

## ***Interactive comment on “Using Low Cost Sensors to Measure Ambient Particulate Matter Concentrations and On-Road Emissions Factors” by K. K. Johnson et al.***

**Anonymous Referee #4**

Received and published: 24 February 2016

This manuscript aims to assess the reliability of 3 low-cost Shinyei-brand PM sensors via comparisons with two other PM measurement methods, a TEOM and an E-BAM, for 3 urban locations. While not clearly explained in the mspt, the output from the Shinyei sensors apparently is uncalibrated (“raw”), so sensor reliability was assessed based on how strong the correlation was between the Shinyei’s raw output and the reference monitor’s reported mass concentration. For a small set of lab experiments, the Shinyei sensors were compared with a DustTrak.

The authors also utilized a low-cost CO<sub>2</sub> sensor (COZIR GC-0010), showing a comparison with a reference monitor for one location. While a “mid-cost” MicroAeth BC sensor was also deployed, no comparisons to reference monitors were made in this

C1

study.

Characterizing the performance of new monitoring devices under field conditions is a worthwhile goal. However, the number of sites measured (3), the small range of concentrations seen at 2 of the sites, and the analytical insights seem insufficient to contribute to this goal. The only insight this reader gained from the manuscript is that the Shinyei sensors are unreliable – the correlation with reference monitors is much too low to view the devices as even semi-quantitative. In addition, given that they fail to capture temporal trends some of the time (p.6, line 25), they cannot be viewed as qualitatively reliable, either. I am not sure if this modest insight (which was already known for one of the models), by itself, merits publication. If it does, then there are a number of other major issues needing to be addressed, as itemized below.

Major issues:

1) The manufacturer reports different lower particle size detection limits for different models (ranging from 0.5 – 1  $\mu\text{m}$ ; page 3, line 28). For a motor vehicle-dominated site, one would expect much of the mass to be submicron – this would lead to differences in measurements for the different models, as well as with the TEOM (which uses a filter with a very high capture efficiency for submicron particles). The mspt needs to discuss how/why a sensor that only detects down to 1  $\mu\text{m}$  would be effective at estimating EF values for combustion emissions.

2) While I did not see this mentioned until p.8 (line 33), it appears that the Shinyei sensors were measuring total PM, with no size cut. I presume that the E-BAM was operated with a PM<sub>2.5</sub> inlet, although this is never stated in the mspt. However, I believe the model 1400a TEOM is designed for PM<sub>10</sub> measurements. The authors need to explain why the plots, and much of the results/discussion, list PM<sub>2.5</sub> as what is being measured or estimated.

3) The discussion of biases in the TEOM measurements seems inadequate (pp. 6-7). There have been quite a few papers since Allen et al’s 1997 paper comparing the

C2

TEOM with other sensors, including those relying on light scattering measurements. As just one example, Karagulian et al (JEM 14:2145, 2012) found an R2 value of 0.75 between the TEOM and a SidePak. It would seem important to note how well previous comparisons between TEOMs and other light scattering instruments have worked, as context for (and comparison with) the measurements from the Shinyei sensors.

4) How does the level of air pre-heating for the Shinyei sensors compare with that for the TEOM? While heating in the TEOM is mentioned as a possible artifact on p.6 (line 30), there is no mention at this point in the mspt that the Shinyei sensors are also heated (even though this was stated back on p.3 line 14).

5) How did the timing of disagreements between PM monitors (p. 6, line 24) relate to the RH measurements (or the temperature measurements)?

6) For the Hyderabad measurements, clearly the R2 value is driven by the highest concentration data points. If only values  $<40 \mu\text{g}/\text{m}^3$  are used, does the R2 become similar to the Atlanta rooftop?

7) The entire “Estimating emission factors” section is inadequately supported, and should be omitted, due to the following concerns:

a. The R2 for this site (between PPD20PV and the TEOM) was 0.18, which is a very uncertain starting point for relying on the PPD20PV values in calculations.

b. As is mentioned in this revised version of the mspt, the light scattering characteristics of the PM will vary between “background urban” and vehicle emissions. So applying a single regression to the Atlanta roadside, where only a subset of the data is believed to be vehicle-dominated, will lead to even more inaccuracies.

c. The EF estimate is based on a 5-min period of elevated PM and CO2 data, even though 1-min measurements were collected over 3 days. Oddly, looking at the CO2 data (Figure 4), it appears that the 5-min period chosen is in the midst of a  $\sim 6$  hr period where CO2 measurements were  $\sim 480\text{-}550\text{ppm}$ . But the synchronous PPD20V

C3

data (Figure 3) show PM levels fluctuating much more frequently over this same time period.

d. No evidence is provided that the approach to choosing this 5-min period involved objective (=statistical) analyses – the authors only say (p.11, line 17) that they chose a morning rush hour (even though there would have been 3 morning rush hours of data), “where both the pollutant and CO2 rose and fell at the same time”. Why weren’t more time periods tested?

e. The lack of information on wind direction, traffic density, and “background” concentration levels (upwind of the roadway) leads to more uncertainties and unsupported assumptions.

f. After obtaining an EF value, and deducing that this would correspond to about 30% of the fuel combustion involving diesel, on a freeway that should have been “dominated by gasoline-fueled vehicles” (p.11, line 4), the mspt says (p.11, line 22) that this EF value “is likely high”.

g. Then, a 2nd EF is found, based on BC and involving a different time period (8-9pm), with no justification for why this period was expected to have high concentrations of roadway emissions. This EF corresponds to 13% diesel combustion – but there is no discussion of whether this EF value is trustworthy or believable. Is it feasible that there could be this large a portion of diesel vehicles making local deliveries, in the evening?

h. I am extremely skeptical that the confidence bounds shown with the EF values are accurate – they seem much too low, and the method used to quantify uncertainty isn’t clearly explained. By SE, are you giving the standard error of estimate, representing the 85% (that is, 1 sigma) prediction interval around each best fit line? But how are you accounting for the fact that, for the roadside site, only 18% of the variability in the Shinyei measurements can be accounted for by the reference PM measurements?

8) On p.14, it is noted that the optics weren’t maintained at all during the “few week”

C4

deployment in India. How do the authors know that the sensors were still performing acceptably near the end of the deployment period? No temporal measurements are shown or were discussed for the India deployment, nor were tests performed to assess how frequently optics maintenance was needed.

9) The use of an exponential (monotonically increasing) eqn to capture a saturation-type effect seems misleading – it gives the erroneous impression that, despite saturation, an accurate mass concentration can still be inferred. If saturation was indeed a problem, then all measurements greater than a certain raw output level should have been omitted from the fitting protocol, and subsequent analyses.

10) The exact same set of data used to determine the best fit line (or exponential) was then transformed, using this best fit equation. It does not seem scientifically appropriate to apply a transformation equation to the exact same set of data that was used to find the equation.

11) In Figure 5, there are noticeable discrepancies between the “raw output” data for the PPD60PV and the  $\mu\text{g}/\text{m}^3$ . In some cases, e.g., around the 12/11 and 12/13 tick marks, large variations in the raw data appear to have been almost completely smoothed out in the  $\mu\text{g}/\text{m}^3$  plot. In other cases, e.g., around the 12/15 tickmark, a large spike has appeared in the  $\mu\text{g}/\text{m}^3$  plot that is completely absent in the raw output plot. Given the monotonic nature of the fitting equation, these changes don't make sense. Is it possible that the dashed line in the last of the 3 sub-figures is instead a plot of PPD20V 1?

12) There are inconsistencies in the numbers reported. For example, the S.E. for the urban roadside TEOM vs Shinyei PPD20V is reported as 7.1 in Figure 3, and listed as 8 in Table 3, but the text (p.8, 2nd line) says that the Shinyei “sensor was within 4  $\mu\text{g}/\text{m}^3$  of the TEOM”. For the Atlanta rooftop, for the PPD60PV the S.E. is 7.1 in Figure 6, vs 17 in Table 3.

13) The added comparisons, in this revised version, between the Shinyei sensors and

C5

a DustTrak in laboratory experiments seems inadequately examined. The advantage of using one source (incense) in repeated experiments is that it allows assessments of variations/inconsistencies between instruments. It also represents a best-case scenario of sorts for correlation strengths. However, a potentially substantial difference between comparing 1-min averages here, vs 1-hr averages in the field, is that lags in instrument response would much more greatly impact the 1-min comparisons. What is the characteristic response time for the Shinyei sensors?

Minor issues:

1) The abstract gives the impression that the study is much more comprehensive than it is. It should be edited to be more straightforward, e.g.

a. Replace “a number of select PM sensors” (line 11) with “three models of PM sensors”

b. Replace “a variety of ambient conditions and locations, including urban background. . .” (lines 11-12) with “a range of ambient conditions at 3 locations: urban background. . .”

c. Likewise, on p.2 (lines 33-34), “a variety of” should instead say “several”, and “include several” should instead say “include 3 models of”

2) P.7, line 21 – the reference to “the way they were assembled in the junction box” is unclear. What was it about the assembly that might have led to a lack of correlation?

3) The figure captions need to be substantially expanded, with the location included for each plot, so that each one will stand on its own.

4) P.9, line 4 – the Williams citation is not included in the reference list.

5) Using a phrase like “most highly correlated” (p.12, line 3) seems inappropriate for an  $R^2$  value of 0.30. “Least poorly correlated”, perhaps?

6) The abstract also gives an overly positive impression of the study's findings, with

C6

the sentence (line 23) “The results of this work show the potential usefulness of these sensors for . . .”

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2015-331, 2016.