Clim. Past Discuss., 11, C2400–C2404, 2015
www.clim-past-discuss.net/11/C2400/2015/


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: 19 November 2015

To Referee#2's major comments.

The Referee's major comments are mostly related to the Referee's lack of understanding, in particular, of the multivariate time series analysis.

1) "I don't see the arguments for preferring Eq. 2 instead of Eq. 1."

There are mathematical arguments in the paper. Here is just one: relations between time series are frequency dependent, the cross-correlation coefficient cannot and does not depend upon frequency; therefore, it cannot describe relations between time series. If you disagree with this, please say it and prove that it is wrong. The same goes for other arguments.



"...why should tree-rings depend on temperatures the previous years and not the same year?"

They do, look at Eq. (2): both x1(n) and x2(n) depend upon the past values of both x1 and x2 and in this way there is a dependence between the left parts of each equation, that is, between x1(n) and x2(n). Also, please read lines 18 and 19, p. 4713. The Referee's statement is wrong.

"...cases where the x2 does respond to x1 at the same time and not just trough the lagged terms. If x1 and x2 are annual values of temperature and tree-rings then why should tree-rings depend on temperatures the previous years and not the same year? "

See the above.

"If serial correlations are important they could be included in Eq. 1, e.g., by letting the noise term be an AR-k process."

Wrong. The proper way to do such things is described in Box and Jenkins (1970, etc); respective techniques are applied in the paper. If the Referee believes that what is done in the paper is mathematically incorrect, such an opinion should be supported mathematically. Otherwise, it is just an unfounded opinion.

2) "The paper insists that the new method is the proper method. I fail to understand that." With all due respect, the lack of understanding by a referee is not an authors' problem. But we agree, the Referee did not understand the method (and wrote a review of it).

"In the example given the new method gives a better fit than the simple regression. However, this should be expected as the number of parameter is larger and as the models are tested on the same data as they are trained on."

According to this logic, the more parameters is included into the model, the better the results will be. Wrong. Example: an AR(1) sequence (a Markov chain) contains just one AR coefficient. Increasing the number of coefficients would make the results of analysis worse, not better.

“A fair way to test the methods and avoid over-fitting would be to test on independent data, e.g., using some kind of cross-validation.”

Please read Box et al (1970, . . ., 2015) and other literature referenced in the paper and in Privalsky (2015) dedicated to the selection of optimal parametric models.

3) “There is actually a lot of relevant discussion about the reconstruction methodologies in the literature that the authors does not mention.”

This a research article, not a review. It does contain a short review of relevant earlier publications. A new method of time series reconstruction based upon multivariate time series analysis in both time and frequency domains is suggested in the paper. Has it been published in climatology before? The Referee has nothing to say about it.

A. Douglass (1936) says that the “similarity between two trees curves . . . is only partly expressed by a correlation coefficient.” – no objections from the Referee.

I. Gelfand and A. Yaglom (1957) say that it is the coherence function (not the cross-correlation coefficient) that describes the mutual information about time series - no objections from the Referee.

“Perhaps the authors should consider using this approach to compare the new method to older ones.”

This is exactly what is done in the paper: comparison with the correlation/regression method by far the most traditional in climatology and paleoclimatology. The authors welcome comments on the paper; suggesting writing a different paper on a different or similar subject is not helpful.

4) “The example with the TSI and SSN seems strange. The strong periodicity would suggest a model with an external forcing.”



This is what is done in the paper. Eq. (12) describes a linear system with SSN as the external forcing and TSI as the output process. The system has a closed feedback loop. Please see, e.g., Bendat and Piersol (1967, . . ., 2010). Also, read p. 4713. And a “strong periodicity” is not necessary for having a system with an external forcing.

“Also the improvement when using the new model seems modest (in particular given the possible over-fitting). There must be a better example.”

We do not have a better example. This one is mathematically correct (as opposed to the regression approach) and gives better results. What is wrong with this?

5) “There is too much jargon as the other reviewer mentioned.”

This is not jargon but mathematical terminology. Note also that this accusation is not supported with any examples, that is, it is groundless.

“Is the historical part and the part about the information rate really relevant?" Absolutely.

“Why not simply say that the connection between x1 and x2 may depend on the frequency and that the simple regression, Eq. 1, may therefore not be sufficient?”

Not may depend but depends; not “may . . . not be” but “is not”. And it is stated in the paper (see p. 4703, lines 15, 16; p. 4706, lines 11, 13).

On Referee’s minor comments:

“Eq. 1 is not complete without information about whether the noise and x1 or the noise x2 are independent.”

This linear regression equation cannot describe multivariate time series, it is valid only for random variables, which do not have correlation functions.

“Section 4.1: If model orders of 32 or 33 are necessary then the model is probably not the correct one. Perhaps an ARMA model would be better.”

Please prove it. Besides, finding the best ARMA model for the SSN and TSI scalar time series is outside the scope of the paper.

“Eq. 12: Perhaps the coefficients could be moved to a table and shown together with their uncertainties.”

We disagree. Most of the coefficients (9) are statistically significant at a confidence level 0.9.

“Caption to Fig. 6 and the text near bottom of page 4715 disagree on the period.”

Please clarify.

“Second point on page 4712: It is not clear what is referred to here. The problem with serial correlations reducing the number of degrees of freedom? Multiple regression?"

What “second point”? There are no problems with “serial correlation” in the paper, the degrees of freedom is not mentioned in it and there are no multiple regressions. If the Referee means Fig. 3, it is just a direct estimate of the cross-correlation function between SSN and TSI.

In conclusion, the lack of understanding and the inability to see what the paper says can hardly serve as a valid basis for rejecting the paper that one does not understand.

---

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