Response to Anonymous Referee #3

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.

The manuscript describes the evaluation of some optical particulate matter sensors in high and low pollution episodes. The topic of very limited novelty as some of the sensors have already been extensively tested in the peer reviewed literature. 2 of the sensors are already not commercially available anymore and follow-up models are being sold. Novelty and generalizable findings would need to be emphasized because right now, there appears little true scientific discussion on fundamentals that would easily be transposable to justify publication of the manuscript on essentially outdated sensors.

Response: We couldn't agree with this general comment. These sensors are not outdated and they are commercially available: as of today, both Grimm 11-D and Plantower PMS5003 are fully available, while Alphasense OPC-N2 is available in limited quantities for existing customers.

The authors should address the following issues

- What is the true novelty here and insights that were not already documented in the existing papers on the Grimm, the alphasense or the PMS?

Response: the true novelty here is the range of ambient air concentrations, which is much wider than any previous work that we could find in existing articles. This unique dataset is relevant for many countries in Eastern Europe, which share the same problem: air pollution with PM as a dominant component. Performance of optical PM sensors greatly depends on composition of PM. In Eastern Europe many houseolds still use coal as the source of energy for heating. Furthermore, network of govermnetal air quality measuremt stations is relativelae sparse. Considering these facts, it is very importnat to systematicaly analyze performance of OPS in this region. Furthermore, wide-range spectrometer, which consists of SMPS and OPC 11-D is for the first time successfully applied in realistic scenario under such ambient air conditions.

- You use in the comparison figures linear regressions with non-zero intercepts, some of these intercepts are substantial! >10 ug/m3 for PM10 (figure 3) both positive and negative. This needs to be explained.

Response: yes, indeed, the intercept values for the lines of best fit in Fig. 3. in the manuscript vary from -17.9 μg/m^3 to 12.7 μg/m^3. But if we take into account that the range of measurements is from 0 to 600 μg/m^3, these intercepts are relatively small. Please have a look at revised pictures where we have added 1:1 line, according to your suggestion. From these graphs we can see that these intercepts are visually close to the ideal situation (1:1 line).

Overall for all the comparison figures, why not indicate a 1:1 line and please do a deeper analysis. It looks like these figures mostly show non linearity with at low concentrations most data points above the lien and at high below or vice versa. There seems to be clear non linearity without any discussion, instead these weird linear regressions with intercepts that are not explained. Even weirder that the authors acknowledge in the text that there is non linearity likely.

Response: thank you very much for this suggestion! We have new graphs below with included 1:1 line. Regarding the linearity, actually it is surprisingly good across the wide range of measurements for all 3 tested devices. Figures 3 and 8 show some non-linear tendency of tested sensors only above 300 μg/m^3 of $PM_{2.5}$.

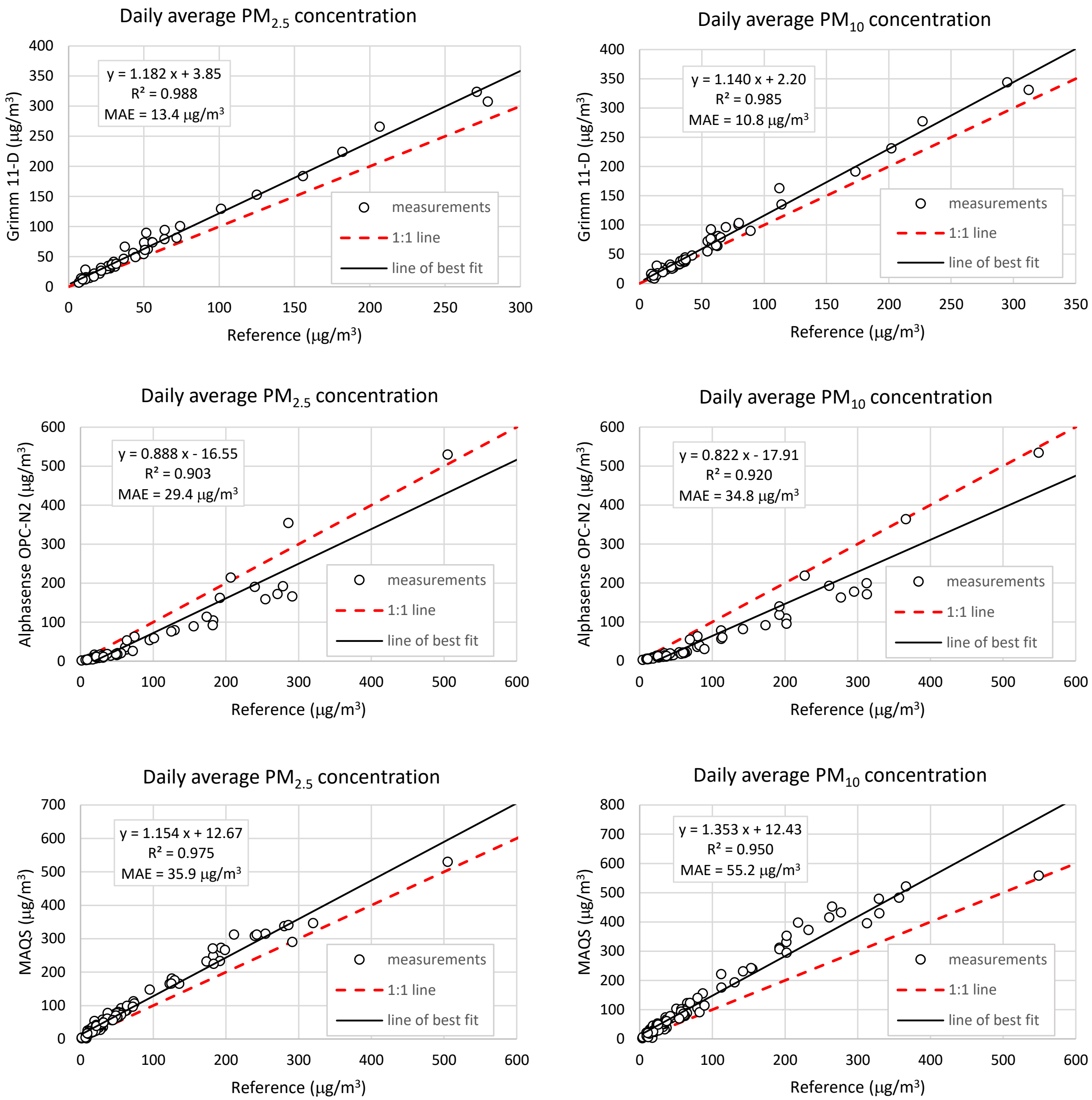


Figure 5. OPS performance during the period of strong pollution (12/2/2019–3/12/2020).

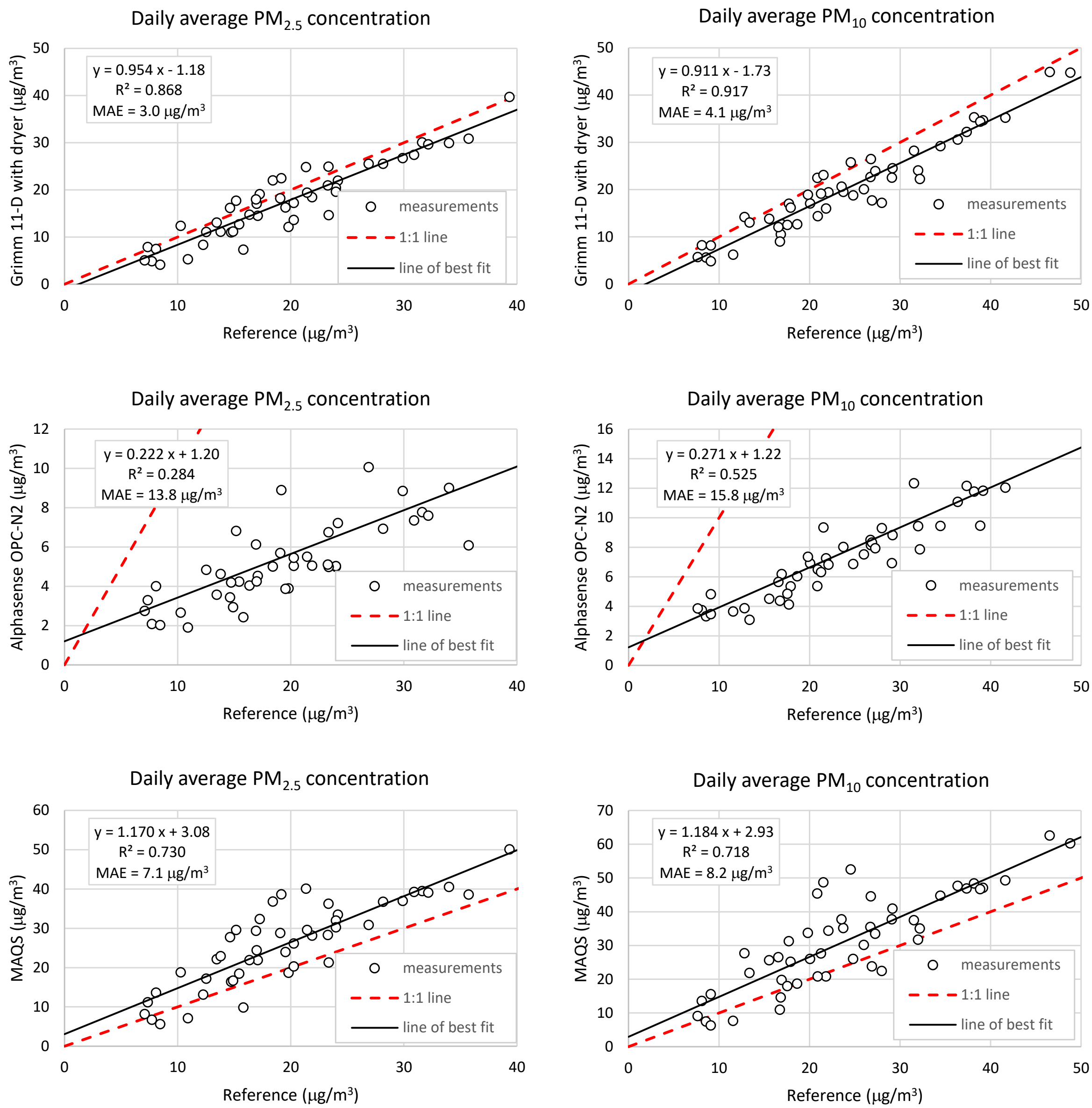


Figure 6. OPS performance during the period of mild pollution (3/13/2020–5/4/2020).

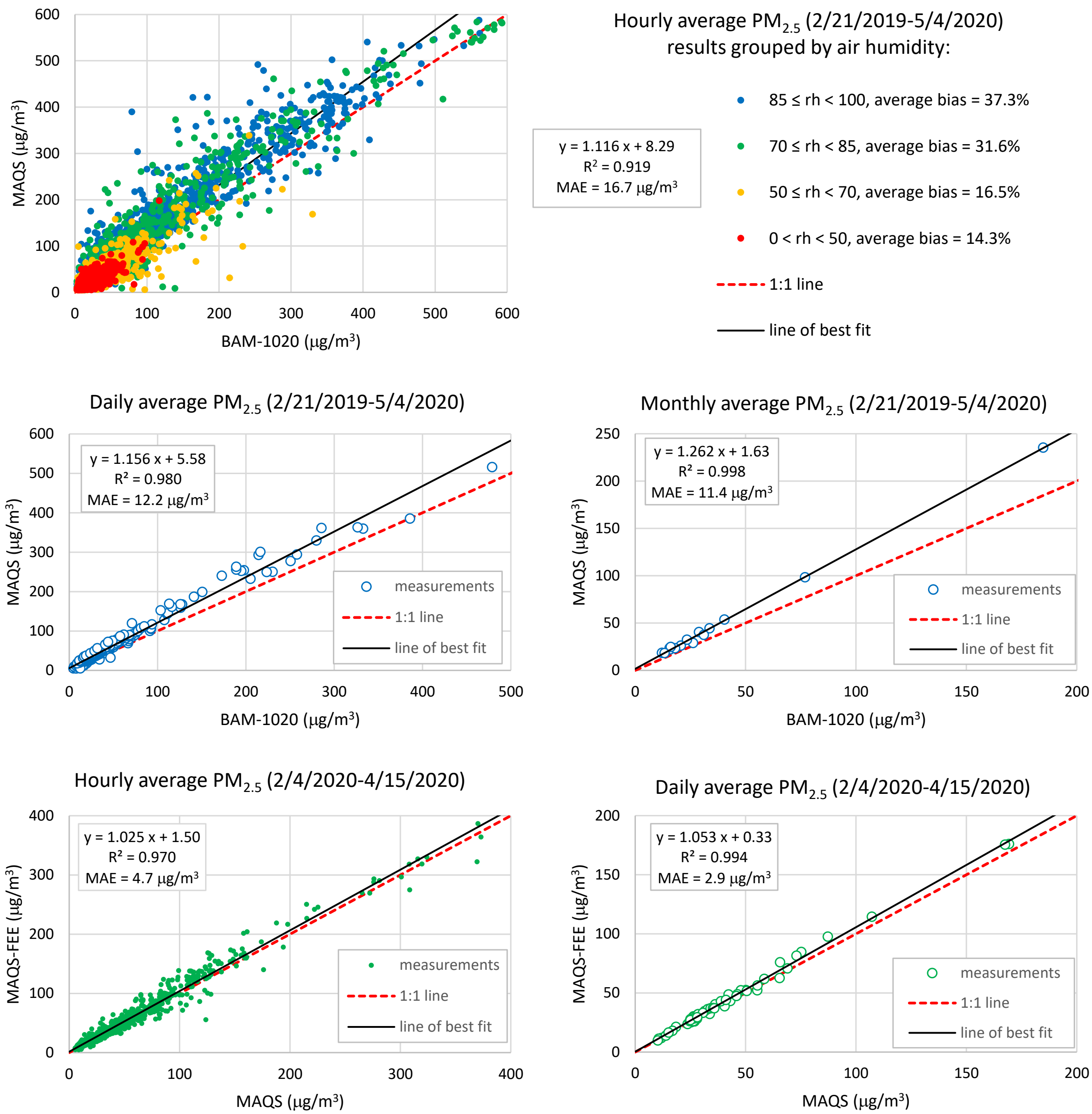
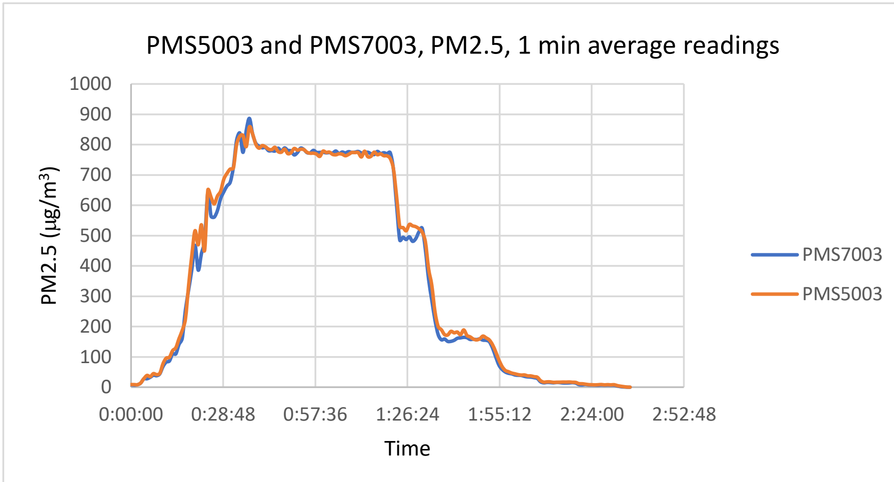


Figure 8. Long-term comparisons of MAQS sensor with BAM operated by US EPA at the nearby location: hourly, daily, monthly average values and comparison of hourly and daily average values of two MAQS sensors: first one (MAQS) at our main facility and second one (MAQS-FEE) at Faculty of Electrical Engineering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

- The Alphasense is now on version OPC-N3 and it is hard to find information on earlier version idem on the PMS5003, they are now at PMS7003. Could you comment if you expect the observations here to be transposable otherwise they are useless.

Response: OPC-N3 is a successor to OPC-N2. By analyzing the specifications of both, we expect that these results are applicable to OPC-N3 as well. Major differences between OPC-N3 and OPC-N2 are internal temperature and humidity sensor in OPC-N3 and slightly lower detection limit, 350 nm (N3) instedad of 380 nm (N2). While these are useful features, they certainly don't make OPC-N2 obsolete.

Plantower PMS7003 is a miniaturized version of PMS5003. The only advantage of PMS7003 is its smaller size. Because of that PMS7003 is preferred if the size is critical, for example the AirBeam project (Mukherjee, 2017, Sensors). But for our research PMS5003 is more appropriate, because it is more rugged and it has larger intake fan. PurpleAir (Tryner, 2020, AE) is good example of integration of PMS5003 into the network of sensors. We have many laboratory test results of PMS7003 and PMS5003 and conclusion is clear: these sensors give the same results. Please have a look at the results of one of our tests, where we compare these two sensors. Burning chamber with incense scents as the source of PM was used.



Comparison of PMS7003 and PMS5003

- The description of the PMS device seems very speculative? This is very weird when a simple google gives clear descriptions of the device (https://www.aqmd.gov/docs/default-source/aq-spec/resources-page/plantowerpms5003-manual_v2-3.pdf)

Response: we have that file, which contains specifications of PMS5003. However, some important information about the sensor is not published there. For example, there are two modes (SM and AE) and the Manufacturer did not publish what is the difference. But we agree that our description of PMS5003 is not good, and we have changed it.

~~It is likely that PMS5003 works in photometric mode, rather than counting individual particles. By disassembling one PMS5003 sensor we have discovered some components: red semiconductor laser which pulsates approximately once per second, photodetector and 32-bit processor.~~ It uses red semiconductor laser, photodetector at 90^0 scattering angle (Kuula, 2020) and 32-bit processor (Cypress CY8C4245, 48 MHz). According to (Tanzer, 2019) PMS5003 is a nephelometer, not the particle counter.

- The discussion of RH impacts is very cursory and given how big the issue is, it would be important to see how results of the sensors agree or disagree as a function of RH. Here it would be critical to discuss that the gravimetric measurements are done at a given RH but how does this RH compare to the sensor measurements.

Response: according to requirements of the standard EN 12341:2014, all filters were conditioned in our gravimetric laboratory at relative humidity between 45% and 50%, and temperature between 19 and 21 ^{0}C. Without dryers on all devices, we can not control these parameters outside. However, this campaign provided us with enough measurements to draw some important conclusions about humidity influence on measurements from OPS. This text is added in section 2:

According to requirements of the standard, all filters were conditioned at relative humidity between 45% and 50%, and temperature between 19 and 21 ^{0}C.

This text is added in section 3.3:

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 pronounced (Kosmopoulos 2020).

Can you comment that you are running the sensors close to their technical specs (95% RH) also at times you actually do run the Grimm D11 outside of specs as the Grimm specs say temperatures above 4 degC (although you also seem to heat the inlet) this is not very clear.

Response: primary aim of this research is the evaluation of instruments in realistic scenario. That includes wide range of all operating parameters. During the campaign we followed all recommendations of manufacturers of devices, especially Grimm, to ensure that instruments are running normally. Rigorous quality assurance procedure was used. All measurements below LLOD, outside of the specifications and with error or warning codes in the logs were discarded.

The introduction needs serious revision. Particulate matter and aerosol is not the same thing. Please eliminate all discussion of aerosol as aerosol is the particles and the gases

Response: we accept this. Instead of “aerosol”, “particulate matter” or other appropriate term will be used consistently through the entire manuscript.

The introduction is very narrowly focused and does not discuss things like the use of TEOM in networks. Also some of the statements should clearly be supported by references

Response: we agree that network of TEOMs is an interesting topic, but it is not related to this work. TEOM is not an optical scattering device.

You are very non quantitative and non rigorous in the text and very imprecise., This needs substantial improvement. E.g. L81: what is meant by extremely high? L85 you know exactly what your lower size limit is, so please state it, L 166 what is mean by a “good” correlation?

Response: L81 “extremely high” refers to measured value of PM2.5 concentration up to 504.9 $\mu g/m^3$. That is extremely high concentration of $PM_{2.5}$. For example, in US AQI categorization, values of $PM_{2.5}$ over 500 $\mu g/m^3$ are beyond air quality index scale. We accept objection about lower size limit in L88, so instead

~~It can detect particles with diameters from few nm up to 1 μm.~~

we have

It can detect particles with diameters from 10 nm up to 1 μm.

R^2 coefficients from 0.90 to 0.99 represent very good correlation in this context (comparison of optical devices to the reference gravimetric method).

The statistical discussion totally lack rigor. L 262 “the correlations for hourly, daily and monthly average values of PM2.5 are 0.919, 0.980 and 0.998, respectively” what does this mean? Followed by “with absolute values overestimated by 20% on average ” how was this obtained? Where is the data? This is not obvious from Fig 8 at all?

Response: Fig 8 is redrawn completely, and now you can see these average bias values on the figure as a function of air humidity. The underlying datasets for this publication are available at

https://doi.org/10.5281/zenodo.3897379

Here are the text changes:

~~During 13.5 months of continuous comparisons, the correlations for hourly, daily and monthly average values of PM2.5 are 0.919, 0.980 and 0.998, respectively, with absolute values overestimated by 20% on average.~~

Based on 13.5 months of continuous comparison of MAQS and BAM-1020, hourly average values give R^2 coefficient 0.919 and MAE 16.7 $\mu g/m^3$. Daily average values produce R^2 coefficient 0.980 and MAE 12.2 $\mu g/m^3$, while the monthly average values give $R^2 = 0.998$ and MAE $= 11.4 \mu g/m^3$ (Figure 8).

- the abstract should not read like an experiential section with study dates etc. These details should not go there instead it should contain quantitative results form the paper.

Response: we accept that, here is the new abstract:

In this paper we evaluate characteristics of three optical particulate matter sensors/sizers (OPS): high-end spectrometer 11-D (Grimm, Germany), low-cost sensor OPC-N2 (Alphasense, United Kingdom) and in-house developed MAQS (Mobile Air Quality System) which is based on another low-cost sensor – PMS5003 (Plantower, China), under realistic conditions of strong and mild urban pollution. Results were compared against a reference gravimetric system, based on Gemini (Dadolab, Italy), 2.3 m^3/h air sampler, with two channels (simultaneously measuring $PM_{2.5}$ and PM_{10} concentrations). The measurements were performed in Sarajevo, the capital of Bosnia-Herzegovina, from December 2019 until May 2020. ~~This interval is divided into period 1 - strong pollution (December 2019 – March 2020) and period 2 - mild pollution (March 2020 – May 2020).~~ This interval is divided into period 1 - strong pollution and period 2 - mild pollution. The city of Sarajevo is one of the most polluted cities in Europe in terms of aerosols: the average concentration of $PM_{2.5}$ during the period 1 was 83 μg/m^3, with daily average values exceeding 500 μg/m^3. During period 2, the average concentration of $PM_{2.5}$ was 20 μg/m^3. These conditions represent a good opportunity to test optical devices against reference instrument in a wide range of ambient particulate matter (PM) concentrations. The effect of an in-house developed diffusion dryer for 11-D is discussed as well. In order to analyze the mass distribution of ~~aerosols~~ particles, a scanning mobility particle sizer (SMPS), which together with the 11-D spectrometer gives the full spectrum from nanoparticles of diameter 10 nm to coarse particles of diameter 35 μm, was used. All tested devices showed excellent correlation with the reference instrument in period 1, with R^2 values between 0.90 and 0.99 for daily average PM concentrations. However, in period 2, where the range of concentrations was much narrower, R^2 values decreased significantly, to values from 0.28 to 0.92. We have also included results of a 13.5 month long-term comparison of our MAQS sensor with a nearby beta attenuation monitor (BAM) 1020 (Met One Instruments, USA) operated by the United States Environmental Protection Agency (US EPA), which showed similar correlation and no observable change of performance over time.