The importance of being early

**Pavithra Parthasarathi · Anupam Srivastava · Nikolas Geroliminis ·
David Levinson**

Abstract   This research quantifies the relationship between the cost of earliness and
lateness by empirically observing commute trips from two different sources. The first
empirical analysis uses individual level travel survey data from six metropolitan regions
while the second analysis uses traffic data from the Twin Cities freeway network. The
analysis conducted in this research provides a method to estimate the ratio of the costs of
earliness to lateness for different datasets. This can be a useful tool for traffic engineers and
planners, to assist them in the development and implementation of improved control
strategies for congested cities. The results also corroborates the hypothesis of earliness
being less expensive than lateness and show that the finding holds steady over time and
across different regions and levels.

Keywords   Earliness to lateness · Congestion pricing · Macroscopic traffic model

P. Parthasarathi (✉) · A. Srivastava
Department of Civil Engineering, University of Minnesota,
500 Pillsbury Drive SE, Minneapolis, MN 55455, USA
e-mail: pavi0006@umn.edu

A. Srivastava
e-mail: sriva056@umn.edu

N. Geroliminis
Urban Transport Systems Laboratory, École Polytechnique Fédérale de Lausanne,
Lausanne, Switzerland
e-mail: nikolas.geroliminis@epfl.ch
URL: http://luts.epfl.ch/

D. Levinson
Network, Economics, and Urban Systems Research Group, Department of Civil Engineering,
University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA
e-mail: dlevinson@umn.edu
URL: http://nexus.umn.edu

Introduction

When travel demand exceeds system capacity, travelers compete for space and time and they cannot all reach the location of their activities at the desired time. Further, travel time uncertainty arises from inability to correctly predict when other users start their trip and which route they choose. Both phenomena result in early or late arrivals. Predicting travelers' behavior attracts high scientific interest as a thorough understanding of choices and reactions to others' choices is still incomplete.

The assumption that the penalty for being early is less than that for being late was put forward by Vickrey (1963) who analyzed how commuters compare penalties in the form of schedule delay (due to peak hour congestion), against penalties in the form of reaching their destination (ahead or behind their desired time of arrival). This assumption has been tested by many researchers since then for various applications, especially in modeling congestion pricing (Arnott et al. 1990) where it is critical to understand the tradeoff between schedule delay and travel delay.

An early empirical analysis of the phenomenon that lateness was more expensive than earliness was conducted by Small (1982) using data from the San Francisco Bay Area to analyze the endogenous scheduling of individual work trips. A logit model formulates the commuters travel schedule as a function of variables such as travel time, cost of early arrival and late arrival at work with their associated penalties, and a fixed penalty for arriving late. Commuters perceive a trade-off between the travel schedules and congestion, and the trade-off is affected by many factors such as the official work start time, occupational and family status, work-hour flexibility and car occupancy. While both earliness and lateness have associated disutilities, the cost of arriving later than the official start time is much higher than arriving earlier.

Hendrickson and Plank (1984) utilize journey to work data from Pittsburgh, Pennsylvania to examine the flexibility of departure times for trips into the Pittsburgh Central Business District. A logit model of mode and departure time choice depends on in-vehicle free flow time, congested time, monetary mode cost, transit access time and minutes of early arrival and late arrival in addition to the quadratic term of the minutes of early arrival and late arrival. The late time is more onerous than early time with the penalty for late arrival to work being much higher ($2.52 for being 5 min late) than the penalty for early arrival ($0.04 for being 5 min early). The coefficients for the quadratic term of late arrival estimated from the model is negative, indicating once an individual is quite late for work, additional late time has little effect.

Noland et al. (1998) distinguish between the costs of early and late arrival in their analysis of the importance of travel time reliability in designing policies to address congestion. Data from a stated preference survey of commuters in the Los Angeles region are used to develop a simulation model combining a supply side model of congestion delay and a discrete choice econometric demand model. The results indicate a shifting of schedules, mainly to offset the increased probability of being late, due to travel time variability. The results from the empirical demand model developed also corrobortate the assumption of lateness being more onerous than earliness.

A stated preference study on travel time reliability explores the tradeoffs travelers make when choosing routes (Tilahun and Levinson 2006). Three alternate utility formulations of reliability are explored in this paper. The first measure uses the "moments" paradigm by calculating the expected lateness and earliness relative to the usual travel time. The second utility formulation of reliability uses the right range of travel time distribution (100th%–50th%) from the median along with the probability of lateness. The

third measure uses the standard deviation of travel time experienced. These three measures are then used to estimate route choice models, based on stated preference data, to understand the role that these variables play in decision making. The results indicate that travelers value reliability highly but the valuation varies with how reliability is measured.

Tilahun and Levinson (2006) and Hollander (2006) review the two distinct approaches used by researchers to understand the attitude of travelers and the tradeoffs they make in response to travel time variability. The first approach, called the *mean-variance approach*, models the the extent of travel time variability directly in the model predicting traveler response, typically a utility or cost function. The second *scheduling approach* assumes that travelers response to travel time variability can be explained indirectly through their trip scheduling considerations, namely through their earliness and lateness considerations when choosing their departure time. The mean-variance approach fails to consider individual preferences but is used widely due to the ease of implementation compared to the scheduling approach, which requires disaggregate input data.

In their internet-based stated preference survey of bus users in the City of York, England, Hollander (2006) utilizes the scheduling approach to understand response to travel time variability and also compared it to the mean-variance approach for the same data. The results indicate that travel time variability can be better explained by the scheduling approach and the use of mean-variance approach underestimates the importance of travel time variability. The results also indicate that the penalty for late arrival is much higher than early arrival and the penalty placed on early arrival is similar to the penalty on travel time itself.

The expected utility theory approach typically used in departure time modeling has been criticized for its lack of behavioral realism and its equal valuation of gains and losses. Prospect theory has been suggested as an alternative due to its ability to value gains and losses differently. A recent study applied the reference point hypothesis of prospect theory to understand commuter departure time decision using morning commute data from Taichung City, Taiwan (Jou et al. 2008). Two reference points, the earliest acceptable arrival time and the work starting time for a given commuter, are defined in the study. Based on these reference points, a commuter experiences a gain when his arrival time falls between his earliest acceptable arrival time and work starting time and a loss when a the arrival time falls beyond this acceptable range. The empirical analysis indicates that most commuters like their arrival times to be close to their work starting times and commuters react asymmetrically to gains and loses.

This research develops new methods to estimate the ratio of the costs of earliness to lateness for different types of datasets, which enable agencies to implement control policies like congestion pricing or metering more accurately. It is shown that the arithmetic value of this ratio is population- and network-specific, but consistent among different dates for the described cities. These results are supported by the detailed analysis of different types of data sets for eight cities with different network topologies and sociological characteristics. Travel survey data from metropolitan areas provide individual travel patterns while loop detector data provide link level traffic flow data. The remainder of the paper is organized as follows: the next section describes the theoretical model used in this paper. The methodologies used to estimate the costs of earliness over lateness using travel survey data and loop detector data are described in the following two sections. The final section of the paper discusses the findings and implications of the study.

Theoretical model and initial observations

Vickrey ([1969]) integrates the earliness-lateness concept in the first structural model of peak period congestion. This seminal paper treats the demand and departure time decision endogenously. To illustrate, consider a model of city traffic where congestion occurs during the rush hour at only a single bottleneck, the entrance to the CBD. This bottleneck has a fixed capacity and when the rate of drivers arriving at the bottleneck is higher than the capacity of the bottleneck, queues develop. A traveler usually experiences a delay cost of waiting in the queue and an earliness/lateness penalty, called *schedule delay*, which is the difference between the actual time passing the bottleneck and the desired time; accordingly, he may adjust his departure time to avoid highly congested periods. Equilibrium obtains when no individual has an incentive to alter his departure time.

The analysis presented in this paper builds on the work conducted by Geroliminis and Levinson ([2009]) to investigate the recurring congestion problem (such as the morning commute) at a network level in an urban area. The analysis combines Vickrey ([1969])'s theory with a recent macroscopic traffic model developed and tested by Daganzo ([2007]) and Geroliminis and Daganzo ([2007]); Geroliminis and Daganzo ([2008]) to analytically estimate an equilibrium solution and to develop a dynamic model of cordon-based congestion pricing. In the above models, equilibrium is defined as a stable choice of departures times, such that no traveler has an incentive to deviate from his choice and cannot find a departure time giving less total trip cost than the one experienced.

The main difference between the network problem and Vickrey's single bottleneck is that capacity (maximum flow) is not constant but a function of average density during congested conditions. For a single bottleneck, queues form when demand exceeds capacity and the system operates at capacity until queues clear. In contrast, at the network level, as more vehicles enter the system the average network flow and the total outflow of the system decrease (as vehicles interfere with each other) and, as a result, the effect of congestion is more significant. The consequence of this property for urban networks is that equilibrium at the network level has distinct characteristics compared with single bottleneck model.

The properties of the network model and the macroscopic traffic model that will be utilized in this paper are briefly described here. For a detailed description of the analytical solution and the assumptions of the models see Geroliminis and Daganzo ([2008]) and Geroliminis and Levinson ([2009]). Geroliminis and Levinson ([2009]) demonstrates that if all travelers have the same values for schedule delay (earliness or lateness) rates and travel delay rates ($ per unit time), the equilibrium travel delay will have a triangular shape with time, increasing with a slope equal to the ratio of earliness over travel delay rate and then decreasing with slope equal to the ratio of lateness over travel delay rate. The first and last individuals (those who depart in the beginning and at the end of the congested period) experience only schedule delay, while the traveler that experiences the maximum travel delay arrives on time (schedule delay is zero). If this elegant analytical derivation holds for real networks, one could easily estimate the ratio of the costs of earliness to lateness for a network if the average (over space) travel delay in a network with time is known, by approximating this delay with a triangular shape during the onset-offset of different peak periods.

The authors presented real data from the morning peak for one day in downtown Yokohama and verified that travel delay has a triangular shape with time and the ratio of the costs of earliness to lateness is about 0.5. We investigate whether the triangular travel delay shape is a data coincidence and if the ratio of the costs of earliness to lateness varies

across different days or has a network-specific value. This is one of the tasks of the current paper.

The part of downtown Yokohama examined in this paper is approximately a 10 km^2 triangle and is congested during the weekdays peaks. Available data included flow and detector occupancy (defined as the fraction of time a detector is occupied by vehicles) measurements every 5 min from 500 fixed sensors and 140 taxis equipped with GPS reporting their position and other data with suitable time stamps (activations and deactivations of parking brake, blinkers/ hazard lights, the beginning and end of all stops lasting more than a few seconds etc.). More details are provided in Sarvi et al. (2003). We analyzed data from many different congested days, both for the morning and evening peak.

In Yokohama, the ratio of the costs of earliness to lateness for the morning peak varies almost uniformly between 0.48 and 0.52 across different days, while for the evening peak the ratio is between 0.36 and 0.43. Figure 1 shows the results for the analysis for one day. This graph shows the number of vehicles in the system with time. Geroliminis and Levinson (2009) find that travel time and cumulative number of vehicles for congested conditions in Yokohama are linearly related with an adequate fit ($R^2 = 0.8$), which means that the ratio of the costs of earliness to lateness obtained from the graph in Fig. 1, would be similar if travel time were used instead. We note that during the evening peak the triangular shape of travel delay is not as good a fit as in the morning peak. This is expected as evening peak includes many non-work trips with different rates for schedule and travel delay, which violate some of the assumptions of the recurring congestion network equilibrium model. According to the US National Household Travel Survey, Daily Trip File, commute trips comprise fewer than 15% of all daily trips in the US, and while a higher share of peak trips, do not constitute a majority (Levinson and Krizek 2008), similar trends are likely seen in Yokohama.
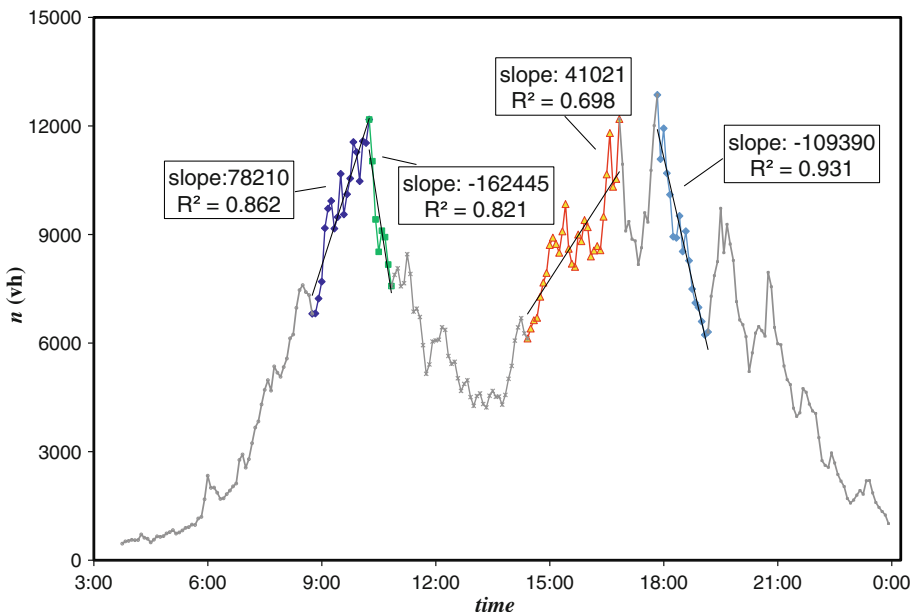

Fig. 1 Estimated earliness and lateness ratios—Yokohama data

The aforementioned papers co-authored by Daganzo and Geroliminis show that traffic in large urban regions can be modeled dynamically at an aggregate level, if these regions can be partitioned in a small number of sub-regions (neighborhoods), which are uniformly congested. These papers show, using a micro-simulation of the San Francisco Business district and a field experiment in downtown Yokohama (Japan), (i) that urban neighborhoods approximately exhibit a Macroscopic Fundamental Diagram (MFD) relating the number of vehicles (accumulation) in the neighborhood to the neighborhood's average speed (or flow) and (ii) there is a robust linear relation between the neighborhood's average flow and its total outflow (the number of vehicles that reach their destinations per unit time) during a day and across days. This means that the traffic model used to describe travel behavior and derive equilibrium conditions for the recurring congestion network problem provides a realistic representation of macroscopic traffic conditions in the area (Geroliminis and Levinson 2009).

Research approach

Travel survey data

Travel surveys are conducted by planning agencies to understand the travel patterns of the residents in the region and to provide a basis to guide highway and transit investment decisions. Surveyed residents are typically asked to maintain a travel diary for at least a day during which they provide information on the trips undertaken with regards to the trip origin and destination, the purpose of the trip, the arrival and departure time of each trip, the travel mode used, etc. In addition socio-demographic information regarding the individual such as age, gender, employment status, occupation type; and household information such as household size, household income, vehicle ownership and type, residence type are also collected (Metropolitan Council of the Twin Cities Area 2003). The travel surveys are typically conducted once every 5–10 years. For example, the Twin Cities travel survey, called the Travel Behavior Inventory (TBI), is conducted about once every 10 years, starting 1949 and has been conducted in 1958, 1970, 1982, 1990 and 2000.

The first empirical analysis conducted in this paper tests the difference in the costs of earliness and lateness, using travel surveys from six metropolitan regions (totaling 11 surveys), available at the Metropolitan Travel Survey Archive (University of Minnesota 2003; Levinson and Zofka 2006). Household travel surveys conducted between 1988 and 2005 are considered for this analysis. The general criteria in the selection of these metropolitan regions was data availability and the ease of separating out the individual commuter data (home-based work) from the overall travel survey data. It is hoped that the wide range in the year of surveys combined with the broad geographical distribution of the survey areas considered, will provide universal validity to our methodology of estimating the ratio of the cost of earliness to lateness.

Methodology

A preliminary analysis of the travel speeds that were estimated from the year 2000 Twin Cities Travel Behavior Inventory (TBI) was used to establish the onset and offset of congestion during the morning and evening periods. The home-based work (HBW) trips for full-time employees, consisting of home to work and work to home trips, were separated out from the overall travel survey data. The survey data provides information on the

commute trip of an individual via data on trip origin and destination (in X & Y coordinates), trip start time, trip end time and trip duration. Only 4,830 HBW records had complete trip and location information. The data in each complete record was augmented by estimates of distance, vehicle kilometers traveled (VKT) and vehicle hours traveled (VHT). The distance between an origin and a destination was estimated from the reported X & Y coordinates. The VKT and VHT were calculated using the estimated distance and reported trip duration.

The procedure to determine the speed profile is as follows: Each HBW record was assigned to a diurnal 30-min time period based on the reported trip departure times. Speed profiles were determined for only those 30 min periods that had a minimum of 30 records. The VKT and VHT of each record that belonged to a common time period were summed and in turn used to calculate the average speed. The use of a minimum number of survey records (≥ 30) is a general criteria to avoid outliers in estimated average speeds, especially in those time periods of extremely low system usage. A profile of the average speeds estimated from the 2000 TBI data is provided in Fig. 2.

A similar analysis of the travel data for full-time HBW trips was conducted for the 2005 Tahoe Regional Planning Agency household travel survey. The unavailability of reported distance between the trip origin and destination, lack of locational data (X & Y coordinates of the trip origin/destination or corresponding TAZ), errors in reported trip distance made it difficult to conduct similar speed profiles for the remaining four metropolitan areas considered in the analysis. The onset and offset of congestion in the morning and evening periods for the remaining metropolitan areas was established using the congested period estimated from the TBI as a starting point, and adjusting the time using data on the number of travelers on the system in each metropolitan area.
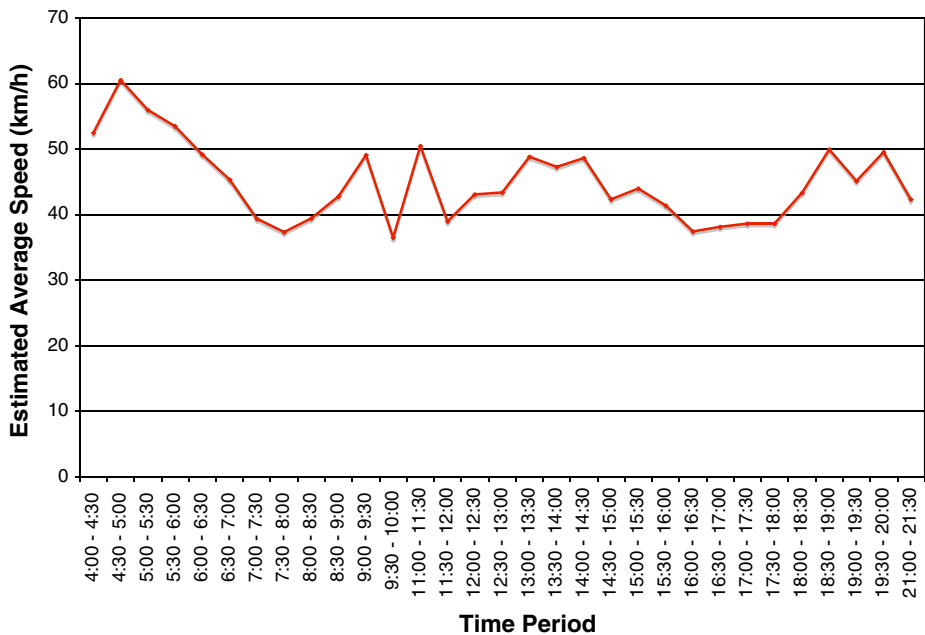

Fig. 2 Estimated speed profile—Twin Cities TBI HBW data

The number of travelers on the system, in each of the six metropolitan area, was estimated from the respective surveys using the reported trip departure and arrival times. This data, stratified into 5-min time periods, was then used to estimate separate regression models for the onset and offset of congestion during the morning and evening periods.

The regression models estimate the number of commuters on the system as a function of time. It is important to note that the number of travelers for the morning peak period were estimated from the *home to work* survey data and the number of travelers on the system during the evening peak period were estimated from the *work to home* survey data for each metropolitan area, except for a few surveys, highlighted in the results, as shown in Tables 1 and 2. Based on the model developed by Geroliminis and Levinson (2009), a ratio is then estimated as:

$$\text{Ratio} = \text{Cost of Earliness/Cost of Lateness} \tag{1}$$

The cost of earliness is the slope coefficient from the regression model for the onset of congestion. The cost of lateness is the slope coefficient from the regression model for the offset of congestion

A plot of the actual commuters from the travel survey data and the predicted commuters estimated using the regression models for the various metropolitan areas is provided in Figs. 3, 4, 5, 6, and 7.

The ratios estimated from the various surveys vary between 0.39–0.76 for the morning peak period and 0.34–0.84 for the evening peak period. Six of the 11 surveys in the morning peak period and four of the 11 surveys in the evening peak period show a value between 0.5–0.65.

The ratios of the costs of earliness to lateness, estimated using the Yokohama data, presented in the 'Theoretical model and initial observations' section of this paper, varies between 0.48 and 0.52 across different days, while for the evening peak the ratio is between 0.36 and 0.43. The ratios of the costs of earliness to lateness estimated from the travel surveys have a broader range compared to the ratios estimated using the Yokohama data. This could be due to the wide geographical and time distribution of the travel surveys considered in this analysis.

In general, the results show that earliness has a different and lower cost than lateness across the various metropolitan areas, especially for the morning peak period. The results for the morning peak period also indicate consistency in the ratios of the costs of earliness to lateness. For example, the analysis conducted using the Twin Cities TBI shows that the ratios estimated for the morning peak period holds steady between 0.55 and 0.58 from 1982 to 2000.

Freeway loop detector data

The second dataset is traffic data obtained from the freeway loop detectors in the Twin Cities metropolitan area. Mn/DOT's Regional Transportation Management Center (RTMC) maintains an exhaustive system of over 4,000 inductive loop detectors placed approximately every kilometer on the mainline freeways and on every entrance and exit ramps. The volume and occupancy data collected from the loop detectors is transferred to the RTMC every 30 s using fiber optic lines (Regional Transportation Management Center 2008). Volume is the number of vehicles crossing the detector in the specified interval, while occupancy is the fraction of time a detector is occupied by vehicles.

Table 1 Empirical comparison of earliness and lateness rates—morning peak period

Dependent variable: number of travelers in the system; independent variable: time in 5-min intervals

Agency	Year	Onset (am)	Earliness	t	R^2	Obs.	Offset (am)	Lateness	t	R^2	Obs.	Ratio
Twin Cities Metropolitan Council, MN	1982[a]	5:00–7:30	1.80	13.48	0.86	30	7:30–9:00	−3.20	−7.97	0.79	18	0.56
	1990	5:00–7:30	6.55	16.14	0.90	30	7:30–9:00	−11.31	−10.92	0.87	18	0.58
	2000	5:00–7:30	3.76	13.32	0.86	30	7:30–9:00	−6.86	−10.59	0.87	18	0.55
Puget Sound Regional Council (PSRC), Seattle, WA	1989	5:00–7:00	1.48	13.63	0.89	24	7:30–9:00	−2.32	−11.75	0.89	18	0.64
	1994	5:00–7:00	1.08	10.56	0.83	24	7:30–9:00	−1.59	−9.37	0.84	18	0.68
	2000[a]	5:00–7:00	0.73	9.14	0.78	24	7:30–9:00	−1.47	−12.65	0.90	18	0.50
Metropolitan Washington Council of Governments (MWCOG), D.C.	1988	5:30–8:00	7.12	16.03	0.90	30	8:00–10:00	−9.39	−18.35	0.94	24	0.76
	1994[a]	5:30–8:00	3.17	10.42	0.79	30	8:00–10:00	−5.37	−19.31	0.94	24	0.59
Kentuckiana Regional Planning and Development Agency (KIPDA)	2000	5:00–7:30	1.39	8.14	0.69	30	7:30–9:00	−3.53	−8.56	0.81	18	0.39
Community Planning Association of Southwest Idaho (COMPASS), ID	2002	5:00–7:30	1.00	10.20	0.78	30	7:30–9:00	−2.06	−8.04	0.79	18	0.49
Tahoe Regional Planning Agency, CA	2005	6:30–8:00	1.89	17.11	0.94	18	8:00–8:30	−4.90	−18.72	0.98	7	0.39

[a] Number of travelers data not separated into home to work and work to home trips for use in regression models due to data limitations

Coefficients are significant at 99% confidence level in all regression models

Table 2 Empirical comparison of earliness and lateness rates—evening peak period

Dependent variable: number of travelers in the system; independent variable: time in 5-min intervals

Agency	Year	Onset (pm)	Earliness	t	R^2	Obs.	Offset (pm)	Lateness	t	R^2	Obs.	Ratio
Twin Cities Metropolitan Council, MN	1982[a]	2:00–5:00	0.94	8.76	0.68	36	5:00–7:00	−1.85	−11.43	0.85	24	0.51
	1990	2:00–5:00	3.77	12.37	0.81	36	5:00–7:00	−6.79	−20.61	0.95	24	0.56
	2000	2:00–5:00	2.59	13.78	0.84	36	5:00–7:00	−4.92	−21.67	0.95	24	0.53
Puget Sound Regional Council (PSRC), Seattle, WA	1989	3:00–5:30	0.98	9.93	0.77	30	5:30–7:30	−1.31	−17.43	0.93	24	0.75
	1994	3:00–5:30	0.81	8.51	0.71	30	5:30–7:30	−1.23	−12.60	0.87	24	0.66
	2000[a]	3:00–5:30	0.36	6.68	0.60	30	5:30–7:30	−0.57	−11.69	0.86	24	0.63
Metropolitan Washington Council of Governments (MWCOG), D.C.	1988	3:30–5:30	5.99	9.81	0.81	24	5:30–7:30	−7.52	−22.44	0.96	24	0.80
	1994[a]	3:30–5:30	3.74	8.14	0.74	24	5:30–7:30	−4.54	−21.84	0.95	24	0.82
Kentuckiana Regional Planning and Development Agency (KIPDA)	2000	2:00–5:00	1.11	8.07	0.65	36	5:00–7:00	−3.12	−12.56	0.87	24	0.36
Community Planning Association of Southwest Idaho (COMPASS), ID	2002	2:00–5:00	0.50	10.30	0.75	36	5:00–7:00	−1.46	−15.66	0.91	24	0.34
Tahoe Regional Planning Agency, CA	2005	4:40–5:30	3.39	13.51	0.96	9	5:30–6:00	−4.05	−21.66	0.99	6	0.84

[a] Number of travelers data not separated into home to work and work to home trips for use in regression models due to data limitations

Coefficients are significant at 99% confidence level in all regression models

Twin Cities Household Travel Survey (2000)
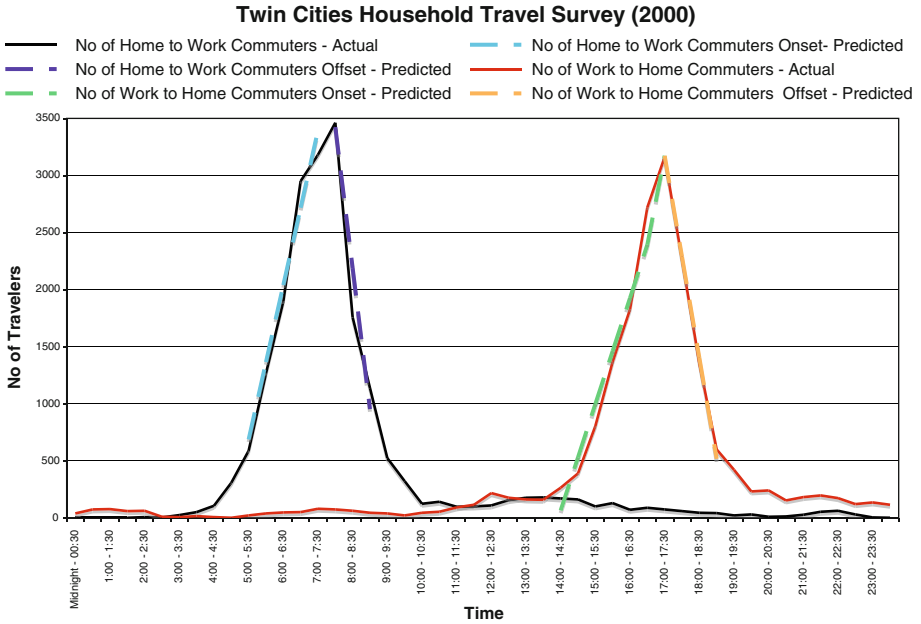
Fig. 3 Actual versus estimated commuters—Twin Cities
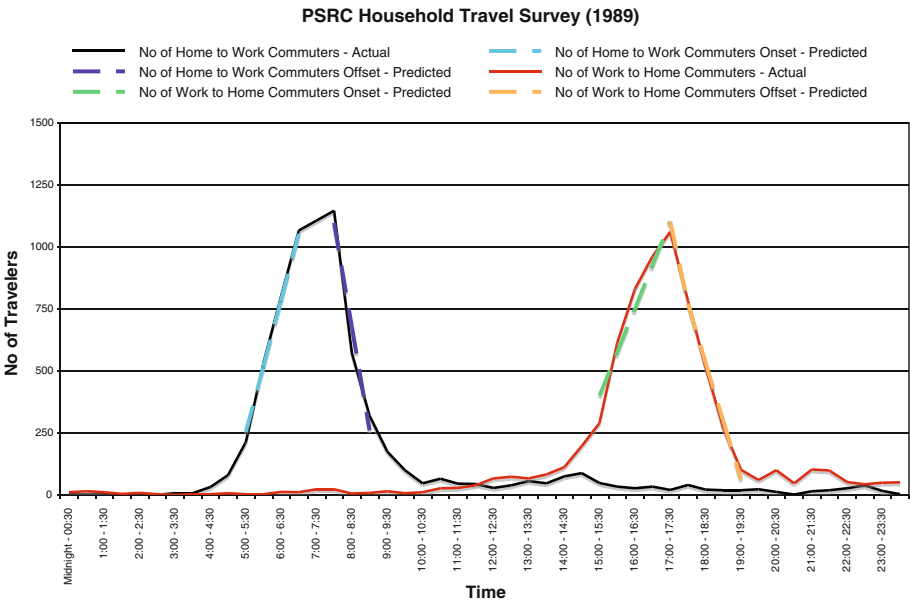
PSRC Household Travel Survey (1989)

Fig. 4 Actual versus estimated commuters—Puget Sound

If occupancy is multiplied by the effective vehicle length, it shows the average density (vehicles/km) in the proximity of the detector. An approximation of the number of vehicles in the network can be obtained if the number of lanes for each detector station
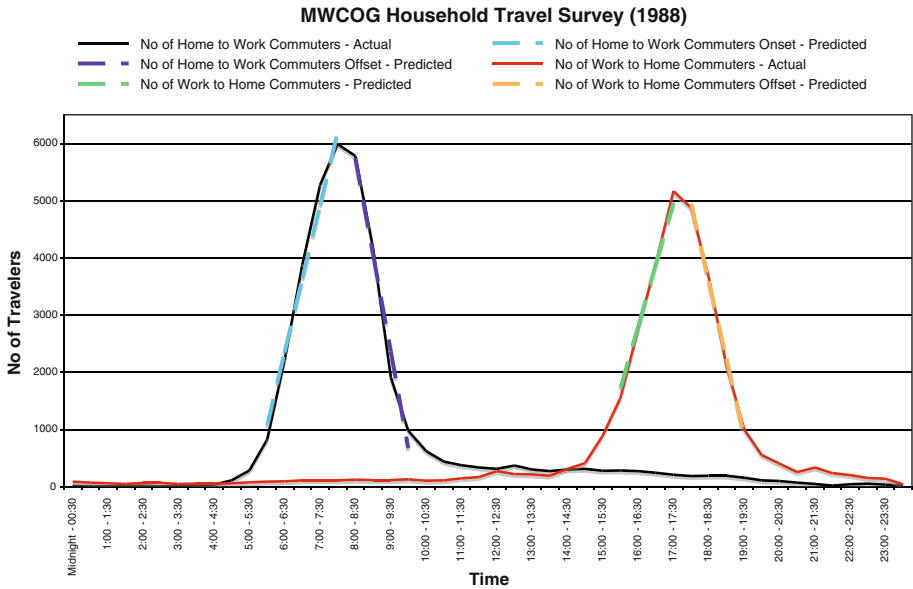
MWCOG Household Travel Survey (1988)

Fig. 5 Actual versus estimated commuters—metropolitan Washington

COMPASS Household Travel Survey (2002)
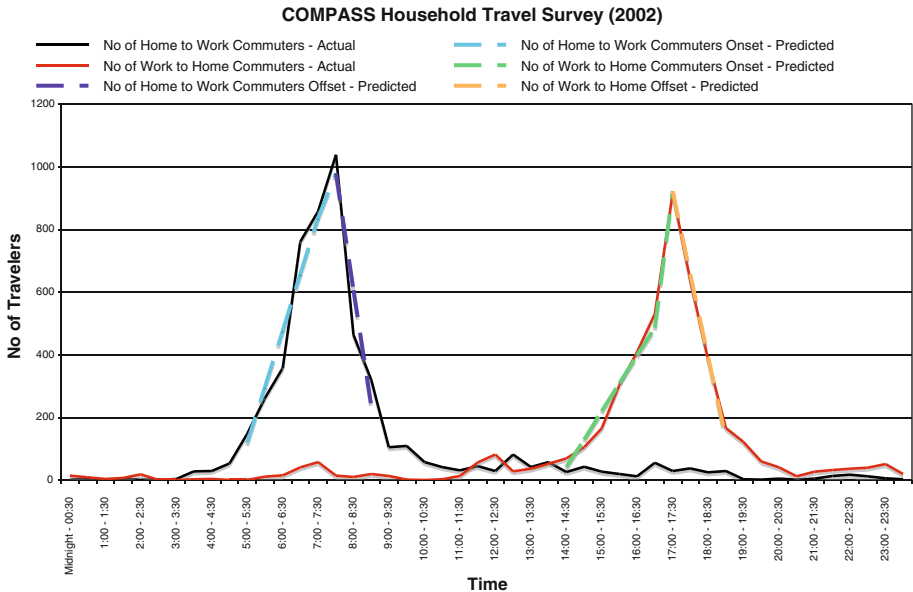
Fig. 6 Actual versus estimated commuters—Southwest Idaho

and the distance between successive detector stations are known. For the Twin Cities freeway network, detectors are almost evenly spaced, which makes the estimation of the number of vehicles less tedious.
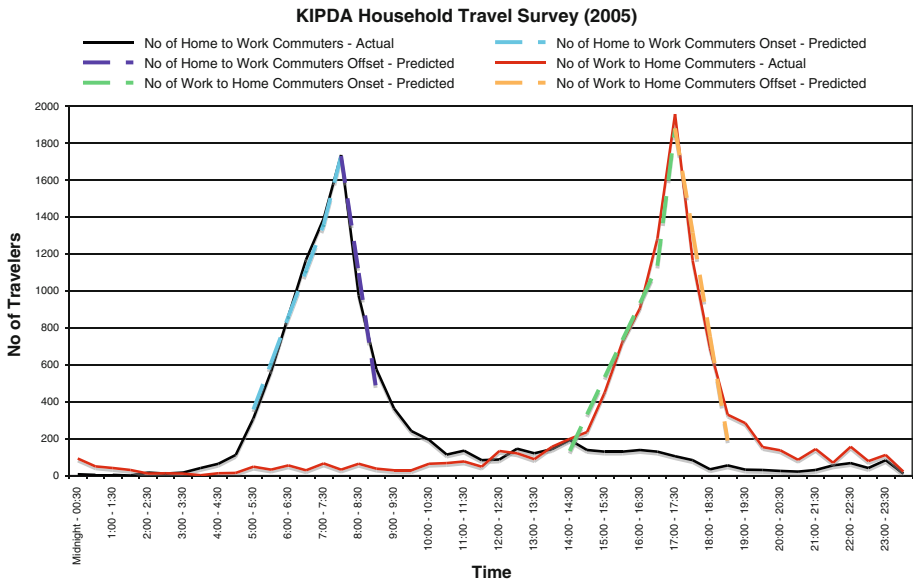
KIPDA Household Travel Survey (2005)

Fig. 7 Actual versus estimated commuters—North Central Kentucky and Southern Indiana

The morning peak period freeway loop detector data is extracted for a selected day in May, 2007 and utilized for analysis in this research[1]. The date (Wednesday, May 23rd, 2007) selection was based on the relative availability of loop detector data over the network. An additional consideration was ensuring that the changes in traffic patterns after the I-35W Bridge collapse did not affect the results of the study. The flows over a sample set of mainline detectors from various sections of the freeway network compare well with other dates (also considered as valid candidates), thereby ensuring consistency in behavior.

The freeway loop detector data was analyzed for presence of rogue values (missing detectors, stuck detectors, malfunctions, and outliers) and cleaned based on the principles of flow conservation. It is important to note that data correction for the selected date of analysis was minimal and fewer than 5% of the loop detectors had to be corrected.

Methodology

The first step in the analysis involves the use of the loop detector data to identify the inputs and outputs to the freeway system. Figure 8 identifies the loop detectors along the freeway network and also defines the boundary of the freeway system (just outside the I-494/I-694 beltway) analyzed in this paper.

The Mn/DOT loop detector data identifies each detector as either an on-ramp, off-ramp or a mainline detector. The on-ramps act as inputs and the off-ramps act as outputs to the freeway system. The input data (vehicle count data) from all on-ramps in the system is aggregated along with the mainline inputs into the freeway system at the defined boundary. This provides the aggregated system input, which is then accumulated to obtain the

---

[1] The Minnesota Traffic Observatory (MTO) maintains a repository of actual freeway traffic data extracted from the loop detectors on the Twin Cities freeway network, available for download at http://data.dot.state.mn.us/datatools/.
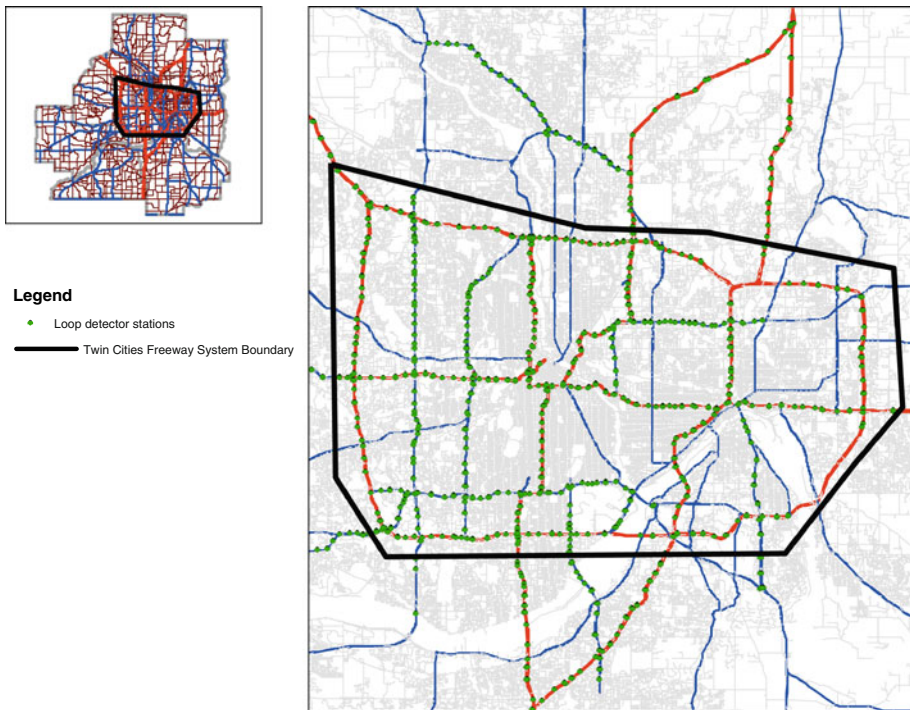
Fig. 8 Twin Cities freeway network—analyzed area

cumulative system input against time. The *cumulative system output* is similarly obtained using the aggregated outputs on all off-ramps inside the defined region, and all outputs occurring at the boundaries of the network.

Instead of plotting cumulative number of vehicles entering/exiting the system versus time t, N(t), an oblique coordinate system was used to plot $N(t) - q_0(t - t_0)$ versus t for the starting time, t_0, and some choice of q_0, as shown in Fig. 9. $N(t)$ is the cumulative number of vehicles up to time t, while q_0 is a constant value with units of flow to highlight the changes in slope. As shown in Fig. 9, the cumulative input starts and ends at 0, which means that q_0 approximately equals the average value of input flow during the entire study period. This transformation magnifies the vertical scale to more clearly observe changes in the slope. This method of plotting the curves was introduced by Cassidy and Windover (1995). Oblique plots are commonly used in analyzing freeway traffic data and better visualize the trends and changes in vehicular traffic, which can usually be masked in a cumulative plot (Muñoz et al. 2000).

The changes in the system in terms of accumulation and dissipation of the vehicles during the onset and offset of congestion are shown in Fig. 10. The '+' sign in the figure indicates an accumulation of vehicles in the network (total input greater than total output), and the '−' indicates a dissipation in the system (total output greater than total input). The cumulative system input and output plot can then be used to estimate the number of vehicles on the freeway network over time (as the difference of the cumulative input and cumulative output curves).

The next step in the analysis is to estimate separate regression models for the onset and offset of congestion. As mentioned previously, the focus of the analysis is the morning
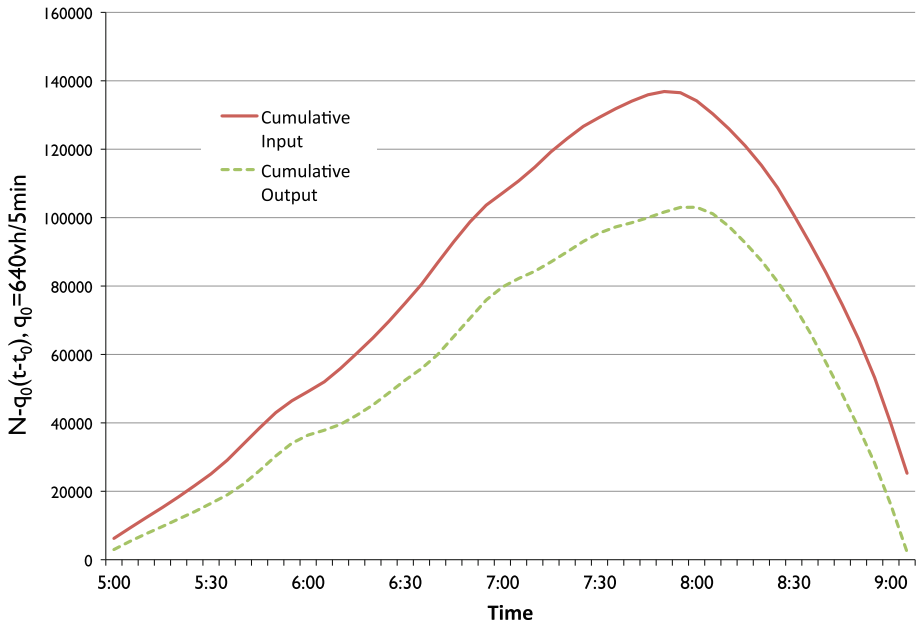
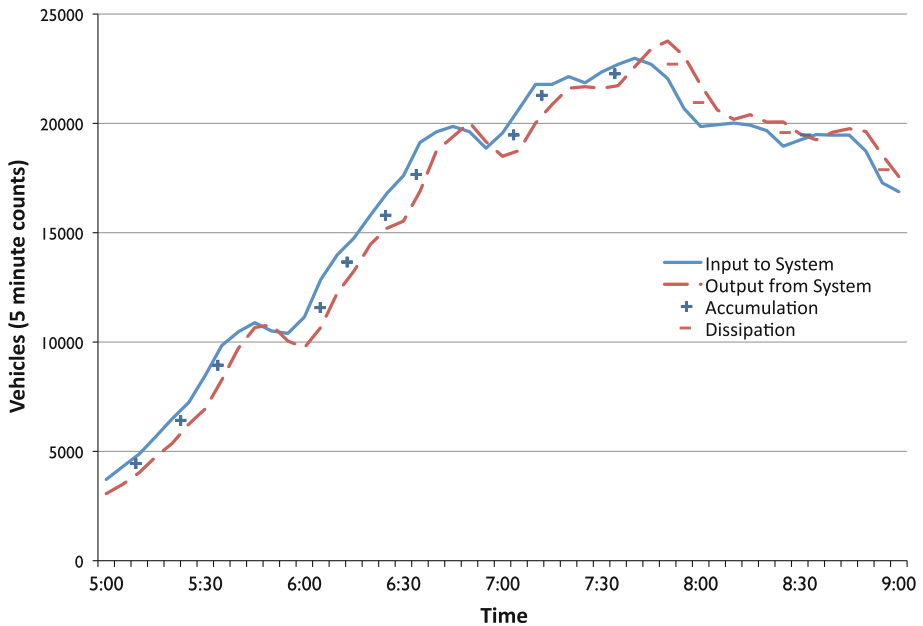Fig. 9 Oblique plot of cumulative system input and output—Twin Cities freeway system

Fig. 10 System accumulation and dissipation—Twin Cities freeway system

peak period traffic on the Twin Cities freeway system and the underlying assumption is that the majority of morning peak period traffic is made up of home to work or commute trips. An analysis of speed profiles is conducted to identify the onset and offset of
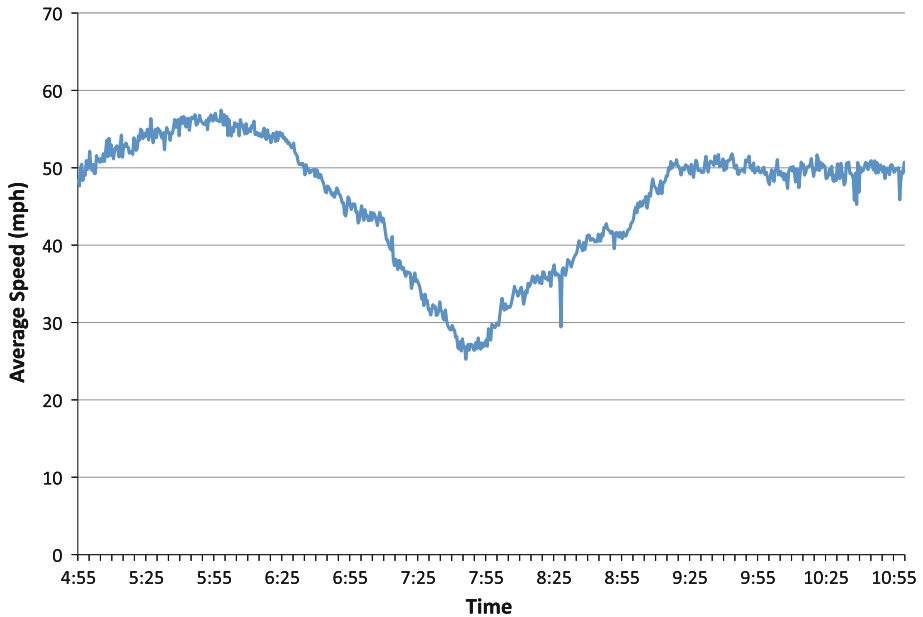
Fig. 11 Estimated average speed over time—Twin Cities freeway system

congestion in the Twin Cities, similar to the speed profiles constructed using the TBI data. A plot of the estimated average speed against time is shown in Fig. 11.

The average speeds estimated using the traffic data from the loop detectors are higher than the speeds estimated from the reported travel survey data. Possible explanations are (Wu et al. 2009; Levinson et al. 2006):

(i)   The differences in drivers' perceived and actual travel time
(ii) The fact that loop detectors on freeways have travel speeds faster than speeds on arterial routes
(iii) Respondents to travel surveys may report door-to-door time, not engine-on to engine-off time

However it is important to note that the comparison of the speed profiles from the TBI data and loop detector data show the same trend of a decrease in average speed of the system users at the onset of congestion and an increase in the average speed at the offset of congestion.

The third step in the analysis is the estimation and use of the Macroscopic Fundamental Diagram (MFD) relating the flow in the system to the average occupancy on all mainline detectors along the freeway network. A detailed explanation of the existence of MFD for traffic in an urban area and the rationale of using the MFD as a tool to describe congestion dynamically is provided in Geroliminis and Daganzo (2008). A plot of the MFD created using the loop detector data for a weekday (3 am–3 pm) is provided in Fig. 12.

The MFD plot for the Twin Cities metropolitan area freeway system shows a well defined relation between average flow and occupancy. When compared with MFDs for arterial networks (e.g. Yokohama) there is more scatter in the congested regime. This phenomenon is due to the spatial heterogeneity in congestion across the freeway network. The MFD for the most congested part of the network (inner ring around downtown Minneapolis) has different data scattering and shape (e.g. during the peak hour the
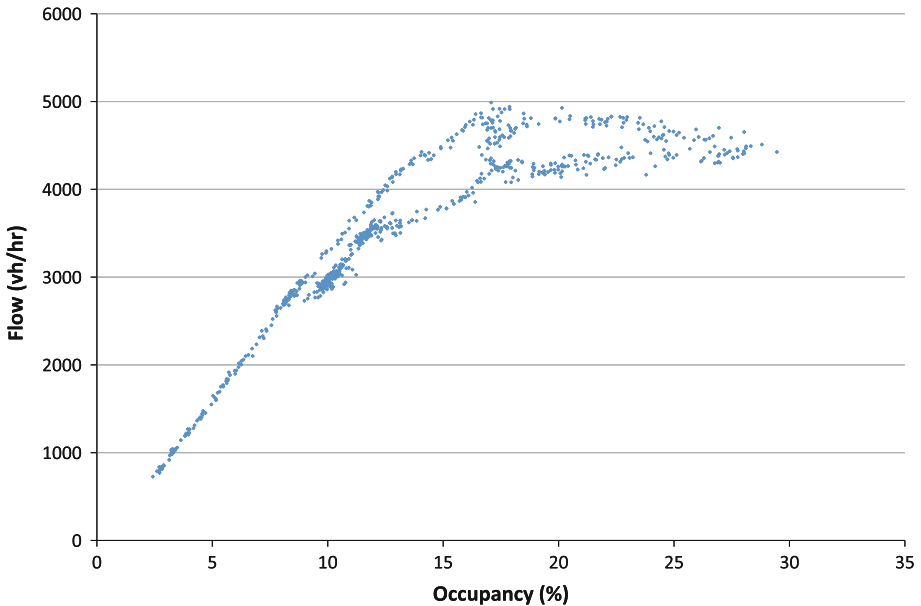
Fig. 12 Macroscopic fundamental diagram—Twin Cities freeway system

maximum value average occupancy in the congested part is 28%, while in the whole network is 16%). But as we show here (Table 3), the difference in the shape and scatter is not significant enough to affect the ratio of the costs of earliness to lateness. We also show that this ratio is the same when estimated for the whole network or only for the inner ring.

Three types of regression model are estimated based on the analysis conducted above. The first two regression models use occupancy data for the congested part and the whole network while the last regression model uses the number of vehicles on the system. The regression based on occupancy is easier to analyze since the required data can be obtained directly from the loop detectors. The regression based on the number of vehicles in the system involves slightly more effort since the data isn't directly collected and needs to be estimated from the loop detector data for on and off-ramps.

The first regression model uses the data on average occupancy over time from all mainline detectors in the Twin Cities freeway network. As in the analysis with travel survey data, separate regression models were estimated for the onset and offset of congestion. A plot of the estimated average occupancy against the actual average occupancy obtained from loop detector data is provided in Fig. 13. The second regression model is also based on the average occupancy data but focuses on the congested portion of the Twin Cities freeway system. The model is estimated the same way as the first regression model and a plot of the results are provided in Fig. 14. Note that the average occupancy for the entire network varies between 2 and 16%, while for the most congested part of the network between 2 and 28%. This happens because most of the low occupancy during the rush hours are located in the outer ring of the study region. Separate regression models are estimated for the entire network and the congested freeway network to account for the presence of spatial heterogeneity identified in the MFD. Such an analysis allows us to see if the relationship between the costs of earliness and lateness as perceived by the system
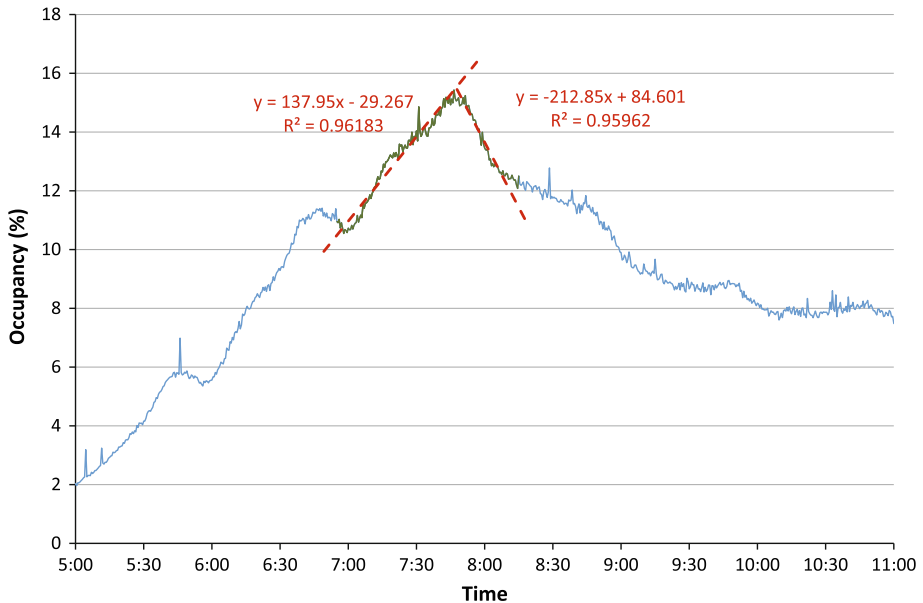
Fig. 13 Regression results—using occupancy on entire system (*y* variable in the vertical axis, *x* variable in the x axis)

users holds under different traffic environments. Building on the analytical model of Geroliminis and Levinson (2009) and the empirical evidence from the Yokohama experiment (see "Research approach" section), we approximate the ratio of the costs of earliness to lateness, with the ratio of the slopes for the onset and offset of congestion from occupancy and number of vehicles time series.

The final regression model in this analysis uses the number of vehicles on the freeway system (obtained indirectly from loop detector data) over time, similar to the models estimated using travel survey data. As in the other models, separate regression models are estimated for the onset and offset of congestion and the plot of the predicted number of vehicles against the actual input data is shown in Fig. 15.

It is important to note that the onset and offset of congestion used in the above three regression analyses is determined by analyzing the speed profiles in combination with the profiles of actual occupancy and number of vehicles on the system respectively. The plot of the results from the above three regression models, shown in Figs. 13, 14, and 15, indicate a good fit for the predicted values against their respective inputs. The ratio of the costs of earliness to lateness estimated from all three regression models using the slopes for onset and offset is presented in Table 3.

The results, obtained using a new methodology, indicate values for the ratio between 0.55–0.65. The ratio estimated from occupancy data are slightly higher than the ratio estimated using the number of vehicles in the system. It is important to note that the ratio of the costs of earliness to lateness is the same in both regressions conducted using the occupancy data, indicating that spatial heterogeneity in congestion doesn't affect how travelers perceive earliness and lateness. Further, the ratio (0.56) estimated using the number of vehicles on the Twin Cities freeway system is very close to the value (0.55) obtained from the estimation of number of travelers on the system using the year 2000 TBI data.
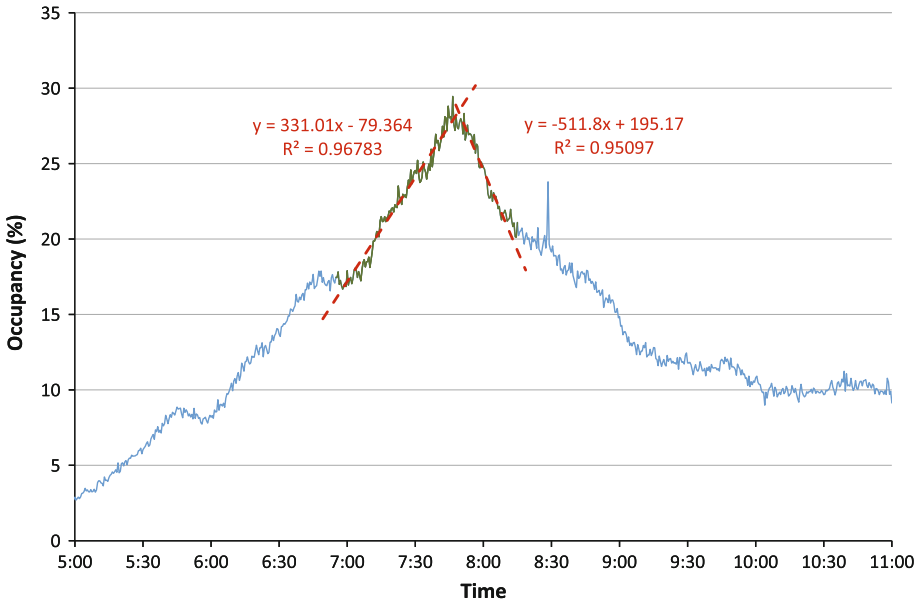
Fig. 14 Regression results—using occupancy on congested portion alone (*y* variable in the vertical axis, *x* variable in the x axis)
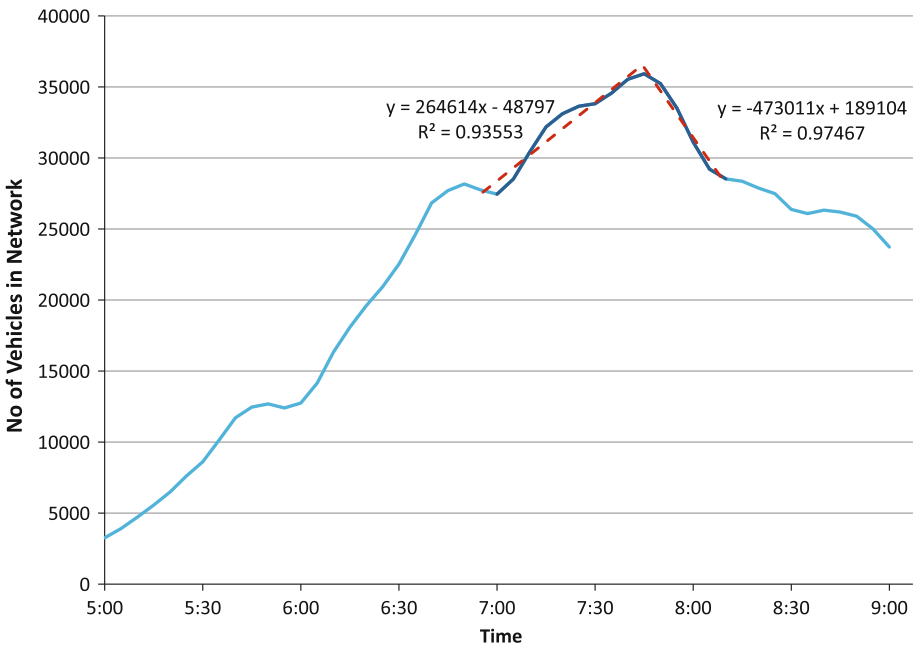

Fig. 15 Regression results—using number of vehicles on entire system (*y* variable in the vertical axis, *x* variable in the x axis)

Table 3 Empirical comparison of earliness and lateness rates—Twin Cities

Data source	System analyzed	Onset (am)	t	R^2	Obs.	Offset (pm)	t	R^2	Obs.	Ratio
Occupancy	Entire Freeway Network	6:55–7:45	49.94	0.96	101	7:45–8:10	−30.44	0.96	41	0.65
Occupancy	Congested Freeway Network	6:55–7:45	54.57	0.97	101	7:45–8:10	−27.50	0.95	41	0.65
Number of vehicles in network	Entire Freeway Network	7:00–7:45	10.77	0.94	10	7:45–8:10	−12.41	0.97	6	0.56

Coefficients are significant at 95% confidence level in all regression models

Conclusions

This research analyzes the importance that commuters place on the costs of earliness and lateness using data at the individual level and aggregate level. The analysis was conducted using existing datasets across different sites (travel survey and traffic data for eight sites). The results indicate that late arrival is more expensive than early arrival, and holds steady over time and across different regions and levels. More importantly, we developed methods to estimate the ratio of the costs of earliness to lateness for different types of datasets. This can be a useful tool for traffic engineers and planners to assist them in the development and implementation of improved control strategies (such as cordon pricing or metering) for congested cities and to appropriately value travel time reliability.

The results presented in this paper are based on a model developed by Geroliminis and Levinson (2009) which focuses on the average behavior of travellers. This simplified model allows us to extract average behavioral characteristics from aggregated traffic data. While it is known that the value of time (and the schedule delay consequently) varies among individuals, our understanding is not complete on how the variability of this and other socioeconomic factors affect the equilibrium solution. Research in this direction is underway.

Results from this survey allow us to assert that the ratio is on average between 0.55 and 0.65 in the Twin Cities (and has other values in other cities). While the use of individual travel survey data overcomes the limitations of just using aggregate data, this study does not focus on the causal factors that explain why earliness is preferred to lateness, or its magnitude. Employing travel survey data on significant subsets of the population should enable examination of whether preference for earliness or lateness varies by spatial, socioeconomic, demographic, employment, or other variables.

References

Arnott, R.J., De Palma, A., Lindsey, R.: Departure time and route choice for the morning commute. Transp. Res. B Methodol. **24B**, 209–228 (1990)

Cassidy, M., Windover, J.: Methodology for assessing dynamics of freeway traffic flow. Transp. Res. Rec. Natl. Res. Counc. **1484**, 73–79 (1995)

Daganzo, C.: Urban gridlock: macroscopic modeling and mitigation approaches. Transp. Res. B **41**, 49–62 (2007)

Geroliminis, N., Daganzo, C.: Macroscopic modeling of traffic in cities. In: 86th Annual Meeting of the Transportation Research Board, pp. 07-0413. Washington, DC (2007)

Geroliminis, N., Daganzo, C.: Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp. Res. B **42**, 759–770 (2008)

Geroliminis, N., Levinson, D.: Cordon pricing consistent with the physics of overcrowding. In: Proceedings of the 18th International Symposium on Transportation and Traffic Theory (2009)

Hendrickson, C., Plank, E.: The flexibility of departure time for work trips. Transp. Res. A **18**, 25–36 (1984)

Hollander, Y.: Direct versus indirect models for the effects of unreliability. Transp. Res. A **40**, 699–711 (2006)

Jou R.C., Kitamura R., Weng M.C., Chen C.C.: Dynamic commuter departure time choice under uncertainty. Transp. Res. A **42**, 774–783 (2008)

Levinson, D., Harder, K., Bloomfield, J., Carlson, K.: Waiting tolerance: ramp delay vs. freeway congestion. Transp. Res. F **9**, 1–13 (2006)

Levinson, D., Krizek, K.: Planning for Place and Plexus: Metropolitan Land Use and Transport. Routledge, New Yrok (2008)

Levinson, D., Zofka, E.: The metropolitan travel survey archive: a case study in archiving. In: Stopher P., Stecher C. (eds.) Travel Survey Methods: Quality and Future Directions, Proceedings of the 5th Intenational Conference on Travel Survey Methods, pp. 223–238. Emerald Group Pub Ltd, Bingley (2006)

Metropolitan Council of the Twin Cities Area: 2000 Travel behavior inventory home interview survey: data and methodology. Metropolitan Council, St. Paul, MI (2003)

Muñoz, J., Daganzo, C., Center, T., University of California (System): Fingerprinting traffic from static freeway sensors. University of California Transportation Center, University of California, California (2000)

Noland, R., Small, K., Koskenoja, P., Chu, X.: Simulating travel reliability. Reg. Sci. Urban Econ. **28**, 535–564 (1998)

Regional Transportation Management Center: http://www.dot.state.mn.us/tmc/tmctools.html. Accessed October 2008

Sarvi, M., Horiguchi, R., Kuwahara, M., Shimizu, Y., Sato, A., Sugisaki, Y.: A methodology to identify traffic condition using intelligent probe vehicles. In: Proceedings of 10th ITS World Congress, Madrid, pp. 17–21 (2003)

Small, K.: The scheduling of consumer activities: work trips. Am. Econ. Rev. **72**, 467–479 (1982)

Tilahun, N., Levinson, D.: A moment of time: valuing reliability using stated preference. J. Intell. Transp. Syst. (2006, in press)

University of Minnesota: Metropolitan Travel Survey Archive. http://www.surveyarchive.org (2003). Accessed Oct 2008

Vickrey, W.: Pricing in urban and suburban transport. Am. Econ. Rev. JSTOR **53**, 452–465 (1963)

Vickrey, W.: Congestion theory and transport investment. Am. Econ. Rev. **59**, 251–260 (1969)

Wu, X., Levinson, D., Liu, H.: Perception of waiting time at signalized intersections. Transp. Res. Rec. **2135**, 52–59 (2009)

Author Biographies

Pavithra Parthasarathi is a Ph.D. candidate in the Department of Civil Engineering at the University of Minnesota. Her research focuses on understanding the linkages between transportation networks, travel behavior, and urban form. Prior to starting her Ph.D. in August 2007, she worked as a transportation consultant for over 5 years.

Anupam Srivastava is a M.S. student in the Department of Civil Engineering at the University of Minnesota. He holds a bachelor's degree from the Indian Institute of Technology (IIT), Kharagpur. His current work involves developing density based ramp metering algorithms, under the supervision of Dr. Nikolas Geroliminis.

Nikolas Geroliminis is the Director of the Urban Transport Systems Laboratory and an Assistant Professor at EPFL, Switzerland. He has a diploma in Civil Engineering from NTUA, Greece and a M.Sc. and Ph.D. from UC Berkeley. Before joining EPFL he was an Assistant Professor at the University of Minnesota. His research interests focus on urban transportation systems, traffic flow theory and control, public transportation and logistics.

David Levinson holds the RP Braun/CTS Chair in Transportation and is the Director of the Networks, Economics, and Urban Systems (NEXUS) research group at the University of Minnesota. His research focuses on understanding the process of network growth, evaluating transportation technology and policy, and modeling transport-land use interactions.