Reply to reviewer #2 comments

Fanourgakis et al. perform a large model intercomparison study focusing on the evaluation of aerosol size distributions, CCN, and Nd with surface observations as a reference. The main conclusion is that models have substantial problems in simulating the concentrations of CCN, with on average strong underestimations. In contrast, Nd as derived applying an off-line parameterisation with prescribed vertical wind is much better correlated to observations-tied estimates.

The study is a very large effort and worth publishing. It is also useful for upcoming assessments such as the IPCC AR6. It is, however, a pity that it was impossible to point to specific parameterisation shortcomings, since obviously neither in the multi-model ensemble nor in the PPE, specific parameterisation choices and/or specific models seemed to systematically correlate better to the observations than others. The only hint is that organic aerosols seem to be more difficult than other types.
I think the authors should better explore the Nd result in a revision.

Reply: We thank the reviewer for the positive evaluation and we will improve the manuscript according the reviewers suggestions.

The first thing that would be very useful and probably not difficult to do is to compare the Nd the models actually compute to the idealised ones computed here off-line. It would be very interesting to see how large the potential biases in the parameterisations of droplet activation are in comparison to the aerosol biases.

Reply: We agree with the reviewer on the interest of evaluating potential biases in the parameterizations of droplet activation used in the models. We see this as a natural follow-up to the present work, but requires considerable additional work to diagnose and provide conclusive statements on – and careful design to get the most useful information. Towards that, we feel that our prior studies (Morales-Betancourt et al., 2014a) provide a good example, where the source of Nd prediction discrepancy for two state-of-the-art parameterizations in the CAM5 global model was unraveled using adjoint sensitivity analysis. The findings of that study, not only lead to identification of parameterization biases, but also pointed to exactly which aspects of the parameterization (i.e., water uptake from large CCN) were not captured adequately, leading to the highly improved droplet parameterizations (Morales-Betancourt and Nenes, 2014b) that was used in the current study.

To clarify this, a relevant comment has been added at the end of the first paragraph of the description of CDNC calculations in section 2.4.: 'Note that evaluation of the differences in CDNC calculations by the different models that are derived both from the parameterizations used and from their input variables would require a different model intercomparison design than here and is planned for the future. Morales-Betancourt et al. (2014a) provide a good example, where the source of Nd prediction discrepancy for two state-of-the-art parameterizations in the CAM5 global model was unraveled using adjoint sensitivity analysis. That study pointed to exactly which aspects of the parameterization (i.e., water uptake from large CCN) were not captured adequately, leading to the highly improved droplet parameterizations (Morales-Betancourt and Nenes, 2014b) that was used in the current study.'

I also wonder why the deviations in Nd are in some cases qualitatively different from the deviations in CCN. Some explanation is needed.

Reply: Indeed Finokalia and Cabauw provide N_d calculations (Figure 8 and related sensitivities) that qualitatively differ from the deviations in CCN0.2 (CCN at 0.2% ss; Figure 2). We explain this pattern by the inability for models to fully capture (overestimate or underestimate) the levels of the largest particles (> 250nm) and/or their hygroscopicity. This is an important conclusion that will be emphasized in this study. These results demonstrate that CCN at selected supersaturations (and respective model-observation discrepancies) cannot be generally used as proxy for cloud droplet number behavior, since the maximum supersaturation that develops inside the cloud (hence droplet number) responds to changes in aerosol and vertical velocity levels. This is even further complicated by the potential for model biases to change even sign across at cloud-relevant supersaturations.

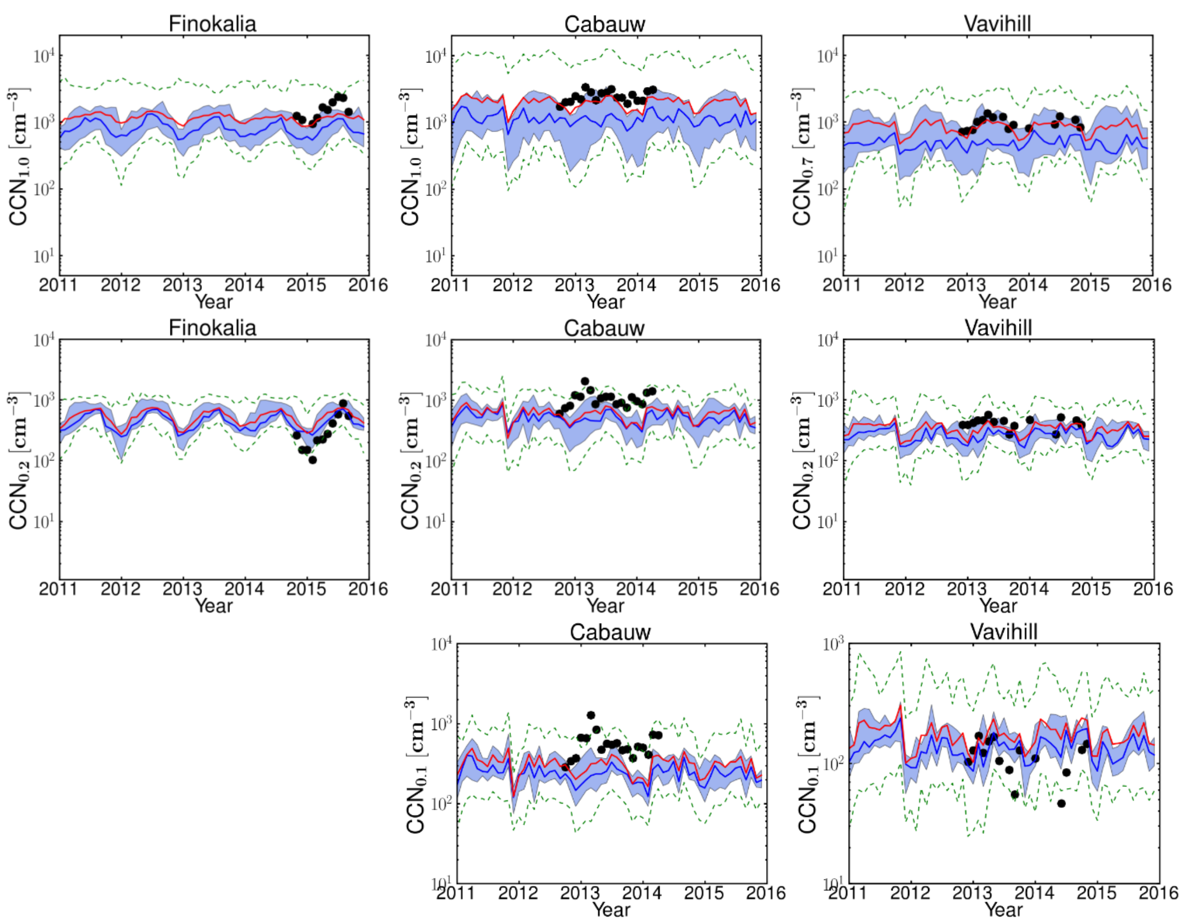
We extended the discussion as follows:

'Based on the behavior described above, one can explain the N_d predicted from simulated and observed CCN spectra. This is straightforward for Jungfraujoch and Mace Head stations. For Cabauw and Vavihill the observed-to-simulated ratio turns from a substantial overestimation in $CCN_{0.2}$ to an underestimation in N_d, and the opposite is found for Finokalia. This can be explained as follows. At both Cabauw and Finokalia, s_{max} derived from observations is very low (approaching in the summer 0.07% at Finokalia and 0.04% at Cabauw; Fig. 8). The models overestimate these low values of s_{max} and such values are indicative of the presence of large particles (>250nm) with sufficient hygroscopicity at these stations that are not captured by the models. Indeed, at Cabauw the available observations of CCN at 0.1% supersaturation show a larger underestimate by the models than for $CCN_{1.0}$ and $CCN_{0.2}$ (Figure below-new supplementary figure S16), also pointing to a model underestimate of the largest particles (>250nm) which induce the very low s_{max}. The overestimate in s_{max} leads to an underestimate in N_d by the models for all seasons except winter at Cabauw when the models at high updraft velocity capture the observationally-derived N_d levels. Furthermore at Finokalia, $CCN_{1.0}$ is underestimated year-around, indicating that, in addition to the largest particles, the very small particles (smaller than 50 nm) that activate at 1.0% supersaturation and/or their hygroscopicity are also underestimated by the models there. On the other hand, larger particles than 120 nm that activate at 0.2% supersaturation are overestimated especially in winter and slightly underestimated in summer. Therefore the global models have significant difficulties in capturing the aerosol size distribution and hygroscopicity at Finokalia – which in turn translate to counterintuitive discrepancies in N_d.

At Vavihill a somehow different behavior is found; the underestimate of CCN at supersaturations of 0.2% and 0.7% changes to an overestimate at supersaturation 0.1% mainly in summer (supplementary Figure S16), indicating an underestimate of fine particles and/or their hygroscopicity and an overestimate of the largest particles and/or their hygroscopicity in particular during summer. This agreement of model results with observations during winter and the overestimate of CCN at 0.1% supersaturation during summer can explain the similar behavior of modelled N_d

The difference between model and observationally-derived $\partial Nd/\partial w$, clearly supports the above statements. Since observations predict a suppressed s_{max} compared to model distributions (Fig 8), water vapor competition effects in the observations are much more severe than in the model, indicating that observations are much more (positively) sensitive to updraft velocity. The opposite trends are seen for activation fraction ($\partial Nd/\partial Na$), given that reductions in aerosol reduce competition effects. The reduced water vapor competition effects at higher updraft velocities, and the trend in CCN error also generally explain why the sensitivities are smaller for the highest updraft velocity. '

At the end of the section we added the following concluding statement: ‘ Because of the discrepancy in sensitivities $\partial Nd/\partial Na$ and $\partial Nd/\partial w$, models may be predisposed to be too “aerosol-sensitive” or “aerosol insensitive” in aerosol-cloud climate interaction studies, even if they may capture average droplet numbers well. This is a subtle, but profound finding that only the sensitivities can clearly reveal and may explain inter-model biases on the aerosol indirect effect. Few published efforts (apart from Morales et al., 2014a and Sullivan et al., 2016) can demonstrate this, and non over a range of models and using a considerable aerosol dataset for evaluation as here performed.’



Supplementary Figure S16. Monthly ensembles for the years 2011-2015 of the CCN number concentration for supersaturation 0.2 % ($CCN_{0.2}$), 0.1% ($CCN_{0.1}$), 0.7% ($CCN_{0.7}$) and 1.0% ($CCN_{1.0}$) when observational data are available for Finokalia, Cabauw and Vavihill.

Specific comments

p4 l12: Is “OA” really defined as a fraction here? - contradiction to p8 l8

Reply: We thank the reviewer for pointing out this misleading sentence. The word ‘fraction’ is removed.

p5 l13: really opposite, i.e. of different sign?

Reply: We changed ‘opposite’ to ‘different’ *p6 l30: Modal aerosol module*

Reply: corrected

p7 l2: Kirkevåg

Reply: corrected

p7 l11: "a few models account for melting and sublimation of ice crystals" - I have the impression this sentence is incomplete.

Reply: We added 'also'. The sentence now reads: 'In addition, while all models account for in-cloud scavenging of aerosols and for the aerosol release from evaporation droplets, a few models account also for melting and sublimation of ice crystals'

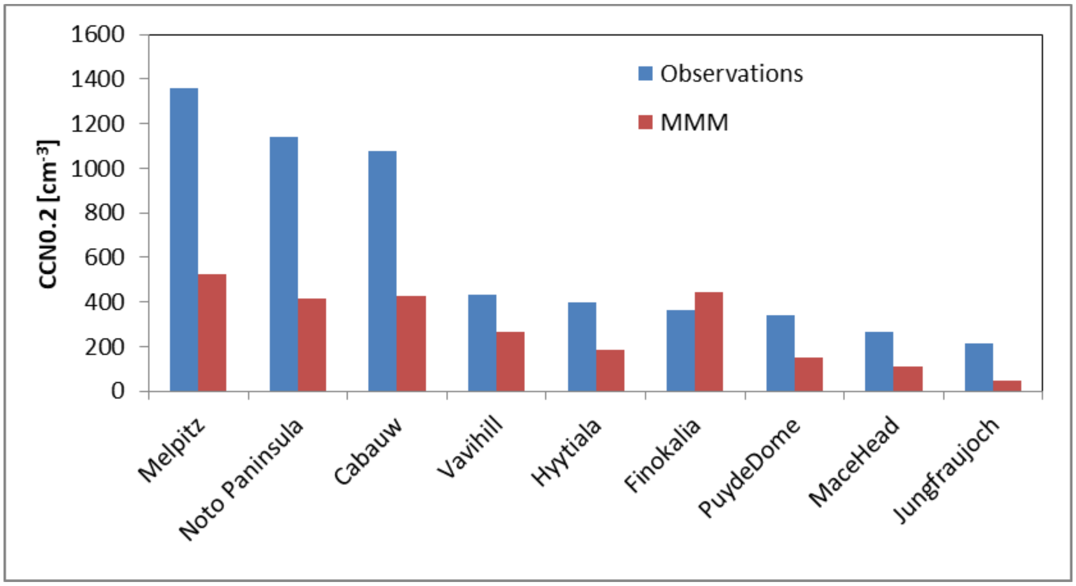
p10 l15: why this very coarse resolution, and not just the resolution of the coarsest-resolved model?

Reply: This is very close to the coarsest-resolved model of 4^{o}x5^{o} (as shown in the supplementary Table S1).

p11 l6: the conclusion that the correlation is captured "satisfactorily" needs quantification: what can be considered "satisfactory"? How do the authors quantify these relative differences? - perhaps a figure like Fig. 3 but for the spatial variability would be useful?

Reply: This part of the discussion refers to the CCN at 0.2% supersaturation. To address this issue, we have plotted the mean $CCN_{0.2}$ as calculated from the observations and as computed from the daily MMM for the days with available observations (figure here-below). In this figure the stations have been ranked based on $CCN_{0.2}$ observations in decreasing levels and this is nicely followed by the MMM with the exception of Finokalia station where the models overestimate $CCN_{0.2}$ as shown in Figure 2 and already discussed in the manuscript.
We also calculated the Pearson linear correlation coefficient (r) for the correlation of observed and modelled $CCN_{0.2}$ at each station as indicators of the ability of MMM to reproduce the temporal variability in the observations and found r values between 0.44 (for Melpitz) and 0.83 (for Mace Head), showing significant covariation of model results with observation.
We have modified the discussion accordingly:

'Despite the quantitative differences in the estimation of the $CCN_{0.2}$ concentrations, models are able to qualitatively capture the relative differences in $CCN_{0.2}$ concentrations between stations, as well as their seasonal variations. Comparing the $CCN_{0.2}$ as calculated from the observations and as computed from the daily MMM for the days with available observations for the stations, we find significant Pearson linear correlation coefficient (r) that vary between 0.44 (for Melpitz) and 0.83 (for Mace Head), showing significant correlation. Furthermore, ranking the stations based on the observed mean $CCN_{0.2}$ levels (supplementary figure S17) we find the corresponding MMM mean follows this station ranking with the exception of Finokalia where, as further discussed, the models overestimate the observed $CCN_{0.2}$ although they capture well (r=0.76) the observed temporal variability.'



New supplementary figure S17: Mean $CCN_{0.2}$ as calculated from the observations and as computed from the daily MMM for the days with available observations. The stations have been ranked based on $CCN_{0.2}$ observations in decreasing levels.

p11 l7: The index of agreement, since it is not a conventional metric, would need to be defined in the main text. The authors further should motivate why this IoA provides extra information that substantially goes beyond NMB and NME.

Reply: We provide such information at the end of section 2. 'For the comparison of model results with observations, a number of statistics variables have been calculated and defined as shown in the supplementary material S3.2. Hereafter we discuss

the Index-of-Agreement (($\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|O_i - \bar{O}| + |P_i - \bar{O}|)^2}$),

the normalized mean bias ($\mathrm{NMB} = \frac{\sum_{i=1}^{N}(P_i - O_i)}{\sum_{i=1}^{N} O_i} \times 100\%$)

and the normalized mean error ($\mathrm{NME} = \frac{\sum_{i=1}^{N}|P_i - O_i|}{\sum_{i=1}^{N} O_i} \times 100\%$),

where M are model results, O are observations and NMB, NME and IoA are used to quantify the performance of the models to reproduce observations. IoA is a measure of the agreement of model results with the observations. In this study we use all three for the evaluation of the capability of the models to reproduce the observations. We now calculate also

the Pearson linear regression coefficient ($r = \left[\frac{\sum_{i=1}^{N}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}(P_i - \bar{P})^2}\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2}}\right]$)

as a measure of the ability of the models results to represent the variability in the observations.

p13 l1: "satisfactorily" - again a quantification would be helpful. Is this significantly better than for CCN0.2?

Reply: The performance of the model to simulate the observed seasonal variability can be quantified by the correlation coefficient. We have now calculated the statistics and the corresponding correlation coefficients for the MMM and we have modified the discussion accordingly to provide this information in the manuscript.

“Comparing model performance for CCN at low supersaturation ($CCN_{0.2}$, Figure 2) and at high supersaturation ($CCN_{1.0}$, Figure 4), $CCN_{1.0}$ is systematically underestimated by the models across all stations. The NME of MMM for $CCN_{0.2}$ ranges from 45% (Finokalia) to 81% (Jungfraujoch) for the different stations with significant correlation coefficients between 0.44 (Melpitz) and 0.86 (Mace Head) indicating that the model are able to simulate the temporal variability in the observations. For CCN at the highest supersaturation with available observations the NME varies from 50% (Finokalia) to 74% (Mace Head) and the correlation coefficients from 0.37 (Melpitz) to 0.78 (Mace Head) (see also supplementary Table S6). These results indicate that $CCN_{0.2}$ is in general better captured than CCN at higher supersaturations, both in absolute values and in temporal variability.”

p14 l5: the + for NME is unnecessary

Reply: we will remove it

p14 l10-12: Why would the authors call an NME of 97% “reasonable”?

Reply: We rephrased this discussion p14-l8-20 of the ACPD version as follows:
“Figure S1 is similar to Figures 2 and 4 but shows particulate SO4, OA mass in PM_1 particles at the nine stations as well as model results for DU and SS. Strong seasonal variations of the SO4 mass of about one order of magnitude are observed and simulated at the alpine site, Jungfraujoch and at the coastal background stations, Mace Head and Finokalia, although winter minima are overestimated by the models at these coastal sites. Smaller and no clear seasonal variation of SO4 is observed at the boreal forest environment of Hyytiälä, the rural background station Cabauw and at the highly polluted Melpitz station during the year. At these three stations, the MMM underestimates the observed annual mean concentration of SO4. Strong seasonal variations of the OA mass are observed and simulated at Mace Head, Finokalia, Jungfraujoch and Hyytiälä, while no distinct seasonal cycle in organic mass is seen at Cabauw and Melpitz and the OA concentrations are underestimated by MMM at all sites. The IoA between the MMM and the observations is between 0.28 and 0.62 for all stations. A detailed analysis of each model separately (Supplementary Table S6) shows that the OA mass concentration is underestimated (mean NMB is -37%) by nine of the models and overestimated by six of them (range of NMB -97% to 216%).

p15 l6: I rather have the impression that it is completely unclear: in 5/9 station, it is indeed longer in winter, but in 4/9 the opposite.

Reply: We agree with the reviewer. We added the scores. The sentence now reads:

‘Depending on the station, the persistence time is longer during winter (5 stations) than during summer (4 stations).’

p15 l7: in 6/9 cases, the MMM is consistent with the summer-winter change as the obs show, in 3/9 cases, opposite. Is that “qualitatively capturing” the change?

Reply: We rephrase the sentence to be more precise as suggested by the reviewer:
‘The average persistence of the $CCN_{0.2}$ number concentrations simulated by the individual models is consistent with the observed change between winter and summer at 6 among the 9 stations.’

p15 l32: it doesn’t hinder an eventual decrease in smax, it implies it.

Reply: The sentence now reads:
'Increases in CCN concentrations tend to increased N_d and associated water vapor depletion in the early stages of cloud formation; this in turn hinders the development of supersaturation and implies an eventual decrease in s_{max}.'

p16 l24: The competition effect cannot explain why the ratio observed/simulated for Cabauw and Vavihill turns from substantial overestimation to underestimation, and why the opposite is found for Finokalia.

Reply: See detailed reply to reviewer's major comments.

p31 Fig. 2 – I don't see the dashed lines for the observations.
Reply: Following reviewer's 1 comment we replaced 'dots and dashed lines' by 'symbols' and we have redrawn the figures to remove the dashed lines that were not clearly seen from the observations.

p34 Fig. 5 – Would it be useful to show the MMM as well?

Reply: Figure 5 has been redrawn to include MMM as suggested by the reviewer. The figure caption and the discussion have been modified accordingly.

References

Morales Betancourt, R. and Nenes, A.: Understanding the contributions of aerosol properties and parameterization discrepancies to droplet number variability in a global climate model, Atmos. Chem. Phys., 14(9), 4809–4826, doi:10.5194/acp-14-4809-2014, 2014a.

Morales Betancourt, R., and Nenes, A.: Aerosol Activation Parameterization: The population splitting concept revisited, Geosci.Mod.Dev., 7, 2345–2357, 2014b.

Sullivan, S.C., Lee, D., Oreopoulos, L., and Nenes, A (2016) The role of updraft velocity in temporal variability of cloud hydrometeor number, Proc.Nat.Acad.Sci., **113**, 21, 5781-5790, doi: 10.1073/pnas.1514043113