Effects of Temporal Data Aggregation on Performance Measures and Other Intelligent Transportation Systems Applications

Alexander Y. Bigazzi, Helene Siri, and Robert L. Bertini

Intelligent transportation systems (ITSs) data are a valuable resource for traffic operations, transportation systems management, performance measurement, and transportation research. Historically, these data are time-aggregated for collection, transmission, and storage, with only mean values saved for traffic parameters for each arbitrary time interval. This convention of aggregation discards valuable information that is necessary for some applications. To understand whether systems should continue the practice of aggregation, this paper investigates how temporal aggregation can affect performance measures and other data applications. The investigation uses disaggregate speed data from loop detectors on a London freeway and vehicle trajectories from video imaging on a California freeway. Aggregating measured speed data greatly reduces the spread in reported vehicle speeds, which will distort estimates of emissions, fuel consumption, and travel delay. Using aggregate data for travel time estimates from sampled speeds results in errors attributable to the constant-speed assumption, group-averaged travel times, and using the arithmetic mean speed (as opposed to the harmonic mean speed) to estimate average travel time. Arithmetic mean speeds consistently underestimate aggregate delay, although estimating a harmonic mean speed from the arithmetic mean speed and speed variance can partially mitigate this effect. Temporal aggregation also affects the identification of traffic state transitions times, the estimation of shockwave speed and shockwave travel times, and the construction of fundamental diagrams. The results of this research will help increase understanding of the ability of ITS data to describe transportation systems, and improve forthcoming sustainability performance measures in the Portland Oregon Regional Transportation Archive Listing data archive at Portland State University.

Because of various historical, technical, and managerial limitations, most ITS data are measured at fixed points and aggregated over predetermined, arbitrary time intervals. The amount of information lost in the temporal aggregation process depends on the interval duration, summary statistics, and methodology of aggregation. With fixed durations and arbitrary start times, aggregation intervals are numb to the quantity, quality, or specific features of traffic flow.

ITS data are initially aggregated during the collection and communication stages, commonly to 20-, 30-, or 60-s intervals. The data can then be aggregated for storage in a data archive. This aggregation, usually anywhere from 1 to 60 min, is motivated by perceived or real limitations in storage capacity or data management for large volumes of disaggregate data.

The output of aggregation is a set of summary measures, often a simple arithmetic mean for each parameter (such as speed or vehicle length) and a count of the number of vehicles. Aggregate data archives can be more robust, however, storing other descriptive statistics such as variance, median, or harmonic mean of traffic parameters. While the exact parameter distribution is still lost, these extra statistics add further detail while avoiding the storage requirements of fully disaggregate data. Another important concern is the detection and handling of missing or erroneous data to avoid masking infidelities or accumulating bias.

Past research has shown an array of relationships between selected performance measures and aggregation. As a few examples, Zietsman and Rilett found variations in emissions estimates of up to 20% for different spatial and temporal aggregations and data sources. They also found travel time errors of up to 40% using aggregate versus disaggregate travel time data from automatic vehicle identification through a corridor. Wang and Liu found slightly better travel time estimates with 5-min aggregated loop detector data than with shorter time intervals when using harmonic mean speeds on short freeway segments. Lindveld and Thijis showed that aggregate travel time estimates from loop detectors are lower when based on arithmetic mean speed than on harmonic mean speed, and suggested a method for correction that is discussed below.

This paper’s objective is to describe the impacts of temporal aggregation on various ITS data applications. It focuses on data from urban freeways, collected from inductive dual-loop detectors and trajectory reconstruction from video imaging. The results will help illuminate the often complex relationships between data aggregation and performance measurement.
DATA SOURCES

The data for these analyses come from two urban freeways with different collection tools. For the first data source, inductive dual-loop detectors collected vehicle speeds on the M4 Freeway in London in November 1998. Five weeks of 24-h data from 3.5 km (2.2 mi) of roadway contains discrete speeds and times for each vehicle passing each loop detector. This data set includes eight successive detector stations (Figure 1a), approximately evenly spaced at 0.5 km (0.3 mi). For analysis, data from 25 weekdays were aggregated to 11 different time intervals: 10, 20, and 30 s, 1, 2, 5, 10, 15, 30, and 45 min, and 1 h.

The second source is the Next Generation Simulation (NGSIM) vehicle trajectory data set from Eastbound I-80 near Emeryville in the San Francisco Bay area (b). Vehicle trajectories were re-created using video image detection at 10 frames per second for three 15-min periods beginning at 16:00, 17:00, and 17:15 on April 13, 2005. Figure 1b shows the detection area of 1,650 ft of roadway over six freeway lanes, including a high-occupancy vehicle lane and an on-ramp. Instantaneous vehicle speeds were calculated from successive detected locations, smoothed with a weighted 1-s running average. Vehicles not detected over the full road section were excluded from travel time calculations.

Figure 2 illustrates the data from the loop detectors on the M4. Figure 2a shows the detected speeds at one station (all lanes) for a 4-h morning peak period, where each circle represents one detected vehicle. Figure 2b converts 1-min aggregated speeds to color bands at each of six stations over the same 4-h period. The speed–color scale is shown at the bottom center of the figure.

Figure 2c illustrates the approximation of constant vehicle speeds over a road section between detectors, where the spot speeds from the bands in Figure 2b are extended up to the next detector station. Figure 2d shows a different approximation, where roadway speeds are assumed to vary linearly between detector stations. This contoured speed plot also uses dashed lines to show the backward and forward shockwaves at the edges of congestion, discussed below.

Figure 3a shows the vehicle trajectories from I-80 for 15 min in one lane. Figure 3b overlays a speed–color grid using space and time aggregation of the trajectory/speed data. Each cell in Figure 3b shows the average of speeds over 1 min along 300 ft of roadway. Figure 3c illustrates spot speeds using 1-min aggregation every 300 ft, simulating point-based detectors. Finally, Figure 3d projects the sampled speeds from Figure 3c along the roadway up to the next sample location, illustrating the endpoint method of travel time estimation.

Lower overall speeds in Figure 3b compared with Figure 3d, seen in the color differences, reveals the effects of spatial speed aggregation versus speed sampling. With speed data recorded every tenth of a second, slower vehicles are more represented in spatial averaging than sampling, in which each vehicle contributes one speed datum to the interval average.

ANALYSIS

Speed Distribution

Some data applications are affected when temporal aggregation discards information about the distribution of speeds within an interval. Analyzing measured speeds from both freeways shows that the spread of the speed distribution decreases at higher base aggregation levels. Figure 4a illustrates the narrowing speed distributions at one location on the M4 for 2 different hours using boxplots. Using aggregated base data consolidates vehicles around central speeds, effectively “averaging out” the higher and lower speed vehicles for both congested and uncongested periods.

Figure 4b illustrates the same effect with the I-80 data from the first 15-min period. The boxplot uses a single arithmetic mean speed measurement for each 5-ft section of roadway over different time intervals, and the histogram assigns the average speed to each vehicle in the interval for direct comparison between aggregation levels. As with the M4 data, while the center of the speed distribution stays roughly constant, more aggregation reduces the spread of speeds and eliminates extreme values.

Higher levels of aggregation in the base data produce speed spreads that are consistently narrower throughout the day. For the M4 data, on average the hourly standard deviations of speed were 12%, 25%, and 43% lower from 20-s, 1-min, and 5-min base aggregated data, respectively, than from disaggregate data.

This narrowing speed distribution is important for emissions estimates because emissions rates typically have a nonlinear relationship with speed, where high and low speeds have greater emission intensity (4). Fuel consumption has a similar relationship with speed, where high and low speeds result in greater fuel consumption per
FIGURE 2 M4 data from November 16, 1998, at six detector stations, all lanes.

FIGURE 3 Trajectory data from I-80, 17:00–17:15, Lane 5.
mile of travel. Consolidating speeds will result in underestimation of these performance measures when using aggregated base data, as compared with using disaggregate data. More generally, any measure with a nonlinear relationship to speed is distorted by time-aggregated base data. For example, safety assessment can be complicated by data aggregation because of the importance of speed to passenger safety.

Travel Time

A common method to estimate travel time on a road section is to extrapolate from a spot speed, assuming constant vehicle speeds over the section that are uniform for all vehicles in a time interval. This method of travel time estimation can lead to several distinct errors, illustrated in Figure 5:

1. Sampling error (I). This error comes not from aggregation but from assuming constant vehicle speeds over the road segment. It is calculated as the difference between a vehicle’s true travel time and the travel time calculated from its single sampled speed.

2. Grouped speed error (II). This error results from applying a single travel time value to a group of vehicles that actually experiences a range of travel times. Ignoring the sampling error by assuming vehicle speeds are constant along the road section, the grouped speed error is the difference between individual travel time estimates and the group-averaged travel time estimates, both from spot speeds.

3. Arithmetic mean speed error (III). This error is the difference between the group-averaged travel time estimates and the group travel time estimate calculated from the arithmetic mean speed. The average of individual travel time estimates is accurately calculated from the harmonic mean speed. Since arithmetic mean speed always exceeds or equals harmonic mean speed, group travel time estimates from arithmetic mean speeds will yield lower values than those from harmonic mean speeds. Figure 5c illustrates the source for this error using 1-min aggregated data from 1 week at all stations on the M4. Assuming constant speeds over a road segment, the harmonic mean of spot speeds is the equivalent of calculating a space mean speed (SMS) for vehicles in a time interval, while the arithmetic mean is the time mean speed (TMS).
These three types of travel time estimation errors are not independent, and can have various inter-effects. They are also artifacts of the spot-speed method of travel time estimation, considered here because of its prevalence of deployment. Other methods of travel time estimation, such as probe vehicle detection through a corridor, will produce a distinct set of errors.

Diverse methods exist for quantifying error in travel time estimation and prediction. Measures of the average error can identify biases in travel time estimation, while measures of the absolute error reflect the size of errors without differentiating between over- and underestimation. Mean absolute percent error (MAPE), used here, shows the average amount of error with respect to total travel time, though it does not account for the natural variability in travel times.

Figure 6a illustrates the magnitudes of the three sources of error on I-80: sampling error (I), grouped speed error (II), and arithmetic mean speed error (III). Travel times were estimated and compared over a 1,500-ft roadway section using sampled speeds at the beginning of the segment and true travel times. To put this into context, free-flow travel time (at 60 mph) across this section would be 17 s, and a 30% error would result in a computed range of 12 to 22 s. The sampling error is independent of the interval width and dominates for short aggregations. Errors II and III increase with aggregation width as more vehicles (and a wider range of speeds) are included in each interval.

Figure 6b shows the three sources of error combined, comparing true travel time to travel time estimates from arithmetic mean speeds. Interestingly, this figure shows that the second and third error sources work to somewhat offset the first, decreasing the total error as more vehicles enter during each interval. Total error decreases until the 5-min aggregation width then increases again, which agrees with past research (6). This result can be explained as more vehicles creating a lateral speed distribution that offsets the longitudinal speed distribution lost from speed sampling (Sampling Error). This boon is overshadowed, however, as very long intervals include more vehicles in different traffic states.

Figure 6c shows the magnitudes of Errors I and II for 60-s aggregations at different freeway segment lengths. The grouped speed error (II) remains relatively constant at different spacings because it is independent of segment length (as is the calculation of Error III). Sampled Speed Error (I), however, increases with longer segments because the individual vehicles’ longitudinal speed variability increases. Even for very short segments (300 ft), Error I exceeds Error II.

Table 1 shows average and standard deviation of travel time errors (in seconds) for the I-80 data at nine aggregation widths, between 1 to 900 s, for all three 15-min periods. Average ground truth travel time on the 1,500-ft corridor was 91 s. Error I leads to a small estimation bias, while Error II does not. As expected, Error III introduces a sizeable underestimation bias that increases at longer aggregations.

To partially correct for the arithmetic mean speed error (III), the SMS can be estimated from the TMS and speed variance, if known.
Equation 1 calculates SMS, $\bar{v}_s$, from the TMS, $\bar{v}_t$, and the speed variance, $\sigma_t^2$.\(^\text{(1)}\)

$$\bar{v}_s = \frac{\bar{v}_t + \sqrt{\bar{v}_t^2 - 4 \times \sigma_t^2}}{2}$$ \(^\text{(1)}\)

This equation applies absolutely if vehicle speed distributions are symmetrical in both time and space. If speed distributions are not symmetrical, it is an approximation.

This approach relies on knowing the speed variance for the interval. When speed variance or standard deviation is not recorded, future work will attempt to estimate an appropriate speed variance from other, known parameters.

Equation 1 can be rearranged to find the ratio of space and time mean speeds:

$$\frac{\bar{v}_s}{\bar{v}_t} = \frac{1}{2} + \sqrt{\frac{1}{4} - CV^2}$$ \(^\text{(2)}\)

where $CV$ is the coefficient of variation, defined as

$$CV = \frac{\sigma_t}{\bar{v}_t}$$

The defined range for $CV$ in Equation 2 is 0 to 0.5. An arbitrary maximum value for the $CV$ of 0.45 captures 99.7% of measured $CV$s for a sample of 1 day’s data on the M4 at all tested aggregation widths.
TABLE 1  Travel Time Estimation Errors from 1,500-ft Section of I-80

<table>
<thead>
<tr>
<th>Lane 3 Agg. (s)</th>
<th>TTi_tr-TT_i_est (I)</th>
<th>TT_i_est-TT_sms (II)</th>
<th>TT_sms-TT_tms (III)</th>
<th>TT_i_tr-TT_sms (I) + (II)</th>
<th>TT_i_tr-TT_tms (I) + (II) + (III)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg.</td>
<td>SD</td>
<td>Avg.</td>
<td>SD</td>
<td>Avg.</td>
</tr>
<tr>
<td>1</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>1.24</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>1.43</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>16.62</td>
<td>1.91</td>
</tr>
<tr>
<td>20</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>24.44</td>
<td>3.65</td>
</tr>
<tr>
<td>30</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>27.55</td>
<td>5.00</td>
</tr>
<tr>
<td>60</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>36.05</td>
<td>8.11</td>
</tr>
<tr>
<td>180</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>44.46</td>
<td>11.31</td>
</tr>
<tr>
<td>300</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>45.66</td>
<td>12.16</td>
</tr>
<tr>
<td>900</td>
<td>5.46</td>
<td>49.05</td>
<td>0.00</td>
<td>54.55</td>
<td>16.37</td>
</tr>
</tbody>
</table>

Note: 91-s average travel time.

Smaller CV values result in small differences between TMS and SMS (a ratio close to 1), while greater values result in increasing differences, up to a maximum of 50% (a ratio of 0.5). Congested periods with low mean speeds tend to have high CVs and thus greater differences between SMS and TMS. These high CVs during congestion, when most delay occurs, lead to high impacts of Error III on delay calculations.

Applied to the M4 speed data, Equation 2 performs well in estimating SMS using the measured TMS and speed variance. Nearly all values are near the true SMS for short aggregations, and at the 5-min aggregation width still 94% of SMS estimates are within 1% of the true value. Accuracy of the equation decreases at longer aggregations as vehicles in different traffic states lead to asymmetrical speed distributions.

**Delay**

Delay is a performance measure often calculated from travel time estimates by subtracting a free-flow travel time. Error in aggregate vehicle delay, when defined as the difference between aggregate delay estimates and disaggregate delay estimates, comes from two sources.

The larger source of error is the arithmetic mean speed error described above for travel time estimates; when aggregate travel time is underestimated the corresponding delay is also underestimated. As with travel time estimates, the arithmetic mean speed error increases at longer aggregations. Since delay is a cumulative measure, the Sampled Speed and Grouped Speed travel time errors have little effect because they don’t produce large travel time estimation biases.

The second source of aggregate estimated delay error from temporal aggregation is due to lost speed distribution. When an interval’s average speed is slightly above the assumed free-flow speed (FFS), delay experienced by vehicles in the interval traveling below the FFS is neglected, underestimating delay in comparison with the disaggregate estimate. This error will occur most often when the mean speed is near the FFS—the least critical time periods of the day from a delay perspective—so this source of error is less critical than arithmetic mean speed error (only leading to 1% to 3% aggregate delay underestimation on the M4).

One method that can be used to compensate for the averaging method error is to estimate the time interval’s SMS using Equation 2. Figure 7 shows the resulting delay estimates using TMS (arithmetic mean speed), true SMS (harmonic mean speed), and estimated SMS (from Equation 2) for 1 week of data at all stations on the M4. For
the purposes of this analysis, a free-flow speed was arbitrarily fixed at 60 mph. The disaggregate delay is indicated by the dashed line at the top. While the delay estimates from both TMS and estimated SMS fall short at higher aggregation levels, the estimated SMS is consistently closer to the disaggregate value.

**Traffic State Transitions**

To look at the effects of temporal aggregation on the identification of shockwaves, data from the M4 and I-80 were analyzed to identify traffic state transitions. Traffic state transitions on the M4 were identified as the local maxima and minima on oblique cumulative speed plots created with disaggregate and aggregate base data, as described by Cassidy and Windover (9) and Cassidy and Bertini (10). All lanes were included at each detector station, and two transition times were identified each day, at the start and end points of the morning congestion shown in Figure 2.

This method of transition time identification produced adequate aggregate results for the M4 data set at the start of morning congestion. For backward-moving shockwaves at the onset of congestion, five detector stations for 25 weekdays and eight different aggregation intervals (from 10 s up to 15 min) produced 1,000 uncongested-to-congested transition times. Compared with disaggregate transition times, 5% of aggregate transition times were outside of the range of one aggregation interval, and all but one of these were for aggregation intervals of 1 min or less. The accuracy errors occurred when the oblique cumulative speed plot did not display a sharp peak. These errors are tied to this method of transition time identification, and different methods will likely have different accuracy results.

In general, the congested-to-uncongested transitions were less abrupt, and more sensitive to accuracy errors using aggregated data. For these transitions, 18% of the aggregate transition times were more than one aggregation interval away from the disaggregate transition time. Again, the short aggregation intervals (1 min or less) experienced most of these errors. For the shorter intervals alone (10-, 20-, 30-, and 60-s aggregations), 10% of transition times were in error for backward-moving waves and 32% were in error for forward-moving waves.

Errors in the identification of transition times can lead to errors in the estimation of shockwave speeds or shockwave travel times. Another potential error for shockwave estimation, though, is the low temporal resolution resulting from long aggregation intervals. The passage of a hypothetical shock is illustrated in Figure 8a, showing disaggregate transition times as red circles, and accurate aggregate transition times as green boxes. Assuming constant shockwave speed, the shocks can be drawn as straight lines connecting transition times at each detector station. Figure 2d illustrates this with dashed black lines at the beginning and end of the congested period.

Because aggregate transition times encompass a range of values (the aggregation interval), uncertainty exists in the estimated shock speed. As shown in Figure 8a, the estimated shockwave speed can vary from \( v_{\text{min}} \) to \( v_{\text{max}} \). The size of the range of possible shock speeds will depend on the aggregation width and the distance between detectors. Longer detector spacing demands less resolution from time intervals for the same shockwave speed uncertainty. Aggregate transition time identification errors could lead to red circles outside of the green boxes (in other aggregation intervals), resulting in a shifting of the estimated shock and a distortion of shockwave speed and shockwave travel time estimates.

The location of a transition within an aggregation interval will depend on the aggregation start time, which is arbitrary. On the scale of short aggregation intervals (say, 5 min and under), the transition time itself on a given day at a given location is also arbitrary. Given enough data points, the transition times within a short aggregation interval will be approximately uniformly distributed across the interval.

For the NGSIM data, shockwaves were manually identified using a time–space diagram, as illustrated in Figure 8b. Disaggregate transitions were identified as abrupt speed changes, or kinks in the vehicle trajectories. The disaggregate shock speed corresponds to the slope of the trend line of the transition points for each vehicle intersecting the shock wave. Aggregate traffic state transitions were identified with colored cells, similar to Figure 3b, using 5-ft freeway sections, and aggregate shockwaves were the trendlines connecting the transition cells.

Figure 9 shows the accuracy of back-moving shockwave speed and shockwave travel time estimation for both freeways. Figure 9a and b use the M4 data over 25 weekdays at two different detector spacings (0.5 km and 2.0 km), excluding “infinite” shock speeds (simultaneous aggregate transition time intervals). While the average shock speeds are similar at increasing aggregation intervals, the error in individual estimates increases, as seen in Figure 9a.

![Figure 8](image1.png)

**FIGURE 8** Illustration of transition shock speed estimation with (a) loops and (b) trajectories.
The MAPE in shockwave travel time estimates increases at longer intervals and for shorter spacing. At 0.5-km detector spacing, shockwave travel times are about 200 to 300 s, so for aggregation intervals wider than 300 s, shockwave travel time estimates are either 0 or equal to the aggregation width. The increased inaccuracy in transition times for forward-moving waves, described above, yielded greater errors in estimated shockwave speeds and travel times, with MAPE values approximately two to three times those for backward-moving shockwaves.

For the I-80 data, Figure 9c and d show the shockwave speed and shockwave travel time errors, respectively, for 10 analyzed shockwaves. Both illustrate increasing error at longer aggregations. The backward-moving shocks on I-80 were short perturbations in congested flow, rather than uncongested–congested transitions as on the M4. The shockwaves covered a shorter distance of roadway, often less than 0.5 km, so the 60-s aggregation width was sufficiently wide to mask some shockwaves and produce several simultaneous transitions along the freeway.

**Fundamental Diagram**

It is very common to plot bivariate scatterplots of fundamental traffic flow parameters from aggregated freeway detector data (11).

Figure 10a shows fundamental diagrams of speed versus flow at two different aggregation widths for 1 day of data at one station on the M4, while Figure 10b, c, and d show fundamental diagrams using three different aggregation widths for data from Lane 5 at an arbitrary location on I-80. Longer aggregations tend to consolidate the data, reducing the spread in addition to the obvious reduction in the number of points. This is due to an averaging of extreme values, where short periods of high or low speeds or flows are averaged into longer intervals with more moderate sustained values. Different levels of data aggregation can change the apparent speed–flow relationship, where short intervals can capture shorter-duration events. Long aggregation intervals will represent more sustained traffic conditions, though they are also more likely to combine multiple traffic states.

**CONCLUSIONS**

Aggregating ITS data discards useful information, the amount of loss depending on the method and summary statistics of aggregation. This research explored the effects of time-aggregating data from freeways, especially with respect to performance measures.

The first aggregation effect, a distortion of speed distributions, will result in underestimation of emissions and fuel consumption esti-
mates based on traffic data. To a lesser extent, the narrowed speed distributions will also lead to lower aggregated delay estimates (as compared with disaggregate estimates).

Temporal aggregation affects travel time estimates through three types of errors: sampling error, grouped speed error, and arithmetic mean speed error. Sampling error dominates for shorter aggregation widths and increases with road segment length, though it is an artifact of spot-speed travel time estimation, not temporal aggregation. The sizes of the other two errors increase with aggregation width, though grouped speed error does not lead to a travel time estimation bias. The combined effects of the three errors produced improved travel time estimates with aggregation widths up to 5 min.

Estimating grouped travel times with arithmetic mean speeds (as opposed to harmonic mean speeds) leads to a systematic underestimation of travel time and delay. When harmonic mean speeds are not recorded, this effect can be partially mitigated by estimating the SMS from the TMS and speed variance. Aggregate delay errors are primarily the result of using arithmetic mean speeds to estimate aggregate travel time, leading to aggregate delay estimates much lower than disaggregate estimates (though estimating an SMS greatly reduces this error).

The largest errors in shockwave identification arise from low time resolution with respect to shockwave travel time. Shockwave speed inaccuracies increase with aggregation width and decrease with detector spacing. Some discrepancies in transition times result from using aggregate data to identify transition times from oblique cumulative speed plots.

Finally, similar to the consolidation of speed distributions, construction of fundamental diagrams is affected by data aggregation. In particular, extreme data points are averaged in to more moderate values at longer aggregation intervals. As with all of the applications described here, employment of ITS data requires awareness or study of the effects of any aggregation used.

To increase the accuracy and utility of ITS data, several steps can improve existing practices. Ideally, data should be collected and archived in their most disaggregate form to avoid limiting possible uses. Costs are associated with this approach, however, in the form of increased data management, storage, and communication capacity. Where data must be aggregated, retaining several summary statistics will decrease the amount of information lost to the aggregation process. Of particular importance for performance measures are harmonic mean and variance. Attention to missing or erroneous data when aggregating, both in the method of aggregation and in recording the fidelity, can also improve the quality and value of ITS data for current and future applications.

This research will help improve forthcoming sustainability performance measures to be added to the Portland Oregon Regional Transportation Archive Listing data archive at Portland State Uni-
versity, which receives and archives loop detector data with a base aggregation of 20 s.

ACKNOWLEDGMENTS

The authors acknowledge Stuart Beale of the Executive Agency of the Department for Transport, United Kingdom; Tim Rees of the Transport Research Laboratory, United Kingdom; and FHWA’s NGSIM project for generously supplying the data used here. A portion of this work was funded by the Department of Civil and Environmental Engineering at Portland State University, the Oregon Transportation Research and Education Consortium, and École Nationale des Travaux Publics de l’État.

REFERENCES


