Peer-review

Park, Lauren. 2025. "Evaluating Macroscopic Traffic Flow Models Using Real Vehicle Trajectory Data." *Journal of High School Science* 9 (1): 52–73.

1.I am unsure as to whether the model you construct is applicable to a range of empirical values that comprises a small subset of the model's boundry conditions (as you yourself mention in the manuscript). Therefore, you either need to apply your model to empirical conditions that encompass the entire model's range or construct a simpler model that will more accurately predict your narrow range of studied parameters. I don't think that your model's mismatch with empirical conditions contributes significantly to the existing corpus of knowledge in this field, despite your describing this as a limitation in the manuscript. Is it possible for you to obtain traffic data from public domain that *does* encompass the entire range of your model? I am sure there must be other datasets in the public domain that are populated with a broader range of data. If not, please provide a simple linear model.

2.Your predicted N(t) looks like it would be superimposble on the velocity-time plot in figure 3. Are you sure you are calculating the number of vehicles and not velocity? Maybe you overlooked to multiply the ($-0.0888\ k(t) + 54.6691$) by k(t) in the equation: $Q(N) = k(t)(-0.0888\ k(t) + 54.6691)$? Please double check your calculations and equations.

3.There are several instances where the time interval is mentioned as "… [7:50:00, 7:35:00]….". Please correct all these instances to 7:50 to 8:35.

4.What is the difference between figures 11 and 14 ? Do they not present the same information ? If not, please describe the difference in the manuscript.

5.I did not understand what was unrealistic in the following paragraph "……Another limitation of our Greenshields flux function is that it implies that the maximum average velocity is 87.98 kph (54.67 mph) which would occur when there are very few vehicles in the study area. This too seems unrealistic since the legal speed limit is 104.61 kph (65 mph)…..".

6.You mention in the manuscript that "…..However, since the data set was collected during the most congested daily time period, it is unlikely that observations could be made when densities are higher than what was observed in the data set…..",yet you also mention "….This covers the majority of the first period of traffic (uncongested phase) which is 900 seconds long. But, beyond that point, the error in prediction generally continues to grow throughout the next two periods, the transition from uncongested to congested conditions phase and the full congestion peak phase…." How do you reconcile these two contradictory statements drawn from the same dataset ?

Dear Reviewer,

Thank you for reviewing my paper. After carefully considering your feedback, I have revised certain parts of the paper and responded to individual comments below. For certain comments, I have highlighted parts of the comment and my response to clearly show which parts of the feedback I am addressing.

Respectfully,

Lauren Park

Rating scale questions

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The paper makes a significant contribution to scholarship.					
The literature review was thorough given the objectives and content of the article.					

Please see my response to comment #1 below. I explain why my paper is a significant contribution to scholarship and I also found a published paper that meaningfully adds to the literature review of my paper.

Open response questions

Comments to author

1. I am unsure as to whether the model you construct is applicable to a range of empirical values that comprises a small subset of the model's boundary conditions (as you yourself mention in the manuscript). Therefore, you either need to apply your model to empirical conditions that encompass the entire model's range or construct a simpler model that will more accurately predict your narrow range of studied parameters. I don't think that your model's mismatch with empirical conditions contributes significantly to the existing corpus of knowledge in this field, despite your describing this as a limitation in the manuscript. Is it possible for you to obtain traffic data from public domain that *does* encompass the entire range of your model? I am sure there must be other datasets in the public domain that are populated with a broader range of data. If not, please provide a simple linear model.

I looked extensively for traffic datasets that encompassed a wider range of densities. As far as I know, such datasets for a single road study area are not publicly available. The NGSIM datasets were the best available because they had detailed traffic data collected at high time frequency which make numerical simulations meaningful. It would be very useful if NGSIM had collected additional data for the study area during other times of the day, for example, 6:00 AM - 6:45 AM (very low traffic density), and 4:30 PM - 5:15 PM (peak evening commute time). Unfortunately, wider range datasets are not publicly available.

The model developed in the paper is linear. Macroscopic traffic flow models require identifying a relationship between traffic flow density and traffic flow velocity in order to (numerically) solve a PDE/ODE. In my paper, I derived a simple linear relationship between velocity and density which is depicted in Figure 5, Figure 6, and equation (5). (Because there is a slight curvature in the empirical data, I had considered using a nonlinear function, such as a cubic-spline, but decided against it. Equation (5)'s R^2 value is 0.9316.)

Other researchers have also utilized the NGSIM datasets for their research papers. I did more literature search and found the following paper:

[12] Fan, S., & Seibold, B. (2013). Data-Fitted First-Order Traffic Models and Their Second-Order Generalizations: Comparison by Trajectory and Sensor Data. Transportation Research Record, 2391(1), 32-43. https://doi.org/10.3141/2391-04

This paper utilizes an NGSIM dataset for a stretch of US 80 freeway near Berkeley, California, which is similar to the study area in my paper. I don't have access to the paper, but I was able to find a preprint on arXiv: https://arxiv.org/abs/1208.0382, which I have included in my reply. The time periods for the US 80 dataset are from [4:00 PM, 4:15 PM], [5:00 PM, 5:15 PM], and [5:15 PM, 5:30 PM]. This is a relatively narrow period of time which does not encompass a wide range of empirical traffic conditions for the study area.

Still, the various models developed in [12] are simulated using the NGSIM dataset. See page 13 of [12]. Interestingly, one of the traffic flow density to velocity relationships in [12] is linear (the Greenshields flux function) which is identical to what is used in my paper. See Figure 2 on page 7 of [12].

I have added [12] onto my paper's list of references since it is similar to the investigation done in my paper. When I did my original background research, I looked at an extensive number of papers, but only listed those that I felt were directly applicable to the work presented in my paper.

Finally, I have slightly revised my paper to clearly separate out *further investigations* from the *discussion*. (See page 19.) I should have done that when I submitted my paper. The latter part of the discussion section was intended to identify how to continue to improve the model. "Recommendations for Future Research" is now a separate section of my paper.

Although the model that I developed did not super accurately reproduce the empirical data, the development of the model, and the analysis of its performance and how to potentially improve it are still valuable. The methodology of utilizing the dataset to calibrate the model, numerically simulate the model, and analyze the model's performance contributes new and meaningful knowledge to macroscopic traffic modeling.

2. Your predicted N(t) looks like it would be superimposable on the velocity-time plot in figure 3. Are you sure you are calculating the number of vehicles and not velocity? Maybe you overlooked multiplying the $(-0.0888\, k(t) + 54.6691)$ by k(t) in the equation: $Q(N) = k(t)(-0.0888\, k(t) + 54.6691)$? Please double check your calculations and equations.

I have carefully reviewed my calculations, equations, and graphs to confirm that they are correct.

a. Figure 3 depicts the empirical velocity.
b. Figure 10 depicts the predicted N(t) starting from 7:5**0**:00 AM.
c. Figure 12 depicts the predicted N(t) starting from 7:5**1**:00 AM.

I have pasted Figure 3 and Figure 10 below for easier comparison to confirm that they are different.

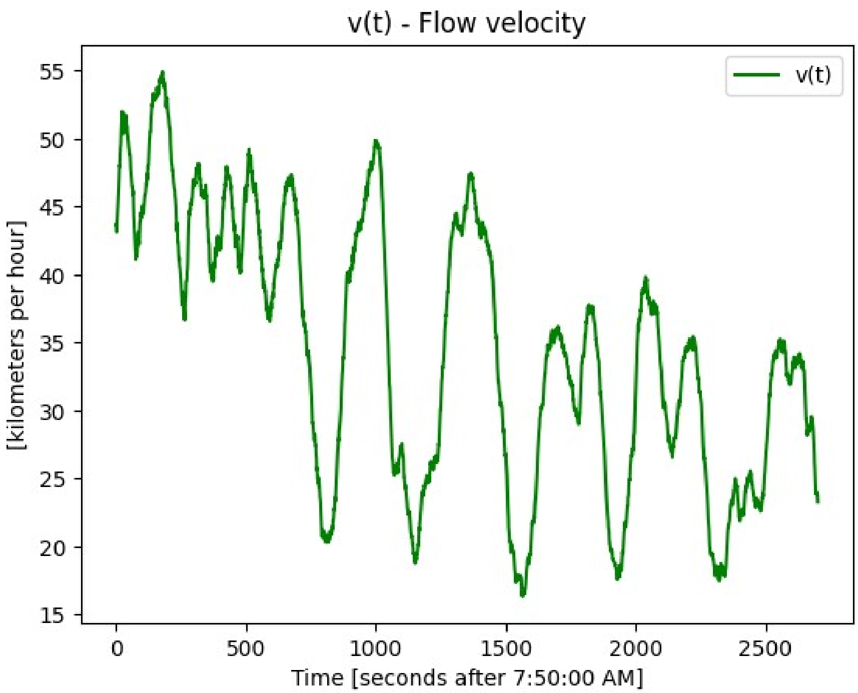

Figure 3. $v(t)$, the flow velocity, the average velocity of vehicles over 1 second intervals.

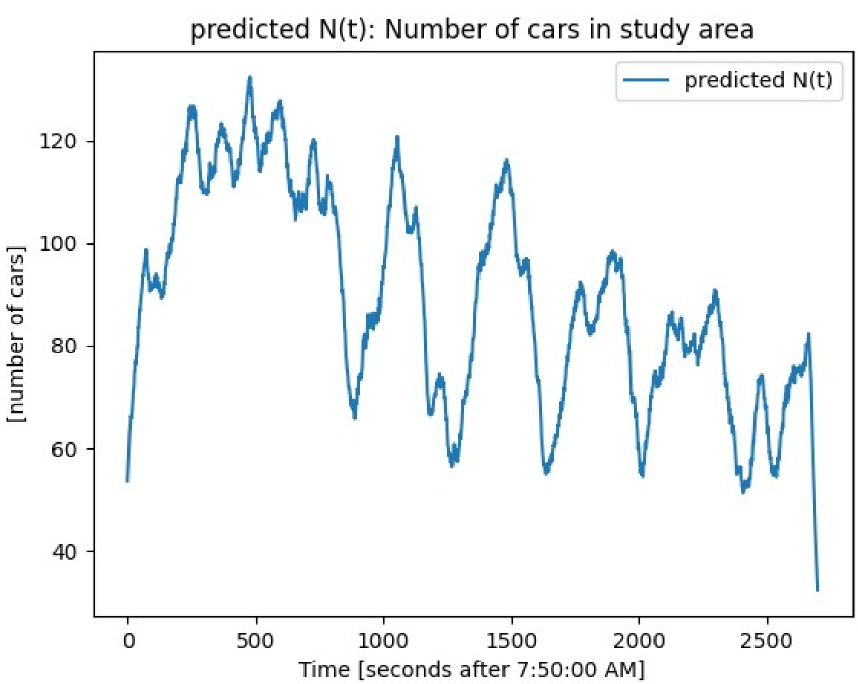

Figure 10. *predicted* $N(t)$, the number of vehicles in the study area predicted by the model.

Note that the maximums and minimums of the two graphs do not align. That is expected. As the velocity increases, the number of cars in the study area (density) should go down - and vice versa. Additionally, if Figure 3 (empirical $v(t)$) is compared to Figure 9 (*empirical* $N(t)$), the inverse relationship is also observed, as expected.

For completeness, I've included Figure 12 below. Figure 10 [seconds after 7:5**0**:00] and Figure 12 [seconds after 7:5**1**:00] are almost identical except that Figure 12 excludes the initial minute (the starting ramp up period of the data collection).

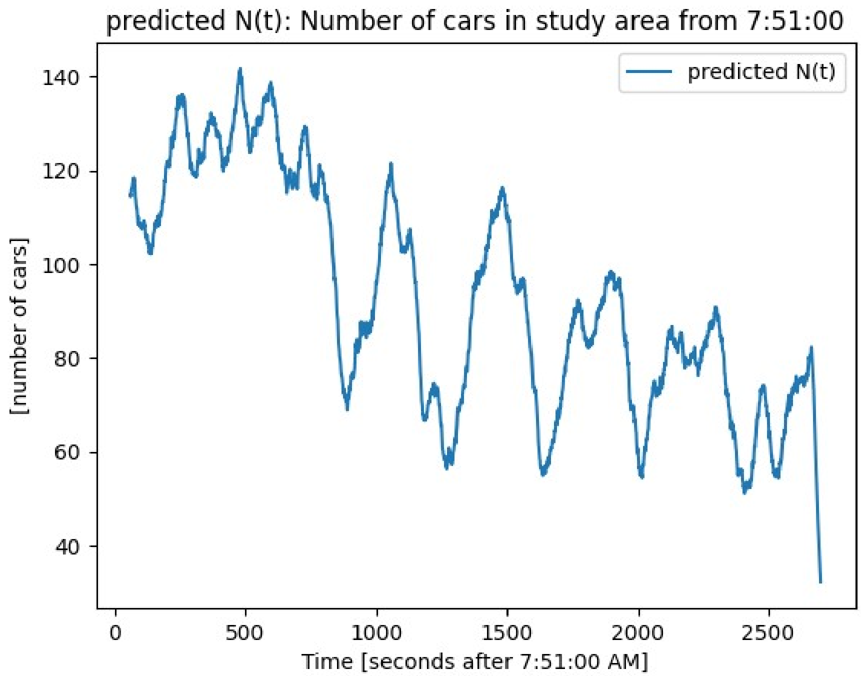

Figure 12. *predicted* $N(t)$ for [7:51:00 AM, 8:35:00 AM].

3. There are several instances where the time interval is mentioned as "… [7:50:00, 7:35:00]….". Please correct all these instances to 7:50 to 8:35.

Thank you for pointing this out. All time interval instances have now been corrected to properly end at … 8:35:00]. Six typos have been corrected.

4. What is the difference between figures 11 and 14 ? Do they not present the same information ? If not, please describe the difference in the manuscript.

Figures 9, 10 and 11 depict information related to the simulation from [7:5**0**:00, 8:35:00].
Figures 12, 13 and 14 depict information related to the simulation from [7:5**1**:00, 8:35:00].

I have added additional sentences to clearly call out that the results of two numerical simulations are being presented in the paper. (See pages 14 and 15.)

5. I did not understand what was unrealistic in the following paragraph "……Another limitation of our Greenshields flux function is that it implies that the maximum average velocity is 87.98 kph (54.67 mph) which would occur when there are very few vehicles in the study area. This too seems unrealistic since the legal speed limit is 104.61 kph (65 mph)…..".

I should have stated clearly that the maximum average velocity should be closer to the speed limit, or likely higher than the speed limit, when traffic density is very low. I have added additional sentences to say this clearly. (See page 18.)

For what it is worth, there is a California state government observation website called Sigalert which reports traffic speeds. The SigAlert website URL for the Southern California area is https://www.sigalert.com/map.asp?lat=34.11304&lon=-118.32408&z=0. The southern end of the study area of this paper is Cahuenga Blvd on US101 Southbound. This location is actually a reporting point for the Sigalert URL above. Cahuenga Blvd point is near the middle slightly above the label “West Hollywood.” If you observe the reported traffic speeds at Cahuenga Blvd during low traffic density time periods (e.g., 7:00 AM (early morning) or 12:30 AM (late at night)) you will routinely observe that the traffic speed is 66 mph.

6. You mention in the manuscript that “…..However, since the data set was collected during the most congested daily time period, it is unlikely that observations could be made when densities are higher than what was observed in the data set…..”,yet you also mention “….This covers the majority of the first period of traffic (uncongested phase) which is 900 seconds long. But, beyond that point, the error in prediction generally continues to grow throughout the next two periods, the transition from uncongested to congested conditions phase and the full congestion peak phase….” How do you reconcile these two contradictory statements drawn from the same dataset ?

Thanks for pointing this out. It does appear that these statements seem contradictory, but they are not. I have reworded this to be more clear. (See page 20.)

Additionally, I have provided an example of what I meant here. I am just pointing out that it is unlikely that observations at other time periods during the day would find traffic densities that are higher than those recorded in the dataset during the study period [7:50 AM, 8:35 AM], which covers the peak morning rush hour commute time. I have added an example that clarifies the text in my paper. (See page 20.)

Thank you for addressing my previous comments. I have some additional comments

1.Below equation 5, you state that the max average velocity is 54.67 miles per hour. However, your graph has units of km/h and I see a point of intersection at approx. 54 km/h. Please check units and the possibility of incorrect filtration through to other calculations in your equations and predictive modeling. For example, this number appears again in “….Another limitation of our Greenshields flux function is that it implies that the maximum average velocity is 87.98 kph (54.67 mph) which would occur when there are very few vehicles in the study area…..” Please check the manuscript thoroughly for implications. I have not checked your other calculations scattered throughout the manuscript that use this incorrect unit and therefore incorrect number.
2.Figure 8: can you plot the points over a 10 second (instead of 1 second) interval on the x axis ? this will make the graph much more readable and understandable. As of now, not much can be interpreted from it because of the x-interval being so small.
3.References should be sequentially numbered in the body of the text. As of now, reference 9 is the starting reference. Please check and renumber as appropriate. You will need to renumber the references in the ‘references’ section as well to correspond.
4.You list 3 ways in which prediction accuracy can be improved. I would like to see you present at least one of the three ways (polynomial, DN flux function or LWR traffic model and see if any one of those models the empirical data better.

Dear Reviewer,

I apologize for the delay in response. I have answered all your questions below.

To answer your question 4, I’ve had to significantly revise my paper. I am submitting two copies of the revised paper. The first is a clean copy. For the second copy I’ve done my best to highlight sections where there have been any changes/additions from the original paper.

Thank you for reviewing my paper.

Open response questions

Comments to author

Thank you for addressing my previous comments. I have some additional comments

- Below equation 5, you state that the max average velocity is 54.67 miles per hour. However, your graph has units of km/h and I see a point of intersection at approx. 54 km/h. Please check units and the possibility of incorrect filtration through to other calculations in your equations and predictive modeling. For example, this number appears again in "….Another limitation of our Greenshields flux function is that it implies that the maximum average velocity is 87.98 kph (54.67 mph) which would occur when there are very few vehicles in the study area….." Please check the manuscript thoroughly for implications. I have not checked your other calculations scattered throughout the manuscript that use this incorrect unit and therefore incorrect number.

Because there are new equations in the revised paper, in this document, I've labeled the equations as (old paper | revised paper).

Equation (5 | 6) in the old paper was written in miles per hour. I have corrected equations (5 | 6) and (6 | 7) to be in km/h. For completeness, here is equation (5 | 6) in both sets of units.

$$v(t)[km/hr] = -0.23\,k(t) + 87.9794 \quad (5 \mid 6)$$

$$v(t)[mi/hr] = -0.0888\,k(t) + 54.6691 \quad (5 \mid 6)$$

The NGSIM data were all recorded using English units. All the work was done in English units and then I found out that JHSS required numerical information to be in MKS units. I converted all variables and data into MKS, but I missed revising equation (5 | 6).

I have verified that Figure 6 and Figure 5 are both correct and in MKS units. Note that both plots are "zoomed in" to show more detail of layout of the data points. (Otherwise the plot would show mostly empty space with the scaled down data points off to one side.) The lower left corner of Figure 6 is approximately (10, 150). Thus, you cannot directly read the y-intercept from the plot.

- Figure 8: can you plot the points over a 10 second (instead of 1 second) interval on the x axis ? this will make the graph much more readable and understandable. As of now, not much can be interpreted from it because of the x-interval being so small.

In the revised paper, Figure (| 13) has been added, a plot of u(t) averaged over 10 seconds. Figure (8 | 12) has been kept, because u(t) averaged over 1 second is what is used in the simulations.

- References should be sequentially numbered in the body of the text. As of now, reference 9 is the starting reference. Please check and renumber as appropriate. You will need to renumber the references in the 'references' section as well to correspond.

In the revised paper, all references in the body text have been corrected and also renumbered in the 'references' section to correspond.

- You list 3 ways in which prediction accuracy can be improved. I would like to see you present at least one of the three ways (polynomial, DN flux function or LWR traffic model and see if any one of those models the empirical data better.

I chose to implement the Daganzo-Newell flux function, which required some significant revisions to the paper. And, the Daganzo-Newell model did have better prediction accuracy than the Greenshields model.

Traffic researchers always use a flux function that is based upon some physical intuition regarding traffic flow. Thus, a polynomial model isn't standard practice.

I cannot implement LWR because I don't have access to a Riemann solver. My research mentor, Jacob Murri, an applied math PhD student at UCLA, advised that I just work with an ODE model that can be solved in Python's SciPy.

Thank you for addressing my comments. Please address some minor comments below:
1.you state "….A second path would be to improve the mathematical modeling of the macroscopic variables, ……Methods to calculate the macroscopic variables as continuous functions would be more realistic than step functions…….. I don't see how that would improve the model, because you are - in effect - interpolating the same data with a smaller step; the function still describes the same mathematical relationship.
2.you state "…..Finally, the empirical data provides detailed location information about the vehicles inside of the study area. Given this, it is possible to build a model that includes spatial dependence inside the study area. Thus, the ODE model (4) which utilizes variables of time, flow and volume would become a PDE that utilizes variables of time, flow, location and density. This more sophisticated model is the Lighthill-Whitham-Richards (LWR) traffic model……." This looks more like a microscopic model that will no longer be applicable to flow simulations. Are you - in effect- saying that macroscopic models are only useful for free flowing traffic and not for peak-time traffic? It may also be that 2100 feet is not enough road span to collect traffic data that will be fed into macroscopic flow models (I realize there may not be data from larger highway spans). Can you provide references that used macroscopic traffic flow models and this short of a road-span? Which is probably why your points are clustered together and you could either recourse to your 'first path' to improve the model or simply collect data over larger spans of highway. Please discuss and include this under your 'first path' recommendation.
3.The references must be in curved parantheses. Please change. To provide differentiation from equations, please insert the word 'equation' in front of all the numbers that describe equations.
4.You actually have the opportunity to build a database of 'synthetic' data points from these two models that incorporates the portions you worked with using actual data. You can upload that to kaggle, where it wil lbe useful for many researchers. There are some synthetic traffic flow datasets on Kaggle but they do not seem to incorporate actual data (at least from a part of the describing function). Obviously, you need not respond to this point in your response.

Dear Reviewers,

Thank you for reviewing my paper. My answers are in blue below.

Comments to author

This paper provides a comprehensive describtion to the methodology, with effective experimental results. It essentially allows readers to clearly understand the objective, methodology, and the process and analysis of the experiments. However, some sentences in the text need improvement.
In the abstract, experimental results are often seen as crucial, as they allow readers to assess the effectiveness of your solution. Therefore, please use a clear and straightforward style to highlight your contribution.
eg. "The predicted volumes were
compared to the observed empirical volumes to evaluate a model's predictive accuracy. The
models are also compared to each other.
You can change it to :
The difference between predicted and observed volumes was used to evaluate predictive accuracy, achieving at (real ACC). In addition, The models were also compared to each other under[eg, (how many scenarios) varying parameter conditions], to display the generalization of the model.

I have revised the abstract. I also added an average error function in the paper (see new equation 9), and calculated its value for each simulation. However, the values don't convey intuitively obvious information. Hence I described the experimental results without directly quoting the average error values in the abstract. For each simulation I did include the average error function value in the results section.

Comments to author

Thank you for addressing my comments. Please address some minor comments below:

- you state "….A second path would be to improve the mathematical modeling of the macroscopic variables,……Methods to calculate the macroscopic variables as continuous functions would be more realistic than step functions…….. I don't see how that would improve the model, because you are - in effect - interpolating the same data with a smaller step; the function still describes the same mathematical relationship.

In macroscopic traffic models, the empirical macroscopic variables are continuous functions of time and space. This assumption is what allows us to write an ODE or PDE that relates between the variables. In my paper, I data model every empirical macroscopic variable as a step function. In the majority of cases the time interval is 1 second: Raw observations that occur in a time interval are averaged or accumulated to calculate the function's value during that interval. So for example, $k(t = 11.6) = k(t = 11.9) = k(t = 11.95) = k(t = b)$ for all b in the interval [11.0, 12.0).

During a numerical simulation, the solver will evaluate the functions in the equation for many different values of t. Because the empirical macroscopic variables are all step functions there will be numerous jump discontinuities during the simulation, e.g., when going across [11.99, 12.01], [123.99, 124.01], [2,777.99, 2,778.01], …. The jump discontinuities may be "small" and not be significant in the calculations of the numerical simulations at the time interval scales as done for my paper. But, that is because the time interval scale of 1 sec was intentionally selected to reduce the size of jump discontinuities. (Alternatively consider if I had chosen to simulate at 1 minute time intervals, then macroscopic variables would be averaged over 1 minute. This may result in significantly greater jump discontinuities.)

If selecting smaller time intervals results in smaller jump discontinuities, you may wonder why I didn't choose a time interval even smaller than 1 second. This is because there are cases where there are not enough raw observations to utilize the selected time interval scale. For example, if I tried to use a time interval of 0.1

seconds, there are intervals in which there are ZERO observations. For example, there are no observations in the interval [11.2, 11.3]. This is also why the time interval for *q(t)* had to be 10 seconds. Thus interpolating at ever finer time interval scales may NOT be possible due to the gaps (e.g., *q(4) = q(95)* = 0) introduced, which produces more jump discontinuities. The mathematical relationships between macroscopic variables do not change at smaller time interval scales. Rather, estimating values for macroscopic variables using step functions eventually doesn't work at small enough time interval scales or big enough time interval scales.

You could linearly interpolate between time steps, i.e. instead of step functions, do connected line segments. But, there are more sophisticated methods, (e.g. Kernel Density Estimation methods – Parzen-Rosenblatt window method for example) which can transform a set of discrete observed data points to approximate a continuous function, thus eliminating jump discontinuities. This should allow numerical simulation at arbitrary time intervals.

Parzen-Rosenblatt seems to be the standard method used by applied mathematicians and data scientists. The macroscopic variables are assumed to be continuous, but in the simulation, empirical variables, such as *u(t),* are step functions. Parzen-Rosenblatt would make the empirical variables continuous, thus making the mathematical modeling more consistent and faithful to the assumptions.

Implementing it would require sophisticated coding and maybe hardware. This is because you have to calculate a weighted average of the raw observations in the neighborhood of a specific point (e.g., k(t = 17.76)), which means that my code would have to have instantaneous access to all 4+ M points in the data set during a numerical simulation.

- you state "…..Finally, the empirical data provides detailed location information about the vehicles inside of the study area. Given this, it is possible to build a model that includes spatial dependence inside the study area. Thus, the ODE model (4) which utilizes variables of time, flow and *volume* would become a PDE that utilizes variables of time, flow, *location and density*. This more sophisticated model is the Lighthill-Whitham-Richards (LWR) traffic model……." This looks more like a microscopic model that will no longer be applicable to flow simulations. Are you - in effect- saying that macroscopic models are only useful for free flowing traffic and not for peak-time traffic?

Both the ODE and LWR models are macroscopic flow models. Both models aggregate individual car observations into continuous field variables, e.g., the density of cars at a given time or location. Neither of these models is microscopic in nature: They do not simulate individual car behaviors and then aggregate them.

No, I am not saying that macroscopic models are only useful for free flowing traffic. The macroscopic models in my paper (and others' already published journal papers) utilize traffic data sets that range from relatively free flowing traffic to dense peak-time traffic. Compared to LWR, the ODE model is simpler in that it treats the entire 2100 feet long study area as being uniform in terms of density. I.e., volume(t) = 2100 * density(t). The macroscopic LWR PDE model is more sophisticated because it models the density within the study area. I.e., the solution of the PDE is a function of two variables: density(t, x) where x is in [0, 2100]. In this case volume(t) = $\int_0^{2100} density(t,x)\,dx$. (But, I don't have access to a PDE solver that can numerically simulate the LWR PDE and its specific boundary conditions.)

It may also be that 2100 feet is not enough road span to collect traffic data that will be fed into macroscopic flow models (I realize there may not be data from larger highway spans). Can you provide references that used macroscopic traffic flow models and this short of a road-span?

The data set utilized is appropriate for analyzing macroscopic traffic models. One example is below, which I did reference in my paper:

Fan, S., & Seibold, B. (2013). Data-Fitted First-Order Traffic Models and Their Second-Order Generalizations: Comparison by Trajectory and Sensor Data. *Transportation Research Record*, 2391(1), 32-43. https://doi.org/10.3141/2391-04

Fan & Seibold also utilized one of the NGSIM data sets. They utilized an NGSIM data set for a 1,640 feet (500 meter) long straight stretch of the I-80 freeway located in Emeryville, California in the San Francisco bay area: I-80 NGSIM data set fact sheet. Their models were all macroscopic traffic flow models.

> Which is probably why your points are clustered together and you could either recourse to your 'first path' to improve the model or simply collect data over larger spans of highway. Please discuss and include this under your 'first path' recommendation.

The data set's points are not "clustered" due to the **length** of the study area (highway). The points are distributed the way they are because of the time of day during which the observations were made [7:50:00, 8:35:00]. That's why there are virtually no points for densities that are less than 150 vehicles per kilometer.

The point being made is that if there were also observations from around 7:00 am, when there is very little traffic – very low density, then the entire set of points would span a wider range of traffic densities, velocities and flows ($v(t), q(t)$, and $k(t)$ respectively), which could potentially result in flux functions which may have more predictive accuracy. Collecting data over a longer span of highway will NOT make the data set's points less "clustered".

- The references must be in curved parantheses. Please change. To provide differentiation from equations, please insert the word 'equation' in front of all the numbers that describe equations.

Thank you for pointing this out. The changes have been made. All references to other scientific articles in the paper's text were converted from [ ]'s to ( )'s. All references in the paper's text to equations were converted to *equation* # style.

- You actually have the opportunity to build a database of 'synthetic' data points from these two models that incorporates the portions you worked with using actual data. You can upload that to kaggle, where it wil lbe useful for many researchers. There are some synthetic traffic flow datasets on Kaggle but they do not seem to incorporate actual data (at least from a part of the describing function). Obviously, you need not respond to this point in your response.

Thank you for the suggestion! I will look at Kaggle after my college applications are submitted.

Accepted.