

This discussion paper is/has been under review for the journal *Climate of the Past* (CP).  
Please refer to the corresponding final paper in CP if available.

# A likelihood perspective on tree-ring standardization: eliminating modern sample bias

J. Cecile, C. Pagnutti, and M. Anand

University of Guelph, School of Environmental Sciences, Guelph, Canada

Received: 22 July 2013 – Accepted: 31 July 2013 – Published: 9 August 2013

Correspondence to: J. Cecile (jcecile@uoguelph.ca)

Published by Copernicus Publications on behalf of the European Geosciences Union.

4499

## Abstract

It has recently been suggested that non-random sampling and differences in mortality between trees of different growth rates is responsible for a widespread, systematic bias in dendrochronological reconstructions of tree growth known as modern sample bias. This poses a serious challenge for climate reconstruction and the detection of long-term changes in growth. Explicit use of growth models based on regional curve standardization allow us to investigate the effects on growth due to age (the regional curve), year (the standardized chronology or forcing) and a new effect, the productivity of each tree. 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.

This class of models describes a new standardization technique, fixed effects standardization, that contains both classical regional curve standardization and flat detrending. Signal-free standardization accounts for unbalanced experimental design and fits the same growth model as classical least-squares or maximum likelihood regression techniques. As a result, we can use powerful and transparent tools such as  $R^2$  and Akaike's Information Criteria to assess the quality of tree ring standardization, allowing for objective decisions between competing techniques.

Analyzing 1200 randomly selected published chronologies, we find that regional curve standardization is improved by adding an effect for individual tree productivity in 99 % of cases, reflecting widespread differing-contemporaneous-growth rate bias. Furthermore, modern sample bias produced a significant negative bias in estimated tree growth by time in 70.5 % of chronologies and a significant positive bias in 29.5 % of chronologies. This effect is largely concentrated in the last 300 yr of growth data, posing serious questions about the homogeneity of modern and ancient chronologies using traditional standardization techniques.

4500

## 1 Introduction

Much of the work in dendrochronology, and dendroclimatology in particular, relies on accurate, unbiased reconstructions of tree growth long into the past. As a result, a great deal of effort has been put into trying to isolate important trends and identify potential biases. However, one major bias called “modern sample bias”, first identified by Melvin (2004), is still largely neglected in applied studies, despite its potential impact on all regional curve standardization chronologies (Brienen et al., 2012a).

Dendrochronologists observed that the older a tree was, the slower it tended to grow, even after controlling for age- and time-driven effects. The result is an artificial downward signal in the regional curve (as the older ages are only represented by the slower growing trees) and a similar artificial positive signal in the final chronology (as earlier years are only represented by the slow growing trees), an effect termed modern sample bias. When this biased chronology is used in climate reconstruction it then implies a relatively unsuitable historic climate. Obviously, the detection of long term trends in tree growth, as might be caused by a changing climate or carbon fertilization, is also seriously compromised (Brienen et al., 2012b). More generally, modern sample bias can be viewed as a form of “differing-contemporaneous-growth-rate bias”, where changes in the magnitude of growth of the tree ring series included in the chronology over time (or age, in the case of the regional curve) skew the final curve, especially near the ends of the chronology where series are rapidly added and removed (Briffa and Melvin, 2011).

Several attempts have been made to address this issue but none have proven fully satisfactory. Melvin (2004) (see also Briffa and Melvin, 2011; Cooper et al., 2012, and Melvin et al., 2012a) attempt to solve the problem by splitting the regional curve into several smaller curves by growth rate as first introduced by Esper et al. (2002) but this approach offers only limited correction as the number of sub-RCS curves is necessarily smaller than the number of levels of growth rate observed by the trees and the reduction in sample size reduces the reliability of each sub-curve. Voelker (2011) took

4501

a different approach, first standardizing the chronologies with respect to age and annual effects, then estimating the linear relationship between tree growth rate and tree age for each species studied from binned growth and age data, which was then used to scale the species (or in some case genus) level chronologies. Interestingly, while the majority of the species/genera analyzed showed a negative relationship between tree age and growth rate, a positive relationship was observed in a few cases, contrary to the predictions of (Briffa and Melvin, 2011). Whether this effect has an ecological basis or simply represents a random quirk of the chronologies examined was not discussed. While revealing, this technique relies on large pre-existing chronologies for the species of interest and assumes a simple common linear relationship between age and tree-specific productivity.

To develop an alternative approach, we make explicit the growth models already used in regional curve standardization. From there, we examine the ecological effects driving the persistent differences in growth rates between sampled trees of various ages and then use regression to obtain an unbiased estimate of the inherent productivity of each tree, the typical growth of the trees at a given age (the regional curve) and the forcing at each year (the standardized chronology). The relationship between this new technique, dubbed “fixed effects standardization”, regional curve standardization (Briffa et al., 1992), flat detrending (Cook and Kairiukstis, 1990) and the more recent signal-free standardization (Melvin and Briffa, 2008; Briffa and Melvin, 2011) is explored along the way.

We conclude with a brief sample of existing dendrochronological records, demonstrating fixed effects standardization, selecting the most appropriate standardization model for each data set and exploring the effects of accounting for tree-level productivity across the globe.

4502

## 2 Growth models

Regional curve standardization makes two central assumptions about the typical growth of trees used in dendrochronological analysis. First, that trees of the same species within the same region follow a certain inherent pattern of growth as they age, given by the regional curve. Second, the growth of each tree in a given year is the product of the expected growth at that age and the common forcing of that year (Melvin et al., 2012a). This common forcing affects all trees equally in proportion to their expected growth, an assumption clearly visible in the division of the raw tree ring series by the regional curve to obtain the standardized chronology, which is by design free of age-driven effects. By doing so, regional curve standardization presumes a model, and hence can be treated as a model-fitting tool (a goal expressed, but not achieved, in Bontemps and Esper, 2011).

Tree ring data can naturally be classified among two dimensions, the year in which the ring was formed ( $t$ ) and the age of the tree when the ring was formed ( $a$ ). Each chronology can be stored naturally in a growth matrix  $\mathbf{G}$ , a form referred to as a “tree-ring array”, in contrast to the traditional form where each column holds the ring widths observed for a single series and the row name denotes the year (a “tree ring table”). Figure 1 shows a conceptual diagram and further explanation of these two alternate organizations of tree-ring data.

If we consider the time effect (standardized chronology) and age trend (regional curve) as the effect vectors  $\mathbf{T}$  and  $\mathbf{A}$  respectively, we can write the implicit growth model of regional curve standardization as follows:

$$G_{ta} = T_t A_a \quad (1)$$

Each element of the growth matrix  $\mathbf{G}$  is a scalar, the product of the corresponding elements of  $\mathbf{T}$  and  $\mathbf{A}$ . Looking at it from the perspective of the entire vectors, we can construct the tree ring-array as the outer product of the effect vectors.

$$\mathbf{G} = \mathbf{T} \otimes \mathbf{A} \quad (2)$$

4503

Obviously, real trees do not follow this growth model exactly, and should be thought of as being drawn from a population described by a probability density function. The product of the time and age effects is the expected value of growth for each ring, the observed data includes an error term  $\epsilon$  to account for stochastic noise and problems with model fit. The simplest way to do so is to assume an additive, normal error term such that:

$$G_{ta} = T_t A_a + \epsilon_{ta}^N \quad (3)$$

However, it is commonly observed that real tree ring width data is strongly heteroscedastic, with variability increasing as the observed growth values grow larger (Biondi, 1993; Meko et al., 2001). Proportional log-normal variability has long been observed in measurements of plant growth (Evans, 1972; Pokharel and Dech, 2012) and reported in dendrochronology (Van Den Brakel and Visser, 1996; Drobyshev and Niklasson, 2010). Furthermore, tree-ring width data is naturally bounded by 0, a fact for which a normal probability density function fails to account. Rather than a normal probability density function, we suggest that most tree ring data may be drawn from a log-normal probability density function instead. A multiplicative log-normal error term accounts for the observed heteroscedasticity while remaining tractable and is consistent with the dendrochronological practice of log-transforming series before analysis.

$$G_{ta} = T_t A_a e^{\epsilon_{ta}^L} \quad (4)$$

To address differing-contemporaneous-growth-rate bias, and thus modern sample bias, we need to examine its cause. Many researchers have observed that during the process of chronology construction, there are persistent differences in growth rates between trees (Brienen et al., 2006; Zuidema et al., 2010). Furthermore, Melvin (2004) observed that the ratio between these series and the common signal (either the regional curve or standardized chronology) is approximately constant. This ratio was termed “error” (later discussed as the ratio between multiple regional curves in Melvin et al., 2012a). In the growth model framework, it can be considered the effect of each

4504

individual tree ( $I_i$ ), and thought of as the inherent productivity of tree  $i$ . Extending our model to incorporate this concept:

$$\mathbf{G} = \mathbf{I} \otimes \mathbf{T} \otimes \mathbf{A} \quad (5)$$

$$G_{ita} = I_i T_t A_a + \epsilon_{ita}^N \quad (6)$$

$$5 \quad G_{ita} = I_i T_t A_a e_{ita}^L \quad (7)$$

From this perspective, it is clear that differing-contemporaneous-growth-rate bias (and thus modern sample bias) is an omitted variable bias! Note that in this case  $\mathbf{G}$  becomes a three-dimensional array, recording the tree  $i$ , age  $a$  and year  $t$  for which each data point was recorded.  $\mathbf{I}$  is a vector much like  $\mathbf{T}$  and  $\mathbf{A}$  but, unlike the others, lacks a natural ordering.

The model in Eq. (1) fails because the assumption that the error in each series is unbiased by which tree the series belongs to is not supported by the data. As series are added to and removed from the chronology, the mean value of  $\mathbf{I}$  in the chronology changes, biasing the observed chronology downwards if the trees were less productive than average and upwards if the trees were more productive than average. The magnitude of this effect only changes when the series present in the chronology changes, the bias observed takes the form of a step function by time and age and is typically stronger near the ends of the chronology (Bowman et al., 2013). Seen from this perspective,  $\mathbf{I}$  is a traditional nuisance parameter (as was  $\mathbf{A}$  originally), whose effects must be accounted for to obtain a reliable estimate of  $\mathbf{T}$ . If  $\mathbf{I}$  is not identically and independently distributed across time and age (as in modern sample bias scenarios), the resulting chronology is skewed.

The choice of mean (arithmetic/geometric) corresponds to the choice of probability density function (normal/lognormal). In many cases, dendrochronologists choose to subtract, rather than divide, the estimated effect vector from the growth data. This corresponds to an additive growth model, similar to:

$$G_{ita} = I_i + T_t + A_a \quad (8)$$

4505

For ring width data, this form of model is almost certainly inappropriate. Growth data is by definition positive and typically multiplicative; the use of additive models can result in negative predicted growth. Note, of course, that the logarithmic transformation of a multiplicative growth model with log-normal noise is equivalent to an additive model with normal noise.

$$G_{ita} = I_i T_t A_a e_{ita}^L \quad (9)$$

$$\log(G_{ita}) \propto I_i + T_t + A_a + \epsilon_{ita}^N \quad (10)$$

These models can still be fit and examined using the techniques throughout the paper. In particular, they can be compared to their multiplicative counterparts using model selection tools (see Section 5 for an exploration of this on real data).

When using real data, it is almost impossible to find a tree-ring measurement for each unique combination of age and year and as a result the tree-ring array is almost never full. The observed array of growth values  $\mathbf{G}$  can be discussed using a weighting array  $\mathbf{W}$ , which describes the number of data points present at each location. By definition, values can only be filled along a single  $t$ - $a$  diagonal for each position of  $i$ , as the tree ages by one year for each calendar year. Similarly, if only complete (pith to bark) tree-ring series are used, the upper-right  $t$ - $a$  triangle is always empty; if a tree is sufficiently old at the appropriate point in the past, it extends the chronology backwards in time a corresponding number of years, leaving the new upper triangle empty again. As discussed in Sect. 4, this unbalanced design complicates analysis and accounts for the improvement of signal-free standardization over traditional regional curve standardization.

These models have a final interesting property: for each solution to a tree ring array, there exist an infinite number of equivalent solutions. The system is singular, any scalar multiple of one of the effects vectors ( $\mathbf{I}$ ,  $\mathbf{T}$  or  $\mathbf{A}$ ) does not change the predicted growth  $\mathbf{G}$  as long as it is counteracted by the reciprocal scaling of a different vector, producing equally likely models with different coefficients. To address this, we fix the geometric mean of the elements of  $\mathbf{I}$  and the elements of  $\mathbf{T}$  to 1 by convention and scale  $\mathbf{A}$

4506

to compensate. This allows, as is traditional,  $A$  to map directly to the typical ring-width increment of a tree at a given age (as in the regional curve) and  $I$  and  $T$  to represent an index of deviation from this expected growth. The chronologies in Melvin et al. (2012a) are scaled in slightly different fashion to reach the same end: the simple comparison of estimates produced by different techniques.

### 3 Ecology of modern sample bias

It has long been known that the life expectancies of trees tend to be negatively correlated with their growth rate, both between and within species of trees (Huntington, 1913; Schulman, 1954; Black et al., 2008; Johnson and Abrams, 2009). Brienen et al. (2012b) explain how this causes modern sample bias via productivity-survivorship bias (in that paper termed “slow-grower survivorship bias”): if slow-growing trees are more likely to survive, they will be over-represented in the oldest sections of the standardized chronology and regional curve, producing a positive skew. When trying to reproduce this effect, it’s helpful to think of the problem in terms of survivorship curves. First, assume each tree follows a particular survivorship curve conditional on its productivity. Formally, the survivorship curve is the probability of surviving at a given age and productivity level and can be written as  $P(S|a, I)$ .

Using Bayes’ theorem, we can look at the distribution of surviving trees of a given age for different values of  $I$  by considering the joint probability of this survivorship curve and the initial distribution of productivity at birth,  $P(I)$ .

$$P(I|a, S) = \frac{P(S|a, I) \cdot P(I)}{P(S|a)} \quad (11)$$

The this expression describes the typical productivity of trees of that age. The fluctuation in the expected value of this expression results in modern sample bias. When  $I$  is not accounted for, ages that are less likely to contain productive trees will be underestimated by the regional curve (Fig. 2). The effect this has on the final standardized

4507

chronology depends on the arrangement of the data, but when all series are complete and alive at the time of sampling (modern) the effect will be precisely opposite that on the regional curve.

For our purposes, survivorship curves are discussed as fast-biased, in which case fast-growing trees are more likely to survive, slow-biased (the reverse) or unbiased. It is of course possible for a survivorship curve to be fast-biased for some ages and slow-biased at others depending on the ecological effects at play. We suggest that there are four broad drivers of productivity-survivorship bias: competitive dominance, the patchy resource effect, ecophysiological limitation and biased disturbances. Many of these effects are both poorly quantified and complex, limiting our ability to predict the direction or magnitude of productivity-survivorship bias in general.

Competitive dominance is the basis of natural self-thinning. It has long been observed that trees suffer increased mortality rates when growing more slowly than their neighbours due to competitive exclusion (Peet and Christensen, 1987). Tree diameter is both allometrically linked to tree height, an important determinant of light competition, and suffers directly when resources are limited by competition. As a result, competitive dominance will increase the typical productivity of the population as competition occurs and slow-growing trees are removed from the population. As Brienen et al. (2012b) suggested, this effect is likely strongest in young trees, especially when the species is shade-intolerant.

The patchy resource effect acts on a larger scale, that of environmental variability in fertility. Resources in a forest ecosystem (water, microclimate and nutrients in this context) are inherently patchy, leading to low and high fertility sites and microsites. Increased site fertility is directly linked to increased competition and accelerated stand closure; as soil nutrients (or microclimate) improve, more resources can be allocated to above-ground biomass and light competition per year (Vanninen and Makela, 1999). Research into self-thinning empirically substantiates this claim and consistently shows accelerated stand dynamics with increasing stand fertility (Elfving, 2010). To understand the effects of this, we consider two stands with different fertility levels. Before

4508

canopy closure occurs, the productivity of trees is positively correlated with the site fertility as expected and no bias occurs. What happens next depends on the sampling protocol selected. The number of trees in the fast-growing population drops much more rapidly than that of the slow-growing population, producing a decline in the typical productivity of the metapopulation as the weightings shift. Thus, increases in the patch-scale variability in fertility tend to increase the bias towards lower productivity trees, producing an artificial positive trend in the time signal.

Ecophysiological limitation driving tree mortality is widespread and lead some to suggest that senescence and physiological limitations may be delayed in slow-growing trees or those on unproductive sites (Chao et al., 2008; Briffa and Melvin, 2011; Brienen et al., 2012b). Stephenson et al. (2011) presents an excellent overview of the major effects of this type by dividing them into four relevant hypotheses: the enemies hypothesis, the growth-defense hypothesis and the growth-hydraulic safety hypothesis and the shade tolerance hypothesis. In the enemies hypothesis, natural enemies are more common to highly productive trees due to their higher energy and nutrient concentrations. The growth-defense hypothesis use the idea of resource limitation; if a tree is using resources to produce growth (especially above-ground biomass), it can't spend them on resource diversification or natural defenses and is thus more likely to die to disturbances or environmental stress. The third hypothesis, the growth-hydraulic safety hypothesis similarly suggests that resistance to hydraulic failure is costly due to increased resistance to water transport and thus trees which invest in proper hydraulic architecture will survive longer than their peers, at least after competitive effects have reduced. Finally, the shade tolerance hypothesis relies on the common trend of reduced growth potential in shade-tolerant and shade-grown trees due to ontological choices in leaf anatomy and biochemistry. While these effects are commonly discussed in terms of between-species differences, there is some evidence to suggest that the genetic variability and phenotypic plasticity present within a species is sufficient to create small effects of this sort (Rötheli et al., 2011).

4509

Ecological disturbances are the final effect shaping survivorship curves, although their effects seem even more context-sensitive. Insect damage, disease, drought and frost mortality, windthrow, herbivory, floods, landslides, fire and harvest are all examples. These events are stochastic and significantly more challenging to model and their preference for trees of different ages and growth rates varies by disturbance type. Some of these events are fast-biased, either due to their disproportionate effect on small trees like herbivory and ground fires. Others, principally windthrow, are more likely to affect larger trees, biasing the survivorship curve towards slow-growing trees. Still other events, catastrophic ones such as landslides, stand-replacing fire, floods and land use conversion don't discriminate at all in terms of growth rate, size or age. Finally, the frequency of disturbance, especially harvesting, may increase with site productivity, limiting the availability of fast-growing old trees.

### Big-tree selection bias

In this framework, big-tree selection bias is quite simple to explain. Large trees are commonly selected for in dendrochronological sampling, in hopes of sampling old trees with long records and to avoid the logistical difficulties associated with coring very small trees. In the case of a minimum diameter cutoff (ignoring the effects of temporal forcing):

$$I_i \geq \frac{D_{\min}}{\sum_{a=1}^{a_i} A_a} \quad (12)$$

As the tree grows with age, the required minimum value for  $I$  becomes less stringent, leading to higher  $I$  values observed for young trees. Even without a strict minimum diameter, there is a bias towards the selection of highly productive young trees (or simply old trees) as long as large total diameter is desired. Because diameter depends on the inherent productivity of a given tree, high  $I$  values are still more important for young trees than old ones, assuming that some fraction of local trees are sampled at each site or subsite.

4510

Big-tree selection bias is not, per se, a concern in even-aged stands. Even though only the largest and most productive trees are selected, there is no bias by age as the experimenter can only choose between trees of the same age at any given point in time. As Brien et al. (2012b) stated, big-tree selection bias can be eliminated through careful experimental design. This is only a small part of the larger modern sample bias problem however, which must be corrected by accounting for the productivity of each individual tree.

#### 4 Estimating growth models

The models introduced in Sect. 2 are quite simple. In all cases, the observed growth data is the product of some number of effects (individual, tree or age driven). Each effect vector is a latent categorical variable, with a separate coefficient for each unique tree, year or age. There are 8 possible models, each with a unique combination of effects, increasing in complexity from the null model ( $G_{ita} = \epsilon_{ita}^L$ ) to the complete model ( $G_{ita} = I_i T_t A_a \epsilon_{ita}^L$ ).

The simplest approach is to find the maximum likelihood solution to the growth model of choice. The error term corresponds to the probability density function used during this process. Once an optimal family is found (using simulated annealing for example), the effect vectors are rescaled to the desired form.

The other approach is to find an approximately optimal solution using the method of moments. The simplest case is a model with only a single effect vector (say  $T$ ). For a normally distributed error term, we can estimate the coefficient  $T_t$  at each year by taking the arithmetic mean of the observed growth values for that year. If we assume a log-normal multiplicative error term, we need to use the geometric mean instead.

The problem of estimating growth models from tree ring data (standardization) can be framed as a regression problem. For the additive normal error term, this is a nonlinear regression with multiplicative categorical variables. In the case of the multiplicative log-normal error term, this is a generalized linear model with a log link function and

4511

no intercept (equivalent to log-transforming the growth model and using traditional linear regression). The “rescaling” that we discussed at the end of Sect. 2 is a dummy variable trap. Because we have multiple latent categorical variables, we can shift the (log)-intercept of each effect as long as it’s counterbalanced by a shift in another variable. The choice to rescale the observed coefficients such that the typical value of  $I$  and  $T$  are 1 corresponds to a special type of contrast. Because this is a regression problem, estimating the coefficients by the methods of moments, least squares or maximum likelihood will all give similar results.

A regression modelling perspective to tree-ring standardization immediately suggests some helpful metrics. If we want to understand the goodness of fit, we can look at the likelihood or  $R^2$  of the model fit. To examine the noise level more directly, we can examine the estimated variance parameters ( $\sigma$ ) in our error term. This is equivalent to the root-mean-square error (RMSE) for the additive normal error term. Comparing  $R^2$  or  $\sigma$  between error families (probability density functions) is not directly meaningful.

Instead, the most appropriate way to decide on the probability density function is to examine the residuals of the model. Using histograms, kernel density estimators or quantile-quantile plots, the residuals should match the desired probability density function. In most cases, visual inspection is the simplest and most reliable approach.

The final, and perhaps most important decision we need to make is which of our 8 (16 if we need to choose an error family as well) is the most appropriate for our data. Simple measurements of goodness of fit, such as likelihood, RMSE or  $R^2$  are not suited to this task. Because the models are nested, adding an additional term will always improve the goodness of fit. The use of these metrics for model selection leads to overfitting. Instead, we need to penalize the use of additional degrees of freedom, as in adjusted  $R^2$  or information criteria do this. Adjusted  $R^2$  has a familiar interpretation. It has the same behaviour as  $R^2$ , in that it increases to a limit of 1 as the model explains all of the observed variability but includes a term penalising the use of additional linear predictors. Adjusted  $R^2$  however obscures the relative strength of evidence for each competing model (Burham and Anderson, 2002, p. 94–96). Information criteria (mostly Akaike’s

4512

information criteria and Bayesian information criteria) use an information-theoretic approach to model selection, balancing the likelihood of each model against the risk of overfitting given a certain number of parameters. Information criteria can be converted into models, which accurately convey the level of support for each competing model, and should be used to understand the relative support of each model.

In some cases, it's not feasible to manually examine the residuals of each of the fit models (as in the case of massive meta-analyses). We can use likelihood (and by extension information criteria) to compare models drawn from competing probability density functions. A model with the true probability density function will tend to have a higher likelihood than the corresponding model with an incorrectly specified probability density function. The use of information criteria will reflect these differences in likelihood, choosing the correctly specified model over competitors. In most cases though, this approach should complement, rather than replace, the visual inspection of residuals. Likelihood approaches to error family selection does not reveal subtler distinctions such as skew or heteroscedasticity which may suggest the use of more sophisticated probability density functions.

The final benefit of these models is the ease of obtaining confidence intervals for the effects estimated. Frequentist confidence intervals, or likelihood based support intervals can be easily extracted from various premade regression packages. One element of each of the effect vectors (typically the first) will not have an estimated confidence interval due to the degree of freedom lost by rescaling the effects. In terms of traditional categorical regression, it is the baseline level. In the case of a normal error term, these intervals need to be scaled with the effects vector as they are converted to their standard form.

#### 4.1 Regional curve standardization and flat detrending

As intended, regional curve standardization fits nicely into this growth model framework. When regional curve standardization is performed, we estimate the model (using

4513

a additive form and normal error term to demonstrate):

$$G_{ita} = T_t + A_a + \epsilon_{ita}^N \quad (13)$$

Similarly, the raw chronology is constructed by fitting the following model:

$$G_{ita} = T_t + \epsilon_{ita}^N \quad (14)$$

Less obviously, fixed effects standardization is also an extension of flat detrending techniques (Cook and Kairiukstis, 1990). If we divide by the mean of each series (the flat detrending line) and then construct the standardized chronology, we're estimating the individual tree and time effects.

$$G_{ita} = I_i + T_t + \epsilon_{ita}^N \quad (15)$$

The use of a normal error term implies the use of an arithmetic mean. If we want to assume a log-normal error term, a geometric mean should be used instead.

But if, for example, regional curve standardization follows a growth model, why must we estimate the age effect first and then the time effect? Do we obtain the same result if the process is reversed? Trivially, the answer is no, as they will be scaled differently. But even when we account for that, the answer is typically no; the order in which the effects are estimated determines the family of effect vectors that is produced.

The reason for this is that standardization, by and large, is done sequentially, rather than simultaneously. In the case of a balanced design (the weight matrix  $\mathbf{W}$  is constant), sequential estimation of the effects works properly (Fig. 3). Removing an estimated effect does not bias our estimation of the other effects because the changes are symmetric across all levels of the other effects (see Appendix A). When the design is balanced, sequential estimation of the effect vectors by taking the mean at each index is an unbiased estimate of the true effect vectors. The family of solutions found is the same regardless of which effects are estimated.

Unfortunately, this is virtually never the case when dealing with real tree ring data. In order to have a balanced design (for regional curve standardization), a chronology

4514



would need to have a completely uniform sample depth by age and time. Unbalanced designs are a fact of life in dendrochronology. When analyzing an unbalanced design sequentially (as in flat detrending or regional curve standardization), changes in one effect vector (such as age), result in changes in the estimate of every other effect vector (such as time). The signals are convoluted and cannot fully be separated.

In fact, this very problem was identified in the work on signal-free standardization referred to as “trend-in-signal bias” (Melvin, 2004; Melvin and Briffa, 2008; Briffa and Melvin, 2011). The solution they proposed was signal-free standardization. By repeating the standardization process, they eventually produces stable estimates of the effects that did not suffer from trend-in-signal bias and better retained low-frequency variability. With slight modifications, signal-free standardization can be expanded to work with models with more than two effects, estimating each in sequence repeatedly (Appendix A). When we do so, we can prove that signal-free standardization converges to an unbiased estimate of the growth model, and properly handles unbalanced designs (Fig. 3).

The power of signal-free standardization comes from its effectively simultaneous estimation of the effects of interest. Because it converges on a unbiased solution to the growth model, the results of signal-free standardization are approximately equivalent to the growth models estimated using more conventional optimization techniques.

## 4.2 Smooth and parametric age effects

Conspicuously absent from all the preceding discussion is the common practice of smoothing the regional curve. In part, this was for convenience. The solutions are much simpler and the symmetry more obvious when age is treated as a categorical variable like time and individuals. But in truth, it is because a much more elegant solution exists. Rather than fit a smooth curve (parametric or nonparametric) to the raw regional curve each time it is estimated, we can simply include a smooth age effect in our model directly (see Bontemps and Esper, 2011 for an example of this principle). A modified negative exponential regional curve for example (Fritts, 1976), would be included by

4515

fitting the following model to your data:

$$G_{ita} = I_j T_t (k_1 e^{-k_2 a} + k_3) e_{ita}^L \quad (16)$$

In the case of a log-normal probability density function, a smooth flexible age trend can be fit using generalized additive models (see Beck and Jackman, 1998 for a gentle introduction). Generalized additive models use either splines or kernels to ensure that the fitted curve is locally smooth. The model used can be represented as:

$$G_{ita} = I_j \cdot T_t \cdot s(a) \cdot e_{ita}^L \quad (17)$$

Where  $s(a)$  is a smooth function of age. As before, we can distinguish between these competing models of standardization using model selection criteria such as AIC. Breadth of imagination and biological plausibility are the only constraints, so long as an appropriate optimization algorithm can be found.

## 4.3 Likelihood ridges in three-effect models

A final complication arises when we attempt to estimate the three-effect (individual, time and age effects) for real tree-ring data. Realistic tree ring data is arranged along time-age diagonal lines in the tree-ring array. Each year that the tree ages, the calendar year advances by one. Because of this structure, a peculiar ridge emerges in the likelihood of three-effect models (Appendix B). For each potential set of parameters ( $\hat{I}$ ,  $\hat{T}$  and  $\hat{A}$ ) there exists a log-linearly related family of solutions ( $\check{I}$ ,  $\check{T}$  and  $\check{A}$ ) that produces the same set of expected values ( $\hat{G} = \check{G}$ ), and hence has the same likelihood. When all three effects are included, each element of the effects vector is scaled by a constant effect  $m$  is raised to a power related to the age ( $a$ ), year ( $t$ ) or birth year of the tree ( $b$ ). Our tree-ring data is stored in a tree ring-array so we can think of this in terms of columns ( $C$ , corresponding to age) and rows ( $R$ , corresponding to time).

$$\check{I}_j = m^b \hat{I}_j = m^{R-C} \hat{I}_j \quad (18)$$

4516

$$\check{T}_t = m^{-t} \hat{T}_t = m^{-R} \hat{T}_t \quad (19)$$

$$\check{A}_a = m^a \hat{A}_a = m^C \hat{A}_a \quad (20)$$

$$5 \quad \check{G}_{ita} = \hat{G}_{ita} \quad (21)$$

By making a priori decisions about the expected distribution of effects, we can distinguish between the “equivalent” solutions by likelihood. In practice, this post-hoc correction seems to work quite well but further investigation and parametrization of plausible effect distributions is still needed.

## 5 Explored published ring-width data

The techniques above leave us with four major questions as to their real world impact:

1. What growth model is most appropriate for tree-ring width data? Which effects (individual, time and age), form (additive or multiplicative) and error term (normal or log-normal) should be used?
2. How common is differing-contemporaneous-growth-rate bias?
3. What is the typical effect of modern sample bias? How much variability is there?

To answer these questions, we turn to the extensive International Tree-Ring Data Bank, version 702 (Grissino-Mayer and Fritts, 1997). 6997 tree ring width data sets were available, of which 5609 (80.2%) could be read into R. From the remaining 5609 data sets, 1200 were selected at random for analysis. 2 of these were excluded from the analysis due to computational constraints (not enough RAM to fit GAM models

4517

on these extremely large data sets). In total, these data sets contained 33 657 series, 6 905 445 unique measurements and spanned from 476–2008 AD, with the bulk of the data occurring after 1500 (see Fig. 4 for an illustration of sample depth across time). Information on the data sets themselves, code used and individual results is available on request.

From our discussion above, we can already rule out sequential standardization techniques, as they are strictly worse than their signal-free (or maximum likelihood or least squares) equivalents. Furthermore, we know that an additive model suggests a normal error term, while a multiplicative model suggests a log-normal error term. In fact, fitting a model with a poor match of form and error term often results in a computation error, such as taking the log of a negative value! Thus, we are left with 16 candidate models ( $2^3$  combinations of effects times 2 model forms), ranging from the null model to the three effect growth model.

Each of these data sets was analysed using fixed effects standardization under each of the 16 competing growth models. As pith offset data was missing for most of the data sets, it was assumed that the first year recorded for each series was the pith of the tree. Because of the inaccuracy in the ages of each ring-width record, this analysis presents an overly pessimistic outlook on the value of including an age effect in tree-ring standardization and may impart an additional negative trend in the time effects of models that do not include an individual effect (Esper et al., 2003; Briffa and Melvin, 2011).

The average goodness-of-fit and model selection criteria were computed for each model and are listed in Table 1. Goodness-of-fit was fairly high, with the average  $R^2$  for the best model for a particular chronology of  $0.63 \pm 0.10$  dropping to  $0.57 \pm 0.11$  for adjusted  $R^2$ . Noise (in terms of the variance geometric standard deviation of the residuals) was around  $0.39 \pm 0.12$  for the best model.

The competing model selection criteria ( $\sigma$ ,  $R^2$ , adjusted  $R^2$ , likelihood, AIC, AICc and BIC) were somewhat consistent, with  $\sigma$ ,  $R^2$  and likelihood preferring the most complex models, BIC preferring less complex models with adjusted  $R^2$ , AIC and AICc

4518

intermediate (Fig. 5). By and large, it seems that  $\sigma$ ,  $R^2$  and likelihood are too liberal (as they have no penalty for added complexity) while BIC may be too conservative in many cases. Of the remaining three choices, adjusted  $R^2$  is neither as powerful or interpretable for model selection purposes as AIC, and AICc is simply a more correct version of AIC.  $R^2$ , adjusted  $R^2$  and  $\sigma$  are also inconsistent across model forms, as they are calculated in different ways depending on the assumed probability density function for the error term. In general, we recommend that model selection in tree ring standardization should be carried out using AICc or BIC (or AIC, as the correction as negligible for typical sample sizes) to allow for the use Akaike model weights.  $R^2$  and  $\sigma$  are still useful and interesting descriptive statistics though, and should be reported for each model as well.

For tree-ring width data, it appears that a multiplicative log-normal error term and model form is far more realistic than the additive normal equivalent. The fair likelihood-based model selection criteria preferred a multiplicative model over the additive equivalent 100 % of the time. On a smaller scale, this fact is easily confirmed by viewing histograms and quantile-quantile plots of the residuals of the fitted standardization model. The choice of probability density function is essential when significance tests, confidence or support intervals are desired.

Within these log-normal models, the flat detrending model was the most popular choice ( $G = IT$ ,  $P(AICc) = 83.5\%$ ,  $P(BIC) = 53.8\%$ ), followed by the full model ( $G = ITA$ ,  $P(AICc) = 14.4\%$ ,  $P(BIC) = 0.4\%$ ), and a productivity-only model ( $G = I$ ,  $P(AICc) = 2\%$ ,  $P(BIC) = 45\%$ ). The traditional regional curve standardization model,  $G = TA$  was selected only once by AICc, and never by BIC. BIC also selected the unstandardized chronology ( $G = T$ ) 5 times, and a strange time-insensitive model ( $G = IA$ ) 10 times. To investigate the effects of modern sample bias, we generated two competing chronologies. In the first (uncorrected) case, the growth model accounted for the effects of time and age using a smoothed age effect and a log-normal error term: The null model was never selected by any model selection criteria, suggesting that null hypothesis tests of tree-ring standardization are almost certainly a foregone

4519

conclusion. Overall, the individual effect had the strongest support, followed by the time effect, with the age effect being weakest across the chronologies as a whole.

$$G_{ita} = T_t s(a) e_{ita}^L \quad (22)$$

The second (corrected) model used adds an effect for the individual productivity of each tree:

$$G_{ita} = I_i T_t s(a) e_{ita}^L \quad (23)$$

This model was estimated using a generalized additive model with a gamma family and a log-link in *mgcv* (Wood, 2001). Smoothing was performed using thin-plate basis splines, with stiffness automatically selected using generalized cross-validation.

The corrected model was preferred 99.9% of the time by AICc, with even BIC preferring it 99.3% of the time (Fig. 6). Model fit statistics were substantially improved for the corrected three-effect model (Table 2).

We can detect the effect of modern sample bias by comparing the ratio of the uncorrected to the corrected age and time effects. For the chronologies sampled, it appears that the sign of modern sample bias varies by chronology. For 355 chronologies (29.5%), the ratio between the uncorrected and corrected time effects had a positive trend (by Spearman's  $\rho$  test for trend), reflecting a bias towards increasing growth by time. The remaining 847 chronologies (70.5%), displayed a negative trend, indicating a systematic underestimate of growth rates in more modern years (Fig. 7). Plotting the typical values of this ratio by time confirms this, revealing a slow, persistent negative bias in uncorrected chronologies over time, accelerating after about 1700 (Fig. 8).

## 6 Discussion and conclusions

Modern sample bias has been shown before (Briffa and Melvin, 2011; Voelker, 2011; Cooper et al., 2012; Melvin et al., 2012b) but the techniques developed in this paper

4520

allow a more detailed and comprehensive examination and correction. Contrary to prevailing opinion (Brienen et al., 2012a), modern sample bias does not always impart a positive bias on the standardized chronology but depends instead on the complex ecological interactions dictating survival and the vagaries of sampling. In fact, in 70.5% of the chronologies analysed, it had a negative effect instead. The ecology and sampling design behind these patterns is interesting in its own right and deserves much finer scale, context-specific study than given here.

D'Arrigo et al. (2008) suggest that modern sample bias may be responsible for the "divergence problem" in dendroclimatology, the widespread reduction in temperature sensitivity of tree-ring chronologies in recent decades. The generally negative trend induced by modern sample bias in recent years certainly suggests that this may be at least part of the problem.

More generally, the theoretical results of this paper clarify, simplify and extend regional curve standardization. Regional curve standardization is a biased implementation of signal-free standardization, while signal-free standardization is itself equivalent to the new effect regression standardization. Working within a regression framework improves the transparency of the standardization process, allows investigators to use classical regression tools such as AIC and, as demonstrated, facilitates investigation of alternate underlying models of tree growth.

The estimates of  $I$  are interesting in their own right, as measurements of tree-level productivity, independent of the effects of tree age and time period. This metric possesses several large advantages over traditional metrics of forest productivity such as site index or direct measurements of annual net primary productivity. While tree ring data may be more laborious than simple DBH and height measurements, it is still relatively cheap and simple to collect and analyze. Most importantly, it is free of confounding age- and time-related effects, a major challenge in most attempts to qualify site or tree quality. Integration of dendrochronological data into more traditional investigations of forest productivity (cf. Pokharel and Dech, 2012) could help control for these confounding effects. Intriguingly,  $I$  describes the productivity of individual trees. As a result,

4521

questions that are cost-prohibitive or impossible to answer using alternate productivity indexes are now feasible. The spatial structure of tree productivity, the effects of microtopography, or the success of tree breeding programs could all be readily quantified from tree ring data using the techniques outlined in this paper.

More extensive empirical validation and testing of the assumptions discussed in this paper is still needed and requires the detailed investigation of many real chronologies. To this end, R source code is available as Supplement and is planned for inclusion in `dplR`, the main dendrochronology package in R shortly (Bunn, 2008). Much of the existing work surrounding variance stabilization and chronology construction (Boreux et al., 2009; Cook and Peters, 1997; Nicault et al., 2010; Bontemps and Esper, 2011) may be significantly clearer in this model-based likelihood framework.

The tools presented in this paper may be useful in other areas of dendrochronology. Tree-specific differences in isotope values have been observed (Hangartner et al., 2012), and while maximum latewood density and other tree-ring proxies are relatively stable with age (Melvin et al., 2012a), microclimatic differences between trees may still drive persistent trends. The growth model presented here suggests that these effects may be segregated, even when using measurements other than ring width, so long as there is reason to believe that persistent effects due to individual trees, calendar year and age are present. Hangartner et al. (2012) encountered this problem, although it was not recognized as such, when attempting to combine partially overlapping tree ring isotope series of differing contemporaneous magnitude, observing the characteristic "jumps" in the final chronology when series began or ended. The approaches examined in that paper are not entirely applicable here; the first three suffer from "segment-length curse" (Cook et al., 1995), masking much of the low-frequency variability, and the fourth is difficult to extend to more than two overlapping series.

Brienen et al. (2012a) show that modern sample bias presents a real challenge to the use of modern tree ring chronologies in climate reconstruction and the detection of long-term shifts in growth, such as those that may result from climate change or carbon fertilization. Differences in productivity are at the root of these problems, and the

4522

techniques introduced here allow us to control for this heterogeneity and thus eliminate this bias.

Fixed effects standardization builds on a long history of dendrochronological standardization techniques but is at once simpler and more flexible. Sequential estimation techniques (classical regional curve standardization or individual series detrending) are grossly and needlessly inefficient and need to be replaced by their signal-free, maximum likelihood or least-squares equivalents in every case. We recommend the use of fixed effects standardization that accounts for differences in productivity in place of classical regional curve standardization in virtually all cases, removing trend-in-signal bias, differing-contemporaneous-growth-rate bias and modern-sample bias.

## Appendix A

### Sequential and alternating standardization algorithms

This appendix aims to clarify and precisely define sequential and alternating methods of estimating tree ring standardization models. In the first section, we show the parallels between flat detrending and regional curve standardization, and show why they fail to produce reliable estimates of the proposed growth models in real world scenarios. The second section precisely defines and revises the signal-free standardization algorithm for categorical effects, and proves that it always produces an unbiased estimate of the underlying growth model.

#### A1 Sequential standardization algorithms

Regional curve standardization (unsmoothed) (Briffa et al., 1992) and flat detrending (Cook and Kairiukstis, 1990) both use sequential estimation of categorical effect vectors to estimate a categorical time effect that is “free” of the effects of age (regional

4523

curve standardization) or individual productivity (flat detrending). The symmetry in the algorithms with respect to  $I \leftrightarrow A$  can be seen when we state their steps:

1. Take the mean of the growth data ( $\mathbf{G}$ ) for each [age/series]. This is the estimated effect vector for [age/individuals] ( $[\hat{A}/\hat{I}]$ )
2. Remove the the estimated effect vector for [age/individuals] from the growth data.
3. Take the mean of the growth data for each time (year). This is the estimated effect vector for time ( $\hat{T}$ ).
4. Rescale the effect vectors to a standard form.

The choice of mean (arithmetic/geometric) corresponds to the choice of error term (normal/log-normal). Similarly, the method of removing effects (subtracting/dividing) corresponds to the form of the model (additive/multiplicative). To limit the tedium of parallel derivations and make the proof as simple to follow as possible, the remainder of this appendix will use additive regional curve standardization with normal noise. The equivalent proofs for multiplicative models and/or log-normal noise are easily found by log-transforming the data and model or using a geometric mean, then applying the same techniques.

We can compare the estimates produced using this technique to the true growth model. In that case, the observed growth data is given by:

$$G_{ita} = T_t + A_a + \epsilon_{ita}^N \quad (\text{A1})$$

We convert the algorithm above into an algebraic expression. Because the data is not necessarily evenly distributed across each cell, we need to take into account the weight array  $\mathbf{W}$  which describes the sample depth at each combination of time and age. We frequently need to iterate over the entire set of observed individuals/ages/years. For this purpose, we define  $t$  to be the set of observed calendar years, such that  $t \in t$  means to iterate over each observed year.  $i$  and  $a$  are defined similarly for individuals and ages.

4524

The first step of the algorithm above, estimating the age effect (regional curve), can be written as:

$$\hat{A}_a = \frac{\sum_{t \in t} G_{ta} W_{ta}}{\sum_{t \in t} W_{ta}} \quad (\text{A2})$$

Removing the regional curve from the growth data (in this case subtracting as the assumed model is additive), we obtain an updated version of the growth matrix  $\mathbf{G}$  that we can denote  $\tilde{\mathbf{G}}$ :

$$\tilde{G}_{ita} = G_{ita} - \hat{A}_a \quad (\text{A3})$$

We then estimate the time effect (construct the standardized chronology) from the updated growth data (standardized series):

$$\hat{T}_t = \frac{\sum_{a \in a} \tilde{G}_{ta} W_{ta}}{\sum_{a \in a} W_{ta}} \quad (\text{A4})$$

Putting the steps together, we can describe the estimated effect vectors in terms of the true effect vectors.

$$\hat{A}_a = \frac{\sum_{t \in t} (T_t + A_a + \epsilon_{ita}^N) W_{ta}}{\sum_{t \in t} W_{ta}} \quad (\text{A5})$$

$$\hat{T}_t = \frac{\sum_{a \in a} \left( T_t + A_a + \epsilon_{ita}^N - \frac{\sum_{j \in j} (T_t + A_a + \epsilon_{ija}^N) W_{ja}}{\sum_{j \in j} W_{ja}} \right) W_{ta}}{\sum_{a \in a} W_{ta}} \quad (\text{A6})$$

4525

When the weight matrix is a constant ( $C$ ) at every position of  $t$  and  $a$ , these equations simplify considerably. First, we can condense our estimate of  $A$ .

$$\hat{A}_a = \frac{C \sum_{t \in t} (T_t + A_a + \epsilon_{ita}^N)}{C|t|} \quad (\text{A7})$$

$$\hat{A}_a = \frac{\sum_{t \in t} T_t}{|t|} + \frac{\sum_{t \in t} A_a}{|t|} + \frac{\sum_{t \in t} \epsilon_{ita}^N}{|t|} \quad (\text{A8})$$

$$\hat{A}_a = \bar{T} + \frac{|t| A_a}{|t|} + \overline{\epsilon_{ita}^N} \quad (\text{A9})$$

$$\hat{A}_a = A_a + \bar{T} + \overline{\epsilon_{ita}^N} \quad (\text{A10})$$

Where  $\bar{x}$  indicates the average over the elements of  $x$ . The estimated age effect vector is the sum of the true age effect and independent, identically normally distributed noise. As expected, only the relationship between the elements of the effects vector are well estimated, rather than their magnitude as the original growth model is singular. For additive models, we can accurately estimate the difference between any two elements of the effects vector, while for multiplicative models we can accurately estimate the ratio. If and only if the weight matrix is balanced, regional curve standardization produces unbiased estimates of the regional curve using the method of moments. Now that we have a simpler expression for the estimate of the age effect vector (regional curve,  $\hat{A}$ ), we can examine the updated growth matrix (standardized series,  $\tilde{\mathbf{G}}$ ).

$$\tilde{G}_{ita} = (T_t + A_a + \epsilon_{ita}^N) - (A_a + \bar{T} + \overline{\epsilon_{ita}^N}) \quad (\text{A11})$$

$$\tilde{G}_{ita} = T_t - \bar{T} + \epsilon_{ita}^N - \overline{\epsilon_{ita}^N} \quad (\text{A12})$$

As desired, the standardized series contain only information about time effects and noise. As most dendrochronologists have probably observed, the resulting standardized series are centered. In this case, because the model used is additive, they are

4526

centered around 0 but in the multiplicative case they will be (multiplicatively) centered around 1 instead. We can find the estimated time effect by substituting the updated growth matrix into Eq. (A4), remembering that the weight matrix is a constant ( $C$ ) in this scenario.

$$\widehat{T}_t = \frac{C \sum_{a \in a} \widetilde{G}_{ta}}{C|a|} \quad (\text{A13})$$

$$\widehat{T}_t = \frac{\sum_{a \in a} T_t - \bar{T} + \overline{\epsilon_{ita}^N} - \overline{\epsilon_{ita}^N}}{|a|} \quad (\text{A14})$$

$$\widehat{T}_t = \frac{(T_t - \bar{T})|a|}{|a|} + \frac{\sum_{a \in a} \overline{\epsilon_{ita}^N} - \overline{\epsilon_{ita}^N}}{|a|} \quad (\text{A15})$$

$$\widehat{T}_t = T_t - \bar{T} + \overline{\epsilon_{ita}^N} - \overline{\epsilon_{ita}^N} \quad (\text{A16})$$

Because we can never know the *true* values for the effect vectors, we can take the total (and thus average) error over the entire growth array ( $\overline{\epsilon_{ita}^N}$ ) to be 0.

$$\widehat{T}_t = T_t - \bar{T} + \overline{\epsilon_{ita}^N} \quad (\text{A17})$$

The estimate of the time effect is again centered and looks very similar to the estimate of the age effect. It is related to the true time effect vector by an unknowable constant offset and contains noise from the data in the corresponding year. As above for the regional curve, the standardized chronology produced using regional curve standardization is an unbiased estimate of the time effect if and only if the weight matrix is balanced. Furthermore, in this case switching the order in which the effects are estimated makes no difference to the family of effect vectors estimated. We can confirm that we've obtained reasonable estimates of the effect vectors by substituting them

4527

back into the growth model to generate predicted values ( $\widehat{\mathbf{G}}$ ).

$$\widehat{G}_{ita} = \widehat{T}_t + \widehat{A}_a \quad (\text{A18})$$

$$\widehat{G}_{ita} = (T_t - \bar{T} + \overline{\epsilon_{ita}^N}) + (A_a + \bar{T} + \overline{\epsilon_{ita}^N}) \quad (\text{A19})$$

$$\widehat{G}_{ita} = T_t + A_a + \overline{\epsilon_{ita}^N} + \overline{\epsilon_{ita}^N} \quad (\text{A20})$$

Examining the residuals ( $\mathbf{G} - \widehat{\mathbf{G}}$ ) can tell us if our estimator is biased.

$$G_{ita} - \widehat{G}_{ita} = (T_t + A_a + \overline{\epsilon_{ita}^N}) - (T_t + A_a + \overline{\epsilon_{ita}^N} + \overline{\epsilon_{ita}^N}) \quad (\text{A21})$$

$$G_{ita} - \widehat{G}_{ita} = \overline{\epsilon_{ita}^N} - \overline{\epsilon_{ita}^N} - \overline{\epsilon_{ita}^N} \quad (\text{A22})$$

Because the noise is assumed to be identically and independently distributed, the expected error is 0, meaning that for uniformly weighted designs, regional curve standardization (and flat detrending) is an unbiased estimator of a fixed effects growth model.

## A2 Alternating standardization algorithms

The signal-free standardization algorithms put forward by (Melvin, 2004; Melvin and Briffa, 2008) and in particular, (Briffa and Melvin, 2011) can be thought of as algorithms to alternately estimate orthogonal effects on growth. These techniques are powerful, removing trend-in-signal bias, but their theoretical foundation is somewhat unclear. In this section, we argue that signal-free-standardization can be reframed as an algorithm which serves to find a least-squares solution to the growth model as a whole.

The algorithm for signal-free (unsmoothed) regional curve standardization is given roughly as:

1. Align all series by their common age and find the mean value at each age.

4528

2. Divide the raw growth data by the regional curve at the same age to produce signal-free series.
3. Find the mean values at each year to produce the signal-free chronology.
4. Repeat steps 1 through 3 until convergence is reached (defined as an approximately zero-variance signal-free chronology).
5. Use this final signal-free regional curve to produce the standardized chronology by dividing the original growth data by the final regional curve.

Instead, we argue that signal-free standardization (as a whole) can be performed using this simpler algorithm:

1. Initialize the working growth data as the original growth data and the estimated effects as the null value (0 for additive models, 1 for multiplicative models).
2. Estimate an effect (via method of moments (taking the mean), least squares or maximum likelihood) and combine it with the previously estimated effect of that type (by adding for additive models and multiplying for multiplicative models).
3. Remove the change in the effect from the working growth data (by subtracting or dividing it from the growth data).
4. Repeat steps 2 and 3 for each orthogonal effect (orthogonal effects have independent sets of predictor variables).
5. Repeat steps 2–4 until convergence is reached, as defined by a zero-signal working growth data (all values of growth are  $\approx 0$  for additive models or  $\approx 1$  for multiplicative models).

Rather than retaining only a single estimated effect at a time (the regional curve/age effect above), we can store our updated estimates of each of the effects and skip the final step where we derive the second effect from the first estimated effect and the

4529

growth data. In this way, we can extend signal-free standardization to include more than two orthogonal effects.

For an additive regional curve standardization growth model with normal noise, we can write the algorithm more concretely.

1. Initialize the working growth data as the original growth data and the estimated effects as 0. Start the step counter at 0.

$$s_A := 0; s_E = 0 \quad (\text{A23})$$

$$\tilde{G}_{ita}^{s_G=0} = G_{ita} \quad (\text{A24})$$

$$\hat{A}_a^{s_E=0} = 0 \quad (\text{A25})$$

$$\hat{T}_t^{s_E=0} = 0 \quad (\text{A26})$$

2. Estimate the age effect by taking the average of the working growth data by age and add it with the previously estimated effect age effect.

$$\hat{A}_a^{s_E+1} = \hat{A}_a^{s_E} + \frac{\sum_{t \in t} \tilde{G}_{ita}^{s_G} W_{ita}}{\sum_{t \in t} W_{ita}} \quad (\text{A27})$$

3. Remove the change in the age effect from the working growth data by subtraction. Increment the step counter by 1.

$$\tilde{G}_{ita}^{s_G+1} = \tilde{G}_{ita}^{s_G} - (\hat{A}_a^{s_E+1} - \hat{A}_a^{s_E}) = \tilde{G}_{ita}^{s_G} - \frac{\sum_{k \in t} \tilde{G}_{ika}^{s_G} W_{ita}}{\sum_{k \in t} W_{ita}} \quad (\text{A28})$$

$$s_G := s_G + 1 \quad (\text{A29})$$



4. Estimate the time effect by taking the average of the working growth data by time and add it with the previously estimated effect time effect.

$$\hat{T}_t^{s_T+1} = \hat{T}_t^{s_T} + \frac{\sum_{a \in a} \tilde{G}_{ita}^{s_G} W_{ita}}{\sum_{a \in a} W_{ita}} \quad (\text{A30})$$

$$s_T := s_T + 1 \quad (\text{A31})$$

5. Remove the change in the time effect from the working growth data by subtraction. Increment the step counter by 1.

$$\tilde{G}_{ita}^{s_G+1} = \tilde{G}_{ita}^{s_G} - (\hat{T}_t^{s_E+1} - \hat{T}_t^{s_E}) = \tilde{G}_{ita}^{s_G} - \frac{\sum_{j \in a} \tilde{G}_{itj}^{s_G} W_{ita}}{\sum_{j \in a} W_{ita}} \quad (\text{A32})$$

$$s_G := s_G + 1; s_E := s_E + 1 \quad (\text{A33})$$

6. Repeat steps 2–5 until convergence is reached, as defined by the working growth data reaching 0 in every locations.

$$G_{ita}^* = 0 \quad (\text{A34})$$

Sequential standardization (regional curve standardization, flat detrending) simply stops this process after one loop through. At each step, the working growth data and the most recent effects combine to produce the original growth data.

$$G_{ita} = \hat{T}_t^{s_T} + \hat{A}_a^{s_A} + \tilde{G}_{ita}^{s_G} \quad (\text{A35})$$

All information is retained at each step. We can confirm that the effect vectors stabilize at our convergence condition by substituting  $\tilde{\mathbf{G}} = 0$  into the updating equations.

$$\hat{A}_a^{s_A+1} = \hat{A}_a^{s_A} + \frac{\sum_{t \in t} 0 \cdot W_{ita}}{\sum_{t \in t} W_{ita}} = \hat{A}_a^{s_A} \quad (\text{A36})$$

$$\hat{T}_t^{s_T+1} = \hat{T}_t^{s_T} + \frac{\sum_{a \in a} 0 \cdot W_{ita}}{\sum_{a \in a} W_{ita}} = \hat{T}_t^{s_T} \quad (\text{A37})$$

4531

The next question is whether the proposed algorithm actually converges to the stopping point. Unfortunately we have not managed, to date, to prove this and welcome proofs to this effect. In practice, the algorithm converges to a stable family of solutions (as determined by comparing the residuals at each step) relatively quickly. Small models and data sets typically take only a 2–5 iterations to stabilize while large models and particularly sparse data sets can take up to 20. Model fit statistics confirm our intuition that the modified signal-free algorithm is a good estimator of the growth model, producing values of likelihood and  $R^2$  extremely close to those obtained via maximum likelihood or least squares optimizers.

Finally, by Eq. (A35), we know that if the working growth data is 0, the estimated effect vectors are an unbiased estimator of the true effect vectors (with an unknowable constant offset).

$$T_t + A_a + e_{ita}^N = \hat{T}_t^* + \hat{A}_a^* + 0 \quad (\text{A38})$$

$$(T_t - \hat{T}_t^*) + (A_a - \hat{A}_a^*) = -e_{ita}^N \quad (\text{A39})$$

$$G_{ita} - \hat{G}_{ita}^* = -e_{ita}^N \quad (\text{A40})$$

$$E[G_{ita} - \hat{G}_{ita}^*] = 0 \quad (\text{A41})$$

Note that the family of solutions found does not depend on the order in which the effects are estimated, unlike in sequential standardization. Because this process is guaranteed to converge (at least for categorical effects), signal-free standardization results in an unbiased least-squares estimate of the original growth model, even in cases where the weight matrix is unbalanced. Signal-free standardization produces least squares estimate for first one effect then the next, to finally converge on a least squares solution for the full model.

4532

## Appendix B

### Likelihood ridges in three-effect models

As we saw before, the fixed effects growth models are singular. There is no single best solution, but instead a family of them. We can deal with this problem by rescaling the coefficients into a standard form. But a similar, less trivial difficulty arises when fitting three-effect (individuals, time and age) models to real tree ring data.

Because each sequential tree ring is both one year older and one year later in time than the preceding one, the tree-ring data for each tree follows a single diagonal along time and age. As a result, the estimated effect vectors can shift away from the true effect vectors in a peculiar way. In this section, we'll use an additive model with no noise for clarity.

$$G_{ita} = I_i + T_t + A_a \quad (B1)$$

We can think of the slice of the tree-ring array in which a single tree is found as a matrix, with calendar years as rows ( $R$ ) and ages as columns ( $C$ ). Each tree was born in a particular year, determining the diagonal ( $k$ ) in which the data is found. The oldest tree will be found along the main diagonal, starting at  $(t_1, a_1)$ , while younger trees will be found in the lower left triangle. The upper right triangle will always be empty unless the chronology contains trees that were born before the oldest complete tree but are missing data near the pith. The relative birth year of tree  $i$  ( $b_i$ ) can be related to the year and age (counting from the beginning of the chronology) in which any particular ring is found.

$$b_i = t_i - a_i \quad (B2)$$

The diagonal of the data is clearly related to the rows and columns in which the data is found.

$$k = R - C \quad (B3)$$

4533

And so then we can link the two perspectives:

$$t_i = C; a_i = R; b_i = -k_i \quad (B4)$$

We can change the coefficient of  $I_i$  without affecting any of the other diagonals, the coefficient of  $T_t$  without affecting any of the other columns and the coefficient of  $A_a$  without affecting any of the other rows. Suppose there is a demonic intrusion (Hurlbert, 1984), that results in each of our estimates of the effect vectors being linearly offset from the best estimate of our effect vectors by an amount related to the birth year, calendar year or age they represent. Let us describe these perturbed effect vectors as  $\check{I}$ ,  $\check{T}$  and  $\check{A}$  for individuals, time and age respectively. These perturbed effect vectors would look like:

$$\check{I}_i = \hat{I}_i + m_i B_i \quad (B5)$$

$$\check{T}_t = \hat{T}_t + m_t t \quad (B6)$$

$$\check{A}_a = \hat{A}_a + m_a a \quad (B7)$$

Which can of course be converted into the matrix perspective:

$$\check{I}_i = \hat{I}_i + m_i(-k) = \hat{I}_i - m_i(R - C) \quad (B8)$$

$$\check{T}_t = \hat{T}_t + m_t C \quad (B9)$$

$$\check{A}_a = \hat{A}_a + m_a R \quad (B10)$$

We can detect this intrusion using least-squares, maximum likelihood or similar if it results in a poor model fit. We can compare the pure predicted growth values ( $\hat{G}$ ) to the corrupted predicted growth values ( $\check{G}$ ) by looking at the difference between them.

$$\widehat{G}_{ita} - \check{G}_{ita} = (\widehat{I}_i + \widehat{T}_t + \widehat{A}_a) - (\check{I}_i + \check{T}_t + \check{A}_a) \quad (B11)$$

$$\widehat{G}_{ita} - \check{G}_{ita} = (\widehat{I}_i + \widehat{T}_t + \widehat{A}_a) - (\widehat{I}_i + m_i B_i + \widehat{T}_t + m_t t + \widehat{A}_a + m_a a) \quad (B12)$$

$$\widehat{G}_{ita} - \check{G}_{ita} = (\widehat{I}_i + \widehat{T}_t + \widehat{A}_a) - (\widehat{I}_i - m_i(R - C) + \widehat{T}_t + m_t C + \widehat{A}_a + m_a R) \quad (B13)$$

5 The best unbiased estimates of our effect vectors cancel:

$$\widehat{G}_{ita} - \check{G}_{ita} = -m_i(R - C) + m_t C + m_a R \quad (B14)$$

We can detect the perturbation if the true and corrupted predictions of growth are different, as the corrupted effect vectors will (by definition) provide a worse fit to the data. Now, if the intruding demon was particularly malicious, it would add precisely  
10 the right amount to each element of the effect vectors such that the predictions of the two model fits were indistinguishable in every cell of the growth data. In that case, the difference between the predictions ( $\widehat{G} - \check{G}$ ) would be 0.

$$0 = -m_i(R - C) + m_t C + m_a R \quad (B15)$$

$$15 \quad 0 = (m_t + m_i)C + (m_a - m_i)R \quad (B16)$$

It turns out, that if you set the slopes of the perturbations up just so, the intrusion is undetectable.

$$m_t = -m_i; \quad m_a = m_i; \quad m_t = -m_a \quad (B17)$$

20 Setting  $m$  to be the arbitrary slope of the perturbation as a whole, we can confirm that the predicted values are identical.

$$\check{G}_{ita} = \widehat{I}_i + m(R - C) + \widehat{T}_t + mC + \widehat{A}_a - mR \quad (B18)$$

$$\check{G}_{ita} = \widehat{I}_i + \widehat{T}_t + \widehat{A}_a = \widehat{G}_{ita} \quad (B19)$$

4535

Solutions to the three-effect growth model of the form given in Eq. (B18) all have the same likelihood, regardless of the value of  $m$  chosen. This creates a likelihood ridge in our solution space. Any residual-based optimizer we choose to use cannot distinguish between the “true” and “corrupted” solutions, leaving us unable to interpret  
5 the reported effects. Residual-driven metrics of model fit for these models are however perfectly valid, the  $R^2$ , likelihood, AIC and BIC for these malformed models are all correct, allowing us to perform model selection even in the face of demonic intrusions.

Similar models, including flexible GAM standardizations and many parametric alternatives (including constant basal area increment and negative exponential age trends)  
10 fall prey to the same problem as they must ultimately generate predictions at the same points. Resolving this problem is vexing, but we can use the sheer implausability of the reported outcomes in our favour.

In many cases, we can assume that the true effect vectors themselves arise from a specified distribution (in this case normal). The demonic intrusion dramatically skews  
15 these distributions, leaving clear proof of its visit. The three effects vectors are each generated through a different process and thus deserve a separate random process, with different levels of variability. From the corrupted effect vectors, we can obtain candidate corrected effect vectors ( $\widetilde{I}$ ,  $\widetilde{T}$ ,  $\widetilde{A}$ ) for any hypothesized level of corruption ( $\widetilde{m}$ ).

$$\widetilde{I}_i = \check{I}_i - \widetilde{m}(R - C) \quad (B20)$$

$$20 \quad \widetilde{T}_t = \check{T}_t + \widetilde{m}C \quad (B21)$$

$$\widetilde{A}_a = \check{A}_a - \widetilde{m}R \quad (B22)$$

When  $\widetilde{m} = m$ , we will have successfully removed the effects of the demonic intrusion, even though the likelihood of all the models are equivalent. For a given value of  $m$ , we  
25 can estimate the likelihood of a single corrected effect coefficient according to which class it belongs to:

$$\mathcal{L}(\widetilde{I}_i) \propto P(\widetilde{I}_i | \mathcal{N}(\mu_{\widetilde{I}}, \sigma_{\widetilde{I}})) \quad (B23)$$

4536

$$\mathcal{L}(\tilde{T}_t) \propto P(\tilde{T}_t | \mathcal{N}(\mu_{\tilde{T}}, \sigma_{\tilde{T}})) \quad (\text{B24})$$

$$\mathcal{L}(\tilde{A}_a) \propto P(\tilde{A}_a | \mathcal{N}(\mu_{\tilde{A}}, \sigma_{\tilde{A}})) \quad (\text{B25})$$

5 Allowing the mean of each normal distribution to be flexible with the estimated effect lets us accomodate the arbitrary scale of any one effect vector. We can compute the overall likelihood of the model as the product of the likelihood of each of the components of the effect vectors. By maximizing this quantity, or rather its logarithm, we can obtain reasonable values for each of our effect vectors that match our distributional ex-  
10 pectations. At this point, all that remains is a fairly trivial one-dimensional optimization.

In practice, this post hoc correction appears to work fairly well. It performs admirably on data sets in which the true effect vectors follow the expected distribution. Its application to real data is rather more approximate, the use of more specialized or empirical distributions for individual, time and age effects would help to ensure its reliability.

15 *Acknowledgements.* We would like to thank Waseem Ashiq and Wayne Bell for their feedback on the ecological causes of modern sample bias and Charles Canham for his helpful discussion and lectures on likelihood techniques.

This research was supported by a NSERC USRA grant to Jacob Cecile, a NSERC PDF grant to Chris Pagnutti and NSERC Engage and Discovery grants to Madhur Anand.

## 20 References

- Biondi, F.: Climatic signals in tree rings of *Fagus sylvatica* L. from the central Apennines, Italy, *Acta Oecol.*, 14, 57–71, 1993. 4504
- Black, B. A., Colbert, J. J., and Pederson, N.: Relationships between radial growth rates and lifespan within North American tree species, *Ecoscience*, 15, 349–357, 2008. 4507
- 25 Bontemps, J.-D. and Esper, J.: Statistical modelling and RCS detrending methods provide similar estimates of long-term trend in radial growth of common beech in north-eastern France, *Dendrochronologia*, 29, 99–107, doi:10.1016/j.dendro.2010.09.002, 2011. 4503, 4515, 4522  
4537
- Boreux, J.-J., Naveau, P., Guin, O., Perreault, L., and Bernier, J.: Extracting a common high frequency signal from Northern Quebec black spruce tree-rings with a Bayesian hierarchical model, *Clim. Past*, 5, 607–613, doi:10.5194/cp-5-607-2009, 2009. 4522
- Bowman, D. M. J. S., Brien, R. J. W., Gloor, E., Phillips, O. L., and Prior, L. D.:  
5 Detecting trends in tree growth: not so simple, *Trends Plant Sci.*, 18, 11–17, doi:10.1016/j.tplants.2012.08.005, 2013. 4505
- Brien, R. J. W., Zuidema, P. A., and Daring, H. J.: Autocorrelated growth of tropical forest trees: unraveling patterns and quantifying consequences, *Forest Ecol. Manag.*, 237, 179–190, doi:10.1016/j.foreco.2006.09.042, 2006. 4504
- 10 Brien, R. J. W., Gloor, E., and Zuidema, P. A.: Can we detect evidence for CO<sub>2</sub> fertilization from tree rings?, *Global Biogeochem. Cy.*, 26, GB1025, doi:10.1029/2011GB004143, 2012a. 4501, 4521, 4522
- Brien, R. J. W., Gloor, E., and Zuidema, P. A.: Detecting evidence for CO<sub>2</sub> fertilization from tree ring studies: the potential role of sampling biases, *Global Biogeochem. Cy.*, 26, GB1025, doi:10.1029/2011GB004143, 2012b. 4501, 4507, 4508, 4509, 4511
- 15 Briffa, K. R. and Melvin, T. M.: A closer look at regional curve standardization of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application, in: *Dendroclimatology V 11: Developments in Paleoenvironmental Research*, edited by: Hughes, M. K., Swetnam, T. W., and Diaz, H. F., Springer, the Netherlands, 113–145, 2011. 4501, 4502, 4509, 4515, 4518, 4520, 4528
- 20 Briffa, K., Jones, P., Bartholin, T., Eckstein, D., Schweingruber, F., Karlén, W., Zetterberg, P., and Eronen, M.: Fennoscandian summers from AD 500: temperature changes on short and long timescales, *Clim. Dynam.*, 7, 111–119, doi:10.1007/BF00211153, 1992. 4502, 4523
- Bunn, A. G.: A dendrochronology program library in R (dplR), *Dendrochronologia*, 26, 115–124, doi:10.1016/j.dendro.2008.01.002, 2008. 4522
- 25 Chao, K.-J., Phillips, O. L., Gloor, E., Monteagudo, A., Torres-Lezama, A., and Martínez, R. V.: Growth and wood density predict tree mortality in Amazon forests, *J. Ecol.*, 96, 281–292, doi:10.1111/j.1365-2745.2007.01343.x, 2008. 4509
- Cook, E. R. and Kairiukstis, L. A.: *Methods of Dendrochronology: Applications in the Environmental Sciences*, Springer, 1990. 4502, 4514, 4523
- 30 Cook, E. R. and Peters, K.: Calculating unbiased tree-ring indices for the study of climatic and environmental change, *Holocene*, 7, 361–370, 1997. 4522

- Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A., and Funkhouser, G.: The “segment length curse” in long tree-ring chronology development for palaeoclimatic studies, *Holocene*, 5, 229–237, doi:10.1177/095968369500500211, 1995. 4522
- Cooper, R. J., Melvin, T. M., Tyers, I., Wilson, R. J. S., and Briffa, K. R.: A tree-ring reconstruction of East Anglian (UK) hydroclimate variability over the last millennium, *Clim. Dynam.*, 40, 1019–1039, doi:10.1007/s00382-012-1328-x, 2012. 4501, 4520
- D’Arrigo, R., Wilson, R., Liepert, B., and Cherubini, P.: On the “Divergence Problem” in northern forests: a review of the tree-ring evidence and possible causes, *Global Planet. Change*, 60, 289–305, 2008. 4521
- Drobyshev, I. and Niklasson, M.: How old are the largest southern Swedish oaks?, A dendrochronological analysis, *Ecol. Bull.*, 53, 155–163, 2010. 4504
- Esper, J., Cook, E. R., and Schweingruber, F. H.: Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability., *Science*, 295, 2250–2253, doi:10.1126/science.1066208, 2002. 4501
- Esper, J., Cook, E. R., Krusic, P. J., Peters, K., and Schweingruber, F. H.: Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies, *Tree-Ring Res.*, 59, 81–98, 2003. 4518
- Evans, G. C.: *The Quantitative Analysis of Plant Growth*, 1st Edn., Blackwell Scientific Publications, Oxford, UK, 1972. 4504
- Fritts, H.: *Tree Rings and Climate*, 1st edn., Academic Press, London, 1976. 4515
- Grissino-Mayer, H. D. and Fritts, H. C.: The International Tree-Ring Data Bank: an enhanced global database serving the global scientific community, *Holocene*, 7, 235–238, doi:10.1177/095968369700700212, 1997. 4517
- Hangartner, S., Kress, A., Saurer, M., Frank, D., and Leuenberger, M.: Methods to merge overlapping tree-ring isotope series to generate multi-centennial chronologies, *Chem. Geol.*, 294–295, 127–134, 2012. 4522
- Huntington, E.: *The Secret of the Big Trees*, Tech. rep., Department of the Interior, Washington, D.C., 1913. 4507
- Hurlbert, S. H. .: Pseudoreplication and the design of ecological field experiments, *Ecological Monographs*, 54, 187–211, 1984. 4534
- Johnson, S. E. and Abrams, M. D.: Age class, longevity and growth rate relationships: protracted growth increases in old trees in the eastern United States, *Tree Physiol.*, 29, 1317–28, doi:10.1093/treephys/tpp068, 2009. 4507

4539

- Meko, D. M., Therrell, M. D., Baisan, C. H., and Hughes, M. K.: Sacramento River flow reconstructed to AD 869 from tree rings, *J. Am. Water Resour. As.*, 37, 1029–1039, doi:10.1111/j.1752-1688.2001.tb05530.x, 2001. 4504
- Melvin, T. M.: *Historical Growth Rates and Changing Climatic Sensitivity of Boreal Conifers*, PhD, University of East Anglia, 2004. 4501, 4504, 4515, 4528
- Melvin, T. M. and Briffa, K. R.: A “signal-free” approach to dendroclimatic standardisation, *Dendrochronologia*, 26, 71–86, 2008. 4502, 4515, 4528
- Melvin, T. M., Grudd, H., and Briffa, K. R.: Potential bias in “updating” tree-ring chronologies using regional curve standardisation: re-processing 1500 years of Tornetrask density and ring-width data, *Holocene*, 23, 364–373, doi:10.1177/0959683612460791, 2012a. 4501, 4503, 4504, 4507, 4522
- Melvin, T. M., Grudd, H. K., and Briffa, K. R.: Potential bias in “updating” tree-ring chronologies using regional curve standardisation: re-processing 1500 years of Tornetrask density and ring-width data, *Holocene*, 23, 364–373, doi:10.1177/0959683612460791, 2012b. 4520
- Nicault, a., Guiot, J., Edouard, J., and Brewer, S.: Preserving long-term fluctuations in standardisation of tree-ring series by the adaptive regional growth curve (ARGC), *Dendrochronologia*, 28, 1–12, doi:10.1016/j.dendro.2008.02.003, 2010. 4522
- Peet, R. K. and Christensen, N. L.: Competition and tree death, *BioScience – American Institute of Biological Sciences*, 37, 586–595, 1987. 4508
- Pokharel, B. and Dech, J. P.: Mixed-effects basal area increment models for tree species in the boreal forest of Ontario, Canada using an ecological land classification approach to incorporate site effects, *Forestry*, 85, 255–270, doi:10.1093/forestry/cpr070, 2012. 4504, 4521
- Rötheli, E., Heiri, C., and Bigler, C.: Effects of growth rates, tree morphology and site conditions on longevity of Norway spruce in the northern Swiss Alps, *Eur. J. For. Res.*, 131, 1117–1125, doi:10.1007/s10342-011-0583-4, 2011. 4509
- Schulman, E.: Longevity under adversity in conifers, *Science*, 119, 396–399, 1954. 4507
- Stephenson, N. L., van Mantgem, P. J., Bunn, A. G., Bruner, H., Harmon, M. E., O’Connell, K. B., Urban, D. L., and Franklin, J. F.: Causes and implications of the correlation between forest productivity and tree mortality rates, *Ecol. Monogr.*, 81, 527–555, doi:10.1890/10-1077.1, 2011. 4509
- Van Den Brakel, J. A. and Visser, H.: The influence of environmental conditions on tree-ring series of Norway spruce for different canopy and vitality classes, *Forest Science*, 42, 206–219, 1996. 4504

4540

- Vanninen, P. and Makela, A.: Fine root biomass of Scots pine stands differing in age and soil fertility in southern Finland, *Tree Physiol.*, 19, 823–830, doi:10.1093/treephys/19.12.823, 1999. 4508
- Voelker, S. L.: Age-dependent changes in environmental influences on tree growth and their implications for forest responses to climate change, in: *Size- and Age-Related Changes in Tree Structure and Function*, edited by: Meinzer, F. C., Lachenbruch, B., and Dawson, T. E., Vol. 4 of *Tree Physiol.*, Springer Netherlands, Dordrecht, 445–479, doi:10.1007/978-94-007-1242-3, 2011. 4501, 4520
- Wood, S. N.: Mgcv: GAMs and generalized ridge regression for R, *R news*, 1, 20–25, 2001. 4520
- Zuidema, P. A., Vlam, M., and Chien, P. D.: Ages and long-term growth patterns of four threatened Vietnamese tree species, *Trees*, 25, 29–38, doi:10.1007/s00468-010-0473-2, 2010. 4504

4541

**Table 1.** Average model fit statistics for the signal-free-standardization models fit to the analysed chronologies.  $\Delta^*/C$  values are calculated relative to the full multiplicative model.

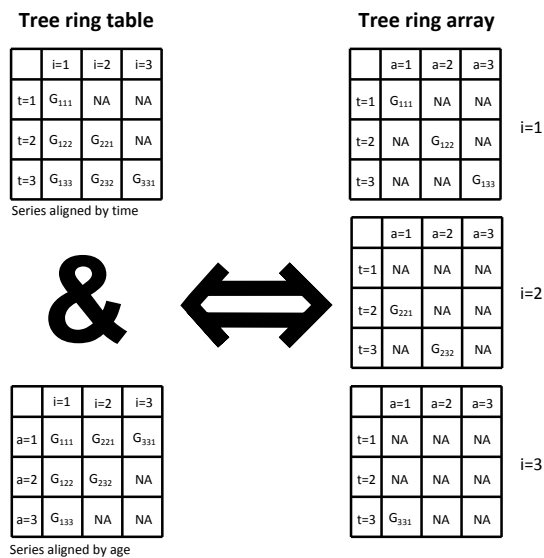
Model	$\sigma$	$R^2$	Adjusted $R^2$	$\Delta AIC$	$\Delta AIC_c$	$\Delta BIC$
$G = I \cdot T \cdot A \cdot e^L$	0.39 ± 0.12	0.60 ± 0.10	0.54 ± 0.12	0 ± 0	0 ± 0	0 ± 0
$G = I \cdot T \cdot e^L$	0.48 ± 0.15	0.40 ± 0.13	0.35 ± 0.14	1800 ± 2300	1800 ± 2300	-300 ± 1900
$G = I \cdot A \cdot e^L$	0.41 ± 0.13	0.56 ± 0.11	0.53 ± 0.12	-80 ± 300	-210 ± 700	-2100 ± 1200
$G = T \cdot A \cdot e^L$	0.46 ± 0.13	0.44 ± 0.14	0.35 ± 0.16	2000 ± 2600	2000 ± 2600	1800 ± 2500
$G = I \cdot e^L$	0.56 ± 0.18	0.22 ± 0.12	0.21 ± 0.12	2500 ± 2800	2400 ± 2800	-1500 ± 2500
$G = T \cdot e^L$	0.51 ± 0.16	0.32 ± 0.13	0.27 ± 0.14	2500 ± 2800	2400 ± 2800	300 ± 2500
$G = A \cdot e^L$	0.56 ± 0.16	0.20 ± 0.13	0.15 ± 0.14	3400 ± 3700	3400 ± 3700	1200 ± 2900
$G = e^L$	0.63 ± 0.19	0 ± 0	0 ± 0	394 000 ± 366 000	394 000 ± 366 000	390 000 ± 366 000
$G = I + T + A + e^N$	65 ± 102	0.62 ± 0.1	0.56 ± 0.12	49 000 ± 41 000	49 000 ± 41 000	49 000 ± 41 000
$G = I + A + e^N$	68 ± 107	0.57 ± 0.11	0.54 ± 0.12	49 000 ± 41 000	49 000 ± 41 000	47 000 ± 40 000
$G = I + T + e^N$	79 ± 121	0.42 ± 0.14	0.38 ± 0.15	50 000 ± 42 000	50 000 ± 42 000	48 000 ± 42 000
$G = T + A + e^N$	78 ± 122	0.44 ± 0.14	0.36 ± 0.17	51 000 ± 43 000	51 000 ± 43 000	51 000 ± 43 000
$G = I + e^N$	93 ± 143	0.23 ± 0.13	0.23 ± 0.13	51 000 ± 42 000	51 000 ± 42 000	47 000 ± 41 000
$G = T + e^N$	88 ± 138	0.32 ± 0.14	0.27 ± 0.15	51 000 ± 44 000	51 000 ± 44 000	49 000 ± 43 000
$G = A + e^N$	93 ± 142	0.21 ± 0.14	0.16 ± 0.15	52 000 ± 44 000	52 000 ± 44 000	49 000 ± 43 000
$G = e^N$	107 ± 165	0 ± 0	0 ± 0	82 000 ± 61 000	82 000 ± 61 000	77 000 ± 59 000

4542

**Table 2.** Average model fit statistics for the GAM models fit to the analysed chronologies.  $\Delta^*IC$  values are calculated relative to the full multiplicative model.

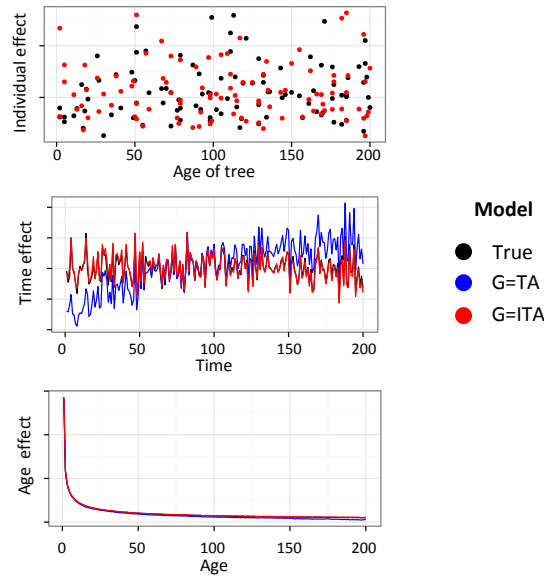
Model	$\sigma$	$R^2$	Adjusted $R^2$	$\Delta AIC$	$\Delta AICc$	$\Delta BIC$
$G = I \cdot T \cdot s(a) \cdot e^L$	$0.50 \pm 0.18$	$0.52 \pm 0.14$	$0.48 \pm 0.14$	$0 \pm 0$	$0 \pm 0$	$0 \pm 0$
$G = T \cdot s(a) \cdot e^L$	$0.57 \pm 0.2$	$0.37 \pm 0.15$	$0.32 \pm 0.16$	$1600 \pm 2300$	$1600 \pm 2400$	$1400 \pm 2200$

4543



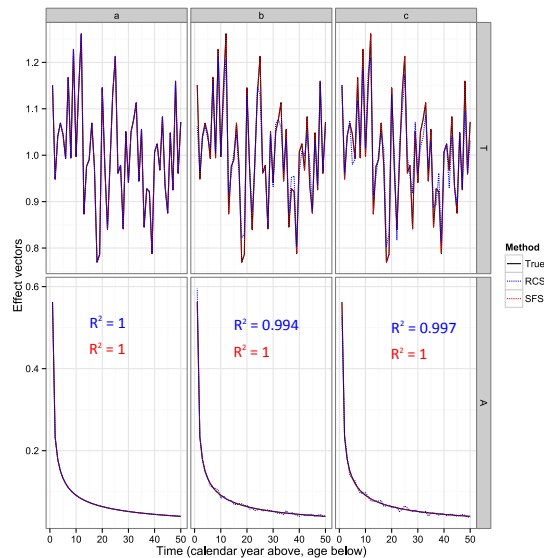
**Fig. 1.** Tree ring data is traditionally stored in a two-dimensional table (top left), with each column representing a different core and each row a different year. When dendrochronologists, wanted to look at the values by age, as when constructing the regional curve, the series need to be realigned so that the rows now match with the age of the rings (bottom left). Instead, we can store information about both time (calendar year) and age of the data in a three-dimensional tree ring array, here shown in slices by core (right).

4544



**Fig. 2.** An artificial modern sample bias scenario. The points in black show the true effects vector for synthetic data generated by a slow-biased survivorship curve. The points shown in colour are estimated effect vectors, estimated using signal-free standardization, with uncorrected estimates ( $G_{ita} = T_t A_a$ ) shown in blue and corrected estimates ( $G_{ita} = I_t T_t A_a$ ) shown in red.

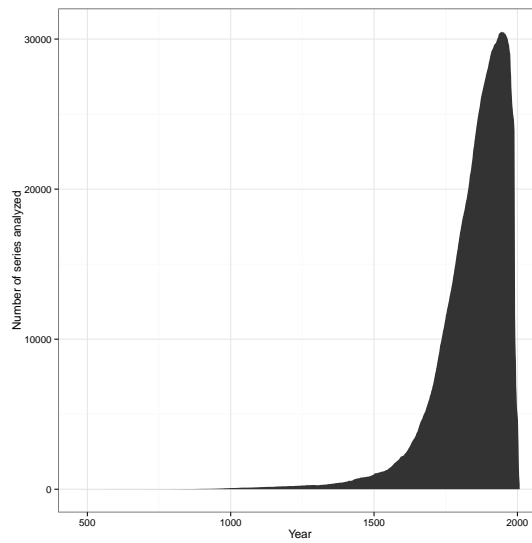
4545



**Fig. 3.** Comparing regional curve standardization and signal-free standardization with different data locations. In all cases, the same time and age effect vectors were used to generate the data. The points in black show the true effects vectors, while the points in colour show the estimated effects vector, estimated via regional curve standardization (blue) and signal-free regional curve standardization (red). **(a)** A complete, balanced design, **(b)** a randomly unbalanced design, data points were retained entirely at random and **(c)** a realistic unbalanced design, data points were retained according to a negative exponential survivorship simulation with uniform random birth years.

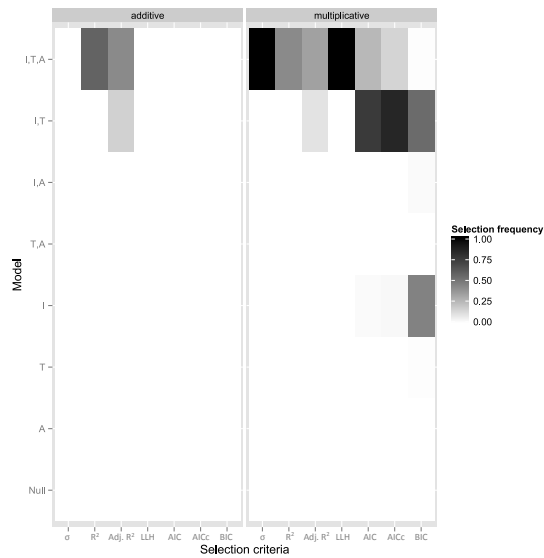
4546





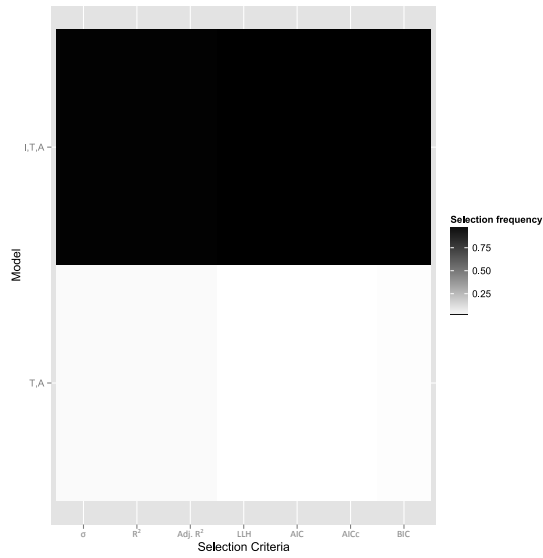
**Fig. 4.** Sample depth of the analysis across time (number of series analysed in each year).

4547



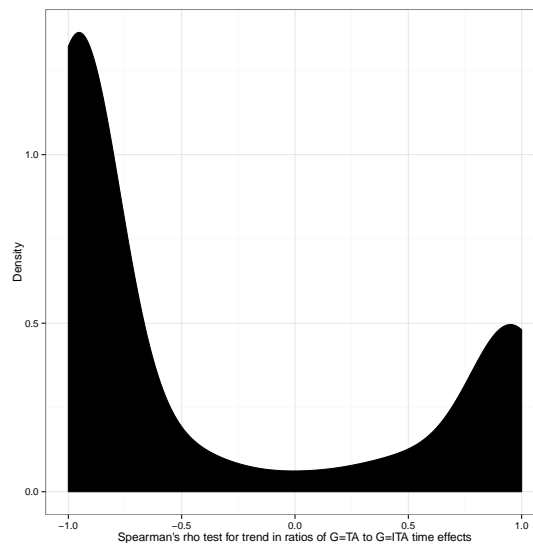
**Fig. 5.** Frequency of model selection over the analysed data set by each of the various model selection criteria. Additive models are shown on the left, and multiplicative models are shown on the right while the y-axis shows which terms are included in the growth model.

4548



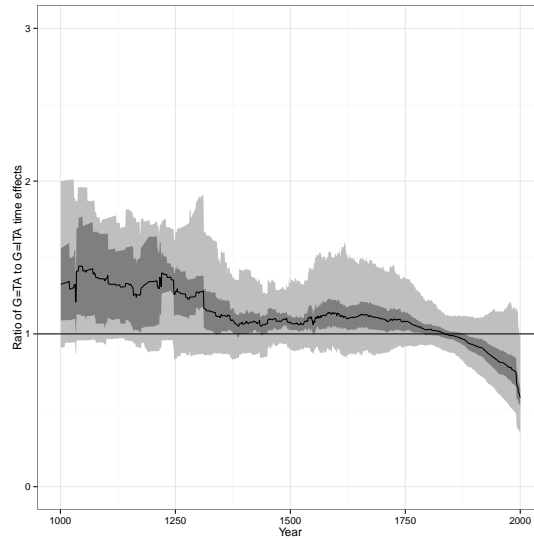
**Fig. 6.** Frequency of model selection over the analysed data set for the corrected ( $G = ITA$ ) and uncorrected ( $G = TA$ ) GAM standardization models by various model selection criteria. The y-axis shows which terms are included in the growth model.

4549



**Fig. 7.** Kernel density plot of the Spearman's  $\rho$  test statistic for trend of the ratio between the time effects of the  $G = TA$  and  $G = ITA$  standardizations. A negative value means that modern sample bias induces a spurious negative trend in the data with time while a positive value means that modern sample bias produces an overly positive estimate of the trend in the data over time.

4550



**Fig. 8.** Ratio between the time effects of the  $G = TA$  and  $G = ITA$  standardizations vs. time. The sign of this ratio corresponds to the direction of bias induced by modern sample bias in each year. The black line show the median value while the dark and light grey bands show the 40–60 % and 20–80 % quantiles respectively. This captures the typical values and their spread in a way that is consistent as the number of chronologies changes.