traffic conditions. Also, level of service criteria were established with the obtained measures in a manner similar to the one mentioned in HCM for level of service of freeways. The following observations were made from this study:

From the sensitivity analysis, the lowest *ASM* was observed in the speed range of 25 to 35 mph, indicating that this speed range indicates highest level of non-uniformity in traffic conditions. Similarly, maximum *CON* was observed in the speed range of 35 to 45 mph. This speed range indicates transient stages in traffic conditions. This speed range also indicates maximum local variations in the traffic conditions. The maximum *ENT* was observed in the speed range of 25 to 30 mph. This speed range indicates the highest randomness in traffic conditions. This was consistent with the other measures. This study used second order statistical measures in performance assessment and level of service estimation by dividing the range of values for each measure into 6 categories from A to F. This method was similar to the one given in HCM for level of service calculation of freeways.

The sensitivity analysis and level of service criteria can be implemented in real time using a standalone module that was developed in this study. The module can also be readily implemented online, which can be applied to any freeway network under study to process traffic contour maps in real-time.

Statistical tests were applied to compare the second order statistical measures for the distributions of congestion regions of all the weekdays. From the results of the tests we can conclude that distributions of congestion regions for all the weekdays behave differently i.e. the characteristics of traffic differed significantly for all weekdays. Although the distribution of measures of all the weekdays differed significantly, it is possible that distributions of any two different days behave similarly. To test these, the weekdays under study are divided into pairs of two days and are tested for similarity in frequency distributions and equality in means. From the results it is observed that only one pair of days: Day 2 and Day 3 consistently exhibited similarity in distributions and equality in means for measures *ASM* and *ENT*. We can also notice that, for a couple of pairs of days though the frequency distributions of measures differed significantly, they exhibited equality in means (Examples: Day 1 & Day 5, Day 3 & Day 4, and Day 4 & Day 5 for measure *CON*) and one pair of days exhibited similarity in frequency distributions with significantly different means (Example: Day 1 & Day 4 for measure *ENT*). However, these conclusions are only true for the freeway network under study and cannot be generalized.

7.3 Future Research

In this study we only used statistical approach to analyze the traffic contour maps. This study can be further extended by using other textural characterization tools such as combination of structural and statistical measures (statistical geometric features), fractals, and morphological operators. Such approaches can enhance the performance measures and can reveal additional properties of traffic behavior which could help better understand the traffic characteristics during transient conditions.

References

Abdullhai, B., H. Porwal, and W. Recker. "Short-Term Freeway Traffic Flow Prediction Using Genetically Optimized Time-Delay-Based Neural Networks." *TRB proceedings,* Transportation Research Board, Washington, D.C., 1999.

Anthony, J. Hayter (2002). Probability and Statistics For Engineers And Scientists. Duxbury.

Banks, J. (2000). Introduction to Transportation Engineering. McGraw Hill.

Bertini, L., M. Leal, and David J. Lovell. (2002). "Generating Performance Measures From Portland's Archived Advanced Traffic Management System Data" 81st Annual Meeting of the Transportation Research Board, Pre-print CD-ROM, Washington D.C.

Chen, C., Karl F. petty, Skabardonis, A., and Jia Z (2001) "Freeway Performance Measurement System: Mining Loop Detector Data", 80th Annual Meeting of the Transportation Research Board, Pre-print CD-Rom, Washington D.C.

Choe, T., A. Skabardonis, and P. Varaiya. (2002). "Freeway Performance Measurement System (PeMS): An Operational Analysis Tool" 81st Annual Meeting of the Transportation Research Board, Pre-print CD-ROM, Washington D.C.

Dahlgren, J., S. Turner, and Reinaldo C. Garcia. (2002). "Collecting, Processing, Archiving and Disseminating Traffic Data to Measure and Improve Traffic Performance" 81st Annual Meeting of the Transportation Research Board, Pre-Print CD-ROM, Washington D.C.

Daniel, K. Papernik., Nanda, D., Robert, O. Cassada, and William H. Morrris. (2000) "An Information Strategy to Enable the Analysis of Performance Measures". 79th Annual Meeting of the Transportation Research Board, Pre-Print CD-ROM, Washington D.C.

Garber, N. and L. Hoel. (1999). Traffic and Highway Engineering. Brooks/Cole Publishing Company.

Gonzalez, C. Rafael. and Woods, E. Richard (2001). Digital Image Processing. Prentice Hall

Highway Capacity Manual. (2000). Transportation Research Board, National Research Council, Washington, DC.

Hsiao C.H., Lin, C.T., and Cassidy M. "Application of Fuzzy Logic and Neural Networks to Automatically Detet Freeway Traffic Incidents." *ASCE Journal of Transportation Engineering,* vol. 120, No. 5., 1994.

Ishak, S. and H. Al-Deek. "Applying Fuzzy ART to Freeway Incident Detection." *Transportation Research Record*. No. 1634, November 1998.

Ishak, S. (2003) "Deriving Performance Measures and Level of Service from Second-Order Statistical Features of Spatiotemporal Traffic Contour Maps. 82nd Annual Meeting of the Transportation Research Board, Pre-print CD-ROM, Washington D.C.

Ishak, S. "Exploring Textural Characteristics of Spatio-temporal Traffic Mpas". To Appear in the ASCE Journal of Transportation Engineering, Vol. 130, No. 1, 2004.

Ishak, S. and H. Al-Deek. "Performance of Automatic Ann-Based Incident Detection on Freeways." *ASCE Journal of Transportation Engineering*, July/August 1999.

James, H. Banks. "Performance Measurement For Traffic Management Systems". 77th Annual Meeting of Transportation Research Board, Pre-Print CD-ROM, Washington D.C.

Khisty, C. and B. Lall. (1990) Transportation Engineering: An Introduction. Prentice Hall.

Kwon, J., Coifman B., and Bickel P. (2000) "Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data", *Transportation Research Record* No. 1717, 120-129.

Liu, H. and H. Motoda. Feature Extraction Construction and Selection, A Data Mining Perspective. Kluwer Academic Publishers, second edition, 2001.

May, A. (1990) Traffic Flow Fundamentals. Prentice Hall.

McNeil, S., W. Wallace, and T. Humphrey. Developing an Information Technology-Oriented Basic Research Program for Surface Transportation Systems. *Report to the NSF and USDOT*, Workshop at the University of Illinois at Chicago, October 15-17, 2000.

Micheli-Tzanakou E. (2000) Supervised and Unsupervised Pattern Recognition, Feature Extraction and Computational Intelligence. CRC Press LLC

Nixon, M. and A. Aguado. (2002) Feature Extraction and Image Processing. Newnes, 2002.

Papacostas, C. and P. Prevedouros. (2001) Transportation Engineering and Planning. Prentice Hall.

Park, D. and l. Rilett. "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks." *Transportation Research Record* N0.1617, Transportation Research Board, Washington D.C., 163-170., 1998.

Rilett, L., D. Park "Direct Foreecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks." *ASCE Journal of Transportation Engineering,* 125(6), 1999.

Ritchie S. and Ruey L.C. "Simulation of Freeway Incident Detection Using Artificial Neural Networks." *Transportation Research C*, vol. 1. No. 3, pp. 203-217, 1993.

Robert, E. Brydia., Shawn, M turner, and Jhy, C. Liu. (1998). "Development of An ITS Data Management System" 77th Annual Meeting of Transportation Research Board, Pre-Print CD-ROM, Washington D.C.

Roess, R., W. McShane, and E. Prassas. (1998) Traffic Engineering. Prentice Hall.

Ruey C. and Ritchie S. "Automated Detection of Lane-Blocking Freeway Incidents using Artificial Neural Networks." *Transportation Research-C* vol. 3, N0. 6, pp. 371-388, 1995.

Ronald, E. Walpole., Raymond, H. Myers., Sharon, L. Myers, and Keying, Ye. (2002). Probability & Statistics For Engineers & Scientists. Prentice Hall.d

Seul, M., L. O'Gorman, and M. Sammon. Practical Algorithms for Image Analysis. Cambridge University Press 2000.

Shaw, T. and Jackson, D. "Reliability Performance Measures for Highway Systems and Segments". (2003). 82nd Annual Meeting of the Transportation Research Board, Pre-print CD-ROM, Washington D.C.

Theodoridis, S. and K. Koutroumbas. (1999) Pattern Recognition. Academic press.

Wright, P. (1996) Highway Engineering. John Wiley & Sons, Inc.

Appendix A: Software module used in the study to develop new second-order statistical measures

```
Private Sub Command1_Click()
    ' Conduct performance measures on traffic data
    Me.Command1.Enabled = False
    ' Parameters Definitions
    Dim inputfile, filename, outputfile As String
    Dim fromstation, tostation As Integer
    Dim fromtime, totime As Integer
    Dim nstations, nminutes As Integer
    Dim timedim, spacedim As Integer
    Dim x As String, i, j, k, i1, j1 As Integer
    Dim direction As Integer
    Dim currenttime, currentstation As Integer
    Dim parameter(3, 2, 3) As Integer
    Dim whichparameter As Integer
    Dim spaceindex, timeindex As Integer
    Dim oldtime, oldstation As Integer
    Dim sum, SumCount As Single
    Dim interval, nintervals, intervaltype  As Integer
    Dim maxval As Integer
    Dim ASMdist As Integer
    Dim Hr, Mi As Integer
    Dim RunID As String
    Dim nbins As Integer
    Dim timelimit As Integer

    '//////////////////////////////////////////////////////////////////////////////////////////
    filename = "d040601.csv"
    RunID = "EB_GL5_MP"
    inputfile = App.Path & "\output\" & filename
    outputfile = App.Path & "\output\TRB\output_" & RunID & "_" & filename
    direction = 1           ' 1 EB      2   WB
    whichparameter = 1      ' 1 Speed   2   Volume  3 Occupancy
    nbins = 10              ' For frequency and cross-classification
    fromstation = 25
    tostation = 40
    fromtime = 7.5 * 60         ' in minutes
    totime = 10 * 60            ' in minutes
    nstations = 3           ' spatial dimension
    nminutes = 5            ' temporal dimension
    timelimit = 10          ' number of time periods at angular variation
    timedim = totime - fromtime + 1
    spacedim = tostation - fromstation + 1
    interval = 5            ' Gray level interval (ex. 5 mph)
```

```
    intervaltype = 0        ' 0 for equal intervals     1 for unequal intervals
    If whichparameter = 1 Then maxval = 60
    ASMdist = 1
    If intervaltype = 0 Then
        nintervals = Int(maxval / interval) + 1
    Else
        nintervals = 5      ' number of unequal intervals
    End If
    ReDim ranges(nintervals, 2) As Single
    If intervaltype = 0 Then
        ' fixed intervals
        For i = 1 To nintervals
            ranges(i, 1) = interval * (i - 1)
            ranges(i, 2) = interval * i
        Next i
    Else
        ' Unequal intervals
        ranges(1, 1) = 0
        ranges(1, 2) = 10
        ranges(2, 1) = 10
        ranges(2, 2) = 20
        ranges(3, 1) = 20
        ranges(3, 2) = 40
        ranges(4, 1) = 40
        ranges(4, 2) = 50
        ranges(5, 1) = 50
        ranges(5, 2) = 75
    End If
    ' ///////////////////////////////////////////////////////////////////////////

    ' Data Extraction from File
    Open inputfile For Input As #1
    ' Sample: 0,0,30,2001-04-02
00:00:30.000,2,68,57,NULL,68,58,NULL,5,3,NULL,4,2,NULL,2,3,NULL,2,1,NULL,40
2000,2,0
    ' hour, minute, second, time, station, ES, ES, ES, WS, WS, WS, EV, EV, EV, WV,
WV, WV, EO, EO, EO, WO, WO, WO, X, X, X
    ReDim Data1(spacedim, timedim, 3), Data2(spacedim, timedim, 3), Data(spacedim,
timedim, 3) As Single     ' three lanes
    Do While Not EOF(1)
        Input #1, x
        currenttime = Val(x) * 60
        Input #1, x
        currenttime = currenttime + Val(x)
        Input #1, x
        Input #1, x     ' ignore
```

```
        Input #1, currentstation

        ' Read data
        For i = 1 To 3
           For j = 1 To 2
              For k = 1 To 3
                 Input #1, parameter(i, j, k)
                 If i = 1 And parameter(i, j, k) > maxval Then parameter(i, j, k) = maxval   '
Set Max Speed
              Next k
           Next j
        Next i

        ' read the last three values and ignore
        Input #1, x, x, x


        ' Store the parameter values in Speed matrix
        spaceindex = currentstation - fromstation + 1
        timeindex = currenttime - fromtime + 1
        If spaceindex > 0 And spaceindex <= spacedim And timeindex > 0 And timeindex
<= timedim Then
           For k = 1 To 3
              Data1(spaceindex, 0, k) = currentstation
              Data1(0, timeindex, k) = currenttime
              Data2(spaceindex, 0, k) = currentstation
              Data2(0, timeindex, k) = currenttime
           Next k
        End If


        ' check to see if this record has the same time and location stamp as previous one
        If oldtime = currenttime And oldstation = currentstation Then
           ' Yes, send data to data2 matrix
            ' Is this record within the range specified?
           If spaceindex > 0 And spaceindex <= spacedim And timeindex > 0 And timeindex
<= timedim Then
              ' Yes, store it
              For k = 1 To 3
                 Data2(spaceindex, timeindex, k) = parameter(whichparameter, direction, k)
              Next k
           End If
        Else
           ' No, send data to data1 matrix
           ' Is this record within the range specified?
```

```
        If spaceindex > 0 And spaceindex <= spacedim And timeindex > 0 And timeindex
<= timedim Then
            ' Yes, store it
            For k = 1 To 3
                Data1(spaceindex, timeindex, k) = parameter(whichparameter, direction, k)
            Next k
        End If
    End If

    ' Ready to move to next record, but store the current time and location first
    ' This is necessary to check if the next record has the same time and location stamp
    oldtime = currenttime
    oldstation = currentstation
Loop
Close #1
' End of data extraction

' Combine data from data1 and data1
For i = 1 To spacedim
    For j = 1 To timedim
        For k = 1 To 3
            If Data1(i, j, k) = 0 Then
                Data(i, j, k) = Data2(i, j, k)
            ElseIf Data2(i, j, k) = 0 Then
                Data(i, j, k) = Data1(i, j, k)
            Else
                Data(i, j, k) = (Data1(i, j, k) + Data2(i, j, k)) / 2
            End If
        Next k
        Data(i, 0, 0) = Data1(i, 0, 1)
        Data(0, j, 0) = Data1(0, j, 1)
    Next j
Next i

' Calculate the average for all lanes
For i = 1 To spacedim
    For j = 1 To timedim
        sum = 0
        SumCount = 0
        For k = 1 To 3
            sum = sum + Data(i, j, k)
            If Data(i, j, k) > 0 Then
                SumCount = SumCount + 1
            End If
            If SumCount > 0 Then Data(i, j, 0) = Format(sum / SumCount, "#.##")
        Next k
```

```
    Next j
Next i

' Interpolate over time
For i = 1 To spacedim
    For j = 1 To timedim
        If Data(i, j, 0) = 0 Then
            ' data is missing at this point
            ' Is this a boundary point?
            If j = 1 Or j = timedim Then
                ' Yes, assume free-flow
                Data(i, j, 0) = 75
            Else
               ' Intermediate point
                ' Begin searching over time for the next valid value
                For k = j + 1 To timedim
                    If Data(i, k, 0) > 0 Then
                        ' Found one at i and k
                        ' Fix the point at i and j
                        Data(i, j, 0) = Data(i, j - 1, 0) + (Data(i, k, 0) - Data(i, j - 1, 0)) / (k - j +
1)
                    End If
                Next k
            End If
        End If
    Next j
Next i

' Interpolate over stations
' Is there a whole station down?
For j = 1 To timedim
    For i = 1 To spacedim
        If Data(i, j, 0) = 0 Then
            ' station i is bad
            ' Is this a boundary station?
            If i = 1 Or i = spacedim Then
                ' Yes, assume free-flow
                Data(i, j, 0) = 75
            Else
                ' intermediate station
                ' Search over space
                For k = i + 1 To spacedim
                    If Data(k, j, 0) > 0 Then
                        ' found one at k, j
                        ' Fix the point at i, j
```

```
                    Data(i, j, 0) = Data(i - 1, j, 0) + (Data(k, j, 0) - Data(i - 1, j, 0)) / (k - i +
1)
                End If
            Next k
        End If
    End If
  Next i
Next j


' Data is now ready for snapshot analysis
' Construct the I matrix from snapshots
ReDim Imatrix(nstations, nminutes) As Integer
ReDim ASMmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim CONmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim IDFmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim AVGmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim VARmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim CORRmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim ENTmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim SUMENTmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As
Single
ReDim DIFFENTmatrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As
Single
ReDim INFO1matrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single
ReDim INFO2matrix(spacedim - nstations + 1, timedim - nminutes + 1, 4) As Single


For i = 1 To spacedim - nstations + 1
  For j = 1 To timedim - nminutes + 1
    ' create a snapshot
    Avg = 0
    For i1 = 1 To nstations
      For j1 = 1 To nminutes
        ' Find which interval
        For k = 1 To nintervals
          If Data(i + i1 - 1, j + j1 - 1, 0) >= ranges(k, 1) And Data(i + i1 - 1, j + j1 -
1, 0) < ranges(k, 2) Then
            Imatrix(i1, j1) = k
          End If
        Next k
        ' Calculate the average of that snapshot from data array
        Avg = Avg + Data(i + i1 - 1, j + j1 - 1, 0)
      Next j1
    Next i1
    Avg = Avg / nstations / nminutes
```

```
        ' Calculate the variance
        var = 0
        For i1 = 1 To nstations
            For j1 = 1 To nminutes
                var = var + (Data(i + i1 - 1, j + j1 - 1, 0) - Avg) ^ 2
            Next j1
        Next i1
        var = var / (nstations * nminutes - 1)


        ' Start calling procedures to calculate the measures of performance

        ' Calculate the PM
         PM Imatrix, nstations, nminutes, filename, maxval, ASMdist, interval, nintervals,
timelimit
        For k = 0 To 4
            ASMmatrix(i, j, k) = ASM(k)
            CONmatrix(i, j, k) = CON(k)
            IDFmatrix(i, j, k) = IDF(k)
            AVGmatrix(i, j, k) = Avg
            VARmatrix(i, j, k) = var
            CORRmatrix(i, j, k) = CORR(k)
            ENTmatrix(i, j, k) = Entropy(k)
            SUMENTmatrix(i, j, k) = SumEntropy(k)
            DIFFENTmatrix(i, j, k) = DiffEntropy(k)
            INFO1matrix(i, j, k) = INFO1(k)
            INFO2matrix(i, j, k) = INFO2(k)
        Next k
    Next j
Next i


' Output to file
Open outputfile For Output As #1
For k = 0 To 4
    Write #1, "TIME/STATION",
    For i = 1 To spacedim - nstations + 1
        Write #1, "AVG" & fromstation + nstations + i - 1,
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, "VAR" & fromstation + nstations + i - 1,
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, "ASM" & fromstation + nstations + i - 1,
    Next i
    For i = 1 To spacedim - nstations + 1
```

```
    Write #1, "CON" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "IDF" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "CORR" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "ENT" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "SUMENT" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "DIFFENT" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "INFOI" & fromstation + nstations + i - 1,
Next i
For i = 1 To spacedim - nstations + 1
    Write #1, "INFOII" & fromstation + nstations + i - 1,
Next i
Write #1,

For j = 1 To timedim - nminutes + 1
    Hr = Int((fromtime + j + nminutes - 1) / 60)
    Mi = fromtime + j + nminutes - 1 - Hr * 60
    Write #1, FormatDateTime(TimeSerial(Hr, Mi, 0), vbShortTime),
    For i = 1 To spacedim - nstations + 1
        Write #1, AVGmatrix(i, j, k),
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, VARmatrix(i, j, k),
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, ASMmatrix(i, j, k),
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, CONmatrix(i, j, k),
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, IDFmatrix(i, j, k),
    Next i
    For i = 1 To spacedim - nstations + 1
        Write #1, CORRmatrix(i, j, k),
```

```
        Next i
        For i = 1 To spacedim - nstations + 1
            Write #1, ENTmatrix(i, j, k),
        Next i
        For i = 1 To spacedim - nstations + 1
            Write #1, SUMENTmatrix(i, j, k),
        Next i
        For i = 1 To spacedim - nstations + 1
            Write #1, DIFFENTmatrix(i, j, k),
        Next i
        For i = 1 To spacedim - nstations + 1
            Write #1, INFO1matrix(i, j, k),
        Next i
        For i = 1 To spacedim - nstations + 1
            Write #1, INFO2matrix(i, j, k),
        Next i
        Write #1,
    Next j
    Write #1,
Next k
Close #1

' Construct histograms
outputfile = App.Path & "\output\Hist_" & RunID & "_" & filename
Open outputfile For Output As #1
Close #1
Histogram "AVG", AVGmatrix, nbins, outputfile
Histogram "VAR", VARmatrix, nbins, outputfile
Histogram "ASM", ASMmatrix, nbins, outputfile
Histogram "CON", CONmatrix, nbins, outputfile
Histogram "IDF", IDFmatrix, nbins, outputfile
Histogram "CORR", CORRmatrix, nbins, outputfile
Histogram "ENT", ENTmatrix, nbins, outputfile
Histogram "SUMENT", SUMENTmatrix, nbins, outputfile
Histogram "DIFFENT", DIFFENTmatrix, nbins, outputfile
Histogram "INFO1", INFO1matrix, nbins, outputfile
'Histogram "INFO2", INFO2matrix, nbins, outputfile

' Construct cross-classification histograms
outputfile = App.Path & "\output\CrossHist_" & RunID & "_" & filename
Open outputfile For Output As #1
Close #1
CrossHistogram "AVG", "ASM", AVGmatrix, ASMmatrix, nbins, outputfile
CrossHistogram "VAR", "ASM", VARmatrix, ASMmatrix, nbins, outputfile
CrossHistogram "AVG", "CON", AVGmatrix, CONmatrix, nbins, outputfile
CrossHistogram "VAR", "CON", VARmatrix, CONmatrix, nbins, outputfile
```

```
    CrossHistogram "AVG", "IDF", AVGmatrix, IDFmatrix, nbins, outputfile
    CrossHistogram "VAR", "IDF", VARmatrix, IDFmatrix, nbins, outputfile
    CrossHistogram "AVG", "CORR", AVGmatrix, CORRmatrix, nbins, outputfile
    CrossHistogram "VAR", "CORR", VARmatrix, CORRmatrix, nbins, outputfile
    CrossHistogram "AVG", "ENT", AVGmatrix, ENTmatrix, nbins, outputfile
    CrossHistogram "VAR", "ENT", VARmatrix, ENTmatrix, nbins, outputfile
    CrossHistogram "AVG", "SUMENT", AVGmatrix, SUMENTmatrix, nbins, outputfile
    CrossHistogram "VAR", "SUMENT", VARmatrix, SUMENTmatrix, nbins, outputfile
    CrossHistogram "AVG", "DIFFENT", AVGmatrix, DIFFENTmatrix, nbins, outputfile
    CrossHistogram "VAR", "DIFFENT", VARmatrix, DIFFENTmatrix, nbins, outputfile
    CrossHistogram "AVG", "INFO1", AVGmatrix, INFO1matrix, nbins, outputfile
    CrossHistogram "VAR", "INFO1", VARmatrix, INFO1matrix, nbins, outputfile
    'CrossHistogram "AVG", "INFO2", AVGmatrix, INFO2matrix, nbins, outputfile
    'CrossHistogram "VAR", "INFO2", VARmatrix, INFO2matrix, nbins, outputfile




    End

End Sub

Private Sub Command2_Click()
    End
End Sub
```

Appendix B: K-S Test output from SYSTAT

ASM

Day 1 and Day 2
KS VAR00001 * GROUP12
Categorical values encountered during processing are:
GROUP12 (2 levels)
1, 2
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	1	2
1	0.00	
2	0.124	0.00

Two-sided probabilities

	1	2
1	-	
2	0.000	-

Day 1 and Day 3
KS VAR00001 * GROUP13
Categorical values encountered during processing are:
GROUP13 (2 levels)
1, 3
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	1	3
1	0.00	
3	0.101	0.00

Two-sided probabilities

	1	3
1	-	
3	0.000	-

Day 1 and Day 4
KS VAR00001 * GROUP14
Categorical values encountered during processing are:
GROUP14 (2 levels)
1, 4
Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	4
1	0.00	
4	0.111	0.00

Two-sided probabilities

	1	4
1	-	
4	0.000	-

Day 1 and Day 5

KS VAR00001 * GROUP15

Categorical values encountered during processing are:

GROUP15 (2 levels)

1, 5

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	5
1	0.00	
5	0.22	0.00

Two-sided probabilities

	1	5
1	-	
5	0.000	-

Day 2 and Day 3

KS VAR00001 * GROUP23

Categorical values encountered during processing are:

GROUP23 (2 levels)

2, 3

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	2	3
2	0.00	
3	0.034	0.00

Two-sided probabilities

	2	3
2	-	
3	0.198	-

Day 2 and Day 4

KS VAR00001 * GROUP24

Categorical values encountered during processing are:
GROUP24 (2 levels)
2, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	4
2	0.00	
4	0.207	0.00

Two-sided probabilities

	2	4
2	-	
4	0.000	-

Day 2 and Day 5
KS VAR00001 * GROUP25
Categorical values encountered during processing are:
GROUP25 (2 levels)
2, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	5
2	0.00	
5	0.155	0.00

Two-sided probabilities

	2	5
2	-	
5	0.000	-

Day 3 and Day 4
KS VAR00001 * GROUP34
Categorical values encountered during processing are:
GROUP34 (2 levels)
3, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	4
3	0.00	
4	0.155	0.00

Two-sided probabilities

	3	4
3	-	
4	0.000	-

Day 3 and Day 5
KS VAR00001 * GROUP35
Categorical values encountered during processing are:
GROUP35 (2 levels)
3, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	5
3	0.00	
5	0.155	0.00

Two-sided probabilities

	3	5
3	-	
5	0.000	-

Day 4 and Day 5
KS VAR00001 * GROUP45
Categorical values encountered during processing are:
GROUP45 (2 levels)
4, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	4	5
4	0.00	
5	0.249	0.00

Two-sided probabilities

	4	5
4	-	
5	0.000	-

CON

Day 1 and Day 2
KS VAR00001 * GROUP12
Categorical values encountered during processing are:
GROUP12 (2 levels)
1, 2
Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	2
1	0.00	
2	0.086	0.00

Two-sided probabilities

	1	2
1	-	
2	0.000	-

Day 1 and Day 3

KS VAR00001 * GROUP13

Categorical values encountered during processing are:

GROUP13 (2 levels)

1, 3

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	3
1	0.00	
3	0.084	0.00

Two-sided probabilities

	1	3
1	-	
3	0.000	-

Day 1 and Day 4

KS VAR00001 * GROUP14

Categorical values encountered during processing are:

GROUP14 (2 levels)

1, 4

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	4
1	0.00	
4	0.053	0.00

Two-sided probabilities

	1	4
1	-	
4	0.006	-

Day 1 and Day 5

KS VAR00001 * GROUP15
Categorical values encountered during processing are:
GROUP15 (2 levels)
1, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	1	5
1	0.00	
5	0.215	0.00

Two-sided probabilities

	1	5
1	-	
5	0.000	-

Day 2 and Day 3
KS VAR00001 * GROUP23
Categorical values encountered during processing are:
GROUP23 (2 levels)
2, 3
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	3
2	0.00	
3	0.045	0.00

Two-sided probabilities

	2	3
2	-	
3	0.03	-

Day 2 and Day 4
KS VAR00001 * GROUP24
Categorical values encountered during processing are:
GROUP24 (2 levels)
2, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	4
2	0.00	
4	0.102	0.00

Two-sided probabilities

	2	4
2	-	
4	0.000	-

Day 2 and Day 5
KS VAR00001 * GROUP25
Categorical values encountered during processing are:
GROUP25 (2 levels)
2, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	5
2	0.00	
5	0.191	0.00

Two-sided probabilities

	2	5
2	-	
5	0.000	-

Day 3 and Day 4
KS VAR00001 * GROUP34
Categorical values encountered during processing are:
GROUP34 (2 levels)
3, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	4
3	0.00	
4	0.093	0.00

Two-sided probabilities

	3	4
3	-	
4	0.000	-

Day 3 and Day 5
KS VAR00001 * GROUP35
Categorical values encountered during processing are:
GROUP35 (2 levels)
3, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	5
3	0.00	
5	0.186	0.00

Two-sided probabilities

	3	5
3	-	
5	0.000	-

Day 4 and Day 5
KS VAR00001 * GROUP45
Categorical values encountered during processing are:
GROUP45 (2 levels)
4, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	4	5
4	0.00	
5	0.222	0.00

Two-sided probabilities

	4	5
4	-	
5	0.000	-

ENT

Day 1 and Day 2
KS VAR00001 * GROUP12
Categorical values encountered during processing are:
GROUP12 (2 levels)
1, 2
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	1	2
1	0.00	
2	0.107	0.00

Two-sided probabilities

	1	2
1	-	
2	0.000	-

Day 1 and Day 3
KS VAR00001 * GROUP13
Categorical values encountered during processing are:
GROUP13 (2 levels)
1, 3

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	3
1	0.00	
3	0.085	0.00

Two-sided probabilities

	1	3
1	-	
3	0.000	-

Day 1 and Day 4

KS VAR00001 * GROUP14

Categorical values encountered during processing are:

GROUP14 (2 levels)

1, 4

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	4
1	0.00	
4	0.042	0.00

Two-sided probabilities

	1	4
1	-	
4	0.055	-

Day 1 and Day 5

KS VAR00001 * GROUP15

Categorical values encountered during processing are:

GROUP15 (2 levels)

1, 5

Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	5
1	0.00	
5	0.235	0.00

Two-sided probabilities

	1	5
1	-	
5	0.000	-

Day 2 and Day 3
KS VAR00001 * GROUP23
Categorical values encountered during processing are:
GROUP23 (2 levels)
2, 3
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	3
2	0.00	
3	0.031	0.00

Two-sided probabilities

	2	3
2	-	
3	0.290	-

Day 2 and Day 4
KS VAR00001 * GROUP24
Categorical values encountered during processing are:
GROUP24 (2 levels)
2, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	4
2	0.00	
4	0.135	0.00

Two-sided probabilities

	2	4
2	-	
4	0.000	-

Day 2 and Day 5
KS VAR00001 * GROUP25
Categorical values encountered during processing are:
GROUP25 (2 levels)
2, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	2	5
2	0.00	
5	0.169	0.00

Two-sided probabilities

	2	5
2	-	
5	0.000	-

Day 3 and Day 4
KS VAR00001 * GROUP34
Categorical values encountered during processing are:
GROUP34 (2 levels)
3, 4
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	4
3	0.00	
4	0.109	0.00

Two-sided probabilities

	3	4
3	-	
4	0.000	-

Day 3 and Day 5
KS VAR00001 * GROUP35
Categorical values encountered during processing are:
GROUP35 (2 levels)
3, 5
Kolmogorov-Smirnov Two Sample Test results
Maximum differences for pairs of groups

	3	5
3	0.00	
5	0.178	0.00

Two-sided probabilities

	3	5
3	-	
5	0.000	-

Day 4 and Day 5
KS VAR00001 * GROUP45
Categorical values encountered during processing are:
GROUP45 (2 levels)
4, 5
Kolmogorov-Smirnov Two Sample Test results

Maximum differences for pairs of groups

	1	2
1	0.00	
2	0.247	0.00

Two-sided probabilities

	1	2
1	-	
2	0.000	-

Appendix C: Two –sample T-Test results from SYSTAT

ASM

Day 1 and Day 2
TEST VAR00001 * GROUP12
Two-sample t test on VAR00001 grouped by GROUP12

Group	N	Mean	SD
1	2058	0.311	0.308
2	2058	0.37	0.327

Separate Variance t = -5.916 df = 4098.6 Prob = 0.000
Difference in Means = -0.059 95.00% CI = -0.078 to -0.039

Pooled Variance t = -5.916 df = 4114 Prob = 0.000
Difference in Means = -0.059 95.00% CI = -0.078 to -0.039

Day 1 and Day 3
TEST VAR00001 * GROUP13
Two-sample t test on VAR00001 grouped by GROUP13

Group	N	Mean	SD
1	2058	0.311	0.308
3	2058	0.362	0.322

Separate Variance t = -5.147 df = 4104.9 Prob = 0.000
Difference in Means = -0.051 95.00% CI = -0.070 to -0.031

Pooled Variance t = -5.147 df = 4114 Prob = 0.000
Difference in Means = -0.051 95.00% CI = -0.070 to -0.031

Day 1 and Day 4
TEST VAR00001 * GROUP14
Two-sample t test on VAR00001 grouped by GROUP14

Group	N	Mean	SD
1	2058	0.311	0.308
4	2058	0.257	0.273

Separate Variance t = 5.915df = 4055.8 Prob = 0.000
Difference in Means = 0.054 95.00% CI = 0.036 to 0.071

Pooled Variance t = 5.915df = 4114 Prob = 0.000
Difference in Means = 0.054 95.00% CI = 0.036 to 0.071

Day 1 and Day 5
TEST VAR00001 * GROUP15
Two-sample t test on VAR00001 grouped by GROUP15

Group	N	Mean	SD
1	2058	0.311	0.308
5	2058	0.337	0.268

Separate Variance t = -2.878 df = 4038.3 Prob = 0.004
Difference in Means = -0.051 95.00% CI = -0.044 to -0.008

Pooled Variance t = -5.147 df = 4114 Prob = 0.000
Difference in Means = -0.051 95.00% CI = -0.044 to -0.008

Day 2 and Day 3
TEST VAR00001 * GROUP23
Two-sample t test on VAR00001 grouped by GROUP23

Group	N	Mean	SD
2	2058	0.37	0.327
3	2058	0.362	0.322

Separate Variance t = 0.789 df = 4113.2 Prob = 0.430
Difference in Means = -0.008 95.00% CI = -0.012 to 0.028

Pooled Variance t = 0.789 df = 4114 Prob = 0.430
Difference in Means = 0.008 95.00% CI = -0.012 to -0.028

Day 2 and Day 4
TEST VAR00001 * GROUP24
Two-sample t test on VAR00001 grouped by GROUP24

Group	N	Mean	SD
2	2058	0.37	0.327
4	2058	0.257	0.273

Separate Variance t = 11.947 df = 3985.0 Prob = 0.000
Difference in Means = 0.112 95.00% CI = 0.094 to 0.131

Pooled Variance t = 11.947 df = 4114 Prob = 0.000
Difference in Means = 0.112 95.00% CI = 0.094 to 0.131

Day 2 and Day 5
TEST VAR00001 * GROUP25
Two-sample t test on VAR00001 grouped by GROUP25

Group	N	Mean	SD
2	2058	0.37	0.327
5	2058	0.337	0.268

Separate Variance t = 3.505 df = 3960.8 Prob = 0.000
Difference in Means = 0.033 95.00% CI = 0.014 to 0.051

Pooled Variance t = 3.505 df = 4114 Prob = 0.000
Difference in Means = 0.033 95.00% CI = 0.014 to 0.051

Day 3 and Day 4

TEST VAR00001 * GROUP34
Two-sample t test on VAR00001 grouped by GROUP34

Group	N	Mean	SD
3	2058	0.362	0.322
4	2058	0.257	0.273

Separate Variance t = 11.189 df = 4003.6 Prob = 0.000
Difference in Means = 0.104 95.00% CI = 0.086 to 0.122

Pooled Variance t = 11.189 df = 4114 Prob = 0.000
Difference in Means = 0.104 95.00% CI = 0.086 to 0.122

Day 3 and Day 5

TEST VAR00001 * GROUP35
Two-sample t test on VAR00001 grouped by GROUP35

Group	N	Mean	SD
3	2058	0.362	0.322
5	2058	0.337	0.268

Separate Variance t = 2.670 df = 3980.7 Prob = 0.008
Difference in Means = 0.025 95.00% CI = 0.007 to 0.043

Pooled Variance t = 2.670 df = 4114 Prob = 0.008
Difference in Means = 0.025 95.00% CI = 0.007 to 0.043

Day 4 and Day 5

TEST VAR00001 * GROUP45
Two-sample t test on VAR00001 grouped by GROUP45

Group	N	Mean	SD
4	2058	0.257	0.273
5	2058	0.337	0.268

Separate Variance t = -9.430 df = 4112.8 Prob = 0.000
Difference in Means = -0.079 95.00% CI = -0.096 to -0.063

Pooled Variance t = -9.430 df = 4114 Prob = 0.000
Difference in Means = -0.079 95.00% CI = -0.096 to -0.063

CON

Day 1 and Day 2
TEST VAR00001 * GROUP12
Two-sample t test on VAR00001 grouped by GROUP12

Group	N	Mean	SD
1	2058	6.569	9.76
2	2058	8.704	12.643

Separate Variance t = -6.066 df = 3866.1 Prob = 0.000
Difference in Means = -2.136 95.00% CI = -2.826 to -1.445

Pooled Variance t = -6.066 df = 4114 Prob = 0.000
Difference in Means = -2.136 95.00% CI = -2.826 to -1.445

Day 1 and Day 3
TEST VAR00001 * GROUP13
Two-sample t test on VAR00001 grouped by GROUP13

Group	N	Mean	SD
1	2058	6.569	9.76
3	2058	7.558	11.228

Separate Variance t = -3.017 df = 4035.8 Prob = 0.003
Difference in Means = -0.989 95.00% CI = -1.632 to -0.346

Pooled Variance t = -3.017 df = 4114 Prob = 0.000
Difference in Means = -0.989 95.00% CI = -1.632 to -0.346

Day 1 and Day 4
TEST VAR00001 * GROUP14
Two-sample t test on VAR00001 grouped by GROUP14

Group	N	Mean	SD
1	2058	6.569	9.76
4	2058	7.322	10.154

Separate Variance t = -2.428df = 4107.6 Prob = 0.015

Difference in Means = -0.754 95.00% CI = -1.362 to -0.145

Pooled Variance t = -2.428df = 4114 Prob = 0.015
Difference in Means = -0.754 95.00% CI = -1.362 to -0.145

Day 1 and Day 5
TEST VAR00001 * GROUP15
Two-sample t test on VAR00001 grouped by GROUP15

Group	N	Mean	SD
1	2058	6.569	9.76
5	2058	6.809	12.64

Separate Variance t = -0.683 df = 3866.6 Prob = 0.494
Difference in Means = -0.241 95.00% CI = -0.931 to 0.450

Pooled Variance t = -0.683 df = 4114 Prob = 0.494
Difference in Means = -0.241 95.00% CI = -0.931 to 0.450

Day 2 and Day 3
TEST VAR00001 * GROUP23
Two-sample t test on VAR00001 grouped by GROUP23

Group	N	Mean	SD
2	2058	8.704	12.643
3	2058	7.558	11.228

Separate Variance t = 3.076 df = 4057.3 Prob = 0.002
Difference in Means = 1.146 95.00% CI = 0.416 to 1.887

Pooled Variance t = 3.076 df = 4114 Prob = 0.002
Difference in Means = 1.146 95.00% CI = 0.416 to 1.887

Day 2 and Day 4
TEST VAR00001 * GROUP24
Two-sample t test on VAR00001 grouped by GROUP24

Group	N	Mean	SD
2	2058	8.704	12.643
4	2058	7.322	10.154

Separate Variance t = 3.866 df = 3930.8 Prob = 0.000
Difference in Means = 1.382 95.00% CI = 0.681 to 2.083

Pooled Variance t = 3.866 df = 4114 Prob = 0.000
Difference in Means = 1.382 95.00% CI = 0.681 to 2.083

Day 2 and Day 5
TEST VAR00001 * GROUP25
Two-sample t test on VAR00001 grouped by GROUP25

Group	N	Mean	SD
2	2058	8.704	12.643
5	2058	6.809	12.64

Separate Variance t = 4.809 df = 4114.0 Prob = 0.000
Difference in Means = 1.895 95.00% CI = 1.122 to 2.668

Pooled Variance t = 4.809 df = 4114 Prob = 0.000
Difference in Means = 1.895 95.00% CI = 1.122 to 2.668

Day 3 and Day 4
TEST VAR00001 * GROUP34
Two-sample t test on VAR00001 grouped by GROUP34

Group	N	Mean	SD
3	2058	7.558	11.228
4	2058	7.322	10.154

Separate Variance t = 0.706 df = 4073.6 Prob = 0.480
Difference in Means = 0.236 95.00% CI = -0.419 to 0.890

Pooled Variance t = 0.706 df = 4114 Prob = 0.480
Difference in Means = 0.236 95.00% CI = -0.419 to 0.890

Day 3 and Day 5

TEST VAR00001 * GROUP35
Two-sample t test on VAR00001 grouped by GROUP35

Group	N	Mean	SD
3	2058	7.558	11.228
5	2058	6.809	12.64

Separate Variance t = 2.009 df = 4057.7 Prob = 0.045
Difference in Means = 0.749 95.00% CI = 0.018 to 1.479

Pooled Variance t = 2.009 df = 4114 Prob = 0.045
Difference in Means = 0.749 95.00% CI = 0.018 to 1.479

Day 4 and Day 5

TEST VAR00001 * GROUP45
Two-sample t test on VAR00001 grouped by GROUP45

Group	N	Mean	SD
4	2058	7.322	10.154
5	2058	6.809	12.64

Separate Variance t = 1.436 df = 3931.3 Prob = 0.151
Difference in Means = 0.513 95.00% CI = -0.188 to 1.214

Pooled Variance t = 1.436 df = 4114 Prob = 0.151
Difference in Means = 0.513 95.00% CI = -0.188 to 1.214

ENT

Day 1 and Day 2
TEST VAR00001 * GROUP12
Two-sample t test on VAR00001 grouped by GROUP12

Group	N	Mean	SD
1	2058	2.777	1.578
2	2058	2.507	1.654

Separate Variance t = 5.362 df = 4105.1 Prob = 0.000
Difference in Means = 0.270 95.00% CI = 1.171 to 0.369

Pooled Variance t = 5.362 df = 4114 Prob = 0.000
Difference in Means = 0.270 95.00% CI = 1.171 to 0.369

Day 1 and Day 3
TEST VAR00001 * GROUP13
Two-sample t test on VAR00001 grouped by GROUP13

Group	N	Mean	SD
1	2058	2.777	1.578
3	2058	2.548	1.626

Separate Variance t = 4.578 df = 4110.4 Prob = 0.000
Difference in Means = 0.229 95.00% CI = 0.131 to 0.327

Pooled Variance t = 4.578 df = 4114 Prob = 0.000
Difference in Means = 0.229 95.00% CI = 0.131 to 0.327

Day 1 and Day 4
TEST VAR00001 * GROUP14
Two-sample t test on VAR00001 grouped by GROUP14

Group	N	Mean	SD
1	2058	2.777	1.578
4	2058	2.886	1.546

Separate Variance t = -2.254df = 4112.2 Prob = 0.024
Difference in Means = -0.110 95.00% CI = -0.205 to -0.014

Pooled Variance t = -2.254 df = 4114 Prob = 0.024
Difference in Means = -0.110 95.00% CI = -0.205 to -0.014

Day 1 and Day 5
TEST VAR00001 * GROUP15
Two-sample t test on VAR00001 grouped by GROUP15

Group	N	Mean	SD
1	2058	2.777	1.578
5	2058	2.198	1.288

Separate Variance t = 12.888 df = 3954.6 Prob = 0.000
Difference in Means = 0.579 95.00% CI = 0.491 to 0.667

Pooled Variance t = 12.888 df = 4114 Prob = 0.000
Difference in Means = 0.579 95.00% CI = 0.491 to 0.667

Day 2 and Day 3
TEST VAR00001 * GROUP23
Two-sample t test on VAR00001 grouped by GROUP23

Group	N	Mean	SD
2	2058	2.507	1.654
3	2058	2.548	1.626

Separate Variance t = -0.812 df = 4112.8 Prob = 0.417
Difference in Means = -0.042 95.00% CI = -0.142 to 0.059

Pooled Variance t = -0.812 df = 4114 Prob = 0.002
Difference in Means = -0.042 95.00% CI = -0.142 to 0.059

Day 2 and Day 4
TEST VAR00001 * GROUP24
Two-sample t test on VAR00001 grouped by GROUP24

Group	N	Mean	SD
2	2058	2.507	1.654
4	2058	2.886	1.546

Separate Variance t = -7.615 df = 4095.4 Prob = 0.000
Difference in Means = -0.380 95.00% CI = -0.478 to -0.283

Pooled Variance t = -7.615 df = 4114 Prob = 0.000
Difference in Means = -0.380 95.00% CI = -0.478 to -0.283

Day 2 and Day 5

TEST VAR00001 * GROUP25

Two-sample t test on VAR00001 grouped by GROUP25

Group	N	Mean	SD
2	2058	2.507	1.654
5	2058	2.198	1.288

Separate Variance t = 6.677 df = 3880.0 Prob = 0.000
Difference in Means = 0.309 95.00% CI = 0.218 to 0.399

Pooled Variance t = 6.677 df = 4114 Prob = 0.000
Difference in Means = 0.309 95.00% CI = 0.218 to 0.399

Day 3 and Day 4

TEST VAR00001 * GROUP34

Two-sample t test on VAR00001 grouped by GROUP34

Group	N	Mean	SD
3	2058	2.548	1.626
4	2058	2.886	1.546

Separate Variance t = -6.844 df = 4103.6 Prob = 0.000
Difference in Means = -0.338 95.00% CI = -0.435 to -0.242

Pooled Variance t = -6.844 df = 4114 Prob = 0.000
Difference in Means = -0.338 95.00% CI = -0.435 to -0.242

Day 3 and Day 5

TEST VAR00001 * GROUP35

Two-sample t test on VAR00001 grouped by GROUP35

Group	N	Mean	SD
3	2058	2.548	1.626
5	2058	2.198	1.288

Separate Variance t = 7.656 df = 3908.7 Prob = 0.000
Difference in Means = 0.350 95.00% CI = 0.260 to 0.440

Pooled Variance t = 7.656 df = 4114 Prob = 0.000
Difference in Means = 0.350 95.00% CI = 0.260 to 0.440

Day 4 and Day 5

TEST VAR00001 * GROUP45

Two-sample t test on VAR00001 grouped by GROUP45

Group	N	Mean	SD
4	2058	2.886	1.546
5	2058	2.198	1.288

Separate Variance t = 15.525 df = 3984.3 Prob = 0.000
Difference in Means = 0.689 95.00% CI = 0.602 to 0.775

Pooled Variance t = 15.525 df = 4114 Prob = 0.000
Difference in Means = 0.689 95.00% CI = 0.602 to 0.775

Vita

Prashanth Kotha was born on March 2, 1980, in Hanamkonda, a small city in India. He obtained his Bachelor of Technology degree in Civil Engineering in 2001 from Kakatiya University, India. He joined the master's program in the Louisiana State University with major in transportation engineering. In December 2003 he will receive the degree of Master of Science in Civil Engineering.