TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Article

Transportation Research Record
2018, Vol. 2672(8) 187–198

DOI: 10.1177/0361198118793273
journals.sagepub.com/home/trr
SAGE

A Study of Bus High-Resolution GPS Speed Data Accuracy

Miguel A. Figliozzi[1] and Nicholas B. Stoll[1]

Abstract

The recent availability of high-frequency transit (HFT) GPS data for buses has allowed the estimation of detailed bus-travel speed profiles between bus stops. HFT data are defined as data comprising GPS vehicle trajectory recorded at or less than 5 s intervals. With HFT data it is now possible to measure changes in bus speed at specific locations of interest, such as intersections, ramps, crosswalks, and so on. Previous research efforts have not compared GPS-based bus speeds with general traffic speeds at specific locations between stops. This research fills this knowledge gap, utilizing accurate stationary radar speed data as a baseline or ground truth data to estimate GPS-based bus speed accuracy. A thorough data analysis of the bus and traffic speed data indicates that HFT speed estimations between stops are accurate and highly correlated with traffic speeds. Time–space speed profiles and regression analysis are utilized to quantify factors that affect HFT speed estimation accuracy. The relative advantages and limitations of the HFT data are presented and discussed. This study concludes that large HFT datasets can be utilized to accurately monitor speeds between transit stops. HFT datasets are suitable to cost-effectively monitor recurrent arterial speed performance for both passenger and transit vehicles.

There are several well-established technologies for collecting speed and travel time data, including vehicle identification sensors—like Bluetooth readers, radar devices, loop detectors, and probe car data. Loop detectors and vehicle identification sensors, once installed, continuously record most (or a share of in the case of Bluetooth readers) vehicles passing specific road sections. Bluetooth readers sample a fraction of the traffic and can estimate average speeds between two readers, but cannot produce speed profiles between the reader locations. The density of Bluetooth, radar, or loop detectors is typically very low on most arterials and most non-freeway network links do not have speed profile data.

This research utilizes high-frequency transit (HFT) bus (probe) vehicle data to estimate arterial street speed and produce speed profiles along a designated arterial. The GPS data are denoted as high frequency when GPS points are obtained at or less than 5 s intervals; the vast majority of transit agencies in the USA are currently storing 30 s interval GPS data or not storing any GPS data outside transit stops. Dedicated probe vehicles can be utilized to collect travel time and other data for designated routes in the network. However, because of cost considerations, government agencies cannot run probe vehicles continuously; moreover, privacy concerns make it problematic to gather data from private vehicles. Private data vendors can provide crowdsourced probe vehicle data, but these datasets can be expensive, and in many cases the data are already aggregated or preprocessed.

The recent availability of HFT data for buses has allowed the estimation of bus-travel speed profiles between bus stops. HFT data are defined as data comprising GPS vehicle trajectory records at or less than 5 s intervals. With HFT data it is now possible to measure changes in bus speed at specific locations such as intersections, ramps, crosswalks, and so on. No previous research efforts have compared GPS-based bus speeds with general traffic speeds at specific locations between stops. This research fills this knowledge gap. This research utilizes accurate stationary-radar speed data as a baseline or ground truth data to estimate GPS-based bus speed accuracy. More specifically, this research is original because it aims to answer two novel research questions: (1) how accurate are the speed measurements obtained utilizing HFT data when compared to state-of-the-art stationary speed sensors; and (2) what are the factors likely to affect the accuracy of HFT speed estimations at specific locations between bus stops?

[1]Transportation Technology and People (TTP) Lab, Department of Civil and Environmental Engineering, Portland State University, Portland, OR

Corresponding Author:
Address correspondence to Miguel A. Figliozzi: figliozzi@pdx.edu

To analyze the accuracy of the HFT speed data, exploratory data analysis, correlations, time–space speed profiles, and regression analysis are utilized. The final sections of this study analyze the impact of transit vehicle frequency and accuracy, as well as advantages and limitations to the HFT speed data. The next section provides a brief literature review.

Literature Review

Previous research efforts with GPS vehicle data have tried to utilize data publicly available from taxi fleets or ambulances; these efforts have focused on estimating speeds between segments or at the network level. Traffic speeds from taxis or ambulances are not calculated at a specific location because of the irregularity of most taxi or ambulance trips. Researchers have used these data from ambulances (*1*) or taxi fleets (*2*, *3*) to estimate hourly average link or path travel times. Other researchers have utilized private data sources from crowdsourced mobile phone data (*4*) or Bluetooth sensors that read and track devices carried via vehicles or people, such as smartphones (*5*). Other researchers have focused on datasets that are sparse concerning low-frequency GPS sampling but use many probe vehicles (*3*, *6*).

Unlike previous research efforts, the HFT data analyzed in this paper are sparse concerning the number of probe (bus) vehicles per hour, per link—the flow of transit vehicles per hour—instead of sparse concerning the GPS sampling frequency. Research indicates that probe vehicle flow rates significantly affect travel time estimation accuracy. Results show that a significant reduction in speed estimation variance is observed when the number of observations per hour increases from 12 to 36 probe vehicles per hour. Minor gains in estimation accuracy are obtained beyond 48–60 probe vehicles per hour (*7*).

Using public buses as probe vehicles *between bus stops* has been studied in the past (*8*, *9*). These early research efforts revealed that when automobiles experience long delays, buses on the same facility are also likely to be delayed but the reverse relationship is not always true, such as when buses dwell at stops because they are ahead of schedule. Previous research efforts in the Portland region have utilized stop-to-stop bus-travel data to assess arterial performance and transit performance (*10*). However, all of these studies (*8–10*) were severely limited by the lack of GPS coordinates between bus stops; the datasets utilized in these studies (*8–10*) were sparse concerning both probe vehicle flow rates and GPS sampling frequency.

The recent availability of 5 s GPS data for buses has removed much of the guesswork involved in estimating bus-travel speed profiles between bus stops (*11*); it is now possible to measure *relative* changes in bus speed at intersections, ramps, crosswalks, and so on (*11*). "Relative" is emphasized because previous research efforts have not compared the actual speeds of transit vehicles and general traffic. For example, Figure 1 shows a time–space

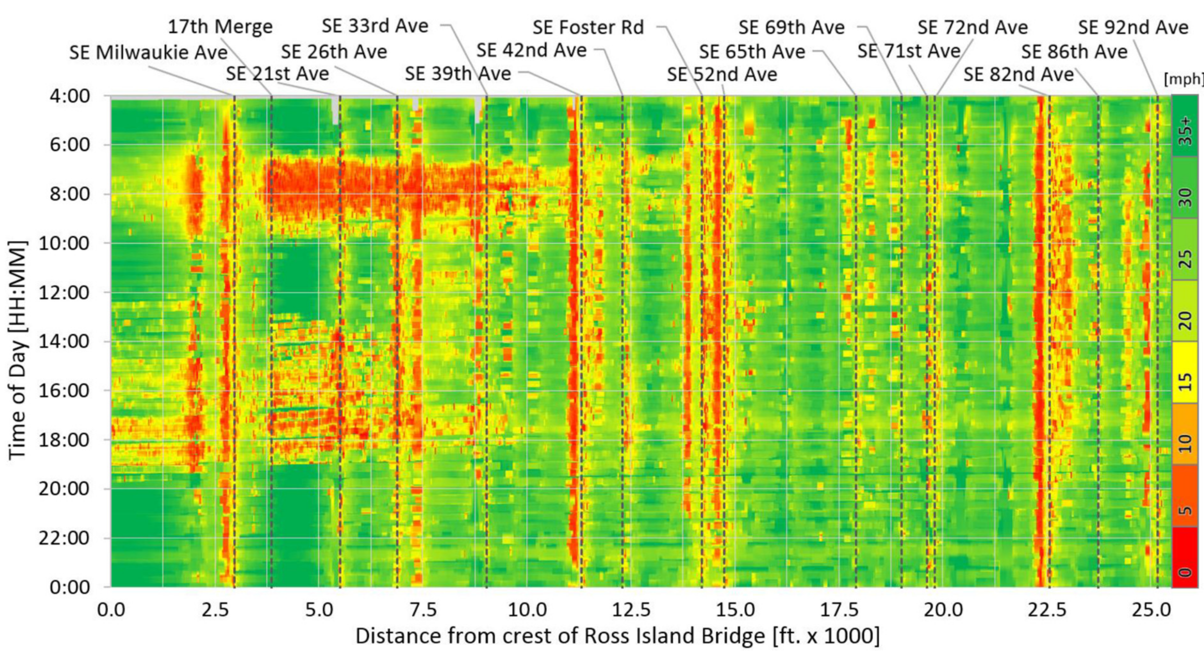

Figure 1. Space–time speed heatmap without dwell times, travel left-to-right, adapted from Glick and Figliozzi (*11*).

speed diagram constructed utilizing HFT data. In Figure 1, signalized intersections are labeled and it is possible to observe delays at some busy intersections like SE 39th or SE 82nd, which are major four-lane arterials. It is also possible to observe peak-hour morning congestion between SE 17th and SE 39th (caused by a bottleneck at a congested downstream bridge access). The most significant limitation of the utilized HFT speed data is that buses do not reflect traffic speeds while servicing customers at a bus stop. However, a recent publication (*11*) showed that *between stops* buses can be utilized to detect recurrent congestion and analyze traffic trends; dwell times or stop service times were removed to create Figure 1—see Glick and Figliozzi for more details (*11*).

Past research has compared speed estimates for link travel times (*9*) but not speeds at a specific location using stationary sensors as a baseline or as ground truth data. No previous research efforts have compared HFT speeds with stationary-speed-sensor data. This research fills this knowledge gap, utilizing Wavetronix Digital Wave Radar (DWR) sensor data as a baseline or ground truth data by measuring on location and with high accuracy. Mid-block Wavetronix sensors are very accurate when properly installed, measuring speeds and volumes with a 1–2% error range (*12–14*) or a spread of ± 1–2 mph when compared against speed guns (*14*, *15*).

HFT bus speed data can even detect bus speed changes when approaching intersections, as well as bus speed changes that are caused by on-ramps and crosswalks (*16*). This research also takes advantage of recent algorithms and data processing strategies to estimate speed percentiles for spot speeds at specific locations (*11*) that may remove the bias that is generated by bus stops (*17*) or traffic signal stops because the buses cannot take advantage of signal progression (*18*). These improvements are key to eliminating the differences between transit and private vehicle speeds by utilizing only transit data from vehicles that are not stopped by an upstream traffic signal and intermediate transit stop. Better and more accurate transit data can then be utilized to visualize transit operations (*19*), transit delay (*20*), and forecast travel times (*21*).

Available HFT Data and Study Location

The HFT data utilized in this research are provided by TriMet, the public transportation agency in the Portland region. On most arterials outside the downtown transit mall, the TriMet HFT data are relatively *sparse* because the flow of vehicles (buses) is, at most, 6–10 vehicles per hour in the majority of arterials and collector streets outside downtown Portland. However, the data have high frequency because GPS data are recorded in 5 s intervals. Reliable data are available for over 220 weekdays per year if holidays and abnormal weeks are removed from the dataset. Therefore, millions of GPS coordinates can be easily obtained over a 6 month period. Other advantages of TriMet's dataset include: (1) the data are freely available; (2) buses travel across congested areas and cover all major arterials and collector streets in the region (even some freeways for express routes); (3) bus data are linked to a physical sensor in the vehicle (measuring wheel motion), making it possible to detect congestion with more accuracy than with GPS data alone; (4) bus stop activity is recorded and GPS data around bus stops can be removed if necessary; and (5) unlike taxis or ambulances, bus-travel paths tend to be extremely regular.

The study location was chosen based on three factors: (1) availability of high-quality Wavetronix data; (2) availability of HFT data; and (3) familiarity with the area. The chosen location, southeast Powell Boulevard, meets the three criteria. The accuracy of the two Wavetronix sensors was previously analyzed as part of a research effort along the same corridor (*22*). Nonetheless, as part of this research effort more recent Wavetronix speed–volume data were reanalyzed to evaluate accuracy of the sensors: no problems or errors were found. Along SE Powell runs the TriMet frequent-service route 9, with approximately 10 min headways during peak hours and 15 min headways during off-peak hours. The researchers are thoroughly familiar with the research area and traffic patterns. This roadway is part of an instrumented corridor where the research team has recently studied the performance of the adaptive traffic signal system SCATS (*22*).

Along SE Powell Boulevard, two Wavetronix sensors are located mid-block near cross streets southeast 24th Avenue and southeast 35th Avenue. The locations of these Wavetronix sensors were chosen to ensure high-quality data and also to best capture free-flow traffic during off-peak hours and severe queuing during peak hours (*22*). At SE Powell and 24th, the distance to the nearest signalized intersections is approximately 655 ft (200 m); at SE Powell and 35th, the distance to the nearest signalized intersections is approximately 1050 ft (320 m). At SE Powell Boulevard and 24th, a Wavetronix sensor measures speed and volumes across a section comprising two through westbound lanes, two through eastbound lanes, and a middle left-turn lane; a similar configuration is present at the intersection of SE Powell Boulevard and 35th. The distances between bus stops vary per location and direction of travel. Distances between the Wavetronix sensors and the bus stops (in feet) are included in Figure 2. Details for each location and travel direction are summarized below:

- Eastbound 24th: The Wavetronix sensor is approximately 50 ft (15.3 m) upstream of the

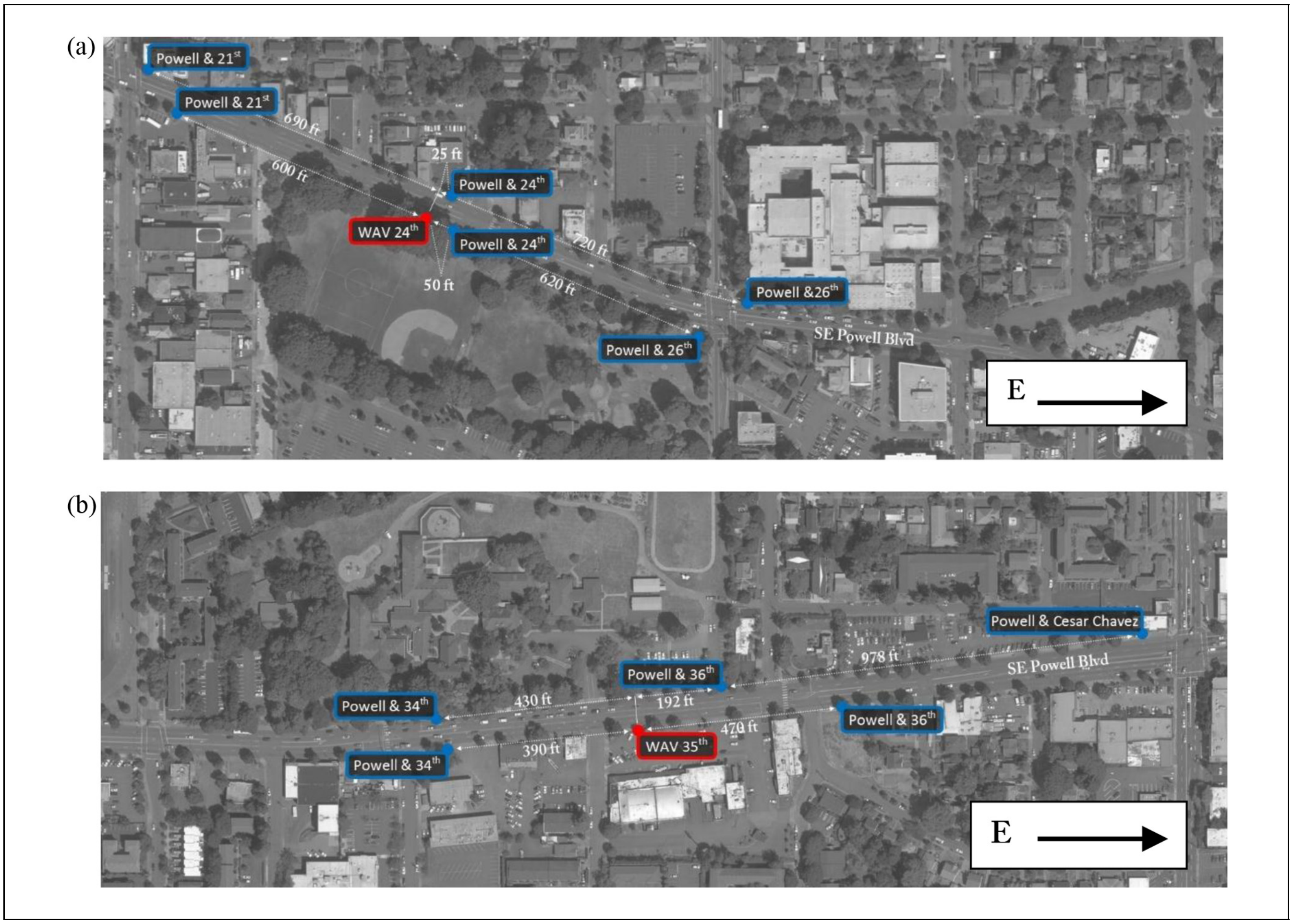

Figure 2. Sensor and bus stop locations: (*a*) SE Powell and 24th, distance between Wavetronix sensor (red) and bus stops (blue); (*b*) SE Powell and 35th, distance between Wavetronix sensor (red) and bus stops (blue).

nearest bus stop (stop ID 4625) and over 600 ft (183 m) downstream of the SE Powell Boulevard and 21st bus stop (stop ID 4622). It is also within close proximity of a crosswalk. At the SE Powell Boulevard and 24th bus stop (stop ID 4625), less than 20% of bus trips service this stop; conversely, more than 80% of bus trips service the stop at the SE Powell Boulevard and 21st (stop ID 4625).

- Westbound 24th: The nearest bus stop is roughly 25 ft (7.5 m) upstream of the sensor (stop ID 4626), but this bus stop is not heavily serviced; roughly 30% of bus trips service this stop.
- Eastbound 35th: There are no bus stops within close proximity to the Wavetronix sensor. The closest upstream bus stop is located 350 ft away at SE Powell Boulevard and 34th (stop ID 4647).
- Westbound 35th: The upstream bus stop is located 190 ft upstream of the sensor, at SE Powell and 36th (stop ID 4649); roughly 50% of the buses service this stop.

Data Processing

TriMet supplied four different datasets: cyclic data, stop event data, stop data, and block data. Cyclic data are recorded at 5 s intervals. This is the probe (bus) vehicle GPS data denominated as HFT data or 5 s resolution data (5-SR data); see Table 1 for a small data sample. The first column, trip number, is used to identify a bus trip. The second column, stop number, is updated whenever the bus services a stop. For example, in Table 1 the bus serviced four different stops. The third and fourth columns record odometer distance in meters and time after midnight in seconds, respectively. The fifth and sixth columns record the latitude and longitude of the GPS locations, respectively.

The stop event and stop datasets contain information on instances where the bus stops—for example, type of stop, dwell time, estimated bus load, boardings, alightings, and so on. Stop event and stop data can be

Table 1. Sample of HFT or 5-SR Dataset

Trip number	Stop number	Distance	Time	Longitude	Latitude
247917030	247917070	65169	30088	45.49713	−122.537
247917030	247917070	65210	30093	45.49712	−122.537
247917030	247917070	65252	30098	45.49711	−122.538
247917030	247917071	65313	30123	45.49707	−122.539
247917030	247917071	65375	30128	45.49704	−122.539
247917030	247917071	65453	30133	45.497	−122.54
247917030	247917072	65531	30138	45.49697	−122.541
247917030	247917072	65608	30143	45.49694	−122.542
247917030	247917072	65685	30148	45.49691	−122.543
247917030	247917073	65763	30153	45.49685	−122.544
247917030	247917073	65832	30158	45.49677	−122.545

combined with the 5-SR data for a given trip number and stop number, but require the bus block dataset to merge all the datasets. The Wavetronix sensors along SE Powell Boulevard record data in 10 s intervals. Wavetronix data provide traffic speed, volume, and occupancy data per lane, aggregated in consecutive 10 s intervals.

The data processing steps, at a conceptual level, can be succinctly described as four major steps:

1. Extract right lane Wavetronix data for the time intervals of interest (only right lane data because buses travel along the right lane on Powell Boulevard).
2. Tag all HFT GPS points *before* reaching and *after* passing a Wavetronix sensor for each direction of travel and location.
3. Match Wavetronix 10 s interval data to the closest *before* and *after* HFT GPS point timestamps.
4. HFT speeds were calculated employing a methodology described in a recently published article (*11*).

Exploratory Data Analysis

The literature review section indicated that stationary sensors, especially Wavetronix radar sensors, are considered to be highly accurate devices for measuring roadway speeds and traffic volumes. Here, Wavetronix data are considered ground truth data and are used as a baseline to measure the quality of HFT speed data.

Figure 3 compares Wavetronix and HFT speed profiles when data are aggregated in 15 min, 30 min, 45 min, and 60 min intervals. The y-axis indicates average travel speed at a given interval and the x-axis is time of day (TOD) between 5:00 a.m. and 22:00 p.m. It is possible to observe that Wavetronix and HFT speed profiles do not overlap at all times, but there seems to be a strong correlation. In almost all cases, HFT speeds are lower than general traffic speeds. A clear gap can be seen at the westbound 24th and westbound 35th locations.

Speed distribution kernel plots indicate that when HFT speeds are lower—for example, at the westbound 24th and westbound 35th locations—the speed distributions are clearly bimodal (see Figure 4). If Wavetronix and HFT speed data are plotted utilizing different colors for specific bus events, it is possible to find some noteworthy patterns. For example, Figure 4 shows the relationship between Wavetronix speeds and HFT 5-SR data, where colors are utilized to differentiate events when the bus door opens (i.e., services a bus stop) at the nearest bus stop downstream from the Wavetronix sensor. The small secondary peaks seen in blue at the westbound locations in Figure 4 are matched by a cloud of low-speed blue points (opened bus doors) that are shown in Figure 5.

Regression Analysis

To analytically identify the key variables affecting HFT speed quality, a regression analysis was undertaken. The following variables were included in the analysis:

- Dependent variable: Error or speed difference (i.e., Wavetronix speed minus HFT speed) for each bus trip. It is important to note that transit vehicle speed is estimated at the location of the Wavetronix sensor and that some locations are near heavily used bus stops, while other locations are farther away from bus stops.
- Independent variables: Variables associated with the vehicle and vehicle activity (e.g., bus load, whether an upstream or downstream stop was serviced, whether bus doors opened, stop service time, and so on), variables associated with traffic conditions (Wavetronix data for traffic speeds, volumes, sensor occupancy, vehicle classification, and so on), and variables related to the nearest GPS coordinates (distance to Wavetronix sensor, time of the day, and so on).

A regression analysis with these dependent and independent variables has not been performed before. To time-effectively find patterns, backward linear regression models were estimated utilizing all the regression variables that were not significantly correlated. A summary of the regression results per location and direction of travel are reported in Table 2. Only coefficients for significant variables are reported, and the coefficient's relative importance is also reported. Relative importance was measured by the contribution of each independent variable to explained variance. The R package "relaimpo" was utilized to estimate all statistics.

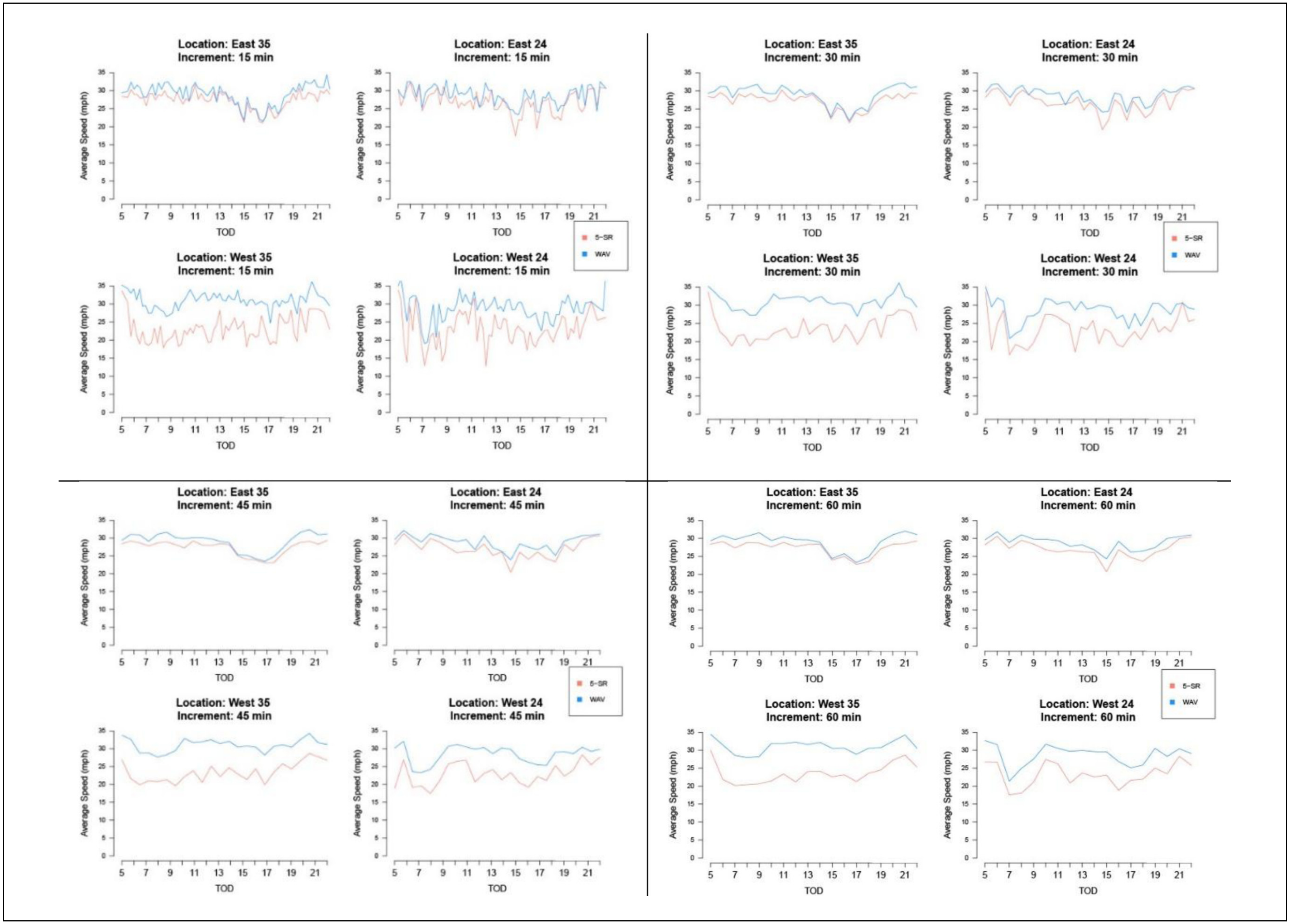

Figure 3. Comparison of average speed profiles by TOD.

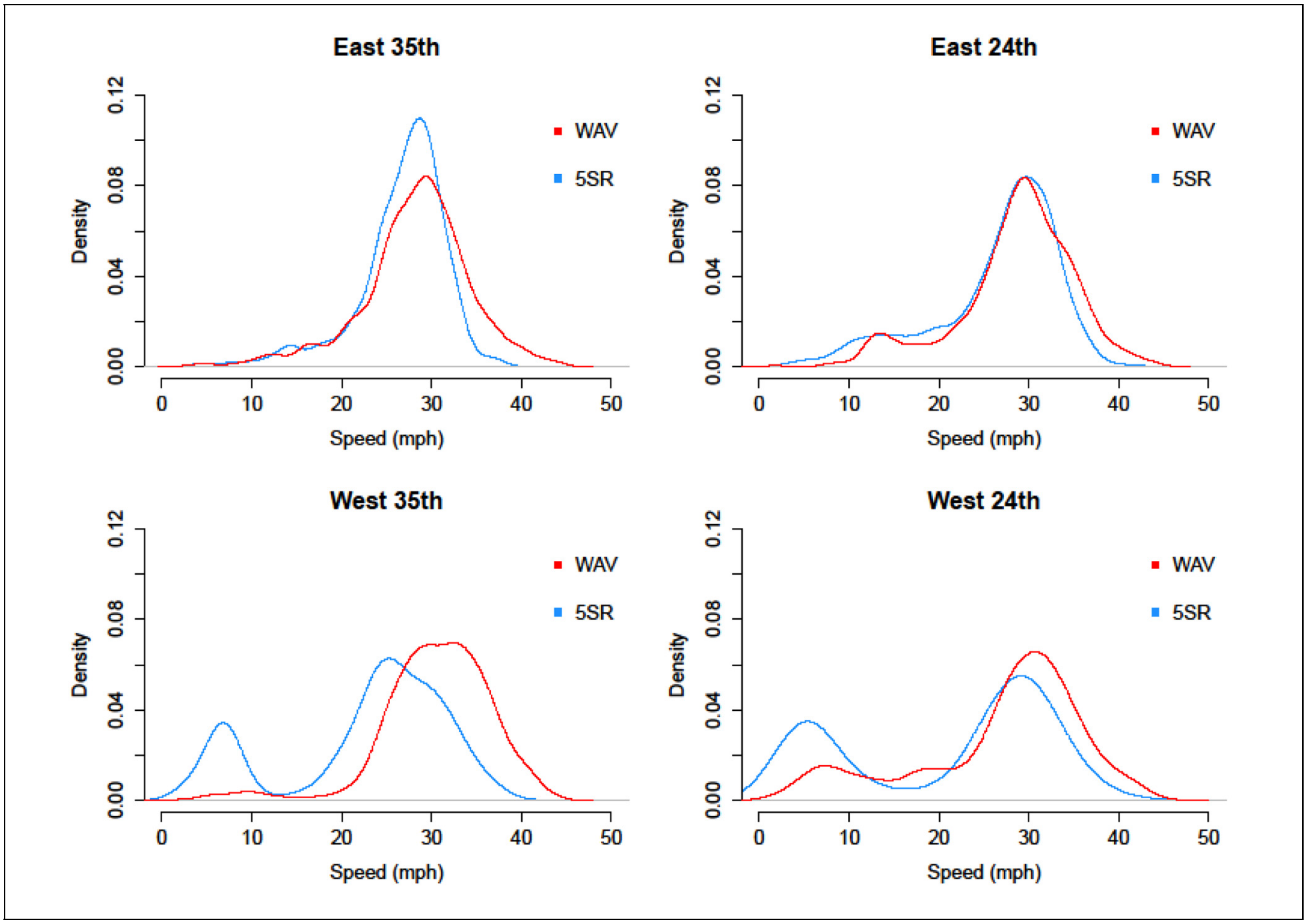

Figure 4. Kernel density plots.

The results of the regression analysis were highly encouraging: (1) All the models had relatively high adjusted R^2 values and (2) there was sign consistence among the key significant variables. The potential interpretation of the key variables is mostly intuitive. Across the board, Wavetronix speed was a highly significant variable, with high relative importance, and a positive sign. The Wavetronix speed variable indicates that, all things being equal, the higher the traffic speed, the larger the error or difference between Wavetronix and HFT speed. This is intuitive because buses are large and heavy vehicles with limited maneuverability and a slower acceleration profile than most passenger vehicles. The speed limit is 35 mph along this section of SE Powell.

The Wavetronix data also show that during free-flow conditions, a significant portion of regular vehicles travel at or faster than the 35 mph speed limit (see Figure 4); owing to their safety training and continuous video monitoring, bus drivers are more likely to obey speed limits than are regular drivers. When traffic conditions are congested, the difference between traffic and HFT speed decreases. This is favorable for the analyst who is interested in finding congestion in specific areas or TODs.

Important differences can also be observed in Table 2. If the bus door opened after passing the Wavetronix sensor, the error or speed difference increases significantly. This variable has a high relative importance only for westbound 24th and 35th; for eastbound 24th and 35th locations, the stop variables are less important. This finding is consistent with the plots shown in Figure 5 (see the difference between red and blue dots at some locations). The other variables—distance between GPS points and the Wavetronix sensors—have mostly negative values.

Error or speed difference tends to decrease as distance increases, holding all the other variables constant. This finding is intuitive because HFT speeds and distance between GPS points are positively correlated. The GPS data are recorded in 5 s intervals; the faster the bus travels, the greater the distance between GPS points. The locations near 24th Avenue have significantly lower R^2 values than the locations near 35th Avenue These bus stops are close to the Wavetronix sensor at 24th Avenue

Table 2. Regression Summary and Comparison

Variable	Westbound 24th		Eastbound 24th		Westbound 35th		Eastbound 35th	
	Coefficient	Relative importance	Coefficient	Relative importance	Coefficient	Relative importance	Coefficient	Relative importance
Intercept	−13.404***	na	−8.189***	na	−14.451***	na	1.202***	na
Sensor speed	0.774***	26.6%	0.655***	49.4%	0.814***	27.1%	0.950***	58.2%
Distance: sensor to point (before)	−0.664***	12.1%	−0.745***	31.7%	0.356***	8.2%	−1.400***	20.8%
Distance: sensor to point (after)	−0.056**	3.9%	−0.141***	6.2%	−0.617***	20.2%	−1.399***	20.9%
Bus door opened (before)	−2.199***	0.9%	−3.158***	2.1%	−7.659***	2.2%	0.200**	0.1%
Bus door opened (after)	15.638***	56.6%	6.150***	10.6%	13.105***	42.3%	na	na
Adjusted R^2	0.687		0.519		0.734		0.914	

$^*p < 0.1$. $^{**}p < 0.05$. $^{***}p < 0.01$.

and this may explain the drop in adjusted R^2 values. It is likely that the available variables failed to capture the changes in errors or speed variability at 24th Avenue.

Traffic performance along an arterial is a complex spatial and temporal phenomenon. The regression results are useful but it is unlikely that a linear regression model can capture all the spatial and temporal subtleties of arterial traffic. Space–time speed profiles are typically more useful to describe traffic patterns, and these are analyzed in the next section.

Time–Speed Profiles and Speed Correlations

The time–speed profiles for each location and direction of travel are presented in Figures 6 and 7. The distances at the bottom of the graph are centered on the location of the Wavetronix sensor. At the bottom of each profile, a graph depicts estimated bus delay and the location of the Wavetronix sensor and nearby bus stops. It is possible to observe significant delays around heavily used bus stops. In addition, the speed profile for westbound 24th shows congestion in the morning hours (see Figure 6*b*); this is realistic as commuter traffic moves toward downtown Portland and generates morning queues at a downstream bridge that propagate all the way up to 24th Avenue and beyond. The same congestion appears in Figure 1 between 3 and 8 miles and between 7:00 and 9:00 a.m.

Table 3 shows correlations between Wavetronix speeds and HFT speeds. The top portion of the table shows the correlations when speeds are measured or estimated at the Wavetronix sensor. At eastbound 24th and eastbound 35th there is a high degree of correlation; at westbound 24th and westbound 35th the correlations are significantly lower, and this can be explained by the proximity of the Wavetronix sensor to heavily used nearby bus stops (see Figures 6 and 7 and regression analysis results). To test this hypothesis, HFT speeds were calculated 200 ft away from the sensor—see Figures 6*b* and 7*b* with a marker for "New Location." At westbound 24th (Figure 6*b*) and westbound 35th (Figure 6*a*) bus stop delays do not influence the HFT speed estimations. Table 3 shows that the correlations increased significantly at the new locations 200 ft away from the Wavetronix sensor at westbound 24th and westbound 35th.

However, moving the location of the HFT data speed estimation does not necessarily improve correlations in all cases. For example, at eastbound 24th (Figure 7*a*), if the HFT speed is estimated upstream, the correlation value decreases. In this example, the Wavetronix sensor is likely capturing delays from peak-hour afternoon queuing that originate at the downstream traffic signal (SE 26th) that are not captured when the HFT speed estimation location is moved 200 ft upstream. Figure 4 shows a slight bimodal distribution for the Wavetronix

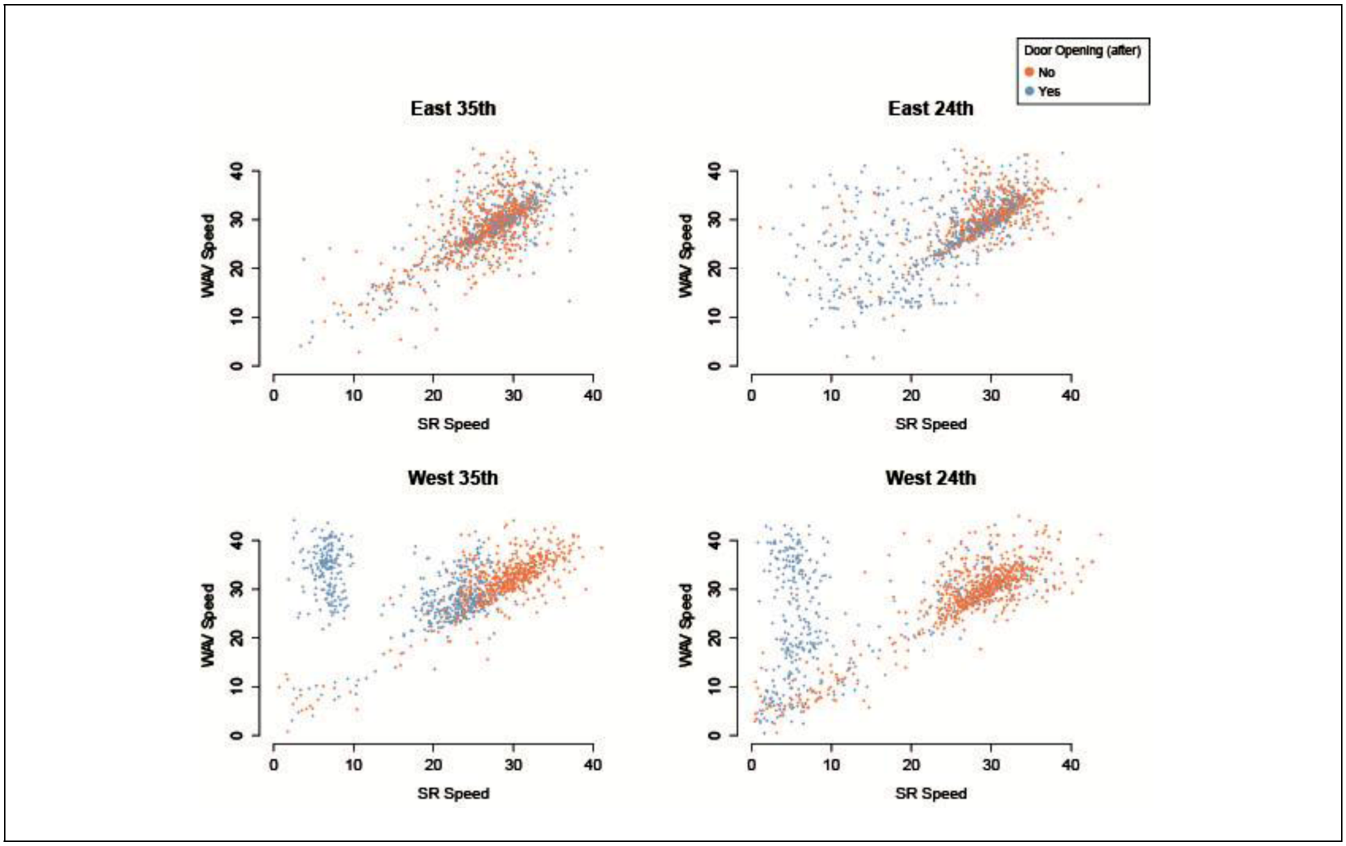

Figure 5. Wavetronix and 5-SR speed conditional on door opening (after point).

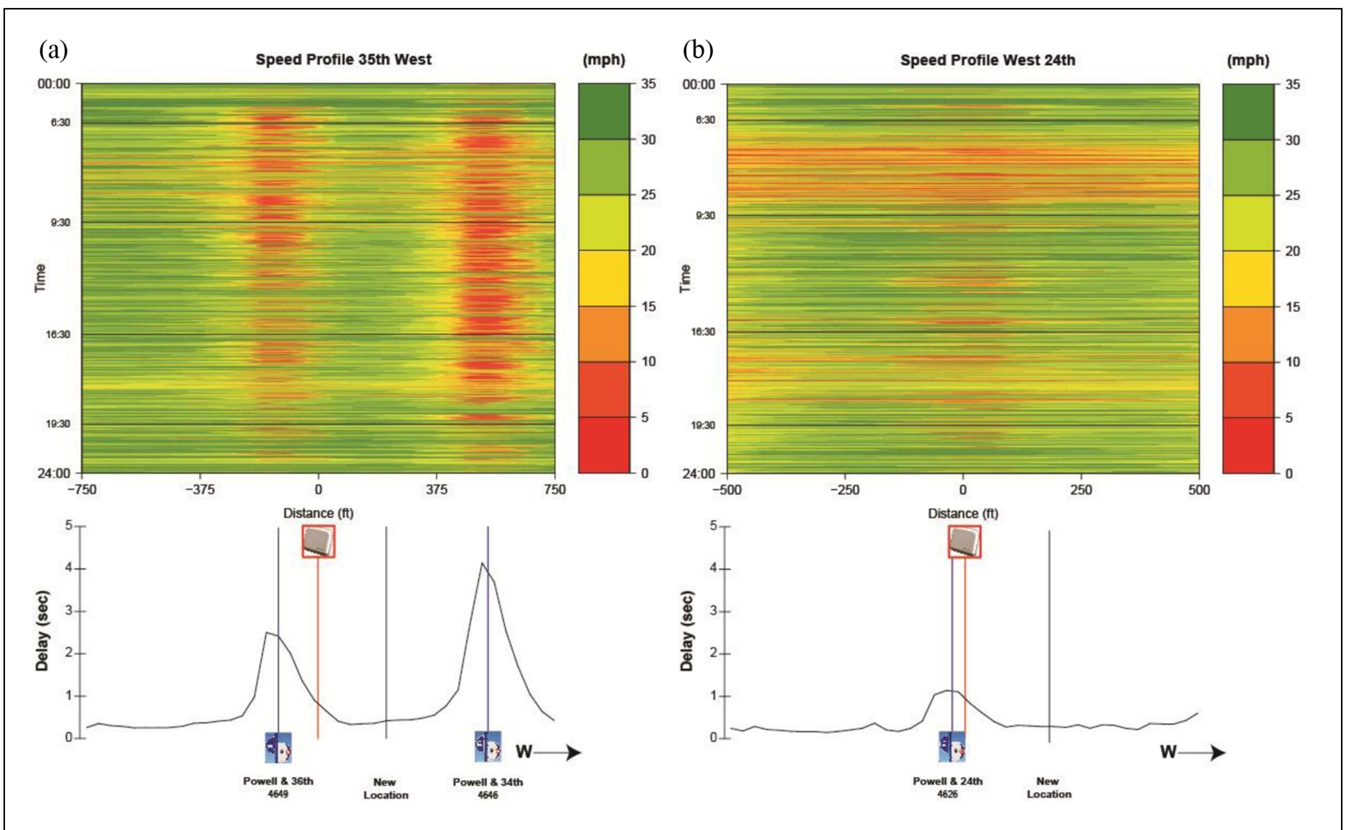

Figure 6. Speed heatmaps and sensor locations for westbound travel: (a) westbound 35th; (b) westbound 24th.

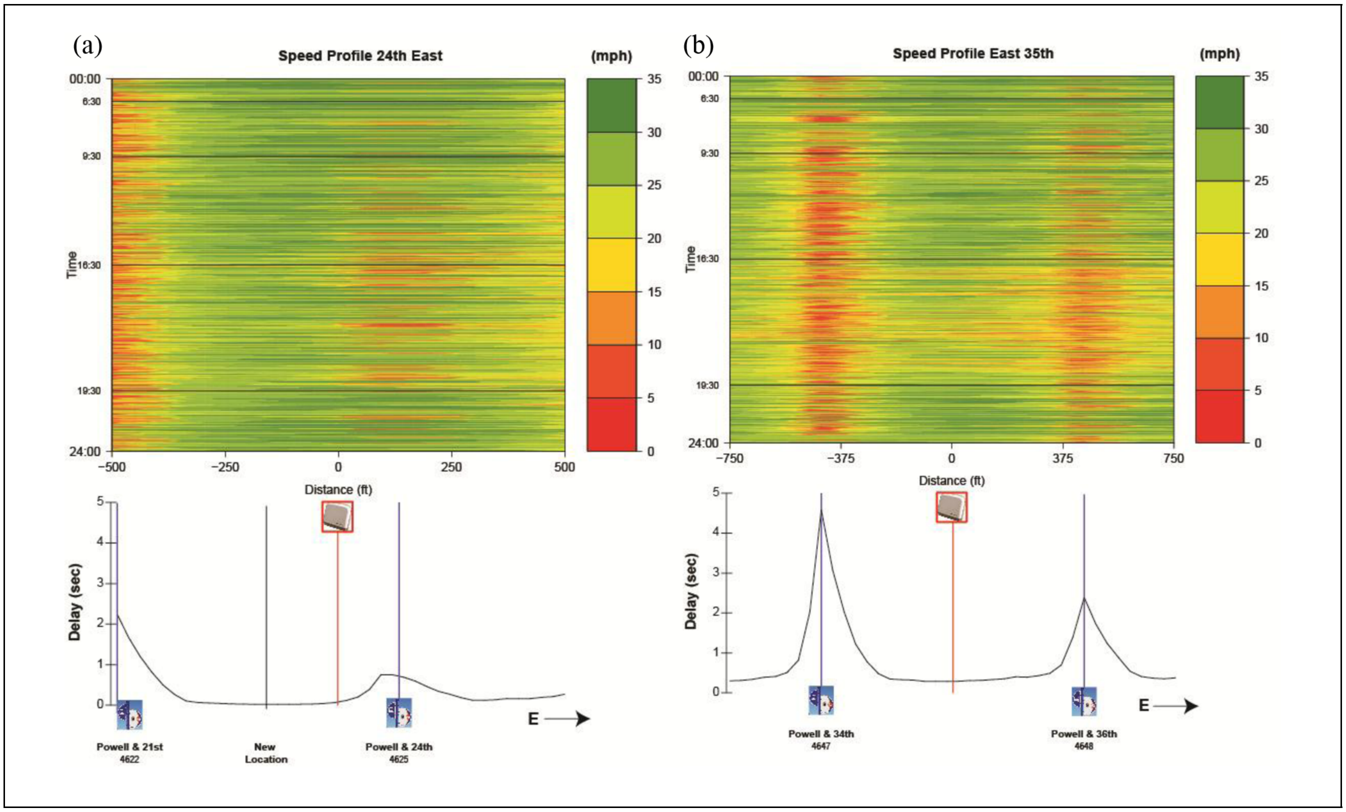

Figure 7. Speed heatmaps and sensor locations for eastbound travel: (a) eastbound 24th; (b) eastbound 35th.

Table 3. Correlations at Wavetronix Sensor and New Locations

Proximity to Wavetronix sensor	Location	Time intervals—Bin sizes 15 min	30 min	45 min	60 min
At the Wavetronix sensor	Eastbound 35th	0.92	0.94	0.96	0.96
	Eastbound 24th	0.80	0.90	0.90	0.94
	Westbound 35th	0.60	0.76	0.68	0.79
	Westbound 24th	0.70	0.74	0.70	0.84
Away from the Wavetronix sensor	Eastbound 24th	0.53	0.60	0.62	0.75
	Westbound 35th	0.84	0.92	0.91	0.94
	Westbound 24th	0.83	0.93	0.92	0.96

speeds at 24th. The low speeds of the second, smaller peak are likely to be associated with delays that originate at the downstream intersection when traffic is heavy. Yellow areas that propagate beyond the Wavetronix sensor can be observed in Figure 7a.

HTF Vehicle Flow Rates and Accuracy

The regression analysis and study of the time–speed profiles and correlations indicate that it is not accurate to include transit vehicles that service stops. It is possible to remove all the trips that service a stop location to increase the accuracy of the HFT speed estimation. It is recommended that HFT speed data are not utilized to evaluate arterial speeds along a segment that is ± 200 ft (± 60 m) around a heavily utilized bus stop without first removing GPS points that are near the bus stop. The bus stop impact area can be readily seen in a time–space speed profile (see Figures 6 and 7).

A reduction in the flow of HFT vehicles has to be offset by a longer data-collection period. The relationships between the frequency of probe data and variance of travel time estimates have been studied in the past (*7*); a significant reduction in variance is observed when the number of observations per hour increases from 12 to 36 probe vehicles per hour (*7*). Minor gains in estimation accuracy are obtained beyond 48–60 probe vehicles per hour. Therefore, it is desirable to obtain a sample size of 50 or more trips per hour. An example can be used to illustrate the tradeoffs. Assuming that nine out of 10 buses service a stop and bus headways are 12 min, to obtain 50 "undisturbed" weekday observations per hour it is necessary to analyze 20 weeks of data. In this case, "undisturbed" means vehicles that do not service a stop and therefore have a useful speed profile. Another example: If bus (vehicle) flow rate is low and equal to one vehicle per hour, in 10 weeks (or less than 3 months), it is possible to collect over 50 observations for a given hour of the day.

The minimum vehicle flow rate to obtain reasonable travel time estimation accuracy has other implications. It is unlikely to find arterials where the HFT flow rate is greater than 20 vehicles per hour per route. Assuming a minimum sample size of 50 more trips per hour, it is not possible to use HFT data to provide accurate real-time speed estimations unless five or six high-frequency bus routes are along an arterial. However, even if the hourly rate is below 50 trips per hour, the HFT data can provide accurate temporal depictions of average-speed travel times—that is, a good indicator for *recurrent* traffic conditions. If the analyst is interested in recurrent traffic conditions, the HFT data can provide accurate estimations of speed profiles along the arterial instead of speeds at specific locations where stationary sensors are located.

Conclusion

The exploratory data analysis, the time–space speed profiles, and the speed correlation graphs indicate that HFT bus speed data have a high degree of correlation with stationary Wavetronix sensors. HFT speed data can be used to cost-effectively analyze bus operations and monitor arterial recurrent speeds and performance with a high degree of accuracy. Better and more accurate transit data can then be utilized to visualize transit operations (*20*), estimate transit delay (*21*), and forecast travel times (*22*).

The high regularity of the HFT data means that in 3–6 months it is possible to accumulate large datasets with enough observations to analyze or detect locations and/or times with recurrent congestion along any section of an arterial. For bus HFT data, the results of this research show that a much shorter HFT data-collection period is necessary to obtain high correlations (greater than 0.90) with Wavetronix speed data when HFT speeds are not influenced by nearby heavily serviced bus stops. Around heavily utilized bus stops, the GPS data can be removed to avoid the bias introduced by slowed buses (*17*, *18*). Likewise, to eliminate differences between transit and private vehicle speeds, only transit data from vehicles that are not stopped by an upstream traffic signal and intermediate transit stop must be utilized (*19*).

HFT bus speed data are reasonably accurate and extremely economical when compared to the cost of purchasing, installing, and maintaining a network of stationary sensors. It is not argued here that large datasets of HFT speed data should replace stationary sensors; rather, HFT speed data can be used to complement stationary sensors and/or reduce the number of stationary sensors that are necessary to cost-efficiently monitor the performance of an arterial street or provide ground truth data and traffic volumes. Stationary sensors are useful to obtain data for a specific roadway section or point, whereas HFT speed data provide speed profiles that can be used to fill the gaps between stationary sensors. HFT data can also be used in real-time applications if the data are used to complement data from stationary sensors. Because of its potential cost-effectiveness, the joint utilization of HFT and stationary sensor data to monitor real-time or recurrent traffic conditions along arterial corridors is a subject that deserves additional research efforts.

Practical applications of this research include that knowing that HFT speed data are accurate allows the direct comparison of locations and the creation of a ranking of delays. Rankings are useful to prioritize locations that need improvements (e.g., transit signal priority or queue jump treatments at signalized intersections) and to produce speed heatmaps that show the impact of general traffic congestion on transit performance.

Acknowledgments

The authors would like to thank TriMet staff members Steve Callas and Miles J. Crumley for graciously providing the datasets used in this analysis and for their support in understanding the intricacies of how the data are structured. The authors would also like to thank Ellen Bradley for carefully editing and proofreading the paper. The authors would also like to acknowledge the support of NITC (National Institute for Transportation and Communities) Transportation Center for funding this research effort.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: MAF, NBS; data collection: Tri County Metropolitan District of Oregon (TriMet); analysis and interpretation of results: MAF, NBS; draft manuscript preparation: MAF, NBS. All authors reviewed the results and approved the final version of the manuscript.

References

1. Westgate, B. S., D. B. Woodard, D. S. Matteson, and S.G. Henderson. Large-Network Travel Time Distribution

Estimation for Ambulances. *European Journal of Operational Research*, Vol. 252, No. 1, 2016, pp. 322–333.
2. Zhan, X., S. Hasan, S. V. Ukkusuri, and C. Kamga. Urban Link Travel Time Estimation Using Large-Scale Taxi Data with Partial Information. *Transportation Research Part C: Emerging Technologies*, Vol. 33, 2013, pp. 37–49.
3. Jenelius, E., and H. N. Koutsopoulos. Travel Time Estimation for Urban Road Networks Using Low Frequency Probe Vehicle Data. *Transportation Research Part B: Methodological*, Vol. 53, 2013, pp. 64–81.
4. Bar-Gera, H. Evaluation of a Cellular Phone-Based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel. *Transportation Research Part C: Emerging Technologies*, Vol. 15, No. 6, 2007, pp. 380–391.
5. Wasson, J., J. Sturdevant, and D. Bullock. Real-Time Travel Time Estimates Using Media Access Control Address Matching. *ITE Journal*, Vol. 78, No. 6, 2008, pp. 20–23.
6. Herring, R., A. Hofleitner, P. Abbeel, and A. Bayen. Estimating Arterial Traffic Conditions Using Sparse Probe Data. *Proc., 13th International IEEE Conference on Intelligent Transportation Systems*, Funchal, Portugal, IEEE, New York, NY, pp. 929–936.
7. Sen, A., P. Thakuriah, X. Q. Zhu, and A. Karr. Frequency of Probe Reports and Variance of Travel Time Estimates. *Journal of Transportation Engineering*, Vol. 123, No. 4, 1997, pp. 290–297.
8. Hall, R., and N. Vyas. Buses as a Traffic Probe: Demonstration Project. *Transportation Research Record: Journal of the Transportation Research Board*, 2000. 1731: 96–103.
9. Cathey, F., and D. Dailey. Transit Vehicles as Traffic Probe Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2002. 1804: 23–30.
10. Bertini, R., and S. Tantiyanugulchai. Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1870: 35–45.
11. Glick, T. B., and M. A. Figliozzi. Traffic and Transit Travel Time Reliability Indices and Confidence Intervals. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. 2649: 28–41.
12. Gates, T., S. Schrock, and J. Bonneson. Comparison of Portable Speed Measurement Devices. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1870: 139–146.
13. Mohammad, H., S. Schrock, and E. Fitzsimmons. Accuracy of Traffic Speed and Volume Data Detected Using Radar Technology. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.
14. Minge, E., J. Kotzenmacher, and S. Peterson. *Evaluation of Non-Intrusive Technologies for Traffic Detection*. No. MN/RC 2010-36. Minnesota Department of Transportation, Research Services Section, Saint Paul, MN, 2010.
15. Sanchez, G. *Evaluation of the Accuracy of Approach Volume Counts and Speeds Collected by Microwave Sensors*. Thesis. Brigham Young University, Provo, UT, 2016.
16. Stoll, N., T. B. Glick, and M. A. Figliozzi. Utilizing High Resolution Bus GPS Data to Visualize and Identify Congestion Hot-Spots in Urban Arterials. *Transportation Research Record: Journal of the Transportation Research Board*, 2016. 2539: 20–29.
17. Glick, T., and M. A. Figliozzi. Measuring the Determinants of Bus Dwell Time. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. 2647: 109–117.
18. Thornton, D., and B. Coifman. Signal Progression Impacts on Transit Buses as Travel Time Probes. *Journal of Transportation Engineering*, Vol. 141, No. 8, 2015, 04015009.
19. Kurkcu, A., F. Miranda, K. Ozbay, and C. T. Silva. Data Visualization Tool for Monitoring Transit Operation and Performance. *Proc., 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Naples, Italy, IEEE, New York, NY, 2017, pp. 598–603.
20. Sidhu, B. S., R. L. Bertini, and A. Pande. Exploiting New High Resolution Transit Data Sources: Developing Improved Data Driven Models of Transit Travel Time and Delay. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
21. Hans, E., N. Chiabaut, L. Leclercq, and R. L. Bertini. Real-Time Bus Route State Forecasting Using Particle Filter and Mesoscopic Modeling. *Transportation Research Part C: Emerging Technologies*, Vol. 61, 2015, pp. 121–140.
22. Slavin, C., W. Feng, M. Figliozzi, and P. Koonce. Statistical Study of the Impacts of SCATS Adaptive Traffic Signal Control on Traffic and Transit Performance. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2356: 117–126.

The Public Transportation Group (AP000) peer-reviewed this paper (18-05735).

Any errors or omissions are the sole responsibility of the authors.