On Variance Adjustments in Tree-Ring Chronology Development

D. Frank[1], J. Esper[1] & E. Cook[2]

[1] *Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland*

[2] *Tree-Ring Laboratory, Lamont-Doherty Earth Observatory, Palisades NY, 10964, USA*

Email: frank@wsl.ch

Introduction

In dendrochronology it is common practice to create a mean-value function as the best estimate of the trees' signal at a site. This averaging process helps eliminate noise particular to individual trees and cores thereby increasing the signal quality. The variance of the mean-value function, however, depends upon the number of series averaged together and their interseries correlation (Wigley et al. 1984). As the number of single series rarely remains constant in dendrochonological or more generally climatic investigations, simple averaging routinely produces changes in variance that are solely a by-product of changes in the number of series. This issue extends into a wide variety of fields; relevant examples from global change studies include the construction of instrumental averages with diminishing numbers of stations and spatial representativity back in time (Jones et al. 1999), averages of proxy networks for climate reconstruction (Esper et al. 2002), and in the construction of individual tree-ring chronologies (e.g., Esper et al. 2005, this volume).

Osborn et al. (1997) theoretically provided and experimentally tested a correction procedure to eliminate variance changes resulting from changing sample replication. The basic correction centers around the use of the effective independent sample size, N_{eff}, which considers the sample replication at every time and the mean interseries correlation (hereafter, rbar) between the samples. This is defined as:

$$N_{eff} = \frac{n(t)}{1 + (n(t) - 1)r}$$

where n(t) is the number of series at time t, and r is rbar. Multiplication of the mean timeseries with the square root of N_{eff} at every time t theoretically results in variance that is independent of sample size. This result can then be further scaled by the square root of 1/rbar to yield an estimate of the mean series in the original units. In the limiting cases, when the rbar is zero or unity, N_{eff} obtains values of the true sample size and unity, respectively. Osborn et al. (1997) extend this basic correction procedure with examples showing possibilities and methods to account for temporal and frequency dependence in rbar. Temporal dependence in rbar may arise from changes in the spatial density of series being averaged together, for example.

In this paper we intend to revisit this topic from a more applied perspective. Specifically, our motivation comes from evaluating the final chronology for the Morocco dataset (see Esper et al. 2005, this volume) corrected with the commonly applied variance stabilization routine incorporated into the program ARSTAN (Cook 1985). After the application of the

“Briffa/Osborn” correction, the variance of this chronology was observed to still increase back in time (Esper et al. 2005, this volume). Corresponding to this increase in variance is an increase in the rbar, which is to a large extent a likely consequence of the higher percentage of correlations computed between cores collected from the same trees (see figure 6 in Esper et al. 2005, this volume). We hypothesize that the correction procedure used in ARSTAN, which does not consider time-dependent changes in the rbar, is limited by this trend in correlation. To more completely assess and potentially improve upon this situation, we developed a routine, following the guidelines of Osborn et al. (1997), to allow for time dependent changes in correlation to be considered in the variance stabilization. We test and compare this time dependent rbar correction method, with the uncorrected average and the sample size correction method that uses a single time independent rbar value (as in ARSTAN).

This paper contains two major themes. In part 1, synthetic datasets are introduced to demonstrate the effects of changing sample size and correlation on the computation of mean chronologies. For each dataset, three mean-value functions and running standard deviations of these mean-value functions are presented. In the second part, we look at the Morocco dataset and evaluate its characteristics with the same array of computations. Results lead us to explore some of the characteristics of the basic tree-ring data from Morocco. We close with a brief discussion and conclusion.

Data

For this study we utilize a set of three synthetic datasets composed of white noise. These datasets possess varying correlation structures and varying sample replication (Fig.1) to allow evaluation of the influence of these features on computation and correction of mean-value functions. The individual series within these datasets have constant variance and no autocorrelation apart from that resulting from chance alone. These three “Cases” should a) illustrate how and when variance inflation occurs and b) how the different variance corrections can or cannot account for the time dependent changes in sample replication and correlation. In Case 1, a dataset with nearly time stable correlations, yet decreasing sample size is studied. In Case 2, a dataset with constant sample replication and increasing rbar back in time is studied. For Case 3, the two basic characteristics of Cases 1 & 2 are combined, for a synthetic dataset that possesses diminishing sample replication and increasing rbar values back in time. This third case most closely mimics the sample replication and rbar characteristics of the Morocco dataset. It should be noted however that as these datasets are generated from a random normal distribution with independence between neighboring observations, they do not fully capture the characteristics of real tree-ring data. Nevertheless, we do believe they represent a good basis for demonstrating and testing some of the features relevant to the computation and variance correction of a mean-value function.

The Morocco dataset is a collection of *Cedrus atlantica* samples taken from living trees in 2002. The individual series, after applying an adaptive power-transformation (Cook and Peters, 1997) and detrending with a 300-year spline, are the basic data utilized in the calculations herein. These data and their treatments are the same as those used for the final chronology shown in Esper et al. 2005 (this volume), although in closing we briefly detail some of the attributes of the raw and ratio detrended Morocco data. The reader is referred to the Esper et al. 2005 publication (this volume) for more details and information about these data.

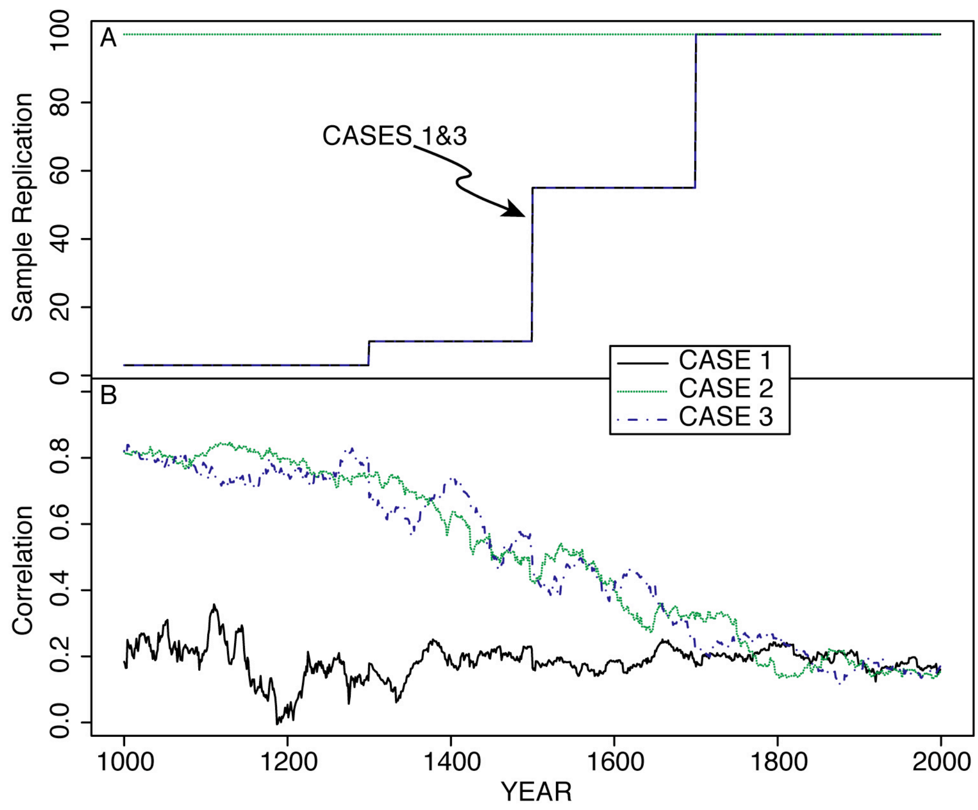

Figure 1: Sample replication (A) and rbar computed in 50-year running windows (B) for the three synthetic datasets. Case 1:stable rbar, decreasing n; Case 2: increasing rbar, constant n; Case 3: increasing rbar, decreasing n.

Methods

For all datasets we computed mean time series with a) no sample size correction, b) with the basic correction for sample size that utilizes a single mean estimate of the rbar, and c) a correction that considers both the sample size and temporal dependence of rbar. Herein, we refer to these chronologies as: UNCORRECTED, MEANr corrected, and RUNNINGr corrected. The MEANr corrected version, following Osborn et al. (1997) should be rather similar to that used in ARSTAN, although there is a range of possibilities in exactly how rbar is determined. For the RUNNINGr correction, we utilized a 50-year window to estimate the rbar at every time t. Rbar is computed as the average Pearson correlation, of all pairs that share at least 25 years of data, within a given window. This diverges from the Osborn et al. (1997) approach, wherein for every time t, the rbar was determined as the average correlation among all data pairs with data during time t, computed over their maximum period of overlap. That is, in the Osborn et al. (1997) approach, the rbar only changes when sample replication changes, and the period of rbar estimation occurs over the period of maximum common overlap and not confined within a running window as in our approach. With the RUNNINGr correction, additional ambiguity exists in the window size used for computation. The final UNCORRECTED, MEANr, and RUNNINGr corrected series are evaluated by simply plotting the resulting chronologies, and by plotting running standard deviations of the mean-value functions computed within 50-year windows.

Results

Variance adjustments in synthetic datasets

Figure 2 shows the final chronologies and running standard deviation for Case 1 which has a decreasing sample replication and a nearly constant rbar of about 0.2. A clear increase in variance is observed in the UNCORRECTED mean, when the sample replication drops from 50 to 10 and then again from 10 to 3. The UNCORRECTED mean is rather insensitive to the change from 100 to 50 series and reflects the asymptotic nature of N_{eff} for larger sample sizes. The MEANr and RUNNINGr chronologies are quite similar and neither shows an increase in variance. Periods when the MEANr chronology has greater variance than the RUNNINGr chronology, correspond to time periods when the running rbar is greater than the mean rbar (Fig. 1b).

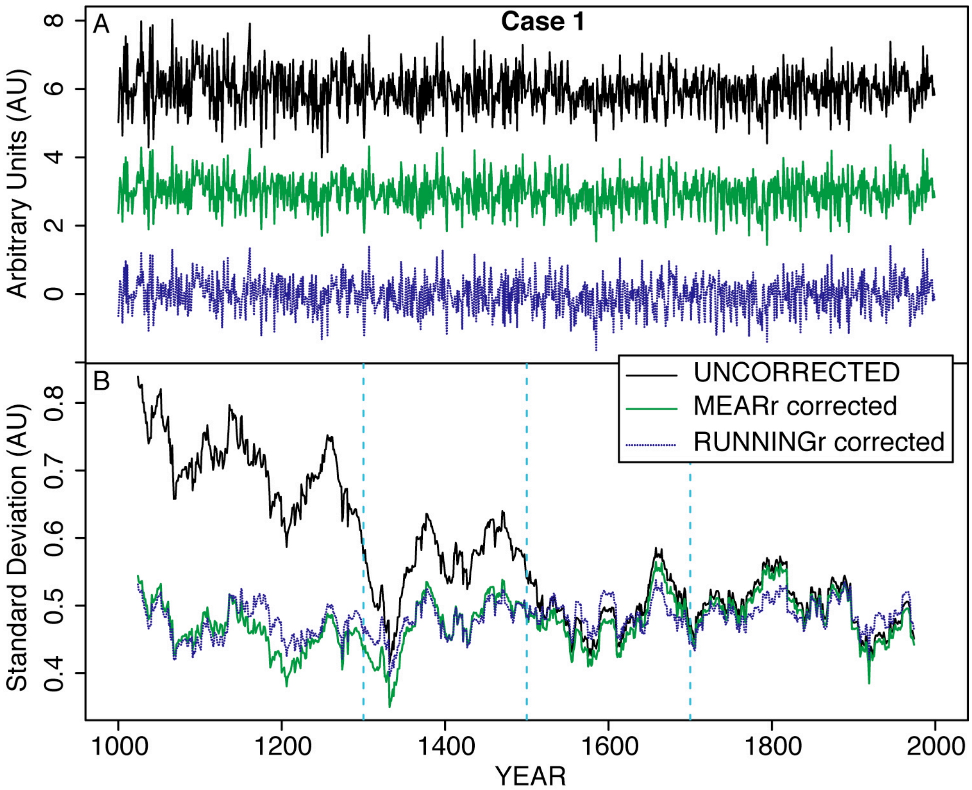

Figure 2: Mean value functions (A) and their running standard deviation (B) computed for the Case 1 synthetic dataset. The times corresponding to replication changes are indicated as vertical dashed lines in B.

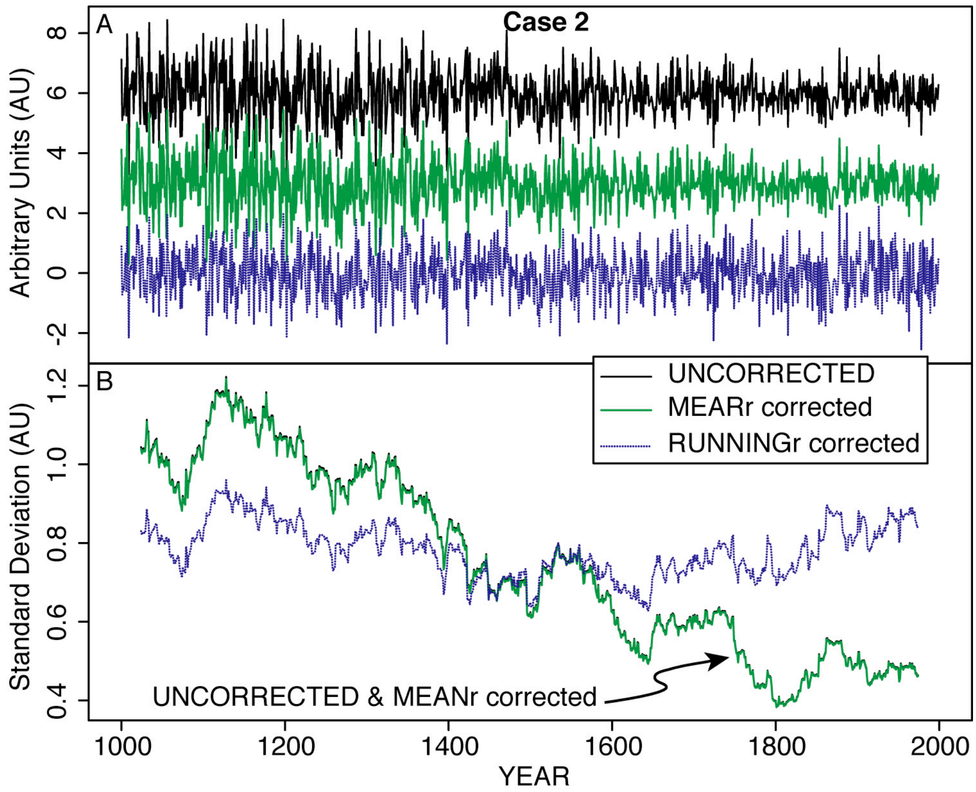

Figure 3: Mean value functions (A) and their running standard deviation (B) computed for the Case 2 synthetic dataset.

For the Case 2 dataset (Fig 3.), where sample size remains constant and rbar essentially increases monotonically (Fig. 1b), the UNCORRECTED and MEANr corrected chronologies are identical, with both possessing substantially greater variance towards the higher rbar values back in time. In contrast, the variance of the RUNNINGr chronology remains reasonably stable over the entire chronology, with deviations representing only the stochastic nature of this dataset. This demonstrates the ability of the RUNNINGr correction to "follow" changes in the underlying correlation structure and subsequently mitigate or eliminate these biases during the computation of the mean-value function. It should be noted, that the time dependent correlation demonstrated by Osborn et al. (1997) would produce the same results as the UNCORRECTED and MEANr corrected versions here, because using their methodology the correlations are computed over the period of individual series overlap, which in this case does not change. However, Case 2 (constant replication, increasing rbar) is perhaps not very realistic for real data characteristics and is used here primarily to illustrate the systematic changes that result from changes in rbar alone.

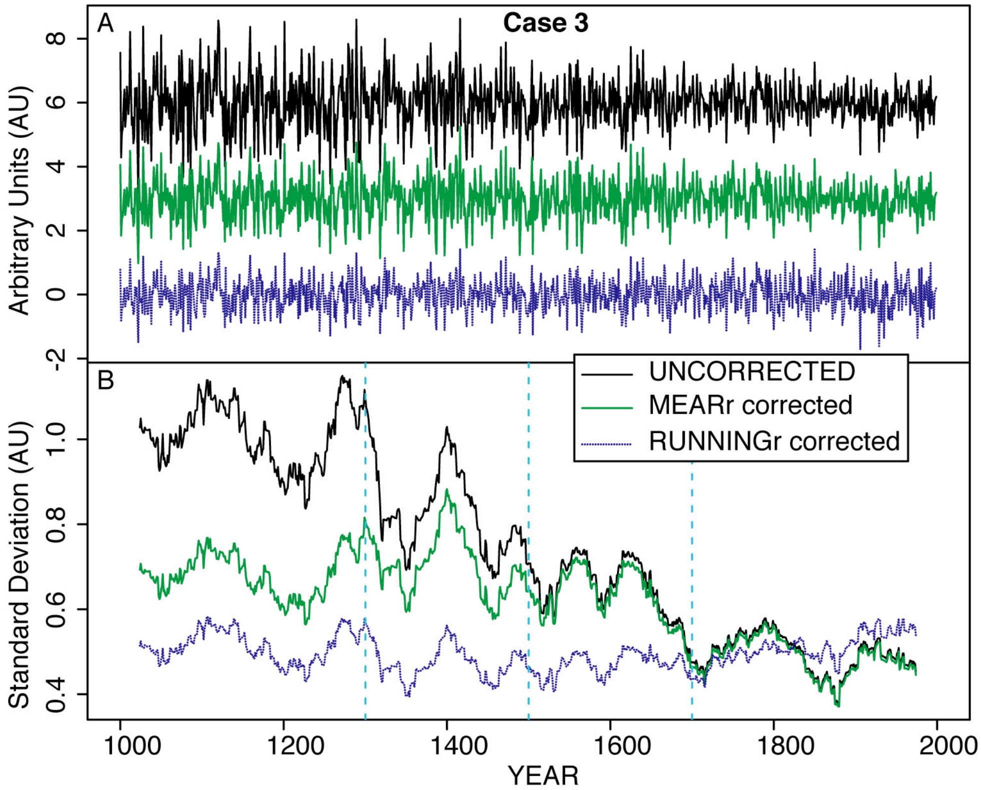

Figure 4: Mean value functions (A) and their running standard deviation (B) computed for the Case 3 synthetic dataset. The times corresponding to replication changes are indicated as vertical dashed lines in B.

In Case 3, as could be inferred, a mixture of the attributes of the mean value functions from Cases 1 and 2 are observed (Fig.4). The variance of the RUNNINGr corrected chronology remains rather stable, as this method successfully considered the time dependent changes in rbar and sample replication. Both the UNCORRECTED and MEANr corrected chronologies show a noticeable increase in variance roughly corresponding to the time when replication drops from 100 to 50 series. It is however likely that the majority of this increase in both series is due to the somewhat steeper increase in correlation at about this time, rather than the decrease in sample replication. In comparison to Case 1, the difference between the variances of the UNCORRECTED and MEANr corrected chronologies in Case 3 is smaller between 1300 and 2000. This is a result of the greater insensitivity to changes in sample replication for higher rbar values (Osborn et al. 1997), which are on average higher for Case 3 than for Case 1. During the earliest period, the variance of the UNCORRECTED chronology shows substantially inflated values, in comparison to the MEANr and RUNNINGr corrected chronologies. In this example, based on the differences between the standard deviation for the different computations during the early portion of the record, it can be estimated that during this early time period, about one third of the standard deviation increase in the UNCORRECTED chronology results from the increase in correlation over the course of this series, and about two-thirds from sample replication changes. Interestingly at around 1400 where the difference in variance between the MEANr corrected and RUNNINGr corrected chronologies is greatest, rbar reaches a local maximum value (Fig. 1b), which results in a decrease in N_{eff}, and hence also the standard deviation for the RUNNINGr corrected series.

Variance adjustments in Cedrus atlantica (Morocco)

We applied the same basic procedure as described above to the *Cedrus atlantica* dataset (see Esper et al. 2005, this volume), with results shown in Figure 5. The raw tree-ring data were subjected to a power-transform to eliminate the heteroscedastic behavior of the raw ring width series (Cook and Peters, 1997) and detrended by taking residuals between the power-transformed data and 300-year-spline fits to eliminate the age-trend. As is conventional for tree-ring indices, the detrended series were rescaled automatically in ARSTAN to have a means near unity. However, it is important to note the variance corrections outlined here assume the individual series to have a mean of zero. To approach this condition, we simply subtracted 1 from all series, resulting in series that are centered approximately around zero. Also included for comparison is the "Briffa/Osborn" variance adjusted version of the Morocco dataset as computed in ARSTAN. To account for the different scalings applied in our calculations and those from ARSTAN, all series were normalized with respect to the 20th century.

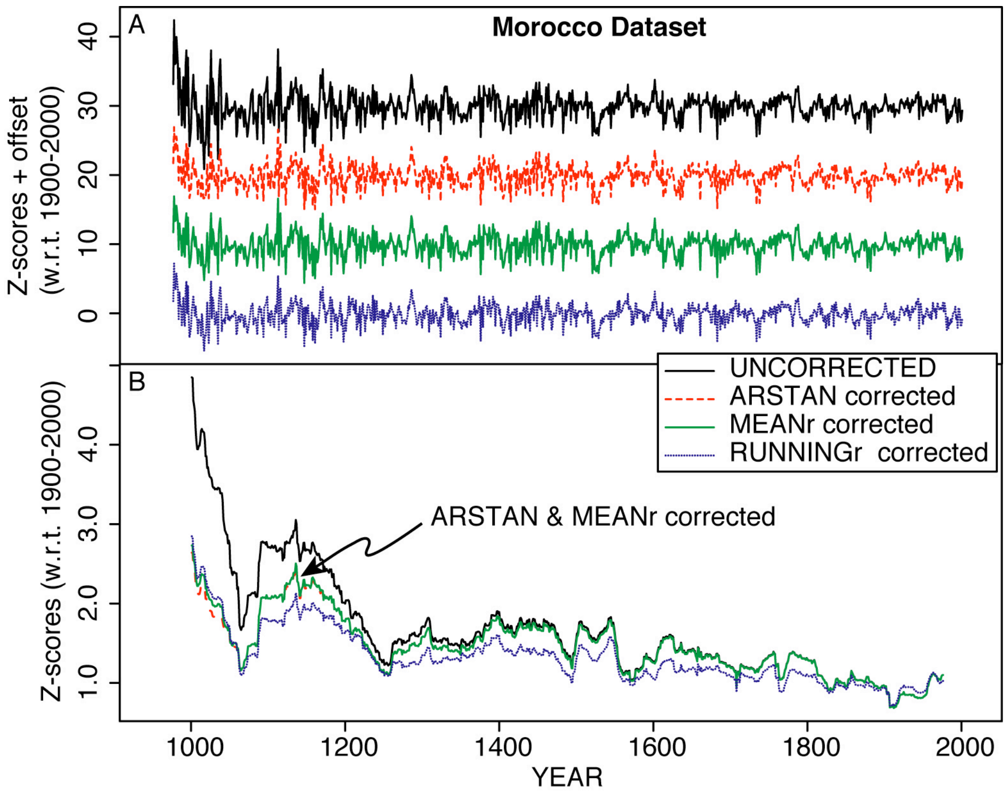

Figure 5: Mean value functions (A) and their running standard deviation (B) computed for the Morocco dataset. Also included is the ARSTAN computed correction. For more details and plots showing sample size and interseries correlations, the reader is referred to Esper et al. (this volume).

The UNCORRECTED, ARSTAN and MEANr corrected versions are nearly identical back to 1400. Prior to this time the UNCORRECTED variance splays apart from the other series, with substantial differences during the first hundred or so years. The ARSTAN and MEANr corrected chronologies are nearly identical, with only slight computational differences present during the early portion of the records. The RUNNINGr corrected chronology possesses the smallest trend in variance, indicating the utility of this method for consideration of the increase in rbar (and decrease in replication) back in time. However, in contrast to the synthetic dataset cases shown above, even the RUNNINGr chronology possesses a noticeable trend in the variance. Although mitigating the variance increase, this result largely refutes our initial hypothesis that the RUNNINGr correction procedure would eliminate the variance trend in the final Morocco chronology, which we assumed was primarily a consequence of the rbar increase towards the early portion of this dataset.

In an attempt to understand the source of this variance increase, we computed running standard deviations for the individual series of the Morocco dataset (Fig. 6). It is evident that the average of the individual standard deviations increases back in time. In all likelihood, this tendency explains most of the variance increase still present in the RUNNINGr computed version of the Morocco chronology. None of the correction methods applied are able to “see” or cope with this apparent variance nonstationarity of the individual series.

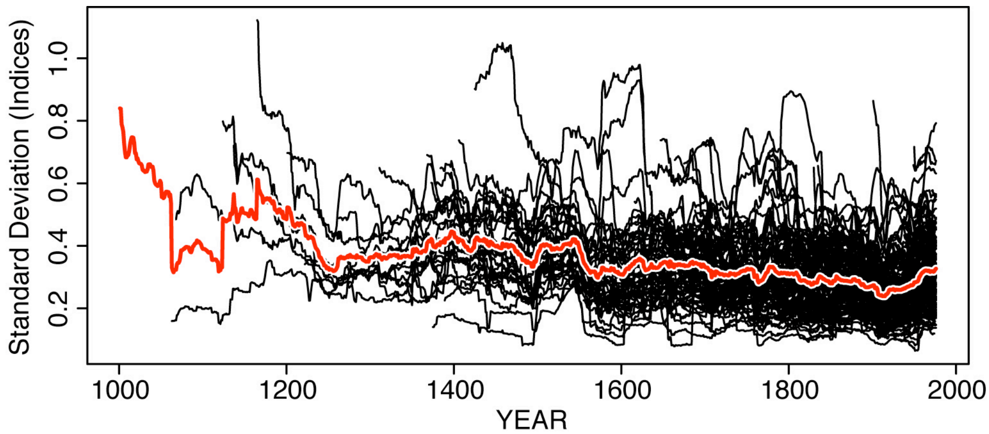

Figure 6: 50-year running standard deviations for the individual series from the Morocco dataset, along with their mean.

To explore if the variance of the individual series may increase with decreasing biological age, as is done for Regional Curve Standardization (Briffa et al. 1992, Esper et al. 2003), we aligned the data by cambial age (although pith-offset data were not used) and again computed the running standard deviations and their mean (Fig. 7). To test the influence of the detrending method, we also detrended the raw tree-ring series using ratios from 300-year-splines instead of power-transformation with residuals. The means from the age-aligned raw and ratio detrended running standard deviations are also shown in figure 7. The standard deviation of the raw series shows the most pronounced trend with biological age. This tendency diminishes with the power-transformed series, and is essentially non-existent with the ratio detrended series. It appears from these results, that these age-related tendencies in variance, as measured here, are not fully removed by the spread versus level calculations used in the adaptive power transformation, yet are removed by the ratio detrending. It thus seems likely, that the biological age-related tendencies in the variance structure of the power-transformed/residual detrended tree-ring data play a central role in the variance increase found even in the RUNNINGr corrected Morocco mean-value function.

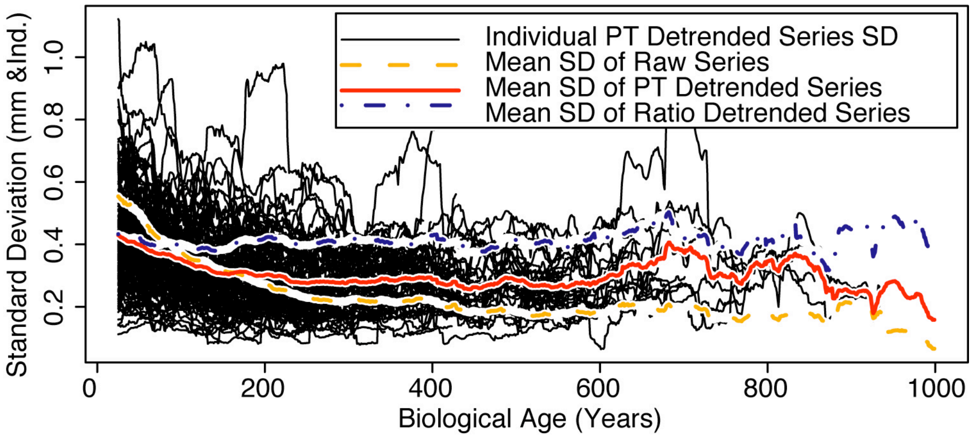

Figure 7: 50-year running standard deviations for the age-aligned individual series from the Morocco dataset, along with their mean. Also shown are mean standard deviations for the age-aligned raw data, and 300-year-spline ratio detrended data.

Discussion and conclusions

Using the synthetic data we demonstrated how an increase in variance can occur from both decreasing sample replication or increasing rbar. Both the MEANr and RUNNINGr variance stabilization methods mitigated variance inflation due to changing sample size, however only the RUNNINGr method can eliminate variance changes resulting from fluctuations in the interseries correlation.

However, with the real dataset, the situation turned out to be more complicated. Although we hypothesized, due to increasing interseries correlations back in time, that the RUNNINGr correction would eliminate the increase in variance found in the Morocco dataset after using the “Briffa/Osborn” correction method in ARSTAN, this was not the case. The RUNNINGr correction mitigated this variance inflation in comparison to the MEANr correction, however, the final RUNNINGr corrected chronology still showed a variance increase. This result was unexpected after tests with the synthetic data. The primary source of this variance increase appears to be that after detrending with a power-transformation and a 300-year spline the variance of the single series (as measured by 50-year running standard deviations) showed a biological age-related component. The greater variance at younger biological ages and the underlying age-structure of this dataset are likely to contribute to the variance trend for all computations and corrections of the mean chronology. Further exploration of this issue is required to more completely understand the origin of this tendency, and to determine if it is unique to this perhaps unusual dataset composed of multi-centennial to millennial length tree-ring series. More efforts to understand the characteristics of ring-width series that make them more or less susceptible to variance nonstationarities in general or after power-transformation are also required.

For brevity we have not digressed even further by showing figures for the Morocco dataset detrended with the ratios method. However, these results (not shown), indicate that the uncorrected chronology possesses a strong increase in variance back in time, which is greatly reduced if the MEANr correction is applied and essentially eliminated if the RUNNINGr correction is performed. If the only criteria for chronology development were the presence of a relatively constant variance, we could suggest at this point that the ratios detrending might be more suitable for this dataset. However, this dataset, with long-lived trees that often possess exceptionally narrow outer rings, is particularly susceptible to index value inflation from detrending curves that enter “the danger zone” (Cook and Peters, 1997). More efforts are needed to untangle costs and benefits in detrending with the power-transformation in comparison to ratios for this dataset, and to understand more generally if other tree-ring sites share these same attributes. For the Morocco dataset presented, with the methods tested, it appears that while the RUNNINGr correction perhaps performed best, the variance structure of the final chronology is still not completely optimal. Exploration of other variance stabilization methods, including an empirically based spline correction (which is also an option in ARSTAN) could be considered. In any case, we can recommend inspection of the individual series variances (perhaps after age-aligning) to determine if there are trends or artifacts that will influence the variance of the underlying chronology. More generally these same issues were seen to effect the variance corrections applied in developing regional means of tree-ring data, when the chronologies themselves were not computed with sample size corrections (Osborn et al. 1997). Similar considerations are easily relevant for the computation of hemispheric-scale temperature reconstructions.

From the synthetic cases, with more stable individual variances, the RUNNINGr correction produced results most consistent with the expectation of constant variance in the mean-value function. Only RUNNINGr types of corrections can mitigate variance artifacts resulting from temporally changing interseries correlations. However, when rbar fluctuates randomly, as in Case 1 presented above, it is possible “overkill” to apply the RUNNINGr correction, when the MEANr correction would perhaps be justified. Variance corrections similar to the RUNNINGr approach used here might be helpful in the computation of gridded or large scale instrumental mean datasets. These datasets not only rely upon fewer and fewer stations back in time, but the locations of the earliest remaining stations tend to be more concentrated in certain areas such as Central Europe or Eastern North America. The RUNNINGr correction applied herein seems to have an additional interesting attribute: during times with greater correlation the variance is diminished. This has implications for the reconstruction of climatologically extreme periods (e.g. the early 19^{th} century in Europe), which are probably not only imprinted on proxy series themselves, but also on the correlation of the individual series. This correction might help mitigate non-linearities resulting from strong pointer years.

It is an objective in dendroclimatology for the variance properties of the mean-value function to most closely represent those of the instrumental target data. In this regard, some success was met using a network to look at the variability of extreme events in the Alps (Frank et al. 2005). To better approach this condition, following the work of Osborn et al. (1997), we have shown herein that corrections for sample size are often needed. Additionally, we have shown examples in which the correction methods routinely applied are additionally complicated by time dependence of rbar and by variance non-stationarity of the raw data. The application or omission of variance adjustments to the mean-value function of time series often has non-negligible consequences. These considerations become critically important in paleoclimatology during the early periods of instrumental or proxy records when sample replication is lowest and spatial, biological, ecological, and sampling representativity are most likely to be unique.

ACKNOWLEDGEMENTS

We thank Kerstin Treydte for helpful suggestions and the Swiss National Science Foundation, Grant # 2100-066628 'Millennia' and NCCR-Climate for support.

REFERENCES

Briffa, K.R., Jones, P.D., Bartholin, T.S., Eckstein, D., Schweingruber, F.H., Karlen, W., Zetterberg, P., Eronen, M. (1992): Fennoscandian summers from AD 500: temperature changes on short and long timescales. *Climate Dynamics 7*: 111-119.

Cook, E.R. (1985): A time series analysis approach to tree-ring standardization. PhD dissertation, University of Arizona, Tucson, AZ.

Cook, E.R., Peters, K. (1997): Calculating unbiased tree-ring indices for the study of climatic and environmental change. *The Holocene 7*: 361-370.

Esper, J., Büntgen, U., Frank, D.C., Nievergelt, D., Treydte, K., Verstege, A. (in review) Multiple tree-ring parameters from Atlas cedar (Morocco) and their climatic signal. In: Heinrich I (Ed.) Tree rings in archaeology, climatology and ecology, TRACE, Vol. 4.

Esper, J., Cook, E.R., Schweingruber, F.H. (2002): Low-frequency signals in long tree-ring chronologies for reconstructing of past temperature variability. *Science 295*: 2250–2253.

Esper, J., Cook, E.R., Krusic, P.J., Peters, K., Schweingruber, F.H. (2003): Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. *Tree-Ring Research 59*: 81-98.

Frank, D., Wilson, R.J.S, Esper, J. (2005): Synchronous variability changes in Alpine temperature and tree-ring data over the last two centuries. *Boreas*, in press.

Jones, P.D., New, M., Parker, D.E., Martin, S., Rigor, I.G. (1999): Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics 37*: 173-199.

Osborn, T.J., Briffa, K.R., Jones, P.D. (1997): Adjusting variance for sample-size in tree-ring chronologies and other regional-mean time-series. *Dendrochronologia 15*: 89-99.

Wigley, T.M.L., Briffa K.R., Jones, P.D. (1984): On the average of correlated time series, with applications in dendroclimatology and hydrometeorology. *Journal of Climate and Applied Meteorology 23*: 201-213.