Response to Anonymous Referee #2

We would like to thank the reviewer for the thoughtful and constructive examination of our paper! Please find below our responses to each comment individually and please note that:

- **Blue bold font** represents comments of the Referee,
- Black regular font represents the response to each referee's comment,
- ~~strikethrough font~~ represents removed text from the manuscript according to referee's comment,
- Red font represents added text in the manuscript according to referee's comments.

General comments: This paper focuses on the performance of 3 OPS devices in a highly polluted area. I think this paper will be helpful to the sensor/air monitoring community as it is at higher ambient concentrations than much previous work with a suite of collocated reference measurements. The authors present a highly valuable dataset. Overall, this is a nice paper with scientific significance, good presentation quality and a few changes to statistical methods/discussion and discussion of previous work will strengthen the scientific quality.

Response: Thank you for the general comments! We will try to implement your suggestions.

Line 114: Can you provide any justification as to why you used the AE channel? The two channels have a nonlinear relationship (Tryner 2020 https://doi.org/10.1016/j.atmosenv.2019.117067) and I wonder if this is some of the reason you have underestimation at high concentration (Figure 3).

Response: The manufacturer of PMS5003 sensor, Plantower, has not explained publicly what is the difference between AE and SM modes, but in private communication they explicitly confirmed that AE is the correct channel for ambient air measurements, while the SM mode is recommended for industrial production workplaces (metal particles or other higher density particles). We have performed laboratory test of these two modes on PMS5003 using the burning chamber and incense scents as the source of PM, and here are the results:

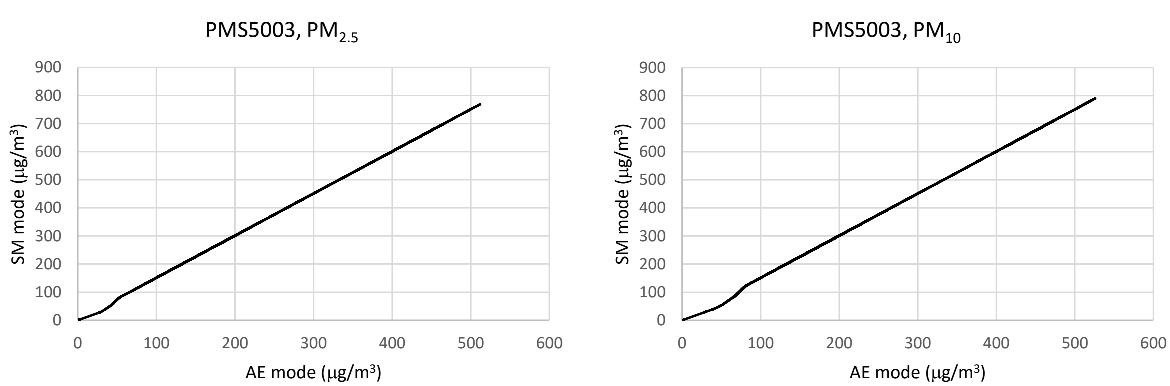


Figure 2. AE and SM modes of PMS5003 sensor.

The relationship between SM and AE modes can be deduced from our test results:

$$\mathrm{SM_{PM2.5}} = \begin{cases} \mathrm{AE_{PM2.5}} \text{ for } \mathrm{AE_{PM2.5}} \leq 30 \\ \text{nonlinear for } 30 < \mathrm{AE_{PM2.5}} \leq 50 \\ 1.5 \times \mathrm{AE_{PM2.5}} \text{ for } \mathrm{AE_{PM2.5}} > 50 \end{cases}$$

$$\mathrm{SM_{PM10}} = \begin{cases} \mathrm{AE_{PM10}} \text{ for } \mathrm{AE_{PM10}} \leq 43 \\ \text{nonlinear for } 43 < \mathrm{AE_{PM10}} \leq 77 \\ 1.5 \times \mathrm{AE_{PM10}} \text{ for } \mathrm{AE_{PM10}} > 77 \end{cases}$$

Investigation of Tryner et al (2020) https://doi.org/10.1016/j.atmosenv.2019.117067 has limited range of PM concentrations, where we can't see this third segment of direct proportionality between SM and AE mode for higher PM concentrations. In conclusion, for low concentrations these two modes are the same, in the narrow, intermediate range the relationship between SM and AE mode is nonlinear, while for higher concentrations SM mode simply gives 50% higher value than AE mode, for both PM2.5 and PM10. If applied to our measurements, SM mode would overestimate actual PM concentrations more than AE mode. These results are added in section 2.

The authors have a heavy reliance on R2 throughout this paper even though it is well known that this is not the best comparison between measurement methods (Bland & Altman "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT" Lancet, 1986; i: 307-310). They do also discuss bias (% difference) but I think it would also be helpful to not rely so heavily on discussion of R2 and add another metric of scatter MAE (or RMSE or another metric the authors prefer).

Response: we agree completely. For all graphs on figures 3, 4 and 8 in manuscript we have added MAE value and 1:1 line. These MAE values are discussed in sections 3.1, 3.2 and 3.6. Furthermore, Figure 8 is modified for some later discussion on humidity influence. Here are the new graphs:

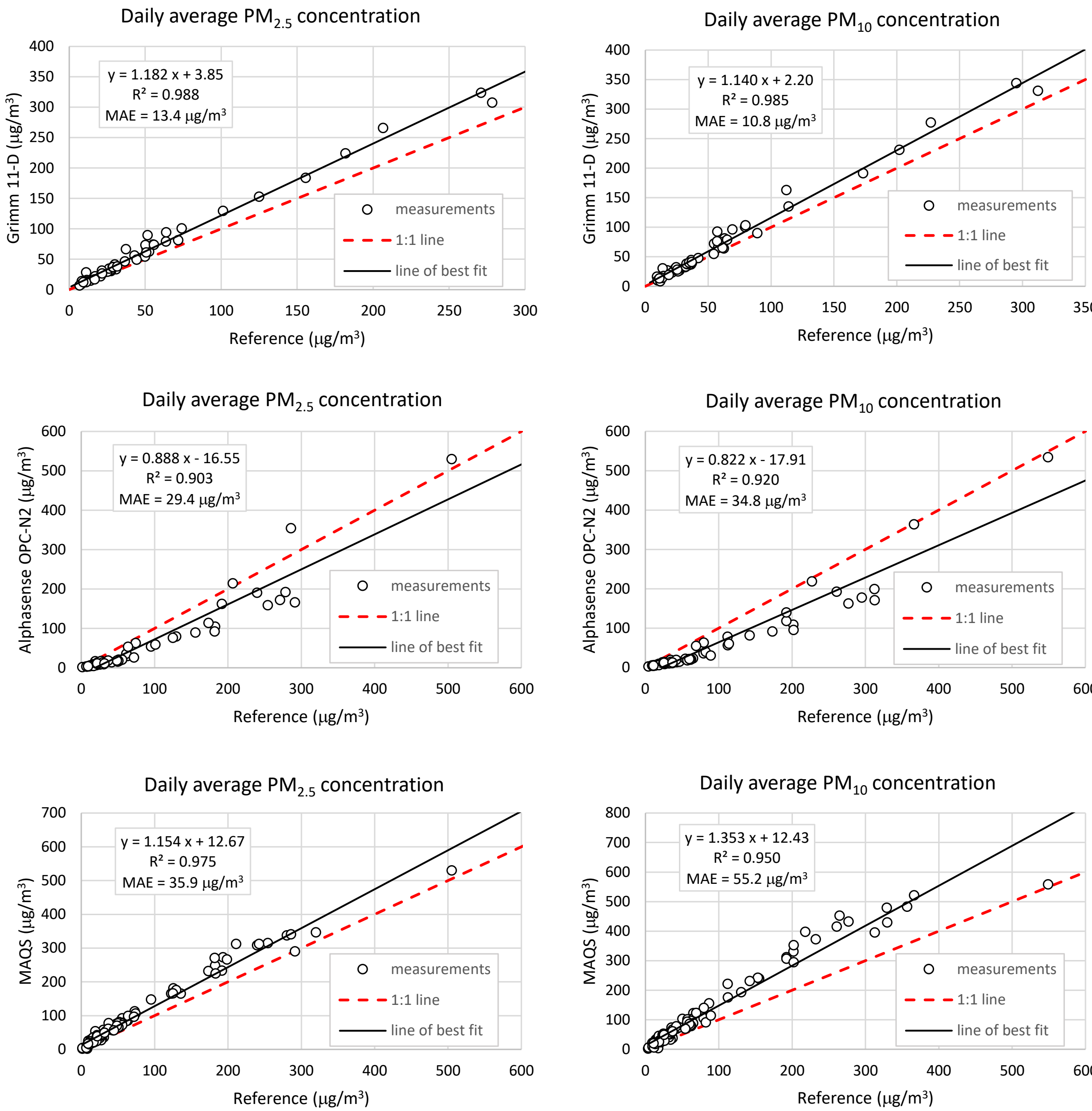


Figure 5. OPS performance during the period of strong pollution (12/2/2019–3/12/2020).

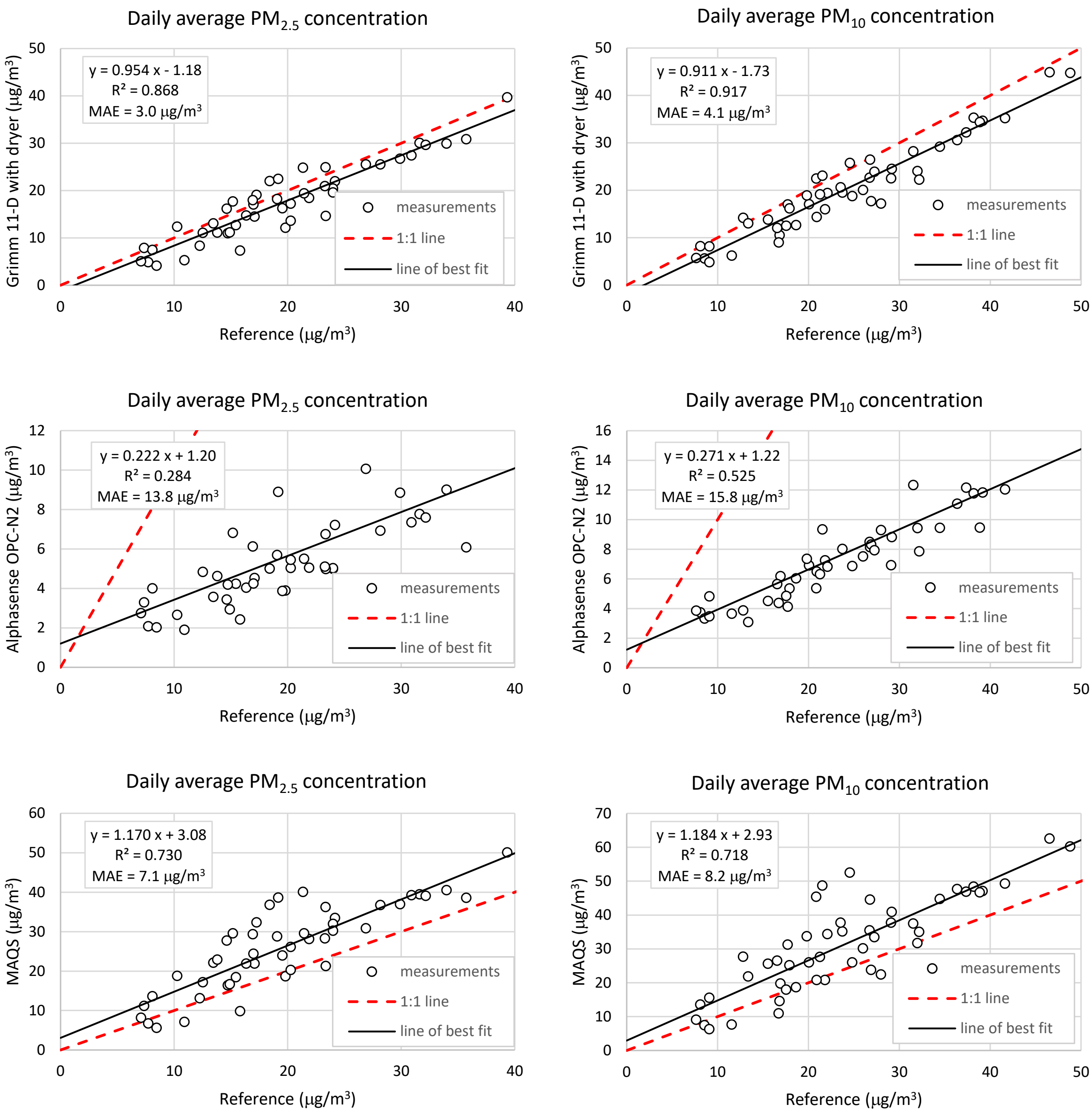


Figure 6. OPS performance during the period of mild pollution (3/13/2020–5/4/2020).

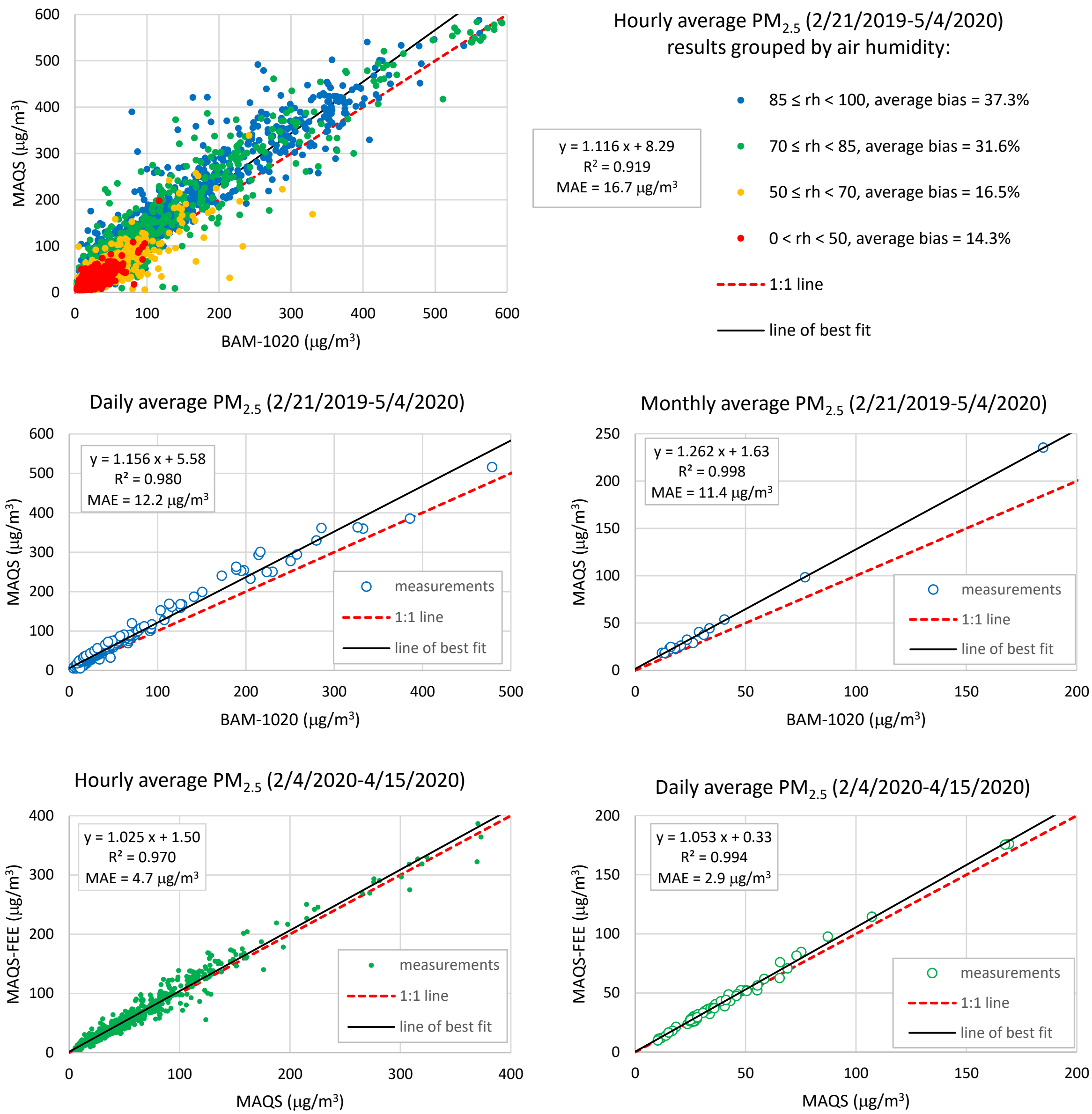


Figure 8. Long-term comparisons of MAQS sensor with BAM operated by US EPA at the nearby location: hourly, daily, monthly average values and comparison of hourly and daily average values of two MAQS sensors: first one (MAQS) at our main facility and second one (MAQS-FEE) at Faculty of Electrical Engineering in the immediate vicinity of BAM-1020.

Line 159-161: I don't think this paragraph provides enough details to understand how you calculated this. I'm guessing this is 3 standard deviations but of what? Just zero concentration experienced in the field? Please elaborate as I think these results will be particularly of interest to the field. It seems like the Bulot paper reports LOD from a bunch of previous work with LCS not just PMS5003/N-2 it might be helpful to strengthen the discussion here. More recent work has also explored the LLOD of PMS5003 sensors (e.g. Magi 2019 https://doi.org/10.1080/02786826.2019.1619915). Also did you want to provide any details on what you do with data below the LLOD (throw out, replace, etc)?

Response: we agree and here are the changes:

Standard deviation (σ) was calculated for periods with near-zero ambient PM concentration and average value of 3σ is estimated LLoD. For PMS5003 our final estimation is 5 μg/m^3. The same value is an estimation of (Magi, 2020), calculated by averaging segmented regressions and (Bulot, 2019) by combining results from several previous studies. This method applied on OPC-N2 yields LLoD of 2 μg/m^3 and 1 μg/m^3 for 11-D. For reference gravimetric system LLoD was calculated using the blank filters, which were treated exactly the same way as real samples (except the sampling of particulate matter), and the calculated value of LLoD is 0.7 μg/m^3. All measurements below LLoD were discarded during the quality assurance phase.

Line 177: Also see recent paper on PMS5003 and large particles that may be helpful (Kosmopoulos 2020 https://doi.org/10.1016/j.scitotenv.2020.141396)

Response: this is a very relevant reference, and we have added it in the revised manuscript. Investigation in Patras, Greece with narrower range of ambient PM concentrations and different aerosol composition with more frequent episodes of Sahara dust. PurpleAir (PMS5003) showed high bias (similar to our results) and bad performance during the Sahara Desert dust episode. We have also registered one intensive desert dust episode on 3/27/2020 from the Aralkum Desert. Here is the new text:

However, different conditions were observed on 3/27/2000 when the dust from Aralkum Desert covered part of Europe, including our test location. During this episode, OPC-N2 performed much better than PMS5003, which wasn't able to determine large fraction of coarse particles correctly (Figure 11). Similar observation about PMS5003 was reported by (Kosmopoulos, 2020), when Sahara dust covered Greece.

The discussion of previous work appears fairly limited. It would be helpful to discuss how the high bias of the PMS5003 and low bias of the OPC-N2 and overall performance compare to studies in other locations as both these devices have been studied fairly extensively.

Response: the following text is added in section 1:

In (Mukherjee, 2017) OPC-N2, PMS7003 and 11-R were compared against BAM-1020 during 12 weeks in the Cuyama Valley, California, USA. Grimm 11-R performed well, while both OPC-N2 and PMS7003 (which is a miniaturized version of PMS5003) produced mediocre performance with heavy low bias. PurpleAir (PMS5003) was tested in (Tryner, 2020) using laboratory and field tests. High bias of PMS5003 was observed. In (Magi, 2020) PurpleAir (PMS5003) was analyzed for 16 months in Charlotte, North Carolina, USA against BAM-1022, high bias of PMS5003 that increases with humidity was reported. High mean bias of PurpleAir (PMS5003) was reported in (Kosmopoulos, 2020) as well.

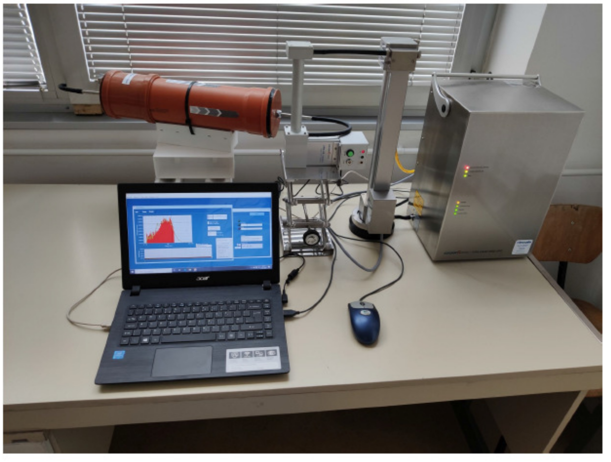
Section 3.3: You only discuss the Humidity influence on the Grimm it would be helpful to discuss the influences on the PMS5003 and OPC-N2 as well.

Response: we have constructed dryers only for 11-D and SMPS (new Figure 1). The design and construction of dryers for small, low-cost sensors, such as OPC-N2 and PMS5003 is in our plans for future work. From the dataset of this study, we can include ambient air humidity as the parameter in Figure 8.

Figure 8 shows the long-term (13.5 months) comparison of MAQS and BAM-1020 with time resolution of 1 hour, together with measured values of ambient air humidity. By averaging all this data we can estimate the influence of humidity on the MAQS sensor: if we sort the measurements by humidity, subset of points where humidity is below 50% has average bias of 14.3%, for humidity range 50%-70%, bias is 16.5%; for humidity range 70%-85% bias is 31.6% and for humidity rang 85%-100% bias is 37.3%. If we subtract bias of least humidity subset from bias of highest humidity subset, we can estimate that humidity influence adds up to 23% on PM2.5 readings from MAQS sensor, which is similar result to the analysis of humidity influence on our 11-D with dryer installed. While this influence can not be neglected, it is still relatively modest. Reason for this is the composition of particles, where we have mostly fine particles below 300 nm, for which hygroscopic growth is less effective (Kosmopoulos 2020).

(a) Air sampler and Stevenson screen.

(b) Devices under test.



(c) SMPS with dryer.

Figure 1. Experimental setup: a) co-located air sampler and Stevenson screen, b) devices under test inside of Stevenson screen: 11-D with dryer, OPC-N2 with SPI adapter and (white-orange) enclosure, MAQS (white enclosure with grey front panel), and c) indoors SMPS with dryer.

It seems like you also have the opportunity to discuss the influence of particle size distribution on the performance of the OPC-N2 and PMS5003 but you have limited your discussion to the Grimm. You mention this briefly in lines 245-249 but it seems like instead of just commenting that small particles could be an issue you can look to see of the OPC-N2 is specifically underestimating more because more of the particles are too small. In addition, both the OPC-N2 and PMS5003 have binned data that could be discussed.

Response: We have data bins for OPC-N2 and PMS5003. However, there are some doubts if the comparison of data bins from PMS5003 and OPC-N2 to 11-D are appropriate, due to the following reasons:

- these devices are constructed differently, 11-D has hydrodynamic focusing (sheath flow), regulated flow rates and high-performance optics, while the low-cost sensors couldn't count individual particles in the same way

- PMS5003 is not a particle counter, it is a nephelometer (Tanzer, 2019) doi.org/10.3390/ijerph16142523

– OPC-N2 has detection limit of 380 nm, PMS5003 from 300 nm, while 11-D can count particles from 250 nm

– when calculating PM mass concentration from individual data bins manufacturers use different weighting factors. Only Alphasense provides values of weighting factors for data bins, assumed particle density for each bin and index of refraction (especially important for correct determination of the particle's diameter). These values for 11-D and PMS5003 are not known.

Taking into account these notes, we present new subsection in the revised manuscript:

3.5 OPS histograms and Aralkum Desert dust

All tested OPS have data bins, with different number of channels, as described in section 2. Figure 11 shows histograms that compare data bins from 11-D, OPC-N2 and MAQS on 1/18/2020 (strong pollution) and 4/16/2020 (mild pollution). It should be noted that we compare here data bins from devices with different specifications and category. As expected, 11-D has ability to count particles below 300 nm, which appear in greatest numbers. Counting efficiency of OPC-N2 is investigaetd in laboratory conditions using PSL particles in (Sousan, 2016a), and the results were good for particles larger than 0.8 μm while for particles with diameter of 0.5 μm OPC-N2 the device showed lower detection efficiency (detection limit of OPC-N2 is 0.38 μm). In our realistic scenario, dominant contribution to the mass comes from particles much smaller than 0.8 μm (Figures 9 and 10) which is not favorable to OPC-N2.

Contrary to OPC-N2, PMS5003 has problems with coarse particle, as indicated in laboratory test (Kuula, 2020). If the fraction of coarse particles is small and steady, PMS5003 performs much better. Ambient conditions in Bosnia-Herzegovina are such most of the time, since the primary source of PM is combustion of coal and biomass. That could explain why PMS5003 performs better than OPC-N2 most of the time. However, different condition were observed on 3/27/2000 when the dust from Aralkum Desert covered part of Europe, including our test location. During this episode, OPC-N2 performed much better than PMS5003, which wasn't able to determine large fraction of coarse particles correctly (Figure 11). Similar observation about PMS5003 was reported by (Kosmopoulos, 2020), when Sahara dust covered Greece.

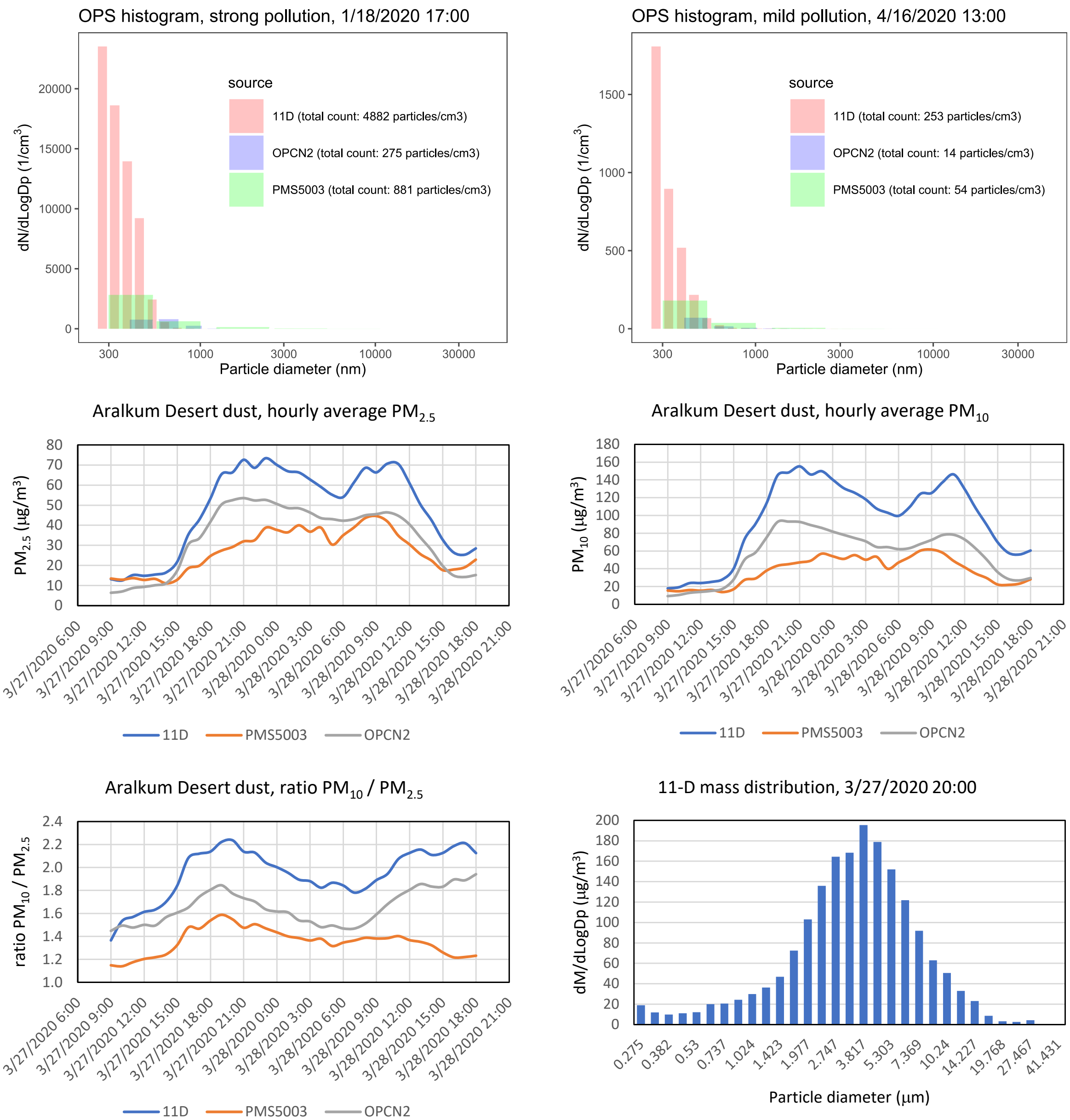


Figure 11. OPS histograms and Aralkum Desert dust episode, hourly average values.

Technical corrections: Line 30: grammatical error "equipped with BAMs" and you should probably spell out what BAM stands for the first time you use it. Line 123,221: missing m on Grimm Line 181: It may be helpful to mention the figure earlier on in the paragraph before discussing the results so that readers can look at the figure and follow along.

Response: we accept all these suggestions and appropriate corrections have been made for the revised text.