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.

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.

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.

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.

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.

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.

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.

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

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?

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

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.