Review of **"Assessing the potential for simplification in global climate model cloud microphysics"**, by Proske et al. submitted to ACP.

General comments. This study used a perturbed parameter ensemble (PPE) framework together with Gaussian process emulation to explore sensitivity of a global climate model to perturbations of four ice microphysical processes. The paper first describes analysis of global results, then a spatial decomposition using spherical harmonics expansion. Overall results show strong sensitivity to the "aggregation" process, while riming has a strong impact on LWP. The authors then discuss implications for the level of complexity of microphysical process representations.

Overall, the paper is well written and addresses an important topic, and is within the scope of *Atmospheric Chemistry and Physics*. This study is interesting. The methodology appears to be sound. The demonstration of this PPE-based approach to investigate cloud microphysical processes as a "proof of concept" is the strongest part of the paper in my opinion. However, I do have several questions regarding the broader context, including the motivation for this approach and interpretation of results. Overall I think it could be acceptable in ACP, but more context is needed and some clarification of points raised below. These are detailed in "specific major comments". I also have a several additional minor comments and a few editorial corrections.

Overall recommendation: *Major revision*

Specific major comments.

1. I don't think "phasing" is the best term to use for the process perturbations. Essentially, you're applying a multiplicative factor directly to the process rates. But "phasing" typically means the relationship between the timing of two or more events, or synchronizing of multiple events. Similarly, the paper also uses the term "phase in and out" which implies some kind of dynamic or periodic change to the multiplicative factors, which is not the case here (they are constant in time and space for a given run, as I understand). Instead, process "perturbing" seems like a more appropriate term than "phasing". I don't think addressing this is required before publication, but I strongly suggest not calling this "phasing" to avoid confusion.

2. My main comment, as noted above, is in the interpretation and broader context of the study. The authors motivate this study by suggesting that results from this approach can be used to inform which processes can be simplified, but the specifics of this are unclear and the paper seems self-contradictory in several places. For example, the argument on lines 320-325 is: "Recognizing it as a threshold process and seeing the gradual response to small deviations from 1.0 in _aggr, it appears that there is potential for a less accurate description of aggregation in the model.", where "it" in this sentence means the "aggregation" process. I don't understand this argument. Moreover, this seems to directly contradict the argument later on lines 385-389 and lines 445-447. These lines state that the model is sensitive to representation of "aggregation" and riming, and therefore these processes "need to be represented accurately"

and should receive attention of model developers, while self-collection and accretion can be simplified. But then this is contradicted again on line 450: "The previous analysis shows that the response of ECHAM-HAM to an inhibition of self-collection or accretion is negligible, while for riming and aggregation at least a less accurate representation can be appropriate." Perhaps this is not what the authors intended to write and it's simply written incorrectly? Either way, this is confusing. See also lines 478-479 in the conclusions section, and lines 481-484, where the latter again argues that aggregation should be the "process of choice" to represent most accurately. Overall, these self-contradictory arguments make the paper feel somewhat incoherent, and this needs to be addressed before the paper can be accepted.

3. I have several other questions and comments about the broader motivation and context for this study. First, a process could still be important even if there is not much sensitivity to some limited set of output metrics in a given climate state (for example, a process might be important for cloud-aerosol interactions even if the impact on LWP, IWP, etc for the current climate state is small). This is alluded to on lines 507-508, but I feel this point should be emphasized more in the paper (e.g. in the introduction). Moreover, if a process is unimportant, it doesn't necessarily "do harm", except for extra computational cost. Computational cost here is negligible for these processes anyway, as discussed on lines 451-455 and Table 1. I appreciate the gain in interpretability with model simplification (stated explicitly on line 460), and this is an important point, but this should be made clear in the introduction when motivating the work. Also, regarding the sentence on line 460, I don't see how such simplification leads to a gain in "comprehensiveness", which I assume means generality; could you clarify this? Furthermore, regarding both computational cost and interpretability (as well as comprehensiveness), it seems likely that other scheme features, particularly the choice of prognostic variables and hydrometeor categories, is more important than process formulations anyway (see major comment #4 below for further discussion of this point).

4. The process formulations themselves only one aspect of scheme complexity, and arguably a rather small part. The choice of hydrometeor categories and prognostic variables is generally more important in terms of determining overall scheme complexity, interpretability, comprehensiveness, and computational cost. Yet nothing is mentioned about this in the paper. In my experience, simplifying the representation of microphysics (ice microphysics in particular) in terms of choice of prognostic variables and hydrometeor categories is likely to have a *much* larger effect on interpretability, cost, etc. than only simplifying the process formulations. To this end, can your approach inform simplification of these other aspects that are likely to be more important overall? As a specific example, one of the main findings of this paper is the importance of "aggregation" (which is probably better termed "ice autoconversion", see major comment #7 below). Yet in simpler schemes that use a single ice category but multiple predicted properties for that category (e.g., Morrison and Milbrandt 2015; Eidhammer et al. 2017) this process is not even included, nor are any other conversion processes between ice categories (e.g. accretion of cloud ice by snow). The critical point is that the separation of self-collection, aggregation, and accretion using separate cloud ice and snow categories is artificial, and the approach ultimately must rely on an ad-hoc choice for how cloud ice and snow are separated. Given that "aggregation" (autoconversion) is the most important

process (of the 4 tested) based on your results, it's troubling that there is little physical constraint for this process. To me, the findings from this study motivate the kind of simplification of using one ice category so that "aggregation" is no longer needed. I feel discussion of this issue is needed, particularly the approach of moving away from separate categories for cloud ice and snow. More broadly, discussion of these other sources of structural uncertainty (categories, prognostic variables, etc.), besides just the process formulations, is needed in the paper.

5. It's not clear what authors mean by aggregation as a "threshold process", since the response to eta_aggr is smooth and gradual (e.g. in Fig. 4). This "threshold" is mentioned many times throughout the paper. I assume this refers to the model not producing snow when eta = 0 for "aggregation", and thus riming and other snow processes also are zero. If so, please state this clearly in the paper the first time the threshold behavior is mentioned. I also question whether this should be called a threshold. I think of a "threshold" as some value that when crossed leads to a large, abrupt change. Is this really what happens? Turning off a process completely doesn't really seem to quality as a "threshold" from this point of view. Thus, my suggestion is to refer to this as something else besides a threshold.

A few other related comments. First, it's not entirely clear that snow should actually be zero when "aggregation" is turned off, if there are other processes which might contribute at least some to snow (see major comment #8 below). Second, this behavior really just reflects assumptions in the scheme, particularly the separation of total ice into cloud ice and snow categories (see major comment #4 above). For example, the shutting off of riming as aggregation goes to 0 must reflect the fact that riming of snow is included but riming of cloud ice is neglected (at least, I think riming of cloud ice is neglected but that isn't clear from the paper). Ultimately this kind of behavior does not necessarily reflect actual underlying cloud physics but rather the separation of ice into cloud ice and snow categories with an ad-hoc conversion between the two called "aggregation" (ice autoconversion).

6. I have a few comments on the process perturbations. The range for the process perturbations (magnitude of eta_i) seems ad-hoc, from 0 to 200%. Following the motivation of understanding sensitivity to process rates and the level of process uncertainty, it seems that the range of eta_i should follow from the level of uncertainty of a given process. However, turning off a process entirely (eta_i = 0) is unrealistic and generally not justified from a physical standpoint or from the standpoint of the actual range of uncertainty of a process. I understand why this might be done as a sensitivity test and "proof of concept" to assess importance of that process, but this point should be clarified in the paper. Also, what determined the upper range of 200%? Was this ad-hoc? Finally, a fairly major simplification of this study is assuming the process perturbations are constant in time and space. Of course, introducing time/space dependence would add more dimensions and increase complexity of the problem considerably, and therefore I can see why constant perturbations were assumed in this study for practical reasons. But from the standpoint of process uncertainty, it seems likely that relative uncertainty in these processes is not uniform in time and space. I suggest briefly mentioning this issue.

7. The term “aggregation” for conversion of cloud ice to snow by ice crystals aggregating with other crystals is confusing, and generally inconsistent with past literature on microphysics schemes. In other schemes this is referred to as “ice autoconversion”. “Aggregation” is confusing because “self collection” also occurs by the physical process of aggregation – the same physical process, except that particles remain within the cloud ice category. Note that some other ice autoconversion schemes also account for growth from cloud ice to snow via vapor diffusion (e.g., Harrington 1995), which can be an important process producing large ice particles. I’m assuming that’s neglected here, and the “aggregation” (or autoconversion) is only formulated via ice-ice collection? Or in some other way? Overall, I strongly recommend the authors refer to this process as “autoconversion” rather than “aggregation”.

8. It’s mentioned that “aggregation” (autoconversion) is the only process that can generate snow. Is this really true? What about freezing of rain? Does that form cloud ice, or what else is done?

9. You might consider briefly mentioning other emulation approaches which have some benefits (as well as some drawbacks) compared to using Gaussian processes for emulation as done here, specifically those based on neural networks. I believe some of these issues are outlined in Watson-Parris et al. (2021), and perhaps you could just mention the issue in a quick sentence or two and cite that paper (already cited elsewhere in your paper).

10. Figure 5. These plots must be showing the column-integrated process rates, but this isn’t stated anywhere. Otherwise the units don’t make sense.

11. There have been several studies (mainly authored by Posselt, van Lier-Walqui, and/or Morales) that have examined sensitivity to microphysical scheme perturbations in a Bayesian framework. These studies have used MCMC directly rather than emulation, which was possible by focusing on idealized modeling with reduced dimensionality (1D or 2D models generally, or small 3D domains). Some examples are Posselt (2016), Morales et al. (2021), He and Posselt (2015), and van Lier-Walqui et al. (2020) (references given at the end of this review). While these studies haven’t focused on global climate, they seem relevant to mention. In particular, the study of van Lier-Walqui et al. (2014) is relevant because they similarly applied constant multiplicative factors to individual microphysical process rates as a way to vary processes and examine the effects of process uncertainty (but in a 1D model, with MCMC).

Specific minor comments.

Line 7. “phasing of a process” isn’t clear and without context most readers will not understand what this means. As argued in major comment #1, I suggest you consider using different terminology than “phasing” altogether.

Line 27. This could give the false impression that Fisher and Koven (2020) concerns cloud microphysics in climate models specifically, when it’s about land surface processes. This is because you say "these processes" in the sentence before on line 25, which refers specifically to cloud microphysics. There's also a paper by Morrison et al. (2020) that discusses similar issues as on lines 23-35, but specifically from the standpoint of microphysics.

Line 66. I don't think that "stripped of detail" means "less accurate". Accuracy implies how close the model is to some benchmark or truth. Thus, my suggestion is to simply state "…where process parameterisations can be stripped of detail to aid the development of simplified models…".

Line 143. So riming of droplets on cloud ice is neglected?

Line 146. But snow number should not change at all for riming, of course. I assume that's the case for the parameterization here?

Line 147. So cloud ice sedimentation is neglected completely? Later on lines 181-182 it seems that sedimentation *is* included, but reword earlier around line 147 to make this clear.

Line 158. Assumed to be *all* liquid at mixed-phase temperatures? Also, better to specifically state the temperature range rather than saying "mixed-phase temperatures" which is imprecise.

Line 164. 15 micron volumetric radius seems very low for a maximum-allowed value. Is this a typo or is this actually the value used? I understand this is only for determining the minimum CDNC, but with such a small maximum radius this implies a rather large CDNC. I realize this isn't a central part of the study but it seems worth pointing out.

Lines 184-186. I think this sentence is confusing: "the gain of ice crystal concentrations in the level into which the ice crystals sediment is restricted to the closest loss of in-cloud ice crystal mass and number concentration in the levels above." Can this be clarified, perhaps using a more mathematical description which would be less confusing?

Line 204. I don't follow the logic that if the response to eta_i follows a sigmoidal curve, then the process need be only represented roughly. Do you mean because the relationship is monotonic? A sigmoidal curve of course could still have a sharp increase with small changes in eta_i implying a strong sensitivity to the given process. Or am I missing something here?

Figure 2. Perhaps mention PPE in the figure itself, not just the caption. i.e., the diagonal arrow could be labeled "Run ECHAM-HAM top generate the PPE" or something like that. Just a suggestion, the authors can take it or leave it.

Line 235. I assume "precipitation" here is surface precipitation?

Line 258-259. This seems like an important point, but could you be more specific what you mean by "disruptive changes" in the CMP processes compared to the changes in Johnson et al.

(2015)? Do you mean you use a larger range, including shutting processes off entirely by setting eta_i = 0? Or something else?

Line 304. This result doesn't seem that surprising – by suppressing "aggregation" you're likely decreasing the snow mixing ratio substantially, in turn reducing riming since the amount of riming depends on the amount of snow. This result likely reflects the somewhat artificiality of neglecting riming of cloud ice (at least, I think it's neglected in the scheme here), and more generally the separation of ice into "ice crystal" or "cloud ice" and "snow" categories (major comment #4). This kind of "thresholding" behavior associated with "aggregation" is discussed a bit below on line 309, but this is unphysical and reflects an ad-hoc separation of ice into cloud ice and snow categories. See major comment #5 above.

Lines 309-310. Snow is not generated by freezing of rain? In a climate model this will be tiny, but it should still be non-zero. Or else, how is rain formed/lofted above the freezing level treated?

Line 322. I disagree with the sentence: "In classical sensitivity studies, where processes are only turned on and off, only the large signal induced by aggregation would have been visible." Clearly there would still be a large signal to turning off the aggregation process completely. A few lines down this sentence also isn't clear: "Recognizing it as a threshold process and seeing the gradual response to small deviations from 1.0 in eta_aggr, it appears that there is potential for a less accurate description of aggregation in the model." To me, discussing a "threshold" process as a "gradual response" is self-contradictory. Not clear exactly what you mean by aggregation as a "threshold process".

Lines 343-345. I'm not sure if this comparison with Lohmann and Ferrachat (2010) is meaningful because here you're choosing an ad-hoc range to vary the eta_i's. And the range extends to unrealistic values, such as setting eta_i = 0 meaning the process is shut off completely. In contrast, I assume Lohmann and Ferrachat (2010) varied scheme parameters over a range of physically plausible values?

Line 346. It's not clear what you mean by varying the autoconversion rate between 1 and 10. Autoconversion rate should have units. Do you mean you vary it by a factor from 1 to 10?

Line 400. I would suggest using a different variable than "f" for the data and complex coefficients. It's confusing to use the same symbol "f" for both.

Line 465. Not sure I'd say you're multiplying the process "effects" by a factor, but rather you're multiplying the process rates.

Editorial comments.

Lines 236-237. Would this sentence be better as: "For the kernel, an additive combination of the linear, polynomial, bias and exponential kernels was used (Duvenaud, 2014)."?

Line 240. I think “perturb” should be “perturbed”, past tense.

Line 270. I think you can remove “thus”.

Line 356. “verion” is misspelled.

Line 396. I think there should be a comma after “decomposition”.

Line 513. It seems “process” should be “processes”.

References.

Eidhammer, T., H. Morrison, D. Mitchell, A. Gettelman, and E. Erfani, 2017: Improvements in global climate model microphysics using a consistent representation of ice particle properties. *J. Climate*, 30, 609-629.

Harrington, J. Y., M. P. Meyers, R. L. Walko, and W. R. Cotton, 1995: Parameterization of ice crystal conversion process due to vapor deposition for mesoscale models using double-moment basis functions. Part I: Basic formulation and parcel model results. *J. Atmos. Sci.*, 52, 4344–4366.

He, F., and D. J. Posselt, 2015: Impact of parameterized physical processes on simulated tropical cyclone characteristics in the Community Atmosphere Model. *J. Climate*, 28, 9857-9872.

Morrison, H., and J. A. Milbrandt, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, 72, 287-311.

Morrison, H., M. van Lier-Walqui, A. M. Fridlind, W. W. Grabowski, J. Y. Harrington, C. Hoose, A. Korolev, M. R. Kumjian, J. A. Milbrandt, H. Pawlowska, D. J. Posselt, O. P. Prat, K. J. Reimel, S.-I. Shima, B. van Diedenhoven, and L. Xue, 2020: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Mod. Earth Sys.*, e2019MS0016898, doi:10.1029/2019MS001689.

Morales, A., D. J. Posselt, and H. Morrison, 2021: Which combinations of environmental conditions and microphysical parameter values produce a given orographic precipitation distribution? *J. Atmos. Sci.*, 78, 619-638.

Posselt, D. J., 2016: A Bayesian examination of deep convective squall-line sensitivity to changes in cloud microphysical parameters*. J. Atmos. Sci*., 73, 637-665.

van Lier-Walqui, M., H. Morrison, M. R. Kumjian, K. J. Reimel, O. P. Prat, S. Lunderman, and M. Morzfeld, 2020: A Bayesian approach for statistical-physical bulk parameterization of rain microphysics, Part II: Idealized Markov chain Monte Carlo experiments. *J. Atmos. Sci.*, 77, 1043-1064.

van Lier-Walqui, M., T. Vukicevic, and D. J. Posselt, 2014: Linearization of microphysical parameterization uncertainty using multiplicative process perturbation parameters. *Mon. Wea. Rev.*, 142, 401-413.