Reviewer comments for:

Mobile Air Quality Monitoring and Comparison to Fixed Monitoring Sites for Quality Assurance

This review is for the above manuscript submitted for publication in Atmospheric Measurement Techniques. The manuscript partially develops and proposes implementation of a new quality assurance procedure to evaluate changes in instrument performance in mobile monitoring of air quality. To do that, the authors use high-temporal resolution (O ~ 1s) mobile-monitoring data collected using regulatory-graded instruments from two campaigns conducted in different regions for very different lengths of time for three pollutants, O3, NO2, and NO. The authors then compare stationary referencing of this data during collocations with the regulatory monitor to the referencing of "vehicle-in-motion" concentrations with regulatory monitoring data (based on distance and road type from the regulatory monitor) for one site, and find similar performance evaluations for a regionally distributed pollutant O3 ($r2 > 0.9$), moderate performance for NO2 ($r2 > 0.5$), and poor performance for a primary pollutant, NO ($r2 < 0.2$) in their new approach. For the second site, the authors do not conduct stationary referencing and only perform the latter "vehicle-in-motion" referencing to the regulatory monitor to estimate optimal temporal "running windows" to identify instrument issues. They calculate that for a 3 km spatial window, a temporal running window of 40 hours for data would allow detection of a systematic measurement drift or sudden instrument or sensor malfunction over the time scale of 7-9 days. While the authors motivate this work to identify and address systematic measurement drift or malfunction, the same is are not demonstrated in their results nor do they show the implementation of this method on any dataset. Additionally, in its current format, the manuscript uses linear models even when they do not seem applicable, especially for NO and NO2. Finally, the deficient analysis of high-concentration plumes further limits the usefulness of the proposed method. I recommend that the authors significantly revise and resubmit this manuscript for further consideration.

1. Deficient motivation and lacking significant findings: The authors begin by motivating this work as to "provide an important tool for ongoing quality assurance during mobile measurement campaigns" [Line 51]. They believe that "through ongoing comparisons of fixed reference site and mobile measurements, it may be possible to identify instrument drift over time or changes in instrument performance that could indicate a malfunction" [Lines 47-49]. The other motivation the authors suggest is that ongoing "mobile-versus-fixed-site comparisons are more scalable than frequent site-by-side parked collocations", "which is particularly important during sustained, multi-vehicle (and fleet-based) mobile monitoring campaigns" [Lines 50, 57, 58]. However, in their findings, the authors conclude and I agree that **this method "is not an absolute method for calibration or instrument verification, as a direct collocated comparison with reference monitors is"** [Line 458]**.** In the two campaigns the authors conducted, they performed daily checks with zero and span gases and in one campaign, conducted direct collocation comparisons. These **standard approaches provide lab-grade confidence in the measurements including with regards to instrument drift and malfunction**

and no replacement for them has been identified in this work. While authors do find that regional pollutants are correlated strongly in “mobile-versus-fixed-site comparisons”, these pollutants are expected to exhibit regional homogeneity and this result is not a significant contribution of this work. Frequently, measurements dominated by secondary pollutants are referenced to nearest (but far-located) reference monitors, both in stationary and mobile monitoring. **What the authors in fact demonstrate is the weakness of the “mobile-versus-fixed-site comparisons” for pollutants with high spatial variability such as NO.** While 15-16 stationary collocations only 20 mins each conducted over 2-3 weeks in the Denver campaign to calibrate against regulatory monitors yield mean r2 of 0.4, this comparison performance drops precipitously to r2 <= 0.2 in hourly averaged “mobile-versus-fixed-site comparisons” (Table 3). This is not surprising, since “spatial coverage from mobile monitoring reveals patterns missed by the fixed-site network”, especially for primary pollutants (Chambliss et al., 2020). Anyway, scaling their standard stationary collocations up to a year (approximately equivalent to the length of the California campaign, also presented here) totals to about 60 hours. In this work, the authors propose sampling in their newly developed approach within a crude spatial scale of 3 km for about 40 rolling hours to identify instrument drift/malfunction. Clearly, there is little advantage to switching to this new approach given the drop of ~50% in the measure of performance in the “mobile-versus-fixed-site comparison” in the only primary pollutant monitored compared to the standard “stationary collocation”. In short, while the authors argue that stationary collocations “ensures comparability only at that specific location and only under the specific atmospheric conditions over which the collocation occurred” (Lines 37-38), **they have demonstrated that stationary collocations perform significantly better than their proposed method.**

2. Use of linear models: As Figures 1 and 4 demonstrate, a linear model seems insufficient to compare measurements of NO and NO2. There is a clear baseline effect, where only a small fraction of variance in concentrations can be explained by variations in the reference site. The extensive dependence on presentation using linear models in the manuscript for the Denver phase further weakens this manuscript. I suggest the authors’ reconsider the presentation approach as well as the feasibility of a “mobile-versus-fixed-site comparison” given visible baselines that are developed primarily based on comparison with the Denver “stationary collocations”.

3. **Studying outlier plume events**: Hyperlocal monitoring has a lot of value given its ability to map pollutant exposures, especially plumes of primary pollutants with high spatial variability. However, prior work suggests that primary pollutant spatial patterns can be predicted well using land-use regressions (Robinson et al., 2019). In contrast, regionally distributed pollutants are not well represented by such regressions but are relatively spatially homogeneous and can be estimated using regional monitoring (Shah et al., 2018). Given these well-established priors, the baseline effect in (2) above, the high-concentration plumes or “outliers” in primary emissions should be studied separately. Similar mobile monitoring work has been done previously and could be referenced (Robinson et al., 2018).

1. Chambliss, S. E., Preble, C. V., Caubel, J. J., Cados, T., Messier, K. P., Alvarez, R. A., LaFranchi, B., Lunden, M., Marshall, J. D., Szpiro, A. A., Kirchstetter, T. W., and Apte, J. S.: Comparison of Mobile and Fixed-Site Black Carbon Measurements for High-Resolution Urban Pollution Mapping, Environ. Sci. Technol., 54, 7848–7857, https://doi.org/10.1021/acs.est.0c01409, 2020.
2. Robinson, E. S., Gu, P., Ye, Q., Li, H. Z., Shah, R. U., Apte, J. S., Robinson, A. L., and Presto, A. A.: Restaurant Impacts on Outdoor Air Quality: Elevated Organic Aerosol Mass from Restaurant Cooking with Neighborhood-Scale Plume Extents, Environ. Sci. Technol., 52, 9285–9294, https://doi.org/10.1021/acs.est.8b02654, 2018.
3. Robinson, E. S., Shah, R. U., Messier, K., Gu, P., Li, H. Z., Apte, J. S., Robinson, A. L., and Presto, A. A.: Land-Use Regression Modeling of Source-Resolved Fine Particulate Matter Components from Mobile Sampling, Environ. Sci. Technol., 53, 8925–8937, https://doi.org/10.1021/acs.est.9b01897, 2019.
4. Shah, R. U., Robinson, E. S., Gu, P., Robinson, A. L., Apte, J. S., and Presto, A. A.: High-spatial-resolution mapping and source apportionment of aerosol composition in Oakland, California, using mobile aerosol mass spectrometry, Atmospheric Chemistry and Physics, 18, 16325–16344, https://doi.org/10.5194/acp-18-16325-2018, 2018.