Clim. Past Discuss., 11, C2284–C2289, 2015
www.clim-past-discuss.net/11/C2284/2015/

Climate of the Past Discussions

Interactive comment on "On reconstruction of time series in climatology" *by* V. Privalsky and A. Gluhovsky

V. Privalsky and A. Gluhovsky

vprivalsky@gmail.com

Received and published: 13 November 2015

1. The statement in the first sentence of our abstract asserting that time series reconstruction should not be made on the basis of cross-correlation and regression is correct and does refer to climate research as a whole. It follows directly from mathematics (theory of random processes and theory of information), as shown in the paper. If it is declared incorrect, the declaration must be proved mathematically.

2. We are not aware of any time series reconstruction studies in climatology based upon multivariate time series analysis in the way we did it. Therefore, we cannot discuss other models of SSN and TSI constructed in this or similar way. A short

review of statistical studies in the area of reconstruction starting from 1909 through 2015, which includes studies that use ARMA models (Guiot, 1985, 1986), is given in the paper, and it takes two pages. Also, this is a comparative study and a "thorough verification on independent data" is not an absolute necessity.

3. According to the Anonymous Referee #1 (the Referee in what follows), "the suggested model, if I understand it correctly ... comes down to the equation

$$TSI(t) = a * SSN(t) + b * SSN(t-1) + c * SSN(t-2)". \quad (1)$$

This equation suggested by the Referee is wrong and, with all due respect, it means, that the Referee has completely misunderstood the method suggested in the paper. The right hand side of the Referee's equation contains only one scalar time series, the arguments in parentheses are incorrect, and it presents just a slightly generalized form of exactly the same traditional linear regression approach that, according to time series analysis and to our paper, SHOULD NOT be used for time series reconstructions. This makes the Referee's general comments and recommendations that follow this incorrect equation groundless because they are given on the basis of a wrong conclusion about the model.

Yet, regarding the general remarks.

1. What is called by the Referee "nothing but standard material of timeseries analysis" is intended not for mathematicians but for climatologists who, judging, in particular, by the Referee's review, are not well-versed in respective theory and methods. As for the M-variate time series part (which takes just 8 lines of text and three simple equations), it is given in the paper to show the reader that the method can be applied in the multivariate case as well $(M > 2)$.

2. Contrary to what the Referee says, there are no predictions of TSI in the paper. Incidentally, if the "TSI memory" were available for the past, there would be no need for its reconstruction.

3. The expression "simulation based on ... a harmonic base model" needs to be clarified.

Regarding the specific comments.

4702, 3. In response to the Referee's question, "mathematically incorrect" in this context means that using cross-correlation coefficient and linear regression to describe relations between time series disagrees with mathematics, specifically, with theory of random processes and theory of information.

4705, 21. The Referee is right, the reference to von Storch et al (2004) should be removed.

4706. This part of the paper is very important because it explains that "the cross-correlation coefficient cannot characterize relations between the components of a multivariate time series". And it does not take "an entire page".

Incidentally, both statements in this comment by the Referee are wrong. The cross-correlation coefficient cannot be "seen as coherence averaged over frequency". And there is no "equivalent formulation of correlation or coherence" in the classical work by I. Gelfand and A. Yaglom, referenced in the paper.

4707. The fact that we are working with a bivariate time series and building its

autoregressive models is mentioned many times in the paper with proper references to mathematical sources. This approach is well known in random processes and accusing us of reinventing the wheel is unfair.

4710, 15. Saying that we demand "perfect" correlation (or coherence) is completely groundless. There are no such requirements in the paper. The basic idea is that the coherence function should be used here, not the cross-correlation coefficient. And the coherence does not have to be equal to 1 everywhere or anywhere (see Fig. 4).

4711, 14."Optimal AR orders are based on which criterion?" AIC for both scalar time series.

4711, 23. According to the Referee, "... explained variance and ... correlation are equivalent measures, so only one should be used". This is not correct, because they provide different information. Suffice it to say that the variance generally has a dimension while the correlation coefficient is dimensionless.

4712, 1. "The cross-correlation function does not look healthy". There is no such term in random processes as a "healthy-looking cross-correlation function". A Kronecker symbol? The cross-correlation function given in the article is obtained directly from observations. And its complicated behavior in our case is one of the reasons why the cross-correlation coefficient should not be used for time series reconstructions.

4712, 8: "time series can just as well be random variables ..." A time series is different in that it is a set of random variables distributed in time; as such, it has a correlation function and a spectral density, which do not exist for random variables.

4712, 13ff. We disagree. These recommendations in the paper are correct, little known in climatology, and relevant to the subject.

4713, 5. Results of analysis of a scalar time series would differ from those obtained from the analysis of a multivariate system that contains the scalar time series.

4713, 24ff. Contrary to the Referee's belief, not only Eq. (11) is not "essentially contained" in Eq. (12) but they are entirely different equations. In fact, the major point of the paper is that it is Eq. (12) that should be used for reconstructions. The text that the Referee erroneously regards as"trivial" contains specific results obtained with the two methods and demonstrates quantitatively the advantages of the time series analysis approach.

4714, 7. Again, there are no predictions in this paper. On top of that, the Referee's comment is not just irrelevant but is also incorrect. The extrapolation function of an $AR(p)$ time series at lead time τ is

$$\hat{\mathbf{x}}_n(\tau) = \mathbf{\Phi}_1\hat{\mathbf{x}}_n(\tau-1) + \mathbf{\Phi}_2\hat{\mathbf{x}}_n(\tau-2) + ... + \mathbf{\Phi}_p\hat{\mathbf{x}}_n(\tau-p), \quad (2)$$

where $\hat{\mathbf{x}}_n(\tau-k) = \hat{\mathbf{x}}_{n-k}$ if $\tau - k \leqslant 0, k = 1, ..., p$.

For example, $\hat{\mathbf{x}}_n(1) = \mathbf{\Phi}_1\hat{\mathbf{x}}_n + \mathbf{\Phi}_2\hat{\mathbf{x}}_{n-1} + ... + \mathbf{\Phi}_p\hat{\mathbf{x}}_{n+1-p}$.

In other words, at any lead time τ, the linear prediction within the framework of the Kolmogorov-Wiener theory always contains all p preceding terms, either observed (if $\tau - k \leqslant 0$) or predicted at larger lead times (see, e.g., Box and Jenkins, 1970; Box et al., 1994, 2015).

4715, 11. "How do you define . . .?" This is a question, not a comment. It is explained in the paper. The next sentence contains a mysterious notation"lag-1 x_1" that makes

the sentence enigmatic. But once again, there are no predictions in the paper.

4716, 6ff. It has been dealt with before (see #2 at the beginning of this response).

4717, 1ff. We disagree. Those two examples show other cases when the proper time series approach works better than the mathematically incorrect approach through the cross-correlation coefficient. Therefore, they are quite relevant for the paper's subject.

In conclusion, we would like to draw the Referee's attention to the following his or her comment:

4703, 12: "statements like ". . . regression approach is generally not correct." should be avoided". Here, the Referee added a period after the word "correct", which is not in our paper. The actual text in the paper is ". . . regression approach is generally not correct for analyzing multivariate time series." This is a ruse and we resent this manner of reviewing.

The authors.

Interactive comment on Clim. Past Discuss., 11, 4701, 2015.