Nonlin. Processes Geophys. Discuss., 2, C560–C564, 2015
www.nonlin-processes-geophys-discuss.net/2/C560/2015/

Nonlinear Processes in Geophysics
Discussions

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 #1 Received and published: 5 October 2015

Review of the manuscript 'Identifying non-normal and lognormal characteristics of temperature, mixing ratio, surface pressure, and wind for data assimilation systems' by A. J. Kliewer, S. J. Fletcher, A. S. Jones, and J. M. Forsythe

Accepted for publication with MINOR REVISIONS

The present article presents a set of statistical tests, two of them for checking non-normality (Shapiro–Wilk, Jarque–Bera) and the third one for checking log-normality

(qui-square) of the temperature, mixing ratio and surface pressure data assimilation (DA) background fields on a grid point basis. The composite (logical conjunction) of the overall accepted test hypotheses (null or alternative) gives a hint for using a DA background term according to normal or lognormal error statistics. From that, it suggests an offline method, with possibilities to become an online one, of optimizing the appropriate error statistics and modify the DA cost function accordingly. This goal is well accomplished in the paper that is well written and succinctly presented. Minor points to discuss

AC: The authors appreciate this review and opinion and have addressed all of the following comments. Hopefully it is the reviewer's decision that the new manuscript is ready for publication in Nonlinear Processes Geophysics. Author comments begin with "AC: ."

1 - Pg. 1367, line 1: The assumption of x and xb being independent is not true. In fact xb is generally strongly correlated with x since xb is a good approximation of x for most of the state vector x conditions. Therefore in order to be consistent, substitute the sentence in the text by '...x and xb assumed to be jointly Gaussian, then the difference of these variables is also a normally-distributed random variable' The independence is too strong and generally not true. In a equal fashion, if log(x) and log(xb) are jointly Gaussian, but not necessarily independent, then log(x/xb) is Gaussian. Change text accordingly please.

AC: The authors agree that the wording of this statement should be clarified. It is our opinion that x_a and x_b are realizations from the same Gaussian distribution. This is a result of assuming the background errors are unbiased. To address this, the manuscript now reads:

"If both x_a and x_b are assumed to be realizations from the same Gaussian distribution then the difference of these random variables is also a normally-distributed random variable."

The original manuscript included further description (page 1367 line 6-10) that the observations x_a and x_b come from the same distribution, guaranteeing the unbiased error assumption. Hopefully the reviewer finds this new description a satisfactory interpretation of the prescribed cost function.

2- Pg 1371. Authors must say that the m(i) are the expected order statistics issued from a given pdf f(x) to be tested. According to the article, f(x) is the normal pdf. What is the theoretical pdf for the test SW under a Gaussian pdf f(x) and for Gaussian realizations of x?

AC: Yes, the text should, and now does, state that f(x) is the normal pdf. Since the theoretical pdf for the test statistic is quite complex, it is beyond the scope of this manuscript. The authors cite Hain for details regarding this hypothesis test (page 1372 line 8) and in Hain the following description is given:

"The last topic we want to investigate in this subsection is the null distribution of W, i.e. the distribution of W under the null hypothesis. Unfortunately, according to Shapiro and Wilk (1965), there is no possibility for giving an explicit form of the null distribution of W for sample sizes n $\geq$ 4. Shapiro and Wilk showed that there exists an implicit form for the distribution of W, for more details of this proof we refer to the original work."

Hain, J.: Comparison of common tests for normality, Thesis, Institut für Mathematik und Informatik, Julius-Maximilians-University at Wurzburg, Germany, 102 pp., 2010.

3 - It is useful to say that vector a in Eq.7 is the Mahalanobis norm of m and SW, being proportional to an inner product, is a concordance measure between a and x.

AC: The authors thank the reviewer for pointing out the presence of the Mahalanobis distance and have included the following note in the section describing the Shapiro-Wilk test.

"It is noted that the vector a contains a term similar to the Mahalanobis distance in the denominator."

4 – The statistical tests are independent. In fact when log-normality is accepted by the qui-square test, in general the SW and JB lead to non-normality. Are there contradicted cases in performed analysis? i.e. cases of accepted normality (by SW or JB) and accepted log-normality (by the qui-square test). Maybe for parameters where normal and log-normal pdfs look quite similar that could happen. Discuss a little this case.

AC: The authors appreciate this idea and investigated the question. This in fact does occur and the following description has now been included in the manuscript (last paragraph of Section 3.1).

"The composite test was constructed to indicate test agreement between the Jarque-Bera, Shapiro-Wilk, and Chiˆ2 tests, i.e. reject the assumption that the data are normally distributed and fail to reject the assumption the data is lognormally distributed. The question arises of whether the tests can suggest that the data follows both a normal and lognormal distribution, i.e. the Jarque-Bera and Shapiro-Wilk test fail to reject the normal assumption and the Chiˆ2 test fails to reject the lognormal assumption. There are numerous examples of this occurring, more prevalent in the lower atmosphere and in the seasonal time domains. In these scenarios the data exhibits low skewness and therefore the normal and lognormal probability distribution fits fall very closely together. This occurs less frequently in the upper atmosphere as values for mixing ratio are closer to zero, and since it is a positive definite variable, has a noticeable skewness. From a data assimilation perspective, it is hypothesized that using a Gaussian or a mixed Gaussian-lognormal system will retrieve similar values for the true state."

This phenomenon was briefly noted in the original manuscript's Conclusions and Discussion section (page 1379 line 26) in the following statement:

"While it is true that a lognormal distribution with a small variance looks very similar to a normal distribution, the detection methods used in this paper attempt to operationally handle large amounts of data similar to the resolution of an inner loop in incremental

data assimilation schemes."

Typos (change to :) Pg. 1367, line 22: Fletcher Pg. 1369, line 23: autocorrelation Fig 1 and Fig. 11: Year 1005 instead of 2005 appears in some panels Pg 1374 line 18: a forecast Pg. 1375, line 11: occurence Pg. 1380, line 18: Bayesian Pg. 1381, line 22: moment

AC: The authors' appreciate the reviewer locating these typos and all have been corrected.

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