General Comments

Development of a new open-source vehicle emissions model has merit. However, I am concerned about the rationale in using the average speed distribution from the State of Georgia in use in this study, also with the language suggesting that it is an improvement in accuracy. I am concerned there is a disconnect in the use of average speed distribution in MOVES and in the CARS model, and don't understand why an average speed distribution could be calculated from the link-level data from South Korea's GSI road shape files. I added substantial questions and comments regarding this issue in the specific comments below. (See Page 8. Line 271-281).

Because I felt this issue must be addressed before going forward with the paper, I stopped my review at this point. And only conducted a cursory review of the remaining aspects of the paper, including results and conclusion.

I have also added many specific comments to remove generalizations that may be not accurate, provide citations behind some of the statements, and to clarify the calculations which I encountered in the abstract, introduction and methods section.

I am willing to re-review the paper if my concerns can be addressed regarding the average speed distributions in the model.

Specific comments (*suggested text in italics***)**

p.1 Line 24 "it can optionally utilize road link-specific average speed distribution (ASD)"

- is it an average speed distribution of the road type, or the individual link? The wording is not clear.

Should it be referred to road-specific average speed distribution? Like on p.1 line 32?

p.2 Line 39-41 "It indicates that the CNG bus is better for the rural area while the diesel bus is better applicable for the urban area for a better ozone control strategy because the rural area is usually NOx limited for ozone formation and urban area is VOC limited region"

Is this backed up with air quality modeling, or assumed based on the reasons given here? In practice, couldn't it be much more complex? If it is not built on analysis, then I think the statement should be re-written as a potential ozone control strategy—which may need to be backed up with more analysis. E.g. what would be the impact on suburban areas of more NOx or VOC emissions?

p.2. line 47. The line about indoor vs. total air pollution makes it seem that ambient air pollution is a relatively minor contributor to public health. I would add more citations to clarify that ambient pollution impacts indoor air quality, or remove the indoor air quality references—as potentially misleading.

For example: Cohen et al. 2017 estimate 4.2 annual early deaths to ambient PM.

Cohen, A. J., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet,* 389 (10082), 1907-1918. DOI: https://doi.org/10.1016/S0140-6736(17)30505-6.

Burnett, et a. 2018 estimate the health burden is closer to 9 million deaths from ambient PM concentrations

Burnett, R., et al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proceedings of the National Academy of Sciences, 115 (38), 9592-9597. DOI: 10.1073/pnas.1803222115.

p. 2. Line 57-59. Is this statement backed up with a citation? If not, I would not say this could be an over-generalized statement is generally always accurate. For example, in areas with persistent cold pooling inversions—modeling the meteorology and chemistry may just or more critical then accurately modeling the emissions—correct?

p.2 60-61. Another over-generalized statement that deserves more context and a citation. For which pollutant? Do you mean for NOx? I don't think this is true for VOCs, and was a little surprised that it was mentioned as true for PM2.5 (because that is not the case with the US NEI for select urban counties).

p. 2 lines 67-69. Are these studies based on air quality modeling or observations? Either way, I would suggest these results are presented in the context of other studies that show that primary PM2.5 is a minor contributor to ambient PM2.5.

Nault et al. 2021 suggests that PM2.5 in Seoul (and other major urban areas across the world) have only a small contribution of PM2.5 from primary PM2.5. Most is secondary formed PM either as secondary organic aerosol or secondary inorganic PM2.5 (ammonium-sulfate or ammonium-nitrate).

> Nault, B. A., et al. (2021). Secondary organic aerosols from anthropogenic volatile organic compounds contribute substantially to air pollution mortality. Atmos. Chem. Phys., 21 (14), 11201-11224. DOI: 10.5194/acp-21-11201-2021.
>
> See also
>
> Jimenez, J. L., et al. (2009). Evolution of Organic Aerosols in the Atmosphere. Science, 326 (5959), 1525-1529. DOI: 10.1126/science.1180353.

p. 3. Line 73. "~~the high quality~~ *highly resolved* spatiotemporal automobile emissions"

- bottom-up emissions inventories by process can give high resolution, but are not necessarily higher quality than top-down methods.

Aren't the vehicle operation processes both physical and chemical? I would recommend stating that you can get more spatiotemporally resolved emissions inventories when pairing process-specific emissions models with resolved vehicle activity data.

p. 3. Line 97-98. They are developed differently to meet *specific user needs?* ~~their own needs~~ based on the types of traffic

activity and emission factors, emission calculation methodologies, and other optional/available.

- I hope the emissions models are not developed to meet the models own needs….but on the model users needs
- Should you mention that each model is developed with different levels of specificity, underlying data sets, and modeling assumptions?

p.3. lines 100-101. This statement is not clear, and not sure it is needed here.

P. 108-109. I disagree about the general statement on the lack of transparency for emission factors. Technical reports that document the emission factors and algorithms for estimating emissions are available here: https://www.epa.gov/moves/moves-onroad-technical-reports

Perhaps, it could be stated that it is high degree of specificity, make it difficult to update and apply to countries outside the USA. (although there are examples of the done in the literature).

Page 6. Line 180-181. What do you mean by traffic density? Are you referring to total VKT as the subtitle suggests? Isn't VKT a measure of traffic flow rather than density?

Page 6. Line 192-193. How is the VIN used in the calculation? Is that to calculate the vehicle age? Could you clarify?

p. 6 line 194-195. VKT with the manufactured year (VKTv,age) is calculated based on the cumulative mileage (Mf) ~~since~~ *between* the last inspection date (Df) and registration date (D0).

- Clarification on the data in the calculation. Is the registration date always at or near age zero? Does the registration data only capture vehicles when they change ownership? Or does that happen more regularly? I think it is important to clarify if the VKT calculated in equation 1, is reflective of most recent years of use, or an average VKT over the lifetime of the vehicle.

- Also, since equation 1 is applied to individual vehicles, shouldn't the subscript reflect individual vehicle in the equations, rather than just vehicle type, and age?

p. 6 lines 205-207. 'nonroad automobile' is potentially confusing. Vehicles in MOVES are classified as onroad or nonroad. When passenger vehicles are operating in driveways, parking lots, I would classify this as off-network automobile emissions.

Can you clarify what are the off-network automobile emissions that are missing? Are those starts? Evaporative? Idling emissions? Or are they spatially allocated to the roadways.

Line 225-227. Should you also mention that emissions are produced from incomplete combustion products that are not controlled from the emissions aftertreatment equipment, such as a three-way catalytic converter for gasoline vehicles?

Line 228- I would not say NOx is similarly produced as SOx, because the source of S is the fuel, not the atmosphere like nitrogen.

Page 7. Line 280-235. Does the age in the emission factor, reflect the impact of model year on the new vehicle emission rate, and age in the DF factor reflect aging effects from the new vehicle emission rate? If so, could that be clarified—otherwise, it is not clear what the purpose of DF is. However, if everything is calculated in terms of age—does that mean an emission rate for a 5 year-old vehicle is the same, for all calendar years? Or does model only work for one calendar year?

Page 8. 247-248. In this equation, is vehicle age used to reflect the model year or technology effect of a new vehicle, or a deterioration effect? If everything is in terms

Page 8. 255 "Figure 3a shows a significant decrease of NOx emissions while speed increases *between 0 and 70 kmh*.

Page 8. Line 260. I would remove word 'constant' speed. That implies that the vehicle emissions model can differentiate between emission factors between constant speeds, and transient sec/sec speeds, which I do not think is the case. My understanding is that you are differentiating between a single average speed, to an average speed distribution. However, the associated emission rates with the average speeds is not changing.

Page 8. Line 262. Remove 'incomplete ICE combustion' higher NOx emissions are not necessary directly related to an issue with incomplete combustion—NOx emissions can be lower when ICE conditions are fuel rich. I would just recommend that a single speed, may not represent the average emission rates, as an average speed distribution with time spend at multiple speeds.

Page 8. Line 271-281.

I believe there may be an issue with using ASD from MOVES for Korea in CARS.

1. Average speed distributions from MOVES and provided from the State of Georgia should be defined. MOVES average speed distributions are intended to calculate the distribution of average speeds across multiple links within the same road type. As well, as to capture the distribution in average speeds in links across time (such as days with traffic incidents or normal travel days). In practice, the data is aggregated from telematics data that calculate average speeds from varying resolutions (1 hz to every 180 seconds). As documented in the MOVES population and activity report.

   USEPA (2020). Population and Activity of Onroad Vehicles in MOVES3. EPA-420-R-20-023. Office of Transportation and Air Quality. US Environmental Protection Agency. Ann Arbor, MI. November 2020. https://www.epa.gov/moves/moves-technical-reports.

   I am not familiar with how the State of Georgia average speed distributions were calculated, but that should be explained if those are used in the study.

2. Why not develop average speed distributions from South Korea's GSI road shape files? The text seems to suggest that it is a problem that there is only one average speed associated with each road link. But this is not a problem. There should be many links of the same road type within a region. Using that data, you can calculate an average speed distribution for that road type in the region. Is it because you want to capture variation in link levels average speeds across different days or hours of the day? If so please clarify.
3. It is not mentioned if the average speed distributions vary by time of day—please clarify, they do in MOVES, which captures the effect of the diurnal traffic pattern on vehicle speeds.
4. The emission rates from the CARS model—are they intended to be associated with a cycle average speed that represents the average speed with a road link (which contains variation of speed within the link)? Or are the emission rates intended to be associated with sec/sec speed data? If the emission rates are intended to be cycle or link average emission rates, then it makes sense to use average speed distribution calculated from the average speeds from many links, rather than using an average speed distribution calculated from sec/sec data.
5. Is using the average speed distribution approach even needed here? If you have the average speed for each individual road link? Why not use that approach? Add an explanation.

6. It is not clear why using ASD from the state of Georgia are more realistic than the using the current data. Also, the explanation is not clear on the development of the inputs. E.g. where did the 2:1 ratio of bins 1 and 2 come from. And where did the additional 2%, 3%, 7%, and 15% come from. Was it added such that the average speeds on each roadway is the same before and after the fix?
7. To me, it seems like this should be better classified as a 'sensitivity' study on the potential impact of using average speed distributions. Since, the data for the average speed distributions used in the study, don't seem to be clearly better than what is there, and seem to be based on quite a few assumptions.

Technical corrections: (*suggested text in italics)*

p. 1. Line 18. "utilize ~~the~~ local vehicle activity data~~base~~"

p.1 Line 20 "to generate a temporally and spatially *resolved* ~~enhanced"~~

p. 2 line 28, "due to ~~its~~ *it having the* longest daily VKT *and relatively high NOx g/km emission rate."*

p. 4. Line 133 "changing specific variables that may be ~~hidden somewhere.~~ *embedded in the code?"*

p.5 line 151 "road link-level"

p.5 line 174. "South Korean traffic databases *from* ~~by~~"

line 224 "hot exhaust emission*s"*

p. 46 Appendix E and F. Speed units should be km/hour. Also, what is the range of the speed bins? That is not clear from the Tables.

p. Appendix F. Single *average* speed for each road type

The grammar should be improved. I recommend that it be reviewed by an English technical editor. I started marking grammatical corrections on the abstract but did not make comments on the rest of the document.