Answer to referee 1:

General comment:

This paper presents a detailed analysis of the impact of introducing electric vehicles on UHI and air pollution in the GTA. The topic is interesting and the paper is well-written. However, before considering publication, it is crucial to clarify the methods used and confirm the reliability of the results, as detailed below:

> The authors are grateful to reviewer 1 for providing significant suggestions and clarifications with sophisticated insights. The manuscript was modified based on the comments. The words and sentences with changes are highlighted in yellow.
>
> Aside from the provided opinions, to deepen the discussion, the authors added new insights of the contribution of NO-titration effect caused by lowered PBL height caused by UHI effect and lowered kinetics of photochemical reactions affected depression of O_3 which are discussed in the sections 4.1.2 and 4.1.3 of the revised manuscript. The results suggested that enhanced NO-titration contributed maximally ~1ppb decrease of O_3 while lowered kinetics of O_3 formation contributed maximally ~0.4ppb decrease of O_3 in the GTA.
>
> Further, total amount of primary emissions of air pollutants in whole analyzed regions (GTA) for the six scenarios of this study were added in Appendix B, which will enhance better understanding of our work to the readers.

Specific comments

The vehicle AH calculated in Sect A.4 appears to be a regional total amount rather than a gridded distribution based on the current description; therefore, how is the vehicle AH spatially distributed to match the innermost WRF domain? Additionally, it would be beneficial to briefly describe in the methods section (e.g., 2.2.1) what parameters/modules are modified in WRF to account for changes in vehicle AH.

> Thank you for the comments. Actually, AH was set to be distributed averagely in the calculated domain, and not matched with the spatial distribution of vehicles. This is because the urban canopy model incorporated to WRF has not implemented the detail category of AH sources. We assumed that the distribution of vehicles is equal in the analyzed domain D3. This might be almost reasonable in the calculated domain because

more than four third of the Greater Tokyo Areas are urbanized while in the non-urbanized regions, there might be overestimation of the reduction of AH. We added this fact to the main article as follows:

Lines 130-134: "Note that the UCM coupled in WRF provides only the average distribution of AH and other urban canopy parameters. In an urbanized region such as GTA, the distribution of on-road vehicles is expected to be distributed equally. Nevertheless, there is also countryside in domain D3 shown in Fig. 1 at the western, eastern, northern, and southern borders of those areas, and the applied reduction ratio might therefore be overestimated. Such a limitation is discussed later in the section."

The reason why the AH from powerplants was not considered in this study also came from the same reason since the powerplants are distributed in the specific area of D3 of which the details are responded in the following answer.

We added new figure (Fig.2) to show the diurnal profile of AH before and after the introduction of BEVs as well as showing the detailed UCM parameters related to AH as follows:

Lines 106-109: "Highlighting the UCM parameters related to AH, Hara et al. provided maximum AH as 29.0 W m^{-2} at 18:00 and 19:00 for three urban area category types (low-density residential, high-density residential, and commercial areas), and this is reflected in the anthropogenic heating diurnal profile of the GTA shown by the black square dots in Fig. 2."

The increase in AH from power plants in the GTA under the BEV scenarios should also be considered.

Thank you for pointing out the lack of information regarding the AH set in this study. As mentioned above, current UCM code implemented in WRF enables us to set only the averaged AH from all sectors in the calculating regions, and there are no options to set UCM parameters for each anthropogenic source (e.g. vehicle, residential, industry, etc.). The distribution of vehicles is expected to be nearly averaged over whole calculated domain of GTA (D3), so the analysis of AH could be done. On the other hand, power plants are condensed in the marine sides of D3 (please see Fig. S3 of the supporting information), and the averaging AH for the power plants could not be evaluated. We added this fact in the revised manuscript as follows:

Lines 134-138: “In addition, the introduction of BEVs is expected to increase AH from power plants due to the increased demand for electricity used in battery charging; however, the increased AH from power plants was not considered in this study. Power plants in GTA are located in specific areas near the bay, and are not distributed throughout the entire area; therefore, it was difficult to consider the effect of AH from power plants and its subsequent UHI effects in the GTA.”

We added this issue in the new section “4.6.1 Limitations and future perspectives” as follows:

Lines 519-522: “Second, also because of the limitations of the parameterization of UCM, the increase in AH from power plants due to the increasing demand for battery charging was not considered. According to Section A.1 of the Appendix A, there would be a 34% increased energy demand for battery charging and the associated AH increase may not therefore be negligible.”

Line 130: "The emission inventories were provided by Chatani et al." This description is not clear. It should clarify which anthropogenic sources are included, and whether natural sources and emissions from countries outside Japan in domain D1 are considered.

We are sorry for the lack of information about the emission inventories. The calculation of chemical transport modeling was conducted for D1, D2, and D3 (D1 is the parent domain of D2, and D2 is the parent domain of D3 by the regional nesting), so the emissions of outside of Japan were also considered. The details of calculated domains and emission inventories were added as follows:

Lines 142-143: “The calculated domains were the same as those used in the WRF model shown in Fig. 1, and regional nesting method was applied for domain D2 (parent domain: D1) and for domain D3 (parent domain: D2). Domain D3 was the analyzed region.”

Lines 148-156: “The emission inventories were provided by Chatani et al. (Chatani et al., 2018) and are briefly summarized as follows. Hemispheric Transport of Air Pollution (HTAP) v2.2 (anthropogenic emissions: Janssens-Maenhout et al., 2015), Global Fire Emissions Database (GFED) v4.1 (biomass burning: van der Werf et al., 2017), AeroCom (volcanos: Diehl et al., 2012), and Model of Emission of Gases and Aerosols from Nature (MEGAN) v2.1 for (BVOC emissions: Guenther et al., 2012) were applied for emissions outside of Japan (D1). The Japan Clean Air Program (JCAP)/Japan AuTo Oil Program (JATOP) Emission Inventory for Vehicle Emission Model (JEI-VEM) (vehicular emissions: Chatani et al., 2011; Shibata and Morikawa, 2021), the Sasakawa Peace Foundation (ship emissions: Sasakawa Peace Foundation, 2022), Japan Meteorological Agency (volcanos: Japan Meteorological Agency, 2022), and MEGANv2.1 (BVOC emissions: Guenther et al., 2012) were applied for emissions inside Japan (D2 and D3)."

Line 133: I am curious whether using only two observation sites for validation is sufficient. For example, the study by Hata et al. (2020), mentioned in Line 167, which was conducted in the same area as this paper, appears to have used observation data from seven sites for validation. Moreover, the model performance from these two sites appears not satisfied. Is it sufficiently credible to discuss the temperature and concentration changes introduced by BEVs under such simulation errors, especially when these changes might be smaller than the simulation errors.

Thank you for the recommendations about the replicability of simulation and observations. The authors thought that it was better to show the minimum results of the modeling performance for readability although we conducted the evaluation of modeling performance for other cities like Hata et al. (2020) in the past. In revision round, based on the suggestion, the comparisons for 7 cities were incorporated to the manuscript. Meanwhile, reviewer 2 suggested that it is better to show the seasonal features of the modeling performance separately, so we amended this issue. Reviewer 2 also suggested to add other meteorological factors, so we added the validation of wind speed, wind direction, and solar radiation, in the revised manuscript.

The discussions of the modeling performance have been getting long, so we decided to move the detail of the comparison between observed and simulated results to Appendix C, the new section implemented in the revised manuscript, with Fig. C1 (ground temperature), Fig. C2. (wind speed), Fig. C3 (wind direction), Fig. C4 (shortwave radiation), Fig. C5 (O_3), and Fig. C6 ($PM_{2.5}$). In the first draft, the results of the modeling performance were shown by scatter plot, but the plots were complicated to see seasonal effect. So, in the revised manuscript, the yearly trend was chosen to clarify the modeling performance. The discussions of the modeling performance are described as follows.

Lines 636-647: "Figures C1–C4 show the correlation between observed and calculated results for the daily ground temperature, wind speed at 2-m height, wind direction, and total solar radiation for winter, spring, summer, and autumn in the seven analyzed sites in 2017. Note that the observed data of daily total solar radiation in Kanagawa, Chiba, Saitama, and Ibaraki were not available, so the remained three sites were compared. Furthermore, in terms of the wind direction shown in Fig. C3, the value was defined by the 16 directions: the value of 0 corresponds to 0° (north), 8 corresponds to 180° (south), and 15 corresponds to 337.5° (north-northwest). The correlations are exhibited using four statistical factors: correlation factor (R), root mean square error (RMSE), normalized mean bias (NMB), and normalized mean error (NME). Overall, the model was found to replicate the observed results well in all seasons. The results for the wind direction for Saitama and Tochigi show low R values of 0.19 and 0.29, respectively, but the time trend and the values of RMSE, NMB, and NME are similar to those of other sites. Unlike O_3 and $PM_{2.5}$, there were no proposed indicators for the statistical values described in later sentences, but the time trend of simulated results shown in Figs. C1–C4 replicated the observed results, and we concluded that the simulated meteorology could be applied in the evaluation of this study."

Lines 661-677: "Figure C5 shows the correlation between the observed and calculated (BASE) results and 8-h daily maximum average (MDA8) O_3 concentrations. As seen in Fig. C5, overall, the modeled O_3 replicated the observed results well, and the R value was more than 0.7, except for that of Gunma. Figure C5(f) suggests a lower correlation for the modeled O_3 with the observed results for Gunma. Emery et al. proposed indicators that can be used to validate O_3 and $PM_{2.5}$ for chemical transport modeling (Emery et al., 2017). According to these indicators, the ideal values of the modeling performance for MDA8-O_3 should be $R > 0.75$, NMB < ±0.05, and NME < 0.15 while the criteria value should be $R > 0.50$, NMB < ±0.15, and NME < 0.25. All the sites except for Gunma fully met these criteria and were close to the described goals. However, Gunma did not meet the criteria for R and NME. This may be because it is located in the countryside, where less primary air pollutant emissions are generated than in highly polluted areas, and the transportation from other regions renders it difficult to predict O_3 by CTM. Nevertheless, most of calculations for the analyzed sites showed good agreements with the observed results, and are thus considered to be acceptable for analyzing this study. Fig. C6 shows the correlation between the observed and calculated (BASE) results of the 24-h daily average (DA24) $PM_{2.5}$ concentrations. Emery et al. proposed that ideal values of indicators used to the modeling performance of DA24 $PM_{2.5}$ should be $R > 0.70$, NMB < ±0.10, and NME < 0.35, while criteria should be $R >$ 0.40, NMB < ±0.30, and NME < 0.50 (Emery et al., 2017). According to Fig. C6, all the modeled results except for those of Kanagawa met the criteria, and some R and NME values were close to the goal. The result of Kanagawa shown in Fig. C6(b) replicated the daily trend of the observed results. This analysis therefore shows that although relatively less accuracy was obtained for simulated $PM_{2.5}$ in Kanagawa, the modeled conditions could be applied in the analysis of this study."

Further, please note that the modeling performance in Gunma was not as good as other cities, so in the revision round, we chose Tochigi, which is also suburban prefecture added from revised manuscript, for the detailed analysis in later section.

Technical corrections:

Line 475: Table A3 should be A1.

Thank you for the clarification. It was improved.

Sect A.1 and A.4 should be combined or linked for easier reading.

Thank you for the suggestion. Sects. A.1 to A.3 were merged for readability.

Line 148: Specify which air pollutants are being referred to.

The primary air pollutants reduced from vehicular exhausts, evaporation, and power plants were added inside the parenthesis of the sentences as follows:

Lines 174-177: "The ALL scenario considers the effects of introducing BEVs and includes emission reductions from engine exhausts (anthropogenic VOCs (AVOCs), particulate matter (PM), ammonia (NH_3), sulfur dioxide (SO_2), and carbon monoxide (CO)), evaporative emissions (AVOCs), emission increases from power plants (NO_X, PM, SO_2, and CO) as a result of battery charging, and UHI changes."

Further, please note that we added the figures of the total emissions of air pollutants in the analyzed domain D3 for each scenario to Appendix B, which will enhance the readers to understand the quantitative amount of the emissions. Following sentence was added to the revised manuscript to connect main part of the article and Appendix B.

Lines 183-185: "Changes in the annual mean primary NO_X, AVOCs, PM, NH_3, SO_2, and BVOC emissions for the scenarios listed in Table 1 are also shown in Figs. B1 to B4 of Appendix B as well as the distribution of changes in emissions described in Figs. S1 to S4 of the SI."

Appendix B demonstrates the summary of the amount of emissions in D3 for each emission scenario with Figs. B1-B4 as follows.

Lines 605-608: "Fig. B1 shows the annual total emissions of NO_X, CO, SO_2, and NH_3 in the four scenarios (BASE, ALL, SEV, and SPP defined in Table 2) in 2017 within the GTA (kt y^{-1}). While all the emissions would decrease due to the introduction of BEV (SEV), the increased emissions from power plants would partly offset or increase total NO_X, CO, SO_2 (SPP) emissions. The relevant decreases in NO_X (~20 %) and CO (~60 %) emissions are shown in ALL and S_{EV} scenarios."

Lines 612-616: "Fig. B2 shows the total annual AVOC emissions for the four scenarios (BASE, ALL, SEV, and SPP, defined in Table 2) in 2017 within the GTA (kt y^{-1}). The species of "Others" in Fig. A2 include alcohol and acetylene. The decrease in AVOC emissions from BASE to ALL scenarios mainly relates to the decrease in exhaust and evaporative emissions from service stations; however, the increase in emissions from power plants is almost negligible. A high decrease in alkane (~23 %) is expected in relation to the introduction of BEVs."

Lines 621-624: "Fig. B3 shows the total annual $PM_{2.5}$ emissions for the four scenarios (BASE, ALL, SEV, and SPP, defined in Table 2) in 2017 within the GTA (kt y^{-1}). Unlike NO_X, CO, and AVOCs, none of the $PM_{2.5}$ components show a relevant decrease following the introduction of BEVs. This means that the $PM_{2.5}$ emitted from vehicle exhaust emissions has almost no effect on total $PM_{2.5}$ emissions."

Lines 628-631: "Fig. B4 shows the total annual BVOC emissions (isoprene and monoterpene) for the two scenarios (BASE and SBVOC, defined in Table 2) in 2017 within the GTA (kt y^{-1}). The emissions in the two scenarios are almost equal, and only ~0.5 % of BVOC emissions decrease from the BASE to SBVOC scenarios. This decrease was related to mitigation of the UHI effect through the introduction of BEVs."

The structure of the paper needs reorganization. The current Discussion content seems more appropriate in the Results section, while the Discussion should provide deeper implications of the results, such as specific emission control policy recommendations. Additionally, discussions on *uncertainties and limitations should be included.*

Thank you for the comment. The results of $PM_{2.5}$ composition, which were previously placed in Discussion content (Sect. 4.2), have been move to Results content (Sect 3.2.3) in the revised manuscript. Instead, we increased the discussions including policy making and limitations of this study as follows.

Lines 496-512: "4.5 Implications for policymaking

The introduction of BEVs as passenger vehicles to the market has been accelerated worldwide. In 2023, China ranked top in the BEV introduction rate (including plug-in hybrid vehicles (PHVs)) at 38%, followed by the European Union (21%), Israel (19%), and New Zealand (14%) (Global EV Data Explorer, 2024). However, in Japan, the BEV introduction rate was only 3.6% in 2023, although the national government has announced the complete substitution of new ICVs to BEVs in the market by 2030's. BEVs are expected to be dominant in Japan and worldwide in the future. According to the results of this study, the introduction of BEVs to the GTA is estimated to be "totally" effective in mitigating O_3 and $PM_{2.5}$ pollution and related premature deaths. Despite these results, ground-level O_3 and $PM_{2.5}$ are expected to increase in some areas depending on the seasons and atmospheric conditions. This means that changes in O_3 and $PM_{2.5}$ should be carefully monitored at a local city level, not only with respect to the introduction of BEVs but also for all emission sources associated with emission reduction strategies. Positive effects are predicted to occur through the reduction in AH followed by the mitigation of UHI, which is a main focus of this study, and the total number of premature deaths caused by O_3 and $PM_{2.5}$ would be reduced. In addition, the decrease in AH and mitigation of the UHI would have a direct impact on reducing health issues, such as heatstroke. The number of deaths caused by heatstroke in the GTA was estimated at 128 in 2017 (Ministry of Health, Labour and Welfare of Japan, 2024); this value is compensated by the decrease in the number of premature deaths (252) attributed to secondary air pollution (175 by O_3 reduction and 77 by $PM_{2.5}$ reduction). Therefore, it is expected that introducing BEVs could mitigate the health impact in the GTA, and may possibly be as effective in other megacities worldwide."

Lines 513-524: "4.6 Limitations and future perspectives

We found out the relationship between ground-level O_3 and $PM_{2.5}$ and mitigation of the UHI effect attributed to the introduction of BEVs in the GTA. Although our study provides new insights, it is of note that two assumptions were made in the setup of numerical simulations. First, the change in AH through introducing BEVs was assumed to be averagely distributed in the calculated region due to limitations in the parameterization of the UCM incorporated in the WRF. This assumption might have caused overestimations of the UHI effects in the countryside; however, as more than half of the area of the analyzed domain (D3) is composed of urban to suburban areas, the effect of overestimation is expected to be limited. Second, also because of the limitations of the parameterization of UCM, the increase in AH from power plants due to the increasing demand for battery charging was not considered. According to Section A.1 of the Appendix A, there would be a 34% increased energy demand for battery charging and the associated AH increase may not therefore be negligible. To account for these issues, it will be necessary to update the program for the UCM model incorporated in the WRF, and collaboration between developers of the WRF and scientists in the field of atmospheric science is recommended in future work."

Additionally, we added the discussions related to the atmospheric chemistry of O_3 formation to the new sections 4.1.2 and 4.1.3. According to the results of sections 4.1.2 and 4.1.3, NO-titration effect caused by lowered PBL height due to mitigation of UHI effect caused maximally ~1.0 ppb decrease of O_3 while the lowered temperature contributed maximally ~0.4 ppb decrease of O_3 via degradation of atmospheric chemical reactions. To the best of our knowledges, there were no previous work which separate the contribution of the effects of NO-titration and photochemical kinetics. In the analysis of section 4.1.3, the calculation of the box model of SAPRC-07 was conducted which was not done in the first draft. The details are as follows.

Lines 316-332: "Compared with the spring and summer, a wider distribution of high ΔTO was observed for winter and autumn, and more relevant NO-titration effects were estimated in these seasons; these were presumed to be caused by the intense decrease in temperature caused by mitigation of the UHI effect. Haman et al. conducted observational and CTM calculations based on Houston (U.S.) from 2008 to 2010 to clarify the relationship of diurnal temperature, PBL, and ground-level O_3 (Haman et al., 2014). The results suggested a positive relationship between ground-level O_3 and the PBL height; this was due to the enhancement of NO-titration caused by relatively weak wind speeds during the lower PBL height which caused NO_X to remain on the ground's surface. Figure 9(e)–(f) shows the seasonal changes in ground-level O_3 between the BASE and SUHI scenarios (ΔO_3). A higher decrease in O_3 is estimated in the central area of GTA, and winter and autumn show a high O_3 decrease; these results are confirmed in Fig. 5. The seasonal variations in the positive increase in O_3 shown in Fig. 9(e)–(f) correspond to ΔTO shown in Fig. 9(a)–(d). However, the distributions of ΔO_3 shown in Fig. 9(e)-(f) do not correspond to the distributions of the H_2O_2/HNO_3 ratio. The results indicate that O_3 reductions caused by the mitigation of UHI effects are attributed by the enhanced NO-titration effect rather than the photochemistry involved in O_3 formation. Enhancement of the NO-titration effect contributes to a maximum decrease in O_3 of ~1.0 ppb. Despite these facts, Haman et al. suggested that the relationship between PBL and the NO-titration effect strongly depends on factors such as the region analyzed and meteorology (Haman et al., 2014). Future work should determine the relationship between UHI effects and O_3 in other regions. The impact of the change in temperature on photochemical reactions is discussed in the next section."

Lines 337-362: "4.1.3 Box model simulation of the temperature dependence of ozone formation chemistry

Section 4.1.2 discusses the decrease in ground-level O_3 caused by the enhancement of NO-titration effect due to the lowered PBL height from mitigation of the UHI effect. Mitigation of the UHI effect causes a local-temperature decrease and is expected to weaken the rate of photochemistry involved in O_3 formation. For example, Coates et al. suggested that O_3 formation is enhanced through the increase in temperature due to the increased reaction rate of VOC oxidation and peroxy nitrate decomposition (Coates et al., 2016). Meng et al. suggested that high temperatures lead to a high HO_2 + NO reaction rate, which increases NO_2 and contributes to high O_3 episodes (Meng et al., 2023). To clarify the effect of temperature changes on the photochemistry of O_3 formation, a box model calculation using SAPRC-07 was conducted in this study. The CO emissions were the highest in the D4 region (as shown in Fig. B1 of the Appendix B) and the CO concentration was set to 1 ppm. The concentrations of the remaining pollutants, including NO_X, SO_2, AVOCs, and BVOCs, were set using the ratio between the annual amount of each emission in the BASE scenario shown in Fig. B1 to Fig. B4 and that of CO. The detail settings of the concentrations of the species for box model simulation are listed in Table S4 of the SI. The concentration of H_2O was fixed to 1.56×10^4 ppm; this was estimated from the value of partial vapor pressure of H_2O at 25 °C with 50 % relative humidity. Solar intensity was defined by J_{NO2}, which is the rate of photodissociation of $NO_2 + h\nu \rightarrow NO + O(^3P)$. In this study, J_{NO2} was set to 0.4 min^{-1} which is the medium value of 0.27–0.54 min^{-1} in an ambient condition within the mid-latitude region of Greece, as reported in a previous study (Gerasopoulos et al., 2012).

The calculated temperature range was 0–35 °C. The black line in Fig. 10 shows the results of the maximum O_3 (ppm) concentration calculated using the box model. The O_3 concentration varied between 0.32 and 0.66 ppm from the calculated temperature range, meaning that temperature is an important factor in O_3 formation. The red circle line in Fig. 10 shows the temperature dependence of the percentage of the change of O_3 concentration per unit temperature ($d[O_3]/dT$ (%/°C). $d[O_3]/dT$ is highest at 15 °C with the value of 3.2 %/°C. The temperature of 15 °C in Japan corresponds to the early spring and late autumn seasons. Assuming that the O_3 concentration is 50 ppb in spring and autumn seasons, the change of 3.2 (%/°C) corresponds to 1.6 ppb/°C. According to Fig. 3, mitigation of the UHI effect causes a maximum temperature decrease of 0.25 °C; therefore, the change in the O_3 concentration attributed to by the photochemical reaction is roughly estimated to be 1.6 ppb/°C × 0.25 °C = 0.4 ppb. In Section 4.1.2, enhancement of the NO-titration effect was attributed to a maximum of a 1 ppb decrease in O_3, indicating that the enhancement of O_3 photochemistry contributed to less than half of that of NO-titration effect when mitigation of the UHI effect occurred through introducing BEVs in the GTA."

Answer to referee 2:

General comment:

This manuscript conducted six parallel simulation experiments using WRF-CMAQ, to reveal the impact of introducing battery electric vehicles (BEVs) on the Urban Heat Island (UHI) effect, air quality, and associated health outcomes in the Great Tokyo Area (GTA) of Japan. The results indicated that replacing internal combustion vehicles with BEVs could contribute a maximum temperature decrease of 0.2 ℃ in the metropolitan GTA. A decline in ground O3 caused by the effects of BEVs is mainly due to reduced atmospheric chemical reactions and lower BVOC. While the impact on PM2.5 formation was more complex, with factors like particle coagulation and chemical reactions playing a role. The significant reduction in premature deaths (175 due to O3 and 77 due to PM2.5) underscores the positive impact of BEVs on air quality and public health in the GTA. Overall, this manuscript provides valuable insights into the environmental and health impacts of transitioning to BEVs in urban areas. The original intent and concept of the research are commendable, but the presentation and language organization are not ideal. Some issues still need to be improved.

The authors are grateful to reviewer 2 for providing significant suggestions and clarifications with sophisticated insights. The manuscript was modified based on the comments. The words and sentences with changes are highlighted in yellow. The language had been checked by the native proofreaders for the original draft, but this time, we asked the proofreaders to carefully check the English according to the opinions from reviewer 2, and the authors also checked the language carefully.

Aside from the provided opinions, to deepen the discussion, the authors added new insights of the contribution of NO-titration effect caused by lowered PBL height caused by UHI effect and lowered kinetics of photochemical reactions affected depression of O_3 which are discussed in the sections 4.1.2 and 4.1.3. The results suggested that enhanced NO-titration contributed maximally ~1ppb decrease of O_3 while lowered kinetics of O_3 formation contributed maximally ~0.4ppb decrease of O_3 in the GTA.

Further, the authors decided to include total amount of primary emissions of air pollutants in whole analyzed regions (GTA) for the six scenarios of this study which will enhance better understanding of our work to the readers. These information are added in the Appendix B of the main manuscript.

Specific comments:

1) For the emission change, the study considered both the reduction by the replacement of ICVs with BEVs, and the increase caused by electricity generation due to the use of BEVs. However, with regard to the AH, it appeared that only the decrease AH by the BEVs was considered, but the AH caused by the electricity generation was not taking into account. According to the calculation, the increase in electricity generation can reach around 30% of the GTA electricity consumption in 2017, which is not a negligible amount. Is there an oversight in the regard?

Thank you for pointing out the lack of information regarding the AH set in this study. As mentioned above, current UCM code implemented in WRF enables us to set only the averaged AH from all sectors in the calculating regions, and there are no options to set UCM parameters for each anthropogenic source (e.g. vehicle, residential, industry, etc.). The distribution of vehicles is expected to be nearly averaged over whole calculated domain of GTA (D3), so the analysis of AH could be done. On the other hand, power plants are condensed in the marine sides of D3 (please see Fig. S3 of the supporting information), and the averaging AH for the power plants could not be evaluated. We added this fact in the revised manuscript as follows:

Lines 134-138: "In addition, the introduction of BEVs is expected to increase AH from power plants due to the increased demand for electricity used in battery charging; however, the increased AH from power plants was not considered in this study. Power plants in GTA are located in specific areas near the bay, and are not distributed throughout the entire area; therefore, it was difficult to consider the effect of AH from power plants and its subsequent UHI effects in the GTA.".

We added this issue in the new section "4.6.1 Limitations and future perspectives" as follows:

Lines 519-524: "Second, also because of the limitations of the parameterization of UCM, the increase in AH from power plants due to the increasing demand for battery charging was not considered. According to Section A.1 of the Appendix A, there would be a 34% increased energy demand for battery charging and the associated AH increase may not therefore be negligible. To account for these issues, it will be necessary to update the program for the UCM model incorporated in the WRF, and collaboration between developers of the WRF and scientists in the field of atmospheric science is recommended in future work."

2) The calculations in the manuscript are mainly based on the statistical results in the year of 2017, including the car numbers and efficiency. However, it is evident that the reduction in UHI effect is closely related to factors such as car numbers, population, and electricity generation efficiency. How can the readers be convinced that the impacts of other factors on the UHI effect are essentially unchanged? Thus, how can conclusions regarding the replacement of ICVs with BEVs be drawn?

> We understand that the introduction of BEV will be completed in the future such as 2040's or later, and at the era, the social properties including other pollution issues (water pollution, soil pollution, etc.), population, economy, lifestyle of the people, etc. will be changed, and the source of UHI (and also primary air pollutants and climate) is expected to be different. Nevertheless, this study focuses on the separated effects of BEV-introduction to the air pollution, and the authors intentionally eradicated the other factors of future trajection to be evaluated the impact of BEVs to local climate and air pollution. If all the future factors are included in this study, it is difficult to follow which factors are important to O_3 and $PM_{2.5}$ pollutions (the main focus of this study). The evaluation of next-generation technologies including BEVs by CTM could be done only for the simplified assumption due to its huge calculation cost, and the mixed factors could be evaluated by more applied field such as life cycle assessment methods, socioeconomical modeling etc., although these methods are too simple to evaluate the detail impacts on climate and air pollution. Thus, this time, we focused only on the BEV's effects on UHI and air pollution. The issues of local climate and air pollutions hold non-linearity, so the results of our study only showed just one of the examples of the effects of BEVs to UHI and air pollution, and similar studies focusing on different features such as human lifestyle, economy, etc. should be accumulated in the future. The authors are hoping this study to be one of the indicators of the pros and cons of the introduction of next-generation technologies to mitigate the urban air pollution which could be used by the policymakers and environmental scientists.

3) The English in this manuscript need be polished, especially in terms of the expression of proper nouns.

> Thank you for the clarification, and we are sorry for improper expressions for some of the words and sentences. We asked the proofreaders to carefully improve the English again, and also the authors checked the English carefully again in the revision process.

4) Figures and their captions should be able to convey their meanings independently of the text. Please modify the captions of the figures so that readers can better understand the content depicted in the figures.

Thank you for the clarification. Some of the abbreviations, which are not relevant (e.g. BEVs, r_{SO4} etc.) written in the captions of the graphs were changed to non-abbreviated form. It is complicated to explain the detail of, for example, the sensitivity scenarios again in the caption, so the words "… five scenarios **listed in Table 1**" were added in the revised manuscript which will be easily followed by the readers.

5) Line 16-19: The sentence is too long to read and understand for the readers. Please rephrase it.

Thank you for the advice. The descriptions were divided into two sentences, and also we used comma to separate the expressions for readability as follows.

Lines 19-21: "The results indicated that mitigating the UHI effect would lead to a reduction in ground-level O_3 formation. This is due to the increased NO-titration effect caused by the lowered planetary boundary layer height, and due to the degradation of photochemistry related to O_3 formation caused by a decrease in temperature and biogenic volatile organic compounds (BVOC). The impact of the increased NO-titration effect on the degradation of O_3 caused by mitigation of the UHI effect was roughly estimated to be more than twice that of the degradation of photochemistry."

6) Line 39-40 and 43-45: The citation should be "Muratori (2018) suggested that ..." and "According to the review of Ulpiani (2021)....". Please check the whole manuscript about the citation form and make corrections.

We are sorry for the ununified expressions of citation. The related sentences were modified as follows:

Lines 37-44: "However, although direct exhaust emissions would be decreased, **previous work** suggested that the power demand would increase due to BEV battery charging (Muratori, 2018), which suggests an increase in emissions of primary air pollutants from power plants attributable to the introduction of BEVs. Nevertheless, anthropogenic heat (AH) from vehicles is expected to decrease through introducing BEVs, as engine exhaust emissions would be reduced. **Further,** the urban heat island (UHI) affects air pollutants through the following, all of which are correlated with each other: (1) changes in the kinetics of O_3 (and $PM_{2.5}$) formation, (2) changes in the air mixing ratio arising from the change in ambient temperature, and (3) changes in biogenic VOC (BVOC) emissions (Ulpiani, 2021)."

7) Line 93: "The first month was treated as the spin-off period...", the authors may want to say "spin-up period".

Thank you for the clarification. The word "spin-off" was modified to "spin-up".

8) Line 95: Do the authors use FNL data as the initial and boundary meteorological fields? Please clarify the specific dataset and its website.

FNL data was used as the initial and boundary condition of WRF. The citation of dataset was included in the first draft, but there were some mis-descriptions in the sentence, so we improved the sentences as follows:

Lines 96-98: "Objective analysis data describing the initial and boundary conditions were obtained at a resolution of 1° × 1° from the FNL Final Operational Global Analysis data provided by the National Center for Atmospheric Research (NCAR: National Center for Atmospheric Research archives, 2024)"

9) Line 98: "The moddelled results..", the authors may want to say "the simulation results...".

Thank you for the clarification. The word "moddelled" was changed to "simulation". Aside from this sentence, there were several sentences which used "modelled", and all the words have been changed to "simulated" according to the suggestion.

10) Line 125: The season division in a year is commonly utilize DJF, MAM, JJA and SON. Why does this study use January to March as winter, and so on for other seasons? Please provide the relevant basis.

We understand that several studies used DJF, MAM, JJA, and SON as the reference seasons. Despite this, since the previous studies, the corresponding author have used January, February, and March as winter, April, May, and June as spring, July, August, and Spring as summer, and October, November, and December as winter (Hata et al. (2022, 2023), Nakamura et al. (2023)) because of two reasons. First, the authors tried to analyze the air quality issues for all the months within the same year (2017 for this study) and second, we wanted to analyze the results of CTM with continuous months. March and December are the transition seasons from winter to spring and autumn to winter, respectively, and both months include high and low temperatures. In 2017 of Tokyo for example, the maximum temperatures of March and December were 13.4 °C and 11.1 °C respectively, while the minimum temperatures of those seasons were 4.2 °C and 2.7 °C respectively. This means, December is relatively more inclined to cold season than March, but March still has the similar feature of coldness. We have prioritized the analysis within the same year, and the seasonal analysis with contentious months since the past.

References:

Hata et al. Urban-scale analysis of nitrogen deposition in Japan: Validation of chemical transport modeling and the sensitivity of anthropogenic nitrogen emissions to dry and wet depositions. *Atmos. Environ.* 2022, 275, 119022

Hata et al. Impact of introducing net-zero carbon strategies on tropospheric ozone (O_3) and fine particulate matter ($PM_{2.5}$) concentrations in Japanese region in 2050. *Sci. Total Environ.* 2023, 891, 164442.

Nakamura et al. Urban-scale analysis of the seasonal trend of stabilized-Criegee intermediates and their effect on sulphate formation in the Greater Tokyo Area. *Environ. Sci. Atmos.* 2023, 3, 1758-1766.

11) Figure 2-4: The simulated and observed values for different seasons are presented in the same subplot, with scatter points of different seasons overlapping, making it difficult to showcase their distribution characteristics. Could the scatter points for different seasons be displayed in separate subplots? Besides, the x- and y- axes are labeled as "observed" and "calculated". Please provide the specific quantities and units for labels in each figure. Lastly, what do the reference lines in the figures represent? Please include the information in the figure captions.

Thank you for the important suggestions regarding simulation performances of WRF and CMAQ. We agree to the opinion in terms of the difficulty to check the seasonal trend from the previous scatter plots. In the revised version, yearly trend graphs of daily-averaged ground temperature, wind speed, wind direction, shortwave radiation, MDA8 O_3, and daily-averaged $PM_{2.5}$ were shown to show the validation of simulated results. The targeted sites have been increased from two sites (Tokyo and Gunma) to seven sites (added Kanagawa, Saitama, Tochigi, and Ibaraki). Because of the increase of the graphs, we decided to move the section of simulation performance to Appendix C: "Appendix C: Validation of simulated meteorology, O_3, and $PM_{2.5}$ with the observed results". Because of the change from scatter plots to yearly trend, reference lines were removed.

12) Line 176-180: The simulation results for PM5 and O3 in Gunma are not as ideal as in Tokyo. Apart from the differences in land-use and chemical factors as described, could this also be attributed to the simulation performance of meteorological factors? The manuscript only provides validations for temperature regarding meteorological factors, while the simulation performances of meteorological factors such as wind fields and boundary layer height, which are relevant to the dispersion and distribution of pollutants, are not presented.

Thank you for the important clarifications. The authors agree to the opinion that two reference sites are not enough to show simulation performance (in the first draft, we intended to minimize the graphs, and only two sites were chosen). As mentioned in above, five sites (Kanagawa, Saitama, Tochigi, and Ibaraki) were added, and totally seven sites were analyzed. Further, according to the suggestion, the meteorological parameters of wind speed, wind direction, and total shortwave radiation were also implemented in the revised manuscript. Although there were a few sites which showed low simulation performance sometimes, overall trend and replicability showed good results. Again, because of the increase of the graphs, we decided to move the section of simulation performance to Appendix C: "Validation of simulated meteorology, O_3, and $PM_{2.5}$ with the observed results". The added discussions in Appendix C are as follows.

Lines 636-647: "Figures C1–C4 show the correlation between observed and calculated results for the daily ground temperature, wind speed at 2-m height, wind direction, and total solar radiation for winter, spring, summer, and autumn in the seven analyzed sites in 2017. Note that the observed data of daily total solar radiation in Kanagawa, Chiba, Saitama, and Ibaraki were not available, so the remained three sites were compared.

Furthermore, in terms of the wind direction shown in Fig. C3, the value was defined by the 16 directions: the value of 0 corresponds to 0° (north), 8 corresponds to 180° (south), and 15 corresponds to 337.5° (north-northwest). The correlations are exhibited using four statistical factors: correlation factor (*R*), root mean square error (RMSE), normalized mean bias (NMB), and normalized mean error (NME). Overall, the model was found to replicate the observed results well in all seasons. The results for the wind direction for Saitama and Tochigi show low *R* values of 0.19 and 0.29, respectively, but the time trend and the values of RMSE, NMB, and NME are similar to those of other sites. Unlike O_3 and $PM_{2.5}$, there were no proposed indicators for the statistical values described in later sentences, but the time trend of simulated results shown in Figs. C1–C4 replicated the observed results, and we concluded that the simulated meteorology could be applied in the evaluation of this study."

Lines 661-677: "Figure C5 shows the correlation between the observed and calculated (BASE) results and 8-h daily maximum average (MDA8) O_3 concentrations. As seen in Fig. C5, overall, the modeled O_3 replicated the observed results well, and the *R* value was more than 0.7, except for that of Gunma. Figure C5(f) suggests a lower correlation for the modeled O_3 with the observed results for Gunma. Emery et al. proposed indicators that can be used to validate O_3 and $PM_{2.5}$ for chemical transport modeling (Emery et al., 2017). According to these indicators, the ideal values of the modeling performance for MDA8-O_3 should be $R > 0.75$, NMB < ±0.05, and NME < 0.15 while the criteria value should be $R > 0.50$, NMB < ±0.15, and NME < 0.25. All the sites except for Gunma fully met these criteria and were close to the described goals. However, Gunma did not meet the criteria for *R* and NME. This may be because it is located in the countryside, where less primary air pollutant emissions are generated than in highly polluted areas, and the transportation from other regions renders it difficult to predict O_3 by CTM. Nevertheless, most of calculations for the analyzed sites showed good agreements with the observed results, and are thus considered to be acceptable for analyzing this study. Fig. C6 shows the correlation between the observed and calculated (BASE) results of the 24-h daily average (DA24) $PM_{2.5}$ concentrations. Emery et al. proposed that ideal values of indicators used to the modeling performance of DA24 $PM_{2.5}$ should be $R > 0.70$, NMB < ±0.10, and NME < 0.35, while criteria should be $R >$ 0.40, NMB < ±0.30, and NME < 0.50 (Emery et al., 2017). According to Fig. C6, all the modeled results except for those of Kanagawa met the criteria, and some R and NME values were close to the goal. The result of Kanagawa shown in Fig. C6(b) replicated the daily trend of the observed results. This analysis therefore shows that although relatively less accuracy was obtained for simulated $PM_{2.5}$ in Kanagawa, the modeled conditions could be applied in the analysis of this study."Finally, please note that the simulation performance of Gunma for ground-level O_3 was not as good as other sites. Therefore, in the revised manuscript, we decided to show the results of sensitivity analyzes of O_3 and $PM_{2.5}$ for Tokyo and Tochigi (instead of Gunma), which are shown in Figs. 5 and 6 in the revised manuscript.

13) Equation 1-2: Are the square boxes in the equations clerical errors? Please verify.

We are sorry for inconvenience of the equations. Microsoft Word was used to write the manuscript, and there is no function to make single superscript in the left side of character when using equation mode. There is only the function of combined suffix and superscript, and thus, square boxes remained. The authors will notify this issue to the editorial staff. We are sorry for inconvenience.

14) Line 517-519: It is described that "The ratio of the decrease in the AH after the introduction of BEVs was calculated to be 0.35." However, in the calculation, a addition of 157.6TJ/d was made to account for AH from other vehicle types. Please make confirmation whether it is decreased to 0.35 of AH through replacement of ICVs with BEVs, or decrease 0.35, and make modification correspondingly.

The authors feel sorry for the complicated expressions in the sentences. Actually, 0.35 is the decrease ratio of AH from the substitution of ICVs to BEVs for cars and small heavy-duty vehicles. On the other hand, the later sentences indicate the "total AH from vehicles" which could be calculated by the summation of AH from non-electrified vehicles (large type heavy-duty vehicles including bus) and from BEVs. The following explanation was added after the equation 505.8 – 451.5 + 157.6 = 211.9 TJ/d as follows:

Lines 599-600: "the calculation of 505.8 - 451.5 corresponds to the AH from the non-targeted vehicles for electrification, and 157.6 corresponds to the AH from BEVs".

15) Line 546: "All the data simulated I the study... " should be "All the simulation data in the study ...".

Thank you for the clarification. The sentence was modified.

Other important changes:

Please note that reviewer 1 suggested to detail the discussion, we tried to clarify how the mitigation of UHI effect affected NO-titration and photochemical cycle of O_3 formation via analysis of the results of CTM and via box modeling method. The finding of these analysis will be beneficial to the field of atmospheric science.

> Lines 316-332: “Compared with the spring and summer, a wider distribution of high ΔTO was observed for winter and autumn, and more relevant NO-titration effects were estimated in these seasons; these were presumed to be caused by the intense decrease in temperature caused by mitigation of the UHI effect. Haman et al. conducted observational and CTM calculations based on Houston (U.S.) from 2008 to 2010 to clarify the relationship of diurnal temperature, PBL, and ground-level O_3 (Haman et al., 2014). The results suggested a positive relationship between ground-level O_3 and the PBL height; this was due to the enhancement of NO-titration caused by relatively weak wind speeds during the lower PBL height which caused NO_X to remain on the ground’s surface. Figure 9(e)–(f) shows the seasonal changes in ground-level O_3 between the BASE and SUHI scenarios (ΔO_3). A higher decrease in O_3 is estimated in the central area of GTA, and winter and autumn show a high O_3 decrease; these results are confirmed in Fig. 5. The seasonal variations in the positive increase in O_3 shown in Fig. 9(e)–(f) correspond to ΔTO shown in Fig. 9(a)–(d). However, the distributions of ΔO_3 shown in Fig. 9(e)-(f) do not correspond to the distributions of the H_2O_2/HNO_3 ratio. The results indicate that O_3 reductions caused by the mitigation of UHI effects are attributed by the enhanced NO-titration effect rather than the photochemistry involved in O_3 formation. Enhancement of the NO-titration effect contributes to a maximum decrease in O_3 of ~1.0 ppb. Despite these facts, Haman et al. suggested that the relationship between PBL and the NO-titration effect strongly depends on factors such as the region analyzed and meteorology (Haman et al., 2014). Future work should determine the relationship between UHI effects and O_3 in other regions. The impact of the change in temperature on photochemical reactions is discussed in the next section.”
>
> Lines 337-362: “4.1.3 Box model simulation of the temperature dependence of ozone formation chemistry
>
> Section 4.1.2 discusses the decrease in ground-level O_3 caused by the enhancement of NO-titration effect due to the lowered PBL height from mitigation of the UHI effect. Mitigation of the UHI effect causes a local-temperature decrease and is expected to

> weaken the rate of photochemistry involved in O_3 formation. For example, Coates et al. suggested that O_3 formation is enhanced through the increase in temperature due to the increased reaction rate of VOC oxidation and peroxy nitrate decomposition (Coates et al., 2016). Meng et al. suggested that high temperatures lead to a high HO_2 + NO reaction rate, which increases NO_2 and contributes to high O_3 episodes (Meng et al., 2023). To clarify the effect of temperature changes on the photochemistry of O_3 formation, a box model calculation using SAPRC-07 was conducted in this study. The CO emissions were the highest in the D4 region (as shown in Fig. B1 of the Appendix B) and the CO concentration was set to 1 ppm. The concentrations of the remaining pollutants, including NO_X, SO_2, AVOCs, and BVOCs, were set using the ratio between the annual amount of each emission in the BASE scenario shown in Fig. B1 to Fig. B4 and that of CO. The detail settings of the concentrations of the species for box model simulation are listed in Table S4 of the SI. The concentration of H_2O was fixed to 1.56×10^4 ppm; this was estimated from the value of partial vapor pressure of H_2O at 25 °C with 50 % relative humidity. Solar intensity was defined by J_{NO2}, which is the rate of photodissociation of $NO_2 + h\nu \rightarrow NO + O(^3P)$. In this study, J_{NO2} was set to 0.4 min^{-1} which is the medium value of 0.27–0.54 min^{-1} in an ambient condition within the mid-latitude region of Greece, as reported in a previous study (Gerasopoulos et al., 2012). The calculated temperature range was 0–35 °C. The black line in Fig. 10 shows the results of the maximum O_3 (ppm) concentration calculated using the box model. The O_3 concentration varied between 0.32 and 0.66 ppm from the calculated temperature range, meaning that temperature is an important factor in O_3 formation. The red circle line in Fig. 10 shows the temperature dependence of the percentage of the change of O_3 concentration per unit temperature ($d[O_3]/dT$ (%/°C)). $d[O_3]/dT$ is highest at 15 °C with the value of 3.2 %/°C. The temperature of 15 °C in Japan corresponds to the early spring and late autumn seasons. Assuming that the O_3 concentration is 50 ppb in spring and autumn seasons, the change of 3.2 (%/°C) corresponds to 1.6 ppb/°C. According to Fig. 3, mitigation of the UHI effect causes a maximum temperature decrease of 0.25 °C; therefore, the change in the O_3 concentration attributed to by the photochemical reaction is roughly estimated to be 1.6 ppb/°C × 0.25 °C = 0.4 ppb. In Section 4.1.2, enhancement of the NO-titration effect was attributed to a maximum of a 1 ppb decrease in O_3, indicating that the enhancement of O_3 photochemistry contributed to less than half of that of NO-titration effect when mitigation of the UHI effect occurred through introducing BEVs in the GTA.”

In the revised manuscript, we also added the summary of the amount of emissions in D3 for each emission scenario with Figs. B1-B4 in Appendix B as follows.

> Lines 605-608: “Fig. B1 shows the annual total emissions of NO_X, CO, SO_2, and NH_3 in the four scenarios (BASE, ALL, SEV, and SPP defined in Table 2) in 2017 within the GTA (kt y^{-1}). While all the emissions would decrease due to the introduction of BEV (SEV), the increased emissions from power plants would partly offset or increase total NO_X, CO, SO_2 (SPP) emissions. The relevant decreases in NO_X (~20 %) and CO (~60 %) emissions are shown in ALL and S_{EV} scenarios.”

> Lines 612-616: “Fig. B2 shows the total annual AVOC emissions for the four scenarios (BASE, ALL, SEV, and SPP, defined in Table 2) in 2017 within the GTA (kt y^{-1}). The species of “Others” in Fig. A2 include alcohol and acetylene. The decrease in AVOC emissions from BASE to ALL scenarios mainly relates to the decrease in exhaust and evaporative emissions from service stations; however, the increase in emissions from power plants is almost negligible. A high decrease in alkane (~23 %) is expected in relation to the introduction of BEVs.”

> Lines 621-624: “Fig. B3 shows the total annual $PM_{2.5}$ emissions for the four scenarios (BASE, ALL, SEV, and SPP, defined in Table 2) in 2017 within the GTA (kt y^{-1}). Unlike NO_X, CO, and AVOCs, none of the $PM_{2.5}$ components show a relevant decrease following the introduction of BEVs. This means that the $PM_{2.5}$ emitted from vehicle exhaust emissions has almost no effect on total $PM_{2.5}$ emissions.”

> Lines 628-631: “Fig. B4 shows the total annual BVOC emissions (isoprene and monoterpene) for the two scenarios (BASE and SBVOC, defined in Table 2) in 2017 within the GTA (kt y^{-1}). The emissions in the two scenarios are almost equal, and only ~0.5 % of BVOC emissions decrease from the BASE to SBVOC scenarios. This decrease was related to mitigation of the UHI effect through the introduction of BEVs.”