Reviewer 3:

6. This manuscript uses large-eddy simulation to analyse the effects of vehicle-fleet electrification on aerosol concentrations. In particular, the PALM LES model and SALSA aerosol module are applied to a Helsinki neighbourhood by examining statistics within two regions of interest (a boulevard and roundabout) under three traffic scenarios, S1 (2022 baseline), S2 (2035) and S3 (2040). The main finding is that particle number (PN2.5) decreases but particle mass (PM2.5) increases the manuscript addresses a problem of great practical importance. It is very clearly and attractively written. Nevertheless, I think the analysis could be improved. Briefly, the authors should show that their results are robust and could not be obtained through simpler means.

Major points:

1. The simulations are restricted to a single winter morning, with data collected from 8:00–9:00 local time being analysed. How representative is this choice? I am unfamiliar with meteorological conditions in Finland, but I imagine that the wind speed and wind direction show considerable temporal variations. If so, then how would the main findings be affected? Ideally the authors should demonstrate that the findings are robust for different meteorological conditions.

Response: We agree with the reviewer that wind speed, direction, and atmospheric stability exhibit temporal variation, and that a single period cannot capture the full range of annual conditions.

The primary aim of this study was to evaluate the possible air quality impacts of different fleet electrification scenarios under a worst-case meteorological scenario. This motivated our initial focus on this specific winter case.

However, to address the reviewer's concern and demonstrate the robustness of our conclusions across varying meteorological conditions, we have conducted a new set of simulations representing summer convective conditions, featuring different wind speeds, wind directions, and enhanced thermal turbulence.

The results from these additional sensitivity runs have been integrated into Section 3.5 in the revised manuscript.

2. The predictions would be more useful if they included estimates of the associated uncertainty. Meteorological variability aside, the aerosol fields also depend on various parameters (e.g. fleet composition, traffic volume and LAD) that are not known precisely. The authors should confirm that the findings are unaffected by the uncertainty.

Response: We agree that input parameters like fleet composition, traffic volumes, and LAD carry inherent uncertainties in future urban planning contexts. However, because this is a comparative scenario analysis, these uncertainties do not affect our core findings.

Baseline and projected fleets are not arbitrary; they are prescribed directly from official municipal planning models for Helsinki.

Our three transition scenarios effectively act as a sensitivity analysis that bounds fleet composition uncertainty.

The spatial layout and LAD of the planned vegetation are kept strictly identical across all scenarios. Any minor parameterization errors in LAD or building drag propagate systematically, meaning they cancel out when calculating the relative percentage differences between scenarios.

Regarding meteorological uncertainty, we have directly addressed meteorological variability through our new seasonal sensitivity analysis presented in section 3.5.

> 3. The sensitivity to the initial size distribution is not considered. According to the text (l. 67), "for S2 (2035) and S3 (2040), the size distributions were derived using the baseline distribution (S1), with the number of particles adjusted according to the reductions in total EF (Table 1)". This seems to imply that the size distribution is determined only by the emission factors (units of particle number per unit length); however, the size distribution should depend on the actual traffic and not just the emission factors (which are essentially normalised quantities). Given the uncertainty in estimates of future traffic, some uncertainty in the size spectra seems unavoidable; the authors should quantify it and confirm that their findings are unaffected.

Response: We thank the reviewer for this critical comment. We agree that the size distribution is highly dependent on actual traffic characteristics, and that future fleet variations introduce unavoidable uncertainties.

As we have explicitly acknowledged in Section 3.6 (study Limitations), a major challenge in current microscale modeling is the complete lack of real-world, measured aerosol size distribution data specifically for fully electric vehicles and their associated non-exhaust emissions.

In the absence of such observational data, scaling the baseline (S1), which is a measured size distribution using the scenario-specific emission factors (EFs) was the optimal approach.

Because this is a comparative study, the relative differences and spatial patterns of pollutant mitigation between the scenarios (S1 vs. S2 vs. S3) are predominantly driven by the shifts in emissions at the source. While the absolute shape of the future size spectrum carries some uncertainty, the opposing physical trends across the transition scenarios, specifically, the reduction in total particle number concentrations alongside the slight increase in particulate mass due to non-exhaust EV emissions remain structurally and physically robust.

We have revised Section 3.6 to further emphasize this limitation, explaining how the lack of EV-specific size distribution measurements affects absolute numbers while clarifying why our comparative scenario insights remain unaffected.

4. If the representativeness of the meteorological conditions, or the insensitivity to the parameters or size distribution cannot be established, then the findings would seem to be essentially qualitative in nature (e.g. electrification decreases PN2.5 but increases PM2.5). In theory, this is fine, but the authors should show that similar conclusions cannot be obtained by simpler means. I suspect that it may be possible to obtain similar results using RANS. Indeed, a RANS-based study, which would allow for changes in meteorology to be examined at minimal cost, could be preferable.

Response: We thank the reviewer for this technical comment. While RANS frameworks are computationally efficient and suitable for broader assessments, they are fundamentally inadequate for the specific objectives of this study for several reasons:

1. Sectional Aerosol Tracking (PALM-SALSA): Resolving size-segregated, multi-component aerosol dynamics across discrete size sections requires accurate representation of local turbulent mixing and surface interactions. Key size-dependent processes, such as particle dry deposition resolved by PALM-SALSA, depend non-linearly on localized velocity gradients and near-surface turbulent fluctuations that steady-state RANS models average out.
2. Transient Urban Turbulence: Pollutant dispersion in complex street canyons is dominated by transient, intermittent turbulent structures (e.g., building wake vortices, ext). RANS models solve only for time-averaged fields and doesn't capture these short-term eddies and systematically misrepresenting street canyon ventilation rates.
3. Microscale hot-spots and exposure: Our primary objective is to provide quantitative spatial guidance for urban planning by identifying pedestrian-level exposure hot-spots. Time-averaged RANS modeling smooths out peak concentration gradients, making it unreliable to evaluate localized exposure zones along the planned boulevard.

We have added a brief paragraph to the Introduction section (line 34–44) of the manuscript to clearly justify the necessity of LES over RANS for this specific aerosol-dispersion application.

5. The authors explain the contrasting responses for PM2.5 and PN2.5 by noting that the increased weight of BEVs leads to more tire and brake wear and an increase in non-exhaust emissions (NEE). Although I do not work in this area, it is my impression that this point is widely understood. It therefore seems to me that the ms would be stronger and of greater scientific interest if the relationships between decreased tailpipe emission, increased car weight, decreased PN2.5 but increased PM2.5 were examined more critically. For example, does PN2.5 always decrease while PM2.5 increases? Are there conditions under which PM2.5 does not necessarily increase? The ms may be viewed as a case study of sorts, which is also of interest, but its general applicability is unclear.

Response: We thank the reviewer for the comment. Even though BEVs are known to increase NEE. There are no studies which would have quantified their possible impact on street-level aerosol concentrations. The need for this kind of studies is evident also from the involvement and interest of the city of Helsinki on the study and its results. Similarly, our results have gained a lot of interest from the research community in conferences as we give indicative values how much the effect from changing emissions might be. Thus, we would argue that there specifically is a need for this kind of studies among scientific community and beyond.

We agree that it would be interesting to examine the relationships of the emissions with vehicle weight and other factor, but that would be out of the scope of the present study.

From a meteorological standpoint, as demonstrated by our new summer convective sensitivity simulations (Section 3.5), ground-level PM2.5 accumulation does not occur uniformly under all conditions. For instance, under unstable summer convective mixing, strong vertical transport enhances atmospheric dilution of non-exhaust mass, resulting in substantially lower surface-level PM2.5 concentrations compared to stable winter inversion conditions. Nevertheless, the direction of the mass trade-off, where non-exhaust emissions cause relative mass increases alongside fleet electrification remains consistent across both thermodynamic regimes.

Therefore, while our numerical estimates serve as directional indicators due to inherent uncertainties in future NEE size distributions, emission factors, vehicle weight evolution, and background concentrations, the underlying physical processes apply broadly to similar urban environment, with rather similar street layout, meteorology undergoing fleet transition strategies.

We have clarified these points in section 3.6 (study limitations).

Specific points

1. l.70 'We predict the fleet composition for years 2035 and 2040 (with a base-line in year 2022)' Some additional details on the prediction would be helpful. LES is the only numerical method mentioned in the introduction.

Response: We thank the reviewer for this suggestion. To clarify this early in the manuscript, we have introduced the fleet composition model (SALAMA) in the Introduction (lines 70–72). Additional detailed information regarding the model and fleet projection methodology has been expanded in Section 2.3 (lines 106–114).

2. l.115 'The temporal and spatial distribution of traffic within the study area was predicted using a multi-layer Long Short-Term Memory (LSTM) network (Leinonen et al., 2024)' The procedure used in the present study should be summarised.

Response: We thank the reviewer for this suggestion. We have expanded Section 2.3 (Methodology) to summarize the LSTM traffic prediction procedure used in this study (line 121–125).

3. Table 2 The root domain seems very large. It is around 30 times longer than the child domain. Is this necessary?

Response: The spatial extent of the root domain was selected based on three key physical and modeling considerations:

1) The root domain is driven by lateral boundary conditions from the MEPS model, which has a native grid resolution 2.5 km. A sufficiently large domain is required to properly ingest and resolve the mesoscale spatial gradients across multiple MEPS grid cells without introducing lateral boundary circulation errors.

2) Inflow from mesoscale models enters as smooth, domain-averaged profiles lacking 3D turbulent structures. A wide root domain provides adequate spatial fetch for the flow guided by the Synthetic Turbulence Generator (STG) to adjust to the underlying aerodynamic roughness and topography, ensuring that fully developed turbulent boundary-layer flow reaches the nested child domains.

3) The greater Helsinki area is characterized by complex land-sea transitions, varied surface roughness (forested areas, low-rise residential, and dense urban canopy), and local thermal contrasts. The extent of root domain captures these upstream fetch influences and coastal boundary-layer developments before the flow enters the target study area around the Vihdintie street canyon.

4. Figure 6 caption. What is meant by 'averaged along the boulevard'? Presumably a 1-D average was taken. If so, is the topography averaged as well?

Response: Yes, a 1D spatial average of the flow variables was calculated along the longitudinal axis (y-direction) of the boulevard and roundabout sections. Regarding the topography, building and terrain grid cells are treated as NaN (masked) in the simulation domain and are excluded from the calculations. We have revised the caption of Figure 6 for clarity.

5. Figure 9 Are these 2-D spatial averages?

Response: No, the boxplots do not represent spatial averages. The underlying data consists of 2D time-averaged fields (08:00–09:00) at 2m height. The boxplots show the spatial distribution across all individual grid cells within the specified domain (boulevard or roundabout) by flattening the 2D grid matrix. We have updated the caption of Figure 9 to make this clear.

6. l.283 'The average PM2.5 increases by 7% and 15%, respectively, within the child domain. This is consistent with Beddows and Harrison (2021), who reported a 21% increase in particulate mass due to fleet electrification. Using a weight-dependent emission model, they attributed this increase to the increased curb weight of BEVs [. . .]' Why are the authors' results consistent with this? Presumably the curb weight enters indirectly through the emission factors. This is fine, but could similar results be obtained directly from the emission factors and traffic volume? If not, this would underscore the usefulness of a CFD study.

Response: In our setup, BEV curb weight is accounted for indirectly through elevated non-exhaust emission factors (brake, tire, and road dust wear). Our results are consistent with the physical mechanism described by Beddows and Harrison (2021), where increased vehicle curb weight (~21%) leads to higher non-exhaust particle emissions. We agree that the original comparison caused confusion, and we have revised this paragraph accordingly.

Regarding whether similar results could be obtained directly from traffic volumes and emission factors, the short answer is No, however multiplying emission factors by traffic volume can estimates total emission rates, but it is fundamentally incapable of predicting microscale dispersion, vertical concentration gradients, or localized human exposure. High-resolution CFD modeling (such as LES) is needed here because it resolves complex street-canyon flow fields, building-wake recirculation, and stability-dependent vertical transport. Without resolving these microscale fluid dynamics, bulk emission calculations cannot capture how for instance mass accumulates in specific ground-level hotspots while being suppressed or vented at different vertical levels. This highlights the necessity of high-resolution CFD over simple emission factor scaling for informing urban planning decisions and guiding microscale air quality assessment.