Reviewer #1

The paper by Mortier et al. reports the results of an interesting study on regional trends in aerosols, focussing on properties which are sufficiently observed by global and regional observational networks. The way the results are presented is appealing, in particular figures 5 and 6. The summary is nicely presented as a series of bullets. The impact of restricted sampling at the monitoring sites on the trends is well explained. I am in favour of publication, but have a number of minor comments that should be addressed before publication.

The authors would like to thank the reviewer for his detailed comments, which contributed to make the manuscript clearer and more intelligible. Below are the answers to the minor comments made by the reviewer.

Minor comments:

- Abstract: Please include a list of the models included in the study.
- A list of the models groups used in this study has been added in the abstract:
    - ***Former version***: The set of observed regional trends has then been used for the evaluation of the climate models and their skills in reproducing the aerosol trends
    - ***Updated version***: The set of observed regional trends has then been used for the evaluation of 11 models (CAMS-reanalysis, 6 AeroCom Phase 3 models, and 4 CMIP6 models) and their skills in reproducing the aerosol trends

- The introduction is too short to my opinion. There is very little information on previous aerosol model-model and model-observations intercomparisons. Activities like AeroCom, CMIP6, should be introduced.
- The introduction has been expanded with references to studies on model-model / model-obs comparisons. The mentioned activities are also introduced in the revised introduction. A reference has also been added regarding the large global AOD increase between 1950 and 1990.

- l89: "Samples are collected every third day". Is this only for the speciated analysis or also for total PM2.5 and PM10?
- Samples are collected every 3rd day for both speciated and PM2.5 and PM10 filters. They all follow the same schedule, and some speciated analyses are actually done on the same filters.

- l102: "The data have been screened" By the authors, by Aas et al. or by the instrument teams? Please give some more details: when is a site regionally representative? Is there a link with the model grid box size? Also for PM the representativity would be good to discuss (in sec 2.1.2).

- The screening has been done by Aas et al., which included instrument teams from the different countries. The words ‘regionally representative’ are referring to the fact that urban sites were not included in this study, as strongly affected by local emissions. This sentence has been removed as it might be confusing with the representativity study presented in this paper.

  Regarding the PM (sec 2.1.2), the following text has been added:
  “The stations are mostly located in remote and rural locations. This ensures a good representativity of the measurements as some chemical species contributing to PM observations (i.e., Organic carbon) can present different seasonality and spatial variability.”

- l117: Since Gliss et al. is in preparation, it would be good to provide more details on the selection/correction procedure.
- Gliss et al., is now publically available in ACPD. The reference has been updated in the manuscript.

- Sec 2.2.1. It is a bit strange to call the CAMS reanalysis a "climate model" (line 120). Maybe it is good to explicitly refer to "climate models, aerosol models and aerosol reanalyses" in the paper or at least clearly explain the types of models/reanalyses.
- Agreed and changed accordingly.

- l129: Please place reference to Inness et al between brackets.
- Done.

- l147: "ESGF nodes". Please explain the acronym and provide a web link or reference.
- An explanation of the acronym has been added as well as a link to the web interface. (Earth System Grid Federation: https://esgf.llnl.gov/)

- l154: ".. into one average time series." This may be removed from the sentence to avoid repetition.
- Done.

- l155: "not as easy to define when combining the trends for individual sites together." Why not? Is there a difference between the trend of a sum and the sum of individual trends?
- We acknowledge that the sentence was not clear enough as the main point was intended to be about the definition of an uncertainty from several individual trends.
  We have clarified and have rewritten the part:
  - ***Former version***: A first advantage of this method is that a single trend can be computed in a given region, with an associated significance and uncertainty, which is not as easy to define when combining the trends for individual sites together.
  - ***Updated version***: One advantage of this method is that a single trend can be computed in a given region, with an associated significance and uncertainty.

It is difficult, apart from a diversity analysis, to define such an uncertainty when combining individual trends

The question raised about the difference between the trend of a sum and a sum of individual trends is relevant though. Differences can actually happen due to the fact that the observations might start at different times in areas associated with contrasted absolute levels.

This is an extreme example illustrating that the sum of individual negative trends (solid lines) could be associated with an overall increasing trends (dashed line) if the representativity of the observation network changes dramatically in time.

- l156: "with our aggregation method, even a station that has not provided a sufficient amount of data for computing a trend at its location can still contribute to the computation of a regional time series." This sounds a bit dangerous. Including incomplete time series that have an offset with respect to the mean will introduce a spurious trend. Please provide more discussion on how much this may impact the trends derived?

The comment is directly connected to the comment below on l180.

- l180: "minimum of 300 valid daily measurements". This is basically one year of data. Why are the authors not more selective, e.g. allowing only stations with measurements for 50% of the time? Why this choice?

The authors acknowledge that taking into account incomplete time series could introduce erroneous trends. This aspect is covered by the representativity study section (3.3) which compares trends computed from incomplete and complete time series. However, the derivation of the trends from the observations is mostly used, in this work, for making a model trends validation dataset, when colocating observation and model data.

The minimum of 300 valid daily measurements has been chosen in order to remove temporary AERONET stations (DRAGON campaigns). This is an arbitrary choice that the authors assume to provide the best compromise, considering the different regions, and different variables, between availability and robustness. For comparisons, AOD trends were computed with the 50% time criteria. This resulted in no available trends in Australia, due to the lack of remaining observations. However, the AOD trends computed with the 300 daily measurements criteria shows a non significant decrease of 1.5%/yr which is assumed to be representative of the trend over the whole period and region, as shown by the representativity study. This

illustrates the fact that in some cases, the partial information provided by the observations remains valuable.

- l160: "Seven regions". Please provide the corner locations of the regions, for instance in a small table. Is there overlap between Asia and NAfrica? Are some stations used in multiple regions?
- A table with the regions coordinates has been added in supplementary.
- In its former version, an overlap was observed between NAFRICA and ASIA, which caused the Aeronet Solar_Village station to be considered in both regions. We have now removed the overlap in the updated version.

- l182: "three valid points". Please specify more explicitly. Is a valid point a daily mean observation for one station? Or something else?
    - ***Former version*:** A minimum of three valid points (daily or monthly depending on the available resolution) is required per month to be present in the overall regional time series
    - ***Updated version*:** A minimum of three valid stations is required to be present in the overall regional time series to produce a valid point. In other words, if the available time resolution is daily, at least three stations need to provide valid data for a certain day in order to produce a valid regional mean for that day.

- l201: "Mann-Kendall test ... Theil-Sen". Please provide a reference.
  The following references have been added to the manuscript:
    - Mann-Kendall: Hamed, Khaled H., and A. Ramachandra Rao. "A modified Mann-Kendall trend test for autocorrelated data." Journal of hydrology 204.1-4 (1998): 182-196.
    - Theil-Sen: Sen, Pranab Kumar (1968), "Estimates of the regression coefficient based on Kendall's tau", Journal of the American Statistical Association, 63 (324): 1379–1389, doi:10.2307/2285891, JSTOR 2285891, MR 0258201.

- l208: "residuals". Please provide a definition of the residuals. Is this computed based on the yearly-mean, regional-mean values?
  We have clarified this in the manuscript: Indeed, the residuals are computed based on the difference between the linear trend and the yearly mean values of the regional time series.

- l228: "model subset of data". What does this mean? Make very clear that this study is based on model data only.
    - ***Former version:*** In order to evaluate [...] two sensitivity studies, focusing on the time sampling and the space sampling, have been conducted using model subsets of data.

    - ***Updated version:*** In order to evaluate [...] two sensitivity studies, focusing on the time sampling and the space sampling, have been conducted using NorESM2 model data subsets.

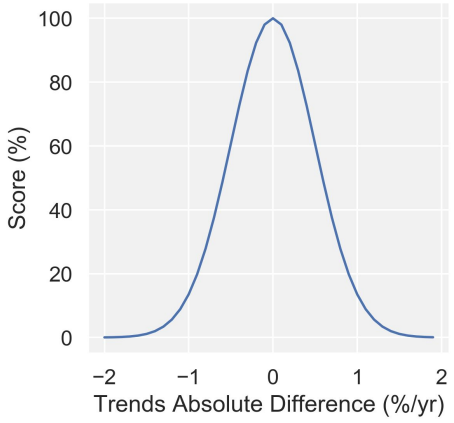
- l230-236: I read the description of the datasets a few times and still I am not sure I understand what is done. Please define the "Ref" and "Exp" very explicitly, maybe even using formula's.

- A sentence has been added before the defintion of the 'ref' and 'exp' datasets.
  The definitions have also been reformulated:
    - ***Former version:***
        - Time representativity study
            - Ref_{time}: Collocation in space and time
            - Exp_{time}: Collocation in space using complete time-series
        - Space representativity study
            - Ref_{space}: Collocation in space using complete time-series (=Exp_{time})
            - Exp_{space}: All grid-points in region using full time-series
    - ***Updated version:***
      The reference dataset corresponds to the model data co-located to the available observations while the experiment dataset uses all model points.
        - Time representativity study
            - Ref_{time}: Model data collocated in space and time with available observations
            - Exp_{time}: Model data collocated in space with available observations using the complete model time-series
        - Space representativity study
            - Ref_{space}: Model data collocated in space with available observations using using the complete model time-series (=Exp_{time})
            - Exp_{space}: All of the model grid-points in the region using the complete model time-series

- l238: sigma = 0.5. Why 0.5? It seems sigma has a unit %/year?!
  As mentioned l240, the choice of Sigma=0.5%/yr is an arbitrary choice which results in a representativity score of 50% when the trends difference (between a reference and an experiment subsets) is of 0.5%/yr. The unit was indeed forgotten, and has been added in the manuscript.

- Eq.2: I'm struggling with the representativity and the normal distribution. A relative trend is expressed as %/year. Therefore this has a dimension. But a normal distribution takes a dimensionless quantity as argument? Therefore Eq.2 does not make sense.
  The authors agree that this part of the manuscript was misleading and have modified the manuscript in order to clarify the representativity strategy.

The normal distribution has been used in order to map the trends difference to a score, expressed in percent, as shown by the following figure.



The equation 2 was misleading and has been removed since it was not further used in the manuscript.

- ***Former version*:** The difference between the relative trends are computed for each parameter and region. Those differences are then converted into a score (\unit{\%}) by using a normal distribution f described by a mean μ = 0 and a standard deviation of σ = 0.5. The choice of these parameters leads to a representativity score of 100% when there is no difference in the trends of a reference and an 240 experiment dataset, while a difference of 0.5\unit{\%/yr} would indicate a representativity score of 50%.
- ***Updated version*:** The difference between the relative trends are computed for each parameter and region. In order to summarize the representativity, those differences are then converted into a score (\unit{\%}) by using a mapping function which has been defined based on a normal distribution. The choice of the parameters describing this function leads to a representativity score of 100\% when there is no difference in the trends computed for a reference and an experiment dataset, while a difference of 0.5\unit{\%/yr} obtained with these two datasets would indicate a representativity score of 50\%.

- Fig. 4: Why is the number of points a coloured region, and not a simple line? Fig. 4: Which model is used for this?
- There was no particular reason for the fact that the points are displayed were colored regions and not simple lines. The Figure has been re-processed with simple lines in order to avoid any confusion.
- NorESM2 was used for the computation of the representativity study since all the variables were available with this model. A sentence in the text and the Figure caption has been added.

- l278: Is there an explanation why PM2.5 is a larger fraction of PM10 in Europe compared to North America?
  The following text has been added in the manuscript:

"This difference in the relative proportion of fine particles against coarse particles in Europe and North America may be due in part to our definition of regions. Putaud et al. (2010) presented a phenomenology of PM data in Europe showing coarse aerosol tended to be highest in southern Europe which in our study is part of the North Africa region. The discrepancy in the relative proportions of coarse and fine aerosol in Europe and North America may be exacerbated by both a decrease in North America of the fine particles concentration due to pollution mitigation strategies coupled with the growth of the coarse mass due to increasing contributions of natural and agricultural sources, particularly in the western half of the U.S. (Hand et al., 2019a)."

- l305: "Collaud Coen". Provide reference.
- Done

- l310: "smoother". What does this mean?
- The word 'smoother' has been replaced by 'more homogeneous'

- l314: "somewhat higher". Do you mean "somewhat more negative" or "less negative"?
- The authors meant 'more negative'. This has been replaced in the manuscript.

- l321: The difference with Collaud Coen deserves more discussion. Is this trend significant? Is the difference understood?
  The following text has been inserted in the manuscript:

  This probably illustrates the difference of methodology which consists of computing the mean of station trends in one case, and the trend of a regional time series in the other case, especially when only few measurements are available. However, as shown by the representativity study (\ref{fig:obs_trends}), the non-significant increase of +0.0\%/yr found, in this study, with the observations is similar to the trend derived over the whole region and using complete time series of the NorESM2 model data.

- l339: second increase: should this be "decrease"?
- Yes indeed, this error has been fixed in the manuscript

- l394: "could be caused by increased wet scavenging". How does this match with the SO4 negative trends?
  Agreed. The hypothesis has been removed from the manuscript.

- l464: The large trend in Arctic? Do the authors have any idea how to explain this?
- The following text has been added to the manuscript:
  "In addition, one finds large and significantly increasing trends in the high Arctic that could be explained by a change in the air mass circulation pattern, or by the increase of open sea, which might contribute to a higher production of sea salt aerosols."

- l542: "brightening Streets et al. (2006); Norris and Wild (2007)." Please place references between brackets.

- Done

- References: Please provide the DOIs for all the cited papers.
- Done

- Olivié, D. et al.: in preparation. Please remove or provide full author list and title. The same remark holds for a few other preprints/in preparation papers.
- These papers were removed from the manuscript