Estimating Uncertainty of Bus Arrival Times and Passenger Occupancies

Vikash V. Gayah, Ph.D., Zhengyao Yu, and Jonathan S. Wood

MNTRC Project 1246

March 2016

SJSU Research Center
210 N. Fourth St., 4th Fl.
San José, CA 95112

Tel // 408.924.7560
Fax // 408.924.7565

transweb.sjsu.edu

Bus transit agencies often provide travelers with current vehicle locations and estimated vehicle travel times to downstream stops using real-time transit information systems. This information has been found to:

- Improve traveler decision-making, and
- Enhance perceived system reliability.

Predictions of bus transit system evolution can enhance real-time passenger decision-making... and improve perceived system reliability.

However, existing travel time predictions fail to provide an indication of the uncertainty associated with these estimates. Since travelers generally believe that these estimates are accurate, this can cause a false sense of precision, which can lead to the types of negative experiences that reduce ridership if and when actual bus travel times differ. Furthermore, no existing models are available to predict individual bus occupancies at downstream stops to help travelers understand if there will be space available to board when the bus arrives. To address these concerns, this research project proposes and tests statistical modeling methods to:

- Estimate both bus travel times and passenger occupancies, and
- Quantify the uncertainty associated with these estimates.

Study Methods

<u>Estimating bus travel times</u>:
Accelerated failure-time survival models are proposed to estimate bus travel times. This technique is typically used to predict the time until an event occurs. In this case, the event being considered is the arrival of a bus to a downstream stop. This modeling method is compared to traditional linear regression models, which appear to be the most commonly used method to predict bus travel times in the existing literature. Confidence intervals for travel times are estimated and compared to actual travel times to assess the different modeling approaches.

<u>Estimating passenger occupancies</u>:
Models currently do not exist in the research literature to predict individual bus passenger occupancies. Therefore, both linear regression and count regression models are considered. Estimates of bus occupancies are compared to actual occupancies to assess the two modeling approaches. Quantile regression models are also proposed to directly estimate the confidence intervals associated with these predictions. This modeling technique does not only present a more comprehensive view of each variable's effects, but also gives better information about uncertainty level associated with each observation.

Case study:

These methods were tested on a campus shuttle route in State College, PA. Data were obtained from the Centre Area Transportation Authority for a period spanning January 2013 to April 2014. These data include archived bus arrival times to each stop available from Automatic Vehicle Locations (AVL) and systems and passenger occupancies from Automated Passenger Counter (APC) systems. Additionally, weather data were obtained from the National Climatic Data Center and the Pennsylvania State Climatologist to incorporate the impact of weather and the environment on bus transit operations.

Photo by Onward State.

Findings

Estimates of bus travel times:

Survival models are found to be as accurate as models developed using traditional linear regression techniques. However, survival models are more beneficial as they provide a better fit of the data being described. In the case study example, the survival models are found to have smaller uncertainties (i.e., tighter confidence intervals) associated with predictions than linear regression models.

Estimates of passenger occupancies:

Linear regression models are found to outperform count regression models, likely due to the additive nature of the passenger boarding process. Various modeling frameworks are tested and the best frameworks are identified for predictions at nearby stops (those within eight stops downstream) and those further away. The quantile model framework is also found to accurately describe the distribution of passenger occupancies and, thus, quantify the uncertainty associated with these estimates.

Policy Recommendations

These findings suggest that bus travel times and passenger occupancies of individual vehicles can be accurately estimated using real-time data that are available to transit agencies. Furthermore, the variability in these estimates can be quantified. This information can easily be integrated into existing real-time transit information systems to enhance traveler decision-making during the trip-planning process.

About the Authors

Vikash V. Gayah, PhD, is an assistant professor of civil engineering at the Pennsylvania State University. Zhengyao Yu and Jonathan S. Wood are PhD candidates studying transportation engineering at the Pennsylvania State University.

To Learn More

For more details about the study, download the full report at ***transweb.sjsu.edu/project/1246.html***

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business. **WEBSITE** transweb.sjsu.edu/mntrc/