Reviewer's report on "Climate SPHINX: evaluating the impact of resolution and stochastic and stochastic physics parameterisations in climate simulation" by Davini et al. Submitted to Geoph. Model Development.

This article reports on a massive experiment using a large number of climate simulations at various resolution with the aim of comparing the performance impact of resolution and the use of stochastic physics. As both these issues are at the heart of ongoing climate modelling research worldwide, the publication of its outcomes is very welcome and will likely benefit this community. However, in its present form the paper is more of an internal report than a research publication. It lacks global insight and fair assessment with regards to other peoples' work and an effort to make an universal piece of work is suggested. While it is clear that the work is done with one specific climate model and using one specific stochastic scheme (rather two), the tune of the presentation makes it feel like this will be the case for all situations regardless of the precise climate model used or even the particular stochastic physics used. In addition, the analysis has a lot flaws and many steps lack scientific rigour.
I thus cannot recommend this paper for publication in its present form. I would advice for reject and resubmit, if that's an option.

Major comments:

1. The first misleading thing is in fact the title. It needs to be changed. Instead of "in climate simulation" at the end, I suggest to use "in the ECMWF model climate simulations".
2. Given that all dynamical cores are perhaps equivalent to some extent, since they are based on a controlled discretization of somewhat the same governing equations, there is hope that the resolution sensitivity part of this work might extrapolate well to other climate models. However, the stochastic physics schemes considered are very particular and arguably questionable since they are essentially simply some black box random number generators that are by no means have any kind of physical justification and/or empirical validation (to the best of the reviewer's knowledge), though to their creators credit, they are the first to appear as such and led the path to stochastic parametrisations to flourish. Now days, there are many stochastic physics schemes that are developed from first principles and in most cases validated (or rather calibrated) using simulation and/or observation data, which actual achieve performances that are much better than what the pioneering ECMWF stochastic physics have achieved, in terms of improving climate simulations. To name a few, I list a couple of published articles that merit citation in this article because the readers of GMD deserve it.
   - Ajayamohan, R. S., B. Khouider, A. J. Majda, and Q. Deng (2016), Role of stratiform heating on the organization of convection over the monsoon trough, Clim. Dyn., pp. 284 1–20, doi:10.1007/s00382-016-3033-7
   - De La Chevroti`ere, M., Mich`ele, B. Khouider, and A. Majda (2015), Stochasticity of convection in giga-les data, Clim. Dyn., doi:doi:10.1007/s00382-015-2936-z
   - Khouider, B., J. Biello, and A. J. Majda (2010), A stochastic multicloud model for tropical convection, Commun. Math. Sci., 8 (1), 187–216
   - Peters, K., C. Jakob, L. Davies, B. Khouider, and A. J. Majda (2013), Stochastic Behavior of Tropical Convection in Observations and a Multicloud Model, J. Atmos. Sci., 70 (11), 370 3556–3575, doi:10.1175/JAS-D-13-031.1.
   - Plant, R. S., and G. C. Craig (2008), A stochastic parameterization for deep

convection based on equilibrium statistics, J. Atmos. Sci., 65, 87?105, doi: 373 doi:10.1175/2007JAS2263.1

- Wang, Y., G. J. Zhang, and G. C. Craig (2016), Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5, Geophys. Res. Lett., 43 (12), 6612–6619, doi:10.1002/2016GL069818, 2016GL069818
- Deng, Q., B. Khouider, and A. J. Majda (2015), The MJO in a Coarse-Resolution 294 GCM with a Stochastic Multicloud Parameterization, J. Atmos. Sci., 72 (1), 55–74, 295 doi:10.1175/JAS-D-14-0120.1.
- Deng, Q., B. Khouider, A. J. Majda, and R. S. Ajayamohan (2016), Effect of Stratiform 297 Heating on the Planetary-Scale Organization of Tropical Convection, J. Atmos. Sci., 73 (1), 371–392, doi:10.1175/JAS-D-15-0178.1
- Jesse Dorrestijn, Daan T. Crommelin, A. Pier Siebesma, Harmen J. J. Jonker, and Frank Selten, 2016: Stochastic Convection Parameterization with Markov Chains in an Intermediate-Complexity GCM. J. Atmos. Sci., 73, 1367–1382, doi: 10.1175/JAS-D-15-0244.1.
- Sakradzija, M., A. Seifert, and A. Dipankar (2016), A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone, J. Adv. Model. Earth Syst.,8, 786–812, doi:10.1002/2016MS000634.
- G. A. Grell and S. R. Freitas: A scale and aerosol aware stochastic convective parameterization. Atmos. Chem. Phys., 14, 5233–5250, 2014 www.atmos-chem-phys.net/14/5233/2014/doi:10.5194/acp-14-5233-2014
- Frenkel, Y., A. J. Majda, and B. Khouider (2012), Using the Stochastic Multicloud Model to Improve Tropical Convective Parameterization: A Paradigm Example, J. Atmos. Sci., 69 (3), 1080–1105, doi:10.1175/JAS-D-11-0148.1.

3. One important point that needs to be specifically addressed is the question of what would be the intake of more carefully designed stochastic physics schemes in the context of the present study?
4. Even in the precise case of the present study, I was hoping to see some simulations where the two stochastic schemes (SPPT v.s. SKEB) were taking apart separately instead of considering them jointly. For instance which of these two leads to the improvements at low resolution and which one is causing the degradation at high resolution.
5. Page 2, Line 22: " a cheaper alternative is to use stochastic parameterisation schemes". This statement is probably true if one is ought to use physically based stochastic schemes as demonstrated in papers from the list provided above. In the case of the scheme considered here, the original purpose was simply to increase the spread in ensemble prediction runs. The idea of them (SPPT and/or SKEB) competing with high resolution models seems to be new to this article. In the reviewer's opinion, these particular schemes are not at all the best suited for the job.
6. Page 3, Line 30: One important pre-requisite to using a stochastic parameterization under grid refinement is that the scheme has scale awareness. However, this issue is completely ignored in this paper except for some minor remark in the conclusion section on the necessity of retuning the stochastic physics at high resolution---it is unfortunate that this much computing time was wasted in brute force computations to arrive toward an obvious conclusion as this one, especially given that the literature about this issue exists (see previous list for a sample)
7. Page 9, Lines 10-13: "unrealistic" in what sense? This is not intuitive at all as one would expect in terms of climate variability. It would probably make sense if we are dealing with an initial value

problem. The fact that the member with smallest RMSE in sea-ice is chosen seems like an ad hoc choice and cannot possibly be a non biased "representative" of the ensemble. Thus the problem. What would one do if the RMSE is not accessible?

Other Comments and typos:

1. Page 1, Line 18: "prediction of climate". Which climate?
2. Page 2, Line 12: "as the jet stream" ... ", the Euro Atlantic". Insert "the" in both places.
3. Page 3, Line 8: which stochastic physics
4. Page 5, Lines 8-9: Are the horizontal correlation lengths and time decorrelation parameters and standard deviations etc physically justifiable? What are the motivations behind?
5. Page 5, Line 22: Are the SPPT and SKEB scheme activated and deactivated simultaneously? Why not also consider cases when only one or the other is activated? (See major comment 4 above).
6. Page 6, Sentence in Lines 27-28: How much impact does this variation in the location of the Gaussian peak has on the simulations? Are we sure we are still comparing apple to apple? What are the reason behind this difference?
7. Page 7, Line 5: "more"-->"more than"; "the majority of them were carried"
8. Page 7, line 20: Do all ensemble members use the same initial data?
9. Page 7, Line 31: What is the RCP8.5 scenario? Provide a reference.
10. Page 8, Line 7: Avoid acronyms in section title, especially when they are not previously defined.
11. Page 9, Line 6: which reconstruction?
12. Page 9, Line 27: explain in what sense the SST pattern is physically consistent with the sea-ice pattern.
13. Page 11, Line 11: "number ... is extremely dispersive"? Could say "Climate aspects are diverse" or "number of diverse climate aspects is large".
14. Page 11, Line 13: "in terms of variability rather than in terms of mean state". This can come at this expense of a hugely deteriorated mean state. Without a proper control of the model's mean state, the sole analysis of the variability by itself is not acceptable.
15. Figure 5, add legends in panels b) and d).
16. Page 11, Line 27: "occur too infrequently ... both observational datasets". This is simply not true. T799&T1279 seem to be as frequent as GPCP at for rain rates larger than 50 mm/day but much less compared to TRRM.
17. Page 12, Line 16-17: How can the stochastic scheme reduce the conversion of vapour to rain. From the description in Section 2, SSPT doesn't care how the parameterization acts on the water vapour field all it does is amplify or reduce the water vapour sink or source according to the random realization of the number xi.
18. Page 12, Line 15: In what sense Raymond et al. 2015 is an adequate reference for MJO studies.?
19. Wheeler and Hendon Index is not at all the proper metric for measuring the "dynamics and thermodynamics" of the MJO. Why not do a Wheeler Kiladis spectral diagram, Hovemuller diagrams, and zonal and vertical structure. These are more more representative metrics for an MJO. The RMM index can be very misleading, see Katherine H. Straub, 2013: MJO Initiation in the Real-Time Multivariate MJO Index. *J. Climate,* **26**, 1130–1151, doi: 10.1175/JCLI-D-12-00074.1