Jam Occupancy and Other Lingering Problems with Empirical Fundamental Relationships

Benjamin Coifman, PhD

Associate Professor
The Ohio State University
Joint appointment with the Department of Civil, Environmental, and Geodetic Engineering, and the Department of Electrical and Computer Engineering

Hitchcock Hall 470
2070 Neil Ave, Columbus, OH 43210
Phone: (614) 292-4282
E-mail: Coifman.1@OSU.edu

Abstract

This paper explores several commonly overlooked factors impacting empirical fundamental relationships that are commonly used to relate the traffic state parameters: speed, flow and concentration. Most of these factors are conceptually simple, but collectively they result in surprisingly large, non-linear distortions of the empirical traffic state measurements. In some cases the impacts are known but underappreciated, e.g., passenger car equivalents and measurement errors arising from sampling artifacts. In other cases the impacts have not been recognized in the literature, e.g., we find that jam occupancy should be about 80%. We also discuss often-overlooked impacts from an inhomogeneous vehicle fleet and non-stationary traffic, both of which can add considerable noise to empirical measurements of the traffic state.

As a result of these distortions, on a freeway we find that over half of the physical distance along the queued regime of the fundamental relationship in the flow-density plane (and flow-occupancy plane) come from speeds below 10 mph. This fact inadvertently gives greater weight to the low-speed measurements because they are spread over a large physical region of the plane while the higher speed measurements are compressed into a narrow sliver. Unfortunately, as we illustrate, the low-speed samples are subject to the largest measurement errors, are the least likely to come from stationary conditions, and they often violate the assumptions used to measure the traffic state. In short, aggregated low-speed data from conventional vehicle detectors should be discounted or possibly discarded when constructing an empirical fundamental relationship.

Introduction

Much of traffic flow theory depends on the existence of a fundamental relationship (FR) between flow, q, density, k, and space-mean speed, v. The dependence on the FR may be explicit, e.g., Lighthill, Whitham, and Richards (LWR) hydrodynamic model [1-2] or implicit, e.g., car following models [3-4]. The FR is commonly characterized in terms of a bivariate relationship between two of the three parameters (in each case the third parameter is determined by the fundamental equation, Equation 1). In our discussion, we will often refer to the flow-density curve, qkFR, one of the three commonly used bivariate realizations of the FR.

$$q = k \cdot v \tag{1}$$

While debate continues about the shape of the FR [5-6], all of the empirically generated FRs use data that average conditions over time and/or space to calculate the traffic state, (q, k, v). Many researchers attribute the lingering ambiguity to the unintentional use of non-stationary traffic conditions [7-9]. It is difficult to measure k directly, so most empirical FR studies commonly use one of two approaches. The first approach uses q and occupancy (occ) from single loop detectors, where: occ is the percentage of the sample period that the detector is occupied by vehicles; occ is proportional to k by the average effective vehicle length, L_{eff}, where a vehicle's effective length is the sum of its physical length and the size of the detection zone. Unfortunately L_{eff} varies from sample-to-sample [10-11] and cannot be measured directly from single loop detectors. The second approach uses q and v from dual loop detectors to derive k via Equation 1. Both approaches assume stationary traffic conditions, which is rare in queued traffic. There is little in the literature about how best to work with occ as a proxy for k or how the loop detector operation can distort the observed traffic states (the most prominent exception being the discussion of using space-mean speed rather than time-mean speed [10, 12]).

This paper takes a different approach to examining the ambiguity in the shape of empirical FRs by considering the microscopic details of aggregating individual vehicle measurements into the macroscopic data commonly used to develop traffic flow theories. In the following section we explore seven commonly overlooked or underappreciated factors of the aggregation process that impact an empirical FR. Most of these factors are conceptually simple, but collectively they result in surprisingly large, non-linear, and often heteroscedastic distortions of the empirical traffic state measurements. Finally, the paper closes with discussion and conclusions.

Sources of Measurement Errors

This section uses hypothetical microscopic examples of the aggregation process to explore seven different factors that distort the resulting macroscopic measurements of the traffic state: First, we consider homogeneous vehicles sampled over a long period of stationary conditions to investigate jam occupancy. Second, we adopt a more realistic sampling period, T, and investigate the sampling errors that arise due to a non-integer number of headways in the sample. Third, we consider the vehicle fleet and see how the FR is highly dependent on L_{eff} (this point is particularly important given the fact that L_{eff} varies from sample-to-sample [10-11]). Fourth, we consider driver behavior. Fifth, we discuss several factors of non-stationary traffic

that will impact the measured traffic state, including acceleration and low-speed vehicles. Sixth, we discuss how averaging across lanes does a poor job addressing these problems. Seventh, we discuss the impact of detector errors.

Homogeneous vehicles, stationary traffic, long sampling period

The two common metrics of concentration are k and occ. Although traffic flow theory typically uses k, it is difficult to measure k directly because most vehicle detectors only monitor a single point along the roadway. So occ is often used as a proxy for density. These two measures are related, but they are not interchangeable. At the extreme: jam density, k_j, is the maximum number of vehicles per unit distance of roadway, while 100% occ is the maximum measurable occ. These two extreme measurements are not synonymous. The nuances of the extreme values of both metrics are not fully appreciated and this oversight has impacted empirical studies that rely on occ to study traffic flow.

Consider the FR in Fig 1 generated under stationary conditions from homogeneous vehicles over long sampling periods with free speed, $v_f = 65\ mph$, capacity, $q_o = 2{,}400\ vph$, and $k_j = 211\ vpm$. The right-hand side of the figure shows the qkFR, which is set to be triangular (see, e.g., [5, 13]). However, the specific shape of the qkFR is not critical to the general findings of this paper. The left-hand side of the figure shows the vqFR, transposed from the commonly used orientation to facilitate direct comparisons between these two forms of the FR. This figure employs the unrealistic assumption of stationary conditions throughout the queued regime because the literature often implicitly assumes that such stationary periods can somehow be observed. We will relax this simplification shortly, but first we identify several relevant features that arise from conventional practice. Equation 2 holds with homogeneous vehicles. Rewriting the equation in terms of jam spacing, s_j, (the reciprocal of k_j), we can solve directly for jam occupancy, occ_j, via Equation 3.

$$k = \frac{occ}{L_{eff}} \tag{2}$$

$$occ_j = \frac{L_{eff}}{s_j} \tag{3}$$

Upon reviewing the literature it is surprising how few researchers even contemplate jam occupancy and we were only able to find publications from five different groups that considered occ_j. Of those that do, neither [14] nor [15] offer any guidance on establishing occ_j. While [16-17] simply say that occ_j should be close to 100% and [18] says it should be equal to 100%. These first four groups use occ_j as an explicit proxy for k_j, but [19] define occ_j differently; they use occ_j to denote the maximum observed occ in a given empirical data set (e.g., q at occ_j is far above zero in their work) and as one should expect, their values vary from site to site.

Returning to Fig 1, $s_j = 25\ ft$ given our choice of k_j, and setting $L_{eff} = 20\ ft$ for a typical passenger vehicle (see, e.g., [11]), we calculate $occ_j = 80\%$. At first this finding may seem counterintuitive; however, consider a platoon of 14 ft vehicles coming to a stop over a single loop detector with a 6 ft detection zone (net: $L_{eff} = 20\ ft$), and $s_j = 25\ ft$. There is a 20/25 chance that a vehicle will stop within the detection zone and 5/25 chance that no vehicle will stop in the detection zone, yielding an expected probability of 80%, corresponding to the calculated occ_j. Of course any given stop should either have 0% or 100% occupancy for its duration, but across many stops the average should be 80%. Using equation 2, we add a second

axes to the abscissa of the qkFR in Fig 1 to also denote the corresponding qoccFR for $L_{eff} = 20\ ft$. This derivation can also explain earlier findings that the empirical qooccFR does not necessarily pass through 100% occ at zero flow [20].

Our choice of k_j is on the high side of what is feasible and one should expect the actual value to vary from location to location as well as over time at a given location. Using the data from [21] we found a median s_j of 28.5 ft for passenger vehicles across shockwaves 1, 3, 5 and 10, yielding occ_j closer to 70%. With longer L_{eff}, s_j would also be higher, so occ_j from Equation 3 could be slightly higher or lower than the values calculated here, but occ_j should still be far below 100% on average.

Realistic sampling period

Even with homogeneous vehicles and stationary conditions, the measurements of q and occ are subject to discretization errors. Fig 2A repeats the triangular qkFR from Fig 1. Using the common sampling period, $T = 30\ sec$, there are only 21 resolvable values of q between capacity and jam, as shown with points on the triangle in the queued regime. Operating agencies typically count vehicles the instant they enter the vehicle detector, but they measure occupancy in whatever sample it occurs. So if a vehicle straddles the boundary between two successive samples, it will contribute to q and occ in the first sample, but only to occ in the second sample. While the resolvable q is discrete, the true range of q is continuous. If the sample period is not equal to an integer number of headways, the non-integer portion will give rise to sampling errors, and push the measured (q, occ) off of the underlying qoccFR.

Fig 2C shows a typical pulse train for homogeneous vehicles passing in stationary conditions. This pulse train has a true traffic state shown with a star in Fig 2A. With a true $q = 660\ vph$, the headway, h, between vehicles is 5.45 sec, which does not divide evenly into our chosen T. The measured q and occ will depend on the relative offset between the pulse train and the start of the sample. The four extreme cases are shown below the pulse train. In T1 the sample begins right before a vehicle enters, so this sample has the highest possible on-time and largest feasible number of rising edges, thus, leading to the highest possible occ and q, respectively. If instead the sample began right after a rising edge, as in T2, occ would not change from T1 but one fewer vehicles would be counted in q. As the offset increases further, occ will remain constant until the end of the sample crosses the falling edge and then it will drop linearly until the end of the sample is right before a rising edge, as in T3. This sample has the lowest possible on-time and thus, the lowest occ. If instead the end of the sample fell right after the rising transition, as in T4, that vehicle would be counted in q while occ would not change from T3. As the offset increases further the occupancy would increase linearly until reaching T1 again (in this case occ does not "flat-line" at the lowest value for a range of offsets, as it did at the highest value, because the true occ is over 50%).

The horizontal lines in Fig 2A show the range of measurable traffic states that arise from the continuous range of true traffic states in the queued regime and the variable offset as was used in Fig 2C. If v is so low that only one vehicle is observed, any value of occ could be measured; thus, at the lowest nonzero speeds measured occ is effectively meaningless. Fig 2B shows the difference between the maximum and minimum feasible occ measurements for the resolvable values of q. The range is at least 7% for all resolvable q below q_o and grows to 100% at the lowest resolvable nonzero q.

The discretization errors in this example are strictly due to sampling. There are no detection errors beyond the fact that there are a non-integer number of headways per sample. The

true traffic state comes from homogeneous vehicles under stationary traffic conditions, and the true traffic state always falls on the triangular qoccFR in this example. Obviously the occ range should increase in real traffic, where headways are non-uniform and thus, a sample could include far more or far less on-time than used in this example.

Inhomogeneous vehicle fleet

A realistic vehicle fleet is inhomogeneous, and thus, the true L_{eff} should vary from sample-to-sample. In which case, at single loop detectors one can no longer use Equation 2 because it is not possible to separate the impacts of L_{eff} from the impacts of v. At dual loop detectors one can measure the L_{eff} in each sample directly and use Equation 2 to calculate k; or alternatively use Equation 1 to calculate k, and then use Equation 2 to solve for L_{eff}.

Fig 3A shows one qoccFR for each of three different homogeneous vehicle fleets. The top curve is strictly cars with $L_{eff} = 20\ ft$ (from Fig 1), the middle curve is strictly single unit trucks with $L_{eff} = 40\ ft$, and the bottom curve is strictly multi-unit trucks with $L_{eff} = 70\ ft$. Each curve is calculated for the given homogeneous fleet under stationary conditions with long sampling periods, thus, the only factor that differentiates the curves is L_{eff}. Note that in this plot occ_j is set to 100% for the two different trucks, which is unrealistically high and represents the case where the vehicles stop with a stop-gap between vehicles equal to the detection zone size [20]. As will be discussed shortly, this scenario provides an upper limit on q_o for the given fleet. Fig 3B repeats the exercise with an additional 10 ft stop-gap between trucks when stopped, and thus, a lower occ_j (we suspect a more realistic stop-gap would be even larger). Among the three curves both q_o and the slope of the unqueued regime decrease as L_{eff} or the stop-gap increases. As observed in [22], both q and occ are a function of L_{eff}. The lower slope in the unqueued regime of the truck curves simply reflects the fact that the trucks are longer and for a given v and q, these longer vehicles are over the detector for more time per vehicle.

Assuming the qkFR for all three groups of vehicles have the same v_f and slope, w, in the queued regime, Fig 3C-D show the corresponding qkFR that were used to generate the curves from Fig 3A-B. Unlike the qoccFR, the unqueued regime from all three curves now fall on top of one another in the qkFR. Note that compared to the cars, k_j is much lower for the two truck curves, dropping by 50% for the single unit trucks and almost 70% for the multi-unit trucks in Fig 3D. Again, this drop arises because the trucks are so much longer than the cars, fewer trucks will fit per unit distance. Since v_f and w are held constant across all three groups, for the longer vehicles q_o must also drop relative to cars and doing so by the same percentages as k_j did. It is quite possible that v_f should be lower for trucks than for cars, which could pull q_o for the trucks even lower than shown in these plots. Although k_j exhibits a large range across groups in Fig 3D, note that the corresponding occ_j in Fig 3B does not.

With an understanding of how different homogeneous fleets exhibit different FR, now consider an inhomogeneous vehicle fleet. Many researchers do not appreciate the full impact of an inhomogeneous vehicle fleet and the complications that arise. We have already seen that as L_{eff} increases for a homogeneous vehicle fleet, the qkFR should shrink, with lower q_o and lower k_j simply because the vehicles are longer, i.e., the true qkFR and its parameters q_o and k_j are functions of L_{eff}. Now consider one of the simplest possible inhomogeneous fleets- a weighted average of vehicles from the three groups shown in Fig 3. If the trucks and cars are mixed together at a constant ratio in stationary traffic and all vehicles travel at the same v, the resulting L_{eff} will be constant and the underlying qkFR will be a weighted average from the three curves in Fig 3D. If the ratio of the three groups of vehicles varies from sample-to-sample, so too will L_{eff}

and the underlying qkFR for the particular combination of inhomogeneous vehicles. Assuming (unrealistically) that each sample remains stationary, the constant v will constrain the traffic state to a line in the qk-plane radiating from the origin with slope equal to v; with different values of L_{eff} falling at different points on this line, thus, creating considerable scatter in the queued regime as a function of the distribution of vehicle lengths in a given sample. For the unqueued regime this scatter all falls on the line at slope v_f, and thus, will appear to be consistent with all three qkFR curves.

In the case of qoccFR the scatter due to different L_{eff} will fall on a roughly vertical line, a fact that is rarely appreciated in the literature and if it is not accounted for, will distort any relationships related to Equation 2. Since the qoccFR do not overlap in the unqueued regime, the unqueued regime will also exhibit scatter in this fashion, e.g., Fig 3B. In any event, the scatter comes before accounting for the sampling issues discussed in the previous section, and those sampling impacts become larger with inhomogeneous vehicles, as is the case in the current scenario.

As the sample L_{eff} changes, so too does the underlying qkFR, e.g., the three distinct curves in Fig 3D. So any empirical observation with inhomogeneous vehicles (but stationary traffic) will actually be averaging from a blurry set of qkFR (e.g., as if each vehicle came from one of the three curves shown in the plot). This fact is the underlying motivation for passenger car equivalents (PCE), but PCE is ultimately a gross correction factor that assumes that all samples have the same L_{eff} and it is not clear how one would reshape the qkFR in response to PCE. A constant PCE is too simplistic to capture the ever-changing mix of vehicles, e.g., as was shown in [10-11], the true L_{eff} can change by a factor of two from sample-to-sample, particularly during low q. In this discussion we have assumed traffic is stationary, if we relax that condition, it is possible that as speeds change over the sample and the low-speed vehicles may come predominantly from one of the three curves in Fig 3D, while the higher speed vehicles come predominantly from a different curve in this plot, further distorting the aggregate traffic state for the sample. If trucks travel slightly slower than cars, that will lead to even greater noise that depends on how the individual vehicles' speed and lengths are correlated. In mixed traffic we suspect that gaps between different vehicle types are likely to be larger, and if so, it would further lower k_j and q_o. If the gap a driver chooses is a function the length of their own vehicle and that of the vehicle ahead, then the necessary PCE becomes quite complicated and is beyond the resolution of conventional aggregated detector data. For example [13] used a PCE that was as specific as possible given the available data, already aggregated to $T = 5\ min$, yet still they found measurably different qkFR in the median and center lanes, suggesting that their PCE did not fully correct for the trucks. Others have sought conditions to eliminate the need for PCE, e.g., in a study of vqFR, [8] explicitly used a lane with a truck restriction to avoid, "the complication of determining and applying passenger car equivalency conversions."

There is no easy solution, though we suspect it would be more constructive to think in terms of the sample distribution of the individual vehicles' effective lengths. What is clear is that the three curves in Fig 3C-D are accurate representations of the true qkFR for the given homogeneous group and at least in theory the true underlying qkFR for an inhomogeneous fleet could vary by the range exhibited in the figure. The three curves converge as they approach occ_j in Fig 3B but remain separate as they approach k_j in Fig 3D. In empirical measurements this reduced noise in qoccFR at high occ is at least partially countered if not exceeded by the increased noise due to the realistic sampling period shown in Fig 2 that was excluded from Fig 3. Combined with a realistic sampling period, a queued sample with $v < 10\ mph$ will have at most

10 veh when $T = 30\ sec$. If all of the vehicles are cars, then that gives rise to Fig 2 discussed above, but if some of the vehicles are trucks, each truck would replace 2-3 cars and would pull the flow downward, with little change to occ.

Inhomogeneous drivers

Similar to the impacts of different L_{eff}, the traffic state will also be perturbed if different drivers behave in different ways. Returning to homogeneous L_{eff} across all vehicles, when a cautious driver in a sample choses a larger gap, the traffic state will move closer to the origin of the qk-plane, with no change to v. Likewise, if an aggressive driver chooses a smaller gap in the sample, the state will move further from the origin. The lighter curve in Fig 4 reiterates the unperturbed states between capacity and jam from the qkFR and qoccFR in Fig 1. Superimposed on top of this curve is the range of feasible states if drivers on average take larger gaps (up to 200%) or smaller gaps (down to 80%) for a given v. All of the perturbed states at the given v fall in a single line radiating from the origin with slope equal to v (in contrast, varying L_{eff} at a constant v will move the measured state vertically in the qocc plane, Fig 3a-b). Note that the left-most line corresponds to v_f and spans the largest range of q; however, in an empirical qkFR this noise is not readily apparent because it blurs with the purely unqueued traffic states at v_f that have a lower q.

Non-stationary traffic

In non-stationary traffic, the measured q and occ are averaged from multiple traffic states, some might not even be on the underlying qkFR, e.g., due to the reasons noted above, as well as accelerations, lane change maneuvers, and so forth (see, e.g., [5-6, 23-25]). The focus of the present work is on the measurement process introducing biases that can confound empirical studies even when the traffic state falls on the underlying qkFR. In that context, consider the lowest speeds. The conventional operation of single and dual loop detectors (as well as most other traffic detectors) assume that acceleration, a, can be neglected [26]. While this assumption probably holds down to about 20 mph, it breaks-down at lower speeds.

The more significant acceleration becomes, the less likely the fundamental equation (Equation 1) holds if acceleration is unaccounted for. At the extreme, consider the vehicles that come to a stop. Fig 5 shows the highest initial speed, V_o, at which a vehicle can come to a complete stop over a detector as a function of its effective length and a. There are three curves, each at a different value of a. There are four lengths highlighted on each curve capturing typical passenger vehicles (physical length of 14 ft) and multi-unit trucks (physical length of 64 ft) as they pass over a single loop detector (6 ft detection zone) or dual loop detector (26 ft detection zone). Ultimately the detectors measure an individual vehicle's time average speed as it traverses the detector. From the equations of motion the time average speed for a stopping vehicle is 1/2 the value of V_o. Such a vehicle could come to a complete stop just past the detector, contributing 0% to occ during the duration of the stop. On the other hand the vehicle could come to a stop over the detector, contributing 100% to occ while stopped. The probability that the vehicle stopped over the detector is equal to occ_j. Since occ_j should be close to 80%, for most of the stops the vehicle will contribute 100% to occ while stopped, pulling the average to the jam state. For the rest of the stops the vehicle will stop just past the detector and contribute 0% to occ, pulling the average to the origin. In either case, a stopped vehicle will not contribute the correct amount of occ_j necessary to satisfy Equation 2. Even if a vehicle does not come to a complete

stop, measured q, occ and v can be distorted due to unaccounted for acceleration while the detector measures the vehicle [26].

If a given sample's $v < 10\ mph$ on a freeway, it is highly likely that some portion of the sample included stopped vehicles and thus, occ should be greatly distorted, as per Fig 5. Combined with the findings from Fig 2 *(that low-speed data exhibit the largest sampling errors in q and occ)* and Fig 3 *(that lower q increases the impacts of +/- one truck in a given sample)* these low-speed data are so noisy that the data should be discounted if not ignored entirely.

Unfortunately, the qk-plane is skewed such that these low-speed samples cover a disproportionately large area and this heteroscedasticity can easily bias any curve fit to empirical data unless the low-speed conditions are properly accounted for. To illustrate this distortion consider the qkFR on the right side of Fig 1 and the corresponding vqFR on left-hand side of the figure. The points on the left-hand side are plotted at 5 mph intervals and translating these horizontally to the curve on the right-hand side of the of the figure, one can see that the highest speeds in the queued regime are compressed to a small range of q. Visually, the qkFR greatly distorts the relationship to v. When speed has dropped by a third ($v = 40\ mph$) flow has only dropped by 10%. When speed has dropped by 85% ($v = 10\ mph$) flow has only dropped by 49%. In other words, the higher-speed data are compressed into a narrow sliver while the low-speed data ($v < 10\ mph$) are spread over more than half of the qk-plane. This large physical area increases the susceptibility to include the extremely noisy low-speed data. Numerous past studies give equal weight to equal sized regions of the qk-plane (or qocc-plane), without regard for this distortion. So instead of discounting the extremely noisy low-speed data, these studies actually give them greater weight than the higher-speed data with less noise.

Finally, within non-stationary traffic there is no guarantee that the observed q and occ actually come from an underlying stationary FR. Several papers offer models and evidence to suggest that the observed traffic state can be systematically pulled above the stationary qkFR [6, 27] or below it [5].

Averaging across lanes

One may be tempted to combine data across lanes to increase the sample size, which may be appropriate for some applications; however, combining lanes does not reduce the impacts of the discretization errors that arise when using realistic sampling periods. For an n-lane road, the resulting average sample comes from n trials of T*q vehicles, each with discretization errors as shown in Fig 2B. The n-lane average does not come from a single sample of n*T*q, which would have finer resolution on q and thus a lower occ rage in Fig 2B.

Even when traffic is near stationary within each lane, averaging across lanes is likely to capture the inhomogeneity between lanes due to different v or different L_{eff}, e.g., [13] shows systematic differences in qoccFR between lanes even after attempting to correct for PCE while [5] found the maximum throughput varied by lane. If the longer vehicles tend to travel in the outside lane, the true qkFR observed in adjacent lanes might look something like the three different curves in Fig 3D and the average across lanes in any given sample will reflect the relative q in the different lanes. The relative q in a given lane should change from sample-to-sample, and if the lane has more trucks, the relative q will be lower than similar conditions without trucks (as per the earlier discussion of the inhomogeneous vehicle fleet). The impacts will be most pronounced whenever a disturbance impacts lanes differently, e.g., if an upstream moving disturbance passes different lanes at different times.

Detector errors

When working with empirical data from vehicle detectors, it is important to keep in mind that there are many measurement errors that can distort the data. These can be split into computation errors and detection errors.

The computation errors may be due to an unknown software bug or because the researcher is not aware of the subtle nuances in a measurement algorithm originally designed for day-to-day traffic management. A case in point is [28], who used dual loop detector data to conclude that conventional single loop detector speed estimation, Equation 4, provided a biased estimate of v. For a decade single loop detectors fell into disfavor as a result of [28]. It was not until [12] used individual vehicle actuations (a resource that is still rarely used) to show that while [28] thought they were working with space-mean speed, they did not realize that the operating agency actually measured time-mean speed. Then [10] showed that furthermore, the operating agency truncated occ to the nearest integer, which greatly distorted the relationships in the unqueued regime. These subtle misunderstandings can have big impacts on the final results and it is quite likely that a given operating agency may process the data in such a way that disrupts the underlying traffic dynamics. For example the data [28] used is far from unique, many other operating agencies calculate the arithmetic average of the individual vehicle speeds (time-mean speed) instead of the harmonic average (space-mean speed), but use of time-mean speed nullifies the fundamental equation and the ability to calculate k from q and v.

$$\hat{v} = \frac{L_{eff} \cdot q}{occ} \tag{4}$$

Even when the measurement algorithms are fully understood, detection errors can still disrupt the measurements. Pulse breakup errors [29] and splashover errors [30] can introduce non vehicle pulses that erroneously increase q. Pulse breakup errors reduce occ while splashover increases occ; [30] also discovered some splashover events simply extended valid vehicle actuations, increasing occ without impacting q. The detector errors can happen more frequently for one type of vehicle, e.g., the trailer on multi-unit trucks typically have a higher ground clearance and thus, a loop detector is more likely to drop out prematurely, decreasing the measured occ for the sample and thus, the given vehicle may have a true effective length shorter than its physical length. The detector responsiveness and sensitivity will also impact the occ measurement and in the case of dual loop detectors, errors in the assumed detector spacing will impact the speed measurements [31]. At a more basic level are the simple misdetections, e.g., when a vehicle passes over a dual loop detector and only one of the paired loops responds to the vehicle. In this case, the vehicle could contribute to q and/or occ without being included in v. Typically the aggregation algorithms used by the operating agencies run in real time, starting to count occ and q as soon as the vehicle enters the detector, then if the actuation is incomplete, the algorithm simply moves to the next vehicle and does not go back to erase the impact of the unmatched pulse on q and occ. One should expect some amount of detection errors to be present in empirical data. These errors can be impossible to detect without the individual vehicle actuations used in [29-31]. Although these high-resolution data are commonly generated in the roadside controller cabinet, they are usually discarded immediately after the data have been aggregated to calculate q, occ and v; and operating agencies rarely record the individual vehicle actuations.

Other detector technologies such as side-fire microwave radar and video image processing are starting to replace loop detectors, but in most cases they also replicate the

functionality of loop detectors. The side-fire detectors typically suffer from occlusion and merging targets together, similar to pulse breakup and splashover errors at loop detectors. Video is also susceptible to changing lighting conditions, yielding less accurate q and occ measurements compared to loop detectors (see, e.g., [32-33]).

Discussion and Conclusions

This paper explored several commonly overlooked factors that impact empirical fundamental relationships. Most of these factors are conceptually simple, but collectively they result in surprisingly large, non-linear, and often heteroscedastic distortions of the empirical traffic state measurements. Although the work used a triangular qkFR to illustrate the findings, the general results should apply no matter what shape the FR takes. The ultimate goal of this work is to help those working with macroscopic traffic data better understand the microscopic source of the noise so that the reader will be equipped to make informed decisions between meaningful displacements and artifacts of aggregation.

We began the analysis by showing that with homogeneous vehicles, stationary traffic, and a long sampling period that occ_j should be around 80% on average; but due to the discrete nature of vehicles, whenever traffic stops over a detector the measured occupancy will either be 0% or 100%. Erroneously assuming $occ_j = 100\%$ will distort any curve fit to empirical data.

By relaxing the unrealistic assumptions of homogeneity, stationarity, and long samples, we found the impacts of noise are greatest at lower speeds, particularly $v < 10\ mph$, for numerous reasons, including:

1) When $v < 10\ mph$, there are at most 10 vehicles in a 30 sec sample, as this number decreases, the impact of any single uncommon vehicle increases (a long truck, cautious driver, etc.).
2) At 10 mph the fractional headways at the ends of a 30 sec sample will cause both q and occ to vary by at least 9%, and this range grows as speed drops.
3) As the traffic approaches the jam state the measured occ becomes completely meaningless, with feasible measurements spanning the entire range from 0% to 100%.
4) Freeway traffic with $v < 20\ mph$ mph is rarely stationary, which undermines many assumptions of the FR.
5) Likewise, with $v < 20\ mph$ many of the assumptions used to measure the traffic state no longer hold, e.g., that acceleration can be ignored. At its worst, a vehicle can come to a complete stop and the stop-time could be reflected in occ but not in v, or vice versa.

As a result of the above factors, we think that empirically collected traffic states with $v < 10\ mph$ should be discounted and treated with great care to avoid spurious results from the increased noise. It is best to avoid stopped traffic altogether when constructing empirical FR. If one only has fixed sample data, it means avoiding samples with low-speeds.

Unfortunately the qk-plane is skewed such that low-speed samples cover a disproportionately large area. We showed that samples with $v < 10\ mph$ are spread over more than half of the qk-plane, while the higher-speed data are compressed into a narrow sliver. So any linear weighting by q, k or occ will disproportionately give greater weight to the extremely noisy low-speed data while inadvertently discounting the higher-speed data with less noise.

One may be tempted to combine data across lanes to increase the sample size. However, combining lanes does not reduce the impacts of the discretization errors and if different lanes

have different L_{eff}, combining data across lanes could amplify some of the sources of noise. Alternatively, moving to a longer T to increase the sample size also increases the chance that multiple traffic states will be inappropriately averaged together [9].

Almost all empirical traffic flow studies use detectors that were actually deployed for day-to-day traffic management. These detectors are rarely optimized for the traffic flow theory needs. When using empirical data from a detector it is important to make sure that the operating agency is calculating the space-mean speed via the harmonic mean of the individual vehicle speeds and that they are not truncating occ. It is also important to verify the tuning and calibration of the detectors to minimize detector errors.

Typically empirical data are already aggregated to 30 sec or longer, but even with the individual vehicle actuations it is difficult to identify stopped vehicles, e.g., a dual loop detector can never measure 0 speed. A vehicle that stops over a dual loop detector for several seconds can still pass the detector with a measured speed above 5 mph (e.g., Fig 5). Though if one has access to the individual vehicle actuations it should be possible to find evidence of stop waves, e.g., find the local minima in the successive vehicle speeds, and then exclude these events from the samples.

When $v > 10\ mph$, assuming the detectors have few errors, we suspect the two largest sources of noise are the inhomogeneous vehicle fleet and inhomogeneous drivers. The former arises due to the ever-changing L_{eff} from sample-to-sample, and thus, the true underlying qkFR also changes from sample-to-sample (e.g., Fig 3), leading to vertical displacements in the qocc-plane. The latter arises from drivers choosing different gaps, which perturbs q, k and occ without impacting v directly (e.g., Fig 4), leading to radial displacements in the qocc-plane.

Of course more research needs to be done. With all of these sources of noise, it is not clear how best to even attempt to fit a FR curve to empirically collected data. Consider the qkFR for example, empirically collected traffic state measurements will be scattered over the qk-plane. Should these data be binned by q, by k, by v, or something else? The fitted curve will differ in all three cases. Furthermore, in the context of Fig 3 and inhomogeneous vehicles, the empirically observed qoccFR will differ from the corresponding qkFR, depending on the distribution of vehicle lengths.

References

[1] Lighthill, M., Whitham, G., (1955) "On Kinematic Waves II. a Theory Of Traffic Flow on Long Crowded Roads, "*Proc. Royal Society of London, Part A*, Vol. 229, No. 1178, pp 317-345.

[2] Richards, P., (1956) "Shock Waves on the Highway," *Operations Research*, Vol. 4, No. 1, pp 42-51.

[3] Chandler, R.E., Herman, R., Montroll, E.,W., (1958) “Traffic dynamics: studies in car following,” *Operation Research*, Vol. 6, No. 2, pp. 165-184.

[4] Gazis, D. C., Herman, R., Rothery, R. W., (1961) “Nonlinear follow the leader models of traffic flow,” *Operations Research*, Vol. 9, No. 4, pp. 545-567.

[5] Coifman, B., Kim, S., (2011) "Extended Bottlenecks, the Fundamental Relationship and Capacity Drop," *Transportation Research- Part A*. Vol 45, No 9, pp 980-991.

[6] Kim, S., Coifman, B., (2013) "Driver Relaxation Impacts on Bottleneck Activation, Capacity, and the Fundamental Relationship," *Transportation Research- Part C*, Vol 36, pp 564-580.

[7] May, A., Athol, P., Parker, W., Rudden, J. (1963) "Development and Evaluation of Congress Street Expressway Pilot Detection System," *Highway Research Board Record No. 21*, pp 48-70.

[8] Persaud, B., Hurdle V. (1988) "Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships," *Transportation Research Record 1194*, p. 191-198.

[9] Cassidy, M. (1998) "Bivariate Relations in Nearly Stationary Highway Traffic," *Transportation Research Part B*, Vol. 32, No. 1, p. 49-59.

[10] Coifman, B. (2001) "Improved Velocity Estimation Using Single Loop Detectors," *Transportation Research: Part A*, Vol 35, No 10, pp 863-880.

[11] Coifman, B., Dhoorjaty, S., Lee, Z. (2003) "Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors," *Transportation Research: Part C*, Vol 11, No 3-4, pp 211-222.

[12] Cassidy, M., Coifman, B. (1997) "The Relation Between Average Speed, Flow and Occupancy and the Analogous Relation Between Density and Occupancy," *Transportation Research Record 1591*, Transportation Research Board, pp 1-6.

[13] Hall, F.L., Gunter M.A. (1986) "Further Analysis of the Flow-Concentration Relationship," *Transportation Research Record, No. 1091*, pp. 1-9.

[14] Budrick, J. (1982). *Evaluation of two isolated responsive ramp control tactics* (No. 96). Illinois Department of Transportation, Bureau of Materials and Physical Research.

[15] Heydecker, B., Addison, J. (2011) "Analysis and Modelling of Traffic Flow under Variable Speed Limits," *Transportation Research Part C*, Vol. 19, No. 2, pp 206-217.

[16] Geroliminis, N., Skabardonis, A. (2006) "Real Time Vehicle Reidentification and Performance Measures on Signalized Arterials," *Proc. of IEEE ITSC 2006*, pp 188-193.

[17] Geroliminis, N., Skabardonis, A. (2011) "Identification and Analysis of Queue Spillovers in City Street Networks," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, pp 1107-1115.

[18] Banks, J. (1999) "Investigation of Some Characteristics of Congested Flow," *Transportation Research Record 1678*, Transportation Research Board, pp 128-134.

[19] Chen, D., Zhang, J., Tang, S., Wang, J. (2004) "Freeway Traffic Stream Modeling Based on Principal Curves and its Analysis," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4, pp 246-258.

[20] Hsu, P., Banks, J., (1993) "Effects of Location on Congested-Regime Flow-Concentration Relationships for Freeways," *Transportation Research Record 1398*, pp 17-23.

[21] Coifman, B. (1997) *Time Space Diagrams for Thirteen Shock Waves*, University of California - PATH, 1997.

[22] Coifman, B., Wang, Y., (2005) "Average Velocity of Waves Propagating Through Congested Freeway Traffic," *Proc. of The 16th International Symposium on Transportation and Traffic Theory*, July 19-21, College Park, MD. pp 165-179.

[23] Cassidy, M., Bertini, R. (1999) "Some Traffic Features at Freeway Bottlenecks," *Transportation Research Part B*, Vol. 33, No. 1, pp 25-42.

[24] Laval, J., Daganzo, C. (2006) "Lane-Changing in Traffic Streams," *Transportation Research Part B*, Vol. 40, No. 3, pp 251-264.

[25] Duret, A., Bouffier, J., Buisson, C. (2010) "Onset of Congestion due to Low Speed Merging Maneuvers within a Free-Flow Traffic Stream: Analytical Solution," *Transportation Research Record 2188*, Transportation Research Board, pp 96-107.

[26] Wu, L., Coifman, B., (2014) "Vehicle Length Measurement and Classification in Congested Freeway Traffic," [in press] *Transportation Research Record*.

[27] Laval, J.A., and Leclercq, L., (2008) "Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model," *Transportation Research: Part B*, Vol. 42, No. 6, pp. 511-522.

[28] Hall, F., Persaud, B. (1989) "Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems." *Transportation Research Record 1232*, Transportation Research Board, pp 9-16.

[29] Lee, H., Coifman, B., (2011) "Identifying and Correcting Pulse-Breakup Errors from Freeway Loop Detectors," *Transportation Research Record 2256*, pp 68-78.

[30] Lee, H., Coifman, B., (2012) "An Algorithm to Identify Chronic Splashover Errors at Freeway Loop Detectors," *Transportation Research-Part C*. Vol 24, pp 141-156.

[31] Lee, H., Coifman, B., (2012) "Quantifying Loop Detector Sensitivity and Correcting Detection Problems on Freeways," *ASCE Journal of Transportation Engineering*, Vol. 138, No. 7, pp 871-881.

[32] Coifman, B., Beymer, D., McLauchlan, P., Malik, J. (1998) "A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance," *Transportation Research: Part C*, Vol 6, No 4, pp 271-288.

[33] Coifman, B., (2006) "Vehicle Level Evaluation of Loop Detectors and the Remote Traffic Microwave Sensor," *ASCE Journal of Transportation Engineering*. Vol 132, No 3, pp 213-226.

List of Figures

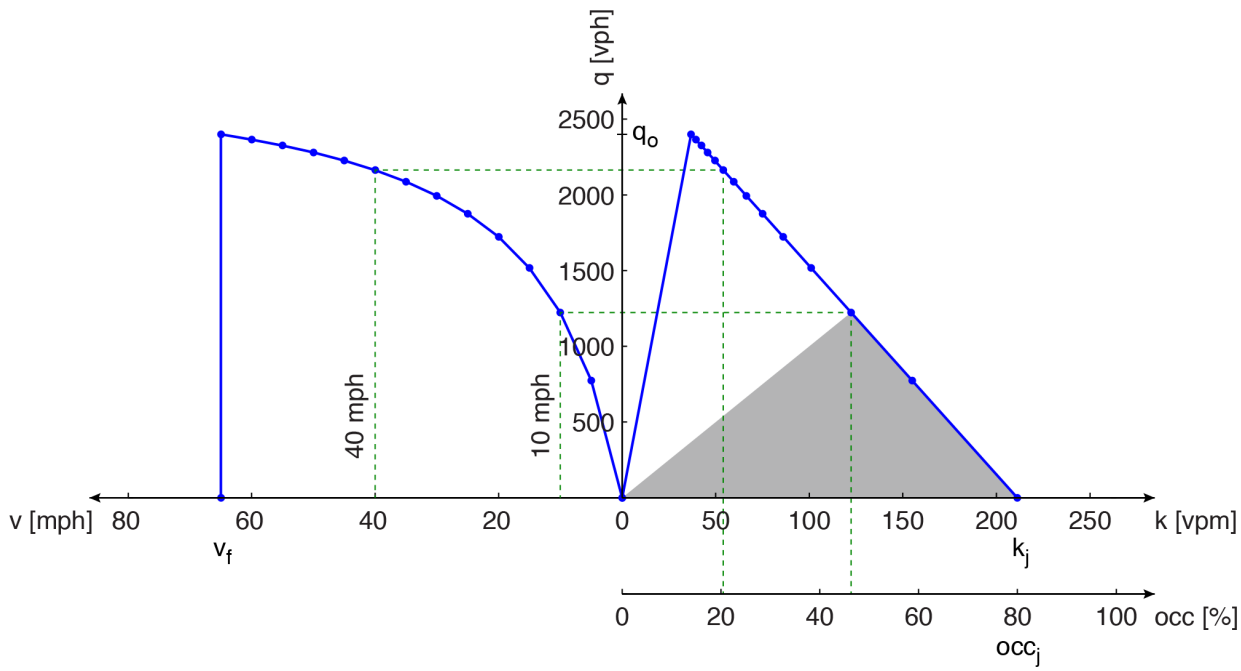

Figure 1, Fundamental relationship between flow versus speed (on the left) and flow versus density or occupancy (on the right) when L_{eff} = 20 ft.

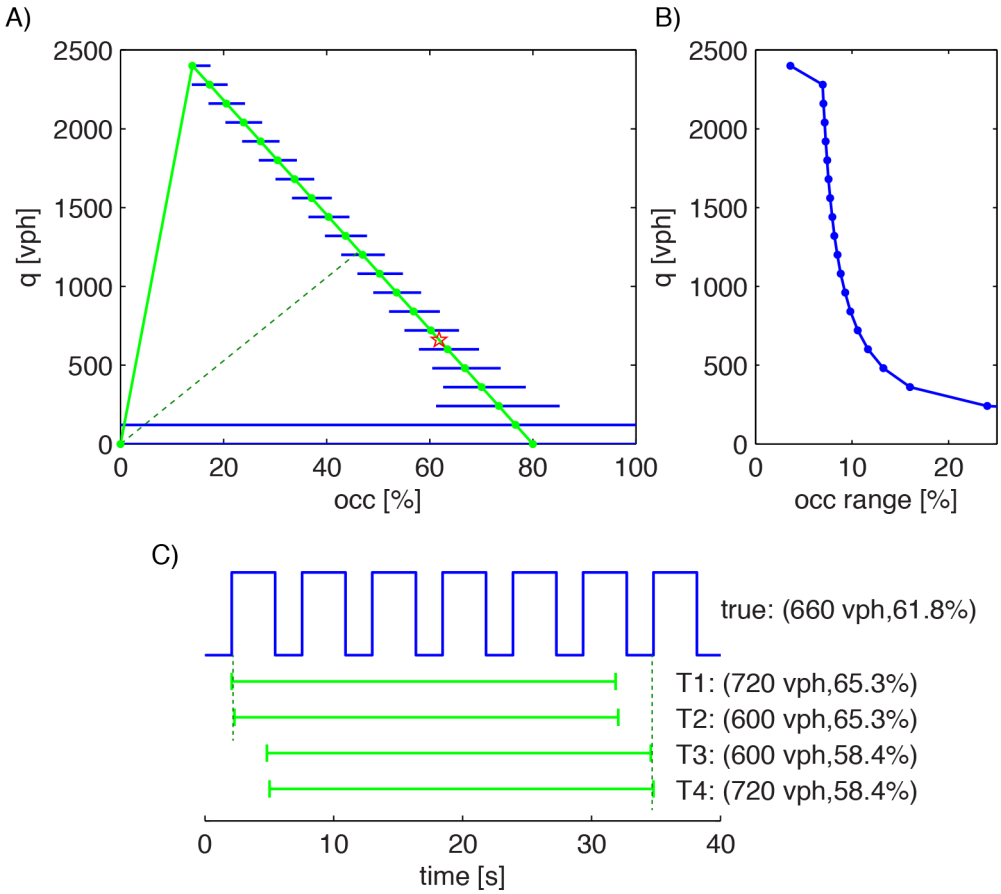

Figure 2, (a) qkFR sampled at T=30 sec when L_{eff} = 20 ft. Resolvable q shown with points in the queued regime and the corresponding range of measurable traffic states with horizontal lines, (b) the difference between the maximum and minimum feasible occ measurements for the resolvable values of q from part a, (c) an example illustrating the range of measurable q and occ due to sampling for a given true q and occ.

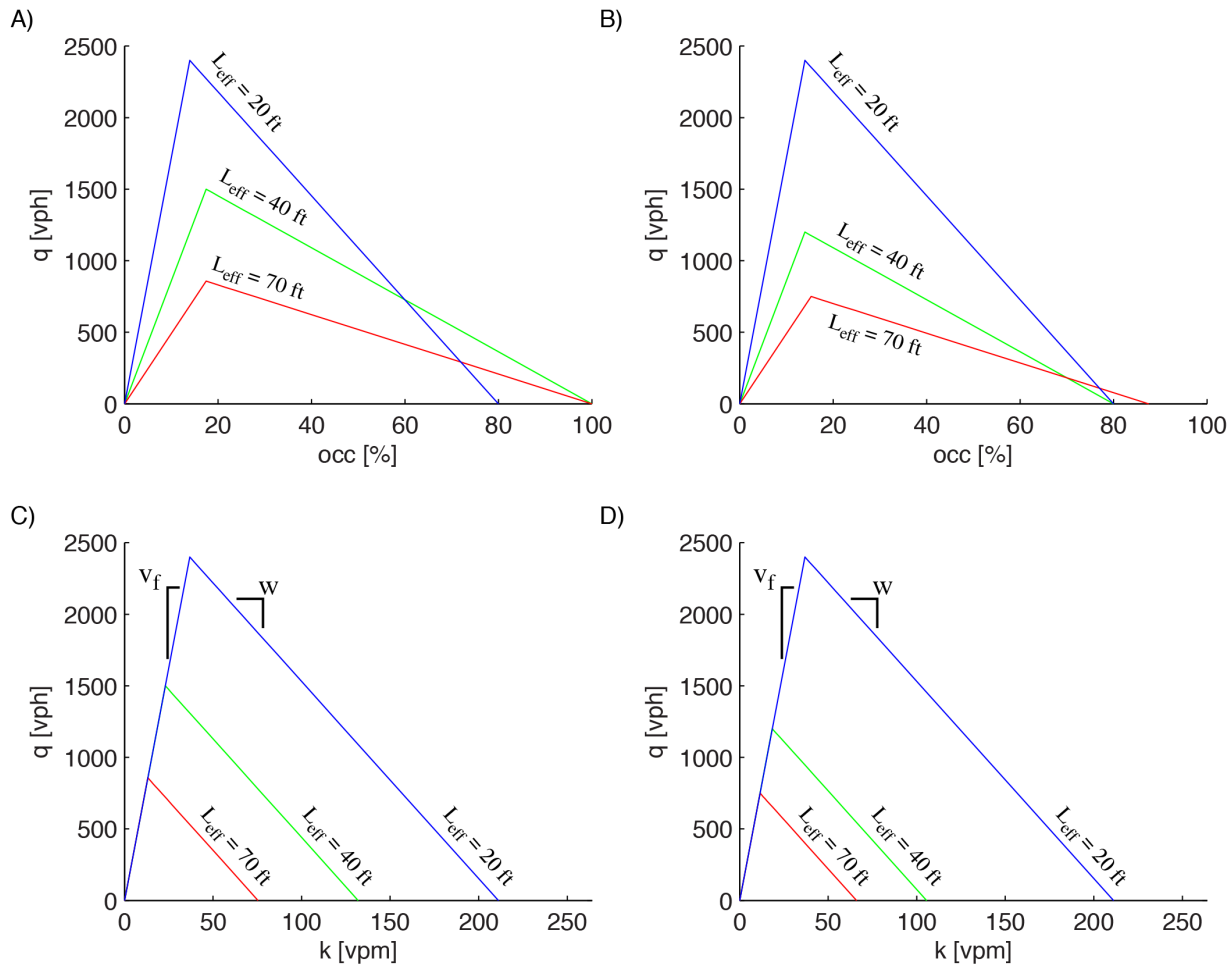

Figure 3, qoccFR for three homogeneous types of vehicles: cars with L_{eff} of 20 ft, single unit trucks with L_{eff} of 40 ft, and multi-unit trucks with L_{eff} of 70 ft, (a) with stop-gap between trucks at jam equal to the detection zone size, and (b) with an additional 10 ft between trucks at jam. (c) The qkFR corresponding to a, and (d) the qkFR corresponding to b.

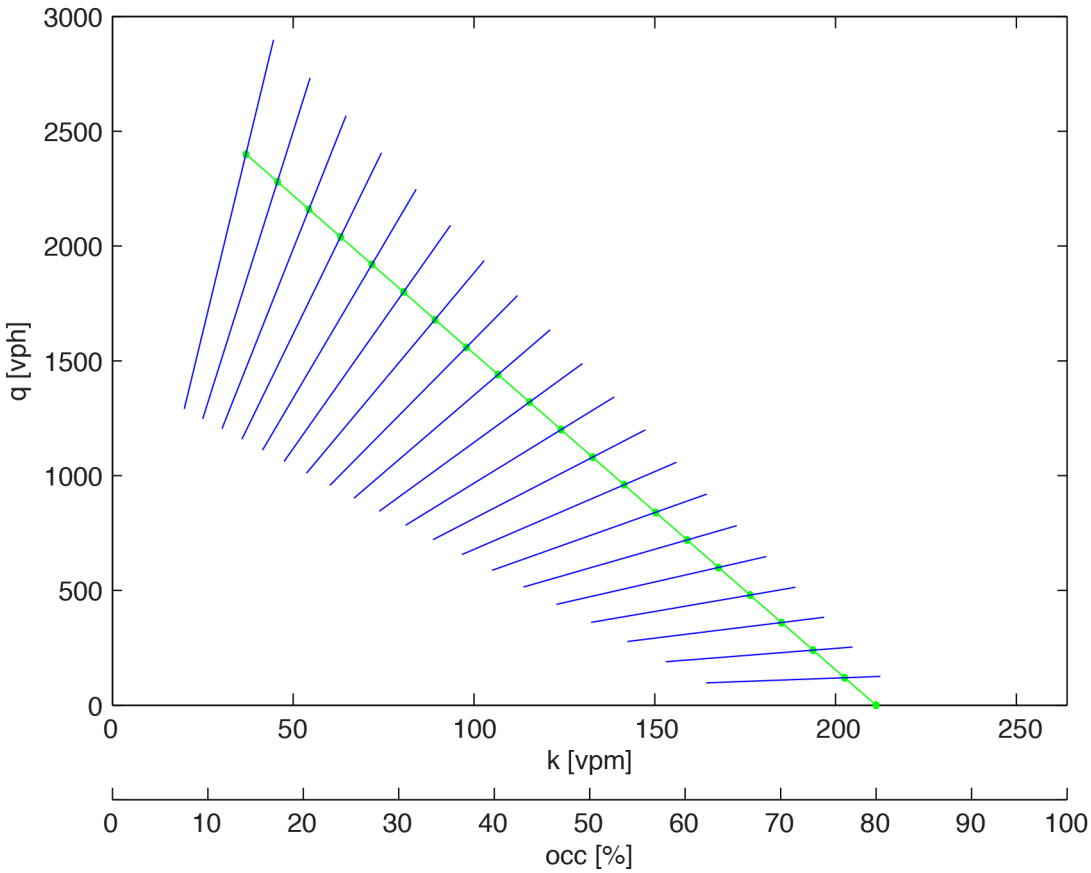

Figure 4, the lighter curve reiterates the unperturbed states between capacity and jam when L_{eff} = 20 ft from Figure 1. Superimposed on top of this curve is the range of feasible states if drivers on average take larger gaps (up to 200%) or smaller gaps (down to 80%) for a given v.

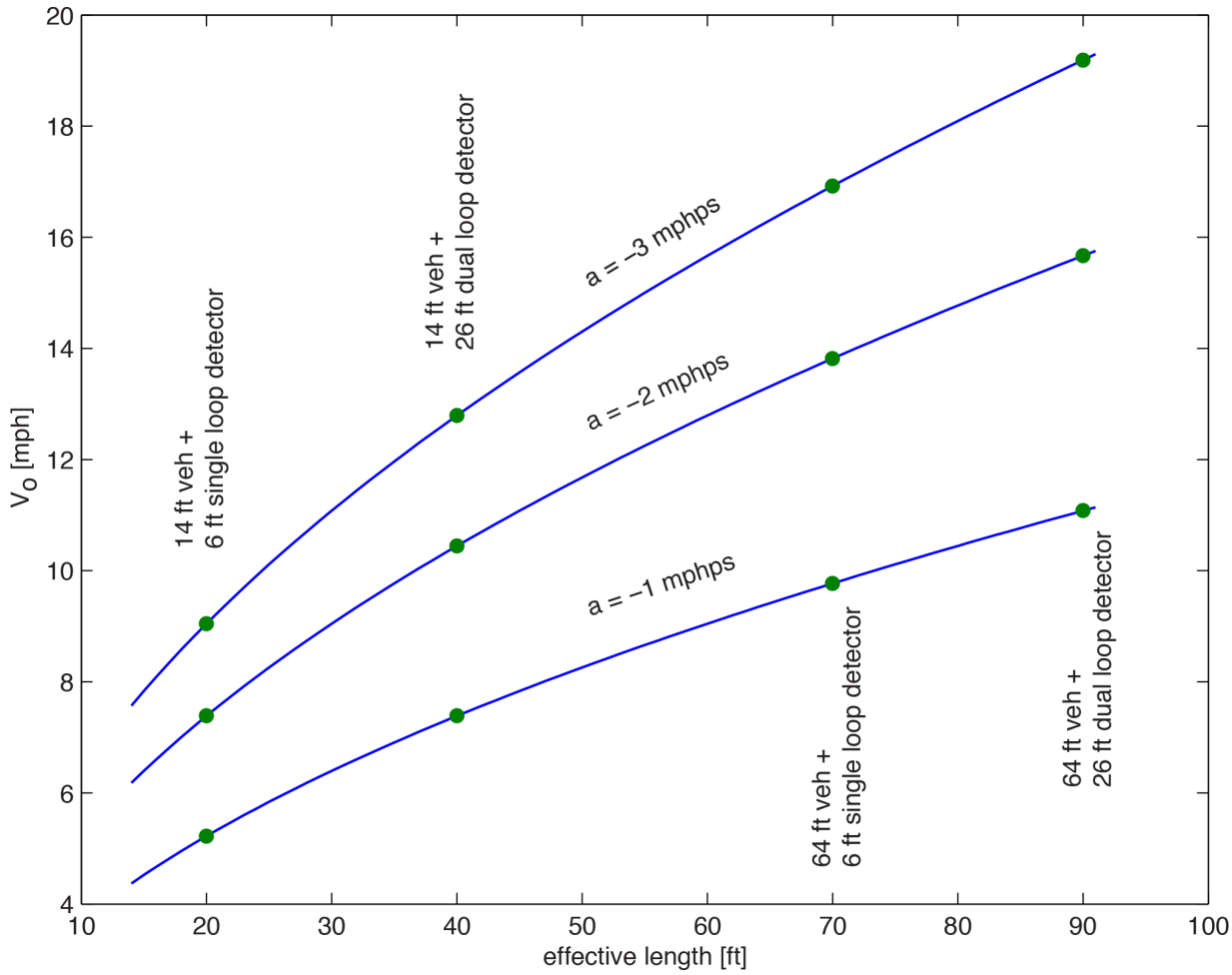

Figure 5, the highest initial speed, V_o, at which a vehicle can come to a complete stop over a detector as a function of its effective length and a.