



Improved EEMD-based standardization method for developing long tree-ring chronologies

Xianliang Zhang^{1,2,3} · Junxia Li² · Xiaobo Liu² · Zhenju Chen²

Received: 26 September 2018 / Accepted: 7 March 2019 / Published online: 27 June 2019
© The Author(s) 2019

Abstract Long tree-ring chronologies can be developed by overlapping data from living trees with data from fossil trees through cross-dating. However, low-frequency climate signals are lost when standardizing tree-ring series due to the “segment length curse”. To alleviate the segment length curse and thus improve the standardization method for developing long tree-ring chronologies, here we first calculated a mean value for all the tree ring series by overlapping all of the tree ring series. The growth trend of the mean tree ring width (i.e., cumulated average growth trend of all the series) was determined using ensemble empirical mode decomposition. Then the chronology was developed by dividing the mean value by the growth trend

of the mean value. Our improved method alleviated the problem of trend distortion. Long-term signals were better preserved using the improved method than in previous detrending methods. The chronologies developed using the improved method were better correlated with climate than those developed using conservative methods. The improved standardization method alleviates trend distortion and retains more of the low-frequency climate signals.

Keywords Standardization methods · Tree rings · Segment length curse · Detrending

Introduction

Millennia-long tree-ring chronologies can be developed by overlapping tree ring series from living trees with dead wood, construction wood or sub-fossil material through cross-dating, extending the records of living trees back in time. The intrinsic growth trends of trees need to be removed from measurements of the width of tree rings to reflect the long-term climate signals recorded within the tree-ring series. In dendrochronology, the methods used to remove growth trends from raw measurements of tree rings are referred to as standardization methods and classified as determination or data adaptive methods. In determination methods, such as the linear regression line and modified negative exponential curve methods (Fritts et al. 1969), a previously defined mathematical model is used to describe the radial growth of trees. Data adaptive methods, such as the smoothing spline method (Cook and Peters 1981), fit the behavior of the observed tree ring width. However, climatic variations or cycles that occur at wavelengths longer than the length of tree ring series are not captured by standardized chronologies because they are identified as

Project funding: This work was funded by the National Natural Science Foundation of China (41601045, 31570632 and 41871027), and the National Key Research and Development Program of China (2017YFD060040301).

The online version is available at <http://www.springerlink.com>.

Corresponding editor: Zhu Hong.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11676-019-01002-y>) contains supplementary material, which is available to authorized users.

✉ Xianliang Zhang
zhxianliang85@gmail.com

✉ Zhenju Chen
zhenju.chen@hotmail.com

- ¹ College of Forestry, Hebei Agricultural University, Baoding 071001, People's Republic of China
- ² College of Forestry, Shenyang Agricultural University, Shenyang 110866, People's Republic of China
- ³ Long-term Forest Silviculture Research Station in Saihanba, Chengde 067000, People's Republic of China

part of the biologically based growth trend (Granger 1966; Cook et al. 1995). Hence, long-term climate signals are lost when developing millennia-long tree ring chronologies using these standardized methods as a result of the segment length problem (Cook et al. 1995).

Cook et al. (1995) described two ways to preserve long-term climate signals in millennia-long tree-ring chronologies. One is simply averaging all the raw ring-width measurements for trees without any age-related trend. For instance, tree-ring series from bristlecone pine, which has a strip-bark growth form, show no age-related growth trend (LaMarche 1974). However, the rarity of strip-bark growth forms has restricted the application of this method to other tree species. Another method is regional curve standardization (RCS; Briffa et al. 1992), although there are many deficiencies in this procedure, which have been addressed by Cook et al. (1995). A signal-free approach was developed to alleviate trend distortion at the end of chronologies (Melvin and Briffa 2008). However, its performance for preserving long-term climate signals is not known. A new method to alleviate the segment length problem and preserve long-term climate signals is thus urgently needed. In addition, part of high-frequency climate signals may be removed by traditional detrending methods. We aim to develop a new method to clearly show that only the lowest frequency signal that is identified as age trend is removed, while frequency signals that are identified as climate signals are retained.

Two authors of the present study, Zhang and Chen (2017), developed a standardization method based on ensemble empirical mode decomposition (EEMD) to precisely remove the tree growth trend. EEMD is an adaptive, data-driven algorithm used to decompose non-stationary signals into their intrinsic modes of oscillation and mean trends (Wu et al. 2011). The mean trend of tree-ring series identified by EEMD can be regarded as the growth trend of the series (Zhang and Chen 2017), but the segment length problem is not resolved by the EEMD-based detrending method of Zhang and Chen (2017). We therefore set out to improve the EEMD-based standardization method to alleviate the segment length problem.

Materials and method

EEMD method

Huang et al. (1998) developed the empirical mode decomposition (EMD) method to decompose non-linear and non-stationary time series into their intrinsic modes of oscillation. The EEMD method alleviates some of the common problems of the EMD method, such as mode mixing, and increases its robustness (Wu and Huang 2009).

The EEMD method is widely used as an effective decomposition method in geophysical studies (e.g., Wu et al. 2007, 2011; Huang and Wu 2008; Fang et al. 2013; Zhang and Yan 2014).

The EEMD procedure is implemented through a three-stage sifting process. (1) All the local extremes in the time series $x(t)$ are identified, and the upper (lower) envelope is obtained by connecting all the local maxima (minima) with a cubic spline. (2) The first component h_1 is calculated as the difference between the data $x(t)$ and the local mean of the upper and lower envelopes. (3) Treating h_1 as the original data, steps 1 and 2 are repeated until the upper and lower envelopes are symmetrical with respect to the zero-mean under certain criteria. The final h_1 is then designated as an intrinsic mode function (IMF₁), the residue of $x(t)$ and h_1 as a new time series, and a second sifting process is applied on the new series. The sifting process is completed when the residual r_n becomes a monotonic function. The time series $x(t)$ can be decomposed into a finite and often small number of IMFs based on the EEMD method:

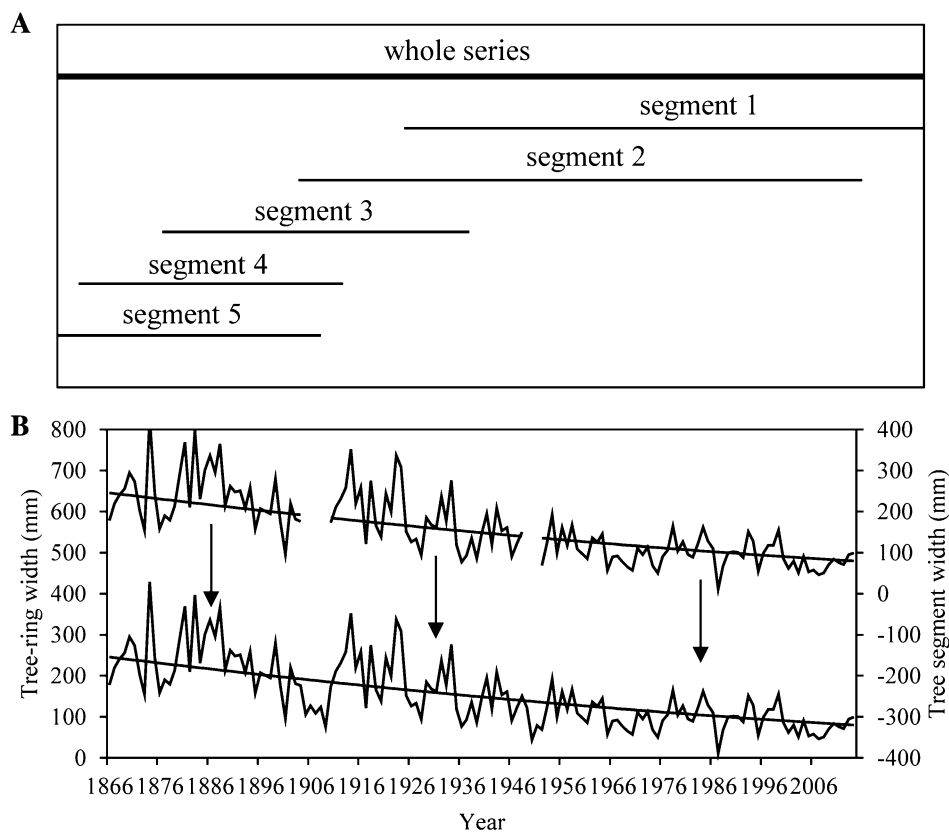
$$x(t) = \sum_{j=1}^n \text{IMF}_j + r_n, \quad (1)$$

where r_n is the residue of the time series $x(t)$ and reflects the mean trend of $x(t)$. Different IMFs represent high- to low-frequency oscillations of the data (Wu and Huang 2009).

A non-stationary time series can be regarded as a mixed signal series and can be decomposed into IMFs on different timescales. Each IMF, involving only one mode of oscillation, represents the intrinsic oscillation mode imbedded in the data. The mean trend remains when all the intrinsic oscillations have been extracted. For a tree-ring series, IMFs represent the external forces at different timescales. The expected growth trend of the tree can be identified when all the external signals have been extracted.

The EEMD method involves averaging numerous EMD runs with the addition of some Gaussian noise. The noise is averaged out by averaging the different decompositions and an estimate of the true decomposition is calculated, together with a confidence estimate. The ensemble number and the standard deviation of the added noise should be assigned in the EEMD procedure (Wu and Huang 2009). The sensitivity of the decomposition of the data to the amplitude of the noise is often small within a certain window of noise amplitude (Wu and Huang 2009). Added noise with a standard deviation of 0.2 and an ensemble size of 500 were used in this study.

Fig. 1 a The whole series is composed of different overlapping segments. **b** The growth trend of a segment should be a segment of the growth trend of the whole series



Improving the EEMD-based standardization method

Identification of growth trends by the EEMD method

The tree-ring width series could be decomposed by the EEMD method into n IMFs and mean trends (Zhang and Chen 2017):

$$x = \sum_{j=1}^n \text{IMF}_j + r_{nk}, \tag{2}$$

where x is the tree-ring width series, IMF_j is the intrinsic mode function and r_{nk} is the mean growth trend of the tree.

Growth trend of the mean value

The length of whole tree-ring series overlaps the length of each segment (Fig. 1a). We used letter G to represent the growth trend of a long tree-ring series, and letter S represents a

segment of this long tree-ring series. The growth trend of S should be a segment of G (Fig. 1b). In realistic examples, a long tree-ring series overlaps tree ring series from different periods. The growth trend of this long tree-ring series reflects the growth trend of trees across different periods. The growth trend of each segment is a segment of the growth trend of the long tree-ring series, and the growth trend of each series should vary at the same timescale with the growth trend of the long tree-ring series. If the growth trend of each series varies at the same timescale with that of the long tree-ring series (e.g. the growth trend of each short series varies at the same timescale with the growth trend of the long series in Fig. 1b), then theoretically, each series could be decomposed into the same number of IMFs as that of the long tree-ring series.

Suppose there are k tree cores, each tree ring series x_k can be decomposed into n IMFs and a mean trend r_n , then the mean value of K tree ring widths can be calculated as:

$$\begin{aligned} \bar{x} &= (x_1 + \dots + x_k)/k = \left[\left(\sum_{j=1}^n \text{IMF}_{j1} + r_{n1} \right) + \dots + \left(\sum_{j=1}^n \text{IMF}_{jk} + r_{nk} \right) \right] / k \\ &\cong \left(\sum_{j=1}^n \text{IMF}_{j1} + \dots + \sum_{j=1}^n \text{IMF}_{jk} \right) / k + \dots + (r_{n1} + \dots + r_{nk}) / k, \end{aligned} \tag{3}$$

where IMF_{jk} is the IMF_j of the k th core and r_{nk} is the mean trend of the k th core.

The mean value of K tree ring widths can be decomposed into n IMFs and a mean trend r_s :

$$\bar{x} = \sum_{s=1}^n IMF_{js} + r_s. \tag{4}$$

Combining Eqs. 3 and 4, we get:

$$\begin{aligned} & \left(\sum_{j=1}^n IMF_{j1} + \dots + \sum_{j=1}^n IMF_{jk} \right) / k + \dots + (r_{n1} + \dots \\ & + r_{nk}) / k \\ & \cong \sum_{j=1}^n IMF_{js} + r_s; \end{aligned} \tag{5}$$

hence, $r_s \cong (r_{n1} + \dots + r_{nk}) / k$ (6)

The mean of the growth trends of K cores were approximately equal to the mean trend r_s , identified for the mean value of k tree-ring widths using the EEMD method.

Development of chronology

The chronology was developed by dividing or subtracting the mean value of k tree-ring widths by the mean trend r_s .

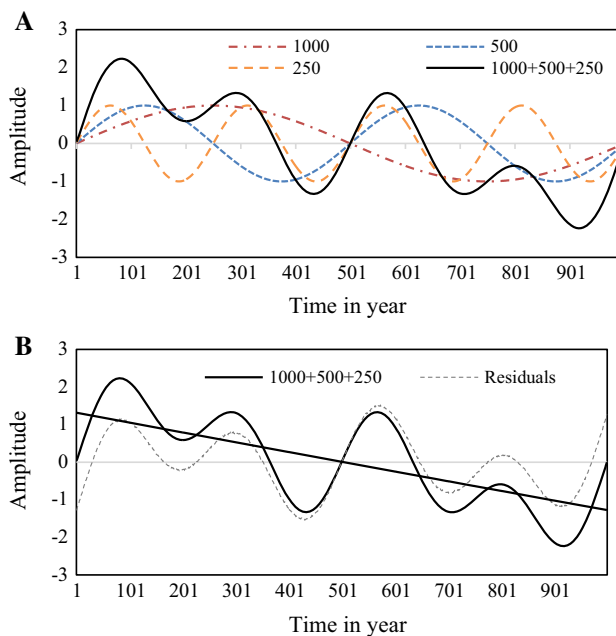
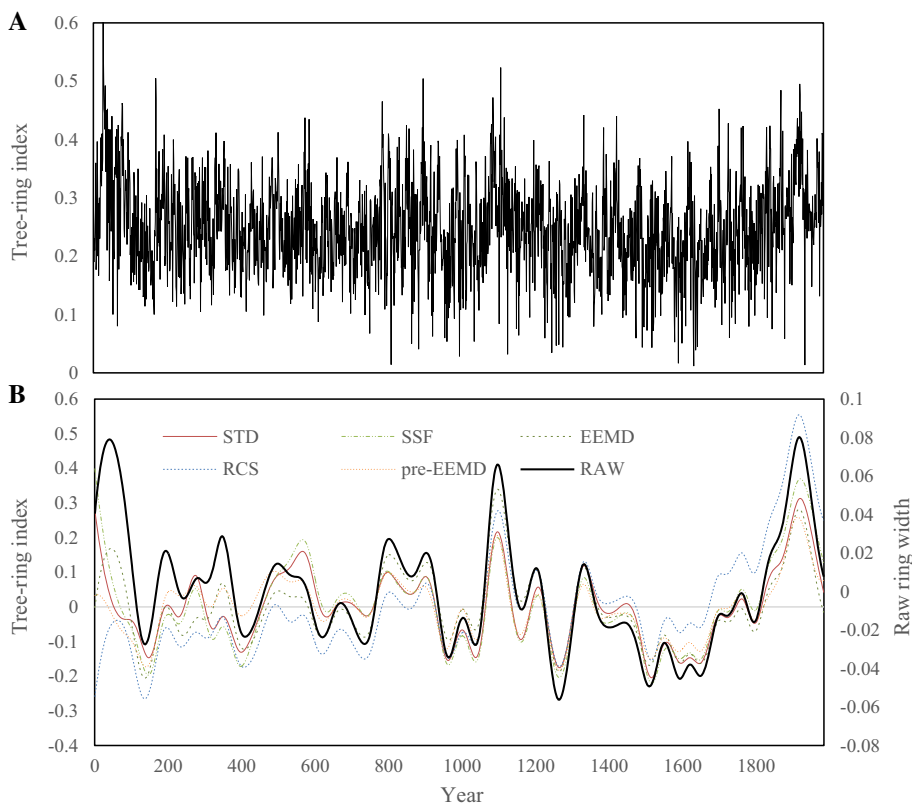


Fig. 2 **a** The segment length problem with sine waves: three pure sine waves with periods of 1000, 500 and 250 years and their combined series. **b** The trend identified with the ensemble empirical mode decomposition method for the combined series and the residuals calculated by removing the trend from the combined series

Fig. 3 The long bristlecone pine series used in the segment length experiment. **a** The series was calculated by averaging all the ring widths. **b** Growth trend identified with different methods: *EEMD* ensemble empirical mode decomposition, *pre-EEMD* previous version of our EEMD method, *RCS* regional curve standardization, *SSF* signal-free method, *STD* linear regression method



Illustrative examples of the preservation of long-term climate signals

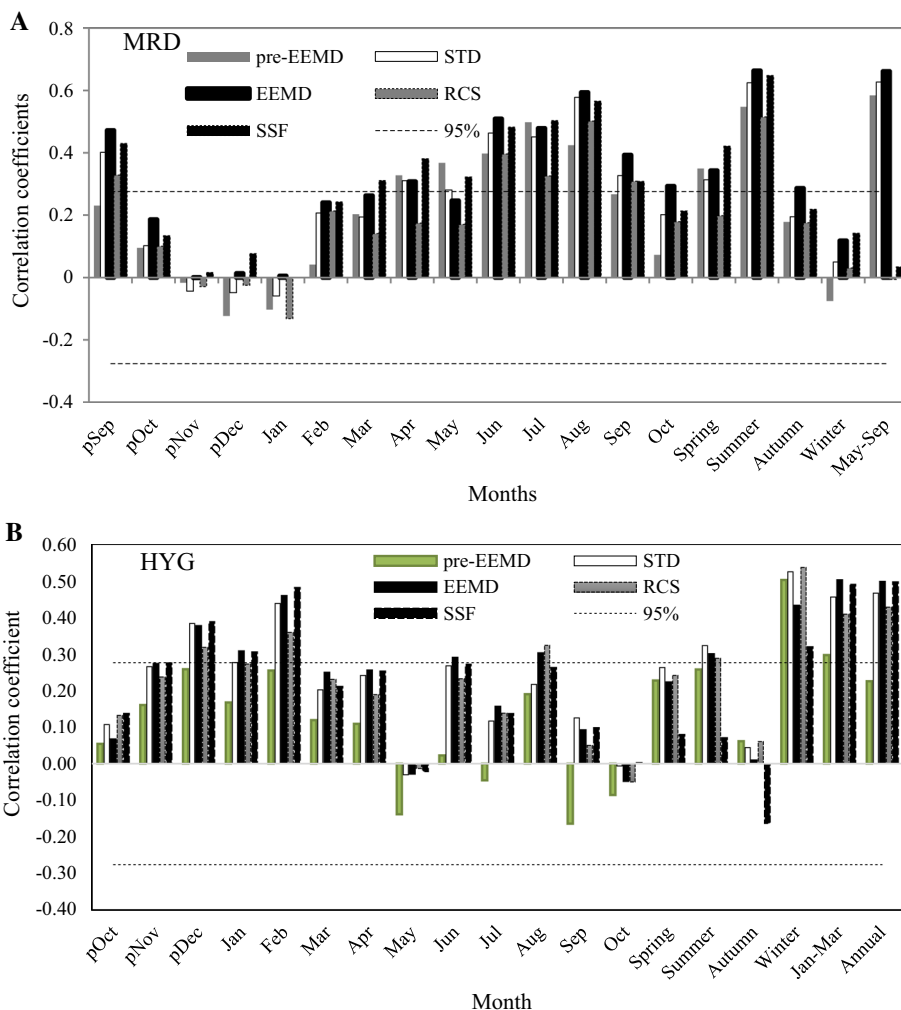
Cook et al. (1995) used three sine waves to illustrate the loss of low-frequency climate signals. The three sine waves, with periods of 1000, 500 and 250 years, had same amplitude (Fig. 2a) and each harmonic had a correlation of 0.577 with the combined series. However, most of 1000-year variance was removed when the whole length of the combined series was fitted by a linear regression curve. The correlation between the residues and the 1000-year harmonic was 0.059. However, the correlation increased to 0.1575 when our improved method was used (Fig. 2b); that is, the 1000-year variance was better preserved when our method was used than when the regression method was used.

The correlation between individual 500-year harmonics obtained with the improved method and the residual is 0.53 (Fig. 2b), which is close to the expected value (0.577) and much higher than that obtained with the linear regression method (0.439). The comparison was conducted when the curve-fitting method was used in ideal situations. Cook

et al. (1995) showed that the 1000-year variance was completely lost when the curve-fitting method was used for the three overlapping 500-year segments. The segment length is generally shorter than the whole length of the chronology. Low-frequency climate signals are difficult preserve with curve-fitting methods, and better results were achieved with our improved method.

We used a more realistic example to address the preservation of long-term climate signals using our improved method. Bristlecone pine has a strip-bark growth form and is assumed to have no age-related trend (Lamarche 1974; Cook et al. 1995). A bristlecone pine ring width series (nev516) was downloaded from the International Tree Ring Data Bank (www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring). The tree ring series was detrended using conservative methods (a linear regression method or modified negative exponential curve; STD), RCS, a signal-free method (SSF; Melvin and Briffa 2008), the previous version of our method (pre-EEMD) and our improved method (EEMD). The results showed that there was a loss of long-term signals with the

Fig. 4 Correlation between monthly mean temperature and five chronology types from the previous October to the current October and the seasonal combination for example sites **a** Moerdaoga (MRD) and **b** Haiyagou (HYG) in China. *EEMD* ensemble empirical mode decomposition, *pre-EEMD* previous version of our EEMD method, *RCS* regional curve standardization, *SSF* signal-free method, *STD* linear regression method



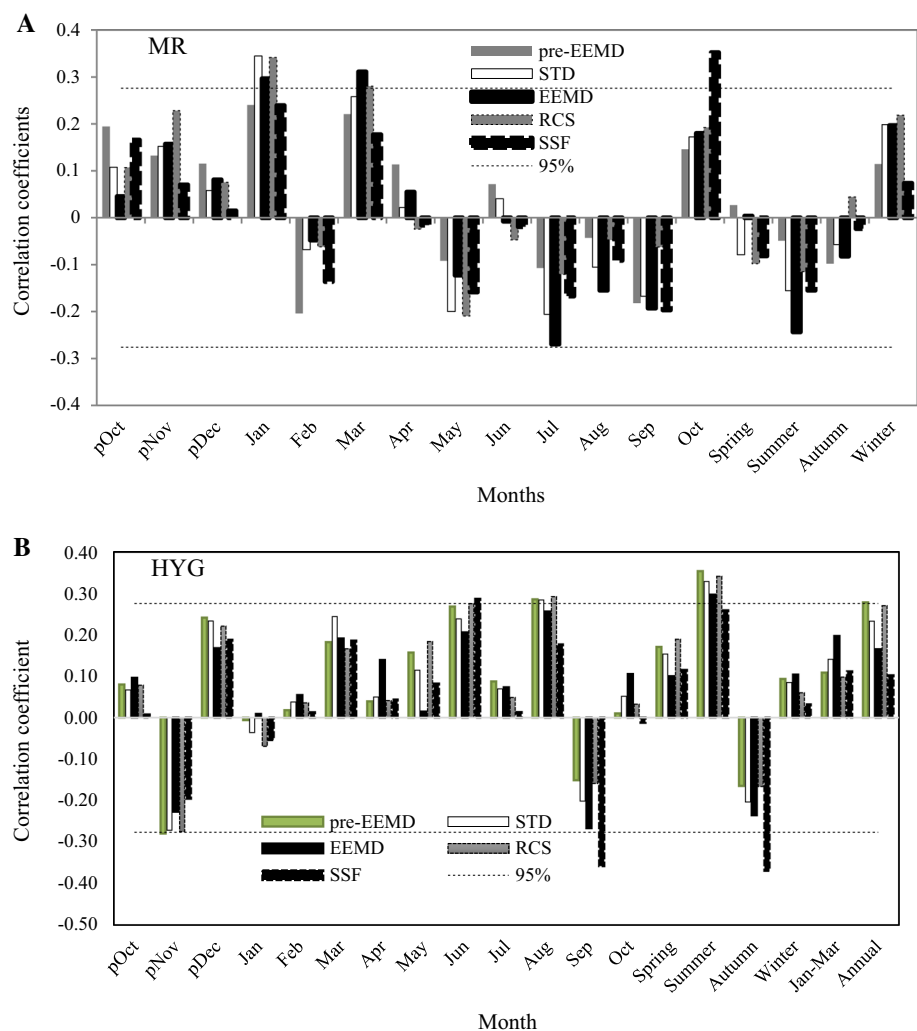
STD, SSF and pre-EEMD methods (Fig. 3). Low-frequency signals were well preserved with the RCS and EEMD methods. However, the long-term trend obtained with the EEMD method had a higher correlation ($r = 0.89$) with that of the raw series than the RCS method ($r = 0.76$). Therefore, the EEMD method showed the best performance among these methods in preserving long-term signals.

Long-term climate signals are the target of century- or millennia-long climate reconstructions based on correlation analysis. The correlation between climate signals and the outputted chronologies determine how the climate signals are reflected in the chronologies. Comparison among five detrending methods (STD, RCS, SSF, pre-EEMD and EEMD) is illustrated by two examples of tree-ring series. One tree-ring series from Moerdaoga (MRD) had previously been used in a temperature reconstruction (Zhang et al. 2011), and another tree-ring series from Haiyagou (HYG, chin070) was selected from the tree-ring data used in Yang et al. (2014) and Zhang et al. (2016). The

correlation was calculated between the site chronologies obtained with the five detrending methods and the climate records from nearby climate stations at each site. Although the correlations were similar among the chronologies obtained with different methods, the EEMD chronology had a higher correlation coefficient with the monthly temperature than the other chronologies in most months (Fig. 4). In addition, The EEMD chronology had good performances in preserving seasonal precipitation signals when compared with other chronologies (Fig. 5). The correlation coefficients between tree growth and climate variables should be at least higher than 0.6 to reconstruct past climate changes using tree-ring series. The improvement in correlation between tree growth and climate variable could make some climate reconstructions possible.

Long-term climate signals are seen as noise in some dendroecological studies (e.g., fire studies), and it is easy to remove long-term climate signals from the chronologies to obtain high-frequency signals. The tree-ring series from HYG (chin070) was used to illustrate how the raw series

Fig. 5 Correlation between monthly precipitation and five chronology types from the previous October to the current October and the seasonal combination for example sites **a** Moerdaoga (MRD) and **b** Haiyagou (HYG) in China. *EEMD* ensemble empirical mode decomposition, *pre-EEMD* previous version of our EEMD method, *RCS* regional curve standardization, *SSF* signal-free method, *STD* linear regression method



can be decomposed into signals at different frequencies (Fig. 6). IMFs at different timescales can be composed of long-term signals (Fig. 6b) and high-frequency signals (Fig. 6c). Different IMFs can also be used to detect the influence of climate oscillations on tree growth. For instance, Lo et al. (2016) used the correlation between IMFs and decadal oscillations in climate (e.g., the Pacific Decadal Oscillation and the El Niño Southern Oscillation) to show how climate influences tree growth at the decadal scale.

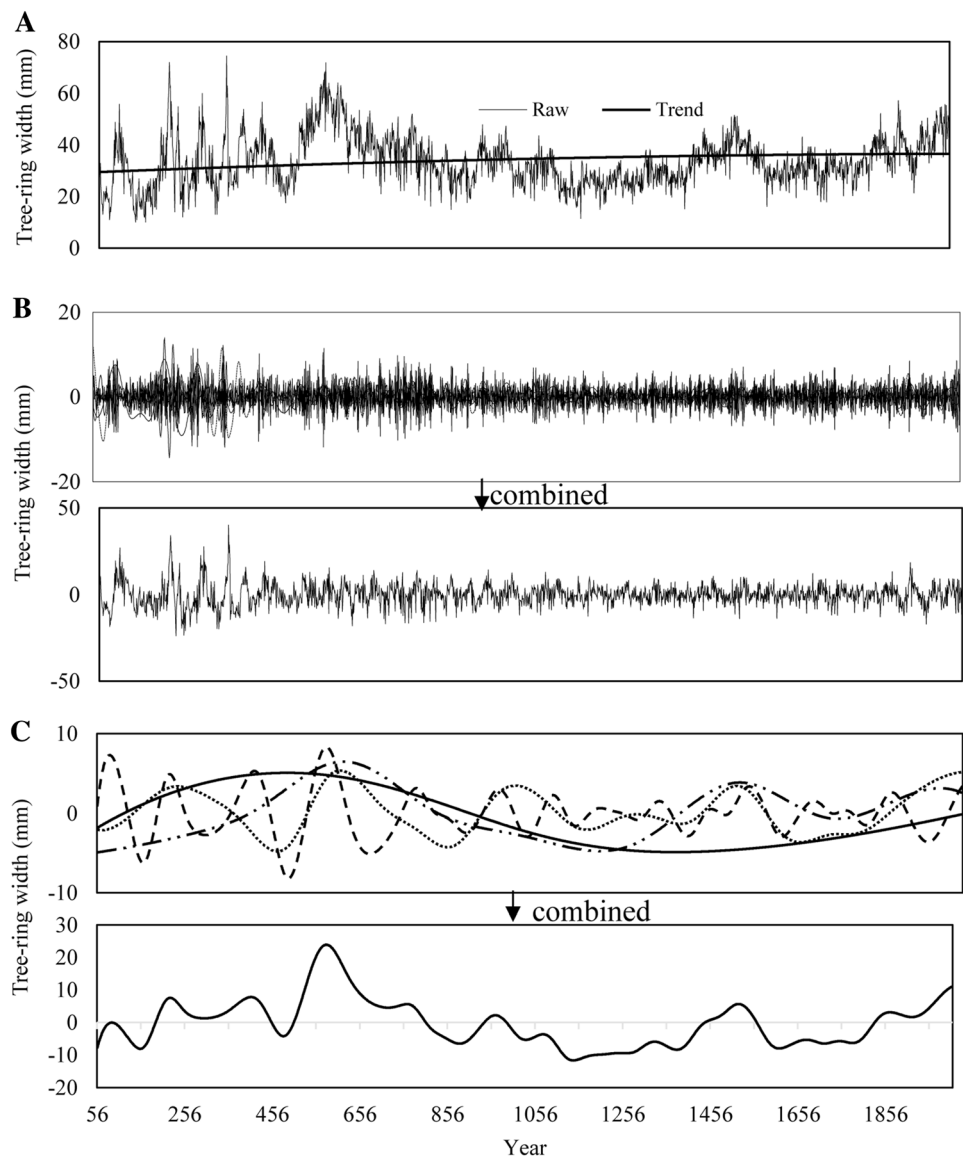
Discussion

The previous version of the EEMD-based standardization method was improved to alleviate the segment length problem when fitting curves to different lengths of tree

cores (Cook et al. 1995). In the commonly used curve-fitting methods, a curve is fitted to a tree ring series to identify the growth trend of the tree core. However, long-term climate signals at wavelengths longer than the segment length of tree ring series are lost due to the segment length problem (Cook et al. 1995).

The EEMD method has advantages in identifying intrinsic variations and mean trend (Huang and Wu 2008). The IMFs identified by the EEMD method are the external signals at different timescales. The mean trend has no intrinsic oscillation and varies on a longer timescale than the IMFs. Based on the improved standardization method, the trend for the mean of the raw measurements was obtained by integrating all the trends of the tree ring series. These trends are regarded as segments of the growth trend of the mean value of whole series. These trends can be averaged as the stacked mean trend because they have

Fig. 6 **a** Raw tree-ring series from Haiyagou (HYG, chin070) in China and its growth trend. **b** High-frequency signals of the raw series and their combined series. **c** Low-frequency signals of the raw series and their combined series. Each component not of interest can be easily removed



similar timescales. The stacked mean trend can be identified by the EEMD method because it varies on the longest timescale and can be separated from the IMFs.

The RCS method preserves long-term climate trends in long tree-ring chronologies (Esper et al. 2002, 2003), but problems such as pith offset, species composition and site composition influence the robustness of this method (Esper et al. 2003), and the segment length problem addressed by Cook et al. (1995). These problems do not appear in our improved method. Moreover, the mean trend identified by the EEMD method is not the regional common growth trend of trees identified in the RCS method. The mean trend cannot be identified by fitting a curve to the mean of the raw measurements because the mean trend should be a mixed signal of all the trends at similar timescales.

Our improved standardization method using EEMD, proposed to alleviate the segment length problem and to retain as much of the low-frequency signal as possible, is a simple method that can be easily applied in dendroclimatic research. A MATLAB program (EemdCrn.m, Appendix S1) and an R software procedure (EemdCrn.R, Appendix S2) to perform the analysis are available as supplementary materials. The codes were also available from the website https://www.researchgate.net/publication/331639377_Detrending_code_for_Matlab_and_R_user.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Briffa KR, Jones PD, Bartholin TS, Eckstein D, Schweingruber FH, Karlen W, Zetterberg P, Eronen M (1992) Fennoscandian summers from AD 500: temperature changes on short and long timescales. *Clim Dyn* 7:111–119
- Cook ER, Peters K (1981) The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. *Tree Ring Bull* 41:45–53
- Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G (1995) The ‘segment length curse’ in long tree-ring chronology development for palaeoclimatic studies. *Holocene* 5:229–237
- Esper J, Cook ER, Schweingruber FH (2002) Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* 295:2250–2253
- Esper J, Cook ER, Krusic PJ, Peters K, Schweingruber FH (2003) Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. *Tree Ring Res* 59(2):81–98
- Fang K, Frank D, Gou X, Liu C, Zhou F, Li J, Li Y (2013) Precipitation over the past four centuries in the Dieshan mountains as inferred from tree rings: an introduction to an HHT-based method. *Glob Planet Change* 107:109–118
- Fritts HC, Mosimann JE, Bortorff CP (1969) A revised computer program for standardizing tree-ring series. *Tree Ring Bull* 29:15–20
- Granger CW (1966) The typical spectral shape of an economic variable. *Econometrica* 34:150–161
- Huang NE, Wu Z (2008) A review on Hilbert–Huang transform: method and its applications to geophysical studies. *Rev Geophys* 46(2):RG2006. <https://doi.org/10.1029/2007rg000228>
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A* 454:903–995
- LaMarche VC (1974) Paleoclimatic inferences from long tree-ring records: intersite comparison shows climatic anomalies that may be linked to features of the general circulation. *Science* 183:1043–1048
- Lo Y, Blanco JA, Guan BT (2016) Douglas-fir radial growth in interior British Columbia can be linked to long-term oscillations in Pacific and Atlantic sea surface temperatures. *Can J For Res* 47:371–381
- Melvin TM, Briffa KR (2008) A “signal-free” approach to dendroclimatic standardisation. *Dendrochronologia* 26:71–86
- Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 1:1–41
- Wu Z, Huang NE, Long SR, Peng C (2007) On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc Natl Acad Sci USA* 104:14889–14894
- Wu Z, Huang NE, Wallace JM, Smoliak BV, Chen X (2011) On the time-varying trend in global-mean surface temperature. *Clim Dyn* 37:759–773
- Yang B, Qin C, Wang J, He M, Melvin TM, Osborn TJ, Briffa KR (2014) A 3,500-year tree-ring record of annual precipitation on the northeastern Tibetan Plateau. *Proc Natl Acad Sci USA* 111:2903–2908
- Zhang X, Chen Z (2017) A new method to remove the tree growth trend based on ensemble empirical mode decomposition. *Trees* 31:405–413
- Zhang X, Yan X (2014) A novel method to improve temperature simulations of general circulation models based on ensemble empirical mode decomposition and its application to multi-model ensembles. *Tellus A* 66:24846
- Zhang X, He X, Li J, Davi N, Chen Z, Cui M, Chen W, Li N (2011) Temperature reconstruction (1750–2008) from Dahurian larch tree-rings in an area subject to permafrost in Inner Mongolia, Northeast China. *Clim Res* 47:151–159
- Zhang X, Yan X, Chen Z (2016) Reconstruct regional mean climate with Bayesian model averaging: a case study for temperature reconstruction in the Yunnan–Guizhou Plateau, China. *J Clim* 29:5355–5361

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.