Anonymous Referee #1

We would like to thank Reviewer #1 for her/his time devoted and the constructive and helpful comments to which we will respond point by point.

This is a very well, and concisely written evaluation of the ability of the CAMS global and regional systems for forecasting a rather particular aerosol event, being a combination of high loadings of dust, biomass burning and sea salt particles. I find interesting to learn that the IFS forecasts excluding data-assimilation already show similar performance as the configuration including data-assimilation. Does this either point at a very good forecast model strength, or rather at deficiencies in the data assimilation setup?

We thank the Reviewer for the general comment. In fact, the use of data assimilation improves the performance of CAMS system in AOD forecast, yet it seems to leave relatively unaffected the PM_{10} and $PM_{2.5}$ forecast near the surface. Both speculations seem valid. On the one hand, IFS is a state-of-the-art model that apparently reproduces the aerosol transport and the induced increase in particulate matter (PM) surface concentrations, without the use of data assimilation. On the other hand, it seems that the data assimilation of a columnar aerosol product (AOD) has no substantial effect on surface PM air quality forecast, for the examined event.

I find it only unfortunate that authors only assess the first-day forecasts. This only a limited view of the forecast capability of the system is presented. It would have been very interesting if the authors would have shown the (likely decay in) forecast capability for the second and third day forecasts, as this would give a better handle for 'citizens and policy-makes' on the reliability of forecasts on a time scale where they are able to take action, which is one of the key objectives of this study. Is it possible for authors to make statements on this?

Apart from this, I only have some small comments.

We agree with the suggestion raised by the Reviewer and thus in the Revised Manuscript (RM) we have included IFS and RegEns day-2 and day-3 forecasts of PM_{10} and $PM_{2.5}$ in Figures 9, 10, S2 and S3. Additionally, we have extended Figure 11 including the results from day-2 and day-3 forecasts. Accordingly, we have modified/added several parts in the RM.

The following paragraph is included in the RM at P11, L21-29:

"The capability of IFS and RegEns systems to forecast the observed PM_{10} and $PM_{2.5}$ surface concentrations two and three days in advance, is finally discussed. As depicted in Figure 9, IFS day-2 and day-3 forecasts reproduce the distinct increases in observed PM_{10} surface concentrations exhibiting similar FGE values but lower correlation scores (in most of the stations) compared to day-1 forecast (Fig. 11c and d). The same applies in the case of $PM_{2.5}$ (Fig. 10), except that the correlation scores for IFS day-2 and day-3 forecasts are not systematically lower than that of day-1 forecast (Fig. 11c and d). As regards the RegEns, although it fairly predicts the observed peaks in PM_{10} and $PM_{2.5}$ up to three days in advance (Fig. 9 and 10), there is a systematic deterioration of its performance in terms of temporal variability over forecast

time. More specifically, the correlation coefficient decreases from day-1 to day-2 forecast and from day-2 to day-3 forecast for almost all examined stations (Fig. 11e)."

The following sentence is included in the Conclusions of the RM at P12, L25-27:

"A deterioration of the RegEns forecast performance is found over forecast time for both PM_{10} and $PM_{2.5}$, characterized by a decrease of the correlation coefficient for the vast majority of the examined stations, which is partially seen in IFS for the case of PM_{10}."

We have modified a part of the Abstract, see P1, L15-17 at the RM:

"Regarding the footprint on air quality, both CAMS global and regional forecast systems are generally able to reproduce the observed signal of increase in surface particulate matter concentrations. The regional component performs better in terms of bias and temporal variability, with the correlation deteriorating over forecast time."

Page 4, l 10L 'day-1 forecasts'. Are these the forecasts initiated at 0h00? Just because the CAMS operational system provides currently two forecasts per day.

Yes. Both CAMS global and regional forecasts used in the paper are initiated at 00:00 UTC. In the RM, we have included the respective information at P4, L8 and P5, L12 as follows: "**(initiated at 00:00Z)**".

Page 4. The authors provide empirical formulae to compute the PM10 and PM2.5 concentrations from the global system, which are crucial for the definition of the contribution of aerosol to the air quality statistics. These equations appear rather empirical, and also different depending on the model version. Can the authors give some more background information as to how these metrics are designed? Also, to what extent does the definition of this metric contribute to differences seen in Figure 8?

These are indeed empirical formulae which are updated when necessary with IFS cycle updates. Regarding the PM_{10} and $PM_{2.5}$ formulae, the factors applied in sea salt bins are used to transform sea salt from 80% relative humidity ambient conditions to dry, while the rest correspond to the fraction of each aerosol type included in PM_{10} and $PM_{2.5}$. More information on the description and evaluation of the aerosol scheme used in IFS is provided by Remy et al. (2019) which we have included in the RM at P4, L18-19 as follows: **"A detailed description and evaluation of the aerosol scheme used in IFS can be found in Remy et al. (2019)."** The differences seen in Figure 8 are likely also due to the definition of PM_{10} and $PM_{2.5}$ in IFS and CAMS regional models, which is already stated in the manuscript as **"The aforementioned inconsistencies are likely due to the different definition of PM_{10} and $PM_{2.5}$ in IFS and each CAMS regional air quality model,.."** (see P10, L12-13 of the RM). A quantitative estimation of this is beyond the scope of this paper, yet that would be an interesting task for a future study.

Page 5, l3: "CAMS regional models assimilate PM10 and PM2.5" : does this hold for all CAMS regional models, or only for some? Also, are PM10 and PM2.5 the modeled tracer fields in the regional models, or are they computed from underlying aerosol composition fields, as is the case

for the global model? In summary, could the authors elaborate a little more on how PM is modeled in the regional system?

Indeed, not all CAMS regional models were assimilating PM_{10} and $PM_{2.5}$ at the time of the event. More specifically, CHIMERE and EURAD were assimilating both PM_{10} and $PM_{2.5}$, MOCAGE only PM_{10}, while SILAM and MATCH were assimilating only $PM_{2.5}$. PM_{10} and $PM_{2.5}$ in the regional models are calculated using underlying aerosol species, as it is the case with the CAMS global model. However, each model is free to choose how exactly this is done, as the aerosol schemes and thus the aerosol species of each model differ. These details have been added in the RM at P5, L3-8: **"Several CAMS regional models assimilate PM_{10} and $PM_{2.5}$ surface observations from various stations of the EEA's (European Environment Agency) Air Quality e-reporting database, but not satellite aerosol products. More specifically, during the period of interest (October 2017), CHIMERE and EURAD were assimilating both PM_{10} and $PM_{2.5}$, MOCAGE only PM_{10} and, finally, SILAM and MATCH were assimilating only $PM_{2.5}$. PM_{10} and $PM_{2.5}$ concentrations in the regional models are calculated using simulated aerosol fields specific to each regional model."**

page 7, line 26: "reducing the bias"-> reducing the error

Done.

page 8, line 1: "the percentage" -> the modeled percentage

Done.

page 10, line 30: "implementation": I would rather write "application".

Done.

page 11, line 9: "The CAMS regional system seems to better predict the: : :." : why not write something like: "for this event the CAMS regional system shows better .."

We have included the phrase **"For the examined event,"** in the beginning of the respective sentence (P12, L20).

References

Rémy, S., Kipling, Z., Flemming, J., Boucher, O., Nabat, P., Michou, M., Bozzo, A., Ades, M., Huijnen, V., Benedetti, A., Engelen, R., Peuch, V.-H., and Morcrette, J.-J.: Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1), Geoscientific Model Development, 12, 4627–4659, https://doi.org/10.5194/gmd-12-4627-2019, 2019