Clim. Past Discuss., 9, C2139–C2155, 2013
www.clim-past-discuss.net/9/C2139/2013/

Climate of the Past
Discussions

CPD
9, C2139–C2155, 2013

Interactive Comment

Interactive comment on "A likelihood perspective on tree-ring standardization: eliminating modern sample bias" *by* J. Cecile et al.

J Cecile

jacobcecile@gmail.com

Received and published: 25 September 2013

We would like to thank Dr. Tom Melvin for his prompt and helpful comments.

We agree with his main suggestion that examples using real chronologies would be extremely helpful in understanding the working of the techniques. The next version of the paper will include basic re-analyses of both the Torneträsk and Yamal chronologies (standardization only, no climate links) with clear, step-by-step explanations to guide the readers through the process.

Most of the remaining comments debate terminology and some of the technical details of the approach taken. Clarifying these terms and details will greatly improve the paper. A number of the points raised are particularly interesting:

1. How do we judge the performance of a standardization technique when the true time signal is unknown? 2. Do dendroclimatologists actually care about the mean tree growth in each year? 3. How do we know which transformation to apply to our tree-ring data? 4. What do we do with missing ("zero-width") rings?

We feel that the modelling perspective offers useful guidance on these issues but suspect that these questions in particular will remain contentious for many years to come.

Dr. Tom Melvin's original comments are reproduced in full below for the convenience of the reader, with our detailed replies in bold.

Sincerely, Jacob Cecile, Chris Pagnutti and Madhur Anand

Interactive comment on "A likelihood perspective on tree-ring standardization: eliminating modern sample bias" by J. Cecile et al. T.M. Melvin (Referee) t.m.melvin@uea.ac.uk Received and published: 30 August 2013

General comments: The title of this paper claims to eliminate "modern sample bias". Modern sample bias arises from sampling trees at one point in time (e.g. cores are taken from living trees) in a situation where the mean growth rates of trees vary independently of climate forcing. The elimination of "modern sample bias" from RCS chronologies would be extremely useful.

To be of value to dendroclimatology this elimination has to be achieved while generating a chronology that includes the long-timescale variance preserved by RCS (contained in the changing index series means over time). The authors do not demonstrate that this basic requirement is met by their "fixed effects standardization". This omission needs to be corrected before the paper is suitable for publication. The paper will be much improved if the authors demonstrate the value of their approach using practical examples preferably based on published chronologies (e.g. Torneträsk or Yamal).

Curve-fitting standardization methods generate series of tree indices for each tree with the same mean value (usually 1.0). This rescaling removes the effects of "differing contemporaneous growth rates" from each tree and removes the mean growth rate of trees over time from the chronology. This effectively high-pass filters the chronology and is a serious problem where the preservation of long-timescale tree growth information is required. Curve-fitting methods eliminate any differences in mean-tree growth rates and thus do not suffer from "modern sample bias".

The RCS method was introduced to try to overcome the high-pass filtering problem inherent in curve-fitting methods. RCS retains the relative growth rates of individual trees in series of tree indices. The presumption is that the variation in tree growth rates referred to as "differing contemporaneous growth rates" is random and that having sufficient samples in each year ensures that the means for each year (chronology values) provide an unbiased estimation of variation in the effects of climate forcing on tree growth over time at all timescales.

The ability of the proposed methods to evaluate the difference between additive and multiplicative models is useful. The evaluation of the signal-free method as accounting for an "unbalanced experimental design" again seems to be useful. The proposed methods appear to be flexible and as such may have considerable potential to improve dendroclimatic methods. The value of these methods to dendroclimatology needs to be evaluated in the context of creating chronologies representing tree-growth rates over time. This manuscript needs to demonstrate that the proposed new method can produce chronologies representing tree growth over time.

Detailed Comments: Abstract: L10 "Including a term for the productivity of each tree accounts for the underlying cause of modern sample bias, allowing for more reliable reconstruction of low frequency variability in tree growth" - but the mean productivity of all trees in each year is vital to the reconstruction of low-frequency variability (see 4th paragraph above).

This is an extremely important and interesting point. We would argue that dendroclimatologists don't actually care about the actual observed growth of the trees in each

CPD
9, C2139–C2155, 2013

year. For if they did, standardization wouldn't exist.

The goal in standardization is not to capture the mean growth of the trees each year (because if so, we could just take a mean) but to make statements about the "typical" growth of trees in that year. If we draw up a very basic conceptual diagram (see attached) of tree-ring growth this becomes clear. The observed growth / ring width data contains all sorts of non-climatic information (microsite quality, ecological events, allometric patterns) that, while very interesting, are usually irrelevant to our climate-focused research question. Because we don't have enough detailed information about the life of these trees, we can't use a real process-based model and first need to segregate the variability in the ring width data according to its source: individual-related, time-related or age-related. This is the process of standardization and the traditional outputs (regional curves and standardized chronologies) are really just latent (abstract intermediate) variables. These help us organize our thoughts and allow us to introduce uncertainty (both noise and modelling uncertainty) at two distinct stages: when the common time / age / individual effects merge together to create the observed growth and in the relationship between the time effect (signal) and the climatic parameters of interest.

If you fit your model correctly (using signal-free optimization, least squares, maximum likelihood etc.), you can fully separate the effect of the inherent productivity of the trees from the time effect because the series will have substantial overlap, allowing the model to distinguish large ring-widths due to a favourable microsite from large ring-widths due to a favourable climate.

L20 “regional curve standardization is improved” – to make this claim you need to show an improvement to chronologies not an improvement to the model fitting. Also you need to do this using an appropriate data set for which RCS was designed (e.g. containing some sub-fossil material). This conclusion is not demonstrably supported by the results shown in this manuscript.

CPD
9, C2139–C2155, 2013

We agree that showing the techniques on a few sub-fossil chronologies would be helpful and will be including two in the next version of the manuscript. See our comment below on the difficulties in assessing improvement in the chronologies directly.

L22 “modern sample bias produced a significant negative bias” the authors have not shown the slope difference is caused by modern sample bias so the validity of this conclusion is not demonstrated.

Modern sample bias arises when the following two conditions are met: 1. There are persistent differences in growth between series that can’t be explained on the basis of time or age (differing contemporaneous growth rates). 2. These persistent differences are non-randomly distributed across time or age within the chronology.

This is "modern sample bias" in the most general sense, caused by the ecology (e.g. slow-grower-survivorship bias) or sampling design (e.g big-tree-selection bias) of the study.

Because the "individual effect" accounts for the differing contemporaneous growth rate biases, a model that includes it cannot suffer from "differing-contemporaneous-growth-rate-bias". As a result, condition 1 is nullified and modern sample bias disappears.

The only differences between the two models (G=ITA, the full model and G=TA, the regional curve standardization model) is the inclusion of an individual effect, all changes in the estimated time effect must be due to differing contemporaneous growth rates and any long-term bias in slope must be due to heterogeneity in the inherent productivities of the series (modern sample bias).

P4501 L12 – The “slow-grower survivorship bias” affects the shape of the RCS curve but this shape change is not directly relevant to “modern sample bias”. (Multiple growth rate based RCS curves can resolve the shape of the growth curves for different growth rates and will enable estimation of the magnitude of this effect). One effect of “slow grower survivorship bias” is to produce a forest of living trees which, although it represents the mean growth rates of trees at the current time, is missing samples (specifically the fast growing trees from earlier time which became large and died) such that rings which grew in earlier periods taken from these living trees do not represent the mean growth rate of trees in the years when those rings grew. Thus sampling living trees produces systematic bias in the earliest years represented.

As we discuss above, modern sample bias (in the general sense) is just the result of heterogeneity in the intrinsic productivities of the series. Slow-grower-survivorship bias is just one of the possible causes of this heterogeneity, through the mechanism you described in your comment.

Assuming proper model estimation, the distortion in the age effect is exactly mirrored by the distortion in the time effect when the chronology is purely modern. Overestimating the regional curve for the old trees causes an underestimate of the standardized chronology in its earliest sections. Shifting to a more staggered chronology design alleviates this problem as the distortion in the age effect is spread out more evenly across the standardized chronology. Hence why "modern sample bias" is most apparent in modern chronologies.

L17 – The “differing-contemporaneous-growth-rate” problem generally applies to all trees and sites over all time, manifest by trees having differing growth rates in the same climate and caused by micro-site variation (e.g. some seeds fall on stony ground). In RCS there is a presumption that this is random with respect to calendar year and ring age and can be removed simply by having sufficient and well distibuted sample replication over time. Care needs to be taken to ensure sample homogeneity over time (e.g. with respect to changing aspect, altitude or soil type) and the presumption is that this avoids any systematic bias. If diameter (or size) based sample selection procedures, as typically used in dendroclimatology, are used to sample a group of living trees in the presence of “differing-contemporaneous-growth-rates” a systematic chronology bias can occur in the living-tree section of a chronology. “Modern sample bias” is thus the result of sampling living trees in the presence of either “slow-grower survivorship bias” or “differing-contemporaneous-growth-rates” (and usually both) and produces a systematic bias whose magnitude does not reduce with increasing replication.

We agree with this comment in full but suggest that a clearer, more general cause of modern sample bias can be given (see our response to the comment on L22).

L25 – “only limited correction” not a useful description for a considerable improvement over previously published standardization methods. It is not wise to discard the use of multiple RCS curves so lightly (at least until fixed effects standardization is shown to solve the problem). The type of models described in this paper could be developed to evaluate the use of multiple RCS curves.

This is a fair contention. The use of multiple regional curves is incredibly powerful and interesting but somewhat more challenging to analyze than classical RCS or fixed effects standardization. They don’t however, directly address modern sample bias but instead solve it as a side effect of their main use: accommodating distinct patterns of growth by age for trees within the chronology. By splitting the data set into multiple sections according to their inherent productivity, you can mitigate the impact of modern sample bias as the persistent differences in growth become much less extreme . But these persistent differences in growth rate still remain to some degree, allowing some more modest problems of modern sample bias to creep back in. There also needs to be more work done in the use of multiple regional curves to understand how and why to assign series to each curve.

Unlike fixed effects standardization, the shape of these regional curves is completely independent, adding a large number of additional parameters (and flexibility) to the model. Fixed effects standardization should be preferred if we know, assume or can prove that only one common age effect is present to minimize the risk of overfitting and completely accommodate differing contemporaneous growth rates.

The application of models to the use of multiple RCS standardization is both feasible and important and is an excellent candidate for further work on this topic.

CPD
9, C2139–C2155, 2013

P4504 L18 “dendrochronological practice of log-transforming series before analysis” this is rarely used for chronology production. Log transformation tends to overcompensate for the positive skew and cannot handle zero values and does not work well in practice.

Rarely used, but not unprecedented. The real problem is to decide which (if any) transformation is most appropriate and how we might judge that.

Fortunately, statisticians and ecologists have been discussing this problem for a while (see Bolker et al. 2009 for a modern overview). In general, the consensus is that with modern statistical and computational tools data transformation for normality / homoscedasticity is unnecessary and should be avoided in favour of modelling non-normal or heteroscedastic noise directly. Data transformation obfuscates the analysis and interpretation of the data while obscuring the true form of the noise (which is often scientifically interesting).

In order to tell which error distribution is appropriate, we need to examine the residuals of our model under a few different assumptions. Multiplicative log-normal noise is a much, much better match (in terms of shape) than additive normal noise for every tree ring width data set we looked at. It also handles the observed heteroscedasticity nicely, it predicts (as we observe) that large rings are generally more variable in absolute terms than small ones. Finally, the domain of data generated using log-normal noise (strictly positive) is a much better fit to the requirements of ring width data (zero or positive) than that of normal noise (negative or positive). Like in most cases where we’re modelling positive-only data, multiplicative and log-normal (or Poisson etc. for count data) is the appropriate choice. We’ll show some diagnostic plots for the real chronologies analyzed to demonstrate this point.

Some tree ring chronologies (MXD, certain isotope traits) may very well be more suited to additive and normal models.

Do you have a citation for the overcompensation for positive skew argument? I’ve heard

CPD
9, C2139–C2155, 2013

Interactive Comment

it before (from Ed Cook at Ameridendro) but the closest paper I can find is Cook and Peters (1997) which is cited briefly in our Discussion. Certainly, if you assume a normal distribution, log-transformation (equivalent to using a log-normal distribution) will underweight positive outliers but if the true distribution really is log-normal it's entirely appropriate.

The presence of true zeros is a strike against naively log-transforming or using a plain log-normal error term. The zeros in ring width chronologies are caused by absent rings (not missing data) and represent a valuable source of information. Missing rings are caused by related but distinct processes from very small rings however, and should be analyzed using a zero-inflated error term (REF). The difference between a ring width of 0.001 mm and 0 mm is not one of magnitude but kind.

L20+ The "persistent differences in growth rates between trees" after age adjustment and over a common period are the "differing-contemporaneous-growth-rates". The average of the growth rates of trees over any period of time is the low-frequency chronology signal being sought for that period. The "inherent productivity of each tree" could be defined as an absolute value or as a fraction relative to the mean growth rate of all trees (the chronology signal) over the common period for that tree. "T" could be defined so as to contain the long-timescale chronology variance although in this paper it appears not to.

The definition of I given is correct. T is in fact the common time signal, both short- and long-timescale, and is equivalent to the "standardized chronology". Hopefully the worked examples and conceptual diagram help make this clear.

P4505 L7-23 "From this perspective, it is clear that differing-contemporaneous-growth-rate bias (and thus modern sample bias) is an omitted variable bias!" The absolute values of growth rates are an important contribution to the chronology mean and not a systematic bias. The growth-rate differences between trees in any year are part of the noise to be removed by the averaging of sufficient samples. The systematic bias created by sampling living trees is the problem (modern sample bias) to be reduced or removed.

See our very first response to the detailed comments (@L10).

L25 - The additive model is generally only used with power transformation in RCS (see Cook and Peters 1997) or with MXD data (e.g. Grudd 2008).

This is useful context, thanks. Additive models are still quite popular for curve-fitting techniques however, even when applied to ring width chronologies.

Of course, if you log-transformed your data ahead of time, the additive normal model is exactly equivalent to the multiplicative lognormal model on untransformed data.

P4508 "Competitive dominance" - The natural "self thinning" has little effect in the dendroclimatic context – only relatively older trees that have successfully passed through the early stages of competition are sampled. The evidence is that slow-growing trees tend to survive longer than fast-growing trees in dendroclimatic samples - although this effect is relatively small.

Slow-growing trees tend to live longer but don't always and the techniques and theory should be flexible to a wide range of sampling designs and ecologies. Self-thinning / competitive dominance patterns are largely relevant to dendroecology but are worth mentioning as they do influence the shape of the survivorship curve heavily for young trees.

P4511 "Big-tree selection bias ..." – building an RCS curve requires that growth rings of the same age are distributed over time and RCS cannot be used on an "even-aged stand". Sampling trees that died at random dates (e.g. typical sub-fossil collections) overcomes the sampling bias but cannot be applied to the living-tree portion of the chronology.

Sorry, we misspoke here. We meant to say a collection of even-aged stands of different ages. This "chronosequence" approach considerably simplifies the influence of ecological / stand development factors. If you're not concerned about modern sample bias or trend-in-signal bias (because you fixed them), regional curve standardization type approaches can be used without issue on purely modern chronologies. The only assumption it makes is that there is a common pattern of growth by age, which is very reasonable for homogenous, even-aged stands.

P4512 Explaining the error in modelling "Gita" is not the primary objective because the variable of interest is a "Tt" which contains long-timescale variance. The standardisation objective is to find the mean growth rate of trees in each year and not to explain the variance of tree rings. This paper needs a clear statement of the criteria for measuring the success of removing the effects of "modern sample bias" whilst retaining mean tree-growth rates.

We agree very strongly with this comment in principle (although again refer back to our differentiation of time effect/signals from mean growth). Unfortunately, there's no way to assess the ability of these techniques to capture the "true" time signal for real tree ring data. This is almost trivially true, for if we knew what the true time signal was, we wouldn't need to perform the experiment at all!

Beyond assessing the ability of a particular standardize model to explain variance in the observed data, there are two approaches to testing the quality of the time signal (time effect / standardized chronology) extracted.

1. Simulate some data using a particular model and see if we can reconstruct the time signal. If the model used for simulation is the same as that used for analysis, this should be quite straightforward. We show this in Figure 3 but would be willing to expand this section if the reviewers and editor feel it would be helpful. Unfortunately, traditional sequential standardization (where the age effect is estimated and removed, then the time effect is estimated) fails at this basic task when the design is at all unbalanced.
2. Comparing the time signal to things that *should* correspond to it. Usually climate data is used for this. Unfortunately this is both fairly involved and subjective and so was

omitted from this already very long manuscript. If the results are poor, the model used is likely inappropriate (assuming the technique passes the check in 1.) and needs to be improved by examining its failures from both a statistical and scientific perspective.

Looking at explained variance in the independent variable is simple, powerful and extremely common in science as a whole. R^2, RMSE, AIC, residuals etc. tell us, objectively, when and where the model fits well and can warn us about over fitting our data. It is not, however, a substitute for scientific expertise: all candidate models need to be both relevant to your research question and scientifically plausible!

P4514 L5 Flat detrending is rarely used as it does not correct for the change of ring width with tree age.

We agree with this statement and feel that is much more useful when combined with regional curve standardization to account for changes by age.

P4514 A comparison between sample chronologies created using SF RCS and equivalent fixed effects standardisation model would be useful – results should be the same if RCS curve smoothing is excluded.

We agree, this will be highlighted in our sample chronologies.

P4515 L22 - RCS smoothing “in part this was for convenience”. If you have sufficient trees (2000+) the RCS curve can be smooth enough not to need any smoothing. With fewer trees there is noise originating mainly from the climate signal. Tree growth in an unchanging climate without noise is expected to be smooth so we can undoubtedly justify smoothing to correct for lack of trees (although negative exponential is not usually suitable - see Melvin and Briffa 2007 Dendrochronologia, 26 (2) for suggestions as to what is required of smoothing).

This is what we’ve seen as well (although some very well behaved chronologies are smooth well before the 2000 tree mark). There’s still a minor risk of overfitting with unsmoothed regional curves / age effects though, where we might mistake noise for true age-related variability and distort our standardized chronologies.

P4517 “Exploring published ring-width data” - Classic RCS (e.g. Briffa 1992) requires a chronology much longer than the age of individual trees (more specifically a chronology containing data from sub-fossil trees). Using RCS methods on chronologies composed solely of living trees sampled at one point in time, where the climate signal distorts the RCS curve is problematic. The signal-free method can reduce this problem provided there is a sufficiently wide variation in tree ages.

You don’t give enough credit to signal-free methods here. Try some tests with no noise and a "true" or categorical age effect (regional curve of the same form as the data generating process, unsmoothed regional curve). This problem is completely resolved and the time and age effects can be fully separated with signal-free standardization as long as at more than one tree age (date of birth really) is present.

It may be that many of your selected sites are not well suited to RCS processing. You would be better using far fewer and more appropriate sites. Including a couple of well replicated sites containing sub-fossil trees and well replicated living tree only sites would be sufficient. If you are correcting “modern sample bias” then your method will not change the sub-fossil portions of the chronologies. Demonstrations should include comparison of chronologies with old and new methods. Torneträsk and Yamal TRW would be suitable sub-fossil data sets (available from http://www.cru.uea.ac.uk/publications/papers) .

We agree and noted that many of the chronologies may be poorly suited to RCS. The relative unimportance of the age effect confirms that common age-driven signals are weak. Case studies should improve the paper a great deal in this respect.

P4519 L19 - Choosing a model on the basis of explaining “G” does not measure the quality of the chronology – the “correction” needs to produce a more accurate assessment of the long-timescale chronology variance in order to be useful.

CPD

9, C2139–C2155, 2013

See @P4512 for a discussion of this point.

P4520 L12 - The ratio comparison detects the change in chronology represented by explaining “I” in the modelling and this may not be change related solely to “modern sample bias”. If you have sufficient sub-fossil samples in a chronology the shape of the RCS curve will not change due to the presence of “modern samples”. The period of changing living-tree counts can be used to isolate the effects of “modern sample bias” on the chronologies. Again, some sample chronologies need to be shown. The difference created by explaining “I” in the modelling needs to be clearly demonstrated (e.g. low-pass filtered for clarity) in chronologies.

"Modern sample bias" could still be present in subfossil samples and distort the estimated age effect. See our earlier comment for more on this point. Seeing the divergence as modern samples are added is however a very useful heuristic and should be evident in the sample chronologies.

P4521 l2 “Contrary to prevailing opinion” - It has not been shown that the change created by the correction is solely related to “modern sample bias” and it remains (until demonstrated) possible that something else is responsible.

See @AbstractL22 for a discussion of this point.

L8 – “D’Arrigo et al. (2008) suggest that ...” does not seem to be right as modern sample bias is associated solely with RCS and divergence is generally associated with curve-fitting chronology construction methods.

Interesting to hear that the divergence problem is most associated with curve-fitting techniques. D’Arrigo et al. (2008) themselves as well as Brienen et al. (2012) both mention that modern sample bias could be a cause of the divergence problem.

L10 – “The generally negative trend” is worrying – no examples are shown and no suggestion as to why other methods produce observations with opposite sign.

This is likely just a problem of using different datasets. Our large sample from the

ITRDB is quite different from "classical" regional curve chronologies and may be fairly unsuited to regional curve models. It's extremely challenging to come to valid conclusions about the ecological / methodological causes of the observed biases at this scale with such little available meta-data. Sample chronologies should help with this.

L15 – "Regional curve standardization is a biased implementation of signal-free standardization, .." only if you do not have a set of trees with a wide enough time range. There is a need to distinguish in your discussion between "systematic bias" and "random noise".

There's an important distinction between bias in the dendrochronological sense (it affects the trend of the time signal) and bias in the statistical sense (our estimates of the model coefficients will never converge to their true values, even with infinite data). Generally, dendrochronological bias is a subset of statistical bias.

We use "biased implementation" here in the statistical sense. Unless you are extremely lucky and have a perfectly balanced design (not merely a "a set of tree with a wide enough time range"), sequential standardization techniques will always be outperformed by their better-optimized equivalents. There is no added "cost" to doing so as the assumptions are identical in each case.

L20–L5 next page "The estimates of I are ..". This is very speculative and not shown in the paper.

This is potentially helpful to inspire future research but somewhat speculative. It will be more clearly marked as such.

P4532 L21 "signal-free standardization results in an unbiased least-squares estimate" not least-squares since the convergence criterion we use is based on minimizing least absolute differences.

The convergence criteria is distinct from the estimation protocol. By taking the mean when estimating the effect of each age / year, it's equivalent to using least squares

CPD
9, C2139–C2155, 2013

estimation of the effect of each age / year because the effects are fully categorical. Refer to Appendix 1 or feel free to ask for a more detailed proof of this if it’s not clear.

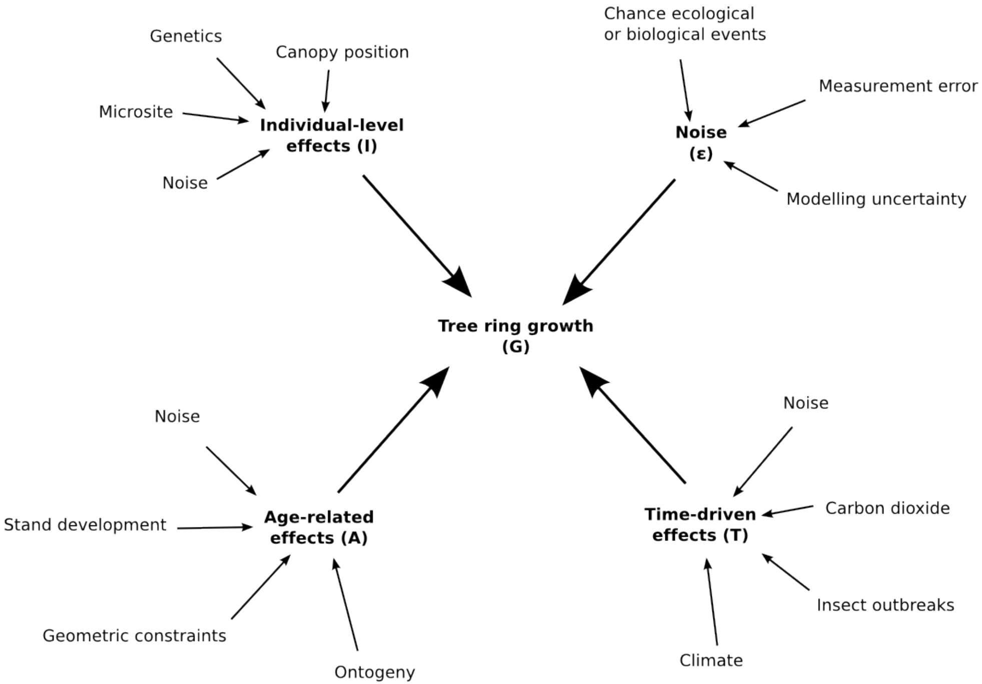
P4540 Melvin 2012a and 2012b are same paper.

Thanks, this will be fixed.

References

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution, 24(3), 127-135.

Interactive comment on Clim. Past Discuss., 9, 4499, 2013.



Supplementary Figure: A conceptual diagram of tree ring growth. The ultimate causes of tree growth (plain text) can be groupesd according to the pattern of variability (common to each individual, time or age) they induce in tree-ring growth. By partitioning the observed variability in tree-ring growth into latent effects, we can seperate time-driven effects (climate, CO2 etc.) from age or individual-related effects. This "time effect" (common temporal signal) is the target of standardization, not "typical growth".

Fig. 1. A conceptual diagram of tree ring growth. The ultimate causes of tree growth (plain text) can be grouped according to the pattern of variability (common to each individual, time or age) they induce.