



**HAL**  
open science

# Extracting a Common Signal in Tree Ring Widths with a Semi-parametric Bayesian Hierarchical Model

Ophélie Guin, Philippe Naveau, Jean-Jacques Boreux

► **To cite this version:**

Ophélie Guin, Philippe Naveau, Jean-Jacques Boreux. Extracting a Common Signal in Tree Ring Widths with a Semi-parametric Bayesian Hierarchical Model. *Journal of Agricultural, Biological, and Environmental Statistics*, 2018, 23 (4), pp.550-565. 10.1007/s13253-018-0330-0 . hal-02410904

**HAL Id: hal-02410904**

**<https://hal.science/hal-02410904>**

Submitted on 3 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Extracting a Common Signal in Tree Ring Widths with a Semi-parametric Bayesian Hierarchical Model

Ophélie Guin, Philippe Naveau, Jean-Jacques Boreux

► **To cite this version:**

Ophélie Guin, Philippe Naveau, Jean-Jacques Boreux. Extracting a Common Signal in Tree Ring Widths with a Semi-parametric Bayesian Hierarchical Model. *Journal of Agricultural, Biological, and Environmental Statistics*, Springer Verlag, 2018, 23 (4), pp.550-565. 10.1007/s13253-018-0330-0 . hal-02410904

**HAL Id: hal-02410904**

**<https://hal.archives-ouvertes.fr/hal-02410904>**

Submitted on 3 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Extracting a common signal in tree ring widths with a semi-parametric Bayesian hierarchical model**

Ophélie Guin, Philippe Naveau, Jean-Jacques Boreux

## **Abstract**

There are numerous statistical challenges involved in the general field of climate reconstructions, including the preprocessing of raw data, often called standardization. This paper focuses on this essential but often overlooked preprocessing stage for one of the most used climate proxy, tree ring widths. One basic premise of dendroclimatology (dendron = tree) is that tree ring widths are assumed to contain relevant information about past climate. By going back to the data source, we focus on improving uncertainty assessments and more accurately identifying a climatic signal.

Tree ring width logarithms measured on a given tree are classically decomposed into an individual age effect and a common signal shared by all trees from the same site. Through informative priors, we assume that the individual age effect component lives on a narrow frequency band. This corresponds to the a priori knowledge that individual trees have a smooth aging process. In contrast, the environmental signal shared by all trees is not assumed to belong to a specific frequency range. From a statistical perspective, the search of this common signal shared by a series of tree ring width logarithms can be viewed at inferring the different components of a specific

additive model. Compared to past dendroclimatology studies, we propose a semi-parametric Bayesian hierarchical model that offers the possibility to capture low and high frequencies in tree ring widths. Our new model is tested on simulated data and applied to *Pinus halepensis Mill.* ring widths recorded in French Mediterranean.

## 1 Introduction

One key issue to understanding past and recent climate changes is to derive, study and apply efficient statistical procedures to reconstruct past records of temperatures and precipitation. Direct measurements of such climatological variables are missing whenever the instrumental record length is shorter than the period of interest. The so-called proxies, i.e. indirect measurements, offer the raw material to reconstruct past chronologies in such situations. Proxies should contain records of past climates, but they are also tainted by important and complex non-climatic factors, e.g. local ecological effects. This explains that most published climate reconstruction results/methods in the statistical literature (e.g., Li et al., 2010b; Wahl et al., 2010; Cressie and Tingley, 2010; Li et al., 2010a; Smith, 2010a; McShane and Wyner, 2011b; Smith, 2010b; Christiansen et al., 2009; Esper et al., 2002; Tingley et al., 2012; Tingley and Huybers, 2013; Werner and Tingley, 2015) generally do not start with raw proxies, but rather have analyzed a series of already pre-processed data products. In this paper, we focus on this pre-processing step for one of most used proxies, tree ring widths.

Since the work of Douglass (1920, 1936), there has been an active and extensive research activity dedicated to the field of dendrochronology (dendron = tree and chronos = time) that study tree rings to analyze temporal and spatial patterns of processes in the physical and social sciences (e.g. Cook and Kairikukstis, 1990; Gornitz, 2009; Cook et al., 1999; Evans et al., 2006; Melvin and Briffa, 2008). One major advantage of dendrochronology is that

annual ring formation makes the time sampling, one ring per year, constant in zones that have a distinct dormant season related to cold weather. Figure 1 shows the lifetime of the seventeen trees that are used in our application. The x-axis corresponds to the years (1867-1993) and the y-axis to the tree label. This graph illustrates the lifetime heterogeneity among individual trees. Some trees like 11 has a short record of a few decades while others like 5 covers more than one century.

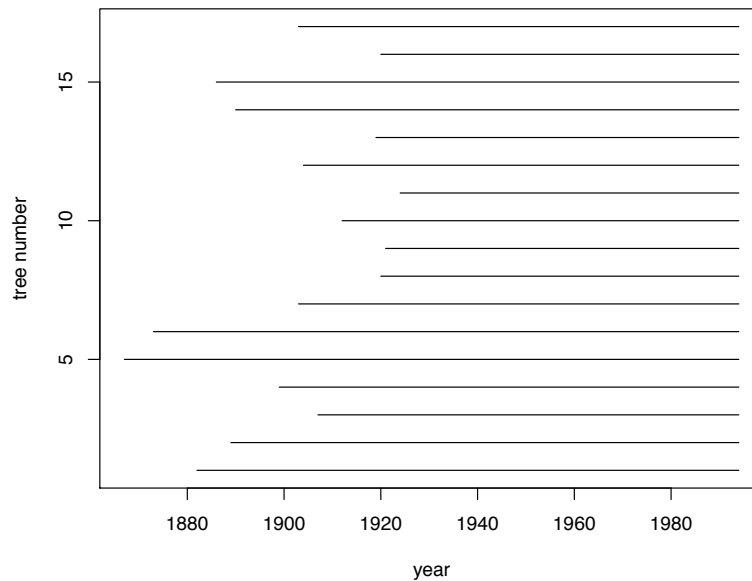


Figure 1: The lifetime of the seventeen trees that has been used in our application, see Section 4.2. The x-axis corresponds to the years (1867-1993) and the y-axis to the tree label.

Typically the number of sampled trees diminishes as one goes back in time. Finding older trees becomes more and more arduous for the field experimenter. This classical issue in paleo-studies implies that the assessment of uncertainty can be non-trivial and should vary in time.

Given tree ring widths measurements from a given site, how should one extract a common signal from this data set ? Our underlying assumption is that the common signal shared by all the trees from a particular site should be due to an environmental factor, possibly climatic but not necessarily. Statistically, we aim to propose and study a Bayesian inference

scheme capable of extracting latent individual and common signals. Essential elements of our analysis are the modeling of varying uncertainties due to tree lifetime heterogeneity, bypassing the need of parametric forms for either individual or common signals and taking into account the prior information given by dendroclimatologists.

## 2 Dendrochronology: data and methods

To present and discuss our approach, we have analyzed a set of seventeen *Pinus halepensis Mill.* French Mediterranean coast where tree ring width measurements were studied by Nicault et al. (2001). This region is climatically characterized by a Mediterranean climate with clear summer droughts. Nicault et al. (2001) identified possible relationships between tree growth measurements and climatic factors in the same geographical region and with the same tree species. Hence this past study provides a referential for our extraction procedure and has been beneficial for discussing and interpreting our approach.

Figure 3 displays seventeen *Pinus halepensis Mill* tree ring width logarithms from the "Rognac" site. The group of seventeen time series illustrates the difficulty of finding a common signal; each time series having its own time length (see Figure 1), its own growth trend and a large variability.

A classical decomposition to represent yearly individual tree ring growths is the following additive model, often called the linear aggregate model (Cook, 1990; Buckley, 2009),

$$\log(\text{ring width}) = F_t + G_t + D_t + \text{unexplained variability} \quad (1)$$

where  $t$  represents a year,  $G_t$  corresponds to the age-related trend due to normal physiological aging processes,  $F_t$  to the climatically-related environmental signal and  $D_t$  to distur-

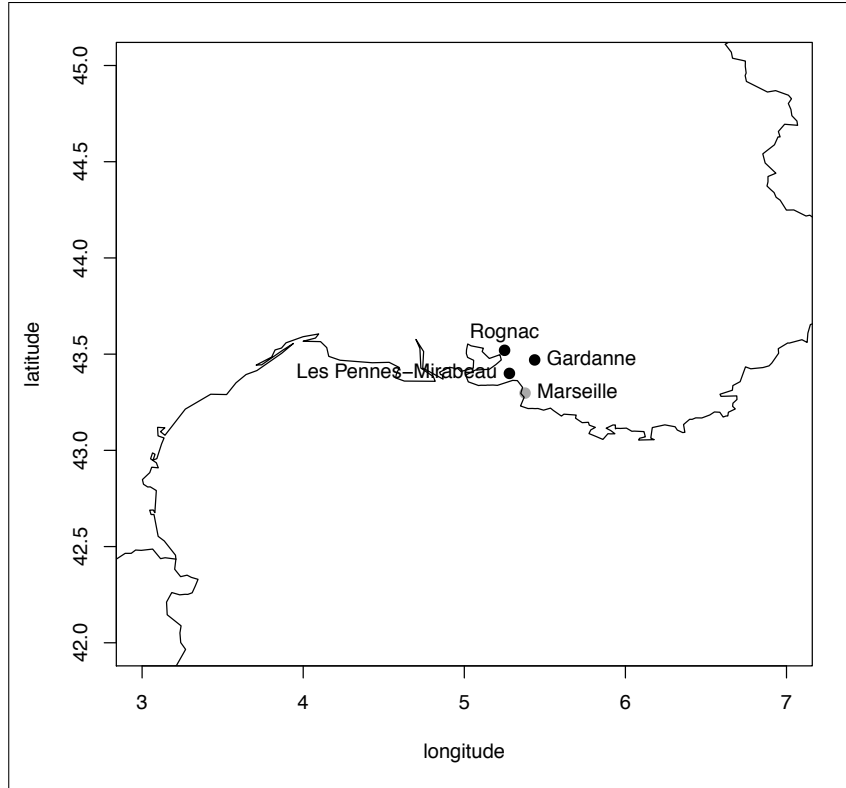


Figure 2: The "Rognac" site located in the South of France, near Marseille, where *Pinus halepensis* Mill tree ring width logarithms shown in Figure 3 were recorded.

bance factors, either within the forest stand or outside of it (e.g., insect outbreaks or fires). In most studies, the site of interest is selected in order to minimize the possibility of internal and external ecological processes affecting tree growth. In this paper, we follow this hypothesis and  $D_t$  is set to zero. Ring width logarithm is used because tree ring multiplicative properties are well known.

Individual trees at a environmentally homogenous site have their own physiological aging process  $G_t$  (e.g., Fang et al., 2010; Cook et al., 1990; Esper et al., 2002). The function  $G_t$  is often represented by parametric curves such as a negative exponential function or more complex smooth functions such as a Hagershoff curve (see Figure 1 in Fang et al., 2010). Hence, one common principle in representing  $G_t$  resides in its smoothness, i.e. it is assumed to be a low frequency signal. Besides this prior knowledge, it is difficult to impose an universal parametric form because it should depend on the tree specie and the

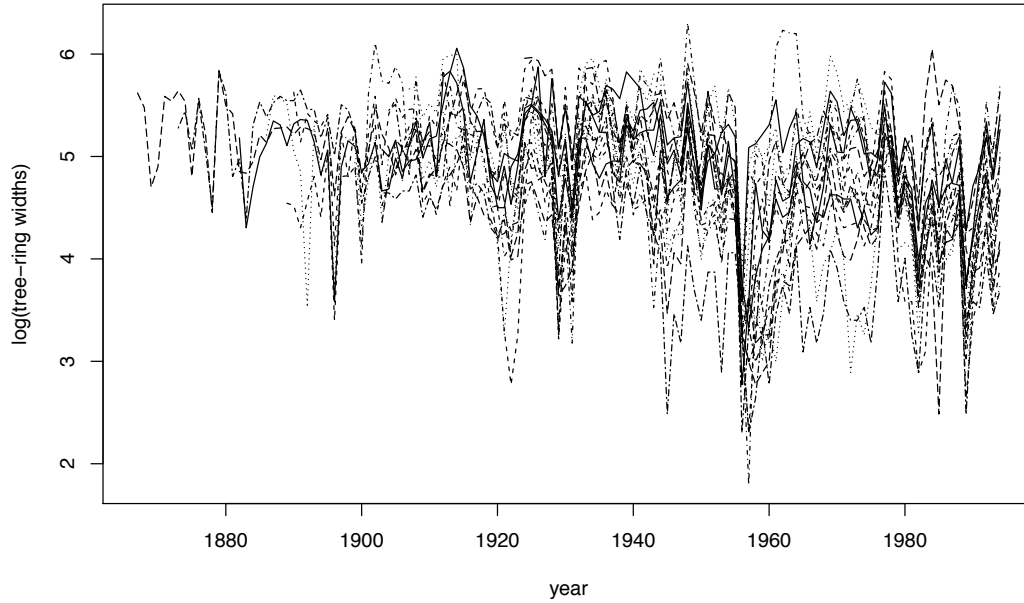


Figure 3: Seventeen *Pinus halepensis* Mill tree ring width logarithms from the "Rognac" site located in Figure 2. The x-axis (years) covers the period 1867 – 1993 and each time series has a different length, see Figure 1.

site of interest and most tree rings have complex temporarily behaviors, see Figure 3. In this paper the low frequency information has been used to guide some of our prior distributions choice within our Bayesian modeling.

The function  $G_t$  in Equation (1) is often considered as a nuisance term to be removed and the main element for most dendroclimatologists resides in finding the component  $F_t$ . This quest leads to the so-called *standardization* problem and remains an object of active research (Melvin and Briffa, 2008; Nicault et al., 2010). Standardization aims at calculating a dimensionless chronological index that reflects an hidden yearly common environmental signal shared by all individual trees. This is encapsulated by Equation (3) from Melvin and Briffa (2008)

$$\text{Ring width} = \text{Chronology index} \times \text{Expected growth} \times \text{Error},$$

which is simply a multiplicative version of Equation (1) (we assume  $D_t = 0$ ) and the so-



called expected growth corresponds to  $\exp(G_t)$ . To eliminate the age-affect  $G_t$  or equivalently the expected growth, an age-related trend is first estimated for each individual tree. This is classically done by implementing a univariate parametric regression (e.g., negative exponential curve (Fritts et al., 1969)) or a semi-parametric one (e.g., Cook and Peters, 1981; Barefoot et al., 1974). Second, each ring width is then divided by the corresponding fitted value obtained from the regression (e.g, see Figure 3 in Fang et al., 2010) and (e.g, see Cook et al., 1990; Esper et al., 2002). This produces the so-called tree indices that should have a mean of approximately equal to one. Third the so-called chronology time series, i.e. the standardized dimensionless index, is calculated as the arithmetic mean of all tree indices for a year.

Although the underlining model beneath this series of statistical steps is similar to a multiplicative model on the raw data, or to an additive model like (1) for the ring width logarithms, the inference aspect of this standardization approach is not clear. Each step is made independently of the previous one. Consequently, calculating valid estimates and confidence intervals of the final output, the dimensionless index, remains challenging. The fact that the variable of interest  $F_t$  is common to all time series implies that the inference scheme should be performed in a jointly manner by treating all trees at once. But univariate inferential techniques have been used at each step and error estimates made of each step are decoupled from each other. This is also another drawback. By construction, the classical standardization scheme takes out all the low frequency information contained in tree rings. This due to the removal of the age-effect. Individually a univariate regression cannot make the distinction between two low frequency components, see  $G_t$  and  $F_t$  in (1). Only, by treating the full set of trees jointly, one can hope to discriminate between a common smooth climate signal and other individual ones. For the practitioner, this drawback is very important. It implies that the classical standardization scheme is only adapted to capture annual variability but not decadal or centennial trends from tree rings. This is also true for other standardization based on ARMA modeling (Guiot, 1987). Recently we (Boreux

et al., 2009) proposed and studied a Bayesian hierarchical model to extract hidden signal but again, it was under the hypothesis that smooth trends have already been removed by a preprocessing of individual tree rings.

The Regional Curve Standardization (RCS) and the Adaptive Regional Growth Curve (Nicault et al., 2010) are attempts to preserve low frequency climatic information contained into tree rings. The former is based on producing a regional biological growth trend obtained by averaging ring widths that have been aligned according to their biological age (not their chronological age). This requires a large number of trees. Another assumption here is that this structural form is the same for each tree and does not vary in time. Coming back to (1), this means that  $G_t$  comes from an unique profile that has been shifted according to the tree age. This is rather strong limitation because individual growth rate trees can differ according to soil conditions, competition and other factors governing productivity. To circumvent this issue, Nicault et al. (2010) proposed to regress tree rings according to biological age, initial and maximum productivities using a neural network. The initial and maximum productivities are defined as the average of the first 10 rings and the maximum value during the first 50 years over an individual smoothed growth profile, respectively. Hence the computation of the predictors is tailored to the application at hand and may be difficult to generalize to other cases without an expert in dendrochronology. In addition, the inference properties of the method are not clear to us because tree rings seem to be used as predictant and as data for building the predictors.

### **3 Model description and its inference**

During the last two decades, Bayesian Hierarchical Models (BHM) have blossomed in climate sciences. One appealing idea in BHMs is to probabilistically decompose a complex climatic process and its relationships to observations in several simple components

throughout a hierarchy of layers. BHM handle efficiently the uncertainty assessment of each layer by clearly identifying prior and posterior distributions of underlining processes. For an introduction to such models, see e.g. Gelman et al. (2003) and the recent book of Cressie and Wikle (2011). Examples of BHM applied to climate issues could be as follows. Berliner et al. (2000) studied long-lead predictions of Pacific Sea Surface Temperatures via Bayesian Dynamic Modeling. Schliep et al. (2010) estimated extreme precipitation from regional climate models by combining BHM and extreme value theory. Tebaldi et al. (2010) characterized uncertainties of future climate change projections using BHM and Sahu et al. (2007) studied space-time ozone modeling for assessing trends. Haslett et al. (2006) investigated the problem of reconstructing prehistoric climates from lake sediment cores by making use of pollen assemblages. Other recent examples of BHM applied in paleoclimate can be found in (McShane and Wyner, 2011b; Berliner, 2011; Craigmile and Rajaratnam, 2011; Davis and Liu, 2011; Haran and Urban, 2011; Holmström, 2011; Kaplan, 2011; McIntyre and McKittrick, 2011; McShane and Wyner, 2011a; Nychka and Li, 2011; Rougier, 2011; Gavin A. Schmidt and Rutherford, 2011; Smerdon, 2011; Tingley, 2011; Wahl and Ammann, 2011; Tingley and Huybers, 2013; Werner and Tingley, 2015).

Schematically, uncertainty in BHM is spread over different layers, usually three. The base level, called the *data layer*, characterizes observations, e.g. tree ring profiles in our case. The second level in the hierarchy, called the *process layer*, models latent processes that drive the growth of such rings, tree-to-tree and regional variations. In this second layer, one can start incorporating temporal processes, e.g. individual age effects and the hidden common environmental factor. The third level, called the *parameter layer*, consists of the information concerning prior parameters distributions that control the process layer.

In dendrochronology, Hooten and Wikle (2007) investigated with a BHM shifts in the spatio-temporal growth dynamics of shortleaf pine. These authors did not work with raw tree measurements but with chronology indices, i.e. already preprocessed and standardized

data. They linked these chronologies with drought information like the Palmer Drought Severity Index. Concerning the standardization issue and BHM, we (Boreux et al., 2009) extracted an inter-annual high frequency signal from detrended tree ring series and consequently, smooth trends were also overlooked. Compared to these past studies, our goal is to add the flexibility of modeling non-parametric trends that can capture low frequency changes for the age effect and higher frequency variations with trend for the hidden common environmental signal.

Denote  $\mathbf{y}_j = (y_j(t_1), \dots, y_j(t_n))^T$  the tree ring width logarithms vector produced by tree  $j$  over the period of interest  $(t_1, \dots, t_n)$ . Equation (1) provides the foundation of our data layer that can be expressed with the common notations used by the Bayesian community as

$$\mathbf{y}_j | \mathbf{g}_j, \mathbf{f}, \sigma^2 \sim \mathbf{g}_j + \mathbf{f} + \sigma^2 \mathcal{N}_n(\mathbf{0}_n, \mathbf{I}_n), \text{ with } j = 1, \dots, p, \quad (2)$$

where the unknown  $\mathbf{f} = (f(t_1), \dots, f(t_n))^T$  represents the hidden common signal, see  $F_t$  in (1), the unknowns  $\mathbf{g}_j = (g_j(t_1), \dots, g_j(t_n))^T$  correspond to the individual age effect for each tree  $j$ , see  $G_t$  in (1),  $\mathbf{0}_n = (0, \dots, 0)^T$  and  $\mathbf{I}_n$  denotes the identity matrix of size  $n$ . Measurement uncertainty is modeled as a zero mean Gaussian vector with covariance  $\sigma^2 \mathbf{I}_n$  and each tree record  $[\mathbf{y}_j | \mathbf{g}_j, \mathbf{f}, \sigma^2]$  is supposed to be mutually independent of each other ("i.e. our observations are assumed to be conditionally independent with respect to the model parameters). In our application shown in Figure 3, the number of tree  $p$  is equal to seventeen and the time period is defined as  $t_1 = 1867$  and  $t_n = 1993$ . The tree length variation displayed in Figure 1 implies that  $\mathbf{g}_j$  starts with a series of missing values for most trees.

To go one step further in our Bayesian hierarchy, we need to define the process layer, i.e. to set priors for  $\mathbf{g}_j$ ,  $\mathbf{f}$ , and  $\sigma^2$ . In contrast to past dendrochronological studies that imposed a parametric form for  $\mathbf{g}_j$  or  $\mathbf{f}$  or both, we opt to describe both functions as semi-parametric

splines viewed within a BHM framework.

Splines modeling was formulated by Reinsch (1967) and developed by many author (e.g., Eubank, 1999; Wand and Jones, 1995; Fan and Gijbels, 1996). Within the Bayesian framework, Kimeldorf and Wahba (1970) demonstrated that specific forms of spline smoothing correspond to Bayesian estimates under a class of improper Gaussian prior distributions on function spaces. For the classical non-parametric regression problem  $\mathbf{y} = \mathbf{f} + \sigma^2\mathcal{N}(\mathbf{0}, \mathbf{I})$ , Wahba (1978) proposed and studied a particular partially improper Gaussian prior for the trend  $\mathbf{f}$

$$\mathbf{f}|\tau^2 \sim \mathcal{N}_n(0, \tau^2\mathbf{K}^-) \quad (3)$$

where  $\tau^2 = \sigma^2/\lambda$  and  $\lambda \geq 0$  is the smooth parameter of the classical penalized sum of squares criterion  $\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int (f''(\mathbf{x}))^2 d\mathbf{x}$  that is minimized over all functions  $f(\mathbf{x})$  such that the integral exists. In (3),  $\mathbf{K}^-$  refers to a generalized inverse of a matrix  $\mathbf{K}$ , with the understanding that an eigenvalue of zero for  $\mathbf{K}$  gives an eigenvalue of  $+\infty$  for  $\mathbf{K}^-$ . In the case of smoothing splines  $\mathbf{K}$  is linked to the penalty  $\int (f''(\mathbf{x}))^2 d\mathbf{x} = \mathbf{f}^T \mathbf{K} \mathbf{f}$ . Hastie and Tibshirani (1990, 2000) showed that this prior covariance matrix  $\mathbf{K}^-$  is equal to  $\mathbf{B}\mathbf{\Omega}^- \mathbf{B}^T$  evaluated at the data. Let  $n_u$  be the number of unique value of  $\mathbf{x}$ , the basis matrix  $\mathbf{B}$  consist of the vector of  $n_u + 2$  cubic B-splines basis functions  $b(\mathbf{x})$  (de Boor, 1978) evaluated at the  $n_u$  sample values  $x_i$  and the penalty matrix  $\mathbf{\Omega}$  has elements  $\Omega_{ij} = \int b_i''(x)b_j''(x)dx$ . Priors for the smoothing parameter or the variances  $\sigma^2$  and  $\tau^2$  belongs to the parameter layer of the Bayesian hierarchy and they have to be fixed. Hastie and Tibshirani (1990, 2000) suggested to use proper inverse gamma priors for the variance components  $\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$  and  $\tau^2 \sim \mathcal{IG}(a, b)$ . In this work, the hyper-parameters of the inverse gamma distribution are (0.0001, 0.0001).

Following the work of Wahba (1978) and Hastie and Tibshirani (1990, 2000), we assume

the same type of priors for  $\mathbf{g}_j$  and  $\mathbf{f}$  (see Equation (3))

$$\mathbf{f}|\tau_0^2 \sim \mathcal{N}_n(0, \tau_0^2 \mathbf{K}^-) \text{ and } \mathbf{g}_j|\tau_j^2 \sim \mathcal{N}_n(0, \tau_j^2 \mathbf{K}^-), \text{ for all } j = 1, \dots, p.$$

At this stage, our model is too versatile and associated with identifiability issues". For example, if all  $\mathbf{g}_j$  are proportional to  $\mathbf{f}$ , it is impossible to distinguish  $\mathbf{f}$  from  $\mathbf{g}_j$ . Additional constraints are needed and these have been chosen according to basic tree ring characteristics. From Fang et al. (2010); Cook et al. (1990); Esper et al. (2002), we know that the individual age effect function  $\mathbf{g}_j$  should be very smooth because individual tree growth is a rather slow and cumulative process. In contrast, we assume that the hidden signal shared by all trees  $\mathbf{f}$  should capture environmental variabilities that correspond to rapid (yearly or decadal) or slow (centennial) changes. This means that the frequency range of  $\mathbf{g}_j$  is assumed to be much narrower than the one of  $\mathbf{f}$ . To illustrate this difference, Figure 4 displays simulations that mimic this phenomenon. The top and middle panels represent a simulated common signal  $\mathbf{f}$  and simulated individual tree growth signals  $\mathbf{g}_j$ , respectively. In this idealized example, one can see that the functions  $\mathbf{g}_j$  do not reproduce the rapid variations seen in  $\mathbf{f}$ . To test the resilience of our method, a slow positive trend was also included into  $\mathbf{f}$  here and this adds difficulties to separate  $\mathbf{f}$  from  $\mathbf{g}_j$ , see Section 4.1.

The smoothness information can be translated into informative prior choice of the smoothness parameters  $\tau_j^2$  for  $j = 0, \dots, p$ . For comparison and interpretation reasons, we substitute  $\tau_j^2$  by a parameter that lives on the interval  $[0, 1]$ ,  $\phi_j = \frac{\tau_j^2}{\tau_j^2 + \sigma_j^2}$ , for all  $j = 0, \dots, p$ . If  $\phi_j$  takes a value near zero, then it means that the curve is very smooth. Beta priors for these parameters seem natural.

To improve identifiability and interpretability, the condition  $\mathbf{g}_1 + \dots + \mathbf{g}_p = 0$  is also imposed. If  $p = 1$ , we have only one tree and it is impossible to distinct  $\mathbf{f}$  from  $\mathbf{g}_1$  (under this case, the latter is set to zero). As the number of trees  $p$  increases, individual growth

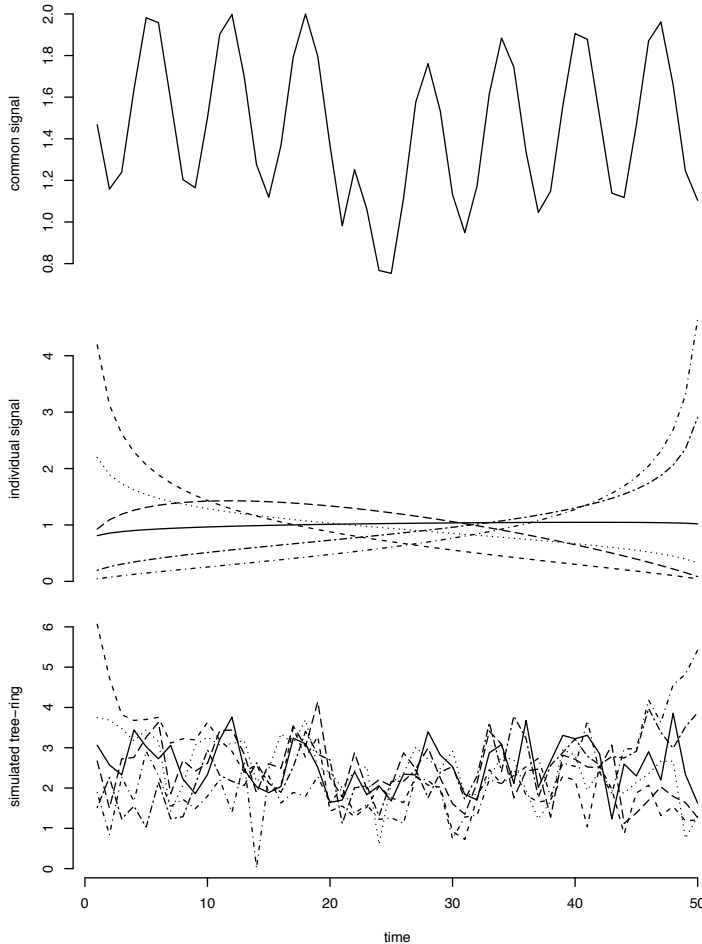


Figure 4: Simulations of logarithmic tree ring widths from the additive model (2). The top panel corresponds to the common signal  $\mathbf{f}$ , the second panel to individual growth tree effect signals  $\mathbf{g}_j$  and the bottom panel to simulated tree ring series  $\mathbf{y}_j$ , respectively. Our objective is to find  $\mathbf{f}$  and  $\mathbf{g}_j$  from the  $\mathbf{y}_j$ 's.

effect  $\mathbf{g}_j$  can become more and more complex. A fundamental case is when all individual growth  $\mathbf{g}_j$  are proportional to each other, say  $\mathbf{g}_j = \alpha_j \mathbf{g}$  with  $\alpha_j \geq 0$  and  $\mathbf{g} \neq \mathbf{0}$ , then imposing  $\mathbf{g}_1 + \dots + \mathbf{g}_p = \mathbf{0}$  implies that  $\alpha_j = 0$ . In other words, if individual growths have something in common, then it should capture by  $\mathbf{f}$  (and not by  $\mathbf{g}_j$ ). This makes sense in our model because  $\mathbf{f}$  will be interpreted as the common signal and the  $\mathbf{g}_j$  will be viewed as individual characteristics (anomalies from the common behavior).

To compute the posteriors of the latent vectors and model parameters, we use Gibbs sampler

and Metropolis-Hasting algorithms. Explicitly posterior distribution for some functions can be derived (Hastie and Tibshirani, 1990, 2000)

$$\mathbf{f}|\mathbf{g}, \lambda_0, \mathbf{y}, \sigma^2 \sim \mathcal{N}_n(\mathbf{B}(\mathbf{B}^T \mathbf{R} \mathbf{B} + \lambda_0 \mathbf{\Omega})^{-1} \mathbf{B}^T \mathbf{s}, \sigma^2 \mathbf{B}(\mathbf{B}^T \mathbf{R} \mathbf{B} + \lambda_0 \mathbf{\Omega})^{-1} \mathbf{B})$$

with

$$\mathbf{s} = \sum_{j=1}^p (\mathbf{y}_j - \mathbf{g}_j), \lambda_0 = (1 - \phi_0)/\phi_0, \mathbf{R} = \text{Diag}(p),$$

and

$$\mathbf{g}_j | \mathbf{f}, \lambda_j \mathbf{y}_j, \sigma^2 \sim \mathcal{N}_n(\mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda_j \mathbf{\Omega})^{-1} \mathbf{B}^T \mathbf{d}, \sigma^2 \mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda_j \mathbf{\Omega})^{-1} \mathbf{B})$$

with  $\mathbf{d} = \mathbf{y}_j - \mathbf{f}$  and  $\lambda_j = (1 - \phi_j)/\phi_j$ . It is also possible to show that  $\sigma^2$  have an inverse gamma posterior distribution. These parameters are estimated with Gibbs sampler. The parameters  $\phi_0$  and  $\phi_j$  don't have standard posterior distributions so we use Metropolis-Hasting algorithm to estimate them (Hastie and Tibshirani, 2000). The Bayesian inference was carried out with the open source R statistical software.

## 4 Data analysis

### 4.1 Simulations results

As described in Section 3 simulated data are used to study model performances. This data are simulated to mimic tree ring widths series as the sum of a common signal  $\mathbf{f}$  and individual tree growth signals  $\mathbf{g}_j$ . The simulated common signal corresponds to the normalize equation :

$$f(t) = \exp \frac{2t}{(1+t)} + 2 \sin(t),$$



were  $t$  represents time and  $\mathbf{g}_j$ 's have a Beta density form with random parameters according to  $j$ . Signals are simulated for 50 time length. The simulated data  $\mathbf{y}_j$ 's, are the sum of this signals with normal uncertainty. These simulations are displayed in Figure 4.

A first analyse study the prior question about  $\phi$ . After discussions with experts in dendrochronology it seems the common signal have to be an annual proxy while individual age effect,  $\mathbf{g}_j$  are smooth trend. In Section 3, Beta priors are selected for the smooth parameters  $\phi$ . Table 1 summaries the posterior distribution of  $\phi_1$  under a variety of prior distribution. We note the prior parameters have not a significant impact over the posterior distribution and this result is aproximatively the same for the others  $\phi_j$ . So we set an non-informative prior  $\phi_j \sim \text{Beta}(1, 1)$  for  $j = 1, \dots, p$ .

Parameters of the prior distribution		Summaries of the posterior distribution	
$\alpha$	$\beta$	Posterior median of $\phi_1$	90% posterior interval for $\phi_1$
1	1	0.00083	[0.00047 ; 0.04853 ]
1	5	0.00175	[0.00076 ; 0.01737 ]
1	10	0.00061	[0.00061 ; 0.02292 ]
1	15	0.00165	[0.00034 ; 0.01060 ]
1	20	0.00567	[0.00153 ; 0.03160 ]

Table 1: Summaries of the posterior distribution of  $\phi_1$  under a variety of conjugate prior distributions.

Concerning  $\phi_0$  prior (common signal f smooth parameter) the question is some different. In a first time we select also a non-informative prior  $\phi_0 \sim \text{Beta}(1, 1)$ . But a good posterior estimation for this parameter depend of the model uncertainty variance. The Figure 5 represents the posterior estimation of  $\mathbf{f}$  compared to the real simulated comun signal for difference noise levels. If  $\sigma^2$  has a too high variance the model has difficulty to detect the exact common signal and it tends to smooth the estimated  $\mathbf{f}$  curve.

To complete this analysis, Figure 6 shows the influence of the Beta prior parameter,  $\beta$ , on the smoothness of the common signal,  $\mathbf{f}$ , while keeping a variance level of  $\sigma^2=0.8$ . As  $\beta$  increases (more prior information), the hidden signal appears to be better captured. The gain after  $\beta=10$  seems marginal and the common signal does not change anymore.

~~To complet this analysis we choose a some important variance level ( $\sigma^2 = 0.8$ ) and we make vary the  $\beta$  parameter for Beta prior over the smooth common signal parameter. Figure~~

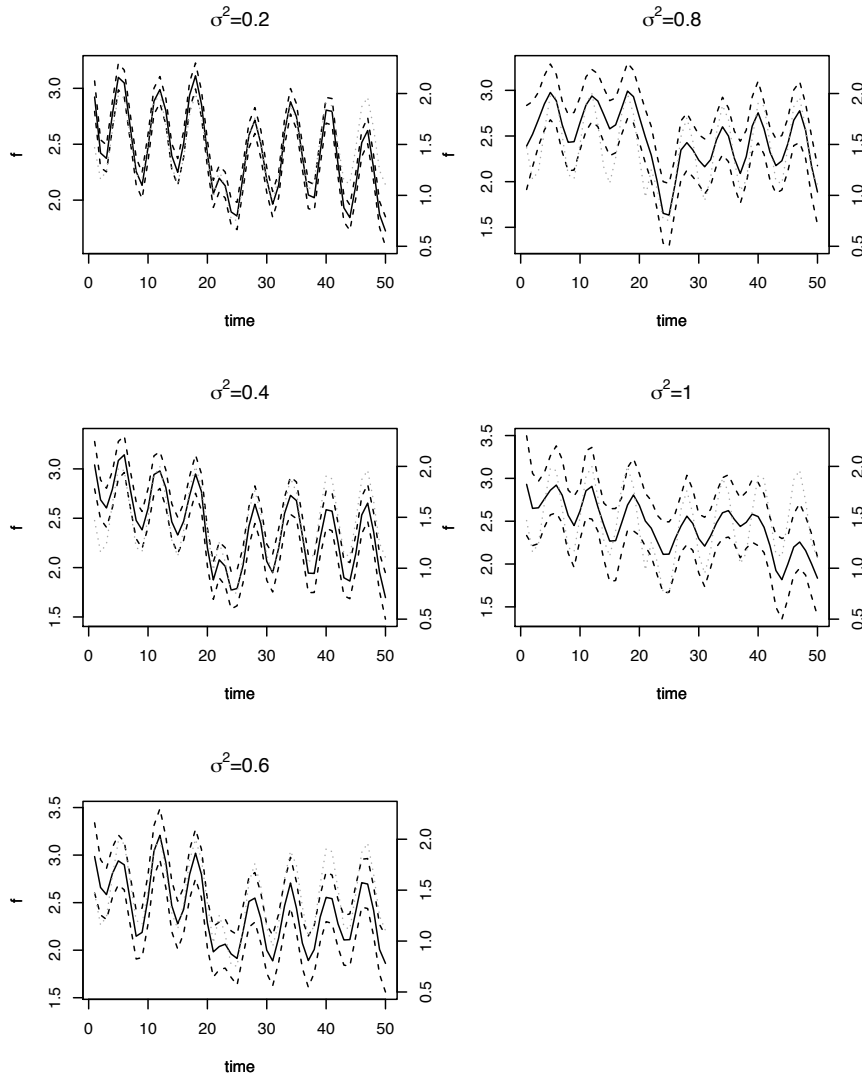


Figure 5: Posterior information about the common signal  $f$  obtained from the simulated tree series shown in the bottom panel of Figure 4 with different noise levels. The gray and black line correspond to the true  $f$  and the estimated posterior median, respectively. The dotted lines represents the 90% credibility intervals.

~~6 represents the common signal  $f$  for these different priors. More the value of  $\beta$  increases (corresponding to more informative prior) better is the estimation of  $f$ . In view of this analysis it seems reasonable to fix a Beta(1, 10) prior for  $\phi_0$ .~~

Different sensitivity analysis concerning the influence of the number of tree were also performed and are available upon request. In a nutshell, around 10 trees in our simulations

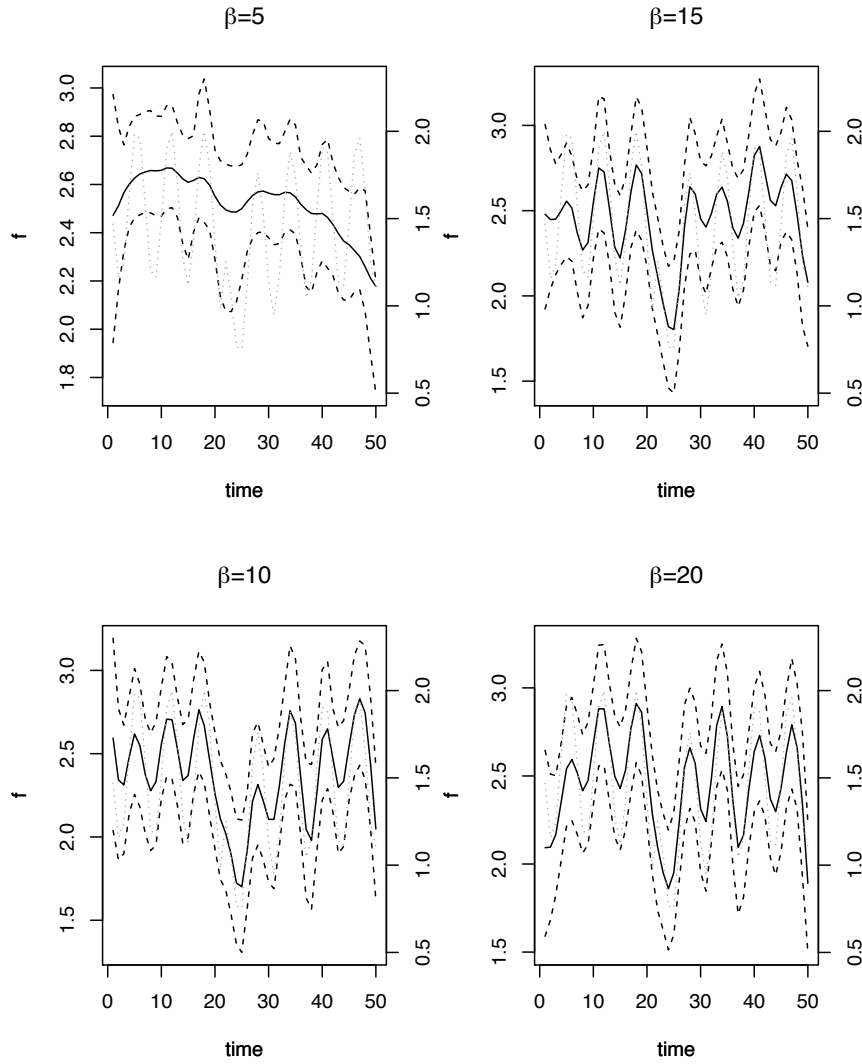


Figure 6: Posterior information about the common signal  $f$  obtained from the simulated tree series with different smooth parameter prior. The gray and black line correspond to the true  $f$  and the estimated posterior median, respectively. The dotted lines represents the 90% credibility intervals.

were necessary to derive reasonable results. However this remark about a minimal number of trees is only valid within the framework of our simulations and it should not be directly transposed to real data because the shapes of  $f$  and  $g_j$  and the variance  $\sigma^2$  strongly depend on the tree species and the site characteristic.

## 4.2 Analysis of seventeen tree ring width logarithms of *Pinus halepensis* Mill

Our model and inference scheme have been applied to the seventeen tree ring width logarithms shown in Figure 3. Our priors were the same that previously discussed. The posterior median (black line) and their associated 90% credibility intervals (dotted lines) of the three individual age trends  $g_1$ ,  $g_2$  and  $g_3$  are displayed in the left panels of Figure 7. The posteriors of the smoothness parameter are centered around 0 - 0.1 (see the right panels). This implies that the inferred curves are fairly smooth. Still this graph reveals a variety of shape with different wiggles, see also Figure 8.

To put our approach into perspective with respect to the RCS method, Figure 8 compares posterior median of individual age effect profiles  $g_j$  that have been aligned according to their biological age (not their chronological age) with the classical global biological RCS trend obtained by averaging ring widths in function of their biological age (gray line). Although a majority of curves has overall the same degree of smoothness, this figure emphasizes the variability among age effect profiles. This tends to indicate that the added flexibility of our modeling approach allows to improve individual age-related growth variability. A strong message from Figure 8 resides in the large variability among the different age effect shapes. Each tree has its own trend and associated uncertainty. And having this information could help dendrochronologists to interpret local tree behaviors.

Figure 9 shows the posterior common signal obtained with our bayesian model. As previously, the black line represents the posterior median and the dotted lines their associated 90% credibility intervals. Our inferred common signal can be compared to the classical RCS method (in gray in Figure 9). The fluctuations to the both signals have the same smoothing of their curve seem has almost the same variations. The major difference is that with the bayesian method the common signal capture low frequencies. The second point is that with this method we have

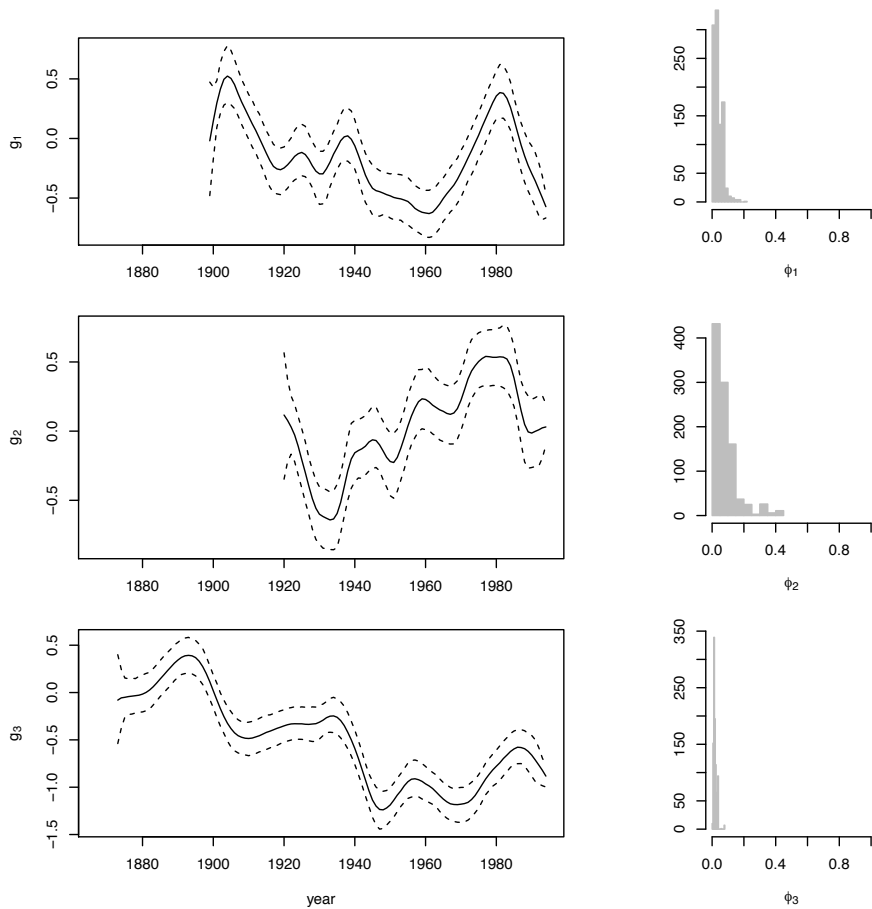


Figure 7: Left panels: posteriors of the three individual age effect trends  $g_1$ ,  $g_2$  and  $g_3$  obtained from our analysis of the seventeen tree ring width logarithms shown in Figure 3. Black lines correspond to posterior medians and dotted lines to 90% credibility intervals. Right panels: posterior pdfs of the smoothness parameters  $\phi_1$ ,  $\phi_2$  and  $\phi_3$ . Note that the y-axes are different in right hand panels.

be able to quantify the uncertainty related to common signal estimation.

## 5 Conclusion

In this paper we introduce a new method to estimate a tree-rings hidden common signal. The proposed model is a semi-parametric Bayesian hierarchical model that offers the possibility to capture low and high frequencies in tree ring widths, and to obtain quantify

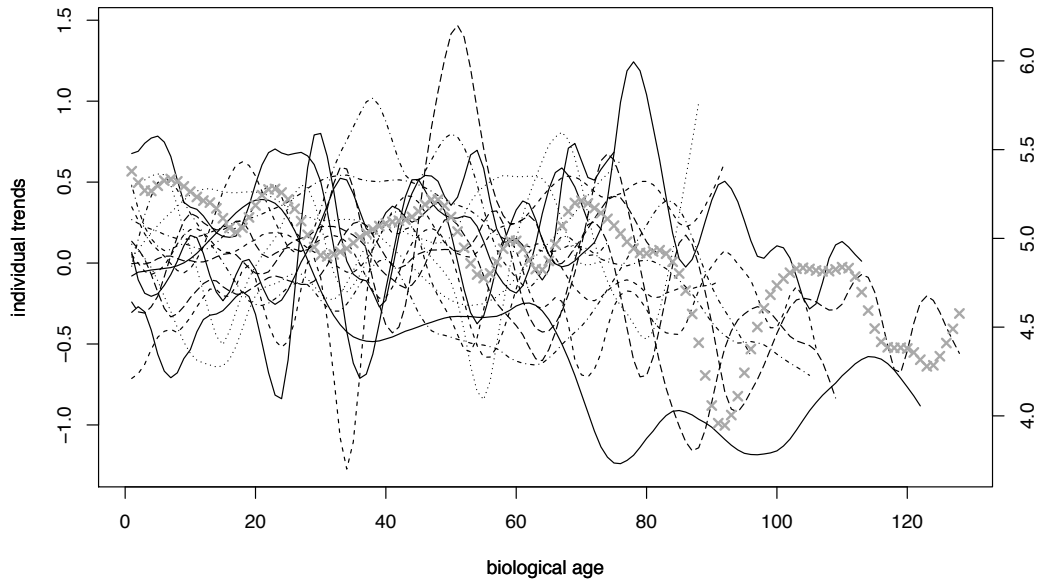


Figure 8: Posterior median of individual age effect profiles  $g_j$  that have been aligned according to their biological age (not their chronological age). The gray line represents the classical global biological trend obtained by averaging ring widths in function of their biological age.

uncertainty over the different estimated signals.

Our model is tested on simulated data and applied to *Pinus halepensis Mill.* ring widths recorded in French Mediterranean. ~~The results are compared with~~ <sup>Overall, our findings do not contradict the classical RCS approach, but it goes one step further</sup> a dendrochronological classical method (Regional Curve Standardization) ~~and the results are not in contradiction.~~ The bayesian method permits to bring further information as credibility intervals or low frequency in common signal contains relevant information about past climate. The more complexe modeling permits also to have different age effect shapes for each tree which can be an interesting information to help dendrochronologists to interpret local tree behaviors.

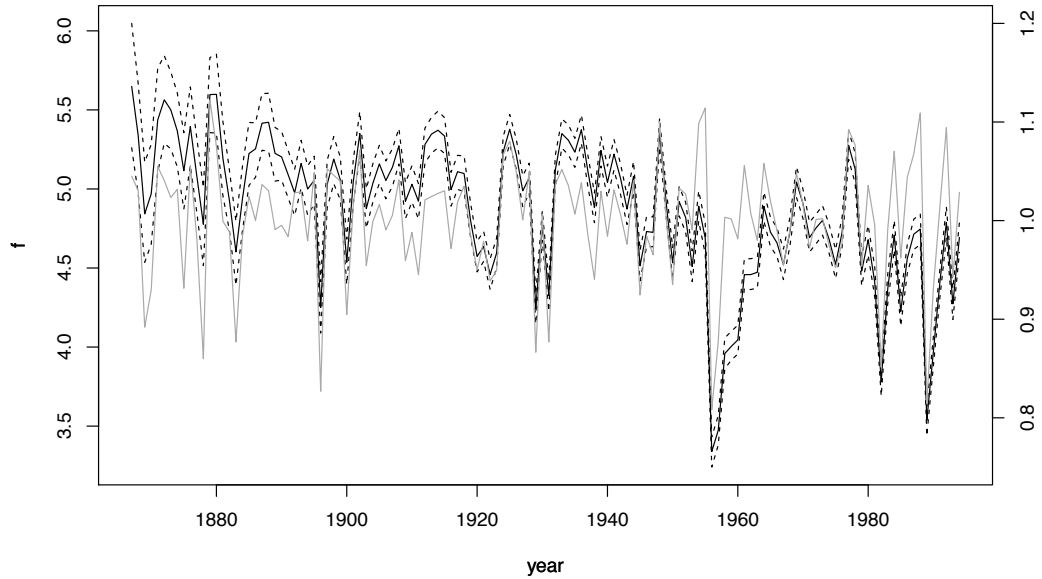


Figure 9: Posterior information about the common signal  $f$  and comparison with RCS chronology. The black line correspond the estimated posterior median and the dotted lines represents the 90% credibility intervals. The gray line is the RCS chronology.

To estimate the contribution of this method for dendrochronologists it will be interested to investigate potential links between extracted signals and climatic variables and try to reconstruct past temperatures for exemple. ~~Different methods were developed for this but they are not adapted to integrate the additional information given with the paper developed method. So a reflexion is required to include, in particular common signal estimation uncertainty, in climatic reconstructions.~~

## References

- Barefoot, A., Woodhouse, L., Hafley, W., and Wilson, E. (1974). Developing a dendrochronology for winchester england. *Journal of the Institute of Wood Science*, 6:34–40.
- Berliner, L., Wikle, C., and Cressie, N. (2000). Long-lead prediction of pacific ssts via bayesian dynamic modeling. *J. Climate*, 13:3953–3968.
- Berliner, L. M. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):45–46.
- Boreux, J.-J., Naveau, P., Guin, O., Perreault, L., and Bernier, J. (2009). Extracting a common high frequency signal from northern quebec black spruce tree-rings with a bayesian hierarchical model. *Climate of the Past*, 5(4):607–613.
- Buckley, B. (2009). *Encyclopedia of Paleoclimatology and Ancient Environments. Encyclopedia of Earth Sciences Series.*, chapter Dating, dendrochronology. Dordrecht, Netherlands: Springer.
- Christiansen, B., Schmith, T., and Thejll, P. (2009). A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, 22:951–976.
- Cook, E. (1990). *Methods of Dendrochronology: Applications in the Environmental Sciences.*, chapter A conceptual linear aggregate model for tree rings, pages 98–104. Kluwer Academic Publ., Dordrecht.
- Cook, E. and Kairiukstis, L. (1990). *Methods of dendrochronology : applications in the environmental sciences / edited by E.R. Cook and L.A. Kairiukstis.* Kluwer Academic Publishers, Dordrecht, Netherlands, Boston.



- Cook, E. and Peters, K. (1981). The smoothing spline : a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. *Tree-ring Bulletin*, 41:45–53.
- Cook, E. R., Briffa, K. R., Shiyatov, S., and Mazepa, V. (1990). Treering standardization and growth-trend estimation, in methods of dendrochronology: Applications in the environmental sciences, edited by e. r. cook and l. a. kairiukstis. *Kluwer Acad*, Norwell:Mass.
- Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K. (1999). Drought reconstructions for the continental united states. *J. Clim*, 12:1145–1162.
- Craigmile, P. and Rajaratnam, B. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):88–90.
- Cressie, N. and Tingley, M. P. (2010). Comment: Hierarchical statistical modeling for paleoclimate reconstruction. *Journal of the American Statistical Association*, 105(491):895–900.
- Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. Wiley.
- Davis, R. A. and Liu, J. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):52–55.
- de Boor, C. (1978). *A practical Guide to Splines*. Applied Mathematical Sciences.
- Douglass, A. (1920). Evidence of climatic effects in the annual rings of trees. *Ecology*, 1:24–32.
- Douglass, A. (1936). Climatic cycles and tree-growth. *Carnegie Institution of Washington publication*, 289(3).

- Esper, J., Schweingruber, F. H., and Winiger, M. (2002). 1300 years of climatic history for western central asia inferred from tree-rings. *The Holocene*, 12:267–277.
- Eubank, R. (1999). *Nonparametric regression and spline smoothing*. Marcel Dekker.
- Evans, M., Reichert, B., Kaplan, A., Anchukaitis, K., Vaganov, E., Hughes, M., and Cane, M. (2006). A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J. Geophys. Res*, 111(G03008).
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall.
- Fang, K., Gou, X., Peters, K., Li, J., and Zhang, F. (2010). Removing biological trends from tree-ring series: testing modified hugershoff curves. *Tree-ring research*, 66(1):51–59.
- Fritts, H., Mosimann, J., and Bortorff, C. (1969). A revised computer program for standardizing tree-ring series. *Tree-Ring Bulletin*, 29:15–20.
- Gavin A. Schmidt, M. E. M. and Rutherford, S. D. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):65–70.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis*. Chapman and Hall, 2nd edition.
- Gornitz, V., editor (2009). *Encyclopedia of Paleoclimatology and Ancient Environments. Encyclopedia of Earth Sciences Series*. Dordrecht, Netherlands: Springer.
- Guiot, J. (1987). *Methods of dendrochronology - 1*, chapter Standardization and selection of the chronologies by the ARMA analysis. International Institute for Applied Systems Analysis, Laxenburg, Austria and Polish Academy of Sciences-System Research Institute, Warsaw, Poland.

- Haran, M. and Urban, N. M. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):61–64.
- Haslett, J., Salter-Townshend, M., Wilson, S. P., Bhattacharya, S., Whitley, M., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G. (2006). Bayesian palaeoclimate reconstruction. *J. R. Statist. Soc. A*, 169, Part 3, pp.(3):1–36.
- Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall, London.
- Hastie, T. and Tibshirani, R. (2000). Bayesian backfitting. *Statistical Science*, 15(3):196–223.
- Holmström, L. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):71–75.
- Hooten, M. and Wikle, C. (2007). Shifts in the spatio-temporal growth dynamics of short-leaf pine. *Environmental and Ecological Statistics*, 14(3):207–227.
- Kaplan, A. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):47–51.
- Kimeldorf, G. and Wahba, G. (1970). A correspondence between bayesian estimation of stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41:495–502.
- Li, B., Nychka, D. W., and Ammann, C. M. (2010a). Rejoinder: Hierarchical statistical modeling for paleoclimate reconstruction. *Journal of the American Statistical Association*, 105(491):910–911.

- Li, B., Nychka, D. W., and Ammann, C. M. (2010b). The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association*, 105(491):883–895.
- McIntyre, S. and McKittrick, R. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):56–60.
- McShane, B. B. and Wyner, A. J. (2011a). Rejoinder of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):99–123.
- McShane, B. B. and Wyner, A. J. (2011b). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):5–44.
- Melvin, T. and Briffa, K. (2008). A “signal-free” approach to dendroclimatic standardisation. *Dendrochronologia*, 26:71–86.
- Nicault, A., Guiot, J., Edouard, J., and Brewer, S. (2010). Preserving long-term fluctuations in standardisation of tree-ring series by the adaptative regional growth curve (argc). *Dendrochronologia*, 28(1):1 – 12.
- Nicault, A., Rathgeber, C., Tessier, L., and Thomas, A. (2001). Observations sur la mise en place du cerne chez le pin d’alep (*pinus halepensis* mill.) : confrontation entre les mesures de croissance radiale, de densité et les facteurs climatiques. *Annals of forest science*, 58:769–784.
- Nychka, D. and Li, B. (2011). Spurious predictions with random time series: The lasso in the context of paleoclimatic reconstructions. discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):80–82.

- Reinsch, C. (1967). Smoothing by spline functions. *Numer. Math.*, 10:177–138.
- Rougier, J. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):96–98.
- Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2007). High-resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association*, pages 1–14.
- Schliep, E., Cooley, D., Sain, S., and Hoeting, J. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13:219–239.
- Smerdon, J. E. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):76–79.
- Smith, R. L. (2010a). Comment: Hierarchical statistical modeling for paleoclimate reconstruction. *Journal of the American Statistical Association*, 105(491):905–910.
- Smith, R. L. (2010b). Understanding sensitivities in paleoclimatic reconstructions. Technical report, Technical report.
- Tebaldi, C., Smith, R. L., and Sanso, B. (2010). *BAYESIAN STATISTICS 9*, chapter Characterizing Uncertainty of Future Climate Change Projections Using Hierarchical Bayesian Models. Oxford University Press,.
- Tingley, M., Craigmille, P., Haran, M., LI, B., Mannshardt, E., and Rajaratnam, B. (2012). Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quaternary Science Reviews*, 35:1–22.

- Tingley, M. and Huybers, P. (2013). Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature*, 496:201–205.
- Tingley, M. P. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):83–87.
- Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society*, 40(3):364–372.
- Wahl, E., Schoelzel, C., Williams, J., and Tigrek, S. (2010). Comment: Hierarchical statistical modeling for paleoclimate reconstruction. *Journal of the American Statistical Association*, 105(491):900–905.
- Wahl, E. R. and Ammann, C. M. (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, 5(1):91–95.
- Wand, M. and Jones, M. (1995). *Kernel smoothing*. Chapman and Hall.
- Werner, J. and Tingley, M. (2015). Technical note: Probabilistically constraining proxy age–depth models within a bayesian hierarchical reconstruction model. *Climate of the Past*, 11:533–545.