Fundamental Diagram Calibration: A Stochastic Approach to Linear Fitting

Brian Phegley
Department of Mechanical Engineering
University of California, Berkeley
Berkeley, CA 94720
phone: (510) 725-2011
brianph@berkeley.edu

Gabriel Gomes
California PATH
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720
gomes@path.berkeley.edu

Roberto Horowitz
Department of Mechanical Engineering
California PATH
University of California, Berkeley
Berkeley, CA 94720
phone: (510) 725-2011
horowitz@me.berkeley.edu

Paper submitted to TRB Annual Meeting 2014
August 1, 2013

4065 words + 7 figure(s) $\Rightarrow$ 5815 'words'

ABSTRACT
A statistical learning methodology is proposed for characterizing and identifying key parameters of the fundamental diagram that describes the dependence of traffic flow (or speed) on traffic density in a roadway section, based on traffic data obtained from a vehicle detection station. The proposed fundamental diagram characterization not only provides the expected value of flow (or speed) given a density measurement, but also a random probability distribution of the flow (or speed) given the density measurement. The former can be used to conduct deterministic traffic flow simulations, while the later can be used to conduct statistical flow simulation studies, by using first order traffic flow models such as the cell transmission model.

INTRODUCTION

As the amount and heterogeneity of real-time traffic data increases, it becomes necessary to develop practical methodologies of relating these data to established concepts of traffic theory, and extracting information in a condensed form, which provides both deterministic and probabilistic descriptions of well-known traffic flow behavior, such as the traffic flow (or speed) versus density fundamental diagram. Of interest are the three main macroscopic properties of traffic – the average speed, the flow, and the density – that define traffic conditions in a section of the roadway at a given time. In this paper we propose a statistical learning methodology for characterizing the dependence of traffic flow (or speed) on traffic density in a roadway section, based on traffic data obtained from a vehicle detection station (VDS), in the form of a mixture of conditional probability density functions (PDF). Such a probabilistic characterization of the fundamental diagram provides both the expected value of flow (or speed) and a PDF of the flow (or speed) given a density measurement.

To begin, we assume that average speed v, flow f, and density ρ in a roadway section, as measured by a VDS, are related by

$$v = \frac{f}{\rho}\,. \tag{1}$$

Of greater concern is the relationship between flow and density. This relationship is often considered to be static and time invariant, and described in the form of a function $f(\rho)$, known as the fundamental diagram.

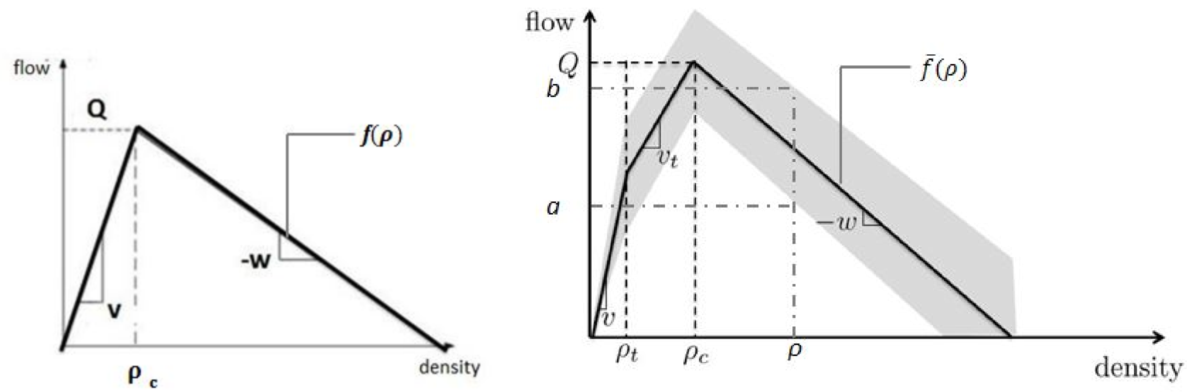
Many forms of the fundamental diagram have been proposed. Perhaps the most well-known is the Greenshields model

$$f(\rho) = v_f \rho (1 - \frac{\rho}{\rho_f}) \tag{2}$$

where v_f and ρ_f are coefficients to be determined. Eq. (2) is clearly a parabolic function of ρ. Another frequently used fundamental diagram function is piecewise affine, as depicted in Fig. 1(a), which is frequently used in first order macroscopic traffic models, such as the Cell Transmission Model (CTM) (*1*). Other functions that have been proposed in the past include logarithmic, exponential, exponential to the quadratic, and various forms of polynomials (*2*). Del Castillo (*3*) contains a review of many of these fundamental diagram functional descriptions, as well as a historical perspective.

A limitation of using static functions to describe the fundamental diagram is that it is difficult to characterize the variability of flow, given a measured value of density, unless the joint flow-density PDF, $\Gamma(f, \rho)$, is also provided. Determining $\Gamma(f, \rho)$ for each VDS is often infeasible. Several methodologies have been proposed to characterize the variability of flow, particularly in the so called congestion regime, from the nominal value provided by the fundamental diagram. Examples include the three-phase theory of Kerner (*4*), which proposes an explicit static function $f_f(\rho)$ for free-flow regime, but uses a so called congested region domain inclusion to describe the flow versus density relation when traffic is congested. Kim and Zhang (*5*) attempted to explain how variation in the drivers efforts to accelerate or decelerate, as well as the spacing between specific cars along the freeway route, lead to variations in values of the fundamental diagram. Sumalee et al. (*6*) use different fundamental diagrams depending on whether a particular roadway section is transitioning between free flow and congestion traffic regimes, or is in steady state of one of the two regimes. None of the above mentioned works provides a methodology for determining, in a statistical sense, what is the conditional expected flow and the conditional probability density function (PDF) of the flow, given a density value.

In this paper, we present a statistical learning methodology for determining a probabilistic description of the fundamental diagram, obtained from traffic flow data provided from a VDS. In contrast to a traditional static piecewise affine fundamental diagram, as the one depicted in Fig. 1(a), the proposed methodology produces a stationary piecewise affine mixture of conditional probability functions of flow, given values of density, as schematically depicted in Fig. 1(b). The graph in Fig. 1(b) is the conditional expected value of the flow given density $\bar{f}(\rho) = E\{f\ |\rho\}$. Also, referring to Fig. 1(b), given a value of density ρ, there is 95 % probability that the corresponding value of flow is contained between the interval $[a,\ b]$, demarcated by the shaded region.



(a) (b)

FIGURE 1 (a) A piecewise affine fundamental diagram with its associated parameters. (b) A stationary piecewise affine mixture of conditional probability functions of flow used in this paper, with the some of its associated parameters

The parameters that characterize the probabilistic piecewise-affine fundamental diagram depicted in Fig. 1(b) are obtained from VDS traffic data, collected through many days, as shown for example by the actual traffic data graphed in Fig. 2. The graph of the conditional expected value of the flow given density $\bar{f}(\rho)$ depicted in Fig. 1(b) has three distinct piecewise affine functions as opposed the the more common graph shown in Fig. 1(a).

$$\bar{f}(\rho) = \begin{cases} v\,\rho & 0 \le \rho \le \rho_t \\ v_t\rho + \rho_t(v - v_t) & \rho_t \le \rho \le \rho_c \\ w(\rho_c - \rho) + Q & \rho_c \le \rho \le \rho_j \end{cases}$$

The density region $[0,\, \rho_t]$ corresponds to the constant speed v free-flow traffic region, where ρ_t will be called the free-flow transition density in this paper. The density region $[\rho_t,\ \rho_c]$ corresponds to the variable speed free-flow traffic region, where ρ_c is the well-known critical density. The region $[\rho_c\ ,\rho_j]$ corresponds to the congested traffic flow region, where ρ_j is the jam density. Q is the expected value of the maximum flow, also known as the capacity. w can be interpreted as the rate at which congestion propagates reverse of the flow direction of traffic at this particular point. As stated above, the shaded regions in Fig. 1(b) demarcate the 95 % confidence interval of measuring a flow value, given a density measurement. Thus, given density ρ, the lower limit a is approximately one standard deviation below the expected flow $\bar{f}(\rho)$, while upper limit b is approximately one standard deviation below the expected flow $\bar{f}(\rho)$. As will be detailed in the next sections, the conditional flow PDFs associated with both the constant speed free flow region $[0,\ \rho_t]$ and the congested region $[\rho_c\,,\rho_j]$ are exponential, while the conditional flow PDF associated with both variable speed free flow region $[\rho_t,\ \rho_c]$ is normal.

This effort to develop this new form of the fundamental diagram depends upon the use of a large number of data for a particular position. Today, there are sensors within the road network that capture this large amount of data on a regular basis. In California, these data for the freeways are stored in one location – the Performance Measurement Systems (PeMS) (7). It is from this source that all the data used here will be extracted. This source contains many years worth of information, but only a select amount will be used. Enough information is used, however, to provide a rich amount of information from which to derive the necessary parameters. It is important to note that a proper understanding of the distribution of data cannot be made without a reasonably large sample size from which to derive the distribution. PeMS provides this source.

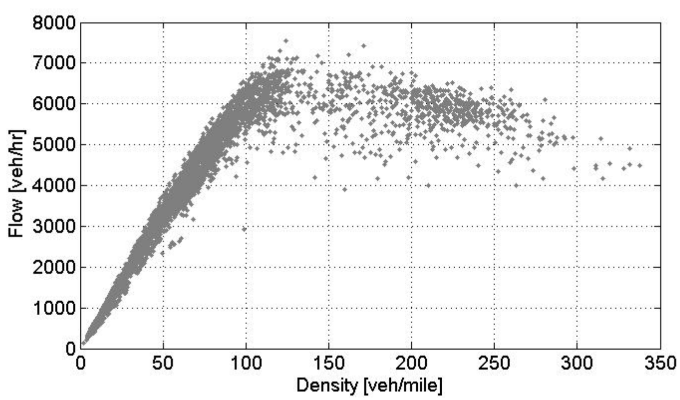


FIGURE 2 A set of real 5-min aggregate data for eastbound Interstate 80 at Richmond for the period of December 1-December 20, 2012. Each dot represents one data reading.

TRANSITION DENSITY AND FREE-FLOW DISTRIBUTIONS

To get through the first steps of establishing the form of the fundamental diagram, one approximates the free flow data to be that of data with density less than that of the data point with the highest flow value. In most cases, this will be approximately accurate, and in any case will designate a starting set to measure the necessary values. Typical in a linear model of the free flow regime is an assumption that the speed remains a constant, that is, there exists a constant free flow speed that is maintained on average by all the vehicles on the road. Non-linear models, however, do not need to make this assumption, and in fact the data does not support this assumption either. Figure 3 makes this more clear with the sample data. This is a speed-density plot of data from the free-flow regime (as approximated in the way discussed). As density increases, the value of speed tends to decay from a higher value at a steady rate.

In order to recognize the detrimental effects of moderate density on the average free-flow speed while maintaining a piece-wise linear system that keeps the parameters easily identifiable

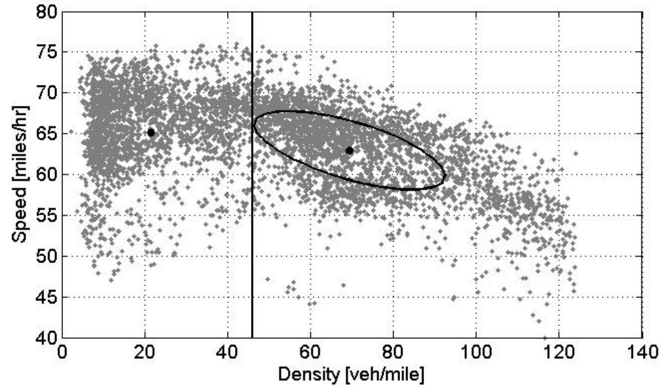


FIGURE 3 The speed-density plot of the Interstate 80 data in the free-flow regime, with the mean of the two distributions marked with solid black dots, the region at which $\Gamma_2 = 1$ as an ellipse, and the transition density location as a solid line.

features, this section will be divided into two regions. The value of density that divides the two regions will be called the transition density. This feature will identify where vehicular interference becomes more prominent, and thereby where the free-flow speed begins to degrade.

Though motivated by the nature of the data, this feature can also be considered in more theoretical terms. In 1998, Newell (*8*) described the effects of a large vehicle or convoy that was moving at a speed that was lower than the prevailing traffic. In essence, he determined that this large vehicle could be considered a bottleneck from the frame of reference of the slow moving vehicle, and that in this frame of reference, the traditional concepts of bottleneck could be applied. When considering that real data is taking averages over time at a particular location, the resulting figure when applied to a linear model is identical to that of the proposed structure. It may thereby be best to assume that what is being observed with this proposed formation is the effects of the slower moving traffic, which act as a continuous bottleneck around which the faster moving traffic has to move around. What is captured in sensors is then the average of this effect.

A calculation of the value of the transition density begins with an estimation of the distribution space that will cover all possible variation of data. The law, limitations of vehicles, and most importantly safety define an upper bound for the velocity at which vehicles can travel. Naturally, almost all valid data points will be below this upper bound. This upper bound holds the free-flow speed at a constant value for a region of low density. As the density increases, however, the moving bottleneck effect has a greater influence than the upper bound, and a more Gaussian form of distribution is identifiable. Thereby, the easiest way to consider the distribution of real data is as two different distributions in the speed-density field. The speed-density distribution that will be used to identify the data of Region 1 of Fig. 3 is given by the following exponential distribution.

$$\Gamma_1(v,\rho|\bar{v},\bar{\rho},\Sigma) = \frac{1}{(v_{max}-\bar{v})\bar{\rho}} e^{(v_{max}-v)/(v_{max}-\bar{v})} e^{\rho/\bar{\rho}} \tag{3}$$

where v_{max} is the maximum speed and $\bar{v}$ and $\bar{\rho}$ are the appropriate average parameters, and Σ is a variable included for notational purposes. Notice that Γ_1 decays in both speed and density, since the goal is to find a limited space distribution that has less of an influence with the larger density. The distribution will be used to identify the speed-density data in Region 2 of Fig. 3 is a Gaussian distribution, centered at $(\bar{v}, \bar{\rho})$ and with variance Σ:

$$\Gamma_2(v, \rho | \bar{v}, \bar{\rho}, \Sigma) = \mathcal{N}([\bar{v}, \bar{\rho}], \Sigma) \tag{4}$$

These two distributions are shown in figure 4.

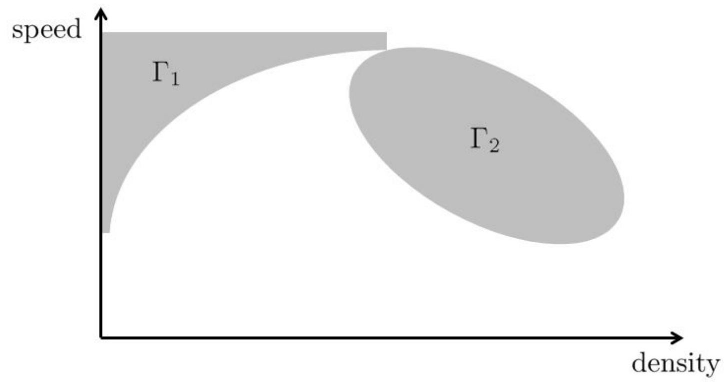


FIGURE 4 A general plot of the one standard deviation space of the two probability distributions assumed by equations (3) and (4).

Using the above two distributions as the base assumption distributions, the Expectation-Maximization (EM) algorithm from (9) for Gaussian mixtures will be adapted to define the following results. The modification in distribution will not effect the convergence of the algorithm because the equation set used here continues to have only one solution. Let π_i be the unconditional probability that a data point is in set i for $i = 1, 2$. Then the conditional probability,τ_n^i of a data point $x_n = [v_n, \rho_n]^T$ being a part of set i given parameters $\mu_i = [\bar{v}_n, \bar{\rho}_n]^T$ is

$$\tau_n^i = \frac{\pi_i \Gamma_i(x_n | \mu_i, \Sigma_i)}{\sum_j \pi_j \Gamma_j(x_n | \mu_j, \Sigma_j)} \tag{5}$$

Through this value, the mean and the appropriate covariance matrix for each distribution (as needed) can be determined

$$\mu_i = \frac{\sum_n \tau_n^i x_n}{\sum_n \tau_n^i} \tag{6}$$

$$\Sigma_2 = \frac{\sum_n \tau_n^2 (x_n - \mu_2)(x_n - \mu_2)^T}{\sum_n \tau_n^2} \tag{7}$$

where Σ_1 is a variable included for clarity to the equations and has no meaning. These values of conditional probability can in turn be used to determine the unconditional probability

$$\pi_i = \frac{1}{N} \sum_n \tau_n^i \tag{8}$$

where N is the number of data points. Using equations (5)-(9) in iteration from an initial estimation of π_i provides an EM algorithm that can converge to the appropriate solution. From the determination of these two sets, the transition density can be identified as the location where the Gaussian distribution, Γ_2 has a significant contribution. In the case of this work, this location has been approximately identified as

$$\rho_t = \min(\rho|\Gamma_2(v, \rho) = 0.1) \tag{9}$$

The equation is highly likely to have a solution because Γ_2 is a Gaussian distribution and thereby always decreasing to zero as $(v, \rho) \to \infty$ from a maximum value at the mean given valid parameters. So given that the value of Γ_2 at the mean is greater than 0.1, continuity implies a solution. From a more intuitive standpoint, the region outside the ellipse generated by $\Gamma_2(v, \rho) = 0.1$ contains all the points with $\Gamma_2(v, \rho) < 0.1$. This means that for all of these points, the probability of a value of from the distribution approaching a small neighborhood ΔA of these points is less than $0.1\Delta A$. What is being decided here is that this qualifies as not a significant contribution to the distribution.

FREE-FLOW REGIME - LINEARIZATION

Now that the transition density has been identified, the actual linearization of the free-flow region must be considered. Typically, this is done through a basic linearization formula. The assumption made with this formula is that the data is distributed according to the formula

$$f_i = v \cdot \rho_i + \epsilon \tag{10}$$

where ϵ is a random, Gaussian distributed variable. Eq. (11) is not valid in this situation, because the distribution is not Gaussian. Instead, as discussed previously, the distribution contains an upper bound which cannot be passed. In order to recognize the bound and use it to create a more accurate model of the distribution, the previous model described in Eq. (3) could be used to determine $\bar{v}$ and applied to this region. Since this distribution was created with some uncertainty in whether these data were contained in the distribution, however, this may not provide the value of v which is being sought here because this value ought to be derived from the deterministic presence of all the data points in this region. Instead, a different kind of linear fit on the flow-density data will be attempted using the following assumption: In the region where density is below the transition value, there exists an upper bound on the speed, v_{max}. Using the values $\tilde{v}_i = v_{max} - v_i$, the set in this region has a value that is approximately an exponential distribution in flow. That is, given a value of the density, ρ, the distribution is approximately exponential with a mean at $\tilde{v}\rho$.

It is a straightforward process to determine the value of $\tilde{v}$ from this assumption. The exponential distribution provides the context of a generalized linear model. At this time, it is worth discussing a simple algorithm to determine a value of $\tilde{v}$ and part of the proof of this equation. The goal in the present linearization is to maximize the likelihood that a given value of $\tilde{v}$ is the correct value. As noted in (*10*), the log-likelihood of any particular value of $\tilde{v}$ given the data set $\{(\rho_n, f_n)\}$ is

$$\begin{aligned} l(\tilde{v}) &= \log \mathcal{P}(\tilde{v}|\{(\rho_n, f_n)\}) && (11) \\ &= \log \prod_{n=1}^{N} \exp(\eta_n \tilde{f}_n - A(\eta_n)) && (12) \\ &= \sum_{n=1}^{N} (\eta_n \tilde{f}_n - A(\eta_n)) && (13) \end{aligned}$$

where $\eta_n = -\nu_n^{-1}$, $\nu_n = \tilde{v}\rho_n$, $A(\eta_n) = -\log(-\eta_n)$, and $\tilde{f}_n = v_{max}\rho_n - f_n$ with the flow f_n, as is the structure of this exponential distribution.

Taking the derivative with respect to $\tilde{v}$,

$$\frac{dl}{d\tilde{v}} = \sum_{n=1}^{N} \frac{dl}{d\eta_n}\frac{d\eta_n}{d\tilde{v}} \tag{14}$$

$$= \sum_{n=1}^{N} (\tilde{f}_n - \nu_n)\frac{d\eta_n}{d\nu_n}\rho_n \tag{15}$$

$$= \sum_{n=1}^{N} (\tilde{f}_n - \nu_n)\frac{1}{\nu_n^2}\rho_n \tag{16}$$

This gradient provides information about approaching the optimum value. An algorithm that follows the gradient will tend to approach the maximum value of likelihood, which is the value desired. Typical of an on-line algorithm of this form is

$$\tilde{v}_{t+1} = \tilde{v}_t + \gamma(\tilde{f}_n - \nu_{nt})\nu_{nt}^{-2}\rho_n \tag{17}$$

where $\nu_{nt} = \tilde{v}_t\rho_n$ and γ is a step size. (*10*) Repeated iterations should approach the correct value of $\tilde{v}$ from which v can be derived.

The remaining region between the transition density and the approximate location of critical density is more reasonably a Gaussian by the structure of Eq. (4) that was proposed earlier. Thereby, a calculation of the line in this region, starting at the end point of the previously determined line (ρ_{end}, f_{end}), can take advantage of the basic linear equation. Let $R = [\rho_1 - \rho_{end}, \rho_2 - \rho_{end}, ..., \rho_N - \rho_{end}]^T$ be the density data values of this region written in vector form and reparametized and $F = [f_1 - f_{end}, f_2 - f_{end}, ..., f_N - f_{end}]^T$ be the flow data values in the same order in vector form and also reparametized. Then the solution is arrived at by

$$v_t = (R^T \cdot R)^{-1} \cdot R^T \cdot F \tag{18}$$

CAPACITY DETERMINATION

Having completed the free-flow region, the capacity of this model has to be identified. The most simple solution would be to take the largest value of flow provided by all of the data and declare it, or some percentage of it, as the capactiy, Q

$$Q = \max(f_n) \tag{19}$$

The problem with this value of Q is the non-representative data. Because these values are being determined from actual data, there is a chance that some data will be above what would be considered the nominal capacity. These data are isolated from the general trend of the data because they mark a particular moment when traffic conditions reached an unusually high value of flow, which would not be a valid point to consider when asking about the general capacity of the road. Given that data sets are relatively dense, however, it is still possible to exclude those points for more

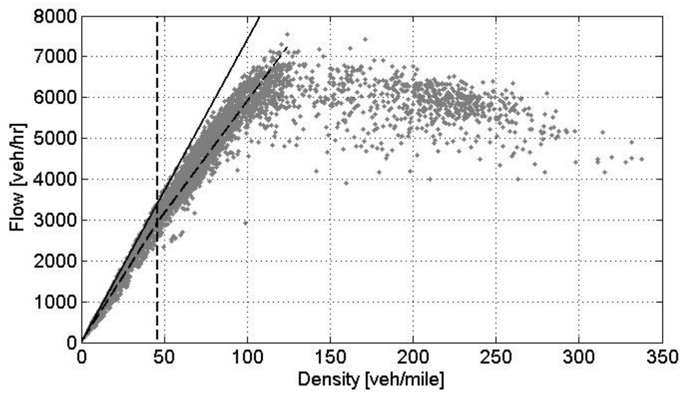


FIGURE 5 The nominal value in the free flow regime shown with the lower line, with the upper bound with the upper line. The transition density value is indicated by the vertical line.

reasonable, lower values of capacity. Given that $x_{max} = [\rho_{max}, f_{max}] = \arg\max(f_n)$, consider the value of

$$\min_{x_{max} \neq x_j} dist(x_{max}, x_j) = \min_{x_{max} \neq x_j} ||x_{max} - x_j|| \overset{?}{>} a \tag{20}$$

where the value of a is pre-set and dependent on the number of data points available. If this equation proves to be true, then the data point falls outside the general range of data, and should therefore be ignored during the calculation of the capacity. The value of capacity should be determined again from the remaining data points.

Again, however, there is the problem of the capacity at this point being the maximum of the majority of the data. The point being a maximum is a problem for this work because the determinant line so far has attempted to follow the nominal value of the data. Given that the values are at a peak in this region, there will be effects upon the congestion regime, as well as a loss in the nominal value of the system. To properly consider all of this, the nominal capacity is intentionally reduced from the maximum data point by some small percentage. Moreover, the range of about 100 veh/hr less in flow that the maximum data point is a reasonable qualification, which is generally a little more than one percent of the maximum flow. From practice then, the most obvious way to develop the nominal point is to declare that

$$Q = 0.98 \max(f_n | \min_{x_n \neq x_j} ||x_n - x_j|| \leq a) \tag{21}$$

Figure 5 shows the results of the calculation of sections 2 and 3. This figure defines the nominal value line in the free-flow regime, as well as the end point of this regime. There is also an upper bound that defines the variation of the data in an understandable fashion. The nominal value of the data becomes the starting point for the congestion regime. Note the extremely high value of the upper limit compared to the actual data near with density near critical. The error supports the breakdown of the exponential distribution, that a Gaussian distribution whose variation is derived from the variation of the data would more accurately model the distribution in this region.

CONGESTION REGIME

Given that the free-flow expected deterministic function has been determined, as well as the capacity, for this paper the congestion region will be defined as all data points that are greater than the density at capacity, that is, the critical density. Unlike the free-flow regime, there is no obvious reason for an upper bound of flow on the congestion region. However, congestion describes the deteriorating conditions of the road, and thereby there is a general upper bound on most cases on these data. Given adequate data from the congestion regime, one can make the assumption that like in the free-flow case, traffic conditions never exceed a state above this upper bound. There may be conditions, for example an unexpected bottleneck in the road, that would cause traffic to come to a state far below this upper bound, and these conditions cannot be ruled out. Thereby, it seems reasonable to use as in the free-flow case an exponential distribution as the distribution of the random variation.

Given this conclusion, the upper bound is defined as the line connecting the maximum flow data point with the maximum flow data point among the data points with the ten largest density values. This should be a reasonable approximation, with data points above this line ignored.

An initial point at the critical density can be, and in this case will be, defined. It may be useful not to define this point and allow for the existence of a capacity drop when entering the congestion regime. The equations are approximately the same, except for the existence of a two-dimensional unknown vector, instead of a one-dimensional vector. The existence of this initial point also emphasizes caution. Because data in the congested region tend to be towards lower density, this can distort the slope of the congestion line if started with poor initial conditions. The value of capacity as the initial condition at the critical density has proven to produce reasonable results.

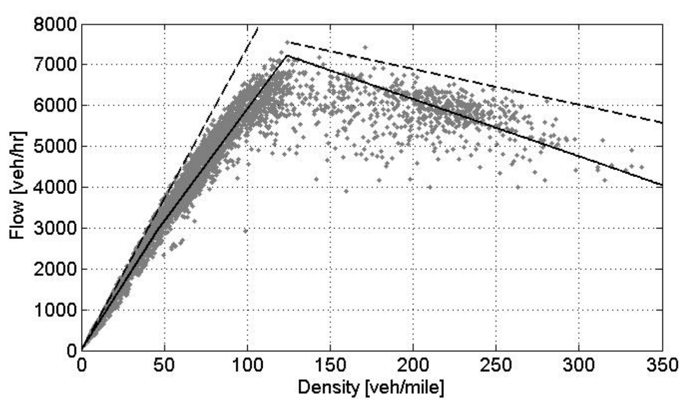


FIGURE 6 The completed fundamental diagram, with the upper bounds defined.

To use an algorithm similar to Eq. (16), some algebraic formulation will be applied to the data. Let the upper bound line be given by the equation $f = -w_u\rho + c_u$. Let the capacity be Q and the critical density be ρ_c. Then the iterated algorithm to find the slope in the congested region, $-w$, is

$$\tilde{w}_{t+1} = \tilde{w}_t + \gamma(\tilde{f}_n - \mu_{nt})\mu_{nt}^{-2}(\rho_n - \rho_c) \tag{22}$$

where γ is a constant step-size,

$$\tilde{w}_t = -w_u + w_t \tag{23}$$

$$\tilde{f}_n = -w_u\rho_c + c_u - f_n \tag{24}$$

$$\mu_{nt} = \tilde{w}_t(\rho_n - \rho_c) - Q - w_u\rho_c + c_u \tag{25}$$

This section completes the description of the fundamental diagram. The results applied to the sample data are shown in figure 6.

EXPERIMENTAL RESULTS

Using the above described steps, the fundamental diagram can be produced for any number of sensor locations. To confirm that such results are accurate for the state of California, several VDS data sets from locations around the state are shown in figure 7. Note that these locations do not necessarily have the perfect measure of sensor data, but a fit can be made to the data that is given.

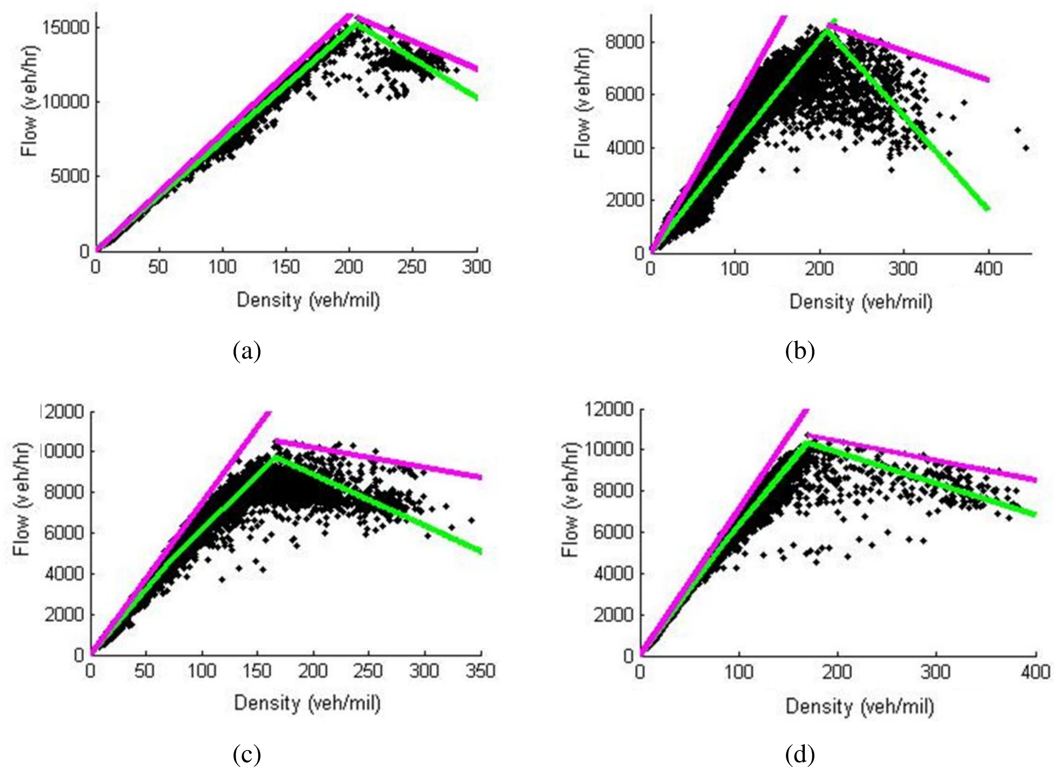


FIGURE 7 Four other sample locations, in order, Interstate 15 southbound near Escondito, State Route 99 northbound in Sacramento, Interstate 210 westbound near Pasadena, State Route 101 northbound near San Francisco.

These data sets contain one month worth of 5-minute data points, over the period of December 1 to December 31, 2012. The data were aggregated from individual lane data to aggregate flow and density parameters. What results is a large number of variations to the pattern of the data, and consequently a wide variation to the value of the parameters. In some cases, the final distance between the upper boundary and the nominal value, which is one standard deviation of the exponential distribution, is within 2000 veh/hr, which indicates that the potential variation of the parameter w is relatively low and the amount of confidence of the nominal value is high. This is true of figures 7a and 7d. There are some situations, however, where the final distance is nearly 3000 veh/hr or larger, where the range of potential values for the parameter w is quite larger, and confidence on the nominal value given cannot be so high. The large variation measure is usually an indication of wide variation in data more than a fault of the algorithm. And in such a situation the algorithm can be useful, because the uncertainty is made obvious from this calculated distance and considered in the model. The visibility of large variation is true of figures 7b and 7c.

CONCLUSION

The paper proposed a new method of estimating a fundamental diagram model that would allow for an expected value deterministic structure while also providing potential identification of variation from the expected value. It used fittings that attempted to be realistic to the provided data and conveyed concisely all the information of the flow-density data within several small parameters. It also allowed the possibility of identifying and retaining knowledge of the upper bounds as a form of variation. Since the variation is taken to be of an exponential distribution, the upper bounds and the mean values would provide information about the standard deviation, and thereby the variance of the data. The model can thereby be both deterministic and probabilistic, and can be used in either context depending on need.

Clearly, the discussion contained here is only the start of the study of the model. A further investigation could be made on how to implement this model with variation onto a simulation. The simulation would add the time element that has been mostly assumed inconsequential in this paper. Another potential direction is through investigating how changes in the environment can change the variation of the data, for example whether it is light or dark, the effects of limited visibility, and precipitation. Likely, there will be an effect on the fundamental diagram, and with the given information, it might be enough to make predictions when the given event occurs in the future.

As traffic congestion continues to be a problem, it is important to be able to make predictions about how traffic will behave in the near future. With this additional knowledge of the fundamental diagram, there is potential to build up a better model to make these predictions of the future.

ACKNOWLEDGEMENTS

This work is partially supported by the California Department of Transportation (Caltrans) through the Connected Corridors California PATH Program and by the National Science Foundation (NSF) through grant CDI-0941326.

REFERENCES

[1] Daganzo, C. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research, Part B*, Vol. 28, No. 4, 1994, pp. 269–287.

[2] Lu, S. W. e., Yadong. Explicit construction of entropy solutions for the Lighthill-Whitham-Richards traffic flow model with apicewise quadratic flow-density relationship. *Transportation Research Part B:Methodological*, 2008.

[3] Del Castillo, F. B., J.M. On the functional form of the speed-density relationship – I: General theory. *Transportation Research Part B:Methodological*, 1995.

[4] Kerner, B. *The Physics of Traffic*. Springer Science and Business Media, Berlin, 2004.

[5] Kim, H., T. A stochastic wave propagation model. *Transportation Research Part B:Methodological*, 2008.

[6] Sumalee, R. Z. T. P. W., A. Stochastic Cell Transmission Model (SCTM): A stochastic dynamic traffic model for traffic surveillance and assignment. *Transportation Research Part B:Methodological*, 2011.

[7] *Caltrans Performance Measurement Systems*, 2012.

[8] Newell, G. A Moving Bottleneck. *Transportation Research Part B:Methodological*, 1998.

[9] Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer Science and Business Media, New York, 2006.

[10] Wainwright, M. *An Introduction to Probabilistic Graphical Models*, 2012. Course Reader, Electrical Engineering 281A, University of California, Berkeley.