Dear Editor,

we appreciate your feedback, and hopefully the manuscript is now in a clearer form. We are looking forward for your opinion. Below you will find our point-by-point response. The revised manuscript and supplementary material (with 'track changes' on) have also been attached to the electronic submission.

On behalf of all authors,

Joel Kuula
Atmospheric Composition Research
Finnish Meteorological Institute
joel.kuula@fmi.fi

In the comments below the line numbers refer to the "authors comments" version of the manuscript.

I don't think that the response to the reviewers is adequate, in regard to the question of how 10 steps in the program of the GP50 might lead to 30 size bins. To be sure, you have said that the 10 steps are not related to the fact that there are 30 size bins (lines 164 to 169).

This is correct; the number of used bins (30) is a computational detail and it is not related to the generation of particles.

However the source of this confusion remains. At line 141 you state that the aerosols are monodisperse, and the implication of this sentence is that this is controlled by the GP50.

Line 141 implies that the GP50 allows the user to freely choose the produced particle size. To be clear, the fact the aerosol is monodisperse is a feature of the VOAG and not GP50.

Thus in the light of your response, one wonders how the different size bins are chosen and/or created. Next at line 158 there is a sentence that still says that the GP50 program (of 10 steps) produced particles sizes in a logarithmically distributed range (0.45 to 9.78 microns). This is still the main clue to the reader as to how you arrived at 30 size bins.

The decision to use 30 bins was purely based on the clarity of the produced figures and statistically sufficient number of measurement points (minimum of 3) belonging to each bin. A statement regarding this was included in the original manuscript, but reviewer #1 considered it to be irrelevant. The statement has been added again to the manuscript.

Section 2.3 data processing: "The decision to divide the data into 30 bins was based on the clarity of the produced figure and statistically sufficient number of measurement points belonging to each bin."

Further, though this may not have been intentional, it could contradict the earlier assertion that your aerosols are monodisperse, since your use of "logarithmically distributed" is ambiguous - it could refer to a continuous statistical distribution. It would be better to say "logarithmically spaced", which is what I think you mean.

Term "distributed" replaced with "spaced".

Thus I cannot work out where your 30 size bins come from (actually I count 28 bins in figure 5), until much later in the paper at line 204. There I see that the bins are actually determined by the APS measurements. But if you are to get logarithmic spacing something else must predetermine what the mean diameter (CMD) should be. So I'm still puzzled.

The reason, why only 28 data points are present in the figures, has been given in the supplementary material; it is because the first and last bin (0.45-0.50 and 8.80-9.73 µm, respectively) of the 30 bins did not in practice contain any measurement points (the size range of the produced particles was approximately 0.55-8.4 µm). By reducing the number of used bins (i.e. widening the width of the bins) it could have been possible to "force" 3 or more data points to each bin; however, this would have compromised the clarity and representativeness of the figures due to the increasing standard deviations and lower size resolution. The lower and upper end of the size range of the 30 bins (0.45 and 9.73 µm, respectively) were chosen according to the theoretical size range of the produced particles. This ensured that the data from different test runs was treated equally.

The relevant section of the supplemental material has been appended at the end of this document.

Another response that I consider inadequate is in regard to the comment where reviewer 1 raises the possibility of there being detections of multiple particles. It is a off-topic to give expected concentrations in China - what really matters here are the concentrations in your experimental setup, at the detectors. I cannot find any estimate of this number in your revised manuscript. Also the assertion that Mie theory gives the total intensity of the scattered light scaling as the sixth power of diameter is too simple here. The scaling is much more complicated because of the rapidly growing forward scattering lobe (with increasing diameter), so the scaling at a particular angle needs to be considered.

The reviewer comment was interpreted to mean that it is unreasonable to assume that the sensors could function as optical particle counters rather than as nephelometers, and that the main reason for this is the particle coincidence resulting from the unsophisticated particle detection configuration. Our response intended to highlight that the stance taken in the manuscript is not, in fact, unreasonable from the technical point of view, and that the assumption of a nephelometer-type functioning is problematic in several different ways; it not only contradicts the previously presented major comment regarding the use of different flow rates, but also undermines all the previous sensor studies, which have attempted to measure size-specific mass fractions. Nephelometers cannot be used to measure sizes of individual particles, and according to the Mie theory, response of a nephelometer type device should be stronger for larger particles and not weaker. This is not what the results of this study showed. The statement regarding the sixth power of diameter is commonly made in aerosol science (see e.g. *W. C. Hinds: Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*) and, although being an approximation, it is, in our opinion, sufficient to prove the point in this case.

The total number concentration of the reference aerosol is irrelevant considering that the limit value for particle coincidence is not known. The maximum concentration can be estimated from the running parameters of the VOAG (listed in Supplemental Table S1), and, in practice, the concentrations were in the range of 30-90 # cm^{-3} range depending on the particle size and the degree of deposition losses. Compared to other aerosol generators (e.g. atomizers and powder generators), the range of concentrations the VOAG produces is very low.

Added to Figure 2 caption: "(concentration range 30-90 # cm^{-3})."

Additional comments:

Referring to an "optical aerosol spectrometer" could confuse. The spectrum determined is a size spectrum, not an optical spectrum. So saying "optical aerosol size spectrometer" would get around this.

Corrected as suggested.

At line 154 "The GP50 was operated in a method-mode, meaning that an automated program was used to dispense the liquids." is a bit confusing because "method-mode" is meaningless to me. Better to say "The GP50 used an automated program for dispensing the liquids."

Corrected as suggested.

In Figure 3, there is nothing that is white, only grey, in spite of the caption referring to something white.

Rephrased as "light grey".

The explanation of figure 4 needs improvement - in particular, in panel c) there is a legend that presumably refers to size bins as determined by the device under test. Things like this need to be be made more explicit.

It is explicitly stated in the figure caption that the figure legends refer to the detection ranges declared by the corresponding manufacturer.

At line 286 - South Coast of which country?

Added: “USA”

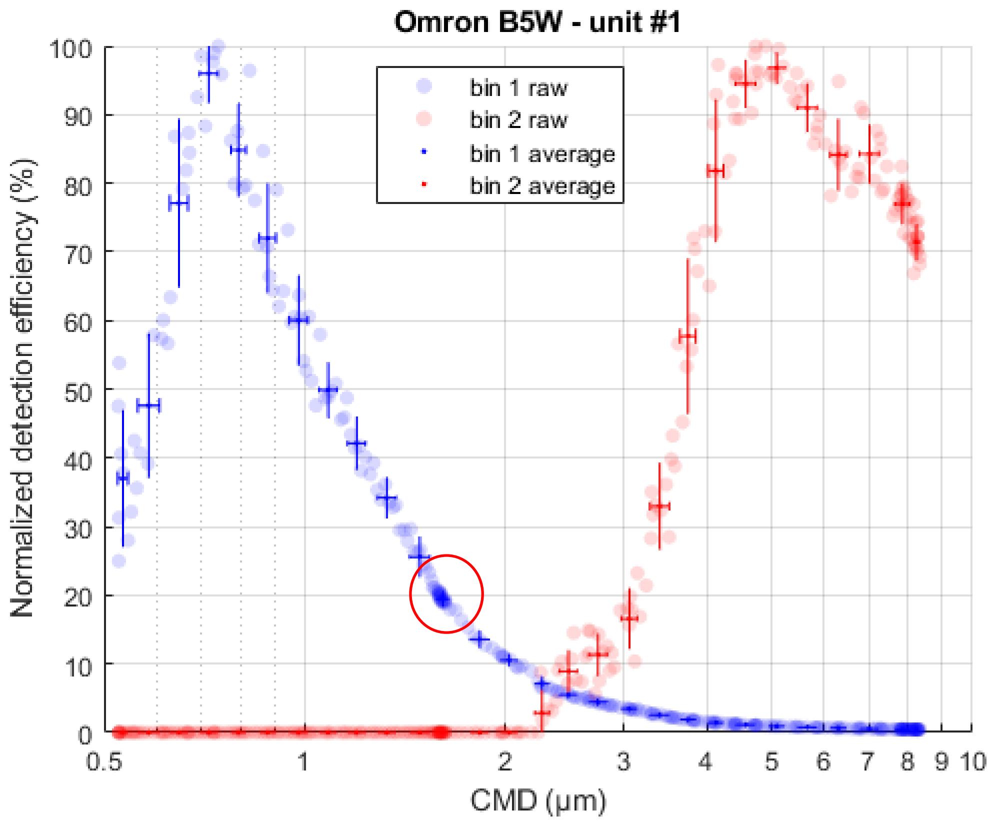
Appendix:

Supplemental material

Detailed description of the data processing method used

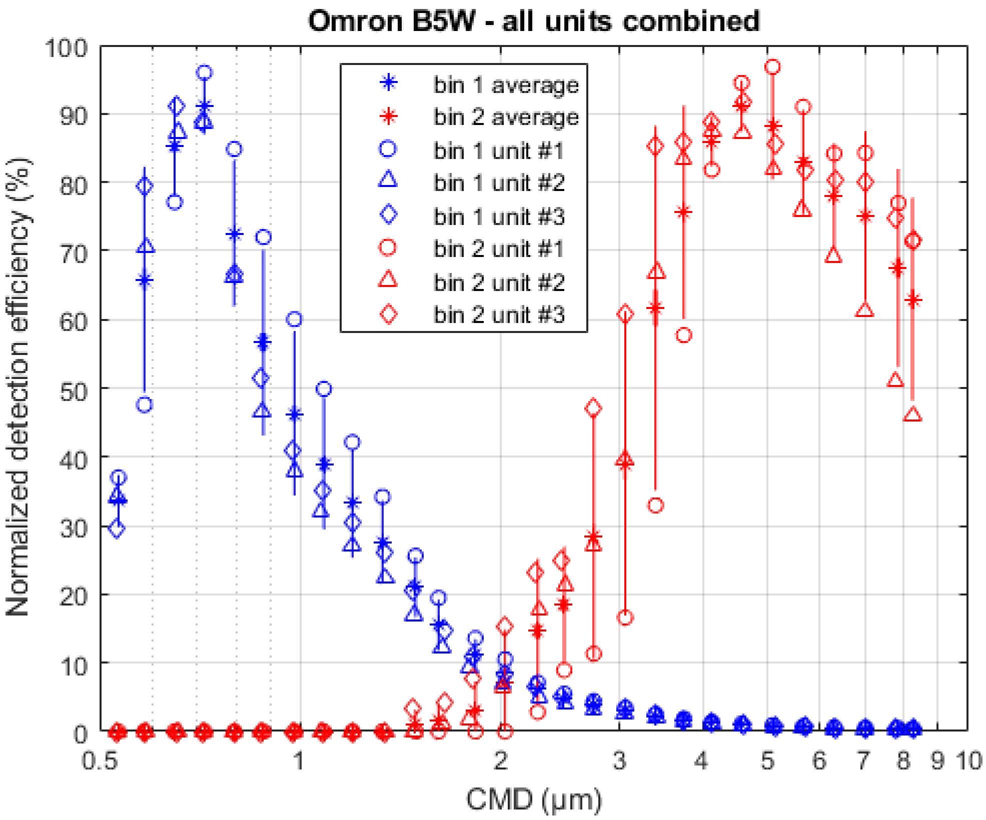
Supplemental Figure S4 shows the normalized and filtered (data points with GSD greater than 1.2 removed) 10-second resolution data of the Omron B5W unit #1 test. Raw data is plotted as transparent bullets and the average values and respective standard deviations (for both CMD and normalized detection efficiency) as solid dots. The raw data was divided into 30 different sections which were logarithmically distributed to 0.45 – 9.73 µm range. This range was the theoretical size range of the produced particles. In the figure, each section corresponds to each solid dot (blue and red), and in this case, a total of 28 dots (for each color) are visible. This is because in practice the first and last section (0.45 – 0.50 and 8.80 – 9.73 µm) did not contain any measurement points.

Despite shown here, the standard deviations of the raw data were not utilized in any form as the final statistical uncertainties were calculated from the average responses of the three individual units. By using the "average of averages", all units had an equal contribution to the final statistics (28 data points each) as in some occasions, the total number of raw data points and the way the points were distributed along different particle sizes varied. See for example the red circle in Fig. S3; for an unknown reason, the speed at which the particle size gradient was evolving decreased momentarily and thus resulted in a cluster of data points. If the raw data would have been used as such, the cluster would have distorted the calculations of average due to the greater number of data points at this specific particle size.

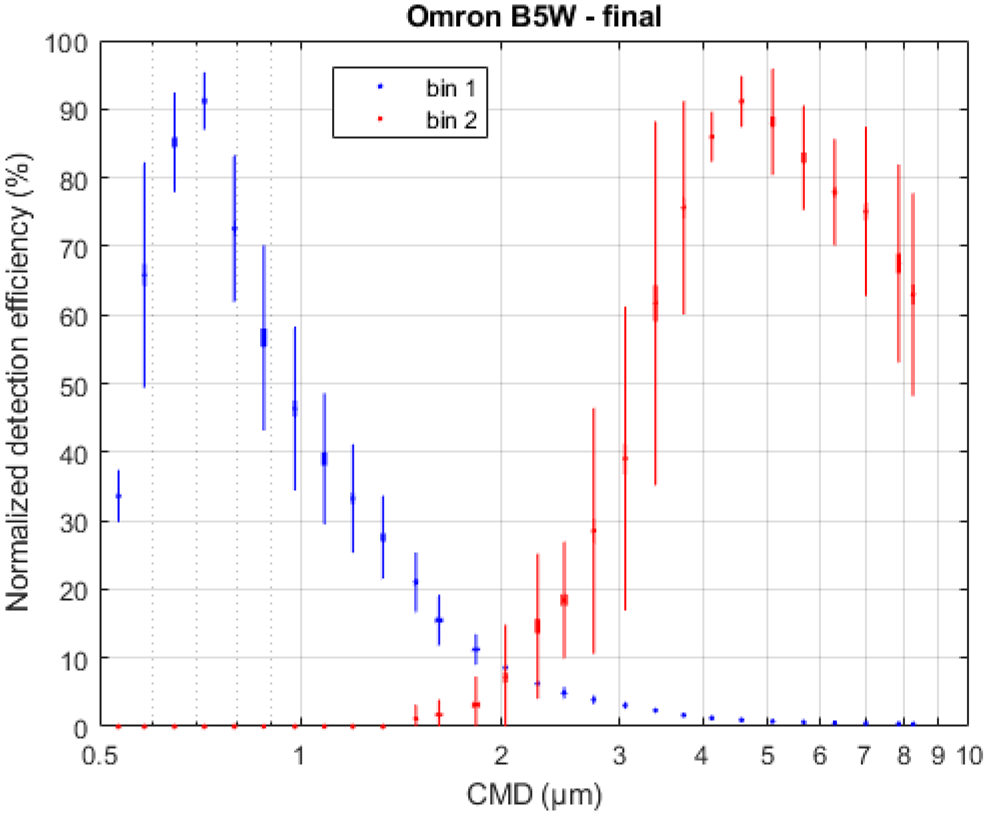


Supplemental Figure S4. Normalized and filtered (GSDs greater than 1.2 removed) data of the Omron B5W unit #1 test run. The raw 10-sec resolution data is shown as transparent bullets and the calculated average values of the 30 different size sections as solid dots (with standard deviations).

The average responses of the three Omron B5W units are shown in Supplemental Figure S5. The circle, triangle, and diamond markers stand for the average responses of the individual units #1, #2, and #3, respectively, and "the average of the averages" (and respective standard deviations) are shown in the figure as star markers. The standard deviations of the average CMDs are negligible compared to the differences observed in normalized detection efficiencies and thus they were not shown in the final manuscript Figure 4f. Supplemental Figure S6, which is essentially the same figure as the final manuscript Figure 4f but with standard deviations of the CMDs included, shows again the insignificance of the CMD standard deviations.



Supplemental Figure S5. Averaged responses of the three individual sensor units.



Supplemental Figure S6. Final normalized detection efficiency of the Omron B5W (with standard deviations).