Response to RC3

The authors have previously performed experiments with the LETKF and a mediumcomplexity AGCM. Previous studies have focused on the effect of localisation. In the present study the authors focus on the Non-Gaussian features of the distribution of certain variables.

I think this is an interesting topic, and the manuscript can indeed become a useful contribution to the community. To be fair, just the fact of having a 1024-member ensemble is a very rich source of information. The manuscript is in general quite clear and well-written.

I have the following comments with respect to the manuscript.

Response: We are really grateful to the referee for the careful review and constructive suggestions.

1. First of all, it is not clear to me how many of the experiments were performed in SPEEDY and how many of the experiments were performed in NICAM. The paper is mainly focused on SPEEDY and the experiments in NICAM are only mentioned in passing in the very last section of the paper. Is this because they were fewer and/or less detailed? Or were exactly the same experiments done in both models? If so, why did you choose SPEEDY?

Response: Although a single experiment with SPEEDY was used in the previous manuscript, we performed an additional experiment to investigate the sensitivity of non-Gaussianity to density of observations following the comment from another referee. We improved section 3 to clarify that an experiment without localization was chosen among five 10240-member experiments in KM16 (ll. 196-199). With NICAM, a single experiment was performed for just one week to get a rough idea of the realistic model's behavior. We revised accordingly to clarify these points (ll. 416-420).

2. I agree with the other reviewers in the sense that more diagnostic metrics need to be considered besides RMSE of the analysis mean. These include: rank histograms, reliability metrics and scores. Also, the relationship between RMSE and spread can be quantified to check for the 'health' of the ensemble system (this is related to the rank histograms too).

Response: We agree. We added rank histograms and continuous ranked probability score (CRPS) at all grid points and the grid points with non-Gaussian PDF in section 4, and discussed them in section 5. Also, we added the pattern correlation between the RMSE and ensemble spread in section 4.

3. I was very interested to see how clustering can occur in these models, and particularly excited about the fact that the set of outliers contained more than one element! In previous works (Amezcua et al, 2012; Anderson 2010) only one outlier was found. I wonder if you could say more about the size of these sets, and the way these outliers appear and disappear as the system

evolves. I guess this has to do with the distribution of maximum and minimum of a sample depending of both (a) the parent distribution and (b) the sample size. Is there anything you can mention about the distribution of the sample maxima?

Response: The outliers randomly appear and seldom affect the analysis accuracy in this study. With the more realistic models and real observations, we do not know how often the outliers are generated and how much they affect the analysis accuracy. Therefore, we would like to investigate them as a future work, and the following sentences were added (ll. 443-445): “These are the results from the simple SPEEDY model. It remains to be a subject of future research how the outliers behave with a more realistic model and real observations.”

4. It is natural to study skewness, kurtosis and KLD as the authors have done. This has been done in a univariate manner only. Can a multivariate analysis be performed? There are some ways to compute higher order moments for multivariate distributions (e.g. Mardia 1970). I am not asking to compute these if it is too hard to do it, but I am wondering if anything could be gained.

Response: We agree. This is the first study to evaluate the non-Gaussianity with huge ensemble size using atmospheric models, so that we evaluate the ensemble in a univariate field. The multivariate analysis would provide more information, and we would like to investigate it as a future work, and the following sentences were added (ll. 453-455): “The measures of non-Gaussianity are evaluated in the univariate field in this study. An extension to multivariate fields with multivariate analysis is remained as a subject of future research.”

5. The LOF method is slightly confusing and I thank the authors for adding the figure to explain it, but I wonder if there is any way to make it even clearer.

Response: The LOF method is complex. We added some additional descriptions of the LOF method in detail using Fig.3 in section 3 (ll. 141-146).

6. SPEEDY is a very "simple" (not in a bad way) model with very low resolution in time and space. It is not surprising, hence, that the source of non-Gaussianity is the parameterisation related to rain. Still, I think the analysis (including the two figures) is important. However, it should be emphasised this conclusion in valid from SPEEDY. Small scale physical processes can generate non-Gaussianity, but they are not represented in SPEEDY. I wonder if you can say anything about this with the results from NICAM?

Response: Following the suggestion, we emphasized that the results were based on the SPEEDY-LETKF system. In addition, we added a case of non-Gaussianity genesis from the advection in the extratropics in section 4 and discussed it in section 5. Also, we added the following sentence (ll. 426-428): “This result implies that the NICAM has various sources of non-Gaussianity such

as smaller scale physical and dynamical processes with various interactions among different model variables,"

7. In the Ensemble Kalman Filter/Smoother, one can separate the sampling effect as having two parts, which can be approximated as being additive (Sacher and Bartello 2008; Amezcua and van Leeuwen, 2018). The more indirect part comes from the gain K coming from the sample. It is a nonlinear function of the sample covariance: bˆ2/(bˆ2+rˆ2) in the univariate case. I wouldl like to know more about the quality of K as the ensemble size changes. Note that this is related to the quality of sample B, but the 'convergence' to the true K may be slower do to the nonlinearity of the relationship.

Response: Kondo and Miyoshi (2016, KM16) investigated the analysis quality of SPEEDY model by changing the ensemble size and localization, and indicated that the analysis accuracy is improved as the ensemble size is increased with broader-scale localizations. Also, Miyoshi et al. (2014, MKI14) showed that the sampling errors were reduced from error correlations based on the B by increasing the ensemble size from 20 to 10240. Therefore, we focus on the non-Gaussianity in this study although it is important to investigate the qualities of B or K. The following sentences were added in section 1 (ll. 47-48, 65-67): "and found meaningful long-range error correlations. In addition, they reported that sampling errors in the error correlation were reduced by increasing the ensemble size.", and "Using the precious dataset of KM16 with 10240 ensemble members, we can make various investigations such as non-Gaussian statistics and sampling errors in the background error covariance. Here we focus on the non-Gaussian statistics in this study."

8. A very simple comment about figure 5: I think it would be easier to visualise if the y-axis had logarithmic scale.

Response: Logarithmic scale y-axis is not appropriate for LOF because we use three thresholds of LOF value = 5.0, 8.0, 11.0.