Dendrochronologia 46 (2017) 77–84

Contents lists available at ScienceDirect

Dendrochronologia

journal homepage: www.elsevier.com/locate/dendro

Original article

Model fitting and evaluation in climate reconstruction of tree-ring data A comment on Steinschneider et al. (2017): Hierarchical regression models for dendroclimatic standardization and climate reconstruction

Matthew R. Schofield*, Richard J. Barker

Department of Mathematics and Statistics, University of Otago, PO Box 56, Dunedin, New Zealand

ARTICLE INFO

Keywords:
Approximation
Bayes
Empirical Bayes
Prediction assessment
Shrinkage

ABSTRACT

Steinschneider et al. (2017) investigate model choices made in the hierarchical climate reconstruction approach of Schofield et al. (2016). We identify two flaws in their approach. The first is the use of an unusual approximation to Bayesian inference that unnecessarily discards important information. The second is that they mischaracterize the robustness of their reconstructions due to overlooking important features of the out-of-sample predictions. We demonstrate how full Bayesian inference can be conducted with no additional effort, providing R/JAGS code. We also show how graphical visualization of the out-of-sample predictions can lead to better understanding and comparison of the models fitted.

1. Introduction

Steinschneider et al. (2017) investigate the use of hierarchical Bayesian models for reconstructing temperature. Using the framework of Schofield et al. (2016) they explore the effect of various modeling assumptions on the reconstruction of held-out temperature observations. These assumptions include partial pooling of parameters, a common variance among trees, and autocorrelation in the observations. They also consider the effect of alternative ageing functions. Among other things, they conclude that including autocorrelation "significantly degrades performance" (p. 182).

We identify two flaws in their approach of Steinschneider et al. (2017). The first is the manner in which the statistical models are fitted. The authors have used an unusual approximation to Bayesian inference that unnecessarily discards important information. The second is the way in which out-of-sample predictions are assessed and interpreted. We argue that important features are missed that have implications for the conclusions drawn regarding the robustness of the various models considered.

It is important to distinguish between the statistical model specified and the method of model fitting used to estimate parameters and predict unobserved random variables. We show how some of the models considered by Steinschneider et al. (2017) are probabilistically identical but differ according to the method of model fitting. Separating these issues allows us to better understand the role of model fitting distinct from model specification. We use this insight to offer general model fitting strategies.

In Section 2 we review model M2 of Steinschneider et al. (2017) and the data we use here. We use model M2 as a base model to evaluate their approach. In Section 3, we first introduce and overview aspects of Bayesian inference. This allows us to critically evaluate the modeling approach of Steinschneider et al. (2017). We conclude in Section 4 with a discussion of the out-of-sample prediction procedures used and the implications for the robustness of the models considered.

2. Data and model

The ring width measurements, denoted y_{it}, are observed for M trees across T years. The variables f_i and l_i denote the first and last years in which tree i was observed. Climate data x_t are available for the final $T - t_0 + 1$ years for which we have observed ring width data. The variable age_{it} denotes the age of tree i in year t. Throughout this manuscript we refer to model M2 of Steinschneider et al. (2017) that is defined as,

$$\begin{aligned}
y_{it} &= \beta_{i0} + \beta_{i1} age_{it} + \eta_t + \epsilon_{it}, && i = 1,\ldots,M,\ t = f_i,\ldots,l_i,\\
\epsilon_{it} &\overset{ind}{\sim} N(0, \sigma_i^2), && i = 1,\ldots,M,\ t = f_i,\ldots,l_i,\\
\eta_t &\overset{ind}{\sim} N(\beta_2 x_t, \sigma_\eta^2), && t = 1,\ldots,T,\\
x_t &\overset{iid}{\sim} N(\mu_x, \sigma_x^2), && t = 1,\ldots,T,
\end{aligned} \tag{1}$$

where ind denotes a conditionally independent random variable and

* Corresponding author.
E-mail address: mschofield@maths.otago.ac.nz (M.R. Schofield).

https://doi.org/10.1016/j.dendro.2017.10.004
Received 7 July 2017; Received in revised form 2 October 2017; Accepted 18 October 2017

the observations x_t are missing for $t = 1, \ldots, t_0 - 1$. The terms β_{i0} and β_{i1} are tree-specific parameters that describe the age-related effect on tree growth. The random variable η_t is the yearly influence on tree growth that is common between trees. The climate variable x_t influences η_t and is weighted by β_2. The parameters μ_x and σ_x^2 describe the marginal model for the climate variable through time.

We use the Torneträsk ring-width data used in Schofield et al. (2016), with 247 series spanning the years 1496–1995. These data slightly differ from that of Steinschneider et al. (2017), who despite claiming to use the same data as in Schofield et al. (2016) consider a series from 1497 to 1997.

We use the same Tornedalen temperature series used by Steinschneider et al. (2017). The only difference is that we do not consider the data from 1996 and 1997. We separate the temperature data into two groups for out-of-sample assessment, with observations from 1906 to 1995 used for calibration and we hold-out data from 1816 to 1905 for validation.

We follow Steinschneider et al. (2017) and apply a Box–Cox power transformation. However, it is not clear how this transformation was applied by Steinschneider et al. (2017) as there are several possibilities. One is to apply the transformation to the raw data y_{it}. Another is to apply the transformation taking into account a model with temperature and tree-specific age effects as in (1). This latter approach is preferable as it is error terms that we assume follow a common normal distribution. This is challenging, as the temperature variable is partially observed and the full model would have to be fit many times to evaluate the best transformation. Instead, we use only the data from the calibration period (we used data from 1816 to 1995) to find the best data transformation ($\lambda = 0.22$) for a linear model with temperate and tree-specific age effects. On the power transformation scale, this is between the log and square root transform. The data are available online as outlined in Appendix C.

3. Bayesian modeling

3.1. Background

Bayesian modeling distinguishes between known (observed) and unknown (unobserved) variables. Given known variables k, we wish to find $p(u|k)$, the probability distribution of the unknown variables, denoted u. For inference, the posterior distribution $p(u|k)$ fully describes our knowledge about u given the information at hand. Typically this is described in terms of summaries of $p(u|k)$, such as the mean and quantiles.

In statistical applications, the knowns are usually data y and the unknowns are parameters θ, leading to

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\,d\theta}.$$

For any particular set of data y, the denominator $p(y)$ is a constant and so we commonly write

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{2}$$

To find the posterior distribution in (2) two things are required; (i) the probability model for y, denoted $p(y|\theta)$ and commonly referred to as the likelihood, and (ii) a prior distribution for θ. The normalizing constant

$$p(y) = \int p(y|\theta)p(\theta)\,d\theta$$

is difficult to find in closed-form for all but the simplest models. As a consequence, it is common to avoid calculation of $p(y)$ and instead draw correlated samples from $p(\theta|y)$ using Markov chain Monte Carlo (MCMC), see Gelman et al. (2014) for details. We use these samples to estimate features of the posterior density, such as the mean or quantiles. There are numerous MCMC algorithms in common use, including Gibbs sampling (Geman and Geman, 1984), the Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and slice sampling (Neal, 2003).

In practice, modeling is often more complicated than presented in (2). It is common that the unknowns include missing data and latent variables in addition to the model parameters. It is straightforward to include such extensions in (2). For example, if we model y in terms of latent variables ξ that themselves depend on parameters θ, then (2) becomes

$$p(\theta, \xi|y) \propto p(y|\xi)p(\xi|\theta)p(\theta). \tag{3}$$

While additional challenges can arise, it is straightforward to sample from $p(\theta, \xi|y)$ using MCMC.

3.2. Models vs methods of estimation

Bayesian inference refers to the approach where we use the posterior density in (2) or (3) to make inference (e.g., see Gelman et al., 2014). Bayesian inference is an extensively studied and well understood approach. It is the only approach to statistical inference that is fully consistent with axiomatic probability theory. Various approaches that approximate $p(\theta|y)$ exist and are typically less well understood. One such approximation is empirical Bayes (Carlin and Louis, 2009).

Empirical Bayes is considered in models such as (3), where ξ is of interest and θ are nuisance parameters. Bayesian inference about ξ in (3) averages across θ

$$\begin{aligned} p(\xi|y) &= \int p(\theta, \xi|y)\,d\theta \\ &= \int p(\xi|\theta, y)p(\theta|y)\,d\theta. \end{aligned}$$

Empirical Bayesian inference does not average over θ but uses the so-called estimated posterior distribution of ξ for inference,

$$p(\xi|\hat{\theta}, y),$$

where $\hat{\theta}$ is some estimate of θ, typically obtained from the marginal distribution $p(y|\theta)$. If the models $p(y|\xi)$ and $p(\xi|\theta)$ are chosen carefully then it can be relatively straightforward to find closed-form expressions for both the marginal distribution $p(y|\theta)$ and estimated posterior $p(\xi|\hat{\theta}, y)$. In this situation, it is fast and easy to fit the model using empirical Bayes, even for large datasets. In contrast, Bayesian inference requires us to sample from $p(\xi|y)$ using a time-consuming method such as MCMC.

Empirical Bayes is usually only considered when approaches for sampling from the posterior distribution are prohibitively difficult or time consuming. Care is needed using empirical Bayes as the variance of $p(\xi|\hat{\theta}, y)$ is smaller than the variance of the true posterior $p(\xi|y)$ (see e.g. Carlin and Louis, 2009, p. 244). Using $p(\xi|\hat{\theta}, y)$ in place of $p(\xi|y)$ will lead to interval estimates that are narrower than they should be. The error in approximation can vary according to the model assumed and the data set used, which places the burden on the researcher to ensure that the approximation is reasonable for their data and model.

3.2.1. Approximate Bayesian methods: empirical Bayes vs Bayes

Steinschneider et al. (2017) used MCMC to generate samples from the posterior density of the parameters in model M2, the model described in (1). They then ignore the MCMC samples of x_t, $t = 1, \ldots, t_0 - 1$ and instead sample from an approximate posterior (p. 179),

$$x_t \sim N\left(\frac{\hat{\sigma}_\eta^2\hat{\mu}_x + \hat{\sigma}_x^2\hat{\beta}_2\hat{\eta}_t}{\hat{\sigma}_\eta^2 + \hat{\beta}_2^2\hat{\sigma}_x^2}, \left[\frac{1}{\hat{\sigma}_x^2} + \frac{\hat{\beta}_2^2}{\hat{\sigma}_\eta^2}\right]^{-1}\right). \tag{4}$$

This is unusual because the x_t values they ignore can be treated as samples from the posterior distribution $p(x_t|\text{data})$ and provide the best information about the unknown climate values x_t that we have, given our data. To then ignore them amounts to discarding information and making inference that is weaker at best and potentially misleading. The