(Following the comments presented by the referees, we have added much explanation and clarification into Secs. 2.2,3,4.1,4.2 and 4.3 of our manuscript. We apologize that this has taken much time and our answer is delayed. The revised sections have been attached.)

(Our response is in the following given in italics)

Clim. Past Discuss., doi:10.5194/cp-2016-27-RC1, 2016

***Interactive comment on* "A universal error source in past climate estimates derived from tree rings" *by* Juhani Rinne et al.**

Anonymous Referee #1

Received and published: 20 April 2016

The manuscript touches upon an important topic within tree-ring research and climate reconstruction, namely sample replication and the ontogenetic or age trend present within tree-ring series (hence referred to as detrending methods). Especially, removing age-related trends while maintaining low-frequency climatic signals is of great inter- est to better understand earths climatic system. Both the fact that a new detrending methodology is presented to address this issue and a case study is provided on a widely used tree-ring record gives this manuscript great potential. However, the poor link to other detrending methods and work that has been done on other relevant tree- ring biases, weakens the message given by the manuscript although very strong state- ments are made. Additionally, for this method to be useful to the community more emphasis should be put on clearly explaining the steps needed to execute this de- trending technique. Because of this I would like to present a few points of discussion in relation to; a) inclusion of other detrending techniques, b) the impact of tree-ring related biases and c) expanding the methodology.

a) *inclusion of other detrending techniques,*

Due to both the title and the context of the paper it appears that the issues in relation to RCS deterending are present within most chronologies and has not been prop- erly addressed. However, recently a lot of work has been done on addressing and resolving detrending related issues, while these methods and sources are not being addressed or mentioned within this manuscript. This for instance includes work on the RCS related issues in Briffa & Melvin (2011 in Hughes et al. Dendroclimatology, Developments in Paleoenvironmental Research) or the work done on comparing multi- ple dentrending methods and their potential to maintain low-frequency signals (Peters et al. 2015 Global Change Biology). Additionally other detrending methods like the signal-free detrending should be considered as they show great potential in being less affected by age-related issues (see Melvin & Briffa 2008 Dendrochornologia).

It was also surprising to see the comparison between the proposed method and the recon- struction performed by Briffa 1992, while multiple comments and revisions have been made on this chronology (see: Melvin et al. 2013 Holocene; Matskovsky et al. 2014 Climate of the Past). The current comparison is therefore very difficult to interpret as it is not clear whether one can state that one is better than the other. Additionally, the fact that the difference between the methods is very small when the sample size is higher than 14 raises more debate on sample replication than on the relevance of the method. To validate the value of the newly proposed method it is therefore crucial to include more dentrending methods and more recent Torneträsk reconstructions.

With the aid of your comments, we will refer to articles on the newer developments. Our work is focused on the estimation of error variances and the error term studied depends only on the distribution of the measurements over the years and age classes. As far as no new trees are added to the sample, that distribution is independent of the reconstruction technique or data source. From this point of view our approach is a new one. The focus is in the error variances. It is sufficient that our reconstruction method works satisfactorily and therefore we see no need to validate our method with other methods.

- Explanatory text on the performance of our method will be included in the revised version of Sec. 4.3.:

"As far no new trees are added, the distribution of the measurements is not changed. Hence S^2 in Eq. (6) will be decreased but Var_{rel} in Eq. (7) is unchanged. The latter term describes the error source studied here and depends only on the distribution of the measurements. Accordingly, it is more generally valid and is independent of developments in the reconstruction methods and their age dependence functions."

- Comparison with Briffa et al. (1992) will be rewritten in the revised version of Sec. 4.1. :

“In Fig. 3a the corresponding anomalies of the Torneträsk reconstruction in Briffa et al. (1992) are compared with our results. The climatologically biased quasi cycle of ca. 350 years is well seen. Especially the recent years are similarly biased showing an underestimation of the temperature. Excluding this recent underestimation, the variants of reconstructions in Melvin et al. (2013, their Fig. 4) show a similar long-term structure. Importantly, the bias is not in the tree ring measurements but in the reconstructions. Hence the bias has essentially remained as such from Briffa et al. (1992) until the present time, in spite of the developments in reconstruction methods. Naturally there can be differences in details because of different computational approaches.“

- Fig. 4 from Melvin et al. 2013, showing different reconstructions, has been attached in the following

b) the impact of tree-ring related biases

Many biases and problem are present within tree-ring data. One of these includes the ontogenetic or age trend. However, many other biases are present like the persistent growth patterns or management related issues.

Different Torneträsk analyses show a similar long-term variation (as far no new trees are added to the sample). Westward and eastward from Torneträsk study area there are no corresponding long-term oscillations in the nearby climatic reconstructions. Such local deviations at Torneträsk (of the order of magnitude of the greenhouse warming) are climatologically impossible and therefore the Torneträsk analyses must contain a bias.

- *The detection of the bias will be explained in the revised version of Sec. 4.1.:*

“In Rinne et al. 2014 it was observed that both in the nearby oceanic (August SST, Norwegian Sea, Miettinen et al. 2012) and continental (Esper et al. 2012) temperature estimates of the long term oscillations clearly and similarly differ from those derived from the Torneträsk data. Such local anomalies are climatologically impossible and therefore the mutually similar long-term oscillations in the Torneträsk reconstructions contain a bias. Accordingly, those reconstructions are suitable for our error studies.

The differences observed are extreme being of the order of magnitude of the greenhouse warming. The long-term bias in

Torneträsk reconstructions is thus detected climatologically. In our computations we estimate the bias as the difference between the Torneträsk reconstruction and the corresponding Esper et al. (2012) reconstruction, the latter having a high number of trees"

Multiple of these biases have been addressed in literature and have been shown to affect low-frequency patterns, which is the main interest of this manuscript. However little attention is provided on either how these biases affect the proposed method or discussed how these could affect the low-frequency signals. Work on both sampling strategy (Nerhbass-Ahles et al. 2014 Global Change Biology) and multiple biases described in literature should be addressed as these are common problems in constructing chronologies (See: Brienen et al. 2012 Global Biogeochemical Cycles; Bowman et al. 2013 Trends in Plant Science; Groenendijk et al. 2015 Global Change Biology; Autin et al. 2015 Dendrochronologia).

The presented general theory is illustrated with the aid of the Torneträsk data. The error source introduced in our article describes the dependence on the distribution of the measurements over the years and age classes. There clearly are other error sources but they do not affect our case. The bias observed in the Torneträsk case is well explained by the error source studied, i.e. by data gaps and paucities in the data.

- The explanation of the bias with the aid of the error source studied will be given in the revised version of Sec. 4.2.:

"To illustrate the combined impact of different error terms, the degrees of freedom are computed with the aid of the approximate formula in Eq. (7) and are shown in Fig. 3c. In the beginning of the data, they grow slower than the number of measurements (Fig. 3c), show rapid changes thereafter and are often very low. Main minima and maxima of the DOF are indicated by full and open circles, respectively. Fig. 3b shows the bias estimate of the reconstruction with b_{max}=270. It is seen, that strong minima (maxima) of the DOF in Fig. 3c well predict following biased (unbiased) values in Fig. 3b and fit with the anomalies in reconstructions shown in Fig. 3a. "

c) expanding the methodology.

For reproducibility it is essential when introducing a new method that all steps and components within the procedure are clear. In general it is therefore important that enough attention is given to the methodology section, which currently is not sufficient to apply this method on other datasets. As an example more information should be provided on how the sensitivity analysis is performed to determine the upper limit of the age classes (bmax). Additionally, how one should determine the other required parameters to perform the computation is unclear (e.g. r1 and rn).

- Different ways to estimate b_{max} will be discussed in the last paragraph of the revised version of Sec. 4.3:

"There is one possibility in some specific cases to regulate the accuracy. Our method implies that an upper limit of the age classes is selected. The higher is the limit, the more measurements are included into the analysis. Simultaneously there will be more longer data gaps. It is to find a compromise between the opposite effects. An illustration is given in Fig. 1, where the cases of b_{max}=270 and b_{max}=370 are compared. In the Tornedalen case study, it was concluded to make use of b_{max}=370 in order to decrease the bias during recent years. That choice was motivated as the hatched line in Fig. 3a is closer to zero. Otherwise it is seen that the changes are weak and the selection of b_{max} is not decisive. If it wished to be made mechanically, a possibility would be to minimize Var_{ave} with respect to b_{max} in Eq. (7). In Esper et al. (2014), the upper limit was taken to be b_{max}=306."

- Determination of r_1 and r_n will be explicitly described in Sec. 2.2.:

"In order to get the average taken over all age classes, the contributions of the missing young and old age classes are needed. These can be estimated as follows. First estimates of measurements are interpolated for every age class where that is possible. Then years with values in age classes of b=20 and b=270 are selected. This makes it possible to compute yearly averages over all age classes of 20 thru 270. Next yearly averages over age classes of 21 thru 270 are computed. Generally this new average is smaller but the variation between the years is strong. By computing the mean and r.m.s.e. of the yearly differences between the averages with and without b=20 we get an estimate of the eliminated age class b=20. In the next step the yearly averages over age classes of 22 thru 270 are computed and used to estimate the impact of missing age classes of b=21 and b=22 to the average of all measurements between b=20 and b=270. The computations are continued by dropping

out more age classes. Similar approach is applied to estimate the impact of missing old age classes.

The estimation turned out to be more complicated if only very young age classes ($b1<\approx 20$) or most of the older age classes ($bn<30$) were missing. To keep the formulae simple, the linear approximation is extended to those cases, too.

The result is that the contributions of the missing young and old age classes to the average taken over all age classes can be estimated linearly by $M_{young}\approx r_1(b_1-1)$ and $M_{old}\approx r_n(b_n-b_{max})$, respectively, where $r_1\approx 0.000128$ and $r_n\approx 0.000170$. Note that M_{old} results in a negative correction. The accuracy of such corrections turned out to be low their variances being roughly $[r_1(b_1-1)]^2$ and $[r_n(b_n-b_{max})]^2$. The application of the corrections are illustrated for an individual year in Fig. 2 for $b_{max}=300$."

The determination of r_1 and r_n can be varied by applying different age classes. In any case the error variances are high and the values of r_1 and r_n cannot be determined accurately in the presence of data gaps and paucities.

What is also vague within this method is how the extrapolation affects the results (see Figure 2). If there is for instance a year with only a few young age classes a large proportion of the mean is determined by the extrapolated older classes, which are heavily dependent upon the assumption you make within this extrapolation procedure (which, if I understand correctly, in the manuscript is proposed as a negative linear relationship which can be highly debated). How these situations affect the method should be described or analysed.

The impact of the missing younger and older age classes is described by the error variances. In the case of data gaps and paucities, the error variances become high and the reconstruction will be less reliable in these instances

Finally, the error estimation and the representation of missing age classes in Figure 3 could be of great value to the community. Especially, visually showing where specific age-classes are lacking could help to detect specific biases and to disentangle whether low-frequency signals are caused by sample replication or climate. Because of this relevance I feel more attention should be given to the methods section on this as currently the description is highly condensed and in some parts not clear.

We wish that the clarifications and explanations to be added into the manuscript make the text clearer.

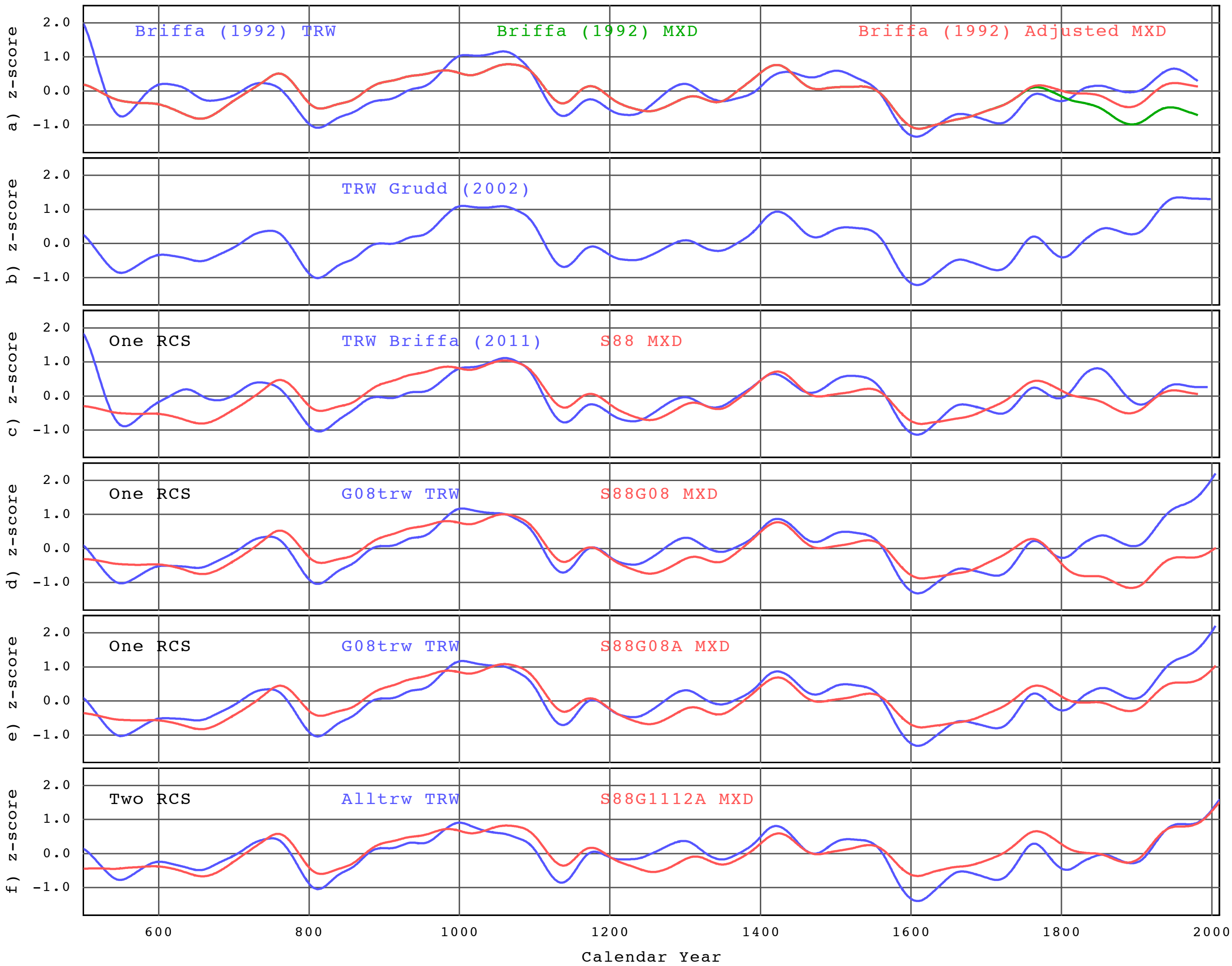
Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2016-27, 2016.

Attachement

Fig. 4 in Melvin et al. 2013 has been referred to in the manuscript. It is attached here (https://crudata.uea.ac.uk/cru/papers/melvin2012holocene/).

There are several reconstructions from several years (e.g. Briffa MXD 1992 red+green in the uppermost panel, Grudd TRW 2002 blue in the next one, Briffa TRW 2011 blue; recent ones in the three lower panels). All of them show similar long-term oscillations that are seen in Briffa 1992 MXD. It is seen that recent advanced reconstruction methods have not changed the situation.

Exceptions are only seen during recent years where the ”divergence” problem is best seen in Briffa 1992 MXD (green line). There are several attempts to correct the problem. No similar variations in the ”divergence” case (systematic underestimation of the temperature) after 1600 AD are visible. The cold bias is similarly in different reconstructions of the order of magnitude of the greenhouse warming.



Changes to the manuscript.

Contents

2.2 Explicit computation formulae

The estimation of parameters r_1 and r_n is described explicitly

3 Torneträsk case study

The original and corrected observations for 1826-38 are presented in Fig. 4. The motivation of the correction is presented in details.

4.1. Bias of the Torneträsk reconstructions

The climatological detection of the bias in the Torneträsk *reconstructions is presented more specifically.*

The use of Briffa (1992) reconstruction in Fig. 3a is explained in more detail.

4.2. Explanation of the bias observed

It is pointed out that the Torneträsk case is only an illustration and application of the general error terms in Eqs. (4) and (5).

The bias in the reconstructions is, on the basis of the description in Sec. 4.1, estimated with the aid of the reconstruction in Esper et al. (2012).

4.3 The performance of the reconstruction method applied

The performance of the reconstruction is described only to show that the error analysis is based on a sound calculation. The resulting temperature estimates are sufficiently satisfactory.

It is explained that the reconstruction methods published in the literature do not impact the error source studied here and are thus outside of the scope of our study.

The selection of the upper limit of the age classes is discussed.

2.2 Explicit computation formulae

(page 4, lines 9-23, the older version of the following new paragraph began with "Assume ...")

In order to get the average taken over all age classes, the contributions of the missing young and old age classes are needed. These can be estimated as follows. First estimates of measurements are interpolated for every age class where that is possible. Then years with values in age classes of b=20 and b=270 are selected. This makes it possible to compute yearly averages over all age classes of 20 thru 270. Next yearly averages over age classes of 21 thru 270 are computed. Generally this new average is smaller but the variation between the years is strong. By computing the mean and r.m.s.e. of the yearly differences between the averages with and without b=20 we get an estimate of the eliminated age class b=20. In the next step the yearly averages over age classes of 22 thru 270 are computed and used to estimate the impact of missing age classes of b=21 and b=22 to the average of all measurements between b=20 and b=270. The computations are continued by dropping out more age classes. Similar approach is applied to estimate the impact of missing old age classes.

The estimation turned out to be more complicated if only very young age classes ($b1$<≈20) or most of the older age classes (b_n<30) were missing. To keep the formulae simple, the linear approximation is extended to those cases, too.

The result is that the contributions of the missing young and old age classes to the average taken over all age classes can be estimated linearly by $M_{young} \approx r_1(b_1-1)$ and $M_{old} \approx r_n(b_n-b_{max})$, respectively, where $r_1 \approx 0.000128$ and $r_n \approx 0.000170$. Note that M_{old} results in a negative correction. The accuracy of such corrections turned out to be low their variances being roughly $[r_1(b_1-1)]^2$ and $[r_n(b_n-b_{max})]^2$. The application of the corrections are illustrated for an individual year in Fig. 2 for b_{max}=300.

The focus here is in the long-term error. Important are the estimates of the error variances due to the missing age classes. Precise values of r_1 and r_n are not needed. The main information included in the results is the dependence of the variance terms on the lengths of the data gaps, b_1-1 and $b_{max}-b_n$. Our reconstruction method is directed to uncover that dependence explicitly in a simple way. Otherwise it is enough that the reconstruction method performs sufficiently satisfactorily. It needs not be an optimal one.

No assumption on the age function is made. The contributions of the young and old age classes can widely vary between years as indicated by their error variances. However, average corrections are needed in order to decrease the impact of systematic errors. By adding the corrections, the yearly average of the interpolated, extrapolated and measured values becomes approximately **…**

3 Torneträsk case study

(page 7, lines 1-6, the changes are given in blue)

Klingberj and Moberg (2003) composited a series of instrumental observations made in the Tornedalen region, some 300 km from Torneträsk. Their construction begins with instrumental data from

Övertorneå (1802-1838). The observation hours were not stated explicitly for 1826-38 and the authors had to assume them. The JJA temperatures are known to be sensitive to the observation hours used. If these are not precisely known, true climatic variations may remain but the average level of the temperature may become wrong. In such cases any kind of support is welcome. During 1826-38 the bias estimate in our reconstructions is rather small (≈ -0.25°C, Fig. 3a) and the DOF are rather high (≈11, Fig. 3c). Our reconstruction can therefore be applicable. If the JJA temperatures in observations are systematically increased by 1°C every year during 1826-38, the smoothed result happens to fit rather well with our reconstruction in Fig. 4. Accordingly, the corrected temperatures are supported by the reconstruction and the anomalous coldness in observations before 1840 in Fig. 4 seems suspicious due to the unknown observation hours. The comparison with instrumental data here - while interesting - is not essential, moreover as in any case temperatures before ca. 1850 may contain biases (Melvin et al., 2013; Grudd, 2008).

(corrected and original observations are now shown in Fig.4)

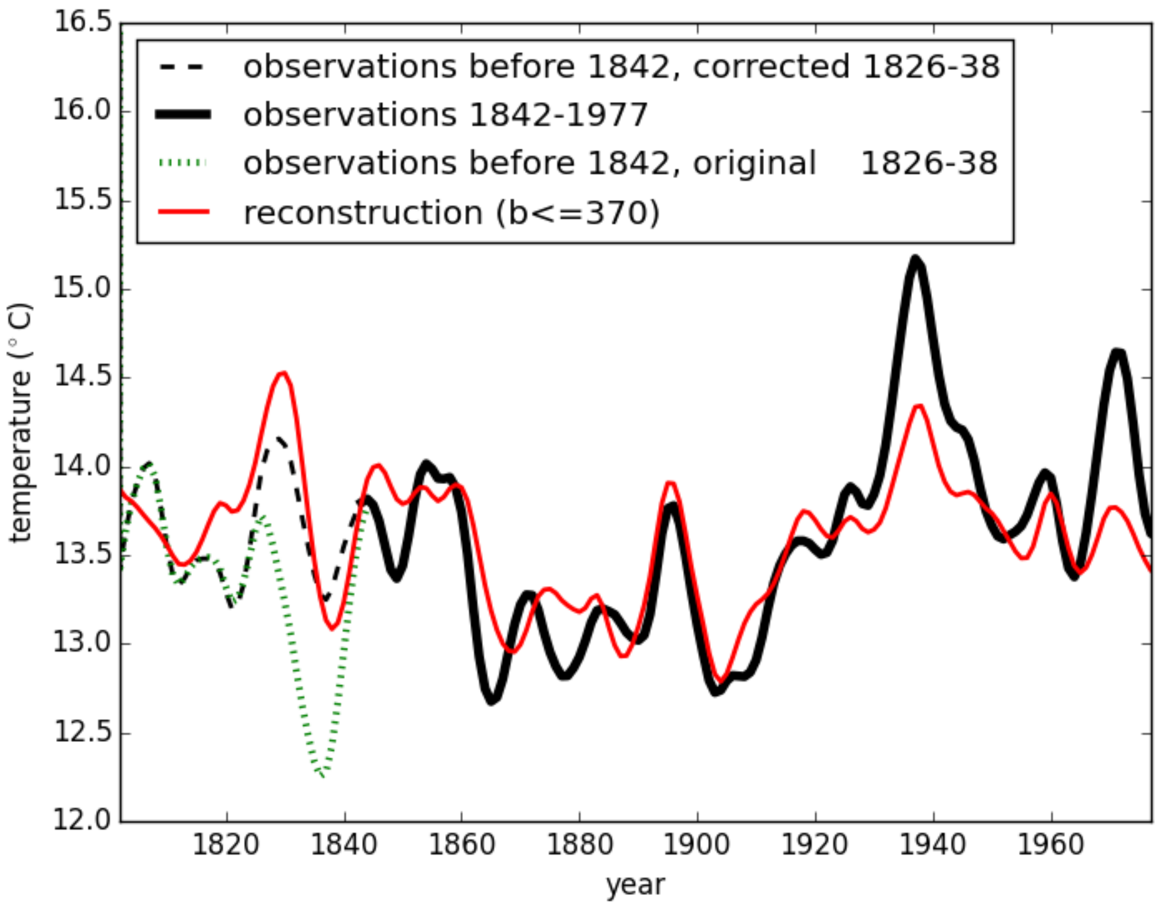


Figure 4. Comparison of Torneträsk temperature reconstruction and Tornedalen temperature observations 1802-1977. Reconstruction mean has been adjusted to fit the corresponding observational mean during 1850-1950 and both curves have been smoothed.

(revised version of the discussion, page 7, line 17 – page 9, line 7)

4.1. Bias of the Torneträsk reconstructions

In Rinne et al. 2014 it was observed that both in the nearby oceanic (August SST, Norwegian Sea, Miettinen et al. 2012) and continental (Esper et al. 2012) temperature estimates of the long term oscillations clearly and similarly differ from those derived from the Torneträsk data. Such local anomalies are climatologically impossible and therefore the mutually similar long-term oscillations in the Torneträsk reconstructions contain a bias. Accordingly, those reconstructions are suitable for our error studies.

The differences observed are extreme being of the order of magnitude of the greenhouse warming. The long-term bias in Torneträsk reconstructions is thus detected climatologically. In our computations we estimate that climatological bias as the difference between the Torneträsk reconstruction and the corresponding Esper et al. (2012) reconstruction, the latter having a high number of trees. In Fig. 3a the corresponding anomalies of the Torneträsk reconstruction in Briffa et al. (1992) are compared with our results. The climatologically biased quasi cycle of ca. 350 years is well seen. Especially the recent years are similarly biased showing an underestimation of the temperature. Excluding this recent underestimation, the variants of reconstructions in Melvin et al. (2013, their Fig. 4) show a similar long-term structure. Importantly, the bias is not in the tree ring measurements but in the reconstructions. Hence the bias has essentially remained as such from Briffa et al. (1992) until the present time, in spite of the developments in reconstruction methods. Naturally there can be differences in details because of different computational approaches.

4.2. Explanation of the bias observed

The error formulae derived are of a general form (Eqs. 4 and 5).

The error variances depend on the relative error terms and the dimensional portions $(b_{max}\, r_1)^2$, $(b_{max}\, r_n)^2$ and S^2. Because the relative error terms depend only on the distribution of the measurements over the age classes and years, they can be applied to any data without knowing the measurements or reconstruction method used.

In Fig. 3 the relative variances are applied for the Torneträsk data with 65 trees. Paucities in the tree population or periods of missing trees (Fig. 1) cause successively data gaps in young, intermediate and old age classes leading to variance variations in Fig. 3d. The increased variance decrease the degrees of freedom and so makes the bias more probable, which is then seen as erroneous long-term oscillations in the reconstructions.

To illustrate the combined impact of different error terms, the degrees of freedom are computed with the aid of the approximate formula in Eq. (7) and are shown in Fig. 3c. In the beginning of the data, they grow slower than the number of measurements (Fig. 3c), show rapid changes thereafter and are often very low. Main minima and maxima of the DOF are indicated by full and open circles, respectively. Fig. 3b shows the bias estimate of the reconstruction with b_{max}=270. It is seen, that strong minima (maxima) of the DOF in Fig. 3c well predict following biased (unbiased) values in Fig. 3b and fit with the anomalies in reconstructions shown in Fig. 3a.

Low values of the DOF (<5) are related to extrema of the bias. High values of the DOF (>14) indicate vanishing bias. The latter ones are not numerous and therefore bias-free cases are seen only temporarily in Fig. 3b.

One year with low DOF is separately indicated in Fig. 3. A change in the tree population is seen in 1356 AD (Fig. 1) where the oldest age class (<270 years) drops suddenly down to b=116 years. As a consequence, variances of the older and intermediate age classes are peaked in Fig. 3d, further lowering strongly the DOF in Fig. 3c. This is reflected in the bias (Fig. 3b) and in the reconstruction (Fig. 3a, smoothed values in Figs. 3a and 3b naturally lag the sudden changes in Figs. 3c and 3d). Correspondingly the optimal distribution of the measurements around 869 and 1500 AD (Fig. 1) are reflected in low error variances (Fig. 3d), high number of the DOF (Fig. 3c) and unbiased temperature estimates (Fig. 3b).

In Fig. 1, there is a repetitive similarity between the two longer periods with nearly continuous samples of trees (AD 1250-1500 and

1600-1800) and two periods without new trees (AD 1500-1600 and 1800-1980). The corresponding terms of the relative variance (Eq. 5) will be discussed in the following.

The large first peaks of variance around 1356 AD (Fig. 3d) result from missing intermediate and old age classes. These damp slowly out and new peaks are seen at 1600 AD. Here they are due to the intermediate (interpolated) and young age classes. The same structure is repeated between 1600-1800 AD and 1800-1980 AD. In this way the paucities due to missing trees in the data cause alternating data gaps leading to varying behavior in temperature reconstructions. Especially biases around 1600 and 1950 can be explained by the error terms connected to the missing younger age classes (due to the missing trees). The latter case is known as “divergence”, a systematic underestimation of the temperatures (Briffa et al. 1998, D’Arrigo et al. 2008). As there is no formal difference between the cases of 1600 and 1950, it is natural to refer to "divergence" in both cases.

4.3 The performance of the reconstruction method applied

The reconstruction method has been designed for estimation of the error variances. A general form of variances is found in Eq. (5). This is applicable as well for tree ring widths as maximum latewood densities. Only three parameters are needed in explicit applications (r_1, S, r_n). Error sources are not corrected but retained in order to study them. For instance, the long-term erroneous oscillations in Briffa 1992 are well reproduced in Fig. 3a during recent years, too. As a summary, the reconstruction method performs as it should.

The reconstruction method gives two values for every year, an average and its error estimate. If the parameters (r_1, S, r_n) were known, yearly confidence limits of the reconstruction could be given. An important part here is the variances due to missing young and old age classes.

To illustrate, the Torneträsk data is applied. Figs. 3a and 4 show that the reconstruction method performs sufficiently satisfactorily and so the error estimates can be seen to be reasonable. On the other hand, the relative errors due to data gaps and paucities are seldom low (e.g. years 869 and 1500 in Fig. 3d). Therefore the reconstruction can be expected to be inaccurate during most years.

If the errors due to the distribution of the measurements are not taken into account, in Eq. (6) Var_{rel}=1. Conventionally the sample accuracy is then characterized by presenting the yearly number of measurements. This practise is followed here. Instead of trying to estimate confidence limits, the sample accuracy is characterized by giving the degrees of freedom (Fig. 3c). The practical application requires that the general error terms in Eq. (5) are approximated and compressed into Eqs. (6) and (7).

In the recent literature new methods to estimate the age dependence function more precisely are introduced (e.g. Briffa et al. 2013; Melvin et al. 2013; Matskovsky and Helama 2014). They improve the reconstruction and the errors in the long-term oscillations may be decreased. Potential error sources in the sampling technique have been detected (e.g. Bowman et al. 2013). As far no new trees are added, the distribution of the measurements is not changed. Hence S^2 in Eq. (6) will be decreased but Var_{rel} in Eq. (7) is unchanged. The latter term describes the error source studied here and depends only on the distribution of the measurements. Accordingly, it is more generally valid and is independent of developments in the reconstruction methods and their age dependence functions.

In our reconstruction method the missing younger and older age classes are taken into account. It is natural that the results resemble to those that use age dependence functions. However, here the impact of the missing younger and older age classes is estimated directly from the measurements and only for the average impact of the missing age classes. The resulting estimates of error variances are high in the cases of data gaps and paucities. Therefore there is here no need for a more advanced reconstruction method because necessarily the error estimates will be high and the estimated impact of missing age classes less accurate.

There is one possibility in some specific cases to regulate the accuracy. Our method implies that an upper limit of the age classes is selected. The higher is the limit, the more measurements are included into the analysis. Simultaneously there will be more longer data gaps. It is to find a compromise between the opposite effects. An illustration is given in Fig. 1, where the cases of b_{max}=270 and b_{max}=370 are compared. In the Tornedalen case study, it was concluded to make use of b_{max}=370 in order to decrease the bias during recent years. That choice was motivated as the hatched line in Fig. 3a is closer to zero. Otherwise it is seen that the changes are weak and the selection of b_{max} is not decisive. If it wished to be made mechanically, a possibility would be to minimize Var_{ave} with respect to b_{max} in Eq. (7). In Esper et al. (2014), the upper limit was taken to be b_{max}=306.