Responses to the interactive comments on "Detection and attribution of aerosol-cloud interactions in large-domain large-eddy simulations with ICON" by Montserrat Costa-Surós et al. to

Anonymous Referee #1

General comments: The authors present analysis of new very high resolution simulations over the EU domain for one day near peak emissions in 1985 and one day in the present decade with relatively low emissions. The authors carefully analyze the high-resolution simulations using satellite and ground-based data. They find that AOD differences, and Nd differences between 1985 and 2013 are reproduced. Changes in cloud macrophysics are too small relative to natural variability to observe. The authors derive an ERFaci for the global mean using scaling with traditional GCMs. The paper is a very nice analysis of cutting-edge new simulations and provides an interesting new evaluation of ERFaci.
We thank the reviewer for their thoughtful summary of our study.

I have the following **major issues** with the paper:

A lot of the paper is given over to ground-based remote sensing. This is fine, but it is not a field that I am very familiar with and I recommend that a reviewer who is an expert be nominated to comment on this. However, I am concerned by the characterization of standard deviation as uncertainty in comparing observations and models (as discussed in specific comments) and I think this needs to be explained more clearly. I am not sure that the authors have made a meaningful comment about the adjustment strength, besides the fact that adjustments are small compared to meteorological variability and are hard to see in one day of data- which doesn't preclude them being important to ERFaci. Critically, I think the scaling to the global ERFaci could be done better (see specific comments below) by expanding the number of GCMs and by showing that the relationship is linear.
We thank the reviewer very much for his/her interesting comments. We will proceed to address all of them in the following specific comments.

Specific comments:

Pg1 Ln11: I kind of follow what the authors are trying to say here, but it is a little easy to lose track. I would suggest not using reference and perturbed to refer to 2013 and 1985 in the abstract. It will be easier to follow which conditions are consistent and inconsistent. Effectively it sounds like the model needs the appropriate year of aerosol data to get the right output, which could be said more succinctly.
We thank the reviewer for his/her appreciation. The sentence has been simplified in the abstract.

Pg2 Ln6: What is large LWP? Is it in or area-mean LWP? I am not sure what I should really be taking away from this result. Is it really a key result that needs to be shown in the abstract?
We really think this has to be said in the abstract since it is an important result. Following your suggestion, we have added "(LWP >200 g m^{-2})" in the sentence, and clarified it is in-cloud LWP (consistent with the observations).

Pg2 Ln16: The results in Rosenfeld 2019 are no longer accurate. There is an errata that revokes most of the findings of the original paper.
We thank the reviewer for the observation. The reference has been removed from the paper.

Pg3 Ln9 "to what extent".
The typo has been corrected.

Pg3 Ln17: Split this into two sentences. CCN is changed and INP is not. Direct effects are not considered. This is really confusing. How is AOD being evaluated if the direct effect isn't considered?

We thank the reviewer for his/her comment, the sentence has been divided into two and the information extended to be more clear since the direct effect (and the semi-direct) is, in fact, considered but no changes are made to it in the different simulations carried out. That means that the changes in the CCN are not affecting the direct (and semi-direct) effect radiative balance in our simulations. We are able to evaluate the AOD because of the additional offline calculations based on COSMO-MUSCAT, where the CCN number concentration of the multi-modal size distribution at a fixed supersaturation is calculated according to Abdul-Razzak and Ghan (2000), as explained in section 2.2.

Pg7 Ln17: Please discuss Song et al 2018 (https://www.geosci-modeldev.) in the context of using COSP for ICON at this resolution. What is used to drive COSP in this case? Is there subgrid variability assumed?

We thank the reviewer for the Song et al. (2018) reference; this study highlights very well the importance of properly including subgrid variability for GCM evaluation via COSP. However, the conclusions of Song et al. (2018) mainly concern the preferred usage of a model-specific sub-grid information (allowed in COSPv2) instead of the COSP sub-column generator, in order to accurately account for the GCM sub-grid cloud and hydrometeor variability (1.9x2.5deg simulations were used to reach these conclusions). This is not relevant for ICON-LEM, where no sub-grid variabilities of cloud and aerosol properties are considered. Consequently, COSP was used without subcolumn varibility and was driven directly by the grid-level ICON outputs. The goal is here to apply the satellite retrieval algorithm on a pixel level, mainly to reproduce the instrumental sensitivity limitations (i.e. which clouds are too thin to be detected, where does the signal saturate?). Note that the 156-m icosahedral outputs were aggregated into a 1-km^2 lat-lon grid (fitting the MODIS resolution) prior to being used in COSP, but we still decided to not to include sub-grid (< 1km) cloud variability to stay consistent with the MODIS retrieval algorithm, which performs its retrievals by ignoring sub-pixel variabilities.

A new sentence has been added into the "Observation / Satellite-based" subsection: "No sub-column variability is used in COSP, consistently with the lack of sub-pixel variability in MODIS retrievals".

Pg9 Ln9: Typo- sentence needs to be reordered. Maybe "AOD is only available over the North Sea region for xx% of retrievals." To reduce ambiguity.

The sentence has been reformulated.

Pg9 Ln17: The authors show a systematic difference in the mean CCN profiles from observations and the CCN used to drive the model. I think the authors are somewhat misusing the uncertainty range. Don't you want uncertainty in the mean, not just the variability, which is what this shows? Shouldn't these be standard error in the mean? Ultimately it seems like there is a 10-30% overestimate in CCN relative to the observations (I assume the standard error in the mean is small). Can the authors convert that to an overestimate in Nd using the nucleation scheme, which is the more relevant quantity in this study?

We appreciate the reviewer's suggestion. In fact, we chose to show the median and the 25-75th percentiles in purpose because we think they better show the variability of the AOD in the region, rather than the mean and the standard deviation. In this sense, we have removed the word "uncertainty" in Table 2 caption to avoid misunderstandings. Regarding the Nd inquiry, it is discussed in Fig. 3.

Pg 12 Ln8: I am not sure that 10% change in LWP is small (am I reading table 3 right?). It's certainly true that variability in LWP due to meteorological variability is large, but this doesn't really tell us anything about the radiative forcing induced by adjustments.
The reviewer is right, 10 % change is not a small change, however it is too small for detection and attribution of LWP changes by satellite, considering current retrieval uncertainties, therefore the change in LWP is not detectable by MODIS on the studied case, and this is what we mean in the sentence. We clarify that the LWP change translates into a substantial systematic effect on the radiation balance and, thus, the aerosol effective radiative forcing.

Pg12 Ln11 Is the cutoff for large LWP? Does this just mean not thin clouds?
The 200 g m^{-2} value refers to the analysis of Fig. 5. At the large-LWP tail of the PDFs, an increase of high LWP values (higher than about 200 g m-2) clearly appears in the perturbed simulation by comparison to the reference and satellite observations. As explained in this paragraph, we attribute this adjustment effect to invigoration of convective clouds as a consequence of higher Nd. This observation is of particular interest because such adjustments could in principle be detectable based on MODIS-like satellite retrievals. We clarify this in the revised manuscript.

Pg12 Ln18: How many more days of simulation would you need to beat down the noise and be able to see the LWP perturbation clearly?
We could only speculate, since such a method hasn't been applied yet to large-domain large-eddy simulations. In GCM analyses, even for nudged simulations, yearlong integrations are necessary. In an LES we believe a shorter analysis is sufficient due to the very much larger amount of independent columns.

Pg 20 Ln29 I think the authors are just calculating the ERF over Europe versus the global mean and coming up with a scaling factor. I think a better approach would be to plot ERF_EU_1985-ERF_EU_2013 versus ERF_global_mean_PD for each CMIP5 model. The way that the authors are doing this assumes linearity in this relationship, which is not necessarily true since the EU in 1985 is so polluted. Based on Carslaw et al. (2013), I am not sure that this calculation should really reduce uncertainty much, but Carslaw et al. (2013) paper implies strong non-linearity in the relationship between local ERF and global-mean ERF. If the authors could increase the number of GCMs beyond 4 and show that the relationship is linear this would be a more robust calculation.
How do the authors deal with the direct effect not being calculated in the simulations for this comparison since it will be in the GCMs (Pg3 Ln17)?

We thank the reviewer for highlighting that this point requires more attention; reviewer #2 had a very similar concern. The reviewer indeed was right that our previous analysis was overly superficial. Fortunately in the meanwhile, the new 6th Coupled Model Intercomparison Project (CMIP6) provided output from the new multi-model ensemble. This is very valuable to the problem here in question since the part of CMIP6 that addresses the radiative forcing (the RFMIP) has one simulation that allows to diagnose the transient ERF due to aerosols. From this new output, we were now able to assess the scaling in a more thorough way. We explain now in the revised manuscript in much more detail the revised procedure to scaling the forcing, and – more importantly perhaps still in response to this reviewer remark – we much better highlight and quantify the uncertainties. The new approach also allows to better isolate the aerosol-cloud interactions.

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., . . . Pierce, J. R. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. Nature, 503(7474), 67-71. doi:10.1038/nature12674