



## USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No 006OY01

### **Research and Education from a Smart Campus Transit Laboratory**

By

Mark R. McCord, Principal Investigator  
Professor of Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mccord.2@osu.edu

and

Rabi G. Mishalani, Co-Principal Investigator  
Associate Professor of Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mishalani@osu.edu

and

Prem Goel, Co-Principal Investigator  
Professor of Statistics  
Ohio State University  
goel.1@osu.edu

Report Submission Date: October 15, 2009



## **DISCLAIMER**

Partial funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT07-G-005 of the U.S. Department of Transportation, Research and Innovative Technology Administration (RITA), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.



## USDOT Region V Regional University Transportation Center Final Report

# TECHNICAL SUMMARY

NEXTRANS Project No 006OY01

Final Report, October 2009

## Research and Education from a Smart Campus Transit Laboratory

### Introduction

For approximately a decade, members of the project team monitored Ohio State University (OSU) campus buses serving four million passengers annually with a “homemade” GPS-based automatic vehicle location (AVL), communications, and information system called BLIS (Bus Location and Information System). We supplied regular, system-wide performance reports to OSU’s Transportation and Parking (T&P) Campus Area Bus Service (CABS), responded to special requests from CABS (generally resulting from customer complaints about service), and conducted research studies that exploited the BLIS data we archived. These research and outreach activities, along with the BLIS archived data, formed the first generation of the OSU Campus Transit Lab (CTL).

Through a joint effort of T&P, the OSU College of Engineering, OSU’s Department of Civil and Environmental Engineering and Geodetic Science, and Clever Devices, Inc., BLIS is being replaced with an advanced, commercial-grade “Smart Bus” system. Clever Devices has equipped large public bus agencies with such systems, but this is the company’s first implementation for a campus bus service.

This substantially upgraded asset and the partnerships surrounding it offer the opportunity to develop the CTL into a unique, valuable, and recognized living lab that can simultaneously support innovative public transportation research, education, and outreach. Obtaining this status will require sustained development that produces benefits to the multiple collaborating stakeholders along the way while keeping them aware of the long term potential. To assist in this sustained development, it is necessary to conduct a multi-faceted effort that implements and manages the underlying physical and institutional infrastructure of the Smart Bus system while simultaneously producing research, educational, and outreach results that exploit Smart Bus data and the CTL.

### Findings

During the reporting period, we

- contributed to the implementation of the “Smart Bus” system on OSU campus buses: This system is now installed. Bugs are still being identified and eliminated, but the system is

functionally in an operational mode, and early indications of user and operator satisfaction appear good.

- developed processes to make Smart Bus automatic vehicle location (AVL) and automated passenger counter (APC) data available in useful formats for research, education, and outreach applications: Further refinements will be required for future applications, but we now have the means to pre-process large quantities of AVL and APC data into formats that can be used by multiple users for a variety of purposes.
- pre-processed a first wave of AVL and APC data and used the data for multiple applications: We used our software to pre-process Smart Bus data into data that served as input to produce empirical measures used in the project reported on here and supported several tasks on a Federal Transit Administration project.
- conceived of and validated innovative measures indicating bus passenger travel patterns that are derived from AVL and APC data and which can be monitored on an ongoing basis in collaboration with OSU T&P: APC data can be used to estimate bus passenger origin-destination (OD) flows. Because the APC data are received on a regular (daily) basis, the OD flows and measures that can be derived from the estimated OD flows and the AVL data (e.g., OD flows by time-of-day and day-of-week, passenger trip distance distributions, expected time on bus conditional on boarding or alighting stop) can be monitored over time. Through field tests and familiarity with CTL routes, we partially validated the measures we produced from the Smart Bus data.
- conducted multiple research studies related to the use of bus AVL and APC data: We improved a method we had previously developed for matching AVL-based bus trajectories to bus schedules by incorporating considerations of bus operating policies. Empirical results using CTL data demonstrated the superiority of the refined method. We also developed an innovative methodological design that we applied to empirical data collected on a CTL route to assess the performance of an easy-to-implement procedure for estimating bus passenger OD flows from available APC data. We found that the procedure worked surprisingly well in our study. In addition, we built a simulation tool based on one of the CTL routes and applied the tool to compare the performance of distance-based to time-based AVL data sampling in terms of the accuracy of estimating bus dwell times. We found that distance-based sampling performed markedly better than time-based sampling in this application.
- implemented new modules focused on the CTL AVL and APC data in two transportation courses: One course is a large (over 100 students), required course for Civil Engineering students, the majority of whom are non-transportation majors. The second course is a smaller (approximately 15 students), elective course for graduate students with a transportation focus. The new modules and quantitative exercises using CTL data exposed the students to the CTL and to the advantages of AVL and APC technologies by addressing practical applications in a familiar and

observable setting. The apparent benefits to the students and to the instructors motivate us to develop additional ways to use CTL data and applications in courses.

- conducted a first wave survey of campus transit bus users' and nonusers' perceptions of OSU's CABS: The motivation for conducting the survey was to develop benchmark information for assessing changes in perceptions, attitudes, and awareness of OSU bus transit service that may be attributable to the implementation of the Smart Bus system. Response rates by demographic group were very good, and we plan to conduct the second wave after use of the Smart Bus-based passenger information system known as TRIP (Transportation Route Information Program) enters steady state. Nevertheless, some of the first wave survey results, such as perception toward the various elements of CABS by demographic group, are already producing information of interest to T&P administrators. Moreover, other results, such as the recognition of the positive impact of a bus system on the environment and on reduced traffic, and the differences in this recognition among different demographic groups, are of general interest to the transit and multimodal transportation community.

## Recommendations

We believe that the multi-thrust approach we undertook during the reporting period was productive in contributing to the sustained development that will establish the OSU Campus Transit Lab (CTL) as a unique, recognized, and valuable infrastructure for research, education, and outreach. Additional development will continue to be required, and we believe that it would be beneficial to proceed in a similarly multi-faceted approach devoted to

- developing the means to collect, process, and make AVL and APC data accessible to multiple researchers and educators on a routine basis,
- using the data to support multiple public transportation related research and educational activities sponsored inside and outside of NEXTRANS,
- conducting additional research studies related to improved bus transit planning and operations that can occur through innovative uses of these data, and
- ongoing monitoring of the bus system in collaboration with OSU T&P to provide benefits to major stakeholders (e.g., to T&P, in terms of better understanding of the service it is providing, and to Clever Devices, in terms of developing new products that can be derived from its technologies), ground research and educational project activities in actual operations, and foster the generation of new research ideas.

## Contacts

*For more information:*

### **Mark R. McCord**

Principal Investigator  
Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mccord.2@osu.edu

### **Rabi G. Mishalani**

Co-Principal Investigator  
Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mishalani@osu.edu

### **Prem Goel**

Co-Principal Investigator  
Statistics  
Ohio State University  
goel.1@osu.edu

### **NEXTRANS Center**

Purdue University - Discovery Park  
2700 Kent B-100  
West Lafayette, IN 47906

[nextrans@purdue.edu](mailto:nextrans@purdue.edu)

(765) 496-9729

(765) 807-3123 Fax

[www.purdue.edu/dp/nextrans](http://www.purdue.edu/dp/nextrans)

NEXTRANS Project No 006OY01

## **Research and Education from a Smart Campus Transit Laboratory**

By

Mark R. McCord, Principal Investigator  
Professor of Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mccord.2@osu.edu

and

Rabi G. Mishalani, Co-Principal Investigator  
Associate Professor of Civil and Environmental Engineering and Geodetic Sciences  
Ohio State University  
mishalani@osu.edu

and

Prem Goel, Co-Principal Investigator  
Professor of Statistics  
Ohio State University  
goel.1@osu.edu

Report Submission Date: October 15, 2009

## **DISCLAIMER**

Partial funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT07-G-005 of the U.S. Department of Transportation, Research and Innovative Technology Administration (RITA), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

## **ACKNOWLEDGMENTS**

The authors acknowledge the invaluable institutional support of Ms. Sarah Blouch, director of OSU's Transportation and Parking Services (T&P), the technical assistance of Mr. Chris Kovitya (T&P) and Mr. Matthew Barber (Department of Civil and Environmental Engineering and Geodetic Science IT), and several helpful discussions with Prof. Nigel Wilson, Mr. John Attanucci, and Dr. Jinhua Zhao of Massachusetts Institute of Technology, Prof. Mark Hickman of University of Arizona, and Prof. Peter Furth of Northeastern University related to the general development and potential of the Campus Transit Lab and to methodological aspects addressed in this report. Prof. Mark Hickman is a member of the project advisory committee. The authors acknowledge the collaboration with and contributions of Prof. Yoram Shiftan of Technion, Israel Institute of Technology (and University of Michigan and The Ohio State University, as Visiting Professor) on the "perceptions and attitudes survey" research activity of this project. They also thank Graduate Research Assistants Mr. Cheng Chen, Mr. Yuxiong Ji, Mr. Jinjian Liang, Mr. Pingbo Lu, Mr. Brandon Strohl, Ms. Fanyu Zhou, and Mr. Honglei Zhu. In addition, the authors acknowledge the financial support offered by the UTC program and The Ohio State University. The views and opinions contained in this report are those of the authors and do not represent the views, opinions, or policies of any other individuals, agency or group.



## TABLE OF CONTENTS

LIST OF FIGURES	5
LIST OF TABLES	6
CHAPTER 1. INTRODUCTION, PROBLEM, AND APPROACH	7
1.1. Introduction	7
1.2. Problem	8
1.3. Approach	8
CHAPTER 2. METHODOLOGY	9
2.1. Smart Bus infrastructure development	9
2.2. Data pre-processing	10
2.3. APC and AVL-based research and outreach activities	10
2.3.1. Matching AVL data to bus schedules	11
2.3.2. Performance assessment of OD estimation from APC data	13
2.3.2.1 Introduction	13
2.3.2.2 The IPF-with-null base procedure	14
2.3.2.3 Design of empirical study	16
2.3.3. Development and application of bus operations simulation	20
2.3.3.1. Simulation structure	21
2.3.3.2. Point-to-point travel time	21
2.3.3.3. Dwell time	22
2.3.3.4. Special point delay	23
2.3.3.5. Validation	23

	4
2.3.3.6. Application of the simulation program	24
2.3.4. Development of outreach products	25
2.4. Educational Use of CTL	27
2.5. Perceptions and attitudes survey	30
CHAPTER 3. FINDINGS	33
3.1. Infrastructure development	33
3.2. Data pre-processing	33
3.3. APC and AVL-based research and outreach activities	34
3.3.1. Matching AVL data to bus schedules	34
3.3.2. Performance assessment of OD estimation from APC data	35
3.3.3. Development and application of bus operations simulation	38
3.3.4. Development of outreach products	39
3.4. Educational use of CTL	45
3.5. Perceptions and attitudes survey	46
3.5.1. Travel mode behavior	46
3.5.2. Perceptions and evaluation analysis	47
CHAPTER 4. CONCLUSIONS	52
REFERENCES	57

## LIST OF FIGURES

Figure 2.4-1: Statement of Campus Transit Lab-based assignment introduced in OSU course <i>CE 570: Introduction to Transportation Engineering and Analysis</i> , Winter Quarter 2009	28
Figure 3.3.3-1: Average absolute dwell time errors, across simulation replications and bus stops, as a function of spatial sampling interval or temporal sampling interval	39
Figure 3.3.4-1a: Expected passenger travel time by CLS boarding stop determined from APC-derived OD matrices and AVL data	44
Figure 3.3.4-1b: Expected passenger travel time by CLS alighting stop determined from APC-derived OD matrices and AVL data	44

## LIST OF TABLES

TABLE 2.3.2-1: Summary of “procedures” used to produce trip level OD volumes	19
Table 2.5-1: CTL transportation first wave survey response rates	31
Table 2.5-2: Response rates comparison with other surveys	32
TABLE 3.3.2-1: Pair-wise performance comparisons between procedures (number of trips in which each outperforms the others)	36
TABLE 3.2.2-2: Relative performance <i>RP</i> summaries across 10 trips	37
Table 3.3.4-1: Normalized OD flow matrix for OSU Campus Loop South route produced from 1003 APC-derived trip level matrices using the IPF-with-null-base procedure	41
Table 3.3.4-2a: Dissimilarity measures for day-of-week pairs of 7-10 AM OD matrices	42
Table 3.3.4-2b: Dissimilarity measures for day-of-week pairs of 2-5 PM OD matrices	42
Table 3.3.4-2c: Dissimilarity measures: 7-10 AM vs. 2-5 PM OD matrices on same day of week	42
Table 3.5-1: Transportation-to-campus mode choices of survey respondents	47
Table 3.5-2: Summary of responses on perception and evaluation questions	49

## CHAPTER 1. INTRODUCTION, PROBLEM, AND APPROACH

### 1.1. Introduction

Approximately a decade ago, a “homemade” GPS-based automatic vehicle location (AVL), communications, and information system – referred to as BLIS (Bus Location and Information System) at that time – was developed for the Ohio State University (OSU) Campus Area Bus Service (CABS), which carries four million passenger riders annually. BLIS was developed and implemented through a collaboration among three groups: planners and operators at CABS, faculty and researchers in Civil and Environmental Engineering, and faculty and researchers in Electrical and Computer Engineering. The third group was primarily interested in the hardware and the integrating software, the second group was mostly interested in effective use of the collected data for service planning, design, and monitoring, and the first group was interested in improved service through the provision of passenger information and the identification and avoidance of service anomalies.

As such, OSU researchers acquired first hand experience with the application of automated data collection systems to public transportation at a time when such applications were in their infancy, and the planners and operators acquired first hand experience with the value of research in extracting meaningful and decision-worthy information from the collected data. This ongoing collaboration demonstrated the value of examining *in-situ* operations to support research and the value of using research to improve operations. Moreover, the potential value of the soft and hard infrastructure resulting from this collaboration in supporting educational activities became apparent. In effect this infrastructure was already serving as a “living lab,” where the CABS operation was large enough to be able to generalize from, yet accessible enough to allow experimentation, whether through the reliance on the automatically collected data or manual field observations. Thus, OSU’s Campus Transit Lab (CTL) was formed.

This historical and mutually beneficial collaboration allowed the case to be made for internal OSU investment from both operating and academic units to replace the increasingly degrading BLIS with a commercial grade system provided by an external contractor. Given the importance of CTL in supporting practice-grounded research and education, part of this investment is now serving as cost-share on this project. CTL is also serving as a critical infrastructure for other externally funded research.

Several desirable characteristics rendered CTL of great interest to Clever Devices, the external contractor. While Clever Devices is an industry leader that has designed and installed the information technologies systems for transit agency in major metropolitan areas, the company had not installed its technologies on a university system. In addition to being able to enter a new market sector, Clever Devices was particularly interested in collaborating with OSU in this major information systems upgrade

because it recognized the value of engaging in the CTL environment, of having proximity to university research, and of gaining access to students with either short or long term employment interests.

The revival of CTL through new technologies and new collaborations – which involve the Department of Civil and Environmental Engineering, the Department of Statistics, the Transportation and Parking Services (which operates CABS), and Clever Devices – is not only critically enhancing the research and education opportunities supported by CTL, but it is also generating unique opportunities for outreach.

## 1.2. Problem

The overarching objective of this project is to assist in the development of the OSU Campus Transit Lab (CTL) as a unique, recognized, and valuable infrastructure for research, education, and outreach both at The Ohio State University and in collaboration with other universities. The specific objectives of this year were to assist in the implementation of the “Smart Bus” system on OSU campus buses, to develop processes to make the data available in useful formats for research, education, and outreach applications, to begin processing the data for multiple uses, to use data for specific applications, and to conduct a survey of campus transit bus users and nonusers that would provide benchmark information for assessing changes in perceptions and awareness of OSU bus transit service that may be attributable to the implementation of the Smart Bus system.

## 1.3. Approach

To assist in developing the CTL into an operational and valuable research, education, and outreach infrastructure, our approach consisted of a mix of activities that can be divided into the following thrusts.

*Thrust 1:* Develop hardware, software, communications protocols, and institutional arrangements that will allow data collected from the Smart Bus system to be processed and used on a routine basis for research, education, and outreach tasks

*Thrust 2:* Pre-process data collected from the Smart Bus system for use in research, education, and outreach tasks

*Thrust 3:* Investigate research questions and develop outreach products involving the use of automatic vehicle location (AVL) and automatic passenger counter (APC) data in bus transit planning and operations

*Thrust 4:* Incorporate the use of CTL data into educational activities

*Thrust 5:* Conduct a survey of the OSU community to provide insights on general and OSU specific attitudes toward transit and real time information systems and that can provide benchmark data to far an eventual investigation of changed attitudes after implementation of the Smart Bus system

## CHAPTER 2. METHODOLOGY

The multiple thrust and sub-thrusts required different methodological approaches. We describe these methodological approaches by each thrust and sub-thrust in this section.

### 2.1. Smart Bus infrastructure development

We needed to address several items to develop the underlying laboratory infrastructure that would allow data to be collected and eventually used for ongoing research, education, and outreach activities. Multiple sensor systems needed to be installed on CABS buses, protocols to transfer data from the data warehouse of the OSU Transportation and Parking (T&P), the operational user of the data, to the project team needed to be arranged, server systems to automatically transfer the raw data files for project use needed to be specified, and software to pre-process the data into easy-to-use formats needed to be developed.

The planning design, and installation of the new set of integrated data collection, communication, traveler information, and enunciation for CABS by Clever Devices took place over a fairly long period that started before proposing and before the approval and commencement of this project. Nevertheless, a substantial part of the effort by this project's research team has been dedicated to continuing with and following through on this process, especially given that the cost-share contribution by OSU to this project takes the form of partially investing in the new system for the purpose of ensuring the availability of the CTL infrastructure to support the research, education, and outreach activities of this project. This effort involved bi-lateral and multi-lateral meetings and discussions with Transportation and Parking Services and Clever Devices to ensure that the CTL capabilities are achieved. In some cases, this led to jointly making system specification decisions.

The OSU Smart Bus system collects and records AVL data at a very high frequency in the daily log-files and APC data at a lower frequency in the bus-state files. Both sets of files are stored on a bus's on-board data units (OBDU). These files are dumped via a wireless channel on to the servers in the Transportation and Parking (T&P) data warehouse after the bus arrives at the depot. These files could contain records for a day or more on all routes the bus served since the time of the last data dump. Other events, such as radio communications and live transmission of bus positions to the control center on a very low space-time resolution, are stored in separate files. The live transmission of bus positions are used in updating the forecasts of next bus arrival time on each stop, which also uses historical data on travel times.

We worked with the T&P IT sector to establish a protocol for automatic transfer of all smart bus data to the CTL servers on an operational basis. This involved several meetings between the project personnel and T&P to understand the data definitions and structures of various files, expected size of these files on

a routine basis, and the formats in which the data could be transferred machine-to-machine. This interaction has also allowed us to have meaningful discussion with IT personnel in Civil Engineering to develop the processor/storage needs of the CTL server and determine how funds from other projects and sources could be leveraged to spec a server that can serve the needs of the CTL and of other data activities and maximize the effectiveness of equipment investments..

## 2.2. Data pre-processing

The data downloaded from the buses to the T&P servers and then transferred to CTL servers must be pre-processed into a form that can be easily accessed for multiple applications by various users, including researchers and undergrad and grad students in different courses. To address this need, we determined the data files that would be used to accomplish the specific applications discussed in Sections 2.3, and 2.4 and the input formats expected by the investigators addressing the applications. These input files would influence the design of the pre-processing software. We have developed, tested, and refined MATLAB codes to transform the data into these formats.

Particularly important was our need to “project” the latitude-longitude vehicle position information contained in the data records onto a description of each transit route, so that a bus record could be associated with a linear distance from some reference location denoting the beginning of the route. We exploited projection logic and codes we developed in previous efforts for a “home-made” automatic bus location system, as well as previously specified way points to define the route structure that receives the projections. However, these projections were not accurate at some locations along the route, since the smart bus AVL data has substantially greater space-time resolution than our old system. We needed to refine the projection logic and codes and specify additional way points to allow more accurate projections of the Smart Bus data. We intend to further improve the projection method for all CABS routes using much more precise shape files of the campus area roads prepared by the Franklin County Engineer’s office and provided by the OSU Facilities Operations and Development Office.

In addition to developing the hardware, software, and communications infrastructure for future use, we wanted to begin processing Smart Bus data this year to initiate its incorporation into research, education, and outreach activities, while we set up automatic data transfer and pre-processing protocols. In this thrust, we determined several activities that would use the Smart Bus AVL and APC data. Activities supported from this project are described in Section 2.3. Other activities, supported from other sources, are described in Section 3.2. We then used the software we developed to process the Smart Bus data for these activities.

## 2.3. APC and AVL-based research and outreach activities

We addressed various research questions and began developing several outreach products that involve the use of APC and AVL.



### 2.3.1. Matching AVL data to bus schedules

The adoption of Automated Vehicle Location (AVL) technology in public transportation provides the capability to amass rich and large data sets that could be used in various planning and operations functions. However, many AVL systems are designed for real-time applications. As a result, the data are not necessarily well archived for use in off-line analyses (Furth *et al* 2006). Even when AVL data are well archived, specific analysis tools must be developed to convert the raw data to meaningful measures. One of the challenges in the development of effective analysis tools is to match the vehicle trajectories derived from AVL data to schedules (Furth *et al* 2006). While the problem applies to both bus and rail applications, this study focuses on AVL data in the case of bus service.

It is important to distinguish between two types of services. For infrequent bus service (i.e., typically headways greater than ten minutes), passengers tend to time their arrivals at the stop according to the schedule. Two common measures of transit reliability in such a case are on-time performance and the variability of schedule deviations, where schedule deviation is defined as the difference between actual bus arrival time and scheduled arrival time. On-time performance can be quantified by the proportion of schedule deviations that fall within an on-time window around the scheduled arrival time. The variability in schedule deviations can be quantified by the standard deviation of the deviations. To produce these two measures, the scheduled trip that a bus follows must be identified.

For frequent transit service, passengers do not generally base their arrivals on a schedule. Rather, they tend to arrive at the stops randomly. One common measure of transit reliability in this case is headway adherence. Headway is the elapsed time between the departure times of two consecutive buses at a specific stop, and headway adherence relates to the regularity of headways. To analyze headways on a route, buses serving the route must have reliably functioning AVL capabilities. Otherwise, it would be difficult to tell whether large headways between two consecutive buses are resulting from bus bunching or an absence of AVL data. Identifying the scheduled trips the buses are following can help distinguish between these two cases and assess reliability more accurately.

If all buses have reliably functioning AVL capabilities, the number of AVL-identified bus trips would equal the number of scheduled trips, assuming that the scheduled number of trips is provided. In this case the actual trips could be matched to the scheduled trips by sorting the AVL-identified trips by time and then matching them to the scheduled trips one-to-one. However, this approach fails in three realistic situations. First the AVL capability could break down on some buses, resulting in missing bus trip trajectories. Second, in some cases bus drivers must initiate the AVL capability on their buses they at the beginning of a run, and the drivers may fail to do so at times Third, congestion or incidents could result in the cancellation of some scheduled bus trips or the introduction of new trips. These situations could result in the loss of a direct correspondence between the AVL-identified trips and the scheduled trips, rendering the matching by this ordering method infeasible or highly prone to errors.

A more sophisticated matching method could be based on calculating the deviation of an AVL-identified bus trip trajectory from all scheduled trips, and then matching the AVL-identified trajectory to the

scheduled trip that produces the lowest deviation. The shortcoming of doing so is that two or more AVL identified trips may be matched to the same scheduled trip. As a result, only one of these trips remains (e.g., the one closer to the scheduled trip) and the other would have to be disregarded.

In light of the above difficulties, prior to beginning this project, the project team developed a method that does not require equality in the number of AVL-identified and scheduled trips and that guarantees that one AVL-identified trip is matched to only one scheduled trip without having to disregard any other AVL-identified trips. The method we originally developed suffered from a shortcoming that arises under commonly applicable conditions. Therefore, under this project we further developed the method to address this shortcoming and arrive at a more robust solution.

Briefly, our method is based on an optimization assignment formulation with the following properties:

- One AVL-identified bus trip can only be matched to one scheduled trip,
- One scheduled trip can only be matched to one AVL-identified trip, and
- A measure of total match error is minimized.

The objective is to minimize the total match error and satisfy the first two properties listed above. The match error could be defined in several ways. In this study, the match error  $e_{ij}$  of assigning the  $i^{\text{th}}$  bus trip in the spot-check table to the  $j^{\text{th}}$  trip in the timetable is considered to be the weighted sum of the absolute value of the difference between the bus crossing times and scheduled departure times, where the sum is taken over all stops as follows:

$$e_{ij} = \sum_{k=1}^N w_{i,j,k} \times |t_{i,k}^{cro} - t_{j,k}^{sch}| \quad (2.3.1-1)$$

and,

$e_{ij}$  = total match error as defined above,

$i$  = index representing the bus trip in the spot-check table,

$j$  = index representing the trip in the timetable,

$k$  = index representing the bus stop,

$N$  = index representing the number of bus stops along the route,

$t_{j,k}^{sch}$  = scheduled departure time at bus stop  $k$  on the  $j^{\text{th}}$  trip in the timetable,

$t_{i,k}^{cro}$  = crossing time at bus stop  $k$  on the  $i^{\text{th}}$  trip in the spot-check table, and

$w_{i,j,k}$  = weight factor at stop  $k$  associated with the sign of  $(t_{i,k}^{cro} - t_{j,k}^{sch})$ .

The term “crossing time” is used to indicate the presence of the bus at the stop, which is estimated using linear interpolation between the last AVL signal before (upstream of) the bus stop and the first signal after (downstream of) the bus stop. The table that includes the crossing times at stops along the route is referred to as the spot-check table. One row in the spot-check table includes the crossing times at all stops for one AVL-identified bus trip. The table that includes the scheduled departure times at all

stops along the route is referred to as the timetable. One row in the timetable includes the scheduled departure times at all stops along the route for one scheduled trip.

The empirical data used thus far is the low resolution AVL CTL data obtained from a “home-made” system previously implemented on several OSU buses. Now that the high resolution AVL CTL data is becoming readily available from the Smart Bus system, the “crossing time” would easily be replaced by either the arrival or departure times. The simulation-based application research activity described below addresses the relationship between AVL resolution and the accuracy of information that could be derived regarding the behavior of buses at stops.

The matching problem is formulated as a network assignment solved using an integer program that minimizes the deviations of the AVL-identified bus trips from the schedule trips under certain constraints that guarantee the properties listed above (Ji et al. 2009). The basic Hungarian algorithm (Hillier et al. 2001) is used to solve this optimization by arriving at a unique solution. In previous validations of this formulation, the solution was found not to be robust. Under this project, the nature of the solution was investigated and source of the resulting lack of robustness was suspected to be related to the specification of the weight factors  $W_{i,j,k}$  in defining the matching error in Equation (2.3.1-1).

The underlying assumption of this formulation is that bus drivers always try to meet the schedule. In the definition of matching error, a negative difference between the crossing time and scheduled departure time represent the case where the bus arrives early to a stop, and a positive difference represents the case where the bus arrives late. In general, if drivers hold at time points when arriving early, early arrivals at stops are less likely than late arrivals. In general, under this operating condition, higher weight factors  $W_{i,j,k}$  should be used for early arrivals than for late ones.

We conducted an empirical study on one of the CTL routes (the Campus Loop South, which is 8.3 km long serving 19 stops). A total of 1,726 AVL-identified bus trips are matched to the schedule using the developed assignment method under two sets of assumptions regarding the weights  $W_{i,j,k}$  of Equation (1). In the first case, equal weighting is specified to early and late arrivals in the total match error of Equation (2.3.1-1). In the second case, the weights associated with early arrivals are specified to be twice those of late arrivals given the holding policy in effect on the route under study.

### *2.3.2. Performance assessment of OD estimation from APC data*

#### *2.3.2.1. Introduction*

Origin-destination (OD) flows constitute one of the most fundamental sets of information used in planning and operating transportation systems. However, OD flows have always been difficult and costly to obtain (Chan, 2007; Furth, et al, 2006). The increasing use of Automatic Passenger Counters (APC) in bus transit systems is yielding comprehensive passenger boarding and alighting data on an on-going basis across the transit network which, although used for other purposes at the transit agencies, offer the potential to determine OD flows on a frequent and comprehensive basis. APC data provide the numbers

of passengers boarding at the various (origin) bus stops and alighting at the various (destination) stops. As such, the APC data provide information related to the OD flows between pairs of stops. This information can only be considered indirect information, however, since the passengers counted as boarding at a stop could have alighted at multiple downstream stops, and the passengers counted as alighting at a stop could have boarded at multiple upstream stops. Still, the indirect information provided by boarding and alighting data can conceivably be helpful in determining OD flows.

Using boarding and alighting data to estimate bus transit OD flows is not a new concept (*e.g.*, Simon and Furth, 1985; Ben-Akiva, et al., 1985), but it is now of greater practical interest because of the availability of APC data. In a Federal Transit Administration project, we are investigating the potential of estimating bus passenger origin-destination flows from APC data. To help guide our FTA-sponsored efforts, we designed and conducted a NEXTRANS study in which we investigated the performance of a simple procedure for determining route-level OD flows (the flows from boarding stops to alighting stops on a bus route where transfers between routes are not considered). Specifically, we investigated the performance of the Iterative Proportion Fitting (IPF) method used with a “null” base matrix. Because it only requires APC data and a specification of the boarding and alighting stops as inputs, the “IPF-with-null-base” procedure can be easily implemented for any route where APC data are collected. However, the “non-informative” nature of the null base matrix may be considered too simplistic to produce good results, and we wished to quantitatively assess the empirical performance of this approach. Our study consisted of collecting true OD flows on OSU bus trips, producing OD flows using the IPF-with-null procedure for the same trips, and comparing the two set of OD flows using an innovative approach that allows a meaningful interpretation of the results. Details on the methodology are provided next.

### 2.3.2.2. The IPF-with-null base procedure

The IPF procedure, which has been referred to by a variety of names, has been widely used in transportation and other fields (Ben-Akiva, et al., 1985). When applied to bus passenger OD estimation, the IPF procedure uses the boarding and alighting volumes to transform an input base OD matrix into an output OD matrix, where the sum of the OD flows from a boarding stop  $r$  to all downstream stops equals the input boarding volume at stop  $r$ , and the sum of the OD flows to an alighting stop  $s$  from all upstream stops equals the input alighting volume at  $s$ .

The IPF-produced OD flows are proportional to base matrix OD flows, with proportionality constants for each row (boarding stop) and for each column (alighting stop). Letting  $q_{rs}^{IPF}$  denote flows between origin (boarding stop)  $r$  and downstream destination (alighting stop)  $s$  that are produced by the IPF procedure using as inputs base OD flows  $q_{rs}^0$ , and a given set of boarding and alighting volumes, the output IPF flows are such that:

$$q_{rs}^{IPF} = k_r^b k_s^a q_{rs}^0 \quad (2.3.2-1)$$

where  $k_r^b$  and  $k_s^a$  are proportionality constants for boarding stop  $r$  and alighting stop  $s$ , respectively, which change from iteration to iteration of the IPF procedure until convergence is achieved.

The IPF procedure was first suggested by Deming and Stephan (1940), in the context of estimation for a contingency table, as a possible solution to the constrained optimization problem consisting of finding estimate  $Q = ((q_{rs}))$  of an OD flow matrix that minimizes the chi-squared distance from a given (or observed from a small survey) base matrix  $Q^0 = ((q_{rs}^0))$ , such that the boarding and alighting volumes determined from the estimated matrix are equal to the observed boarding and lighting volumes. That is:

$$\min_{q_{rs} > 0} \sum_r \sum_s \left[ (q_{rs} - q_{rs}^0)^2 / q_{rs}^0 \right], \quad (2.3.2-2)$$

subject to the constraints

$$\sum_s q_{rs} = b_r, r = 1, \dots, k \quad (2.3.2-3a)$$

$$\sum_r q_{rs} = a_s, s = 1, \dots, k, \quad (2.3.2-3b)$$

where,  $B = \{b_r, r = 1, \dots, k\}$  represents a specified (measured) vector of boarding volumes, and  $A = \{a_s, s = 1, \dots, k\}$  represents a specified (measured) vector of alighting volumes. However, Stephan (1942) showed that the IPF procedure provides an approximate solution to this optimization problem. The convergence of the IPF procedure was first proved by Feinberg (1970). Mosteller (1968) pointed out that the OD flows estimated from the IPF procedure, starting with a base matrix  $Q^0 = ((q_{rs}^0))$  and subject to a given set of boarding and alighting totals, retain the interaction structure of the base matrix, in that the odds ratios of the base matrix and the determined matrix are the same, i.e.,

$$\frac{q_{ij}^0 q_{hk}^0}{q_{ik}^0 q_{hj}^0} = \frac{q_{ij} q_{hk}}{q_{ik} q_{hj}} \quad \forall i \neq h, j \neq k. \quad (2.3.2-4)$$

In the context of determining bus route OD flows, it can be shown (Furth and Navik, 1992) that the estimates produced from the IPF procedure using a null base matrix as input are equivalent to those that are produced from a special case of a method given by Tsygalnitsky (1997), where it is assumed that any passenger on board when the bus arrives at a stop is equally likely to alight at that stop.

As seen from the above summary, in addition to the boarding and alighting volumes, which can be collected from an APC technology, the base matrix is an essential input to the IPF procedure. A base matrix can be considered to be the best OD matrix available to the planner that could be used as the set

of starting OD flow values in the iterative procedure. It could be developed from historical data, a planning model, expert opinion (although eliciting expert opinion for the extremely large number of OD pairs in a transit system would be operationally difficult), or some combination of these sources. If the base matrix  $Q^0$  is consistent with boarding and alighting volumes used in the constraints (2.3.2-3a) and (2.3.2-3b), the optimal solution to the above problem is  $Q^0$ , and  $q_{rs}^{IPF} = q_{rs}^0$  for all OD pairs. Since boarding and alighting volumes are strictly determined from the true OD flows, it follows that if the true OD flow matrix is used as the base matrix, the IPF solution is the base (true) matrix. Of course, if the base matrix reflects different boarding and alighting volumes than the observed inputs, the output matrix will differ from the base matrix.

In the absence of any informative base information, a null OD matrix reflecting equal flows across OD pair can be used as input base matrix. That is:

$$q_{rs}^0 = q^0 \text{ for all } r,s. \quad (2.3.2-5)$$

It follows from constraints (2.3.2-3a) and (2.3.2-3b) and equation (2.3.2-4) that if the base matrix  $Q^0$  is replaced by a scalar multiple, the optimal solution would remain unchanged. Therefore, for operational purposes, the null base can be arbitrarily constructed to consist of unit flows ( $q_{rs}^0=1$ ) for all feasible OD pairs  $rs$ . For computational reasons, the IPF procedure will be more computationally efficient if the average flow is considered in the null base for each OD pair, *i.e.*,  $q_{rs}^0 = \sum_r b_r / N = \sum_s a_s / N$ , for all feasible OD pairs  $rs$ . (To specify feasible OD pairs, it is assumed that travelers do not board and alight at the same stop and only travel downstream along the route.) Alternatively, in the context of a “normalized” OD matrix, where the matrix provides the proportion of total flow traveling from a specified origin to a specified destination, the base matrix entries can be set to  $q_{rs}^0 = 1/N$ , where  $N$  is the number of feasible OD pairs. The normalized OD matrix can be interpreted as the probability that a random passenger travels from the specified origin to the specified destination. In this sense, the null (normalized) base matrix implies that any feasible OD pair is equally likely to be the one traveled by a random passenger. The null base matrix can, therefore, be considered a “non-informative” prior distribution in Bayesian terminology (Berger, 1985).

### 2.3.2.3. Design of empirical study

In our empirical study, we collected true OD passenger flows and the corresponding boarding and alighting volumes for each of a set of bus trips. We then used the IPF procedure to calculate the OD matrix for each bus trip, using the trip-level boarding and alighting volumes and a null base matrix as inputs. The quality of the flows produced was assessed by comparing each determined bus trip OD matrix to the corresponding observed true OD matrix. To put the performance in perspective, the performance of other approaches used to produce OD matrices was also assessed. Some of the approaches would be expected to perform worse, and others would be expected to perform better than the IPF procedure using the null base matrix.

It can be shown that the solution to the optimization problem defined by objective function (2.3.2-2) and constraints (2.3.2-3a) and (2.3.2-3b) does not change if the base OD flows, the boarding volumes, and the alighting volumes are all divided by the total volume, except the problem is converted to the determination of the normalized OD matrix. (As mentioned above, the “normalized” OD matrix is the matrix that provides the proportion of passenger trips, rather than the number of passenger trips, using the OD pair. It is formed by dividing the OD matrix indicating the numbers of OD trips by the total number of trips in the matrix.) Considering normalized matrices focuses the comparison on determining OD patterns in the form of proportions or probabilities and controls for any effect of volume on the analysis. Therefore, we based our comparisons on normalized OD matrices for each bus trip.

We adapted a procedure described by Simon and Furth (1985) to collect data on ten trips of OSU’s Campus Loop South (CLS) bus route between 8 and 10 a.m. on weekdays during the winter quarter (January through mid March) of 2009. Two data collectors rode CLS buses, with one person stationed near the front door and one stationed near the rear door. The data collectors distributed cards indicating the boarding stop to passengers as they boarded the bus and collected the cards as the passengers alighted. By filing the cards collected according to the alighting stop and bus trip, the cards could be used to determine both the empirical OD flows and the corresponding empirical boarding and alighting volumes for the various bus trips. This approach allowed us to collect OD flows on all origin-destination pairs on a bus trip with only two data collectors.

CLS travels in a loop pattern, serving twenty stops, four of which are located in a “West Campus” parking lot. (At the time that data were collected for the study described in Section 2.3.1, CLS served nineteen stops. A twentieth stop was recently added to the route.) For the purposes of this study, the four West Campus stops were aggregated into a single pseudo-stop, which we considered as the first boarding stop for the ensuing bus trip and the last alighting stop for the just-completed bus trip. (Because of service provided by other bus routes and the trip patterns derived from campus activities, it is rare that a passenger would board upstream of the West Campus parking lot for a destination downstream of the lot. Only 6 of the 702 passenger trips were observed with such an OD pattern, and they were omitted from the empirical data used in the study reported here.) In this way, for each of the 10 trips, the data consisted of volumes for 18 boarding stops and 18 alighting stops and OD flows for each of the 153 feasible OD pairs.

We use  $Q_j^{true}$  to denote the matrix of true (normalized) OD flows for trip  $j$  and  $q_{rs,j}^{true}$  to indicate the true (normalized) flow between boarding stop  $r$  and alighting stop  $s$  on trip  $j$ . Using the empirical boarding and alighting volumes for trip  $j$  with the IPF procedure and a null base matrix  $Q^{null}$  as input, we determined a trip-level OD matrix  $Q_j^{IPF}(Q^{null})$  with elements  $q_{rs,j}^{IPF}(Q^{null})$  for each trip.

For comparison purposes, we considered trip-level OD flow matrices produced by other “procedures.” We summarize these procedures in Table 2.3.2-1 and describe them briefly here. We motivate and explain the matrices further in McCord et al. (2009).

- As discussed above, the null matrix  $Q^{null}$  represents a “non-informative” estimate of the normalized OD flows, where it is assumed that a random passenger on the trip was equally likely to travel on any of the feasible OD pairs.
- The refined null matrix  $Q^{ref-null}_j$  refines  $Q^{null}$  by using the boarding and alighting data on trip  $j$ . Specifically, if no passengers boarded at a stop on trip  $j$ , there could be no OD flow to any of the downstream destinations on the trip and, similarly, there would be no flow on trip  $j$  to stops that had no recorded alighting volume on the trip. In  $Q^{ref-null}_j$ , zero probability (proportion) is assigned to all such OD pairs, and the equal probabilities are recalculated based on the reduced number of feasible OD pairs.
- The matrix  $Q^{IPF}_j(Q^{null})$  produced from the IPF-with-null base procedure has been discussed above.
- As also explained above, we obtained true OD flows for ten empirical trips. To represent the results produced from an on-board survey (with 100% sample), we produced the normalized flows from this set of OD flows. Results from on-board surveys would be used to predict flows on future trips. As such, when considering estimating the OD flows on trip  $j$ , we held out the true trip  $j$  flows when forming  $Q^{on-board}_j$ .
- The onboard survey should provide a better estimate of the OD flows than would the null matrix. Therefore, we would expect that the matrix  $Q^{IPF}_j(Q^{on-board}_j)$  produced from the IPF procedure using the on-board survey matrix as input would perform better than the matrix produced when using the null matrix as input.
- As presented above,  $Q^{true}_j$  represents the matrix of true normalized OD flows on trip  $j$ , where the true flows were obtained by the data obtained in our data collection effort.

We produced the OD matrices determined by the different “procedures” summarized in Table 2.3.2-1 for each of the ten trips for which we collected empirical data. We then compared these matrices to the true trip-level normalized OD matrices. To assess the performance of procedure  $m$  in determining the true normalized OD matrix  $Q^{true}_j$  on trip  $j$ , we computed two different scalar measures of performance  $P_j^{i[m]}$ ,  $i = 1, 2$ .

Performance measure  $P^1$  consists of the sum of the squared differences between the normalized OD flows produced by procedure  $m$  and the true normalized OD flows:

$$P_j^{1[m]} = \sum_r \sum_s [q_{rs,j}^{true} - q_{rs}^{[m]}]^2 \quad (2.3.2-6)$$



where,  $q_{rs,j}^{true}$  is the true (observed) normalized OD flow on trip  $j$  from boarding stop  $r$  to alighting stop  $s$  and  $q_{rs,j}^{[m]}$  is the normalized OD flow from  $r$  to  $s$  on trip  $j$  determined by procedure  $m$ . Larger values of  $P^1$  represent poorer performance.

TABLE 2.3.2-1: Summary of “procedures” used to produce trip level OD volumes

Procedure	Notation	Description
Null	$Q^{null}$	Equal probabilities across all theoretically feasible OD pairs, zero otherwise; constant across all trips.
Refined null (R-null)	$Q_j^{ref-null}$	Equal probabilities across all APC-determined feasible OD pairs, zero otherwise; determined from boarding and alighting volumes on trip $j$ .
IPF-with-null base (IPF-null)	$Q_j^{IPF}(Q^{null})$	IPF procedure using APC data on trip $j$ as inputs and the null matrix for a base.
On-board survey (OBS)	$Q_j^{on-board}$	Normalized flows based on all observed trip OD flows excluding those of trip $j$ .
IPF with OBS (IPF-OBS)	$Q_j^{IPF}(Q^{on-board})$	IPF procedure using APC data on trip $j$ as inputs and the on-board survey matrix for a base.
True	$Q_j^{true}$	Observed (true) normalized OD flows for trip $j$ .

The sum of squared differences  $P^1$  is a commonly used measure of performance in comparing vectors in general applications, but it does not incorporate the spatial nature of OD flows. For example, assigning flows from an origin to an erroneous destination close to the correct destination may be considered less onerous than assigning the erroneous flows to a destination farther away.

To incorporate a spatial dimension in measuring performance, we developed a second measure  $P^2$  based on passenger distances traveled (PDT) derived from the OD matrices. Specifically, we used the route distance between origin-destination pairs and the (normalized) OD flows  $Q_j^{[m]}$  for trip  $j$  to produce the distribution of PDT values for trip  $j$  and formed the cumulative distribution  $F_j^{PDT,[m]}(x)$  of these PDT values. The matrix  $Q_j^{true}$  of true OD volumes for trip  $j$  provides  $F_j^{PDT,true}(x)$ , the true cumulative distribution of PDT for trip  $j$ . The second measure of performance  $P_j^{2[m]}$  is then the absolute value of the areas between the true cumulative distribution function and the distribution function derived from the trip level OD matrix obtained from procedure  $m$ :

$$P_j^{2[m]} = \int \left| F_j^{PDT,true}(x) - F_j^{PDT,[m]}(x) \right| dx. \quad (2.3.2-7)$$

Like  $P^1$ , smaller values of  $P^2$  indicate better performance. However,  $P^2$  incorporates spatial considerations through its use of distance. On the other hand,  $P^2$  would not reflect a large error in estimated flow for an OD pair with a given distance that is compensated by an error of similar magnitude in the opposite direction (*e.g.*, an underestimate compensated by an overestimate) for an OD pair with similar distance. Therefore, we considered both measures  $P^1$  and  $P^2$  in assessing empirical performance.

Both  $P^1$  and  $P^2$  can be used to represent ordinal performance of the procedures. That is, a procedure with smaller value of  $P_j^1(\cdot)$  performs better than a procedure with larger value in determining the OD flows on trip  $j$  according to the sum of squared differences measure  $P^1$ , and similarly for the passenger distance traveled measure  $P^2$ . To provide a more meaningful quantification of the performance of the procedures in determining OD flows on trip  $j$ , we defined a measure of relative performance  $RP_j^{[m]}$ . For trip  $j$ ,  $RP$  quantifies the improvement in the OD matrix, according to the reduction in performance measure  $P^1$  or  $P^2$ , produced by procedure  $m$  from that of the null matrix, as a proportion of the corresponding performance measure for the null matrix on that trip. Comparisons are made with respect to the null matrix, since the null matrix is the most basic estimate of the OD flows, where only the structure of the route (yielding the feasible OD pairs) is used. In this way, we defined  $RP_j^{i[m]}$  as:

$$RP_j^{i[m]} = [P_j^{i[Null]} - P_j^{i[m]}] / P_j^{i[Null]} \quad (3.2.2-8)$$

According to the  $RP$  measure, the relative performance of the null matrix  $Q^{null}$  would be zero for any trip, and the relative performance of the true matrix  $Q^{true}$  would be one for any trip. Thus,  $RP$  values close to zero would indicate little improved performance, relative to what could be produced from the null matrix.  $RP$  measures close to one would indicate a relatively large improvement in performance.

### 2.3.3. Development and application of bus operations simulation

Given the complexity of actual bus operations, certain problems are not possible to characterize and solve analytically. An effective alternative in such situations is the use of simulation. Our interest in developing a campus lab motivates us to be able to investigate transit operations under complex, realistic conditions. Therefore, we are developing a bus operations simulation tool. The simulation is bus specific and reflects the stochastic nature of operations. This year, we derived the actual parameters of a first version of the simulation tool from previously collected CTL data and partially validated results with additional data collected this year. We then used this version of the bus simulation tool to assess the effect of AVL sampling on the accuracy of bus dwell times, an important variable for transit planning and operations that can be estimated from AVL generated data. This section describes the developed simulation and its application to the AVL sampling analysis.

### 2.3.3.1. Simulation structure

The bus simulation model is based on the OSU Campus Loop South route and is partitioned into three components: point-to-point travel times over space, dwell times at bus stops, and delays at special points. Point-to-point travel time is the time a bus needs to transverse a certain space that does not include bus stops or some special points. Dwell time is the time interval between the arrival of a bus to a stop and its departure from that stop. Delay at special point is the delay caused by certain locations along the bus route such as signalized and un-signalized vehicular and pedestrian intersections that form bottlenecks in the roadway network. At this point, only delays caused by major signalized intersections have been incorporated in the simulation tool.

The three components collectively determine the bus trajectory. Space along the route is represented by contiguous sections, where the boundaries between sections signify either a bus stop or a special point. Furthermore, sections are subdivided into small contiguous segments (a segment length of 5 meters is used as a default). For each section, a point-to-point time is simulated. A simulated bus moves along a section accordingly until it encounters a stop or a special point at the end of a section. A dwell time or a special point delay is simulated at that point. A simulated bus comes to a stop at bus stops (to reflect the operating policy in place on the OSU bus system) but not necessarily at a special point. Both dwell time and special point delays are added to the travel time when the bus crosses the boundary and just before it starts traversing the next section. In what follows the simulation of point-to-point travel times, dwell times, and delays are described.

### 2.3.3.2. Point-to-point travel time

Point-to-point travel time is the time a bus needs to transverse a certain section along the route that does not include bus stops or special points. A point-to-point travel time is simulated for each section based on empirical data. A time-based simulation is adopted whereby a bus speed is generated following a certain time-step (a time-step equal to one second is used as a default) and then advanced along the route based on the generated speed. A generated speed depends on the previously generated speed and empirical historical data observed on the segment where the bus is located at the instant of the simulated event.

The dependence between two consecutive simulated speeds is captured through constraints set on the acceleration of the bus. The default maximum acceleration is set at  $+2 \text{ m/s}^2$  and the default minimum acceleration is set at  $-4 \text{ m/s}^2$ . That is, if the most recently simulated speed is  $v$ , then the speed simulated one time-step later must fall within the range of speeds defined by the maximum and minimum accelerations and the previous speed  $v$ . Given this permissible speed range, to simulate the current speed, a speed is randomly drawn from the conditional speed distribution derived from empirical AVL-based data for the segment where the simulated bus is located. The conditioning is performed on the determined speed range.

The simulated bus is then advanced along the route in accordance with the time-step and the newly generated speed. At that point the process repeats itself. That is, for the new location, a speed range is determined from the previous speed and the acceleration constraints, and the new speed is simulated by drawing from the empirically derived conditional speed distribution for the segment on which the newly simulated bus position is located.

#### 2.3.3.3. Dwell time

Dwell time is the time interval between the arrival and the departure of a bus to and from a bus stop, respectively. The simulation generates dwell times from sets of stop-specific dwell time values estimated from empirical data. The empirical dwell times are estimated from the CTL AVL data using one of several possible methods. The inputs to these methods include the latest AVL signal upstream of the stop and the earliest signal downstream of the stop. The signal data include location, time and speed.

Given the operating policy in place, buses must come to a complete stop at the bus stop even if there are no passengers wishing to alight or board. This stopping behavior could be captured in one of two ways: (i) The bus runs at a constant speed from the location of the last upstream AVL signal before the stop until it arrives at the bus stop at which point the speed drops to zero instantaneously; it then instantaneously changes its speed from zero to some constant speed until it arrives at the location of the first downstream AVL signal. Or, (ii) the bus decelerates from a certain speed from the location of the last upstream AVL signal to a speed of zero when arriving at the bus stop; and then accelerates from a speed of zero when departing the stop to a certain speed at the location of the first downstream AVL signal. These two scenarios are referred to as the “without-acceleration” and “with-acceleration” models, respectively. In both cases, the trajectory of the bus is projected from the location and time point of either AVL signal to the stop. The dwell time is then calculated as simply the difference between the projected departure and arrival times at the stop.

In the with-acceleration case, the speed out of the first signal is assumed constant followed by a default acceleration of  $-2\text{m/s}^2$  such that the bus comes to a full stop at the location of the bus stop. If this constraint cannot be met under this projection assumption, the bus is assumed to follow an acceleration from the given speed at the location of the upstream signal that ensures a full stop at the location of that stop. Similarly, when projecting from the downstream signal a constant speed is assumed going into the signal preceded by a default acceleration of  $+2\text{m/s}^2$  if the constraint of zero speed at the stop can be met. Otherwise, the necessary higher acceleration is calculated such that the bus achieves the given speed at the location of the downstream signal.

In both acceleration models, speed information at either AVL signal is required. The instantaneous speeds reported with each AVL signal (along with location and time) are prone to high measurement errors. Therefore, a form of smoothing is adopted. More specifically, historical bus speed values along the route are averaged within the contiguous segments constituting each section (as defined above). As such, each route has a speed look-up table with some speed value associated with each segment. (In a more general implementation, the look-up table could be time period specific.)

Three alternative average speeds are adopted in generating three look-up tables for a bus route: average instantaneous AVL speed, harmonic average of instantaneous AVL speed, and average arc speed. In the first case, the average instantaneous AVL speed is the simple average of the historical instantaneous AVL speeds falling within each segment. In the second case, the harmonic average, instead of the simple average, is adopted. Since speeds of zero cannot be used in calculating the harmonic average, the reported zero speeds could either be ignored or replaced with a very low speed. In this study, zero speeds are replaced with 0.2 m/s. In the third case, first historical arc speeds are calculated from the location and time of consecutive pairs of AVL signals not separated by bus stops or special points. The location of the calculated arc speed is set to be the midpoint of the two respective AVL signal locations. The look-up speed for a segment is then determined as the simple average of the historical arc speeds whose locations fall within that segment.

The estimated dwell times based on each of the six combinations of speed look-up tables and acceleration models constitute the historical dwell times based on which the dwell time distribution is determined. It is one of these six empirically derived distributions that is drawn from when simulating the dwell times in the simulator.

When the bus is simulate to be at a stop where a holding policy is in effect, once a dwell time is simulated, the resulting departure time is compared to the scheduled departure time. If the calculated departure time is less than the scheduled departure time by more than a short threshold, then the simulated bus is held until the schedule departure time. Therefore, in deriving the dwell time distributions for each segment from which dwell times are drawn, only empirical dwell times that do not reflect any bus holding are considered. (That is, only empirical dwell times that correspond to bus departure times larger than scheduled departure times are considered.)

#### *2.3.3.4. Special point delay*

Similar to the simulation of dwell times, special point delays are also simulated by generated delays from empirically derived distributions for each special point. In this model, major intersections along the route are treated as special points. As in the case of dwell times, special point delays need to be determined to produce the respective distributions to sample from in the simulation. These delays are calculated from two AVL signals, one upstream and one downstream of a special point, along with the speed look-up tables in a manner similar to that used in the calculation of the dwell times. However, while in the case of dwell times the bus is constrained to come to a full stop at the location of a bus stop, this constraint is not applied in the case of the special points, given that buses do not always stop at these special points.

#### *2.3.3.5. Validation*

The three speed look-up tables, coupled with the two acceleration models (used to calculate the dwell times and special point delays that are, in turn, used to derive the respective distributions that generate the dwell time and special point delay components of the simulator) result in a total of six possible

simulation scenarios. We conducted a validation exercise where we compared the simulated bus travel times between two departures from consecutive bus stops under each of the scenarios to observed travel times. More specifically, we compared the mean and variance of simulated and actual travel times and identified the scenario that most closely matched the actual travel times.

#### 2.3.3.6. Application of the simulation program

We envision using the simulation program to address multiple questions of interest that will arise relating to bus operations. This past year, we wanted to apply the system to assess the ability of a simple method to estimate bus times at a stop as a function of data frequency. The time the bus is stopped at a bus stop is an important measure of performance for off-line planning and real time operations of bus systems. Understanding stopped time patterns allows planners to understand where drivers rush to make up time or where there is excess time on the route so that the bus can wait to get back on schedule. We will call these stopped times “dwell times” for this study, although dwell times are often measured as the duration of time from when the bus doors open after arriving at a stop until they close before departing the stop. The dwell times we will consider will also include any time the bus may wait at the stop with doors closed after all passengers have boarded and alighted. This extra stopped time (referred to as “holding time”) would generally occur at predetermined time-points (which are stops designated for possible holding) when a driver is ahead of schedule at that time-point. As discussed above, note that due to the structure of simulating dwell times where holding is simulated separately, only dwell times where no holding is taking place are used in empirically deriving the dwell time distributions from which the time a bus spends stopped at a stop before holding is simulated. Our method of estimating dwell times was motivated by our experiences with the previous, “home-made” AVL system. Like many systems, location data in our system were transmitted and recorded on a limited basis to reduce communication costs. Time-stamped locations were to be communicated every 100 meters of travel, or every 3 minutes of elapsed time, whichever occurred first. Such data would not allow a definitive determination of the bus dwell times, and we wished to estimate the times to understand behavior of dwell times – including the spatial (across stops) and temporal (across time at the same stop) variability in the dwell times – for the bus system so as to understand transit operations better and to calibrate our simulation program.

As a simple means of estimating dwell times, we assumed that we could develop a “look up” table that provided the average speed for a bus to traverse 80-m long spatial intervals, where intervals were non-overlapping and, taken together, spatially covered the entire route. Given these “look up” speeds, the location of any AVL signal, and the location of the considered bus stop, the time that would elapse between the time the bus sent an AVL signal upstream of the stop and the time that the bus would arrive at the stop could be determined. Adding this time to the time of the AVL signal would yield an “arrival time” estimate at the stop. Similarly, the time that would have elapsed between the time that the bus departed from the stop and the time that it sent a downstream AVL signal could be determined. Subtracting this time from the time of the signal would yield a “departure time” estimate from the stop. Subtracting the arrival time from the departure time would produce the estimated dwell time.

We obtained the “look up table” speeds by running the simulation model many times, assuming that time-stamped AVL locations were generated at some spatial or temporal sampling frequency (see below), determining the average speed between a pair of signals by dividing the difference between the linear distances by the difference between the corresponding time stamps for the pair of signals, and associating this speed with the midpoint of the signal distances. Then, we used the harmonic mean of all such generated speeds in 80-m spatial cells as the speeds corresponding to the cells that were used to estimate arrival and departure times.

In our simulation, we assumed that signals were generated either with a spatial sampling interval of  $y_1$  meters or a temporal sampling interval of  $y_2$  seconds. We would generate the true location and time of the vehicle at a very fine resolution by using the simulation program. We then sampled the generated locations and times of the vehicle at the specified sampling interval and determined the last sampled upstream signal before a stop and the first sampled downstream signal after a stop. Based on these sampled signals, we used the “look up table” speeds described above to estimate the arrival and departure times at the stop and, consequently, the estimated dwell time at the stop. The simulation generated true dwell times at the stops. We used the absolute value of the difference between the simulated and true dwell times to produce a measure of performance for the dwell time accuracy.

We ran 1000 replications of the simulation, where the replications generated consecutive bus trip covering the entire route. We sampled the simulated true locations and times in 10-m spatial increments and 3-sec temporal increments. The results are presented in Section 3.3.3.2.

#### *2.3.4. Development of outreach products*

In addition to conducting research activities related to the use of APC and AVL data, we wish to exploit the data and the results of our research investigations to produce quantitative information that we will use, in collaboration with OSU Traffic and Parking Services (T&P), to monitor performance of the OSU bus service. In Section 2.3.2 we described the use of the Iterative Proportional Fitting (IPF) procedure to produce origin-destination (OD) flows from the boarding and alighting data recorded by the APC system. (We are presently investigating and refining other methods for future use.) We intend to produce OD matrices for OSU T&P on an ongoing basis. In this first year effort, we wanted to apply the codes we developed in our FTA-supported project in a batch mode to produce multiple OD matrices using the Smart Bus APC data collected on the Campus Loop South (CLS) route and synthesize the multiple matrices.

In addition we identified various applications that rely on the OD estimates to develop and monitor travel patterns. We intend to establish benchmark patterns developed from these applications and monitor the patterns over time in collaboration with OSU T&P. In this first year, we developed several concepts and produced preliminary results using the first wave of Smart Bus data on the CLS route.

One concept we developed is derived from an aspect we are addressing in our FTA project. In that project we are developing methods to automatically indicate periods of homogeneous OD flows from the APC data. In the NEXTRANS project reported on here, we borrowed concepts from these methods,

which are still under development, to test exogenously specified periods for similarity of OD patterns on the CLS route. We exploited the codes we are developing in the FTA project to develop the ability to:

- automatically segment the route-level OD flow matrices produced from the APC data into specified periods
- aggregate the route-level OD matrices into normalized period-level OD matrices (matrices indicating the proportion of passengers using the specified OD pairs)
- calculate “dissimilarity” measures between pairs of aggregated normalized OD matrices to indicate matrices that are similar to each other and matrices that are very different. The measure used here is based on the chi-squared statistic for a pair of OD matrices, divided by the degrees of freedom, which is equivalent to Cramer’s measure of association between two probability distributions, described in 3.5. The dissimilarity measures are defined so that greater values indicate less association.

The dissimilarity measures can be recalculated and monitored over time to determine changes in travel patterns. Our goal in this first year was to use the first wave of APC data to demonstrate the concept on the CLS route.

In addition to determining periods of similar OD patterns, we were interested in developing the ability to monitor the distribution of bus passenger trip distances, where the trip distances are derived from the APC-derived OD matrices and the distances between stop pairs. We collected true OD data during Winter and Spring quarters using the technique described in Section 2.3.2. From these data, we noticed a higher percentage of short (less than one mile) bus trips in Winter quarter than in Spring quarter. The hypothesis is that the better spring weather entices more people to walk these shorter distances, rather than ride the bus for such trips. We wished to validate that our APC-derived OD estimates reflected this change in travel pattern. If it did, we would have more faith in our ability to use the APC-derived OD matrices to monitor the passenger distance distributions over time. We used the IPF-with-null base procedure (see Section 2.3.2) to produce aggregate Winter quarter and Spring quarter OD matrices from the observed boarding and alighting volumes and investigated whether these matrices exhibited the reduction in short trips observed in the true OD flow data.

The final concept exploiting the OD estimates that we developed this year was inspired by the course assignment developed for *CE 570* (see Section 2.4 below). Specifically, we developed the means to combine the estimated OD flows derived from the APC data with the bus travel time and dwell time information contained in the AVL data to determine the expected time that a passenger travels on the bus, conditional on the passenger’s boarding stop, and the expected travel time on the bus, conditional on the passenger’s alighting stop. The goal of the first year effort was to demonstrate the concept with the first wave of APC and AVL data collected and to prepare for benchmarking and ongoing monitoring in the future. We therefore produced these times from the Smart Bus APC and AVL data on the CLS route.



## 2.4. Educational use of CTL

The inclusion of the Smart Bus system in the CTL has begun providing a unique infrastructure for research and outreach projects and has, therefore, provided an important educational experience for several graduate and undergraduate research assistants. Several of these students are writing theses or MS reports related to the CTL. The CTL can also be used to enhance coursework. Although the Smart Bus system has only recently been installed, we sought ways to incorporate the data and information we have been collecting and will continue to collect in existing courses. We identified two OSU courses, *CE 570: Introduction to Transportation Engineering and Analysis* and *CE 873: Urban Transportation Demand Analysis*, in which we could use data and results produced from our first year efforts described above.

*CE 570* is a course required of all undergraduate Civil Engineering students. Some of these students choose “transportation” as their major area of specialization, but the vast majority of students choose other areas of Civil Engineering as their “major area.” For most of these students, *CE 570* is the only transportation course taken in their undergraduate program. The course covers multiple topics in transportation engineering. However, the various topics are covered at a level deep enough that, in addition to learning basic concepts and terminology, students are expected to conduct mathematical and logical analysis so as to gain insights for design, planning, or operations. The course is offered once each year and has had recent enrollments of approximately 100 students per offering.

A previously existing module of *CE 570* covered the estimation of expected travel times for a public transportation system. Calculations had been conducted analytically for a system with dedicated right-of-way. To supplement this module, we developed an assignment in which students used the OD data described in Section 2.3.2 and specially collected bus travel and dwell time data to determine empirical expected passenger “line haul” times (times aboard the bus). The students were given stop-to-stop expected travel times for the OSU Campus Loop South (CLS) route, expected dwell times at the stops on the route, and the average numbers of passengers per trip who boarded in the West Campus area and alighted at each downstream stop. (These average numbers of passengers were derived from the passenger OD matrices as described in Section 2.3.2.). The students were then requested to calculate the expected times for a passenger boarding at the OSU West Campus area to arrive at each main campus bus stop, the probability that a passenger boarding at West Campus would alight at each of the main campus stops, and the expected time on the bus (line haul time) for a random passenger boarding at West Campus. To motivate the assignment, the campus transit lab, as it existed at the time and as it is envisioned, was presented in a lecture format to the students, and the advantages of AVL and APC technologies for data collection were emphasized. The statement of the assignment is presented in Figure 2.4-1.

*CE 873* is a graduate level course devoted to the theory of discrete choice modeling as applied to transportation choices. This course requires statistical and econometric analysis, mathematical, logical, and domain specific analysis, and computer work with model estimation software. The course emphasizes the use of the binary, multinomial, and nested logit models for modeling discrete choice. Graduate students majoring in Civil Engineering, City and Regional Planning, and, occasionally,

Geography take this course. *CE 873* is a required course for students enrolled in the Dual Masters Degree in Urban Transportation Planning program. The “dual degree program” is a specially designed program in which accepted students can receive M.S. degrees in Civil Engineering and in City and Regional Planning in less time than it would take to pursue these degrees separately. *CE 873* is an elective course for the students not enrolled in the “dual degree program,” who make up the majority of the class. The course is offered every other year and has had recent enrollments of between 10 and 15 students per offering.

**The Ohio State University** **Winter 2009**  
**CE 570: Transportation Engineering and Analysis** **Dr. Mark R. Mc Cord**

Take-home problem: This problem will count toward 2 points of your 50-point final exam.

Date Handed Out: 10 March 2009

Due Date: 17 March 2009 (slip under my office door if you cannot turn it in during my office hours)

You may and are encouraged to discuss your approach with others, but you are to write your own response. *Make sure your write up is clear and neatly organized.* I do not need to see details, but I must be able to *easily* follow what you are doing.

Make sure to make and keep a copy before submitting.

Use the data below, which we collected on OSU’s Campus Area Bus Service (CABS) Campus Loop South route, to perform the following tasks. (For the simplicity of this assignment, all four stops on west campus have been grouped to form one stop, called the West Campus stop.)

1. For each stop in the system, determine the expected time *in minutes* that it takes a passenger to arrive at his or her stop once the bus has left West Campus. (That is, find the expected line haul time from West Campus to each stop.)
2. For each stop in the system, determine the probability that a passenger who boarded at the West Campus stop will alight at that stop.
3. Using your answers from part 1 and part 2, find the expected line haul time for a random passenger boarding at the West Campus stop. (The expected line haul time will represent the expectation considered across all possible alighting stops.)

Figure 2.4-1: Statement of Campus Transit Lab-based assignment introduced in OSU Course *CE 570: Introduction to Transportation Engineering and Analysis*, Winter Quarter 2009

**Expected Stop-to-Stop Travel Times**

From Stop	To Stop	Time (sec)
West Campus	Blankenship Hall	105.8
Blankenship Hall	Ag Campus Eastbound	73.7
Ag Campus Eastbound	St John Arena Eastbound	58.0
St John Arena Eastbound	Drake Union	106.4
Drake Union	Cannon Dr. & 12th Southbound	104.8
Cannon Dr. & 12th Southbound	Med Center Dr. & Cannon	128.8
Med Center Dr. & Cannon	Med Center Dr. & 9th	40.6
Med Center Dr. & 9 <sup>th</sup>	10th Ave. & Neil Ave.	76.3
10th Ave. & Neil Ave.	Mack Hall	68.3
Mack Hall	Hale Hall	108.3
Hale Hall	Ohio Union Northbound	29.2
Ohio Union Northbound	Arps Hall	92.8
Arps Hall	North Dorms	84.9
North Dorms	Woodruff & Tuttle	80.8
Woodruff & Tuttle	St. John Arena Westbound	77.7
St. John Arena Westbound	Ag Campus Westbound	57.0

**Expected Dwell Time & Average Alighting Passengers who Boarded at the West Campus Stop**

Stop	Dwell Time (sec)	Average Number of Passengers Alighting
Blankenship Hall	5.8	0.2931
Ag Campus Eastbound	10.2	1.2446
St John Arena Eastbound	12.3	0.2733
Drake Union	20.0	1.0559
Cannon Dr. & 12th Southbound	12.5	3.3760
Med Center Dr. & Cannon	5.9	0.5258
Med Center Dr. & 9th	9.1	1.1770
10th Ave. & Neil Ave.	16.5	1.3230
Mack Hall	17.7	1.1740
Hale Hall	13.3	0.1837
Ohio Union Northbound	23.0	0.1575
Arps Hall	36.3	0.7382
North Dorms	20.2	0.5732
Woodruff & Tuttle	5.9	0.0262
St. John Arena Westbound	10.0	0.0682
Ag Campus Westbound	30.6	0.2883

Figure 2.4-1 (continued)

Although most of the examples and assignments discussed in the course pertain to transportation mode and destination choice, other uses of discrete choice models are presented. In addition to a series of model estimation assignments, students develop, with the help of the instructor, a 3-4 week project in which they estimate logit models to gain insights on some problem of their choosing. The projects are generally performed in groups. During the offering in this past year, a group of two students estimated binary logit specifications to investigate the effect of different factors on the performance of the bus passenger OD estimation described in Section 2.3.2. Specifically, the students determined the

percentiles of the  $P^2$  measure presented in Section 2.3.2 from the multiple route-level OD matrices produced. They then estimated binary logit models, where the dependent variable was a binary indication of whether the  $P^2$  measure for the estimated OD matrix was greater or less than the specified  $P^2$  percentile value. The independent variables investigated consisted of the different bases (null or on-board survey) used in the IPF procedure, the trip volume, the day of the week, and the time of day.

## 2.5. Perceptions and attitudes survey

A two-wave survey of the OSU community was designed, and the first wave was undertaken to study factors that influence transportation choices and traveler satisfaction, in general, and to develop insights on individual *preferences and perceptions of transportation options* as impacted by the provision of passenger information. OSU's CABS and the Smart Bus system are used *as a case study*. The first wave of the survey took place before the provision of real-time passenger information, and the results provide benchmark data for investigating possible changes in perceptions and attitudes resulting from the implementation of the Smart Bus system, which are to be captured by data collected in the second wave of the survey.

During the planning phase, we made several iterations on possible study designs, the issues to be addressed in the study, and modes of administering the survey questionnaire. Given the relative high costs for conducting the survey via intercept modes, it was decided that a web based survey would be the most cost-effective way to obtain the data of interest. A contract was signed with the Ohio State University Statistical Consulting Service (SCS) to implement our survey design on-line. First, the pilot version of the questionnaire was implemented in order to test the format and the wording of the questions. Based on the feedback from potential subjects, we finalized the survey questionnaire. SCS then coded the final version for online implementation. SCS obtained a random sample of e-mail addresses of undergraduate and graduate students from the OSU Office of the Registrar and of the faculty and staff from the Office of Human Resources for inviting the sample of subjects to participate in the survey.

The questionnaire consisted of 9 demographic questions, 10-13 questions (the number depends on a subject's response on certain questions, which would then prompt follow-up questions) dealing with subject's mode of transportation to and on campus, and 14 questions about his or her perceptions and evaluation of CABS service, safety, and externalities, such as CABS' role in contributing to reduction of traffic on campus or making the campus "green". In all, there were up to 36 questions that a respondent could answer. It was estimated that a subject would require no more than 8 minutes to complete the survey.

We encountered some delays in administering the survey. The research study involves responses from human subjects, and the survey protocol required approval by The OSU Institutional Review Board (IRB) for human subject research. The application for approval required the project investigators to complete the CITI training before the application could be submitted. The research protocol describing the process to be followed to ensure that the privacy of the respondents would be protected was submitted for

approval on Oct. 30, 2009. The submitted protocol also included the e-mail message to be sent to the invited survey participants and the web-based questionnaire. The research team received an exemption from continued oversight by the board on Nov 7, 2008. SCS administered the survey soon after. Subjects were given approximately six weeks to complete the survey. The survey response rates by category of participants are provided in Table 2.5-1

Table 2.5-1: CTL transportation first wave survey response rates

Group	Surveyed	Responses	Response Rate
Faculty	4480	1233	27.52%
Staff	4479	1758	39.25%
Grad Students	2994	571	19.07%
UG Students	5999	837	13.95%
Overall	17,952	4,399	24.5%

To understand the adequacy of these response rates, we compare them with response rates to other OSU surveys. Recently, a survey was conducted regarding attitudes and perceptions of OSU undergraduate students on global warming. A random sample of 24900 undergraduate students was selected and 3570 responded to this survey. The 14.3% response rate is similar to the response rate of undergraduate students in our survey. Response rates from faculty, staff, graduate students, and undergraduate students for our survey and other OSU surveys are provided in Table 2.5-2, where it can be seen that our response rates are comparable to two other surveys that were devoted to information technology. In addition, our survey had similar undergraduate response rates as a recent survey devoted to Transportation and Parking issues at OSU, but substantially higher response rates from faculty, staff and graduate students. (It should be noted that undergraduate students formed the demographic group most related to the issues of the Transportation and Parking survey.) Thus, we believe that our surveys findings and conclusions have a high degree of validity.

The data collection process was fairly smooth. At the end of December 2008, SCS provided us with a data dictionary and the raw response data, with no identifiers on respondents or non-respondents.

Table 2.5-2: Response rates comparison with other surveys

Group	Fall 2008 CTL Transportation Survey	OIT 2009 CIO Technology Poll Questionnaire	OIT 2008 CIO Technology Poll Questionnaire	T&P 2008 COTA-CABS Survey
Faculty	27.52%	27.0%	26.0%	10.5%
Staff	39.25%	37.2%	33.1%	21.4%
Grad Students	19.07%	19.9%	23.2%	8.2%
UG Students	13.95%	13.2%	17.2%	13.4%
Overall	24.5%	23.4%	24.3%	13.8%

## CHAPTER 3. FINDINGS

We present the findings by thrust in this section.

### 3.1. Infrastructure development

After the planning and design process was completed, the installation of the integrated technologies for CABS by Clever Devices commenced in late summer of 2008 and continued for approximately one year before most of the issues and bugs were addressed. All 28 buses in the CABS fleet are fully equipped with the necessary hardware and software, the bus drivers are trained to use the on-board display, the real-time communication of AVL data to the operating center is on-going, the bus arrival time forecasting algorithm is functioning, and electronic messages signs at 10 major stops are in full operation. As of Autumn quarter 2009, the system with its integrated components is for the most part operating reliably, and use of the passenger information component is sharply increasing.

Initially we obtained Smart Bus data for one bus on one route for one day. The data, which was supplied in .csv format, allowed us to understand the data structures and the size of files that could be expected to be produced on a daily basis. We obtained a better understanding of data storage needs for the large amount (> 1 TB) of Smart Bus data that will be generated in the next two years. We are now finalizing the details that will make up our server order. The iterations we undertook this past year were also helpful in helping to establish the relationships between IT personnel in T&P and in Civil Engineering that will be required to implement the automatic data transfer protocols.

The pre-processing software seems to work well and the computational times for a day's worth of data are very low (in the order of seconds to tens of seconds). We will need further refinements when we start pre-processing data for other routes, but the experience we gained by working on the Campus Loop South route should reduce the time spent on the learning curve in the future. The value of the data produced is demonstrated in the use of the data in the following sections.

### 3.2. Data pre-processing

We were successful in using the codes we developed to process a small portion of Smart Bus APC and AVL data into formats that can be exploited by multiple users. Specifically, we processed the data for use in two aspects of our Federal Transit Administration (FTA) project, for outreach tasks described in Section 2.3.4, and for development of future APC- and AVL-related research and educational investigations. We are anticipating that the data transferred on a routine basis will be pre-processed on a regular basis and stored into a data base for access by the multiple users.

As described above, in our FTA project we are investigating and developing methods to estimate bus passenger origin-destination (OD) flows from APC data. We were successful in processing the first wave of Smart Bus APC data into formats that could be used in our FTA projects for various purposes:

- to serve as a set of inputs to different OD estimation methods for comparison of outputs
- to develop and test methods that automatically determine periods of homogenous OD patterns
- to determine inputs to simulation programs that are used to compare the accuracy of the different OD estimation methods

A second thrust of our FTA project is to explore the potential of using bus AVL data to indicate traffic conditions on surface streets. In that project, we developed an approach to detect indications of recurring congestion from AVL-derived bus speeds and used AVL data previously collected from our “home-made” AVL system to validate this method on the Campus Loop South route. After overcoming the “projection problem” mentioned in Section 2.1, we recently processed a first wave of Smart Bus AVL data into a form that can be used by the researchers on the FTA project to confirm the promising results with this new set of data. If these validation tests are successful, we will extend our empirical scope to include other OSU bus routes using Smart Bus AVL data that we are now collecting and which we will be collecting in the future.

In section 2.3.4 we described various measures derived from OD flow matrices that we wish to monitor in collaboration with OSU Traffic and Parking Services. We used our software to process the first wave of APC data into formats that allowed determination of these measures, as explained in Section 3.3.4.

The data used in the educational contexts discussed in Sections 2.4 and 3.4 were generated from the field-based data collection effort described in Section 2.3.2. In the future, we wish to generate the data for these and other course assignments from the Smart Bus APC and AVL data. We were successful in processing the first wave of Smart Bus data into formats compatible with the formats used in the exercises.

### 3.3. APC and AVL-based research and outreach activities

#### *3.3.1. Matching AVL data to bus schedules*

In Section 2.3.1 we described a potential improvement to the mathematical formulation of the approach we had previously developed to match AVL-based bus trajectories to bus schedules. In this revised formulation, weights on the deviations between empirical and scheduled times at a bus stop are adjusted to reflect the presence of a holding policy that was in effect on the CTL route for which we obtained AVL data.

We conducted an empirical study of the equal-weight formulation (which ignores the holding policy) to the revised formulation using AVL data from 1,726 CTL bus trips. Compared to the results produced in the equal weight case, the revised formulation produced 81 fewer mismatches of trajectories to schedules.



In addition to validating the expected improvement offered by our revised formulation, we also examined the conditions that are particularly prone to produce matching error when equal weights are assumed as a means to identify further methodological refinements. The occurrence of matching errors resulting from the assumption of equal weights becomes more pronounced when excessive delays in bus operations occur. Under such conditions, the likelihood of early bus arrivals to stops is much lower than the likelihood of late arrivals, whereas using equal weights would imply equal likelihoods. When the discrepancy between the two sets of likelihoods increases, the chances of encountering matching errors increases as well.

In our empirical study, we arbitrarily specified the weights so that they differed by a factor of two. The empirical results demonstrate that even this rather arbitrary specification can improve performance appreciably. However, our analysis motivates, additional analysis focusing on the sensitivity of the improved performance to the weighting and the development of an operational means to specify the weights in a more meaningful manner.

### 3.3.2. Performance assessment of OD estimation from APC data

In Section 2.3.2, we presented the methodology used to assess the performance of the easy-to-implement, but rather simplistic, IPF-with-null base procedure of estimating passenger OD flows from APC data. Our assessment is based on using the sum of squared difference measure  $P^1$  and the cumulative passenger distance traveled-based measure  $P^2$  to quantify the difference between the normalized OD flow matrices produced by the IPF-with-null base procedure and the true normalized OD flow matrices, calculating the relative performance  $RP$  defined in equation (2.3.2-8), and comparing these quantified measures to those obtained when using the other procedures listed in Table 2.3.2-1.

In Table 3.3.2-1, we present the number of trips (out of the 10 trips for which we collected empirical data) for which a procedure from Table 2.3.2-1 outperformed another procedure from the table by performance measures  $P^1$  and  $P^2$ . Not surprisingly, the OD matrices produced from the IPF-with-null base procedure (IPF-null) outperformed the null matrix (Null) or the refined null matrix (R-null) for all 10 trips. We also see that, according to  $P^1$ , the OD matrix produced by the IPF-with-null base procedure outperformed the on-board survey matrix (OBS) for all 10 trips. On the other hand, we note that when using  $P^2$ , the matrices determined from the on-board survey performed better than the IPF-with-null base matrices on all 10 trips, highlighting the value of using multiple measures for summary comparisons. Similarly, it is surprising that, according to  $P^1$ , the matrices produced when using the IPF-with-null base procedure were better than those produced when using the IPF procedure with the supposedly better base determined from the on-board survey (IPF-OBS) for 3 of the 10 trips. Again, when using  $P^2$ , the results are more in line with intuition: using the better base produced better results on all 10 trips.

Given these results, one might believe that  $P^2$  is a more appealing measure than  $P^1$ . However, when comparing the results produced by the null matrix to those produced by the refined null matrix, it is seen that measure  $P^1$  produces more intuitive results than  $P^2$ . Compared to the null matrix, the refined

null matrix uses additional information, namely, the boarding and alighting data. Therefore, one would expect the refined null matrix to be better than the null matrix. This is the case for all 10 trips when using  $P^1$ , but for only 4 of the 10 trips when using  $P^2$ . The seemingly paradoxical result that the null matrix does better than the refined null matrix when measuring performance by  $P^2$  but not by  $P^1$  can be explained when looking closely at the spatial pattern of the true trips in the empirical data.

TABLE 3.3.2-1: Pair-wise performance comparisons between procedures (number of trips in which each outperforms the others)

Procedure	Null	R-null	IPF-null	OBS	IPF-OBS
Based on performance measure $P^1$ (sum of squared differences)					
Null	–	0	0	0	0
R-null	10	–	0	0	0
IPF-null	10	10	–	10	3
OBS	10	10	0	–	0
IPF-OBS	10	10	7	10	–
Based on performance measure $P^2$ (passenger distance traveled)					
Null	–	6	0	0	0
R-null	4	–	0	0	0
IPF-null	10	10	–	0	0
OBS	10	10	10	–	0
IPF-OBS	10	10	10	10	–

Table 3.2.2-1 illustrates that, unsurprisingly, using the IPF-with-null base procedure clearly did better than using a null matrix or a refined null matrix as an estimate of the OD matrix, and that the OD matrices produced from the IPF procedure were mostly improved when using a better base. Whether the matrices produced when using the IPF-with-null base procedure did better than those produced directly from an on-board survey is not clear from the table: They performed better 10 of 10 times according to  $P^1$  but 0 of 10 times according to  $P^2$ . In addition, the table shows that, according to both

measures, using the on-board survey to produce a base for the IPF procedure always performed better than using the on-board survey directly.

In Table 3.2.2-2, we show *RP* summaries across the ten trips by procedure based on each of the two performance measures. (As discussed above, the *RP* of the null matrix is zero for all trips, and the *RP* of the true matrix is 1; therefore, neither of these measures is included in the table.) Using the refined null matrix – *i.e.*, using the boarding and alighting data to possibly improve the null matrix – produced an improvement of at most 19% of the possible improvement, whereas using the simple IPF-with-null base procedure with the boarding and alighting data improved performance by between 60% and 89%.

TABLE 3.2.2-2: Relative performance *RP* summaries across 10 trips

Procedure	Average	Minimum	Maximum
Relative performance measure using $P^1$ (sum of squared differences)			
R-null	0.10	0.03	0.16
IPF-null	0.70	0.60	0.89
OBS	0.33	0.21	0.48
IPF-OBS	0.74	0.55	0.88
Relative performance measure using $P^2$ (passenger distance traveled)			
R-null	-0.04	-0.21	0.19
IPF-null	0.68	0.60	0.78
OBS	0.80	0.72	0.87
IPF-OBS	0.89	0.83	0.92

Comparing across procedures, the *RP* results indicate that using the IPF-with-null base procedure markedly improved performance compared to using the null matrix or the refined null matrix directly (*i.e.*, without transforming these matrices with the IPF procedure). Using the better base obtained from the on-board survey as input to the IPF procedure improved performance further, but the marginal improvement is less marked. As mentioned above when discussing the results in Table 3.2.2-1, directly using the matrix obtained from on-board survey to determine the OD flows performed better than using the results produced from the IPF-with-null base procedure according to the passenger distance traveled measure  $P^2$ . However, according to Table 3.2.2-2, the improvement in performance was slight. When considering the sum of squared distances measure  $P^1$ , the decrease in performance, compared to that produced by the IPF-with-null base procedure, is of greater magnitude.

### 3.3.3. Development and application of bus operations simulation

The preliminary validation of the simulation revealed that the simulation scenario where the dwell time and special point delay distributions are derived from dwell times and delays calculated using the harmonic average speed look-up tables in general resulted in the more accurate results in terms of matching travel times along the route.

As discussed in the application portion of Section 2.3.3, we determined the absolute value of the difference between the estimated and (simulated) true dwell time for each simulation replication, for each of the 19 stops on the route, and for each spatial or temporal sampling interval considered. We averaged these absolute values across the replications and stops for a given sampling interval to summarize the dwell time estimation error for the sampling interval.

We present these results in Figure 3.3.3-1. The curve on the left of the figure corresponds to the spatial sampling interval indicated on the left vertical axis, whereas the curve on the right corresponds to the temporal interval indicated on the right vertical axis. We arranged the heights of the left and right vertical axes to correspond to equal quantities of data generated by the corresponding spatial and temporal sampling intervals. For example, the spatial sampling interval of 100 m on the left axis would generate 83.2 AVL points per bus trip on the 8320 meter long route. The 100 m height on the left axis corresponds to a 26 sec height on the right axis, since a 26 sec sampling rate would produce the same number of AVL signals on an average bus trip, where the average running time of the trip is 36 minutes.

To illustrate the interpretation of the figure, consider a 100-m spatial sampling rate. Entering the left curve (that produced when simulating spatial sampling), the corresponding average (absolute) dwell time error is approximately 4 seconds. Entering the curve produced when simulating temporal sampling (the right curve) at the same height (*i.e.*, at the temporal sampling interval of 26 seconds, which corresponds to the same number of AVL points generated per bus trip), the error is approximately 16 seconds. The spatial sampling approach produced much smaller average error than did the temporal sampling approach (4 seconds, compared to 16 seconds). The shift to the right of the temporal sampling curve indicates that the spatial sampling approach outperforms the temporal sampling approach for all “equivalent” intervals. (We are presently investigating the slight nonmonotonic behavior of the temporal sampling curve at relatively large values of temporal sampling.)

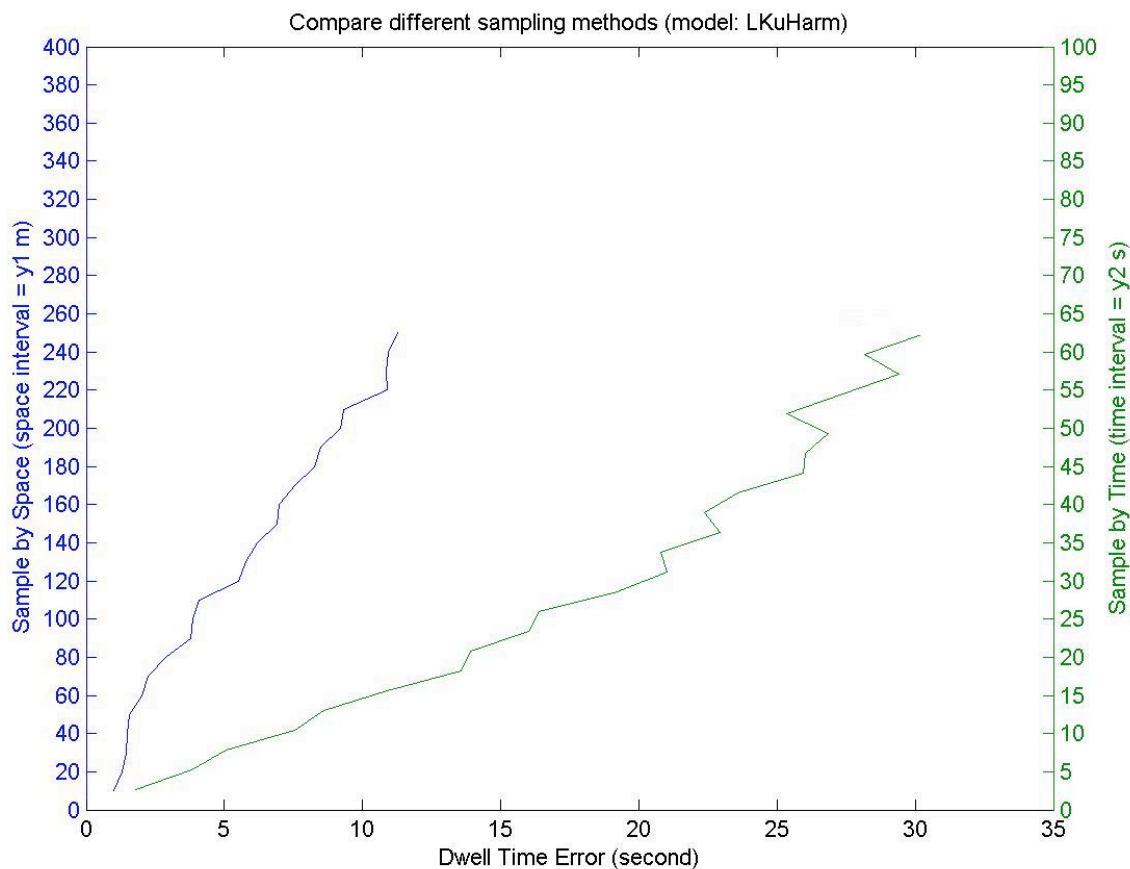


Figure 3.3.3-1: Average absolute dwell time errors, across simulation replications and bus stops, as a function of spatial sampling interval or temporal sampling interval

#### 3.3.4. Development of outreach products

As explained above, we processed the first wave of Smart Bus APC and AVL data into information that could be used to produce quantitative products and measures that we plan to monitor in collaboration with OSU Traffic and Parking (T&P). This past year, we also developed preliminary versions of these products and measures, which we will soon present to T&P. We report on these results in this subsection.

With the help of research assistants from our FTA project, we used the data processed in the NEXTRANS project to produce 1003 trip level origin-destination (OD) flow estimates on OSU's Campus Loop South (CLS) route. We aggregated these into a normalized matrix that provides the probability that a passenger chosen at random from this set of 1003 trips used the designated OD pair. The normalized matrix is presented in Table 3.3.4-1.

To investigate the potential of using the measure discussed in section 2.3.4 to investigate the similarity of OD matrices, we considered a morning, 7-to-10 AM period, an afternoon, 2-to-5 PM period, and determined normalized OD matrices from the 1003 APC-derived trip level matrices that fell in the appropriate period by day of the week. In Table 3.3.4-2a, we show the dissimilarity measure values between the OD matrix for the morning period of one day of the week and the OD matrix for the morning period of another day of the week. In Table 3.3.4-2b, we present the dissimilarity values for day-of-week pairs for the afternoon periods. Larger values correspond to more dissimilar matrices. In comparing the values in the two tables, we notice that the values obtained when comparing Friday afternoon matrices to the other afternoon matrices are noticeably larger than when comparing any other pair, indicating that the greatest day-of-week difference in passenger trip pattern is associated with Friday afternoon.

To support the use of the dissimilarity measure, we determined the value of the measure between a matrix produced in the morning on a given day of the week and the afternoon matrix for the same day of the week. A large proportion of morning CLS riders park in the remote West Campus lot and ride the bus to main campus. In the afternoon, they travel from main campus to the West Campus lot. As such, the expectation is that the OD patterns would be very different for these two times of day, and the dissimilarity measure would, therefore, be large. The dissimilarity measures determined in this way, presented in Table 3.3.4-2c, are indeed much larger than those in the previous two tables, supporting the validity of this measure.



Table 3.3.4-2a: Dissimilarity measures for day-of-week pairs of 7-10 AM OD matrices

Day of week	Mon	Tue	Wed	Thu	Fri
Mon	0	1.3	0.86	1.09	0.86
Tue	1.3	0	1.01	0.78	0.86
Wed	0.86	1.01	0	1.15	0.89
Thu	1.09	0.78	1.15	0	1.01
Fri	0.86	0.86	0.89	1.01	0

Table 3.3.4-2b: Dissimilarity measures for day-of-week pairs of 2-5 PM OD matrices

Day of week	Mon	Tue	Wed	Thu	Fri
Mon	0	1.46	1	1.2	2.63
Tue	1.46	0	1.11	0.8	2.15
Wed	1	1.11	0	1.05	1.77
Thu	1.2	0.8	1.05	0	1.74
Fri	2.63	2.15	1.77	1.74	0

Table 3.3.4-2c: Dissimilarity measures: 7-10 AM vs. 2-5 PM OD matrices on same day-of-week

Time Period	PM (14:00-17:00)					
	Weekday	Mon	Tue	Wed	Thu	Fri
AM(7:00-10:00)	Mon	6.95	6.49	7.09	6.2	5.3
	Tue	6.16	5.63	6.04	5.17	4.41
	Wed	6.58	6.34	6.78	5.79	4.61
	Thu	5.46	4.67	5.24	4.57	3.91
	Fri	4.83	4.68	5.06	4.4	3.77

As mentioned in Section 2.3.4, we are also proposing to monitor the distribution of bus passenger trip distances over time, where the distribution is produced from the APC-derived OD matrices and the distances between stops. To support the ability of monitoring this distribution, we validated the ability of the APC-derived OD matrices to capture the decrease in the proportion of short bus trips from Winter to Spring quarters which we had observed in the true OD data. We produced quarter-specific OD



matrices using the IPF-with-null base procedure described in Section 2.3.2 and boarding and alighting data obtained from the manual, onboard procedure described in that section. Based on the true OD data collected, the proportion of passenger trips that were less than one mile was 5.5% lower in Spring quarter than in Winter quarter. Using the matrices produced from the boarding and alighting data, the Spring quarter proportion was 3.5% lower than the Winter quarter proportion. Given the simplicity of the IPF-with-null base procedure, we consider this to be fairly good agreement and encouraging of the potential to monitor trip distance distributions from the APC data.

The final OD matrix-derived measure that we developed this year was the expected passenger time on the bus, conditional on boarding or alighting stop. Figures 3.3.4-1a and b, we present the expected passenger time on the bus by boarding and alighting stop, respectively. When considering the time by boarding stop (Figure 3.3.4-1a), we notice the highest time, as expected, from the first boarding stop (stop 4, which is an aggregation of the remote West Campus parking lot stops). We see a cluster of mostly decreasing times for boarding stops 5 through 9, which are stops that progressively approach Central Campus, and a second cluster of approximately similar times for the remaining stops, serving Central Campus.

The clusters of times in Figure 3.3.4-1b for alighting stops 4-10, 11-12, 13-20, and 21 (which is the aggregated West Campus parking lot alighting stop) are similarly compatible with the sections of campus served by the CLS route. More specifically, in the morning the most common origin for destinations up to stop 10 is stop 4 (the west campus parking lot). Thus, travel times progressively increase for alighting stops 5 through 10. The travel time function plateaus at alighting stops 11 and 12. This result is consistent with the expectation that some trips destined to these stops are more likely to be originating from stops closer than stop 4, resulting in lower travel times for some travelers (while those originating from stop 4 have longer travel times). Travel to stops 13 through 20 mostly reflects local short trips, as travelers destined to these stops from the remote stop 4 are much more likely to take an alternative route (Campus Loop North) which reaches these stops more quickly; thus, the lower travel times associated with alighting at these stops. Finally, travelers heading to stop 21 (the West Campus parking lot) are engaged in substantially longer trips because of the relatively long distance from stop 20 to 21.

The compatibility between the preliminary results produced and our knowledge of the characteristics of the CTL route gives us confidence that the approach can produce meaningful measures, which can be monitored through time. We will soon be presenting these concepts and preliminary results to the OSU T&P director and operations staff.

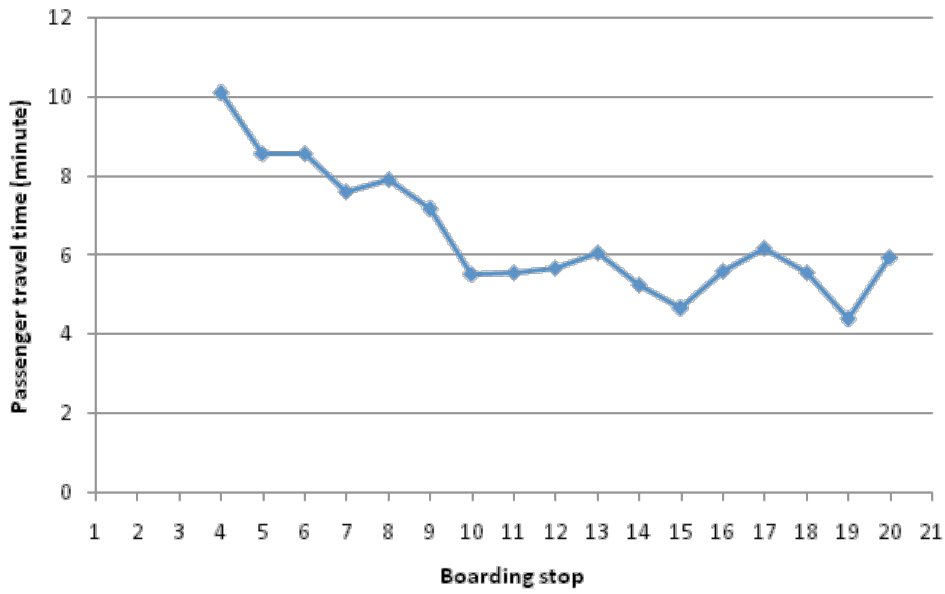


Figure 3.3.4-1a: Expected passenger travel time by CLS boarding stop determined from APC-derived OD matrices and AVL data

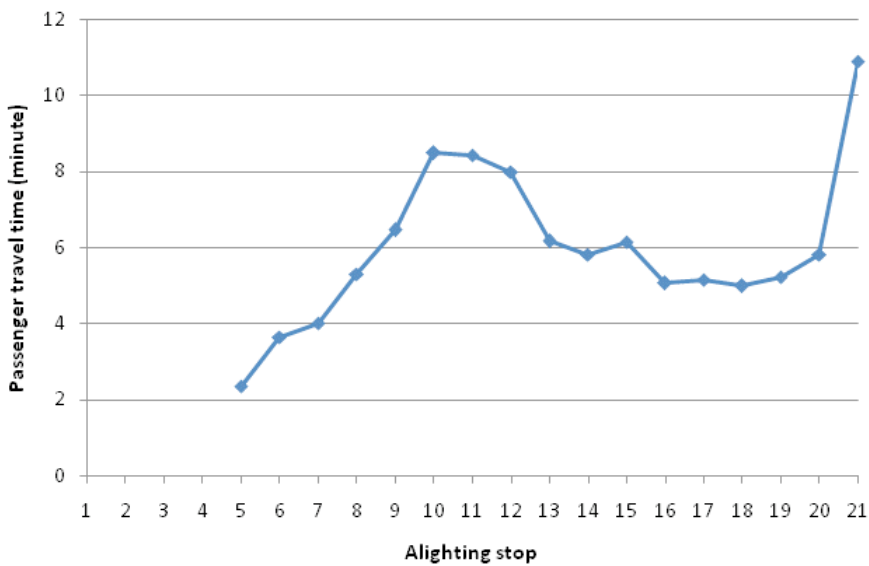


Figure 3.3.4-1b: Expected passenger travel time by CLS alighting stop determined from APC-derived OD matrices and AVL data

### 3.4. Educational use of CTL

As described in Section 2.4, CTL data and results were incorporated in two OSU transportation courses in this past year – a large course (*CE5 70*) required of all Civil Engineering undergraduate students, including both transportation majors and non-majors, and a graduate course (*CE 873*) for students interested in transportation systems from multiple departments.

There were 105 students enrolled in *CE 570* this past year. The lecture used to describe the campus lab and the assignment consumed approximately 30 minutes. The lecture was given at the end of the quarter, and the extra time required was obtained from the “slack” built into the course to accommodate special topics such as this.

Incorporating the CTL material into the course was considered successful. Most students received full credit on the assignment, indicating that the material was successfully presented and received. Nevertheless, there were enough students, even those who performed well in other aspects of the course, who made mistakes that were sufficiently similar in nature to illustrate a lack of understanding on a specific concept – namely, the need to integrate various components of passenger travel time and conditional mathematical expectations to form the unconditional expectation of passenger travel time from origin stop to destination stop. This concept is important to understanding the analytical method covered and is not specific to the empirical component introduced for the first time this past year. In this way, incorporating the CTL component was valuable in highlighting this previously unnoticed relative difficulty. (We call this a “relative” difficulty, since most of the students did not seem to have trouble with this concept.)

There were 13 students in *CE 873* this past year. Two of these students undertook the CTL-based project described in Section 2.4. The estimation results produced in the project mostly revealed what we had already found in our focused investigation described in section 3.3.2. Specifically, the results showed a significant improvement when using the IPF procedure with the boarding and alighting data rather than simply using the boarding and alighting data through the “refined null” matrix, and a further significant improvement when using the matrix derived from the onboard survey, rather than the null matrix, as the base in the IPF procedure. The results also revealed some aspects we had not considered. Whereas the results did not indicate a day-of-week effect, they did show a slight time-of-day effect: Matrices produced from morning trips were slightly better than matrices produced from afternoon trips. The results also indicated, although weakly, that matrices produced from trips with higher volumes performed better than those produced from trips with lower volumes. We will consider these results, produced in an educational context, in our future research.

The students who undertook the project seemed to gain the intended insights into model estimation and interpretation as much as, or more so, than those who undertook other projects. To obtain the results summarized above, the students estimated multiple specifications, where the results of one specification were used to inform subsequent specifications. This type of open-ended use of the model is one of the primary objectives of the extended project.

In addition to the impact on the two students who undertook this project, we were happy with the more general educational impact of this project. The student teams present the setting, design, and results of their projects orally to the rest of the class and the instructor. In their oral presentation, the students involved with the CTL project did an excellent job of communicating the concept of the CTL, the data collection effort, and the various components of the OD estimation procedure in their oral presentation. This was the first time most of the students in the class were exposed to the CTL. A prerequisite course for *CE 873* exposes the students to the IPF procedure in a different context. The *CE 873* project allowed the students to see this procedure applied in a practical application (estimating bus passenger origin-destination flows) using empirical data collected in a local setting (the OSU campus) with which they are familiar. This type of reinforcement is considered particularly valuable.

### 3.5. Perceptions and attitudes survey

In this section, we report on some interesting survey responses to travel behavior, perception, and evaluation questions. All the reported results are based on analysis that indicates statistical significance where applicable. Our findings on a few perceptions and evaluation issues are similar to those obtained in the 2008 OSU Transportation and Parking Services (T&P) survey, which was conducted for a very different purpose. However, our survey produced results that we consider of interest to the general transit community that were not addressed in the 2008 T&P survey. Therefore, a description of the methodology and findings will be presented in a paper under preparation for possible publication.

Approximately 30% and 4%, respectively, of the undergraduate and graduate respondents live on-campus. Thus an overwhelming proportion of our subject population commutes to campus. In addition, the respondents represent a cross-section of the campus community, which is spatially distributed across the large OSU campus (consisting of a core surrounding by spread-out areas) as well as various academic and administrative groups.

#### 3.5.1. Travel mode behavior

Some interesting highlights about the travel mode behavior of our respondents are as follows:

- Approximately 60% of respondents never use CABS, and approximately 30% ride CABS occasionally, where as only 10% ride CABS regularly.
- Approximately 90% of the respondents have a car in the Columbus area.
- Most of the respondents drove a car to campus as shown in Table 3.5-1, while those living on-campus walked to their campus destination.
- Approximately 69% of the respondents who do not have a car in Columbus consider CABS is valuable or highly valuable to their travel needs compared to approximately 36% of those who have cars.
- Approximately 49% of the respondents were familiar with one or more routes on CABS service, whereas approximately 45% knew that CABS existed, but were not familiar with any of its routes. Approximately 6% did not know that CABS existed.

- Approximately 23% of the respondents who never use CABS were familiar with one or more routes on CABS service, as compared to approximately 85% of the respondents who use CABS only occasionally and 99% of the respondents who use CABS regularly.

Table 3.5-1: Transportation-to-campus mode choices of survey respondents

Travel Mode	Drive alone	Share a car	COTA	CABS	Bike	Walk
%	67	7	4	3	5	13
Counts	2825	286	176	133	190	539

### 3.5.2. Perceptions and evaluation analysis

The survey contained fourteen statements designed to elicit respondents' attitudes toward CABS. The respondents were asked to respond to each question using a 5-point scale, labeled as 1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree, and 5: Strongly agree. The statements are paraphrased below:

- EQ1- Having CABS service reduces the amount of car traffic on campus...
- EQ2- Providing bus service around campus should be part of OSU's efforts to promote a green campus...
- EQ3 - CABS offers service that is valuable to my travel needs...
- EQ4 - I feel safe walking to CABS stops...
- EQ5 - I feel safe waiting for CABS buses...
- EQ6 - I feel safe riding CABS buses.
- EQ7 - CABS bus drivers are professional...
- EQ8 - CABS buses are comfortable...
- EQ9 - CABS routes are reasonable...
- EQ10 - My travel time to reach my destination using CABS is reasonable...
- EQ11 - My waiting time for CABS buses is reasonable...
- EQ12 - Accessing information about CABS service (e.g., routes, frequency of service, hours of operation) is easy...
- EQ13 - CABS is reliable...
- EQ14 - Overall, I am satisfied with CABS...

We note that twelve of these statements can be classified into three perception categories:

- Category 1: Environmental Issues (EQ 1 - 2)
- Category 2: Safety Issues (EQ 4 - 6)
- Category 3: CABS Service Quality Issues (EQ 7 - 13)

The other two statements concern the value of CABS to individual travel needs (EQ3) and an overall evaluation of CABS (EQ14).

Response rates to EQ 1-3 were very high (greater than 75%). The other statements – which relate to specific aspects of CABS, such as safety, CABS service quality issues, and overall evaluation of CABS – received much lower response rates, since many of the respondents use CABS rarely or not at all. The distribution of the 5-point responses across the fourteen statements is summarized in Table 3.5-2. Each statement is associated with three rows in the table. The first row, in which EQ# appears, lists the five possible respondent responses. The second and third rows provide, respectively, the proportion and number of individuals responding to the statement.

Some of the interesting observations that can be made based on this table and from an investigation of numerical association between individual responses (described in more detail below) are the following:

- CABS' value to individual travel needs received a lower rating than did other evaluation items. 35% of respondents do not believe CABS is valuable to their travel needs (those who choose 1 or 2), while only 39% believe CABS is valuable (those who choose 4 or 5).
- CABS received its highest rating in response to its contribution to promoting a green campus. Only 3% of respondents do not recognize CABS' role in promoting a green campus (those who choose 1 or 2), while 90% recognize such a role (those who choose 4 or 5).
- Responses to statements about safety issues (EQ 4,5,6) are closely associated with each other; that is, an individual respondent is likely to provide a similar rating to all three of these statements. Among these three issues, safety of walking to a CABS stop and safety of waiting for a CABS bus have stronger association than do the two other pairs.

Table 3.5-2: Summary of responses on perception and evaluation questions

EQ1	1	2	3	4	5
	0.02	0.08	0.18	0.42	0.3
	75	285	623	1492	1067
EQ2	1	2	3	4	5
	0.02	0.01	0.07	0.36	0.55
	79	47	257	1394	2130
EQ3	1	2	3	4	5
	0.11	0.24	0.26	0.2	0.19
	375	805	883	678	642
EQ4	1	2	3	4	5
	0.01	0.04	0.19	0.41	0.34
	39	112	537	1157	944
EQ5	1	2	3	4	5
	0.01	0.05	0.2	0.4	0.33
	37	136	558	1103	908
EQ6	1	2	3	4	5
	0.01	0.02	0.17	0.39	0.41
	27	41	449	1055	1106
EQ7	1	2	3	4	5
	0.01	0.03	0.24	0.43	0.29
	30	81	597	1073	716

EQ8	1	2	3	4	5
	0.01	0.05	0.27	0.47	0.21
	27	120	671	1179	535
EQ9	1	2	3	4	5
	0.01	0.06	0.24	0.49	0.21
	35	138	587	1225	512
EQ10	1	2	3	4	5
	0.03	0.09	0.27	0.44	0.16
	72	228	659	1075	401
EQ11	1	2	3	4	5
	0.04	0.12	0.32	0.4	0.12
	92	282	778	977	284
EQ12	1	2	3	4	5
	0.03	0.12	0.29	0.4	0.17
	70	299	759	1023	432
EQ13	1	2	3	4	5
	0.02	0.04	0.27	0.47	0.2
	38	96	676	1173	494
EQ14	1	2	3	4	5
	0.01	0.04	0.26	0.5	0.19
	38	95	669	1282	498

In addition, we have found the following grouped by pertinent category:

*Satisfaction with CABS' service*

- People who used CABS more often are more satisfied with CABS service.
- People who come to campus by CABS, COTA (the Columbus area transit service) or walk are more satisfied with CABS service than those who travel by car or bike.
- People who spent more than one hour/day on the Internet are slightly more satisfied with CABS' information accessibility than those who spent less than one hour/day on the Internet.

*CABS' contribution to environment and traffic reduction*

- More than half of the people who said CABS had little value to their travel needs nevertheless expressed an appreciation of CABS' positive environmental contribution and of its contribution to reducing traffic on campus.
- Undergraduate students had a slightly lower appreciation of CABS' positive environmental contribution than did the other groups.
- The frequency of using CABS had little impact on people's positive appreciation of CABS' environmental contribution.
- People appreciated CABS' positive environmental contribution more than CABS' traffic reduction contribution.
- People who use or have used metropolitan public transportation (MPT) appreciated CABS' positive contribution to the environment and to traffic reduction more than those who do not or have not used such public transportation.

*CABS usage*

- People who used or are using MPT were more likely to use CABS on campus.
- Conditioning on CABS users, whether or not a person uses MPT did not affect the distribution of frequency of using CABS.
- People who came to campus by CABS, COTA or walked were more likely to use CABS while on campus than those who came to campus by other modes like the car or bike.
- People who came to campus by bike were more likely to use CABS while on campus than those who drove to campus, but less likely to use CABS while on campus than those who came to campus by CABS, COTA, or walked.

*CABS safety*

- People feel safer when riding CABS than when walking to a CABS stop or when waiting for a CABS bus.
- People feel equally safe when they are walking to a CABS stop or waiting for a CABS bus.
- While waiting for a CABS bus, a longer waiting time tends to make people feel less safe.



Overall evaluation

The responses to the overall evaluation item (EQ14) are closely associated to EQ 9, 10, 11, 13. Thus, a person is likely to give higher overall evaluation to CABS if he or she appreciates the reasonableness of CABS routes, travel times, waiting times, and reliability.

## CHAPTER 4. CONCLUSIONS

Despite the delays encountered in implementing this large “Smart Bus” system and exploiting the rich data it generates for multiple academic purposes – or, perhaps, in light of the seeming inevitability of such delays – we have made significant progress in this past year toward developing a Campus Transit Lab (CTL) that can serve as a flagship, living lab for research, education and outreach. Furthermore, we have proceeded with various research activities resulting in insightful findings of value to researchers, transit planners, and transit operators. T&P personnel have told other universities of the benefits of the Smart Bus system to its operations. They have also emphasize the value they are deriving from taking advantage of the system to deepen their partnership with the “academic side” of the university. (The conference in which many of these remarks were made publicly occurred after the end of the reporting period for this year’s project. We will report on the conference in later reports.)

It took much longer than expected for OSU Transportation and Parking Services (T&P) to complete the bidding process, select the contractor and then work with the contractor to install and test the integrated information technology “Smart Bus” system on OSU’s Campus Area Bus Service (CABS). Testing identified many startup glitches, as might be expected with a project of such a magnitude and complexity. Furthermore, some elements continue to be implemented and refined. For example, just recently, after the close of our reporting period, but before submission of this report, T&P introduced the text messaging capability of the CABS traveler information system referred to as “Transportation Route Information Program” (TRIP), <http://tp.osu.edu/cabs/trip.shtml>. And, while 10 electronic message signs are installed and are in operation at the major stops on the system, another set of signs are planned for installation in the near future. The formal announcement of TRIP was originally planned to be made in Spring 2009, but doing so was postponed until Autumn 2009 in light of the delays in addressing the glitches. In fact, a note on the TRIP site still states that it “is currently going through testing, and bus information displayed may not be accurate.”

In addition, given the complexities of using automated data collection systems, it has taken the project team longer than anticipated to obtain a preliminary set of reliable AVL and APC data that can be used for research, education, and outreach. Furthermore, developing an understanding of the AVL and APC data, and validating it with the ground truth collected by our team on selected bus trips added to the effort in implementing the soft infrastructure to support research, education, and outreach. The formats of the data from bus AVL and APC sensors are primarily driven by operational needs. As a result, the data sets are often not available in a readily usable form for other than operational needs and contain fields that are cumbersome to interpret. Delays and time-consuming iterations seem inevitable when implementing a large project in a new environment. Indeed, we have been present when T&P personnel have told other universities that they should expect unforeseen delays if they plan to

implement such a system. We would similarly conclude that obtaining, processing, and using data from such a system for research, education, and outreach will likely encounter unanticipated delays.

We have made progress in developing the means to transfer and store data automatically and on an ongoing basis and in “working out the bugs.” We are especially pleased with our ability to process the data we have received. Through a combination of conceptual design and trial and error, we have successfully processed APC and AVL data into a general form that can be used by multiple users for a variety of applications. There are still several fields that are yet to be deciphered, and there exist some elements of information (*e.g.*, the route the bus is serving) that we believe must be present in the data structure but have yet to incorporate in an automatic fashion. (For the time being, we have developed somewhat labor-intensive means of identifying the bus route from the data.) We will be clarifying many of these issues in the next iteration of interactions with the contractor in the near future.

We have worked around present start-up difficulties, and the APC and AVL data we have collected and processed this year have been valuable in several ways. We have converted the data into preliminary indicators of OSU bus passenger activity that we will monitor in collaboration with OSU Traffic and Parking Services (T&P). We have also processed the data into formats that can be used in the course exercises we piloted with manually collected data in this past year. And, the data have supported and enlightened multiple aspects of a Federal Transit Administration project we are conducting.

The conceptual development of the educational and outreach activities we produced this past year were important steps toward broadening the use and appeal of the CTL. We implemented CTL-based concepts into two very different courses. The course instructors believe that the ability to introduce the concepts in a practical setting on a system that is “just outside the doors of the classroom” offers a unique and valuable means of exposing the students to the advantages of AVL and APC technologies for bus transit planning and operations, and of advanced data collection technologies for transportation systems in general. Using empirical CTL data in quantitative exercises is also valuable as an understandable means of reinforcing general principles presented in courses. Although we manually collected the data in the exercises this past year, now that we can produce the data from the AVL and APC systems in a reliable and repeatable manner, we will be using Smart Bus data in the future. Moreover, the success of these educational efforts is motivating us to develop and implement additional CTL-based educational modules in the upcoming year.

To indicate our progress in understanding and processing the Smart Bus data, we mentioned above our ability to produce APC- and AVL-derived measures of passenger activity which we will use for monitoring service with T&P. Equally important was our development and validation of these measures. To our knowledge, monitoring measures of dissimilarity in origin-destination (OD) flow matrices, trip distance distributions, and expected bus passenger travel time conditional on boarding or alighting stop have not been proposed previously. Indeed, the measures are all derived from OD flow matrices, which previously could not be produced on a routine basis before the relatively recent availability of spatially and temporally extensive APC data. Because of the ongoing and spatially extensive APC data that are now available, these matrices can now be produced for monitoring purposes, and we see the CTL as

means to demonstrate the value of doing so. Since the APC-derived OD matrices are only approximations of the true OD matrices, the quality of the approximations would affect the quality of the measures we developed. Therefore, the correspondence of the empirically estimated measures to “ground truth” data we collected manually and to our knowledge of general passenger activity on the OSU campus system is noteworthy as a preliminary validation of the ability to produce meaningful measures on an ongoing basis.

We also conducted multiple research activities related to the CTL that make important contributions. The method we had developed to match AVL data to bus schedules are unique, and the refinements we made appear to have generated interest in the academic community. The refined approach designed to incorporate the effects of bus holding operating policies and resulting behavior into our mathematical formulation produced much improved results. The matching problem becomes more complicated when more complex real-time interventions are involved. Examples include short-turning and schedule swapping (via overtaking) in mid-route. Such interventions need to be identified using additional information for the matching method to produce reliable results. Incorporating such information in the formulation of the assignment-based matching method would be a valuable extension of interest to both the research and practice communities.

Our empirical study on the quality of OD matrices determined from a procedure that can be readily implemented to take advantage of existing APC data is important. The empirical results were derived from only ten bus trips on one route. However, since there are very few empirical studies where true OD information is obtained, ours can be considered to be large. It was surprising that the seemingly simplistic IPF-with-null-base procedure produced results of similar quality to those obtained from an on-board survey. Since on-board surveys have traditionally been the primary means of directly obtaining OD matrices, this roughly equivalent performance indicates that much could be gained from using readily available APC data, even when applying the simple IPF-with-null-base procedure. In addition, it was found that the results of the IPF procedure were markedly improved when using the on-board survey derived matrix, rather than the null matrix, as a base. The result is a strong indication that combining on-board survey information with increasingly available APC data can lead to OD matrices that are markedly better than those presently available.

In addition, we believe that the design of the study – through the use of multiple metrics, the development of the intuitively appealing relative performance summary, and the direct comparison to other reference procedures – makes an important methodological contribution. The developed methodology should be used in next steps on this project with more extensive empirical data and on other routes to investigate if the relatively good performance of the IPF-with-null base procedure we observed is dependent on specific travel patterns or route structure. Other methods aimed at estimating OD flows that are more general and, thus, more complex than the IPF procedure are worth investigating, and we are developing these concepts in a FTA sponsored project for applications on larger urban transit systems. We will apply those concepts to APC data on OSU routes being studied in CTL as a stepping-stone in this regard. This type of coordinated support activity highlights the role of CTL

as a test bed that can support multiple, yet distinct efforts and serve as a launching pad where proven methods can be applied to larger systems.

We also believe we have developed a useful tool in our bus operations simulation. While additional validation is underway, the initial validation results indicate that the simulator is capturing CABS operations realistically. Especially once further validation is completed and possible refinements are developed, we expect to use the simulation framework to address multiple research questions on this project in the future. For example, this simulation would be effective in evaluating various approaches to schedule design. There is a re-emerging interest in headway-based schedules, which can now gain traction in light of the prevalence of real-time AVL systems. Our simulation tool would also be valuable in investigating and evaluating the performance of various real-time operations strategies, which automated monitoring and communications systems are also making feasible,

Our application this year of the simulation program to compare AVL signal sampling on a distance-based approach to sampling on a time-based approach in terms of the accuracy in dwell time estimation showed the superiority of the distance-based approach. Whether in the context of providing real-time passenger information or in supporting off-line analyses, sampling is a critical component of AVL systems. However, to our knowledge, such comparisons are not available in the literature. Dwell time estimation would only be one component to consider, of course, but it was illuminating to note the extent to which the distance-based approach outperformed the time-based approach. We are presently developing a behavioral explanation of this result, and we are investigating the simulation-based results in finer detail – for example, comparing distance- to time-based sampling performance separately at stops with long dwell times and at stops with short dwell times, or investigating the impact of the characteristics (length, variability in travel time) of the sections immediately upstream or downstream of the stop on the comparative performance.

Finally, our survey results are proving informative. The original intent of the survey was to serve as a benchmark (the “before” case) for perceptions and attitudes in a before- and after-implementation of the Smart Bus and passenger information system. We still intend to use it as such. However, the results are providing interesting information on attitudes of users and nonusers toward CABS, the OSU bus system. We believe some of our findings, such as perception toward the various elements of CABS by demographic group, will be interesting to T&P administrators. (We will report in the future on our discussion of the survey results with T&P, which occurred after the close of this reporting period.) However, some of the results – such as the recognition of the positive impact of a bus system on the environment and on reduced traffic, and the difference in this recognition among demographic groups – are of general interest to the transit and multimodal transportation community.

In summary, we believe that the multi-thrust approach we undertook this past year was productive in leading to the establishment of the OSU Campus Transit Lab (CTL) as a unique, recognized, and valuable infrastructure for research, education, and outreach. More progress is required, and we believe that it would be beneficial to proceed in a similarly multi-faceted approach devoted to:

- developing the means to collect, process, and make available AVL and APC data on a routine basis,
- routinely making the data available to multiple users,
- using the data to support multiple research activities (sponsored inside and outside of NEXTRANS), new educational activities, and ongoing bus system monitoring in collaboration with OSU T&P,
- conducting several research studies related to improved bus transit planning and operations that can occur through innovative uses of these data.

## REFERENCES

- Ben-Akiva, M.E., Macke, P.P, and Hsu P.S. (1985) Alternative Methods for Estimating Route-Level Trip Tables and Expanding On-Board Surveys. *Transportation Research Record*, No. 1037, pp. 1-11.
- Berger, J.O. (1985) *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York, 2nd edition.
- Bishop, Y.M.M, Fienberg, S.E., and Holland, P.W. (1977) *Discrete Multivariate Analysis*. The MIT Press, Boston, MA.
- Cascetta, E., and Nguyen, S. (1989) A unified framework for estimating or updating Origin/Destination matrices from traffic counts. *Transportation Research*, Vol. 22B, pp. 437-455.
- Chan, J. (2007) Rail OD Estimation and Journey Time Reliability Metrics Using Automated Fare Data. MS thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Deming, W.E., and Stephan, F.F. (1940) On a Least Square Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known. *Annals of Mathematical Statistics*, Vol. 11, pp. 427-444.
- Feinberg, S.E. (1970) An Iterative Procedure for Estimation in Contingency Tables. *Annals of Mathematical Statistics*, Vol. 41, pp. 907-911.
- Furth, P.G., Hemily, B., Muller, T.H.J. , and Strathman, J.G. (2006) Using archived AVL-APC data to improve transit performance and management. *Transit Cooperative Research Program Report 113*, Transportation Research Board, Washington, D.C.
- Furth, P.G., and Navick, D.S. (1992) Bus Route O-D Matrix Generation: Relationship Between Biproportional and Recursive Methods. *Transportation Research Record*, No. 1338, pp. 14-21.
- Hillier, F.S., and Gerald J.L. (2001) *Introduction to Operations Research*. McGraw-Hill.
- Ji, Y., Mishalani, R.G., and McCord, M.R. (2009) Assignment-based Methodology to Match Automated Vehicle Location Transit Bus Trip Trajectories to Schedules. CD proceedings of the 11th International Conference on Advanced Systems for Public Transport, Hong Kong.
- McCord, M., Mishalani, R, Goel, P., and Strohl, B. (2009) Empirical Comparative Assessment of the IPF Procedure for Determining Bus Route Passenger OD Flows, Submitted for possible presentation at Transportation Research Board 2010 Annual Meeting and publication in *Transportation Research Record*.
- Mosteller, F. (1968) Association and Estimation in Contingency Tables. *Journal of the American Statistical Association*, Vol. 63, pp. 1-28.

Simon, J., and Furth, P.G. (1985) Generating a Bus Route O-D Matrix From On-Off Data. *Journal of Transportation Engineering*, Vol. 111, pp. 583-593.

Stephan, F.F. (1942) Iterative Methods of Adjusting Sample Frequency Table when Expected Margins are Known. *Annals of Mathematical Statistics*, Vol. 13, pp. 166-178.

Tsygalnitsky, S. (1977) Simplified Methods of Transportation Planning. MS thesis, Massachusetts Institute of Technology, Cambridge, MA.