Response to RC2

This manuscript presents some fascinating and novel analysis of the distributions generated by an ensemble assimilation system and a low-order atmospheric model using ensembles of unprecedented size. It is generally well-written and presents a series of results that expose a number of novel aspects of the problem. It is a great addition to the ensemble literature and motivates the need for non-gaussian DA procedures rather than just using ever larger Kalman-class ensemble filters. The only weakness is that, in some places, the authors may be a bit too bold in extrapolating their results. What they have documented is the distributions from their system using a simple model and an LETKF. It is unclear whether the non-gaussian aspects of the resulting distributions are fundamental to the observing system and model, or if they might be very different with a more general data assimilation system. As an example, it is not clear if a non-gaussian DA system would result in more (or less) gaussian distributions if applied with the same nature run and observations. The authors do provide hints that the low-order model may not be a very good proxy by showing a few results with a more realistic model which look very different from the results for the SPEEDY model used here. Looks like lots of room for additional interesting studies in the future.

Response: We are really grateful to the referee for the careful review and constructive suggestions. Following the suggestions, we revised the manuscript to emphasize that the results were based on the SPEEDY-LETKF system in section 5, and to discuss the dependence of non-Gaussianity on observing systems (ll. 402-409).

Specific comments:

1. Line 29: "disappear 'naturally"' Not clear what it means to be 'natural'. Do you just mean that they disappear after a while? Are there any places where things disappear 'unnaturally' in contrast?

Response: The phrase "disappear naturally" was removed.

2. Line 30: "1000 ensemble members may be necessary…" This is a tricky argument. Maybe it takes many fewer if the DA system isn't assuming and enforcing (to some extent) gaussianity for increments. Or maybe there is a whole bunch more 'detailed' structure out there in that case.

Response: Following the suggestion, the sentence was revised accordingly (ll. 29-33): "Sensitivity to the ensemble size suggests that approximately 1000 ensemble members be necessary in the intermediate AGCM-LETKF system to represent the detailed structures of the non-Gaussian PDF such as skewness and kurtosis; the higher-order non-Gaussian statistics are more vulnerable to the sampling errors due to a smaller ensemble size." In addition, the sentence was added (ll. 76-78): "This study also discusses how many ensemble members are necessary to represent non-Gaussian PDF without contaminated by the sampling error, since in general higher-order non-Gaussian statistics are more vulnerable to the sampling error due to a limited ensemble size."

3. Line 34: "find the optimal initial state" Definition is too limited. Goal is to find some representation of a pdf. A subcase would be an optimal state.

Response: Following the suggestion, the sentence was revised (ll. 35-37). "Data assimilation is a statistical approach to estimate a posterior probability density function (PDF) using information of a prior PDF and observations. Based on the posterior PDF estimate, the optimal initial state is given for numerical weather prediction (NWP)"

4. Line 53: Actually Anderson shows that outliers occur in WEAKLY nonlinear situations, not in strongly nonlinear cases.

Response: Revised as suggested (l. 57).

5. Line 62: "non-Gaussianity will degrade the analysis." Compared to what?

Response: The sentence was revised as follows (ll. 67-69): "Since the Gaussian assumption makes the minimum variance estimator of the EnKF coincide with the maximum likelihood estimator, the non-Gaussian PDF may bring some negative impacts on the LETKF analysis."

6. Line 82: There is an issue with your definition of kurtosis throughout. The standard definition of kurtosis (check wolfram, Wikipedia) has a value of 3 for a normal distribution and the term beta_2 is usually used for this. The excess Kurtosis is this value minus 3 and has a value of 0 for normal distribution. You should make sure that both your symbols and definition are clear throughout.

Response: Following the suggestion, we replaced the kurtosis with sample excess kurtosis (eq. 2).

7. Line 91: "the PDF is considered to be non-Gaussian". This is a magic number here. Give some insight on why you picked it or make it clear that this will be discussed below.

Response: We added the following sentences (ll. 100-104): "The histogram with KL divergence $D_{KL} = 0.01$ looks approximately Gaussian while the other three histograms with larger D_{KL} values show significant discrepancies from the Gaussian function. The skewness and kurtosis measure the degrees of symmetry and tailedness, respectively, while the KL divergence D_{KL} is more suitable for measuring the degrees of difference between a given PDF and the fitted Gaussian function."

8. Line 125: k = 20 is another magic number. Give some insight into why you picked it.

Response: The sentences about choosing *k* were revised as follows (ll. 151-155): "Breunig et al. (2000) suggested that choosing *k* from 10 to 20 work well for most of the datasets. If we choose *k* too small, some objects deeply inside a cluster have a large *LOF*, and the LOF method does not work. In fact, using the dataset of KM16, *k* = 10 showed this problem, while *k* = 20 did not. Therefore, we chose *k* = 20 in this study. Similar to the SD method, the LOF method is applied to a one-dimensional dataset consisted of 10240 ensemble members.

9. Line 157: "No localization was applied, yielding the best analysis accuracy." Do you know this to be the case? If so, state explicitly what localizations you tried (in the previous work I assume). It is surprising that no localization would be optimal since localization can also protect against nonlinear relations which are certainly occurring in the presence of non-gaussian marginal distributions that you are examining.

Response: We added the following phrase (ll. 196-199): "No localization was applied, yielding the best analysis accuracy as shown by KM16 who compared five 10240-member experiments with different choices of localization: step functions with 2000-km, 4000-km and 7303-km localization radii, a Gaussian function with a 7303-km localization radius, and no localization."

10. Line 181: "extremely large" Compared to what?

Response: Following the suggestion, the sentence was revised (l. 221-222): "At grid point B, although the PDF seems to be closer to Gaussian, skewness $\beta_1^{1/2}$ and kurtosis β_2 are much larger than those at grid point A."

11. Line 230: "their instability is mitigated in the model." I had trouble with this sentence which seemed to say that the model generated instability and mitigated it at the same time. Just needs clarification.

Response: We apologize for our mistake. The word "unstable" was wrong and was replaced with "stable" (l. 272).

12. Line 250 and elsewhere: "propagates" I think that this may be a poor word choice for how marginal non-gaussianity is generated.

Response: Revised accordingly (ll. 295, 412).

13. Line 286: "Gaussian filters cannot produce accurate analysis when significant non-Gaussianity exists." I am uncomfortable with this statement although I hope it is qualitatively

accurate. Still, I don't think you have documented it. Kalman filters are still optimal in certain senses even for non-gaussian priors. You have provided no evidence that a non-gaussian filter would produce significantly different analyses in cases with significant non-gaussian priors. I think this statement should have a lot of caveats and suggest the need for more research.
Response: We agree and removed the sentence.

14. Line 288: "non-gaussianity of the atmosphere" First, the atmosphere doesn't have a PDF, so it can't be non-gaussian (caveat classical physics). Second, you have used such a simple model that it is difficult to say too much about more realistic models with similar observing systems, so be careful to not be too strong here.
Response: We removed "of the atmosphere". As shown in the previous responses, we emphasized that the results were based on the SPEEDY-LETKF system in section 5.

15. Line 294: At least in some cases, "non-Gaussianity is explained by the convective instability." However, your results from the more realistic model suggest it is not the only cause, and your study here only looks at a few points in detail so cannot provide strong conclusions. In addition, it is easy to demonstrate that any advective problem will generate non-gaussian distributions when the advecting flow is uncertain and the quantity being advected has gradients. This is certainly introducing non-gaussianity in all atmospheric models. See also line 302 which I think is too strong of a statement.
Response: We agree. The sentences were revised accordingly (ll. 373-374, 382-383): "With the SPEEDY model, the genesis of non-Gaussian PDF in the tropics is mainly associated with the convective instability.", and "Therefore, convective instability is a key to non-Gaussianity genesis in the tropics in the SPEEDY model." We also added a case of non-Gaussianity genesis from the advection in the extratropics in section 4.

16. Line 339: I believe these results are actually consistent with the results in Anderson (2010). The last section of that paper looks at results in a low-order atmospheric model, similar to SPEEDY but dry, and finds that outliers occur at local points (not for all state variables), that they form and then normally quickly disappear, and that sometimes multiple outliers can occur at the same point. Even for Lorenz-96, outliers do not form for all variables simultaneously. For instance, ensemble member 1 may be an outlier for state variable 2, while member 10 is an outlier for state variable 12.
Response: We removed "Anderson (2010)" from the sentence (l. 434) and added the following sentence (ll. 433-434): "Anderson (2010) also reported similar results using a low-order dry atmospheric model."