The authors have tested a heavy-duty vehicle in Japan on a dynamometer and using a PEMS to evaluate how driving force and season influences emissions of CO2 and NOx. The literature review is not sufficient to show why this work is novel, especially as the authors measure one vehicle only. The methodology is not transparent sufficiently to allow key outputs to be replicated, namely the transient emissions maps. There are issues with the figures in the SI which need to be addressed, including the transient emissions maps and the correlation analysis.

The authors are grateful to the reviewer, Dr. Bishop, to taking a time and giving us the important insights, suggestions and clarifications. We modified the manuscript carefully based on the opinions. The added and modified sentences based on reviewer 1's opinion were highlighted in yellow. To address the novelty and validity of the experiments and analysis (using only one heavy-duty vehicle), following sentences were added in the article.

Lines 80-81: "To the best of our knowledge, this is the first study in which PEMS measurement results are applied in the development of an estimation model for vehicular exhaust."

Lines 327-331: "The experiment as well as the detailed analysis were conducted only for one heavy-duty vehicle in the Japanese market. In future, it is expected that further studies would be conducted to obtain the variability of real-world vehicular exhaust emissions; after the measurements results for the consistent number of vehicles have been obtained, the emissions inventory based on real-world measurements should be evaluated for use in policy making regarding air quality treatments."

Line 20: Define long-term and short-term

According to the suggestion, definitions and examples were added as the following sentences.

Line 20-23: "The air pollution caused by long-term air pollutants, which are chemically stable components such as CO2, and short-term air pollutants, which are reactive chemicals such as NOx, volatile organic compounds (VOCs), and photochemical oxidants, is a cause of significant concern in many countries, including the United States, the European Union, China, India, and Japan (Akimoto et al., 2015; Costa et al., 2014; Ravindra et al., 2016; Sullivan et al., 2018; Yang and Wang, 2017)."

Line 23: What year is the NASA reference?

The webpage of NASA is updating continuously, so the concrete year was not defined in the citation. Meanwhile, we added the corresponding year of 1 degC increase from average temperature between 1951 and 1980 as "… in 2020" in Line 25.

Line 23: Better to use 'climate change' instead of 'global warming'

Thank you for the suggestion. The terms were substituted by 'climate change'.

Line 25: This reference is 10 years old – can you find a newer source to support your point?

According to the suggestion, we added IPCC report, "Summary for Policymakers", published online in 2018 as the citation which describes the potential increase of global temperature in the near future.

Line 25: Is 'photochemical oxidant' a single species? It seems this would be a group of chemicals

Thank you for pointing out the lack of information. Photochemical oxidants are composed of several components including ozone, PAN etc. and approximately 90% of oxidants are ozone. The sentence was modified as following.

Lines 29-32 "In addition, photochemical oxidants, mainly composed of ozone, are well-known short-term air pollutants, generated by the photochemical reaction of NOx and volatile organic compounds (VOCs) (Sillman, 1999). The concentration of photochemical oxidants in the atmosphere is a significant concern for humans, animals, and crops in many countries (Chappelka and Samuelson, 1998; O'Neill et al., 2004; Wang et al., 2017)."

Line 41: I don't believe this to be the case – lab tests are set to standard conditions to allow repeatability over all tests. The narrow test conditions means the results may not align with what we see in more varied real-world conditions

The authors understood what the reviewer wanted to mention. There were several studies focusing on extreme temperature conditions such as exhaust measurements on the very low temperature conditions, but the laboratories which can conduct the kind of experiments were very limited mainly because of the construction cost of measurement setup. Meanwhile, we added the following sentences to further explain the pros aspect of PEMS experiment.

Lines 44-45: "(recently however, temperature variable chassis dynamometers were adopted only in a limited number of laboratories (Clairotte et al., 2013; Ko et al., 2017))".

Lines 52-54: "Road temperature and gradient might vary from one season and location to another; therefore, real driving emission measurements are more suitable for a better understanding of the real-world driving emission (RDE) verification."

Line 44: What is the EPA reference year? Catalytic converters operate based on stoichiometric combustion in a spark ignition engine, and are (I believe) relatively independent of the ambient Temperature

Thank you for pointing out the lack of information, the reference year 2010 was added in the citation. Also thank you for pointing out the mistake of the sentence. Ambient temperature affects to the amount of cold-start emission especially because of the temperature dependence of engine coolant temperature and the urea injection time. These discussions are mentioned in section 3.1 in the main article. We corrected the sentence as following.

Lines 47-50: “It has been noted that environmental temperature considerably influences the amount of exhaust emissions (detailed explanations regarding this observation are provided in section 3.1), leading to the release of a large amount of pollutants (including NOx) into the atmosphere in low ambient temperature conditions (the MOVES2010 Report by the U.S. Environmental Protection Agency, 2010).”

Line 45: Might be better to say proportional to fuel burn, since exhaust treatment technology mitigates the effect of driving conditions on tailpipe emissions

Thank you for the suggestion. Based on the suggestion, the sentence was modified as following.

Lines 51-52: “…, which is presumed to have a negative effect on fuel consumption (or CO2 emission) and other exhaust emissions.”

Line 70: What are the 2016 Japanese regulations?

The detail of the regulation was added in Table S1 of supplementary information.

Line 81: How were the lab test conditions modified to reflect different seasons?

We did not reflect the seasonal dependencies by laboratory tests. The purpose of chassis dynamometer measurement includes two folds: First, understanding the differences of velocity and acceleration profiles between lab test and RDE, and 2nd, to verify as the same condition of the tested vehicle in each tested dates because the experiments were conducted in four different dates and we needed to assure the vehicle performance to be the same condition to compare the RDE results in the different tested dates. The results are mentioned in section 3.3 of main article. To mention those reasons, following sentence was added in the main article.

Lines 75-79: “The purpose of this study is two-fold. First, chassis-dynamometer-based and RDE measurements using PEMS were conducted on a heavy-duty Japanese vehicle to determine the importance of RDE specific factors, including the ambient temperature and road gradient, among others. Second, the obtained experimental results were analysed based on two parameters, i.e., the driving force and vehicle speed, to develop an analytical method to evaluate the amount of CO2 and NOx emissions from the vehicle in an arbitrary driving condition."

Line 90: Why was EGR measured only in spring and summer?

In the first two experiments, autumn (November 2018) and winter (January 2019), EGR ratio was not taken in account to be measured. After we analyzed the experimental results of those two seasons, we realized that it was important to measure EGR ratio and the measurement of EGR ration from OBD was conducted for remained two seasons.

Line 91: I assume the route was the same across all days and seasons?

Thank you for clarifying lack of information. Exactly the route was the same in all the experiments. This was added as following.

Line 119: "The driving route was the same across all days and seasons."

Line 96: What is the justification for a 5 second smooth?

We added the 5 points smoothing flow as the following sentence.

Lines 111-115: "In equation (1), vehicle speed was smoothed using the 5-point average of the speeds at neighbouring time steps. We determined the averaging number of vehicle speed as 5 points based on two concepts. First, the averaging number of vehicle speed should be minimized as much as possible to maintain the high resolution time step. Second, the dispersion should be sufficiently lower than the measured vehicle speed. It is also worth noting that the averaging number of vehicle speed may depend on the measurement tools, and the 5-point value was suitable in this study."

Line 99: Worth explaining the central difference method and justifying its use here

Central difference method is a mathematical discretization formula of the differential form which is described in equation (2) of main article. The vehicle speed measured by the experiments was not the continuous value, so the discretization was needed to obtain acceleration. We think that central difference method is just a mathematical form of discretization method and no justification can be added. Please reconsider our insights.

Line 120: Rolling resistance and aerodynamic drag should be derived from coast down tests – was this done in your 2012 work referenced?

First, the authors are feeling sorry about the incorrect reference. The resistance parameters were not cited from our work in 2012. The parameters were set by using the same value of heavy-duty vehicle measured in our previous study (but there is no official citation), because the parameters were not derived from coast down tests in this study. The sentence was modified as following.

Lines 152-154 "While the coast down test to determine μr and μa was not conducted in this study, based on our previous study, μr and μa were set at 0.0089 and 0.0027, respectively, given that we used the same type of heavy-duty vehicle that was tested in our previous study."

Line 121: You have switched from km/s2 to m/s2 units for acceleration. How is the 0.139 m/s2 threshold determined? Line 122: How was the test mass determined?

The unit of vehicle acceleration defined in section 2.3.1 is described as km/s2 because the monitored vehicle speed was in the unit of km/s, on the purpose of unifying the unit. On the other hand, m/s2 is used in section 2.3.3 because acceleration parameter is used to calculate driving force, the unit is defined in N(=kg m/s2).

0.139 m/s2 corresponds to 0.5 km/(h s). The unit km/(h s) can be considered as the differential of vehicle speed (km/h) by time (s), dv/dt. In this study, we defined 0.5 km/(h s) or less than this value as almost no acceleration because this value is considered to be low enough to be considered as zero. The boundary threshold, 0.5 km/(h s) was used in the previous study (Yoshizumi et al., 1980), and the reference was also added in our manuscript.

Yoshizumi, K et al. Automotive Exhaust Emission in an Urban Area. SEA. Tech. Pap. Ser. 1980, 800326, p17.

Line 132: What is the justification for smoothing the altitudes? Altitudes are already smoothed to be a constant value within each mesh.

We added the following explanation about how the smoothing was conducted.

Lines 168-170: "…This smoothed value in the vicinity of ~5-m meshes was determined by varying the averaging mesh number (e.g., from 1, 2 …), while carefully checking to ensure that the noise remained negligible relative to the height change."

Line 133: Similarly, what is the justification for the 7m smoothing to determine road slope?

We added the following sentence as justification of smoothing value.

Lines 172-173: "This smoothed value, in the vicinity of ~7-m meshes, was determined using the same method that was applied to altitude data, as described above."

Line 150: I disagree - exhaust temperature varies more than coolant temperature

Thank you for pointing out the irrelevant description. The sentence was modified as following.

Lines 189-191: "Figure S1 suggests that in the cold-start situation, the engine coolant temperature was proportional to the exhaust gas temperature. Further, in general, engine coolant temperature is usually used to control the EGR system."

Lines 206-207: Should define the torque and speed ranges

According to the suggestion, the sentence was modified.

Lines 224-225: "…divided into six torque ranges (0–100, 100–200, 200–300, 300–400, 400–500, and 500–600 N m) and three engine rotation ranges (500–1000, 1000–1500, 1500–2000 rpm)."

Line 204: There is no method to recreate the transient emissions table

The authors were convinced about the opinion above. All the raw data of temperatures, NO_x and CO_2 emissions, road gradients etc. have been sorted to be available to the readers. The data is vested in the *Tokyo Metropolitan Government* and we do not have the right to open the data to the public by online repository (after the discussion with the government). Therefore, the data could be available by the request to the corresponding author. The statement of data availability was added as the following sentence.

Line 333: "Data availability. All the RDE data measured in this study were available by the request for the corresponding author."

Line 206: Engine out emissions are related to driving force (and fuel used), but tailpipe NOx is decoupled from engine out emissions due to active management by the SCR

Thank you for the clarification. To explain the precise meaning and background of transient emission table, the related sentences were improved as following.

Lines 249-253: "The formulation method was based on the assumption that the amount of emissions such as those of NOx and CO2 from engine exhaust depend on the driving force and vehicle speed at any given moment, consequently resulting in the transient NOx and CO2 emissions from the tailpipe. Because this dependence cannot be formulated mathematically owing to the non-linear relationship between the detoxification tools of EGR and urea-SCR, an emission table containing the parameters of the driving force, vehicle speed, and amount of emissions was employed."

Line 213: Earlier, you said ambient temperature had an important role to play in emissions

At this sentence, the authors intended to mention that temperature is not critical to "CO2" emission, and the term "CO2" was added in the revised manuscript.

Line 218: Some evidence is needed for this. The air conditioner will manifest as some additional load, *in the same way as a heater during the winter.*

Unfortunately, the behavior of air conditioner was not monitored in this study. For this reason, we added additional information as below. Thank you for your reconsideration.

Line 262: "…while it was not used in the winter season, …"

Lines 265-266: "Despite this, the trend of the air conditioner was not monitored in this study, and such discussions are one of the possibilities."

Fig 1: There are two trips per day in each of the seasons – does this graph show the average of those two trips? Were all of these cold starts, with engine coolant temperature from the same starting point

Thank you for pointing out the lack of information. The time trends shown in Fig.1 are AM test results: this information was added in the caption of Fig.1. The start point of coolant temperatures were different in each season, this is also one of the reasons why the seasonal dependency happened. We added this discussion as following.

Lines 194-195: "…because the initial coolant temperature is higher in the hotter seasons."

Fig 2: How is this graph determined? We don't know what the ambient temperatures were in each of the eight tests (two tests per day, four seasons)

Thank you for pointing out the lack of information. The detail of temperatures and plots were added in the caption of Fig.2 as following.

"Each plot was obtained from AM and PM tests conducted in four seasons (the experiments were conducted for a period of 5 days in each season). Average ambient temperatures were determined using the average value of the temperatures recorded during each driving test."

Fig 3: How many dyno tests were done? Were the ambient conditions of the PEMS test replicated here?

One chassis dynamometer test was conducted before the PEMS experiment in four seasons, so totally four dyno tests were conducted. The ambient conditions were almost the same between all the tested environments. We added this information in Lines 109-110.

Lines 109-110: "n all the laboratory measurements, room temperature was set to be approximately 25 °C."

Fig 4: Why is the area of the EGR + SCR plots (third column) larger than the No EGR + SCR and EGR only plots? They should all occupy the same area because vehicle speed and driving force doesn't change across the three columns

Driving force and vehicle speed are different between three phases. No EGR+SCR results were obtained from first 10 min after the driving started. On the other hand, EGR+SCR results were obtained from 30 min after the driving started. The driving root after 30 min included high speed and road gradient environments, and therefore, EGR+SCR results holds wide range of driving and speed compared with other phases.

Fig 5: As Fig 4 Fig 6: The R2 value for these graphs might be high, but there is large variation about the 1:1 line

The authors added the following sentence to refer the dispersion of predicted amount of exhaust emissions as for the reviewer's opinion.

Lines 277-284: "The dispersion of the plots might indicate the limitation associated with the application of the transient emission table in the estimation of NOx and CO2, as well as other presumable vehicular exhaust emissions. The dispersion resulted from the fact that the transient emission table modelled in this study simulated real-time exhaust emissions based on driving force and vehicle speed. Real-time prediction is associated with several uncertainties that cannot be taken into account by the two parameters, driving force and vehicle speed, such as transient high acceleration, emission control systems settings by the manufacturer, etc. Nevertheless, the predicted results shown in Figure 6(a) and (b) include high linearity even for the real-time measurement results. The results also highlight the possibility of being applied in emission inventory evaluation in future."