Nonlin. Processes Geophys. Discuss., 2, C565–C573, 2015
www.nonlin-processes-geophys-discuss.net/2/C565/2015/

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Identifying non-normal and lognormal characteristics of temperature, mixing ratio, surface pressure, and wind for data assimilation systems" *by* A. J. Kliewer et al.

A. J. Kliewer et al.

anton.kliewer@colostate.edu

Received and published: 7 December 2015

Anonymous Referee #2 Received and published: 26 October 2015 The authors study the probability distribution of some parameters of the GFS model forecast (mixing ratio, temperature, wind field). They try to combine different tests to detect whether a probability distribution is non-normal and lognormal in particular. Finally they discuss where and when the GFS forecast follows a non-normal distribution. The main objective is to provide this information so that a data assimilation system can adapt and uses assimilation schemes specifically for lognormal distributions. Instead of using a single test, they combine the results of three statistical tests (Shapiro-Wilk, Jarque-Bera, a Chiˆ2-

test) together to decide whether a variable at a given location is lognormal distributed.

Unfortunately, I cannot recommend this manuscript for publication in Nonlinear Processes Geophysics. I have the following major comments about this manuscript, essentially about the level of innovation, the robustness of the results and about the way the conclusions are supported by the statistical tests:

AC: The authors appreciate the reviewer's comments and opinion. Please consider our responses to the previous points you have raised. Hopefully they have been adequately addressed and the reviewer will now view our manuscript suitable for publication in Nonlinear Processes Geophysics. Author comments begin with "AC: ."

1. First of all, it is unclear to me if by combining the results of three standard statistical tests together in a single test, the authors reach a sufficient level of innovation to justify a publication in Nonlinear Processes Geophysics. The scientific advancement of this manuscript is not on a par with what I usually see in manuscripts published in this journal (and other journals with a similar scope for that matter).

AC: The goal of this research is develop some initial testing procedures to determine the probability distribution of random variables pertinent to data assimilation systems. Since each test that is included in the composite test evaluates different features of a sample, the goal was to exploit the capabilities of current and reliable hypothesis testing procedures. Given the vast amounts of data that is available, along with the proven potential of a mixed Gaussian-lognormal data assimilation system (Kliewer et al. 2015, Fletcher and Jones 2014), the composite test is utilized as a novel way to identify locations where the histogram of the data clearly favor a lognormal distribution as opposed to a Gaussian fit. The creation of new hypothesis test entirely, along with synthetic testing and evaluation, is beyond the scope of what we intend to accomplish; the use of the composite test here is merely a means to an end. Future research could include the development of an online testing procedure with the composite test serving as a validation metric or running offline between cycles.

The authors affirm that the conclusions brought forth by utilizing the composite test are of value to the scientific community. It is a simple testing procedure capable of identifying non-Gaussian and lognormal distributions in atmospheric random variables which can be used to determine an optimal data assimilation scheme to reduce retrieved errors (Kliewer et al. 2015, Fletcher and Jones 2014).

2. It is also unclear how much better the new approach works. The authors do not attempt to determine the probability of the type I or type II error of the combined test. Further work could be done either on a theoretical level or with synthetic data with known distribution. This leaves the reader with the question, if it is really better to use 3 tests instead of one and by how much is it better?

AC: Since the result of a hypothesis test is either "Reject" or "Fail to Reject" the null hypothesis, the composite test was designed to capture test agreement. We acknowledge that statistics can be skewed by outliers, sample sizes, methodology, and a host of other factors. To address these issues different hypothesis tests have been proposed in order to reach a binary conclusion. Since each test is designed to evaluate a different characteristic of the data set, the authors combined a few of the more powerful tests (Razali and Wah 2011) to be able to quickly identify locations where the time series appears to come from a lognormal distribution. The presented histograms convey the quality of the composite test results. This research is hopefully the jumping-off point for further investigation into the temporal and spatial probability distributions for variables used in data assimilation.

3. The question where a variable follows a Gaussian distribution and how strong the non-Gaussian behavior is are actually two different questions. The statistical tests (in particular Shapiro-Wilk, Jarque-Bera) are designed to answer the first one. But for variables which are always larger than a given threshold value, we already know that these variables do not follow a Gaussian distribution. If the sampling size could be made arbitrarily large, then one would end up rejecting always the null hypothesis (rejecting the thesis that a variable follows a Gaussian distribution), for the mixing ratio

in particular (since it is always positive). It should have been clear from the start that the mixing ratio does not follow a Gaussian distribution. The authors present the results of the test for a data set spanning one year. The presented maps are not robust if the data from increasingly long time spans were considered as the normality test would tend to reject more often the null hypothesis (simply because the probability to have negative mixing ratio is exactly zero and the pdf of a Gaussian distribution is always strictly positive). Another way to look at this is to assume that a variable follows a distribution which is close to a Gaussian distribution but has a very small skewness. Since the skewness is not exactly zero, for a sufficiently larger sampling size n (equation 8), the Jarque-Bera test will end up to rejecting the null hypothesis. The question should rather be whether the deviation from Gaussianity is too large to apply a assimilation scheme like the Kalman Filter of 4D-Var.

AC: The composite test defined by the three hypothesis tests have a significance level of $\alpha = 0.01$. This stringent condition is meant to identify the most clearly non-Gaussian distributions found in the time series. The significance level is always a matter of personal choice and system tuning. All of the hypothesis tests include the sample size as a parameter, and as such aid in determining whether or not the null hypothesis is rejected. Due to land topography, jet streams, the water cycle, El Nino/La Nina effects, etc., the authors do not necessarily believe that given a large enough climatology, all positive definite atmospheric variables, at all vertical pressure levels across the globe will exhibit a non-Gaussian structure. It is proposed that this assumption be checked systematically and for this the composite test is shown to be a valid metric.

In Fletcher (2010), histograms of column water vapor when a boundary cloud is present, containing roughly 4000 points, taken from the Atmospheric Radiation Measurement in Oklahoma show a seasonal dependence. Here, a lognormal distribution better approximates the data in the winter months as opposed to a Gaussian fit in the summer. So it is shown that a large sample of a positive definite random variable can be best approximated by a Gaussian distribution for certain seasons.

Moreover, the results in this manuscript show that the distribution for mixing ratio and temperature can follow a Gaussian, lognormal, or bi-modal distribution depending on the season and vertical pressure level. Therefore it is not always the case that a positive definite random variable cannot be approximated by a Gaussian distribution given a large enough sample.

The year 2005 was chosen because of previously found evidence of a lognormal distribution in water vapor mixing ratio, and the "seasonal" time domains were evaluated to determine any changing dynamics of the observed variables. In fact, several examples of this type of phenomenon have been demonstrated with histograms at individual points. One year of data is sufficient for each of the hypothesis tests that have been used, and they do not all indicate non-Gaussian behavior because of the magnitude of n. Future work can explore multiyear records with inter-annual oscillations as this research is meant to demonstrate the implementation of the tests.

4. Data assimilation is cited at several places as the application for this study where an assimilation scheme could change depending on the distribution of a variable at a given location and time. However, it is not clear to me if it is really feasible to change the assimilation schemes from normal to lognormal distributed variables from one grid point to the next. This would lead to unrealistic discontinuities at the transition zones. A clearer view on how the results of the detection method will be used later during the assimilation is necessary in my opinion in order to ensure that the results can be used.

AC: The authors would like to emphasize that the composite test is not proposed to be simply inserted into a real-time data assimilation system as is in order to improve weather forecasting. The composite test was constructed to leverage existing, robust, statistical testing procedures and report test agreement. As noted in the original manuscript, these testing procedures can be used offline between cycles or windows (page 1368 line 10). This initial research will hopefully be the precursor and quality assessment to an online, spatial and temporal probability estimation and detection system.

The reviewer raises an interesting point in considering how a transition zone between distributions must be addressed. This is something that will certainly have to be investigated in a real-time system.

Knowing the correct probability distribution of an atmospheric random variable has been shown to impact the quality of a data assimilation analysis. In Kliewer et. al. (2015), it is shown that a lognormal-based DA scheme outperforms a Gaussian system for lognormally-distributed random variables and is now referenced in the manuscript. The following text is now included in the introduction of the manuscript:

"In Kliewer (2015), a 1D-VAR Gaussian-based retrieval system is updated to include a mixed Gaussian-lognormal cost function for lognormally-distributed random variables. Synthetic, lognormally-distributed random numbers are generated to represent the true state for water vapor mixing ratio and the lognormal-based retrieval scheme shows to have substantially smaller errors than the Gaussian system. Observations from September 2005 from the Advanced Microwave Sounding Unit (AMSU) A and B are also used to retrieve mixing ratio and temperature estimates through a mixed Gaussian-lognormal based retrieval system and are seen to have smaller final innovations and are more comparable to the Microwave Surface and Precipitation Products System (MSPPS) than the Gaussian system."

Kliewer, A.J., S.J. Fletcher, A.S. Jones and J.M. Forsythe, 2015: Comparison of Gaussian, Logarithmic Transform and Mixed Gaussian-Lognormal Distribution-based 1DVAR Microwave Temperature-Water Vapor Mixing Ratio Retrievals, Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.2651.

In the paragraph following this new reference, this paper is again referenced, emphasizing that a mixed Gaussian-Lognormal data assimilation scheme can be used and has shown to offer an improvement over a Gaussian-only system.

5. When the null hypothesis cannot be rejected, one cannot conclude that the null hypothesis is true (see also below). This assumption is made at several places within

the manuscript, especially when deciding if a variable is lognormally distributed or not.

AC: Yes, the authors completely agree with this comment and this was outlined in the Hypotheses section of the original manuscript. Unfortunately, a few subsequent conclusions deviated from this framework and have since been corrected. The current manuscript now claims a variable is lognormally-distributed only if a histogram clearly represents this as opposed to a positive result from the composite test. In the latter case, it is stated that there is a potential for the variable to be lognormally-distributed.

Additional comments

- page 1399, line 26: the minimizing solution is usually noted x_a while x is the free parameter of the cost function.

AC: Page 1366 contains the cost function (Eqn. 1) and the variable is now x_a throughout the manuscript.

- page 1367, line 22; Fletchera -> Fletcher (I suppose)

AC: It should read "Fletcher" and this typo has been corrected.

- page 1369, line 18: The validity of statistical tests relies critically on the assumption whether the samples are independent. I think that a paper focused on statistical tests should verify this assumption (even if it is not difficult to find other papers in the literature which do not verify this assumption).

AC: To check for independence, the authors analyzed the sample autocorrelation since we are considering a time series. The authors omitted these figures in lieu of the histograms of the actual data, which more directly convey the relevance of this research. The original manuscript addressed this point (page 1369 line 22) with the following paragraph:

"The samples' autocorrelation has been checked in order to verify the iid assumption for the hypothesis tests. While there is some autocorrelation in the samples, we attempt

to minimize its effect by choosing such a small α."

- page 1371, line 14: "This design ensures that the data truly is lognormally distributed without a false positive." I guess (hope) that this is a typo here. A test without any false positive is of course not possible (unless would be set to zero).

AC: Yes, unfortunately this is a typo and it has been corrected. The authors appreciate the noted need for clarification. The manuscript now reads:

"The goal of this design is that the data truly is lognormally distributed and attempts to reduce the possibility a false positive."

- page 1373, line 3: "If the distribution is normal, then asymptotically the Jarque–Bera (JB) test statistic has a chi2 distribution with two degrees of freedom...".Can you be more specific to what the term "asymptotically" refers to? Which parameter needs to tend to infinity? The sampling size?

AC: Yes, as the sample size n tends to infinity, the test statistic has a Chiˆ2 distribution. The JB statistic is essentially the sum of squared normal random variates where the normality comes from the Central Limit Theorem. The text is now corrected as follows:

"If the distribution is normal, then asymptotically, as the sample size n tends to infinity, the Jarque-Bera (JB) test statistic has a Chiˆ2 distribution with two degrees of freedom and. . ."

- page 1375, line 10: "Therefore the composite test concludes that almost 29% are lognormally distributed." The authors make often this shortcut in their reasoning. However, if the null hypothesis (variable follows a lognormal distribution) is not rejected, it does not mean that one can conclude that the variable follows a lognormal distribution (because the sampling size could also be too small to reveal deviations from the lognormal distribution).

AC: The original manuscript reads: "For example, the entire year of forecasts of mixing ratio at 300 hPa has almost 99% of points coming from a non-normal distribution as concluded by the Shapiro-Wilk and Jarque Bera tests, and almost 29% of points cannot be determined to not come from lognormal distribution as per the Chi-squared test. Therefore the composite test concludes that almost 29% are lognormally distributed.

The first sentence of this conclusion agrees with the reviewer's point that the interpretation of a result of a hypothesis test is incredibly important, and unfortunately the second sentence did overstep the hypothesis tests' framework. This sentence has been changed to the following:

“Therefore the composite test concludes that almost 29% are not normally distributed and simultaneously are not seen to be not lognormally distributed.”

- page 1381, line 1: "It has been shown in Fletcher and Jones (2014) that there is a negative impact on the performance of a normal distribution only incremental 4-D-VAR when lognormal forecasts are assimilated ". I guess that the authors wanted to say "... when lognormally distributed observations are assimilated". Otherwise I do not understand this sentence.

AC: The authors have corrected this sentence as suggested to reflect that lognormally-distributed observations are assimilated.

Interactive comment on Nonlin. Processes Geophys. Discuss., 2, 1363, 2015.