Abstract—Accurate freeway travel time estimates are critical for transportation management and traveler information—both infrastructure-based and in-vehicle. Infrastructure managers are interested in estimating optimal freeway sensor density for new construction and retrofits. This paper describes a concept developed from first principles of traffic flow for establishing optimal sensor density based on the magnitude of under- and over-prediction of travel time during shock passages when using the midpoint method. A suggested aggregate measure developed from vehicle hours traveled (VHT) is described for a reasonable range of detector densities. Extensions of the method to account for both recurring and nonrecurring congestion are included. Finally some suggestions for future research are described.

I. INTRODUCTION

Accurate travel time estimation is important for transportation management and information. In the U.S., freeways account for 3% of the national highway mileage, but accommodate more than 30% of the vehicle miles traveled (VMT). The alleviation of congestion on urban freeways is receiving heightened attention, and transportation agencies are applying improved management strategies to reduce congestion and improve travel time reliability. Incident management and traveler information systems can be implemented at relatively low cost. However, such systems rely upon accurate measurement of traffic parameters such as flow, speed, travel time, and delay. Usually these data are measured by fixed sensors (loop detectors, video cameras, radar sensors, etc.) or by mobile data sources such as automatic vehicle identification (AVI) toll tags or automatic vehicle location (AVL) probe vehicles. In order to answer the question “how much detection is needed,” this paper focuses on the common use of fixed sensors (such as loop detectors) as a basis for formulating an optimal detector placement strategy.

Fig 1 illustrates a hypothetical space-time (x-t) plane of length ℓ and time interval \( t_1 \). A set of vehicle trajectories is shown (in grey) with that of vehicle \( i \) being highlighted in black. If equipped with AVL or any other data logging system, vehicle \( i \)'s trajectory could be plotted and all information necessary to completely describe its path would be known, including its actual travel time and its speed (slope of trajectory) at any point. If one assumes a free flow speed, the free flow travel time can be computed and the delay (actual minus free flow travel time) for vehicle \( i \) is known. In practice it is more common to use sensors at fixed points (such as point \( x_1 \)) to measure speed and subsequently extrapolate that speed.
over a segment. In Fig 1 one can observe how a speed measured at \( x_1 \) is extrapolated over a segment of length \( l \) resulting in the calculation of the extrapolated travel time. As shown, the estimate does not perfectly match the actual travel time. The magnitude of this difference, as a function of detector density, is the topic of interest in this paper. Ignoring sensing errors, the passage of a shock represents the “worst case scenario” for travel time prediction, so the methods presented in this paper can be interpreted as a form of robust decision analysis about sensor spacing.

II. TRAVEL TIME ESTIMATION FRAMEWORK

A fundamental traffic flow relation is assumed, as shown in Fig 2. Congested state C (flow \( q_c \) and speed \( v_c \)) and uncongested states A, B, and D (flows \( q_A, q_B, q_D \), speed \( v_f \)) are shown. Below the flow-density \( (q-k) \) diagram is an \( x-t \) plane showing a bottleneck (either recurrent or nonrecurrent) at location \( b_n \). Following the rules of macroscopic traffic dynamics, there is a transition between uncongested state A and congested state C, defined by a shock of velocity (slope on diagram) \( v_{AC} \). Fig 2 shows that for an arbitrary highway segment (separated by the two dashed lines) transition AC is bounded by a rectangle as the shock passes. Upon bottleneck deactivation at time \( t_{deact} \), transition CD occurs, from congested state C to uncongested state D, marked by a backward-moving recovery wave of velocity \( v_{CD} \). Transition DA is separated by a forward-moving recovery wave of velocity \( v_f \).

Fig 3 shows an \( x-t \) plane for an actual freeway corridor, with the \( y \)-axis as distance and the \( x \)-axis as time (4:00 to 20:00). The figure shows speed as a “color,” for a 23-mile corridor (northbound I-5 in Portland, Oregon) on Feb 8, 2007. Mileposts are shown on the left \( y \)-axis, and the locations of five variable message signs (VMS) are overlaid with their milepost locations shown on the right \( y \)-axis. Downtown Portland is located at approximately milepost 300. There are more than 500 lane and ramp sensors on the Portland area freeways, at about 138 locations, with an average sensor spacing of 1.24 miles. Nearly all of Portland’s sensors are located just upstream of on-ramps, as part of the region’s ramp metering system. The Oregon Department of Transportation (ODOT) displays travel time information to Downtown on VMS 1, 2 and 3 during the AM peak period. ODOT and others are interested in increasing sensor density toward some “optimal” value, in order to improve the accuracy of travel time information in a cost-effective manner.

For the section between milepost 294 and 295, one can observe state A (uncongested) from 4:00 to 7:00. This is followed by transition AC (uncongested to congested) marked by the upstream propagation of a queue through the section. A line with slope \( v_{AC} \) has been superimposed on the figure. Next, the section is in state C (congested) until after 8:30. This is followed by transition CD (congested to uncongested), marked by the passage of a backward moving recovery wave. A line with slope \( v_{CD} \) has been superimposed on the figure. Traffic remains in state D until after 9:00 when a forward moving wave (at speed \( v_f \)) marking transition DA passes through the section. The remainder of the day was marked by uncongested conditions in the section (e.g. state A). This pattern can be typical for both recurrent and nonrecurrent congestion. In the case of the 1 mi section mentioned here, these transitions only occurred once. However, looking further downstream at another next section (for example between milepost 299 and 300), one can observe that there were two such transitions on this particular day.
III. CALCULATING TRAVEL TIMES DURING STATIONARY CONDITIONS

Fig 4 illustrates that travel times during regimes A and D (fully uncongested), C (fully congested), and during transition DA (uncongested) can be estimated by:

\[ t_{f} = \frac{\ell}{v_f}, \quad t_c = \frac{\ell}{v_c} \]

where \( t_f \) is the free flow travel time and \( t_c \) is the congested travel time. If the sensor where speed is measured is located at the center of the segment, this is called the midpoint method. Travel time estimation over a segment can be performed accurately if the traffic state within the segment is either fully uncongested or fully congested. In addition, within the segment is either fully uncongested or fully congested. In addition, \( t_c \) is an upper bound on actual travel time through the segment and \( t_f \) is a lower bound on the actual travel time. Empirical studies have shown that travel time estimation during stationary conditions (either uncongested or congested) is usually quite accurate [1,2]. For simplicity, in this paper, the midpoint method will be used; however, other methods have been developed for improving travel time estimates [3,4] and can be studied further later.

IV. TRAVEL TIME ESTIMATION DURING TRANSITIONS

As initially illustrated in Fig 3, there are two basic transition types in freeway traffic flow: uncongested to congested (e.g. AC) and the reverse (e.g. CD). These transitions can occur multiple times in a given section as queues propagate and dissipate and sometimes combine with one another. Queues may be caused by recurrent bottlenecks or by incidents. Travel time estimation based on current measured conditions can result in both overprediction and underprediction. Underprediction will be considered to have a higher weight than overprediction, since travelers whose travel times are much longer than predicted at their entry to a segment will be more likely to be dissatisfied. A hypothetical segment of length \( \ell = 1 \) mi will be considered for comparison purposes. For the sake of numerical examples a range of five sensor spacings \( s \) (0.1 to 1 mi) will be used and assumed traffic flow parameters will be: \( q_d = 2000 \) vph, \( q_c = 1800 \) vph, \( v_f = 60 \) mph, \( v_c = 30 \) mph, \( v_{CD} = -17.1 \) mph, and \( v_{AC} = -7.5 \) mph.

A. Underpredicting Travel Time During Regime AC

Fig 5 illustrates transition AC from uncongested conditions with vehicles traveling at \( v_f \) to a congested state with vehicles traveling at \( v_c \). The figure illustrates a backward moving shock passing through the segment of length \( \ell = 1 \) mi, and sensor spacing \( s \), at a speed \( v_{AC} \). Vehicle \( j_1 \) is the last vehicle to pass through the section entirely at \( v_f \) and vehicle \( j_3 \) is the first vehicle to pass through the section entirely at \( v_c \). All vehicles between \( j_1 \) and \( j_3 \) travel at average speeds between \( v_f \) and \( v_c \). The figure also shows that the sensor continues to record speed \( v_f \) until time \( t_c \), a lag time \( \alpha = -s/2v_{AC} \) after the shock enters the section. After vehicle \( j_1 \) and until \( t_c \), drivers expect a free flow trip time through the entire section while their actual trip time will be higher. For example, driver \( j_2 \) enters the section an instant before time \( t_c \) and expects a speed of \( v_f \) through the entire section, but experiences a longer actual travel time \( z \). For vehicles after \( j_f \) entering the section before time \( t_c \), travel time is underpredicted by an amount equal to the difference between the expected travel time (dashed trajectory in the figure) and the actual travel time (solid trajectory). Vehicle \( j_2 \) experiences a travel time of \( z \), and the maximum travel time underprediction (\( u_{max} \)) of:

\[ u_{max} = z - t_f = \frac{\ell(v_c - v_{AC}) + \frac{1}{2} s(v_f - v_c)}{v_c(v_f - v_{AC})} - t_c \]

For example, if \( \ell = s = 1 \), the driver of vehicle \( j_2 \) expects a travel time of 1 min, but actually experiences a 1.56 min trip, which is more than 50% longer than expected. If one assigns zero weight to any travel time overprediction (between time \( t_f \) and the time the shock reaches the upstream end of the section), for now one can neglect any flow entering the section after time \( t_c \). The remainder of this section only considers this scenario, since it is assumed that traffic management officials want to avoid giving drivers false expectations of low travel times, when they actually experience longer ones. One can quantify the predicted...
and actual vehicle-hours traveled ($VHT_{\text{pred}}$ and $VHT_{\text{act}}$) solely for vehicles experiencing underprediction (superscript $u$) over the hypothetical segment:

$$VHT_{\text{pred}}^u = \frac{q_A}{v_f} \left( \frac{\ell}{v_f} - s \right)$$

$$VHT_{\text{act}}^u = \frac{q_A}{2} \left( \frac{\ell}{v_f} - s \right)$$

For $\ell = s = 1$, the predicted VHT/mile is 2.78 veh-hr/mile, yet the actual VHT/mile is 3.55 veh-hr/mile, a 22% error over the collection of vehicles entering transition AC before $t_c$. Vehicles experience actual underpredictions between 0 and 0.56 min.

In order to extend these calculations to arrangements with greater sensor density, Fig 6 illustrates how predicted and actual VHT will change with increased sensor density. The increased sensor placement reduces the lag time $\alpha$ so vehicles entering the segment receive the congestion message sooner. This reduces the magnitude of the VHT composed of travel time underprediction, and the figure indicates this using the darker shaded areas. One can see that as the lag time decreases, the VHT of traffic impacted by travel time underprediction will decrease in this 1 mi section. For now the issue of overprediction is set aside.

Table 1 shows the values of $u_{\text{max}}$ and lag time $\alpha$ for regime AC as a function of detector spacing $s$. Table 1 also shows the predicted and actual VHT/mile for a range of $s$. To reiterate, For this hypothetical 1 mi segment, vehicles entering the section prior to $t_c$ will expect free flow conditions, but will instead experience progressively longer travel times (the lag time gets shorter with increased detector density). For this set of vehicles, the actual VHT is higher than the predicted VHT. Even with sensors at 0.1 mi spacing, the VHT error is 7% and vehicles expecting a 1 min travel time can actually experience 1.16 min, a 16% underprediction. The gap between the predicted and actual VHT grows with larger sensor spacing. For the range of sensor spacing considered, the VHT error falls between 7% and 22%. For the average ODOT sensor spacing of 1.2 mi, the VHT error is 24%.

Here, more detection is better if the goal is to minimize underprediction. Spacing between 0.25 and 0.50 mi would keep $u_{\text{max}}$ below 33% of $t_c$. Table 1 also includes a sixth row with $s = 0$, which is equivalent to some form of ubiquitous sensor coverage.

### Table 1: Segment Travel Time Statistics During Regimes AC and CD

<table>
<thead>
<tr>
<th>Regime</th>
<th>Underprediction</th>
<th>Overprediction</th>
<th>AC Total</th>
<th>Underprediction</th>
<th>Overprediction</th>
<th>CD Total</th>
<th>Regimes AC and CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$ (mi)</td>
<td>Lag</td>
<td>Time</td>
<td>mile</td>
<td>%</td>
<td>VHT/</td>
<td>VHT/</td>
<td>%</td>
</tr>
<tr>
<td>0.00</td>
<td>0.56</td>
<td>4.00</td>
<td>2.78</td>
<td>3.55</td>
<td>22%</td>
<td>4.44</td>
<td>3.95</td>
</tr>
<tr>
<td>0.10</td>
<td>0.16</td>
<td>0.40</td>
<td>0.78</td>
<td>0.84</td>
<td>5%</td>
<td>8.88</td>
<td>6.91</td>
</tr>
<tr>
<td>0.25</td>
<td>0.22</td>
<td>1.00</td>
<td>1.11</td>
<td>1.23</td>
<td>10%</td>
<td>7.77</td>
<td>6.27</td>
</tr>
<tr>
<td>0.50</td>
<td>0.33</td>
<td>1.33</td>
<td>1.30</td>
<td>1.46</td>
<td>11%</td>
<td>7.40</td>
<td>6.04</td>
</tr>
</tbody>
</table>

Note: Positive error percentages indicate underprediction, while negative error percentages indicate overprediction.
actual VHT is 3.95 veh-hr, an overprediction of 13%. Most drivers would be pleasantly surprised by the shorter travel time. With increased sensor density, the percent error in VHT increases to a maximum of 27% overprediction for \( s = 0.10 \).

If one assumes that under- and overpredicted VHT can be added together numerically (with one canceling out the other), the total predicted VHT in regime AC would be \( VHT_{\text{pred}} + VHT_{\text{act}} \), and the actual VHT in regime AC would be \( VHT_{\text{act}} + VHT_{\text{act}} \) (7.50 veh-hr). Here one time unit of underprediction is canceled out by a time unit overprediction. As shown in Table 1, when the two components are added, for \( \ell = s = 1 \), the effect is still a 4% underprediction in VHT. For decreasing values of \( s \), the aggregate effect is overprediction, up to 23% with \( s = 0.10 \). Fig 7 illustrates this graphically, where the optimal sensor spacing might be close to 0.8 mi, which is where the VHT error line crosses from overall underprediction to overprediction. For the situation where drivers might assign a higher “price” to underprediction than to overprediction, Fig 7 shows, with the line series labeled “penalty,” the percent error in VHT when the underpredicted VHT is weighted at 3\( \times \) that of overpredicted VHT. From this example, one can see that optimal sensor spacing might be close to 0.6 mi. Table 1 also includes a column where the total VHT error is considered by adding the absolute value of the differences in actual and predicted VHT (also shown in Fig 7). The absolute error ranges between 17% and 25% for the range of detector spacing considered.

When viewing Fig 7 and Table 1 together, for regime AC it can be seen that when \( s \approx 0.5 \), the underprediction is 14%, the overprediction is 20%, the aggregate error is -11% (overprediction), and the absolute error is 18%. This may be reasonable for user acceptance.

C. Predicting Travel Time During Regime CD

Fig 8 illustrates transition CD from congested state C to uncongested state D (see also Figs 2 and 3). Vehicles to the left are traveling at \( v_C \), and a backward-moving recovery wave passes through the section at speed \( v_{CD} \). The sensor receives the “uncongested” signal at time \( t_r \), which occurs a lag time \( \alpha' \) after the wave enters the section. As shown in the figure, travel time is overpredicted for vehicles entering after vehicle \( j_1 \) and before time \( t_r \), and the VHT for these vehicles can be calculated as:

\[
VHT_{\text{pred}} = \frac{q_C}{v_C} \left( \frac{\ell}{2} + \frac{s}{2} \right)
\]

\[
VHT_{\text{act}} = \frac{q_C}{2} \left( \frac{\ell}{2} + \frac{s}{2} \right) \left( \frac{v_f - v_{CD}}{v_{CD} v_f} \right)
\]

Vehicles entering after vehicle \( j_2 \) expect free flow travel times but experience higher travel times (underprediction). Predicted and actual VHT are:

\[
VHT_{\text{act}} = \frac{q_C}{v_{CD} v_f} \left( \frac{\ell}{2} + \frac{s}{2} \right) \left( \frac{v_f - v_{CD}}{v_{CD} v_f} \right)
\]

Table 1 shows the under- and overprediction results for regime CD. In this case, the percent VHT error for underprediction increases with increased detection. Fig 9 shows that when the errors are simply added (allowing overprediction to cancel out underprediction) an optimal \( s \) would be about 0.4 mi. Applying a 3\( \times \) penalty to underprediction results in an optimal \( s \) of about 0.8 mi. Adding the absolute values of the under- and overprediction results in errors in the 17-19% range.

D. Combined Effects of Transition Regimes AC and CD

Fig 3 shows a freeway section between mileposts 294 and 295, where it is known that freeway travel time estimation can be performed accurately using the midpoint method throughout the entire day except during regimes
AC and CD. It is possible to consider just the impact of VHT under-prediction during the transition regimes. Fig 10 shows the additive effects of only underpredicted VMT (data from regimes AC and CD in Table 1). The optimal s when considering only underprediction is 0.5 mi (minimum is 16.5% error).

Table 1 also shows the overall additive effects of the underprediction and overprediction that occurs in regimes AC and CD. As shown for $\ell = s = 1$, the aggregate effect is a 2% VHT overprediction error. This error increases to 12% VHT overprediction for $s = 0.10$. This is also shown graphically in Fig 11. Since drivers may assign a higher value to over-prediction, a $3\times$ penalty is applied to the underprediction error before it is added to the over-prediction error, resulting in a net 14% underprediction for $s = 1$ mi, and a 3% overprediction for $s = 0.1$ mi. As shown in Fig 11, this reveals an optimal $s$ of 0.33 mi (zero error). It is not clear how differently drivers actually value underprediction versus overprediction, so this is just a sample, and could be the topic of further research. The total error applying the sum of the absolute value of the underprediction and overprediction error is also shown in Table 1 (18-22% error) and in Fig 11.

V. CONCLUSIONS

This paper addressed the question: “how much detection do you need?” toward accurate estimation of freeway travel time using the midpoint method. It is false to assume that detection decisions are made in isolation of other issues, and in fact, sensors are usually placed to enable operation of ramp metering (e.g. Portland) and traffic monitoring (e.g., counting, speed maps, and incident detection). Freeway travel time estimation is often a useful side benefit that can be leveraged from an existing sensor network. It is also false to assume that in the future some combination of fixed infrastructure based sensors and vehicle based sensing (e.g. AVL) will not provide answers and improvements. However, this analysis has taken the question of sensor density in some degree of isolation which has resulted in some helpful outcomes. Another issue that has been left for further research is the question of where to optimally place sensors, beyond simply a question of spacing. The use of other travel time algorithms beyond the midpoint method should also be explored further. The optimal placement of sensors in relation to known bottlenecks, and high incident locations will be examined in the future.

ACKNOWLEDGEMENT

Galen McGill of ODOT posed the sensor density question and supports our freeway travel time research. Dennis Mitchell and Jack Marchant of ODOT provide the data. K. Tufte, S. Kothuri, B. Zielke and R. Fernandez of PSU assisted in development of this work. D. Lovell also helped in the attempt to think in the space time plane.

REFERENCES