

# ***Interactive comment on “Detecting instabilities in tree-ring proxy calibration” by H. Visser et al.***

H VISSER, U BÜNTGEN, R D’ARRIGO and AC PETERSEN (authors)

[hans.visser@pbl.nl](mailto:hans.visser@pbl.nl)

Received and published: ## 2010

L.s.,

We will respond to the comments of the three reviewers in order of publication date:

- Reviewer 1, V.V. Shishov
- Reviewer 2, A. Nicault
- Reviewer 3, L. Kutzbach

## **Reply to reviewer 1**

We thank the reviewer for his remarks on the paper. The first remark deals with the way ring widths have been standardized in the publication D’Arrigo et al. (2009) on page 233, line 5 of the paper. We agree with the reviewer that the process of standardization plays an important role in proxy calibrations. If due to an incorrect standardization procedure spurious trends enter into the standardized index chronology, the SRF approach will trace these trends by showing them in the trend component  $\mu_t$  (see model (1) in the paper). The consequence will be a decision not to use this particular chronology in a reconstruction study. Perhaps a more appropriate standardization technique would have prevented us to take this negative decision.

However, how do we know what the correct standardization technique is? On this topic a vast literature exists. Methods are, among many others, exponential functions, RCS, splines, age banding and the signal-free approach, as suggested by the reviewer.

Our main answer to the reviewer is that we present a new technique in proxy reconstruction which allows us to detect instabilities between tree growth and a climate proxy. That is the theme of this paper. To illustrate the SRF technique we have chosen two articles from the recent peer-reviewed literature: D’Arrigo et al. (2009) and Büntgen et al. (2008). Thus, we did not perform any standardization procedure on the data published in D’Arrigo et al. (2009). That was not the goal of our paper. Perhaps the misunderstanding comes from the fact that the first authors of these two studies are also *co-authors* of this paper.

Although not the theme of our paper, our advise on standardization would be to perform a sensitivity analysis on different tree species (as done in both articles), on different ring-width indicators (TRW or MXD), as done in D’Arrigo et al. 2009) or on different standardization techniques (RCS, splines, signal-free, age banding, etc.). This latter approach is followed in Büntgen et al. (2008) and e.g. in Esper et al. (2010).

*We propose not to change the text.*

The second remark deals with the properties of the residuals of the estimated SRF models (model (1) in the paper). The reviewer states that these residuals - in Kalman-filter terms one-step-ahead predictions or *innovations* – should be normally distributed. Additionally, he asks if it is possible to test the residuals for normality. If they follow another probability distribution, other approaches would be more suitable.

Our answer is that it is an attractive property for innovations if they follow a normal distribution. In that case the Kalman filter yields the minimum mean square error estimates (MMSE). See page 242, line 18, in the paper. However, it is not a **necessary condition** for the Kalman filter. If the innovations do not follow a normal distribution, the filter still yields the minimum mean square linear estimator (MMSLE). See page 242, line 20, in the paper. For details on normality or non-normality in the context of the Kalman filter and structural time series models please refer to Harvey (1989, page 111).

In conclusion, the normality of innovations is *not* a necessary condition for the Kalman filter. However, if the innovations are normal, the estimators have stronger statistical properties, i.e. MMSE. Therefore, we totally agree with the reviewer that it is advisable to test any innovation series for normality. One side remark: if normality holds,  $2\text{-}\sigma$  confidence limits can be interpreted as 95% confidence limits.

Although we did not make a remark on normality in sections 3.1 and 3.2, we have tested all models, summarized in Tables 1 and 2, on normality. To this end we prefer to use *normality plots*. These plots show visually if the innovation series are normally distributed. In that case the innovations will lie on a straight line. The advantage of such a graph is that we can judge the deviations of the straight line. The disadvantage is that some subjectivity is involved in this judgment. We did not perform a formal test on normality, such as the Kolmogorov-Smirnov goodness-of-fit test or the Shapiro-Wilk goodness-of-fit test, because these tests will reject the Hypothesis of normality for long time series (N large). *In practice* it is sufficient that the *shape* of the innovation distribution is close to normality, whatever the length of the sample period is.

For the eight models presented in Tables 1 and 2, three innovation series showed perfect normality, three series showed reasonable normality and two series showed moderate normality. We found these results good enough to assume normality in all eight models.

We note that there is another property of the innovation series, not mentioned by the reviewer, which is more important and even a **necessary condition** for Kalman filter estimates: the innovation series should be *white noise*. We tested this property by plotting the autocorrelation function (ACF) for the innovation series, along with  $2\text{-}\sigma$  confidence limits for all time lags plotted (we chose a maximum lag of 20 years). Furthermore, we made for all series a so-called lag-plot. That is a scatterplot between the innovation pairs  $(v_{t-1}, v_t)$ , with  $t$  any year in the sample period. This graph gives a visual impression of the presence or absence of any coherence between subsequent innovations. Thus, the lag-plot presents a visual presentation for serial correlation.

*We propose to*

- Add to page 229, line 6: (MMSE, normally distributed noise processes)
- add on page 242 the reference Harvey (1989, p. 111) for details on normality/non-normality.
- name the abbreviation MMSLE in line 20 of page 242 and name in line 18 that MMSE is for normally distributed innovations only.

- add some lines at the end of Sections 3.1 and 3.2 on test results on normality and whiteness of innovations, along with a mentioning of how these inferences were made (normality plot, ACF, lag plot).

## Reply to reviewer 2

We thank the reviewer for her remarks and her compliment. All minor remarks will be implemented.

## Reply to reviewer 3

We thank the reviewer for his detailed and elaborate comments, and his positive judgment of the value this paper.

### *General comments*

All comments are treated below in the specific comments. We strictly follow the reviewers comments, numbered from 1 up to 29.

### *Specific comments*

**Comment 1)** As for initial values of noise variance: we have chosen the approach of a diffuse or non-informative prior (implemented in the TrendSpotter software). This means that we set the initial covariance matrix to the unity matrix with large numbers on the main diagonal. Thus, we simply ‘tell’ the filter that we have no information whatsoever at the first iteration. See Harvey (1989, page 121) for details.

The consequence of this approach is that the filter needs some iterations to arrive at stable state-space estimates. For the models presented in Tables 1 and 2 of this paper we have chosen for a transient period of 20 years. The consequence of a ‘diffuse prior’ is that the innovations series starts after these 20 years. See Harvey (1989, page 256) for details.

One can wonder why this transient period is not seen in Figures 1 through 4 in our paper. The reason is simply that these graphs do not show the *filtered* estimates for  $\mu_t$  and  $\alpha_t$ , but the *smoothed* estimates.

In conclusion, due to the diffuse prior we do not have the problem of initialization of noise variances. We simply let the filter start without any information. The consequence is the presence of a transient period. The length of this transient period can be found by plotting filtered estimates, rather than smoothed estimates.

*We propose* to add some lines to Appendix A on using a diffuse prior at the start of the filter iterations. We will mention the period of 20 years for the examples in this paper. This with reference to Harvey (1989).

As for the use of the symbols  $\eta_{i,t}$  and  $\eta_t$  in Appendix A (lines 9 and 13, page 242): we agree that these symbols are very alike. We are not in favor of using a different Greek symbol for these noise variances. An alternative would be to replace the notation  $\eta_t$  and  $\sigma_\eta$  by  $\eta_{m+1,t}$  and  $\sigma_{\eta,m+1}$  since there are  $m+1$  noise variances in total.

*We propose* to change the notations on page 242 in this way (thus using ‘m+1’).

As for tests on normality and serial correlation: see our answer to reviewer 1. *We propose* to add text to Sections 3.1 and 3.2 with respect to these two items.

**Comment 2)** The reviewer has right: the data cannot ‘choose’. What we meant to say, in a popular way, is “letting the model choose”. As set out above: the initial values are diffuse. The process of maximum likelihood optimization the transient period is left out (here the first 20 years of the sample period). *We will change* ‘data’ to ‘model’.

**Comment 3)** Okay. We meant to say that the ‘model without intercept’ is best avoided. But the sentence suggests differently. *We will change* the wording here.

**Comment 4)** Agree. No changes needed.

**Comment 5)** We agree with the reviewer that a climate envelop is a reasonable first step if proxy-growth relations appear to be stable. We were too optimistic in suggesting that this climate envelop will help for situations where instabilities are found at the end of the calibration period (i.e. the divergence problem). We will return to this point in our reply to ‘Comment 11’ of the reviewer.

We propose to add the reference Fritts (1976, p. 15), next to Loehle (2009). Furthermore, we will add the following sentence at line 23 of page 232: ‘We will return to this point in Section 4.1.’

**Comment 6)** Okay. We will add a sentence explaining that  $\mu$  and  $\alpha$  come from model (1).

**Comment 7)** Okay, we will add a sentence on hypothesis testing and refer to Harvey (1989, p. 236) for more details. Furthermore, we will remark that the symbol ‘ $\alpha$ ’ used here, should not be confused with the symbol for the weighting factor in model (1).

**Comment 8)** Same answer as given to the first remark of reviewer #1. We have taken two examples from the recent literature and take the decisions of the authors of these articles ‘as they are’. We propose not to change the text.

**Comment 9)** The reviewer asks: is this right? Our answer is: yes, that is correct. We will add an extra line explaining the term ‘explained variance’.

**Comment 10)** Agreed. For some reason, the text on the example shown in Figure 2 has been omitted. We will add three lines explaining Figure 2.

**Comment 11)** As stated in ‘comment 5’, we agree with the reviewer. We were too optimistic here. In fact, the presence of instabilities at the end of the calibration period is bad news. Such an index chronology could better not used in an unbiased reconstruction of past conditions, as formulated by the reviewer. Loehle (2009) gives a clear (mathematical) argumentation.

*We propose* to keep Section 4.1 into text since the omission of data over recent decades has been a means to avoid instabilities in a number of articles in the literature. We will keep the text in Section 4.1 unchanged up to the middle of line 15 on page 237. The text in lines 15-21 will be replaced by the following new text:

This argument is in line with the Uniformitarian principle as formulated in the Introduction. The principle implies that the same *kinds* of limiting conditions affected the same *kinds* of processes in the same ways in the past as in the present; only the frequencies, intensities, and localities of the limiting conditions affecting growth may have changed (Fritts, 1976). Loehle (2009, p. 241) comes to a similar conclusion, using a mathematical approach: ‘if a reconstruction already shows divergence, it is an indication that recent temperature are already in the non-linear zone; such reconstructions should not be used for evaluating past climates.’

Loehle also suggests an argument *in favor* of omitting data over recent decades. If we could address the cause of recent instabilities to an external factor, only *recently* in operation (air pollution, soil acidification, etc.), we could truncate the calibration period. However, this introduces a new problem: how could we uniquely attribute instabilities to such (anthropogenic) drivers? E.g., in case of the example shown in Figures 3 and 4 we do not have such a unique clue.

In conclusion, we feel that the omission of data over recent decades is not a sufficient means to accept a specific chronology for use in a reconstruction network.

**Comment 12)** See our text above.

**Comment 13)** Okay. We will replace the sentence in lines 10-11 by the following:

Unfortunately, this shortening does not guarantee stability in the past, as discussed in Section 4.1.

**Comment 14)** Agreed. We propose to alter the conclusions 2, 3 and 4. Conclusion #3 becomes conclusion #2. Conclusions #2 and #4 change to:

3. Stochastic response functions are ideally suited to localize instabilities over time. However, if these instabilities occur in recent decades (‘divergence’) and if the cause of these instabilities can not be traced/attributed to drivers which are only in operation *during these recent decades*, the omission of recent decades in the calibration period is not a valid means of generating an unbiased reconstruction network.
4. Two examples have been discussed as illustrations of the potential application of stochastic response functions to climatic reconstructions. For both examples, we find screening results that are only partly comparable to those found using other methods of validation ( $R^2$ , RE, CE). It is unclear if the stochastic response methodology would filter out *more* proxies than these traditional methods. Clearly, much more analysis is necessary to evaluate the various screening methods.

**Comment 15)** Okay, we will add a few extra lines explaining  $R^2$ , RE and CE.

**Comment 16)** Okay.

**Comment 17)** Okay.

**Comment 18)** Okay.

**Comment 19)** We will add: (TRW or MXD)

**Comment 20)** Okay.

**Comment 21)** Okay.

**Comment 22)** Okay.

**Comment 23)** Okay.

**Comment 24)** The term ‘correlation coefficient’ is equally common.

**Comment 25)** Agreed. Square symbol will be added.

**Comment 26)** Okay, we will use the suffix ‘m+1’ for  $\eta_t$ :  $\eta_{m+1,t}$ .

**Comment 27)** Since the variables  $I_t$  and  $\mu_t$ , shown in Figure 1a, have no unit (normalized index), the trend difference has no unit either.

**Comment 28)** Okay.

**Comment 29)** Okay.