Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
CORSO DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXIX

TREE-RING BASED PALAEOCLIMATE RECONSTRUCTION USING A HIERARCHICAL BAYESIAN MODEL

Coordinatore: Ch.mo Prof. Nicola Sartori

Supervisore: Ch.mo Prof. Claudio Agostinelli

Co-supervisore: Dr. Luigi Spezia

Dottorando: Elisa Carraro

Padova, 31 luglio 2017

To Fr. Marco.

Table of Contents

Abstract

Reconstructing climatic variables in the pre-instrumental period is important to get a correct understanding of the climatic change issues. The main aim of this research is to present a model to reconstruct summer mean temperature and moisture on the basis of a dendrochronology, using a Bayesian approach. As a second step, a procedure for the estimation of monthly observed climatic variables on the basis of annual dendrochronology is illustrated, along with an application to an Italian location.

Sommario

Ricostruire le variabili climatiche nel periodo pre-strumentale è importante per una corretta comprensione delle problematiche legate al cambiamento climatico. Lo scopo principale di questa ricerca è presentare un possibile modello per la ricostruzione della media estiva di temperatura e umidità a partire da una dendrocronologia, utilizzando un approccio di tipo bayesiano. Viene poi illustrata una procedura per stimare variabili climatiche osservate mensilmente sulla base di dendrocronologie annuali, con un'applicazione a un sito italiano.

x

Acknowledgments

List of Symbols and Notation

I_n Identity matrix of size n.

$\mathbb{I}_A(\cdot)$ Indicator function of the set A.

$N(a, b)$ Univariate Gaussian distribution with mean a and variance b.

$N_k(a, B)$ $k-$variate Gaussian distribution with mean vector a and variance-covariance matrix B.

$IW_k(\nu, S^{-1})$ Inverted-Wishart distribution with ν degrees of freedom, positive-definite $k \times k$ scale matrix S, implicit dimension k, mean equal to $(\nu - k - 1)^{-1}S$.

$IG(a, b)$ Inverted-Gamma distribution with parameters a and b, mean equal to $b/(a - 1)$ for $a > 1$ and variance equal to $b^2 / \left[(a - 1)^2 (a - 2) \right]$ for $a > 2$.

$x(t)$ Variable x at time t.

$E(x)$ Mean vector of the random vector x.

$\text{Cov}(x)$ Variance-covariance matrix of the random vector x.

$F_{norm}^{-1}(\cdot)$ Quantile function of the standard Gaussian distribution.

$x \perp y$ Variables x and y are independent.

$\mathbf{0}_k$ Null vector of length k.

Chapter 1

Introduction

1.1 Overview

The topic of climate change is tightly bound to the problem of climate reconstruction in the past centuries. A better understanding of climate dynamics in fact requires studies of the period before instrumental measurements are available. Statistics can give an important support to palaeoclimatological studies, and this work, hopefully, would like to give a little contribution in this direction. This introductory Chapter describes the problem, presenting the main contributions in the literature and giving an overview of the thesis' structure.

Let us first report the definition of climate given in the Glossary of the synthesis report published by the Intergovernmental Panel on Climate Change (IPCC, 2014):

> Climate in a narrow sense is usually defined as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period for averaging these variables is 30 years, as defined by the World Meteorological Organization. The relevant quantities are most often surface variables such as temperature, precipitation and wind. Climate in a wider sense is the state, including a statistical description, of the climate system, which is made up of the atmosphere, the hydrosphere, the cryosphere, the lithosphere and the biosphere and the interactions between them.

Measurements of climatic variables are available only since the mid XIX century, so in order to evaluate climate dynamics in the proper perspective, past climatic behaviour needs to be taken into account (Bradley, 2013).

Palaeoclimatology is the study of climate before the period of instrumental measurements. Since no measurements are available there is the need to study other kinds of natural phenomena that are climate-dependent and therefore provide a so-called proxy record of climate. Different kind of proxy data can be used in palaeoclimatic reconstruction, for example ice cores, lake or marine sediments, pollen, tree-rings and others. These types differ in several respects like their spatial and temporal coverage, their resolution, that is the degree of detail each record can provide (seasonal, annual, decadal, . . .) and that determines the value it has for reconstruction, and the possible temporal lag between the climate variation and its appearance in the proxy record.

To extract the paleoclimatic signal from proxy data, the record must first be interpreted or calibrated. Calibration means using modern climatic measurements and proxy materials to establish
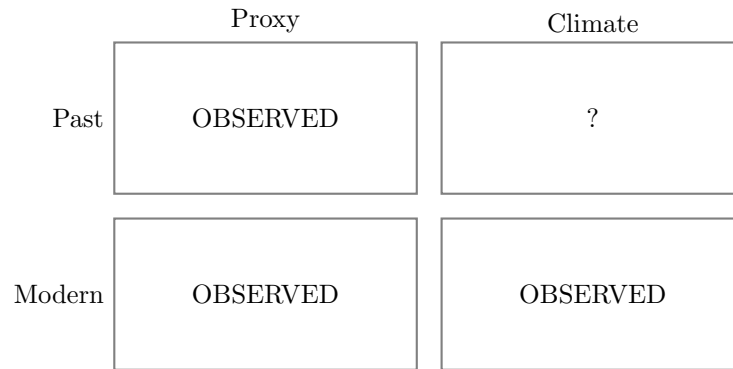
	Proxy	Climate
Past	OBSERVED	?
Modern	OBSERVED	OBSERVED

Figure 1.1: The typical datset.

the climate-proxy relationship.

The uniformitarianism principle is assumed, which means to suppose that the modern climate-proxy relationship has operated in the whole period of interest. The typical situation is illustrated in Fig. 1.1: the calibration period corresponds to the bottom line, where both proxy and climate data are available.

Palaeoclimate reconstructions can be performed globally or locally, and can be based on single or several proxy variables.

Some contributions regard multi-proxy global reconstructions, like the paper by Mann et al. (1998), who reconstruct Northern hemisphere mean annual temperatures back to 1400, calibrating a multi-proxy data network by the dominant patterns of temperature variability in the instrumental record, obtained by principal components analysis. Their approach has drawn criticism, mainly for the method employed, as summarized in McShane and Wyner (2011).

Barboza et al. (2014) propose an approach based on hierarchical Bayesian models, obtaining directly posterior draws of Northern Hemisphere annually averaged temperature anomalies back to 1000 AD, starting from 25 proxy series of different type, that are combined in a weighted averaged (called reduced proxy) in its turn modeled as a function of the latent temperature process. Temperature is modeled as a function of the external forcings solar irradiance, volcanism and greenhouse gases, and prior distributions are assigned to the model parameters.

Concerning local reconstruction we can cite Tolwinski-Ward et al. (2015), who propose a hierarchical Bayesian approach to the simultaneous reconstruction of temperature and soil moisture from width and isotopic dendrochronologies on a Californian site, introducing a time-varying climatic response in order to overcome the uniformitarianism principle.

Since it is climate that has influence on proxies, Bayesian inference is the natural framework to deal with these kind of problems, because it allows to start from the distribution of proxy conditionally on climate, and then inverting this 'forward model' to make inference on climate conditionally on proxy data. Moreover, the output of a Bayesian analysis is a sample of values from the target process, which can then be elaborated to answer several questions of interest.

A paper that gives an excursus on existing reconstruction methods is Lee et al. (2008), which proposes also a reconstruction method based on Kalman filter algorithm, that incorporates an estimation of the climate response to external forcings. Li et al. (2010) illustrates a Bayesian hierarchical model to reconstruct Northern Hemisphere temperatures on a multi-proxy basis. A paper that discusses the scientific and statistical challenges posed by reconstruction problems is Tingley et al. (2012).

This work founds reconstruction on tree-ring widths proxy, whose main peculiarities are illustrated in Chapter 2; an application to a site in Italian central Alps, studied also in Coppola et al. (2013), is shown in Chapter 4.

The remainder of the thesis is as follows. Chapter 3 presents first a model for reconstruction, where both climatic and proxy data have annual resolution, while Paragraph 3.2 illustrates how calibration is performed when tree-ring widths data have annual resolution and climate is observed monthly. Results obtained from simulations are presented for both models. Chapter 5 presents some ideas for future developments. Appendix A explains the statistical background necessary to understand the work done in the thesis: the main concepts about Bayesian inference, the Gibbs sampling and the Metropolis-Hastings algorithm, the fundamentals of the theory about state-space models and the Kalman filter algorithm. The last paragraph of Appendix A is devoted to Gibbs sampling for state-space models, that is the algorithm used in the reconstruction procedure.

1.2 Main Contributions of the Thesis

The main contributions of the thesis can be summarized as follows:

1. development and application of a Bayesian hierarchical model for palaeoclimate reconstruction on an Italian location;

2. development of an R package based on Fortran code for the estimation of the parameters and latent variables in the annual model.

Chapter 2

Tree-rings and climate

This Chapter explains how tree-ring series used in palaeoclimatological studies are obtained from raw data.

Tree-rings are the only kind of proxy which can be accurately dated to an individual year and can provide records spanning more than a thousand years. Observing a cross-section of a tree (Fig. 2.1) we can notice an alternation of lighter and darker bands, that are made up of large, thin walled cells (earlywood) and thick-walled cells (latewood). Earlywood and latewood comprise an annual-growth increment, more commonly called a tree-ring (Bradley, 2013, Chap. 13). From the same picture we can see that several variables can be measured on a tree-ring, like the maximum and minimum density and the earlywood or latewood width or density. In this work we use only measurements of the whole ring width.
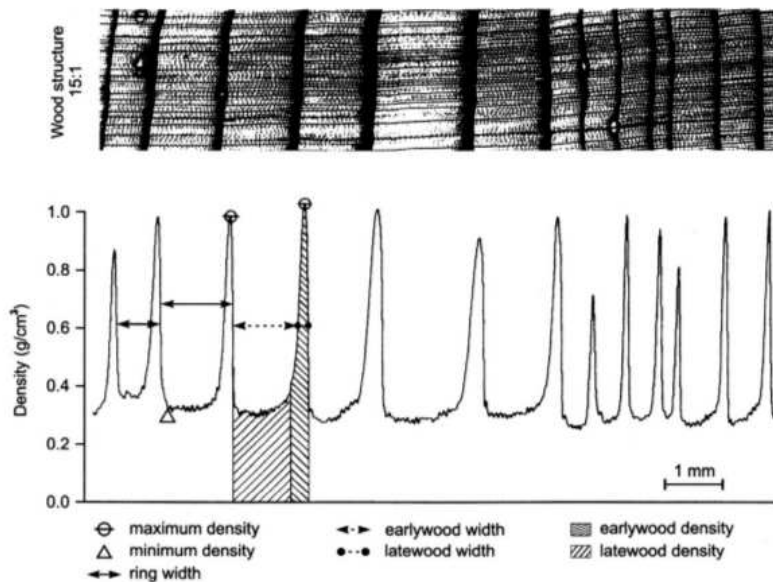
Figure 2.1: Tree-rings' structure. *From* Bradley (2013).

In dendroclimatological studies, trees are sampled in sites where their growth is limited by stressing conditions related to the climatic factors under study. In this way the variability of the ring-width can reflect how limiting the climate was to growth. Such trees are defined *sensitive* and this is referred to as the *principle of limiting factors*. When climatic factor do not vary sufficiently to limit the growth process, there may be little difference among the rings, and such

trees are defined *complacent* (Fig. 2.2). The most common types of climatic stress are moisture and temperature stress.
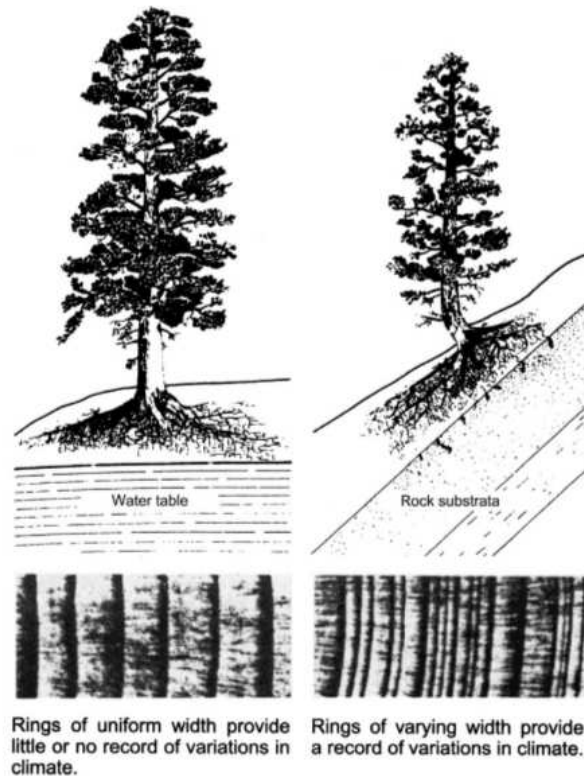

Figure 2.2: The difference between a complacent tree (left) and a sensitive tree (right). *From* Fritts (1976).

The procedure to build-up a tree-ring series for a site, that is a *dendrochronology*, is made up of the following steps:

1. Sample selection: trees are sampled by removing two or tree cores of wood from the radius of the tree, on at least 20 trees per site.

2. Cross-dating: ring-width sequences from each core are compared, so that characteristic patterns of ring-width variation are matched and the age of the sample is established (Fig. 2.3).

3. Standardization (in a climatological sense): the width of a ring depends on many variables, like the tree species and age, in addiction to the climatic factors. In order that the different cores can build-up a chronology, it is necessary to remove the growth function peculiar of each tree. This procedure, named standardization, has the consequence that some of the low frequency climatic information is removed, so that tree-rings rarely provide information at frequencies greater than a few hundred years. Growth functions are removed by fitting a curve (commonly a negative exponential function) to the data and dividing each measured ring-width value by the corresponding value on the growth curve. A further step could be to estimate an auto-regressive model of order 1, to eliminate the influence that one year has on the following, by using the residual series.
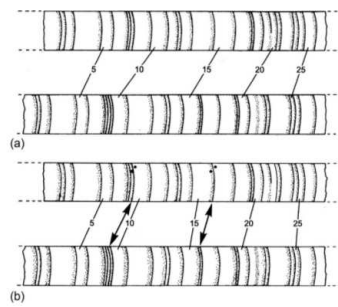
Figure 2.3: Cross-dating of tree-rings. Panel (a): counting of the rings shows a lack of synchrony in the patterns of the two cores. Panel (b): dots in the upper core show where a ring is missing and the patterns can be matched. *From* Fritts (1976).

4. Chronology construction: the indexes obtained from the previous step are averaged among specimens to produce a mean chronology for that site (Fig. 2.4).
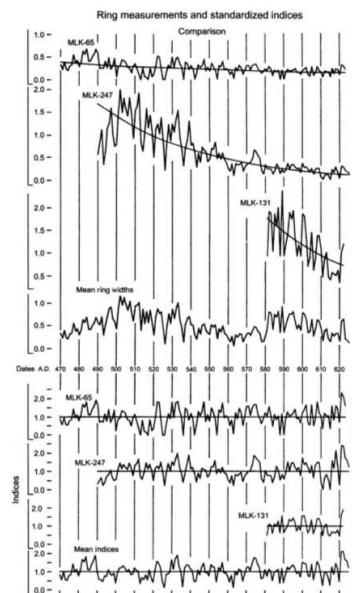
Figure 2.4: Chronology construction. The upper part of the figure shows three series and their mean without standardization, when the mean is influenced by age. The bottom part shows the series after standardization, when they can averaged among specimens differing in age, to produce the series of mean indexes in the bottom line. *From* Fritts (1976).

The chronology used in this work is obtained as outlined above, and it is considered reliable, without taking into account of the uncertainties introduced by its construction.

Chapter 3

Model estimation

In this Chapter we present a model for the estimation of annual climatic variables from annual tree-ring series in Section 3.1, and a model for the estimation of monthly climatic variables from annual tree-ring series in Section 3.2. Some results from a simulation study for both the models are illustrated in Section 3.3.

3.1 Annual model

We consider an annual model in which the climatic variables are the summer mean temperature $T(t)$ and the summer mean moisture $M(t)$ for year t, computed by averaging the corresponding monthly variables of June, July and August (JJA).

It is convenient to label the years in the calibration period $\mathcal{C}$ from 1 to n_C and the years in the reconstruction period $\mathcal{R}$ from 1 to n_R. Let $n = n_C + n_R$. We replace climatic variables with transformations suitable for the model definition, called *anomalies*, similar to what is done in Tolwinski-Ward et al. (2015). Temperature anomaly $x_1(t)$ is a function of temperature $T(t)$:

$$x_1(t) = \frac{T(t) - \mu_T}{\sigma_T}, \ t \in \mathcal{R} \cup \mathcal{C}, \tag{3.1}$$

where μ_T and σ_T are the mean and standard deviation of the values $T(t)$ in the calibration period.

Moisture anomalies $x_2(t)$ are computed from moisture values $M(t)$:

$$x_2(t) = F_{norm}^{-1}\left(\hat{F}\left(\frac{M(t) - M_n}{M_x - M_n}\right)\right), \ t \in \mathcal{R} \cup \mathcal{C}, \tag{3.2}$$

where $M_n = 0.001$ and $M_x = 0.999$ are the minimum and maximum allowable soil moisture values, $\hat{F}$ is a non-parametric estimate of the cumulative distribution function of normalized moisture in the calibration period, F_{norm}^{-1} is the quantile function of the standard Gaussian distribution. The relationship between tree-ring widths and climatic variables on year t is:

$$y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \epsilon(t), \qquad \epsilon(t) \sim N(0, \psi^2), \ t \in \mathcal{R} \cup \mathcal{C}, \tag{3.3}$$

while the climatic variables follow a vector autoregressive model of order 1 (VAR(1)):

$$x(t) = \Phi x(t-1) + \eta(t), \qquad \eta(t) \sim N_2(0, \Sigma), \quad t \in \mathcal{R} \cup \mathcal{C}, \tag{3.4}$$

where $x(t) = (x_1(t), x_2(t))^\top$.

In the estimation procedure we retain the estimates of Φ whose eigenvalues are in modulus less than 1, to ensure the stationarity condition is satisfied (Tsay, 2005).

In the calibration period we can estimate anomalies from climatic variables; on the other hand, in the reconstruction procedure we first get estimates of anomalies conditionally on tree-ring widths observations and then we retrieve temperature and moisture by inverting (3.1) and (3.2).

In the Bayesian context, prior distributions are assigned to the parameters Φ, Σ, $\beta = (\beta_0, \beta_1, \beta_2)^\top$ and ψ^2.

Let
$$\Phi = \begin{pmatrix} \phi(1,1) & \phi(1,2) \\ \phi(2,1) & \phi(2,2) \end{pmatrix}$$
and λ_1, λ_2 be the eigenvalues of Φ. The prior distribution for Φ is

$$p(\Phi) = \begin{cases} \prod_{i,j=1}^{2} p(\phi(i,j)) & \text{if } |\lambda_k| < 1 \text{ for } k = 1,2, \\ 0 & \text{otherwise,} \end{cases}$$

where $p(\phi(i,j)) = N(0,10)$ for $i,j = 1,2$.

As prior distribution for Σ we choose an Inverted-Wishart distribution:

$$p(\Sigma) = IW_2(\nu_0, S_0^{-1}) \text{ with } \nu_0 = 4, S_0^{-1} = \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}^{-1}.$$

In this case the mean is
$$S_0 = \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix},$$

in fact we suppose that the higher the temperature the less the soil moisture. In previous studies we made on about 20 Italian datasets, we have always observed a negative but not strong correlation between temperature and soil moisture, deeming the value -0.4 reasonable also for this location.

The prior distribution for β is $p(\beta) = N_3(\mathbf{0}_3, \Sigma_0)$ with

$$\Sigma_0 = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}.$$

The prior distribution for ψ^2 is an Inverted-Gamma distribution, $p(\psi^2) = IG(a,b)$, where hyperparameters are chosen by considering the relationship with the moments. Since $E(\psi^2) = b/(a-1)$ and $\text{Var}(\psi^2) = b^2/((a-1)^2(a-2))$, then

$$a = E^2(\psi^2)/\text{Var}(\psi^2) + 2,$$
$$b = (a-1)E(\psi^2).$$

We express our prior beliefs on $E(\psi^2)$ on the basis of previous studies made on Italian datasets, in which estimates $\hat{\psi}^2$ of ψ^2 were obtained by the residual mean square after fitting a linear model like (3.3) to the data by ordinary least squares method. We choose a quite high value for the variance with respect to the mean, in order to get a vague distribution.

The model (3.3)–(3.4) implies that $y(t)$ and (Φ, Σ) are conditionally independent given $x(t)$:

$$y(t) \perp (\Phi, \Sigma)|x(t), \ \forall t \tag{3.5}$$

and that

$$(x(t), \Phi, \Sigma) \perp (\beta, \psi^2), \ \forall t. \tag{3.6}$$

We also assume that the parameters are independent to each other, so:

$$p(\Phi, \Sigma, \beta, \psi^2) = p(\Phi)p(\Sigma)p(\beta)p(\psi^2). \tag{3.7}$$

Let $X = \{x(t) : t \in \mathcal{R} \cup \mathcal{C}\}$ the process level variables and $Y = \{y(t) : t \in \mathcal{R} \cup \mathcal{C}\}$ the data level variables. We are interested in the posterior distribution

$$p(X, \Phi, \Sigma, \beta, \psi^2|Y) \propto p(Y|X, \Phi, \Sigma, \beta, \psi)p(X, \Phi, \Sigma, \beta, \psi). \tag{3.8}$$

Let us consider the first factor on the right-hand side of (3.8):

$$p(Y|X, \Phi, \Sigma, \beta, \psi) = p(Y|X, \beta, \psi^2),$$

because of (3.5).

The second factor on the right-hand side is:

$$p(X, \Phi, \Sigma, \beta, \psi^2) = p(X|\Phi, \Sigma)p(\Phi)p(\Sigma)p(\beta)p(\psi^2),$$

because of (3.6) and (3.7). The posterior distribution can then be written:

$$p(X, \Phi, \Sigma, \beta, \psi^2|Y) = \frac{p(Y|X, \beta, \psi^2)p(X|\Phi, \Sigma)p(\Phi)p(\Sigma)p(\beta)p(\psi^2)}{\int \dots \int p(Y|X, \beta, \psi^2)p(X|\Phi, \Sigma)p(\Phi)p(\Sigma)p(\beta)p(\psi^2)\,\mathrm{d}X\,\mathrm{d}\Phi\,\mathrm{d}\Sigma\,\mathrm{d}\beta\,\mathrm{d}\psi^2}.$$

The integral in the denominator is too difficult to be computed, so we need an approximate solution. A good algorithm to solve this problem is the Gibbs sampler (Section A.2):

1. Set the initial values:

$$\Phi^{(0)} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \Sigma^{(0)} \sim IW_2\left(4, \ \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}^{-1}\right), \ \beta^{(0)} = (0,0,0)^\top, \ \psi^{2(0)} = 0.8.$$

2. For iteration $i = 1 \dots, I$:

    (a) sample $\Phi^{(i)}$ from $p(\Phi|\Sigma^{(i-1)}, \beta^{(i-1)}, \psi^{2(i-1)}, X_C, Y_C)$;

    (b) sample $\Sigma^{(i)}$ from $p(\Sigma|\Phi^{(i)}, \beta^{(i-1)}, \psi^{2(i-1)}, X_C, Y_C)$;

    (c) sample $\beta^{(i)}$ from $p(\beta|\Phi^{(i)}, \Sigma^{(i)}, \psi^{2(i-1)}, X_C, Y_C)$;

    (d) sample $\psi^{2(i)}$ from $p(\psi^2|\Phi^{(i)}, \Sigma^{(i)}, \beta^{(i)}, X_C, Y_C)$;

    (e) sample X_R from $p(X|\Phi^{(i)}, \Sigma^{(i)}, \beta^{(i)}, \psi^{2(i)}, Y_R)$,

with $X_C = \{x(t) : t \in \mathcal{C}\}$, $Y_C = \{y(t) : t \in \mathcal{C}\}$, $X_R = \{x(t) : t \in \mathcal{R}\}$, $Y_R = \{y(t) : t \in \mathcal{R}\}$.

In steps (2a)–(2d) we condition on data from the calibration period to estimate Φ, Σ, β and ψ^2, while in step (2e) we use the Gibbs sampler for state-space models to perform the reconstruction, as explained in Section A.4, where X corresponds to the vector of temperature and moisture anomalies in the reconstruction period, $m = 2$ and a_0 and P_0 are estimated as the mean vector and variance-covariance matrix of the anomalies in the calibration period.

From now till the end of the Section we will compute the full conditional distributions for steps (2a)–(2d). For simplicity we omit the subscript C from X_C and Y_C. Letting $x(-1) = (x(2), \dots, x(n_C))$, the full conditional distribution of Φ is:

$$p(\Phi|\Sigma, \beta, \psi^2, X, Y) \propto p(X|\Phi, \Sigma)p(\Phi)$$
$$\propto [p(x(-1)|x(1), \Phi, \Sigma)p(x(1)|\Phi, \Sigma)]\, p(\Phi).$$

From (3.4), $p(x(t)|x(t-1), \Phi, \Sigma) = N_2(\Phi x(t-1), \Sigma)$, while the marginal distribution of $x(1)|\Phi, \Sigma$ is $N_2(0, \Gamma_0)$ with $\Gamma_0 = \sum_{i=0}^\infty \Phi^i\Sigma(\Phi^i)^\top$ (Tsay, 2005). To estimate Γ_0 we stop the summation when

$S(\Gamma_{i+1} - \Gamma_i) = S(\Phi^{i+1}\Sigma(\Phi^{i+1})^\top) \leq 10^{-6}S(\Gamma_i)$, with $S(A)$ the sum of the coefficients of matrix A in absolute value.

So we have

$$p(\Phi|\Sigma, \beta, \psi^2, X, Y) \propto \prod_{t=2}^{n_C} \exp\left\{-\frac{1}{2}(x(t) - \Phi x(t-1))^\top \Sigma^{-1}(x(t) - \Phi x(t-1))\right\}$$

$$\cdot (\det(\Gamma_0))^{-1/2} \exp\left\{-\frac{1}{2}(x(1))^\top \Gamma_0^{-1} x(1)\right\}$$

$$\cdot \prod_{i=1}^{2}\prod_{j=1}^{2} \exp\left\{-\frac{1}{2}\frac{(\phi(i,j))^2}{10}\right\} \mathbb{I}_F(\Phi) \Big/ p(F),$$

where $\mathbb{I}$ is the indicator function of the set $F = \{\Phi : |\lambda_k| < 1 \text{ for } k = 1, 2\}$, λ_k are the eigenvalues of Φ.

Since it is not possible to draw values directly from this distribution, we use a Metropolis algorithm within the Gibbs sampler (Section A.2) where the target is the full-conditional distribution:

1. Choose a value for the variance δ^2 of the proposal distribution and an initial value for Φ, $\Phi^{(1)}$. Set a counter for the acceptance rate, $a = 0$.

2. For $s = 1, \ldots, I$ (number of iterations):

   (a) Generate a candidate matrix $\Phi^{(*)} = (\phi_{i,j}^*)$ with $\phi_{i,j}^* \sim N(\phi_{i,j}^{(s)}, \delta^2)$, $i, j = 1, \ldots, 2$, and check for its stationarity. If stationarity is not reached, set $\Phi^{(s+1)} = \Phi^{(s)}$.

   (b) Compute the acceptance ratio:

   $$r_\Phi = \frac{p(\Phi^{(*)}|\Sigma, \beta, \psi^2, X, Y)}{p(\Phi^{(s)}|\Sigma, \beta, \psi^2, X, Y)} = \frac{p(x(-1)|x(1), \Phi^*, \Sigma)p(x(1)|\Phi^*, \Sigma)p(\Phi^*)}{p(x(-1)|x(1), \Phi^{(s)}, \Sigma)p(x(1)|\Phi^{(s)}, \Sigma)p(\Phi^{(s)})}.$$

   This expression is an approximation that does not take into account of $p(F)$.

   (c) Generate $u \sim U(0,1)$. If $u < r_\Phi$ set $\Phi^{(s+1)} = \Phi^{(*)}$ and $a = a + 1$, otherwise set $\Phi^{(s+1)} = \Phi^{(s)}$.

3. Compute the acceptance rate as $a_r = a/M$.

We tune δ^2 to have $0.2 < a_r < 0.5$. Values not included in this interval reveal problems in the mixing of the chain.

The full conditional distribution of Σ is

$$p(\Sigma|\Phi, \beta, \psi^2, X, Y) \propto p(X|\Phi, \Sigma)p(\Sigma)$$

$$\propto p(x(-1)|x(1), \Phi, \Sigma)p(x(1)|\Phi, \Sigma)p(\Sigma)$$

$$\propto (\det(\Gamma_0))^{-1/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(x(1)(x(1))^\top \Gamma_0^{-1}\right)\right\}$$

$$\cdot (\det(\Sigma))^{-(n_C + \nu_0 + 2)/2}$$

$$\cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\left((S_0 + S_\Phi)\Sigma^{-1}\right)\right\},$$

with $S_\Phi = \sum_{t=2}^{n_C}(x(t) - \Phi x(t-1))(x(t) - \Phi x(t-1))^\top$.

We can write

$$p(\Sigma|\Phi, \beta, \psi^2, X, Y) \propto f(\Sigma)q(\Sigma),$$

with $f(\Sigma) = (\det(\Gamma_0))^{-1/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(x(1)(x(1))^\top \Gamma_0^{-1}\right)\right\}$ and

$$q(\Sigma) = (\det(\Sigma))^{-(n_C+\nu_0+2)/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\left((S_0 + S_\Phi)\Sigma^{-1}\right)\right\},$$

that is proportional to a density of an Inverted-Wishart distribution with parameters $\nu_1 = \nu_0 + n_C - 1$ and $S_1^{-1} = (S_0 + S_\Phi)^{-1}$. This suggests to use a Metropolis-Hastings algorithm, taking $IW_2(\nu_1, S_1^{-1})$ as proposal distribution.

In this case the acceptance ratio is

$$
\begin{aligned}
r_\Sigma &= \frac{p(\Sigma^{(*)}|\Phi, \beta, \psi^2, X, Y)}{p(\Sigma^{(s)}|\Phi, \beta, \psi^2, X, Y)} \cdot \frac{q(\Sigma^{(s)})}{q(\Sigma^{(*)})} \\
&= \frac{(\det(\Gamma_0^{(*)}))^{-1/2} \exp\left\{-\frac{1}{2}(x(1))^\top (\Gamma_0^{(*)})^{-1} x(1)\right\}}{(\det(\Gamma_0^{(s)}))^{-1/2} \exp\left\{-\frac{1}{2}(x(1))^\top (\Gamma_0^{(s)})^{-1} x(1)\right\}}.
\end{aligned}
$$

The full conditional distribution of β is

$$p(\beta|\Phi, \Sigma, \psi^2, X, Y) \propto p(Y|X, \beta, \psi^2)p(\beta).$$

Letting $\mathring{X}$ the $n_C \times 3$ matrix whose $t-$th row is $(1, x_1(t), x_2(t))$, we have:

$$p(Y|X, \beta, \psi^2) = N_{n_C}(\mathring{X}\beta, \psi^2 I_{n_C}) \propto \exp\left\{-\frac{1}{2}(Y - \mathring{X}\beta)^\top (Y - \mathring{X}\beta)/\psi^2\right\},$$

$$p(\beta) = N_3(\beta_0, \Sigma_0) \propto \exp\left\{-\frac{1}{2}(\beta - \beta_0)^\top \Sigma_0^{-1}(\beta - \beta_0)\right\}.$$

We can expand the products as follows:

$$(Y - \mathring{X}\beta)^\top (Y - \mathring{X}\beta)/\psi^2 = Y^\top Y/\psi^2 - 2\beta^\top \mathring{X}^\top Y/\psi^2 + \beta^\top \mathring{X}^\top \mathring{X}\beta/\psi^2,$$

and

$$(\beta - \beta_0)^\top \Sigma_0^{-1}(\beta - \beta_0) = \beta^\top \Sigma_0^{-1}\beta - \beta^\top \Sigma_0^{-1}\beta_0 - \beta_0^\top \Sigma_0^{-1}\beta + \beta_0^\top \Sigma_0^{-1}\beta_0.$$

And we can write the full conditional distribution for β as:

$$p(\beta|\Phi, \Sigma, \psi^2, X, Y) \propto \exp\left\{-\frac{1}{2}\left[-2\beta^\top \left(\frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1}\beta_0\right) + \beta^\top \left(\frac{\mathring{X}^\top \mathring{X}}{\psi^2} + \Sigma_0^{-1}\right)\beta\right]\right\}. \quad (3.9)$$

If we denote with β_1 and Σ_1 the mean vector and the variance-covariance matrix of the full conditional distribution, we should be able to recognize in (3.9) the quadratic form

$$(\beta - \beta_1)^\top \Sigma_1^{-1}(\beta - \beta_1) = \beta^\top \Sigma_1^{-1}\beta - 2\beta^\top \Sigma_1^{-1}\beta_1 + \beta_1^\top \Sigma_1^{-1}\beta_1.$$

Only the first two terms depend on β and from the correspondence with (3.9) we get

$$\Sigma_1^{-1} = \frac{\mathring{X}^\top \mathring{X}}{\psi^2} + \Sigma_0^{-1}$$

and

$$\Sigma_1^{-1}\beta_1 = \frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1}\beta_0,$$

so that the full conditional distribution of β is $N_3(\beta_1, \Sigma_1)$ with

$$\Sigma_1 = \left(\frac{\mathring{X}^\top \mathring{X}}{\psi^2} + \Sigma_0^{-1} \right)^{-1}$$

and

$$\beta_1 = \Sigma_1 \left(\frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1} \beta_0 \right).$$

The full conditional distribution of ψ^2 is:

$$p(\psi^2|\Phi, \Sigma, \beta, X, Y) \propto p(Y|X, \beta, \psi^2) p(\psi^2)$$

$$\propto (\psi^2)^{-n_C/2 - a - 1} \exp\left\{ \left[-\frac{1}{2}(Y - \mathring{X}\beta)^\top (Y - \mathring{X}\beta) - b \right] / \psi^2 \right\},$$

that is the kernel of an Inverted-Gamma distribution of parameters

$$a_1 = n_C/2 + a \text{ and } b_1 = \frac{1}{2}(Y - \mathring{X}\beta)^\top (Y - \mathring{X}\beta) + b.$$

3.2   Monthly model

In this Section we develop a model for the estimation of monthly climatic data from annual tree-ring widths in the calibration period. To deal with monthly data, we introduce the notation $z(h)$ for the variable z in month h, with $h = 1, \ldots, 12n_C$ in the calibration period. Sometimes we need to show that the variable refers to month j of year t, so we can write $z(12(t-1) + j)$ with $j = 1, \ldots, 12$. Temperature anomalies are computed according to the monthly equivalent of expression (3.1), that is

$$x_1(h) = \frac{T(h) - \mu_T}{\sigma_T}, \ h \in \mathcal{C}, \tag{3.10}$$

where μ_T and σ_T are the mean and standard deviation of the monthly values $T(h)$ in the calibration period. The expression for moisture anomalies is

$$x_2(h) = F_{norm}^{-1} \left(\hat{F} \left(\frac{M(h) - M_n}{M_x - M_n} \right) \right), \ h \in \mathcal{C}, \tag{3.11}$$

where M_n and M_x are the same as before and $\hat{F}$ is estimated on monthly data in the calibration period.

The monthly model relates monthly climatic variables to annual tree-ring widths:

$$y(t) = \beta_0 + \sum_{j=1}^{12} \sum_{i=1}^{2} \beta_{i,j} x_i(12(t-1) + j) + \varepsilon(t), \qquad \varepsilon(t) \sim N(0, \psi^2), \ t \in \mathcal{C}, \tag{3.12}$$

while the climatic variables follow a seasonal autoregressive model

$$x(h+1) = \Phi_1 x(h) + \Phi_S x(h-11) + \Phi_2 x(h-12) + \eta(h+1), \qquad \eta(h+1) \sim N_2(0, \Sigma), \tag{3.13}$$

where $\Phi_2 = -\Phi_1 \Phi_S$, $x(h) = (x_1(h), x_2(h))^\top$ and $h \in \mathcal{C}$.

This corresponds to the model

$$(I_2 - \Phi_1 B)(I_2 - \Phi_S B^{12}) x(h+1) = \eta(h+1), \ \eta(h+1) \sim N_2(0, \Sigma),$$

where B is the backward operator, so that the stationarity condition is that the eigenvalues of Φ_1 and Φ_S are in modulus less than 1.

After the estimation we retrieve temperature and moisture from anomalies, by inverting equations (3.10) and (3.11).

We are interested in the posterior distribution $p(X, \Phi_1, \Phi_S, \Sigma, \beta, \psi^2 | Y)$, where $X = \{x(h) : h \in \mathcal{C}\}$ and $Y = \{y(t) : t \in \mathcal{C}\}$. Model (3.12)–(3.13) implies the assumptions that $y(t)$ is conditionally independent on Φ_1, Φ_S and Σ given $x(12(t-1)+1), \ldots, x(12(t-1)+12)$:

$$y(t) \perp (\Phi_1, \Phi_S, \Sigma) | (x(12(t-1)+1), \ldots, x(12(t-1)+12)), \ \forall t;$$

that $X, \Phi_1, \Phi_S, \Sigma$ are independent on β and ψ^2

$$(x(h), \Phi_1, \Phi_S, \Sigma) \perp (\beta, \psi^2), \ \forall h;$$

and that the parameters are independent to each other, so that

$$p(\Phi_1, \Phi_S, \Sigma, \beta, \psi^2) = p(\Phi_1)p(\Phi_S)p(\Sigma)p(\beta)p(\psi^2).$$

The posterior distribution can be written as:

$$p(X, \Phi_1, \Phi_S, \Sigma, \beta, \psi^2 | Y) \propto p(Y|X, \beta, \psi^2)p(X|\Phi_1, \Phi_S, \Sigma)p(\Phi_1)p(\Phi_S)p(\Sigma)p(\beta)p(\psi^2). \tag{3.14}$$

It is not possible to sample directly from it. As in the annual case, we generate values alternatively from the full conditional distributions:

1. Set the initial values:

$$\Phi_1^{(0)} = \Phi_S^{(0)} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \Sigma^{(0)} \sim IW_2\left(4, \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}^{-1}\right), \ \beta^{(0)} = \mathbf{0}_{25}, \ \psi^{2(0)} = 0.8.$$

2. For iteration $i = 1 \ldots, I$:

   (a) sample $(\Phi_1, \Phi_S)^{(i)}$ from $p(\Phi_1, \Phi_S | \Sigma^{(i-1)}, \beta^{(i-1)}, \psi^{2(i-1)}, X_C, Y_C)$;

   (b) sample $\Sigma^{(i)}$ from $p(\Sigma | \Phi_1^{(i)}, \Phi_S^{(i)}, \beta^{(i-1)}, \psi^{2(i-1)}, X_C, Y_C)$;

   (c) sample $\beta^{(i)}$ from $p(\beta | \Phi_1^{(i)}, \Phi_S^{(i)}, \Sigma^{(i)}, \psi^{2(i-1)}, X_C, Y_C)$;

   (d) sample $\psi^{2(i)}$ from $p(\psi^2 | \Phi_1^{(i)}, \Phi_S^{(i)}, \Sigma^{(i)}, \beta^{(i)}, X_C, Y_C)$.

In order to determine the full conditional distributions we examine the factors of (3.14) separately, omitting the subscript C of X_C and Y_C for simplicity. Letting $\mathring{X}$ an $n_C \times 25$ matrix whose $t-$th row is $(1, x(12(t-1)+1), \ldots, x(12(t-1)+12))$ the first factor of (3.14) is:

$$p(Y|X, \beta, \psi^2) = N_{n_c}(\mathring{X}\beta, \psi^2 I_n)$$
$$\propto (\psi^2)^{-n_C/2} \exp\left\{-\frac{1}{2}(Y - \mathring{X}\beta)^\top (Y - \mathring{X}\beta)/\psi^2\right\}.$$

Let $x(-(1:13)) = (x(14), \ldots, x(12n_C))^\top$ and $x(1:13) = (x(1), \ldots, x(13))^\top$. The second factor then is:

$$p(X|\Phi_1, \Phi_S, \Sigma) = p(x(-(1:13))|x(1:13), \Phi_1, \Phi_S, \Sigma)p(x(1:13)|\Phi_1, \Phi_S, \Sigma)$$
$$\approx p(x(-(1:13))|x(1:13), \Phi_1, \Phi_S, \Sigma)$$
$$\propto (\det(\Sigma))^{-\frac{12n_C - 13}{2}}$$
$$\cdot \exp\left\{-\frac{1}{2} \sum_{h=14}^{12n_C} (x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))^\top\right.$$
$$\left. \Sigma^{-1} (x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))\right\}.$$

We do not know the marginal distribution of $(x(1), \ldots, x(13))^\top$, so in this case we approximate the required distribution with $p(x(14), \ldots, x(12n_C)|x(1), \ldots, x(13), \Phi_1, \Phi_S, \Sigma)$.

Now we start considering the bits referred to the prior distributions. We choose the same prior distribution for both Φ_1 and Φ_S, given the respect of the stationarity condition:

Let

$$\Phi_a = \begin{pmatrix} \phi_a(1,1) & \phi_a(1,2) \\ \phi_a(2,1) & \phi_a(2,2) \end{pmatrix},$$

with $a \in \{1, S\}$ and λ_1, λ_2 be the eigenvalues of Φ_a. The prior distribution for Φ_a is

$$p(\Phi_a) = \begin{cases} \prod_{i,j=1}^2 p(\phi_a(i,j)) & \text{if } |\lambda_k| < 1 \text{ for } k = 1, 2, \\ 0 & \text{otherwise,} \end{cases}$$

where $p(\phi_a(i,j)) = N(0, 10)$ for $i, j = 1, 2$.

The prior distribution for β is:

$$p(\beta) = N_{25}(\beta_0, \Sigma_0) \propto \exp\left\{-\frac{1}{2}(\beta - \beta_0)^\top \Sigma_0^{-1}(\beta - \beta_0)\right\},$$

with $\beta_0 = \mathbf{0}_{25}$ and $\Sigma_0 = 10 \cdot I_{25}$.

The prior distributions for Σ and ψ^2 are the same of the annual model:

$$p(\Sigma) = IW(\nu_0, S_0^{-1}),$$

with $\nu_0 = 4$ and

$$S_0^{-1} = \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}^{-1},$$

so

$$p(\Sigma) \propto (\det(\Sigma))^{-(\nu_0+3)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}(S_0 \Sigma^{-1})\right\},$$

while

$$p(\psi^2) \propto (\psi^2)^{-a-1} \exp\left\{-\frac{b}{\psi^2}\right\}.$$

From previous studies, we deemed reasonable to set the parameters to have $E(\psi^2) = 0.5$ and $\operatorname{Var}(\psi^2) = 8$, that is $a = 2.031$ and $b = 0.516$.

The full-conditional distribution for (Φ_1, Φ_S) is:

$$p(\Phi_1, \Phi_S | \Sigma, \beta, \psi^2, X, Y) \propto p(X|\Phi_1, \Phi_S, \Sigma)p(\Phi_1)p(\Phi_S)$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{h=14}^{12n_C}(x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))^\top \right.$$

$$\left. \Sigma^{-1}(x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))\right\}$$

$$\cdot \prod_{i=1}^2 \prod_{j=1}^2 \exp\left\{-\frac{1}{2}\frac{(\phi_1(i,j))^2}{10}\right\} \mathbb{I}_{F_1}(\Phi_1)$$

$$\cdot \prod_{i=1}^2 \prod_{j=1}^2 \exp\left\{-\frac{1}{2}\frac{(\phi_S(i,j))^2}{10}\right\} \mathbb{I}_{F_S}(\Phi_S) \Big/ p(F_1 \cap F_S),$$

where $F_1 = \{\Phi_1 : |\lambda_{1,k}| < 1 \text{ for } k = 1, 2\}$, $F_S = \{\Phi_S : |\lambda_{S,k}| < 1 \text{ for } k = 1, 2\}$, $\lambda_{1,k}$ are the eigenvalues of Φ_1, $\lambda_{S,k}$ are the eigenvalues of Φ_S.

This distribution does not show a known form, so we use a Metropolis algorithm, where the proposal distribution is Gaussian and the acceptance ratio is

$$r = \frac{p(X|\Phi_1^*, \Phi_S^*, \Sigma)p(\Phi_1^*)p(\Phi_S^*)}{p(X|\Phi_1^{(s)}, \Phi_S^{(s)}, \Sigma)p(\Phi_1^{(s)})p(\Phi_S^{(s)})},$$

where, as in the annual case, we use an approximation dropping $p(F_1 \cap F_S)$.

The full conditional distribution of Σ is:

$$p(\Sigma|\Phi_1, \Phi_S, \beta, \psi^2, X, Y) \propto p(X|\Phi_1, \Phi_S, \Sigma)p(\Sigma)$$

$$\propto \det(\Sigma)^{-\frac{12n_C-13}{2}}$$

$$\cdot \exp\left\{ -\frac{1}{2} \operatorname{tr} \left(\sum_{h=14}^{12n_C} (x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13)) \right.\right.$$

$$\left.\left. \cdot (X(h) - \Phi_1 X(h-1) - \Phi_S X(h-12) - \Phi_2 X(h-13))^\top \Sigma^{-1} \right) \right\}$$

$$\cdot (\det(\Sigma))^{-(\nu_0+3)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}(S_0 \Sigma^{-1}) \right\}.$$

This is an Inverted-Wishart distribution of parameters $\nu_0 + 12n_C - 13$ and $(S_0 + S_\Phi)^{-1}$, with

$$S_\Phi = \sum_{h=14}^{12n_C} (x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))$$

$$\cdot (x(h) - \Phi_1 x(h-1) - \Phi_S x(h-12) - \Phi_2 x(h-13))^\top.$$

The full conditional distribution of β is

$$p(\beta|\Phi_1, \Phi_S, \Sigma, \psi^2, X, Y) \propto p(Y|X, \beta, \psi^2)p(\beta)$$

$$\propto \exp\left\{ -\frac{1}{2} \left[-2\beta^\top \left(\frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1}\beta_0 \right) + \beta^\top \left(\frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1}\beta_0 \right) \beta \right] \right\}.$$

This expression is proportional to a Gaussian density $N_{25}(\beta_1, \Sigma_1)$ with mean vector β_1 and variance-covariance matrix Σ_1, where

$$\Sigma_1 = \left(\frac{\mathring{X}^\top \mathring{X}}{\psi^2} + \Sigma_0^{-1} \right)^{-1}$$

and

$$\beta_1 = \Sigma_1 \left(\frac{\mathring{X}^\top Y}{\psi^2} + \Sigma_0^{-1}\beta_0 \right).$$

The full conditional distribution of ψ^2 is:

$$p(\psi^2|\Phi, \Sigma, \beta, X, Y) \propto p(Y|X, \beta, \psi^2)p(\psi^2)$$

$$\propto (\psi^2)^{-n_C/2-a-1} \exp\left\{ \left[-\frac{1}{2}(Y - \mathring{X}\beta)^\top(Y - \mathring{X}\beta) - b \right] / \psi^2 \right\},$$

that is the kernel of an Inverted-Gamma distribution of parameters

$$a_1 = n_C/2 + a \text{ and } b_1 = \frac{1}{2}(Y - \mathring{X}\beta)^\top(Y - \mathring{X}\beta) + b.$$

3.3 Simulation study

3.3.1 Annual model

We performed a simulation study on the annual model, studying the influence on the results of two factors:

1. the variance δ^2 of the proposal distribution used for the estimation of Φ;

2. the direction of the series. A first idea to perform back-casting can be to reverse the series and proceed as we were going forward; we would like to check if this can have an influence on the results.

We chose two levels for each factor, obtaining therefore four different scenarios:

1. $\delta^2 = 1/10000$ and forward series;

2. $\delta^2 = 1/10000$ and backward series;

3. $\delta^2 = 1/1000$ and forward series;

4. $\delta^2 = 1/1000$ and backward series.

For each scenario we have simulated 500 datasets, using the same set of values, that in their turn were estimated from real data:

$$\Phi^{(sim)} = \begin{pmatrix} 0.386 & 0.039 \\ 0.011 & 0.255 \end{pmatrix}, \ \Sigma^{(sim)} = \begin{pmatrix} 0.867 & -0.242 \\ -0.242 & 0.562 \end{pmatrix},$$
$$\beta^{(sim)} = (-0.016, -0.163, 0.593)^\top, \ \psi^{2^{(sim)}} = 0.819.$$

The reconstruction period is 7895 years long, while the calibration lasts 85 years. For each dataset we saved 10000 estimated series, drawn from a total number of 201000 iterations, after discarding the first 1000 for the warm-up and taking one value each 20. The performance of the procedure can be tested by evaluating the correlation between the estimated series and the simulated one, but it can depend on the values used to simulate. To get a basis for comparison we first compute a sample of 1000 correlation among the simulated series. From results in Fig. 3.1 it seems that we cannot expect high correlations.

In Figures 3.2(a)–3.2(d) the results for the four scenarios. We computed the median of the correlations between the 10000 estimated series and the simulated one, and we draw the boxplot of the 500 correlations. We can see no substantial differences among the scenarios, suggesting that the direction of the series has little influence and that the two values proposed for δ^2 do not engender different performance, in fact the acceptance rate is in both cases quite high, between 0.7 and 0.9.

We report in Figure 3.3 the result of a cross-validation experiment, in which we reconstruct the first 60 years of the calibration period on the basis of the last 25. The procedure seems able to catch the overall pattern of the series.

3.3.2 Monthly model

We simulate a dataset with a calibration period of 200 years using the parameters below, estimated from real data, measured on the Italian Alps (12.04°E, 46.32°N). The tree-ring widths chronology is made available by F.H. Schweingruber on the website of the United States National Oceanic and
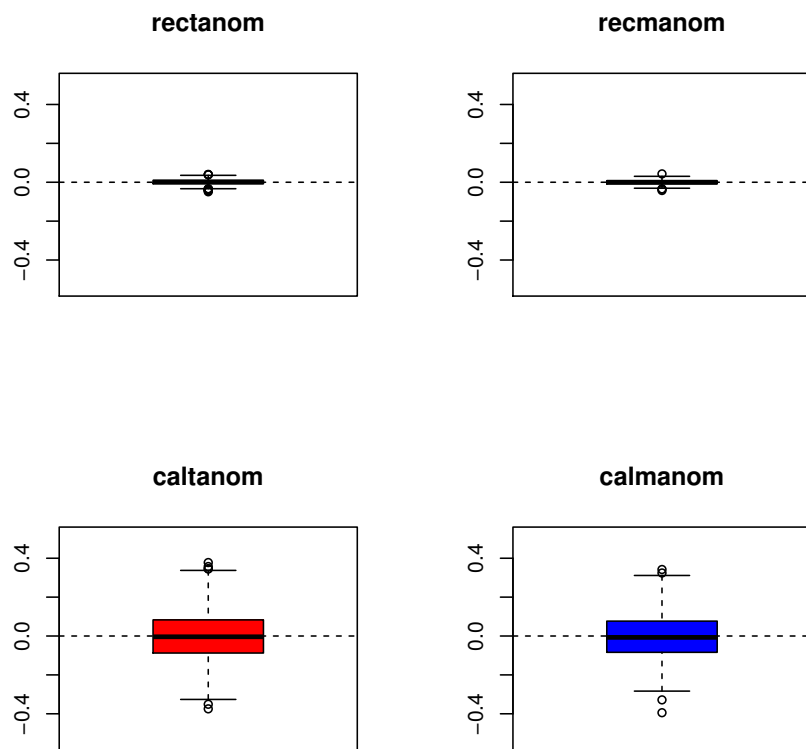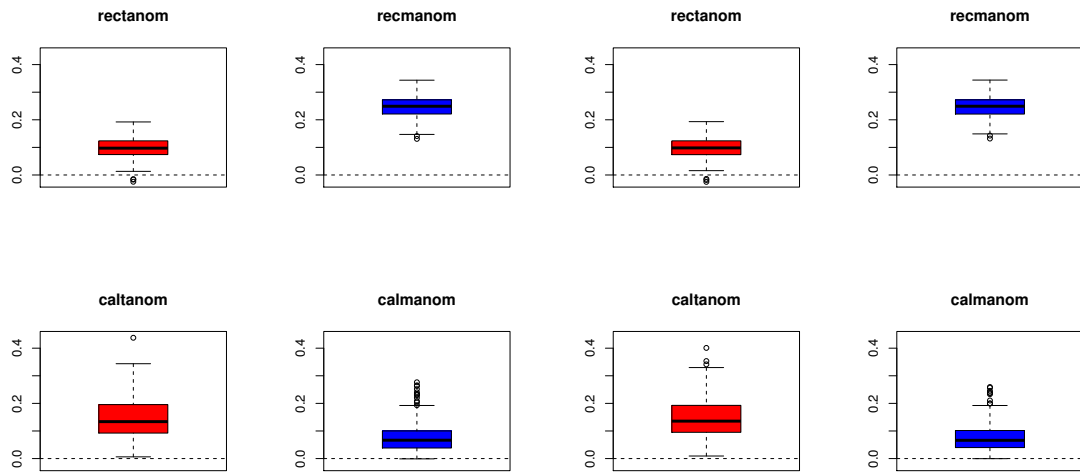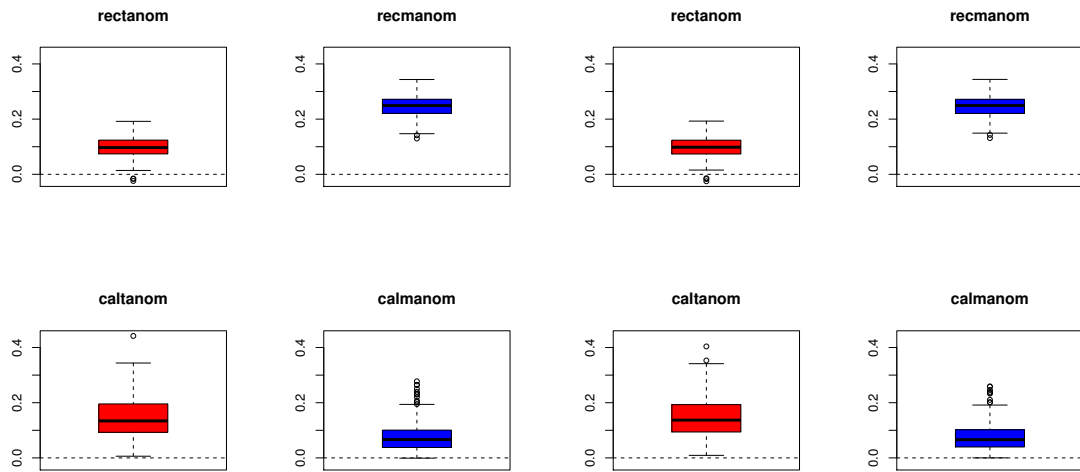
Figure 3.1: Annual model: sample of the process correlations.

(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

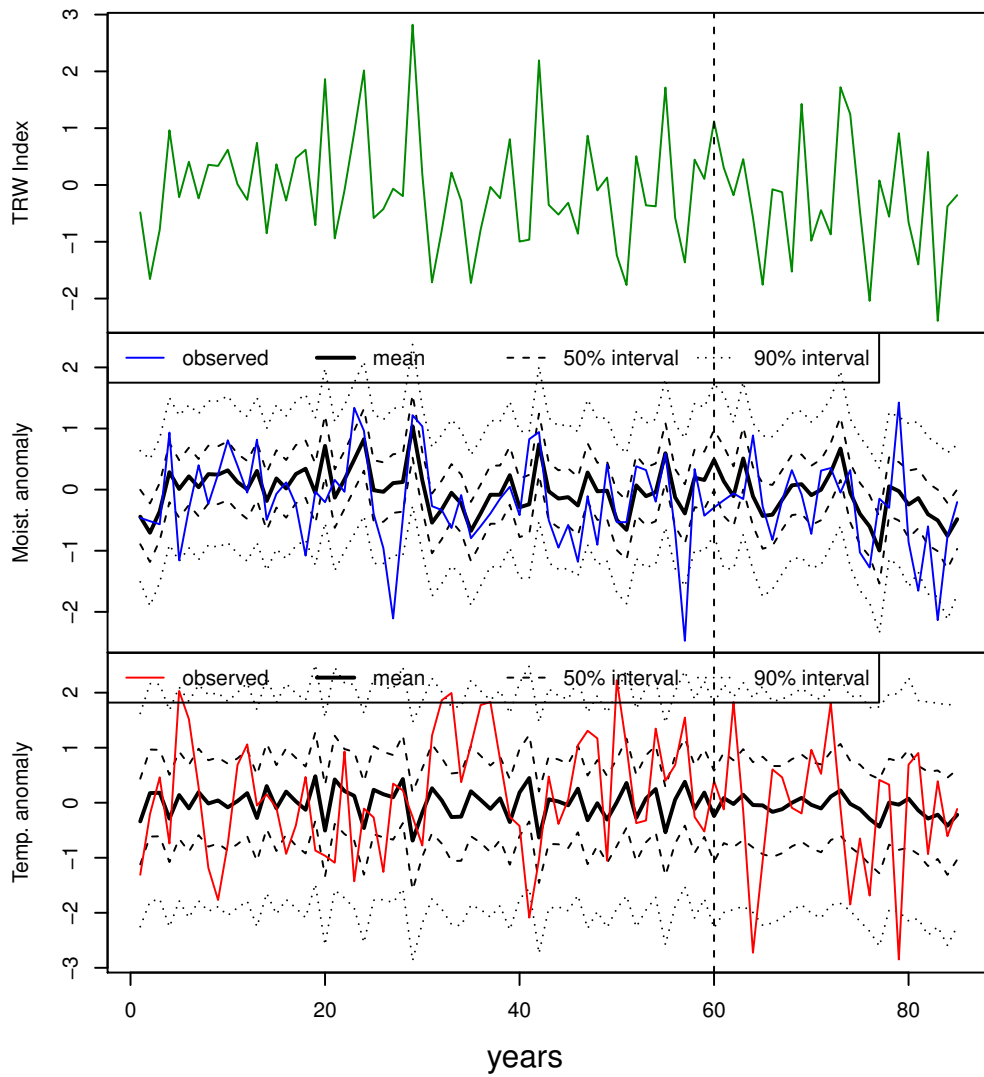Figure 3.2: Annual model: median correlations.

Figure 3.3: Annual model: crossvalidation experiment.

Atmospheric Administration (`https://www.ncdc.noaa.gov/paleo/study/4390`); temperature is retrieved from the Climatic Research Unit (CRU) Database of the University of East Anglia (Harris et al., 2014), while moisture is computed via the leaky-bucket model implemented in the R package VSLiteR (Tolwinski-Ward, 2015; Tolwinski-Ward et al., 2011; Huang et al., 1996).

- $vec(\Phi_1^{sim}) = (0.218, 0.037, -0.009, 0.553)^\top$;

- $vec(\Phi_S^{sim}) = (0.003, -0.015, 0.041, -0.003)$;

- $vec(\Sigma^{sim}) = (0.949, -0.105, -0.105, 0.404)$;

- Intercept $\beta_0^{sim} = 0.002$;

- Coefficients of temperature anomalies:

  $(\beta_1^{sim}, \dots, \beta_{12}^{sim}) = (0.148, 0.091, -0.097, -0.09, 0.177, 0.125, 0.163, -0.037, 0.067, 0.121, -0.087, -0.136)$;

- Coefficients of moisture anomalies:

  $(\beta_{13}^{sim}, \dots, \beta_{24}^{sim}) = (-0.049, 0.243, 0.066, -0.08, -0.027, 0.003, -0.182, 0.011, 0.094, -0.194, 0.11, 0.091)$;

- $\psi^{2^{sim}} = 0.404$.

We report in Figure 3.4 the result for the first 300 months, that seems satisfactory. The remaining periods show a similar behaviour and the corresponding graphs are reported in Appendix B, Figg. B.1–B.7.
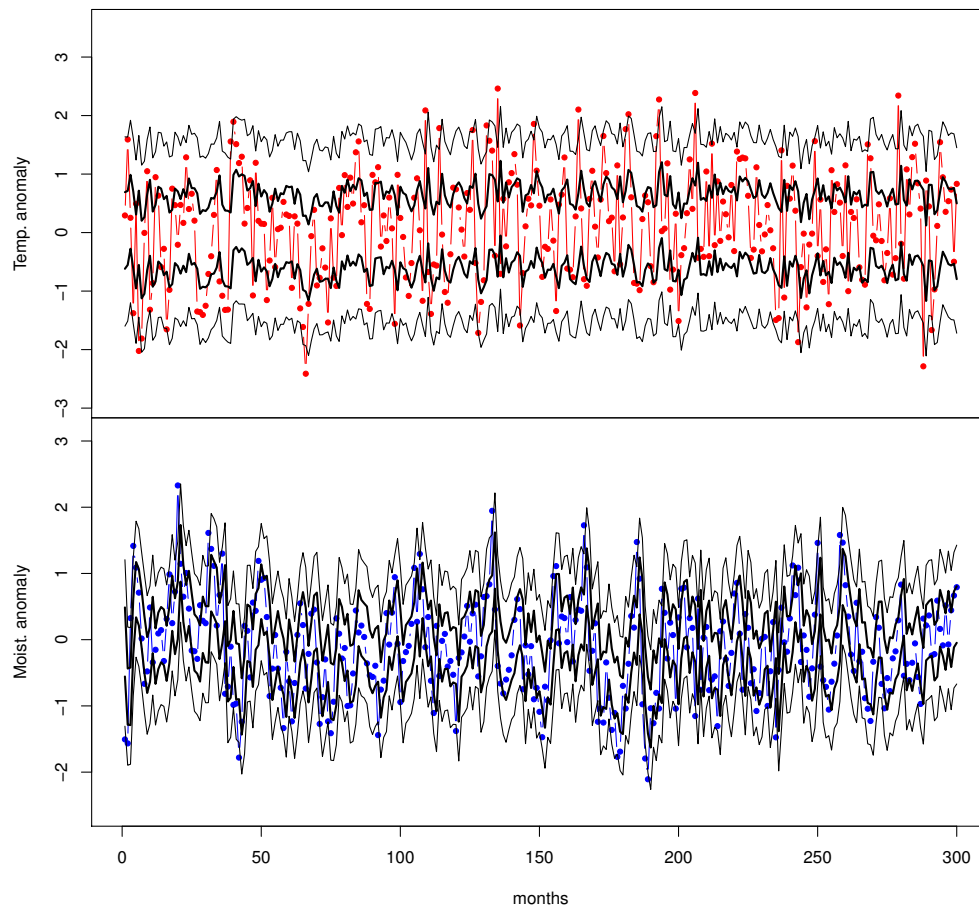
Figure 3.4: Simulated series in the calibration period, months from 1 to 300. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.

Chapter 4

Data

In this Chapter we present an application of the methodology outlined in Chapter 3 to a location in the Italian Central Alps named Val d'Avio (Fig. 4.1). We consider as climatic variables temperature and moisture in Section 4.1. Temperature is retrieved from the Climatic Research Unit (CRU) Database of the University of East Anglia (Harris et al., 2014) that has a resolution of $0.5° × 0.5°$. Moisture is computed via the leaky-bucket model implemented in the R package VSLiteR (Tolwinski-Ward, 2015; Tolwinski-Ward et al., 2011; Huang et al., 1996). The tree species is European larch.

For this site another kind of moisture measurement, the PDSI index (Dai et al., 2004), is available. In Section 4.2 we present the application for temperature and PDSI, while in Section 4.3 we do a comparison among different methods of reconstruction.

The tree-ring widths chronology and PDSI data for this location were kindly provided by University of Pisa.

They built the site chronology starting from raw data series `ital040`, `ital041` , `ital042` and `ital043`, available on the International Tree-Ring Data Base (`https://www.ncdc.noaa.gov/paleo/study/19875`). Raw data on their hand were built in a dendro-climatic perspective, taking into account of the altitude, the characteristics of the site and the correlation with temperature.

The series were standardized in a climatological sense, dividing them by the theoretical values given by a negative exponential function. An auto-regressive model of order 1 was fit to the standardized series, and residuals were used to build the mean site series, in order to eliminate the influence between consecutive years.

Tree-ring widths indexes are positive, so we standardize the chronology to get an index that varies on the real line, and that can be used as response variable in equation (3.3) or (3.12).

4.1 Site A - Val d'Avio: temperature and moisture

The tree-ring widths chronology in this area (Coppola et al., 2013) is related to temperature and moisture referring to the cell 10°– 10.50°E, 46°– 46.50°N of the CRU database, and we take as annual datum the mean temperature and moisture on June, July and August (JJA).

The observed series of tree-ring widths, summer-mean temperature and moisture in the 1901–2008 period are in Fig. 4.2. We can notice an increase in temperature since 1980. Scatterplots for the variables in pairs are reported along with histograms in Fig. B.8 of Appendix B. Sometimes climatic variables on the whole calibration period do not show the best correlation with tree-ring widths series. We compute in Tab. 4.1 the correlation on some 50-years subsets of the period, to
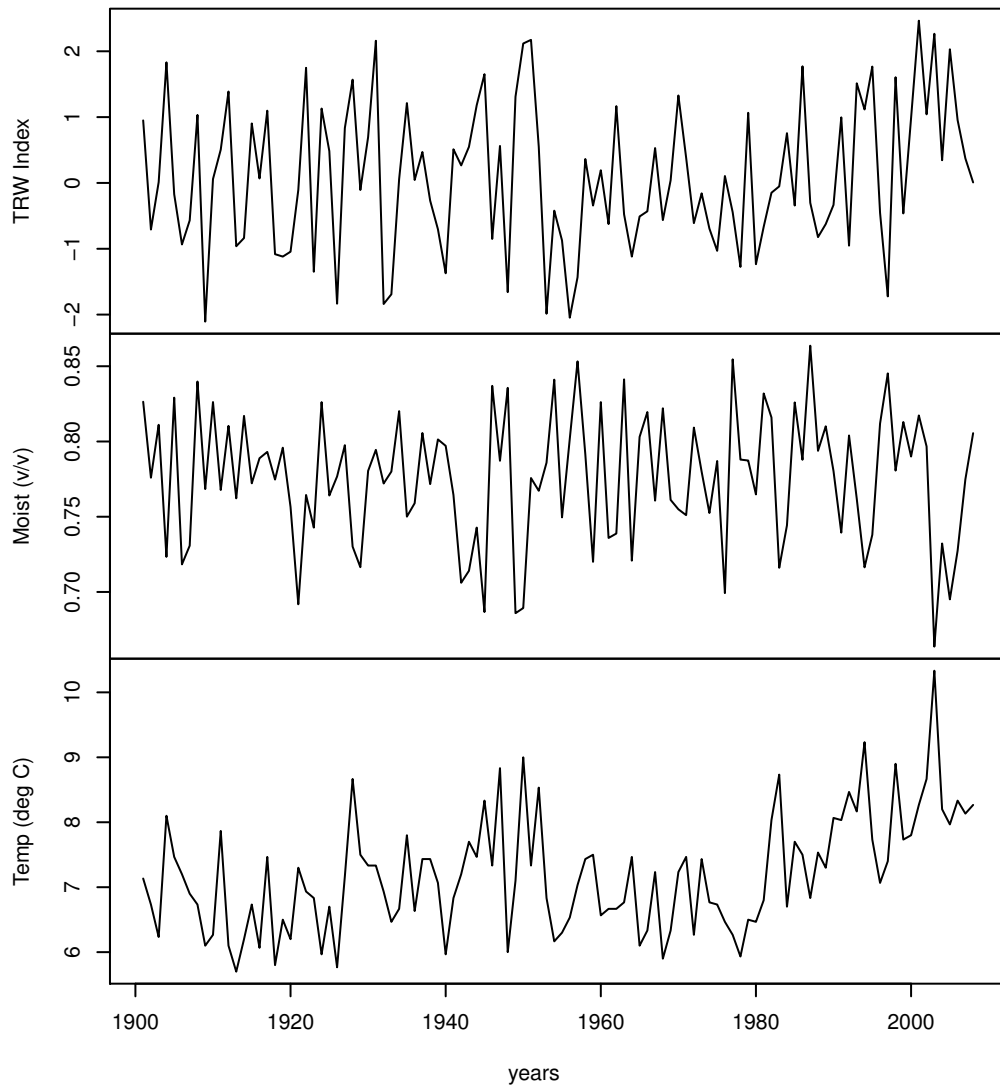
Figure 4.1: Val d'Avio site.

Figure 4.2: Site A. From top to bottom: Tree-Ring Widths index, JJA-mean moisture and JJA-mean temperature series, observed from 1901 to 2008.

	Temp	Moist
1901–1950	0.58	-0.21
1916–1965	0.56	-0.33
1931–1980	0.56	-0.43
1946–1995	0.45	-0.47
whole	0.53	-0.32

Table 4.1: Site A. Correlations between tree-ring widths series, temperature and moisture, on different 50-years subsets of the calibration period.

check if it can be improved, and we decide to use for calibration the years from 1931 to 1980.

Estimated series are in Fig. B.9. The annual algorithm seems not able to catch properly the raise in temperature in the recent years.

Let us estimate climatic variables from 1901 to 2008 with the monthly algorithm. We use the same calibration period as before. Results for monthly series are in Fig. B.10 and B.11 (first year is not shown because its estimate depends too much on the algorithm's starting values), while in Fig. B.12 and B.13 we compute the JJA average from monthly estimates, to compare it with the results from the annual algorithm. The monthly algorithm catches the behaviour of the series better than the annual one, and it is able also to show the raise in temperature level from 1980 on.

4.1.1 Annual algorithm diagnostics

After some trials, we set the variance of the proposal distribution for Φ at 0.0125, obtaining an acceptance rate of 0.412. The acceptance rate for Σ is 0.943, too high. To get a proper one, we should act on the hyperparameters of the Inverted-Wishart distribution mentioned on page 13, to raise its variance. Since the treatment of the variance is difficult and the results of the reconstruction and diagnostics are satisfactory, we decide to keep this value anyway.

	mean	sd	5%	25%	med	75%	95%
$\Phi(1,1)$	0.15	0.16	-0.11	0.05	0.15	0.25	0.41
$\Phi(2,1)$	0.05	0.11	-0.13	-0.02	0.05	0.12	0.24
$\Phi(1,2)$	-0.05	0.22	-0.41	-0.20	-0.06	0.09	0.31
$\Phi(2,2)$	-0.08	0.16	-0.34	-0.19	-0.08	0.03	0.19
$\Sigma(1,1)$	0.99	0.20	0.71	0.85	0.96	1.10	1.37
$\Sigma(2,1)$	-0.35	0.11	-0.55	-0.41	-0.34	-0.27	-0.18
$\Sigma(2,2)$	0.51	0.11	0.37	0.44	0.50	0.57	0.70
β_0	-0.10	0.13	-0.30	-0.18	-0.10	-0.01	0.11
β_1	0.50	0.14	0.26	0.40	0.50	0.59	0.73
β_2	-0.34	0.20	-0.68	-0.48	-0.34	-0.21	-0.01
ψ^2	0.81	0.17	0.57	0.69	0.78	0.90	1.11

Table 4.2: Site A: parameters' estimation summary (temperature/moisture model, annual algorithm).

We collected a final sample of size 10000 after 1000 iterations for the burn-in and thinning by 20. Some statistics on the parameters' values are reported in Tab. 4.2. For each parameter we investigated the plot of the chain values and the histogram, boxplot, normal quantile plot, empirical auto-correlation function plot and running mean plot of the final sample, to detect problems in terms of exploration of the parameter space, autocorrelation and stationarity. Results show now particular issues, and we collected some of the graphs in Figg. B.14–B.17.

4.1.2 Monthly algorithm diagnostics

In this case we set the variance of the proposal distribution for Φ at 0.00025, obtaining an acceptance rate of 0.415.

We collected a final sample of size 10000 after 2000 iterations for the burn-in and thinning by 60, with no evidence of problems in terms of exploration of the parameter space, autocorrelation and stationarity. Diagnostic pictures for some of the parameters are shown in Figg. B.18–B.22.

In Fig. B.23 there are the box-plots of the monthly coefficients of temperature anomalies, while in Fig. B.24 the ones for moisture anomalies. We can notice the seasonality and that the effect of temperature anomalies on tree-ring widths is substantially positive in June and August, but negative in July.

Tab. 4.3 shows a summary of the parameters' estimates.

	mean	sd	5%	25%	med	75%	95%
$\Phi_1(1,1)$	0.20	0.05	0.13	0.17	0.20	0.23	0.28
$\Phi_1(2,1)$	-0.03	0.08	-0.17	-0.08	-0.03	0.03	0.11
$\Phi_1(1,2)$	0.04	0.02	0.01	0.03	0.04	0.05	0.06
$\Phi_1(2,2)$	0.64	0.03	0.59	0.62	0.64	0.66	0.69
$\Phi_S(1,1)$	0.89	0.02	0.86	0.88	0.89	0.90	0.92
$\Phi_S(2,1)$	-0.19	0.05	-0.27	-0.22	-0.19	-0.15	-0.10
$\Phi_S(1,2)$	-0.04	0.02	-0.07	-0.05	-0.04	-0.03	-0.01
$\Phi_S(2,2)$	0.09	0.04	0.02	0.06	0.09	0.12	0.16
$\Sigma(1,1)$	0.13	0.01	0.12	0.13	0.13	0.14	0.14
$\Sigma(2,1)$	-0.02	0.01	-0.04	-0.03	-0.02	-0.01	-0.00
$\Sigma(2,2)$	0.49	0.03	0.44	0.47	0.49	0.51	0.54
β_0	-3.17	1.25	-5.19	-4.00	-3.19	-2.34	-1.06
β_1	-0.00	0.44	-0.72	-0.29	-0.01	0.29	0.72
β_2	0.14	0.34	-0.42	-0.09	0.13	0.36	0.69
β_3	-0.45	0.39	-1.10	-0.71	-0.46	-0.19	0.19
β_4	-0.51	0.39	-1.16	-0.78	-0.51	-0.25	0.13
β_5	0.32	0.67	-0.77	-0.13	0.32	0.77	1.40
β_6	1.40	0.64	0.34	0.99	1.41	1.83	2.43
β_7	-1.18	0.73	-2.38	-1.66	-1.18	-0.70	0.01
β_8	2.29	0.60	1.30	1.90	2.30	2.70	3.26
β_9	-1.03	0.50	-1.85	-1.35	-1.03	-0.70	-0.20
β_{10}	0.51	0.52	-0.36	0.16	0.51	0.86	1.35
β_{11}	-0.31	0.60	-1.29	-0.70	-0.31	0.09	0.68
β_{12}	0.07	0.44	-0.66	-0.22	0.07	0.36	0.79
β_{13}	-0.17	0.16	-0.44	-0.27	-0.17	-0.06	0.09
β_{14}	0.28	0.20	-0.05	0.15	0.28	0.41	0.61
β_{15}	0.24	0.21	-0.11	0.10	0.24	0.38	0.58
β_{16}	-0.19	0.26	-0.63	-0.37	-0.19	-0.02	0.23
β_{17}	0.30	0.37	-0.31	0.05	0.30	0.55	0.91
β_{18}	-0.78	0.33	-1.34	-1.00	-0.78	-0.57	-0.23
β_{19}	-0.61	0.25	-1.03	-0.77	-0.61	-0.44	-0.19
β_{20}	0.21	0.23	-0.17	0.06	0.21	0.36	0.59
β_{21}	0.10	0.18	-0.19	-0.01	0.10	0.22	0.39
β_{22}	-0.46	0.15	-0.72	-0.57	-0.46	-0.36	-0.21
β_{23}	0.33	0.17	0.06	0.22	0.33	0.44	0.61
β_{24}	0.12	0.15	-0.12	0.02	0.12	0.22	0.36
ψ^2	0.37	0.10	0.23	0.29	0.35	0.42	0.56

Table 4.3: Site A: parameters' estimation summary (temperature/moisture model, monthly algorithm).
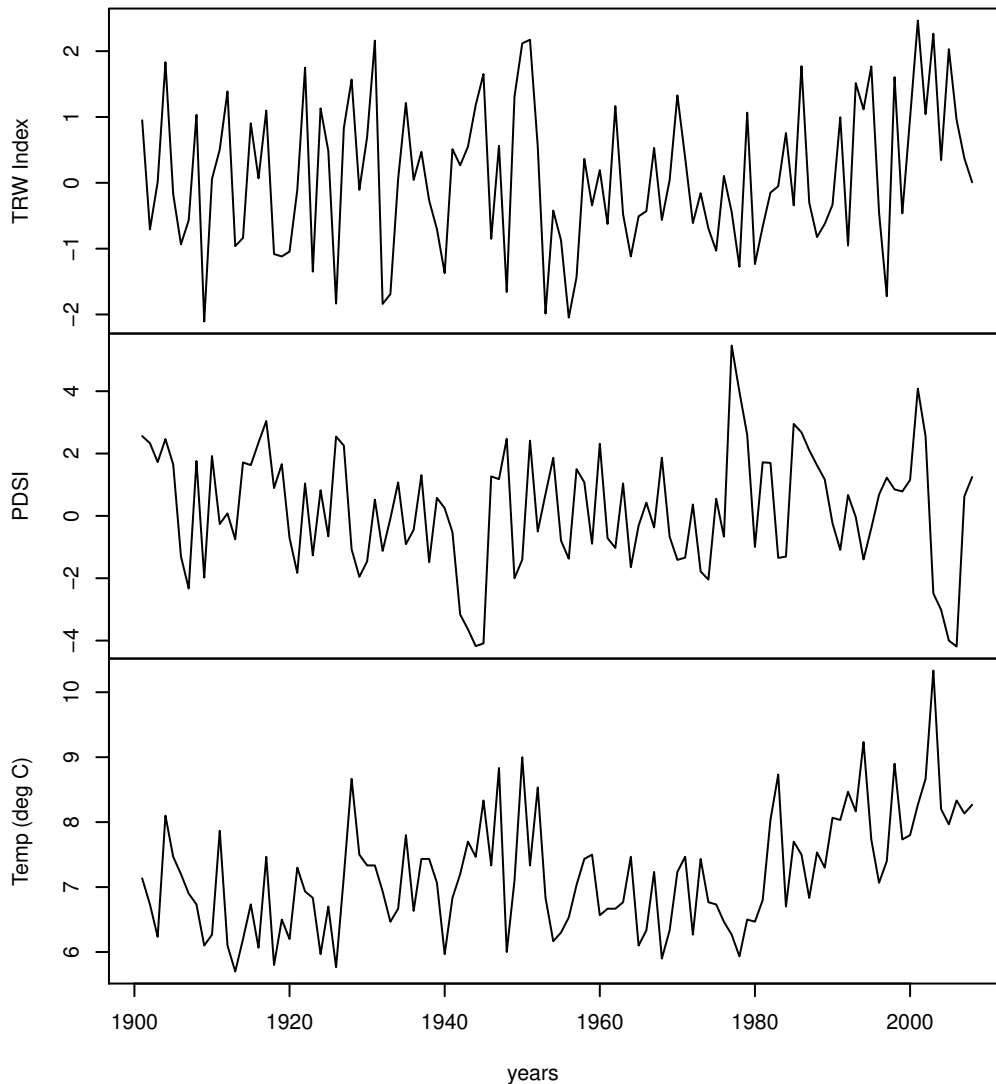
Figure 4.3: Site A. From top to bottom: JJA-mean temperature, JJA-mean PDSI and Tree-Ring Widths index series, observed from 1901 to 2008.

4.2   Site A - Val d'Avio: temperature and PDSI

On site A we have also observations of the PDSI index, a variable that can be used in place of moisture anomalies (Dai et al., 2004).

The series of tree-ring widths, summer-mean temperature and PDSI in the whole calibration period are in Fig. 4.3. We can notice an increase in temperature from 1980. Scatterplots for the variables in pairs are reported along with histograms in Fig. B.25. Sometimes climatic variables on the whole calibration period do not show the best correlation with tree-ring widths series. We compute in Tab. 4.4 the correlation on some 50-years subsets of the period, to check if the correlation can be improved, and we decide to use for calibration the years from 1931 to 1980.

Estimated series are in Fig. B.26. The annual algorithm seems not to be able to catch properly the raise in temperature in the recent years.

	Temp	PDSI
1901–1950	0.58	-0.07
1916–1965	0.56	-0.16
1931–1980	0.56	-0.23
1946–1995	0.45	-0.17
whole	0.53	-0.09

Table 4.4: Site A. Correlations between tree-ring widths series and each of the climatic variables on different 50-years subsets of the calibration period. (PDSI)

Let us estimate climatic variables from 1901 to 2008 with the monthly algorithm. We use the same calibration period as before.

Results for monthly series are in Fig. B.27 and B.28, while in Fig. B.29 and B.30 we compute the JJA average from monthly estimates, to compare it with the results from the annual algorithm. The monthly algorithm catches the behaviour of the series better than the annual one, and it is able also to show the raise in temperature level from 1980 on.

4.2.1 Annual algorithm diagnostics

After some trials, we set the variance of the proposal distribution for Φ at 0.01, obtaining an acceptance rate of 0.414. The acceptance rate for Σ is 0.94, leading to the same reasoning done in Section 4.1.1.

	mean	sd	5%	25%	med	75%	95%
$\Phi(1,1)$	0.08	0.15	-0.17	-0.02	0.08	0.19	0.33
$\Phi(2,1)$	0.44	0.28	-0.02	0.26	0.45	0.63	0.89
$\Phi(1,2)$	-0.10	0.08	-0.23	-0.16	-0.10	-0.05	0.03
$\Phi(2,2)$	0.40	0.15	0.16	0.31	0.40	0.50	0.63
$\Sigma(1,1)$	0.96	0.20	0.69	0.82	0.94	1.07	1.32
$\Sigma(2,1)$	-0.71	0.27	-1.18	-0.87	-0.69	-0.52	-0.32
$\Sigma(2,2)$	3.15	0.65	2.25	2.70	3.07	3.51	4.32
β_0	-0.09	0.13	-0.31	-0.18	-0.09	-0.01	0.12
β_1	0.61	0.14	0.37	0.52	0.61	0.71	0.85
β_2	0.00	0.08	-0.12	-0.05	0.00	0.05	0.13
ψ^2	0.85	0.18	0.61	0.73	0.83	0.95	1.17

Table 4.5: Site A: parameters' estimation summary (temperature/PDSI model, annual algorithm).

We collected a final sample of size 10000 after 1000 iterations for the burn-in and thinning by 20. Some statistics on the parameters' values are reported in Tab. 4.5. For each parameter we investigated the same plots as in the case of moisture, with no evident problems in terms of exploration of the parameter space, autocorrelation and stationarity. Graphs for a subset of the parameters are gathered in Figg. B.31–B.34.

4.2.2 Monthly algorithm diagnostics

In this case we set the variance of the proposal distribution for Φ at 2e-04, obtaining an acceptance rate of 0.258. We collected a final sample of size 10000 after 2000 iterations for the burn-in and thinning by 60, with again no evidence of problems in terms of exploration of the parameter space, autocorrelation and stationarity. Diagnostic pictures for some of the parameters are shown in Figg. B.35–B.39.

In Fig. B.40 there are the box-plots of the monthly coefficients of temperature anomalies, while in Fig. B.41 the ones for PDSI anomalies. We can notice the seasonality and that the effect of temperature anomalies on tree-ring widths is substantially positive in June and August, but negative in July.

Tab. 4.6 shows a summary of the parameters' estimates.

	mean	sd	5%	25%	med	75%	95%
$\Phi_1(1,1)$	0.20	0.04	0.13	0.17	0.20	0.23	0.27
$\Phi_1(2,1)$	-0.03	0.10	-0.20	-0.10	-0.03	0.04	0.13
$\Phi_1(1,2)$	0.00	0.01	-0.01	-0.00	0.00	0.01	0.02
$\Phi_1(2,2)$	0.90	0.02	0.87	0.89	0.90	0.92	0.93
$\Phi_S(1,1)$	0.90	0.02	0.87	0.89	0.90	0.92	0.93
$\Phi_S(2,1)$	-0.01	0.06	-0.11	-0.05	-0.01	0.03	0.10
$\Phi_S(1,2)$	-0.01	0.01	-0.03	-0.02	-0.01	-0.00	0.01
$\Phi_S(2,2)$	0.01	0.04	-0.06	-0.02	0.01	0.04	0.08
$\Sigma(1,1)$	0.13	0.01	0.12	0.13	0.13	0.14	0.15
$\Sigma(2,1)$	-0.00	0.01	-0.03	-0.01	-0.00	0.00	0.02
$\Sigma(2,2)$	0.76	0.04	0.69	0.73	0.76	0.79	0.83
β_0	-2.27	1.44	-4.63	-3.22	-2.28	-1.35	0.09
β_1	0.06	0.53	-0.81	-0.30	0.06	0.41	0.94
β_2	0.51	0.35	-0.07	0.28	0.51	0.73	1.08
β_3	-0.77	0.45	-1.51	-1.07	-0.77	-0.47	-0.01
β_4	0.02	0.46	-0.74	-0.29	0.02	0.32	0.77
β_5	0.42	0.54	-0.46	0.06	0.42	0.78	1.31
β_6	2.37	0.63	1.34	1.96	2.37	2.79	3.38
β_7	-0.74	0.76	-1.98	-1.25	-0.75	-0.24	0.50
β_8	1.15	0.65	0.06	0.72	1.15	1.58	2.20
β_9	-1.21	0.60	-2.20	-1.61	-1.21	-0.81	-0.21
β_{10}	1.04	0.60	0.04	0.64	1.04	1.44	2.00
β_{11}	0.13	0.58	-0.83	-0.27	0.13	0.52	1.08
β_{12}	-0.10	0.47	-0.86	-0.42	-0.11	0.21	0.68
β_{13}	-0.07	0.12	-0.27	-0.15	-0.07	0.01	0.13
β_{14}	0.02	0.18	-0.27	-0.10	0.02	0.14	0.32
β_{15}	0.43	0.27	-0.02	0.25	0.43	0.61	0.88
β_{16}	-0.09	0.25	-0.49	-0.25	-0.09	0.08	0.33
β_{17}	0.15	0.23	-0.23	-0.00	0.15	0.30	0.53
β_{18}	-0.06	0.21	-0.40	-0.20	-0.06	0.08	0.28
β_{19}	-0.47	0.23	-0.84	-0.62	-0.47	-0.32	-0.09
β_{20}	-0.07	0.25	-0.48	-0.24	-0.07	0.09	0.34
β_{21}	0.22	0.22	-0.14	0.07	0.22	0.37	0.58
β_{22}	-0.35	0.16	-0.61	-0.46	-0.35	-0.24	-0.09
β_{23}	0.22	0.15	-0.03	0.12	0.22	0.31	0.46
β_{24}	0.02	0.12	-0.18	-0.06	0.02	0.10	0.21
ψ^2	0.47	0.13	0.30	0.38	0.45	0.54	0.72

Table 4.6: Site A: parameters' estimation summary (temperature/PDSI model, monthly algorithm).

4.3 Comparison with traditional methods

In this Section we compare the results given by the Bayesian algorithms shown above, to reconstructions obtained by more traditional methods applied in palaeoclimatology, that are based mainly on regression. We first outline the underlining theory, thereafter comparing the results

for temperature, moisture and PDSI. The first traditional method is linear regression, and it is detailed in Bradley (2013), Chapter 13, in two steps:

1. response function analysis: select the climatic variables that can be more effectively reconstructed, on the basis of their correlation with the tree-ring series.

2. regression: perform reconstruction by using a linear regression model with the climatic variable as the predictand and the tree-ring series as the predictor.

Letting $Y(t)$ be the tree-ring width on year t and $K(t)$ a climatic variable that in our case will be in turn temperature, moisture anomaly or PDSI, we have:

$$K(t) = \beta_{K,0} + \beta_{K,1}Y(t) + \zeta_K(t), \qquad \zeta_K(t) \sim N(0, \sigma_K^2), \tag{4.1}$$

and the estimated series is:

$$\hat{K}(t) = \hat{\beta}_{K,0} + \hat{\beta}_{K,1}Y(t). \tag{4.2}$$

The $100(1 - \alpha)\%$ prediction interval for $K(t)$ is

$$(\hat{K}(t) - S_{pred}t_{(n_C-2,\alpha/2)}, \hat{K}(t) + S_{pred}t_{(n_C-2,\alpha/2)})$$

with

$$S_{pred} = \sqrt{\hat{\sigma_K^2}\left(1 + \frac{1}{n_C} + \frac{(Y(t) - \bar{Y})^2}{\sum_{j=1}^{n_C}(Y(j) - \bar{Y})^2}\right)}, \quad \hat{\sigma_K^2} = \frac{\sum_{j=1}^{n_C}(Y(j) - \bar{Y})^2}{n_C - 2},$$

where $t_{(n_C-2,\alpha/2)}$ is the value of the Student's t distribution with $n_C - 2$ degrees of freedom that puts $\alpha/2$ probability in the upper tail, and n_C is the number of observations in the calibration period.

We can select the best calibration period from Tables 4.1 and 4.4. It turns out to be 1901–1950 for temperature, 1946–1995 for moisture and 1931–1980 for PDSI.

The output is reported in Tables 4.7, 4.8 and 4.9 for temperature, moisture and PDSI respectively.

	Estimate	Std. Error	t value	p-value
(Intercept)	5.0861	0.3924	12.96	0.0000
Y	1.8711	0.3761	4.97	0.0000

Table 4.7: Site A: temperature. Parameters' estimation using least squares in the linear model.

	Estimate	Std. Error	t value	p-value
(Intercept)	1.3705	0.3841	3.57	0.0008
Y	-1.4068	0.3828	-3.67	0.0006

Table 4.8: Site A: moisture anomaly. Parameters' estimation using least squares in the linear model.

The value of adjusted R-squared is 0.3264 for temperature, 0.2033 for moisture anomaly and 0.03244 for PDSI, so tree-ring widths explains only part of the climate variability, and particularly for the PDSI the relationship with the tree-ring series is poor. Diagnostic plots of the residuals are shown in Figures B.42, B.43, B.44 of Appendix B and they do not highlight particular problems, taking into account of the sample size.

	Estimate	Std. Error	t value	p-value
(Intercept)	1.6923	1.1310	1.50	0.1411
Y	-1.8391	1.1312	-1.63	0.1106

Table 4.9: Site A: PDSI. Parameters' estimation using least squares in the linear model.

Let us now consider a family of methods that are named scaling methods, used for multiproxy reconstruction in Ahmed et al. (2013). In these methods the tree-ring series is standardized in the statistical sense, and the climatic variable is estimated as:

$$\hat{K}(t) = \mu_K + \sigma_K Y(t), \tag{4.3}$$

where μ_K is the mean of the climatic series in the calibration period and σ_K can be determined in different ways (Lee et al., 2008):

1. Direct regression: estimating the model

$$K(t) = \sigma_K Y(t) + \epsilon(t).$$

2. Inverse regression: estimating the model

$$Y(t) = \sigma_K K(t) + \epsilon(t),$$

    in which case we compute $\hat{K}(t) = Y(t)/\hat{\sigma}_K$.

3. Variance matching: σ_K is the standard deviation of the climatic series in the calibration period, so in that period $\mathrm{Var}(K) = \mathrm{Var}(\sigma_K Y)$.

We compare these three variants by plotting estimated values and 50% and 90% confidence intervals, obtained from 5000 bootstrap replications, for each variant. Results are in Appendix B in Figures B.45, B.46 and B.47 for temperature, B.48, B.49 and B.50 for moisture and B.51, B.52 and B.53 for PDSI.

The inverse regression algorithm gives implausible estimates for the variables, overestimating the variability. Direct regression estimates have a plausible range, while the variance matching algorithm seems to have the best performance. In the following we compare its results to simple linear regression estimates and our Bayesian model results. Figures B.54, B.55, B.56 refer to temperature, B.57, B.58, B.59 to moisture and B.60, B.61, B.62 to PDSI.

The Bayesian annual algorithm and the traditional linear regression model give similar results in the reconstruction period. During calibration, the variance matching algorithm gives better results with more variability in the estimates, and the monthly Bayesian algorithm seems good in following the pattern of the observations.

We also compute the correlation between the reconstructed series obtained from different methods, splitting them in observed (1901–2008) and reconstruction period (1549–1900).

In Tables 4.10 and 4.11 we give the correlations for temperature. If we consider the correlation with the observed series, the worst method is the annual Bayesian, while others are almost equivalent. Linear regression and scaling methods give perfectly positively correlated estimates, more in agreement with the monthly than the annual Bayesian estimates.

In the reconstruction period traditional methods and the Bayesian algorithm give highly correlated values.

	LR	DR	IR	VM	BA	BM
OBS	0.53	0.53	0.53	0.53	0.38	0.52
LR		1.00	1.00	1.00	0.21	0.29
DR			1.00	1.00	0.21	0.29
IR				1.00	0.21	0.29
VM					0.21	0.29
BA						0.95

Table 4.10: Site A: calibration temperature. Correlation among results from different methods: OBS is the observed series, DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, BM is monthly Bayesian, LR is linear regression.

	DR	IR	VM	BA
LR	1.00	1.00	1.00	0.99
DR		1.00	1.00	0.99
IR			1.00	0.99
VM				0.99

Table 4.11: Site A: reconstruction temperature. Correlation among results from different methods: DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, LR is linear regression.

In Tables 4.12 and 4.13 we give the correlations for moisture. Regarding the agreement with the observed series, monthly Bayesian algorithm seems to have the best performance. The annual Bayesian algorithm has very low correlation with the observations, because it underestimates the variability of the series. Variance matching method gives negative correlation because it does not take into account of the sign of the proxy-climate relationship, but among the scaling methods it seems to give the best results in catching the range of the observed values. In the reconstruction period, methods substantially agree, apart from the variance matching algorithm.

In Tables 4.14 and 4.15 we give the correlations for PDSI. Also in this case the first table says that the monthly Bayesian algorithm is the most suitable for the estimation of the observed series, followed by the annual Bayesian. Traditional methods are in agreement among them, apart from variance matching, but not too much in agreement with Bayesian methods. Regarding reconstruction there is substantial agreement among the methods, apart from variance matching algorithm.

	LR	DR	IR	VM	BA	BM
OBS	0.34	0.34	0.34	-0.33	-0.06	0.80
LR		1.00	1.00	-1.00	-0.22	0.35
DR			1.00	-1.00	-0.22	0.35
IR				-1.00	-0.22	0.35
VM					0.22	-0.35
BA						-0.20

Table 4.12: Site A: calibration moisture. Correlation among results from different methods: OBS is the observed series, DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, BM is monthly Bayesian, LR is linear regression.

	DR	IR	VM	BA
LR	1.00	1.00	-1.00	0.98
DR		1.00	-1.00	0.98
IR			-1.00	0.98
VM				-0.98

Table 4.13: Site A: reconstruction moisture. Correlation among results from different methods: DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, LR is linear regression.

	LR	DR	IR	VM	BA	BM
OBS	0.10	0.10	0.10	-0.10	0.26	0.97
LR		1.00	1.00	-1.00	-0.01	0.06
DR			1.00	-1.00	-0.01	0.06
IR				-1.00	-0.01	0.06
VM					0.01	-0.06
BA						0.31

Table 4.14: Site A: calibration PDSI. Correlation among results from different methods: OBS is the observed series, DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, BM is monthly Bayesian, LR is linear regression.

	DR	IR	VM	BA
LR	1.00	1.00	-1.00	0.87
DR		1.00	-1.00	0.87
IR			-1.00	0.87
VM				-0.87

Table 4.15: Site A: reconstruction PDSI. Correlation among results from different methods: DR is direct regression, IR is inverse regression, VM is variance matching, BA is annual Bayesian, LR is linear regression.

Chapter 5

Conclusions

The problem of palaeoclimate reconstruction is really complex and shows many facets. Traditional approaches based on regression methods assume independence among the observations, without taking into account of the temporal structure of the series, and consider climate as the predictand, that is estimated using tree-ring data as predictors.

The Bayesian approach on the other hand, permits to express the climate-proxy relationship in the proper way, with the proxy series explained on the basis of climate, and then to reverse this relationship, determining the climate configuration that most likely has contributed to the generation of the observed proxy.

This thesis consider a Bayesian framework as well, first showing a model for climate and proxy data with the same (annual) resolution, and then trying to extend it to the estimation of monthly climatic data from annual tree-ring widths observations. The use of an auto-regressive model in the process level explicitly accounts for the temporal dimension.

We have supposed that tree-ring width in year t depends linearly on summer mean temperature and moisture of the same year in the annual model, and on monthly temperature and moisture of the same year in the monthly model. This does not take into account of the time lag in the growth-response of the tree to climate and of the fact that during winter the tree is quiescent, so it could be more realistic to set the dependence of $Y(t)$ on months from September of year $(t-1)$ to August of year t.

Moreover we assume that the linear relationship between proxy and climate does not change through time, and that it persists also in the past, but this could be a limit if the linear relationship is weak, and a non-linear climate-proxy relationship could have occurred in the non-observed period.

Estimation of the parameters in the process-level equation requires to check for the eigenvalues of the matrix Φ. We could improve this step by generating directly matrices that respect the stationarity condition, through their spectral decomposition.

The most natural thing to be done is to complete the work performing monthly reconstruction when only the proxy series is available; this means to extend equation (A.4) to the monthly case, taking into account that the factorization now is more complicated, because of the dependencies in the process level equation (3.13):

$$p(X|Y^n) = p(x(1), \ldots, x(12n), Y^n)/p(Y^n)$$

$$= p(x(12n)|Y^n) \cdot \prod_{l=1}^{11} (p(x(12(n-1)+l|x(12(n-1)+l+1), Y^n))$$

$$\cdot \, p(x(12(n-2)+12)|x(12(n-1)+1), x(12(n-1)+12), Y^{n-1})$$

$$\cdot \prod_{h=1}^{11} (p(x(12(n-2)+h)|x(12(n-2)+h+1), x(12(n-2)+h+12), x(12(n-2)+h+13), Y^{n-1})$$

$$\cdot \prod_{t=1}^{n-2} \prod_{j=1}^{12} (p(x(12(t-1)+j)|x(12(t-1)+j+1), x(12(t-1)+j+12), x(12(t-1)+j+13), Y^t)).$$

If more than one chronology is available for a certain site (e.g. earlywood or latewood density measurements) a further extension is to build-up a multivariate model, to exploit all the information available. In this case $y(t)$ in the observation equation would be a vector of proxy variables.

Another issue related to tree-ring proxy data is that the process of constructing chronologies from raw measurements introduces a further level of uncertainty. An idea to overcome this problem is the direct use of raw data (Schofield et al., 2016) or the introduction of a further level, where the tree and instrumental observations are conditioned on the corresponding error-free latent process, as explained in Tingley et al. (2012).

Appendix A

Statistical background

Bayesian inference seems particularly suitable in the treating of palaeoclimate reconstruction problems. In Section A.1 we set the general ideas underlying this framework, following mainly Gelman et al. (2013) and Davison (2003). In Sections A.2, A.3 and A.4 we outline the algorithms employed in the estimation and reconstruction procedure.

A.1 Bayesian inference

The context of the scientific problem under study suggests what data are needed and what kind of probability models can be employed in the statistical analysis.

If we can express our beliefs about a set of parameters θ using the prior probability density $p(\theta)$, and we suppose to have data y following the probability model $p(y|\theta)$, we can update our knowledge about θ by using newly available data y, through Bayes' theorem:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)\,\mathrm{d}\theta}. \tag{A.1}$$

We can note that $p(y|\theta)$ can be interpreted as the likelihood function, if the focus is placed on θ, so that in terms of θ we have

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

It is useful to note that any quantity that does not depend on θ cancels from the denominator and numerator of (A.1). This implies that if we can recognise which density is proportional to (A.1), regarded solely as a function of θ, we can read off the posterior density of θ.

The marginal distribution of y is the denominator of (A.1) and does not depend on θ. Gelman et al. (2013) name it the *prior predictive distribution*, because it is not conditional on previous observations and it is the distribution of an observable quantity. In addition to the posterior distribution $p(\theta|y)$, in Bayesian analysis there is also interest in the distribution of an unknown observable variable from the same process, $\tilde{y}$. After data y have been observed, the posterior

predictive distribution of $\tilde{y}$ given y is

$$
\begin{aligned}
p(\tilde{y}|y) &= p(\tilde{y}, y)/p(y) \\
&= \frac{\int p(\tilde{y}|y, \theta)p(\theta)p(y|\theta)\,\mathrm{d}\theta}{\int p(\theta)p(y|\theta)\,\mathrm{d}\theta} \\
&= \int \frac{p(\tilde{y}|y, \theta)p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)\,\mathrm{d}\theta}\,\mathrm{d}\theta \\
&= \int p(\tilde{y}|y, \theta)p(\theta|y)\,\mathrm{d}\theta \\
&= \int p(\tilde{y}|\theta)p(\theta|y)\,\mathrm{d}\theta,
\end{aligned}
$$

where the last step follows by assuming conditional independence of $\tilde{y}$ and y given θ.

One of the central issues in Bayesian analysis is the choice of the prior distribution. Conjugate densities are particularly useful because of their simple closed forms. Often the class of conjugate priors is too restrictive with respect to the needs of the analysis, but since mixtures of conjugate densities are conjugate, the number of possibilities increases. The Bayesian approach incorporates in the analysis the information about a parameter even when the information is not precise, which leads to a non-informative or vague prior.

Once the posterior density is available, it can be used to calculate the probability of any event of interest. But some summary quantities may be useful. For example, if $\theta = (\psi, \lambda)$ is a vector, and we are interested in ψ, the marginal posterior density

$$
p(\psi|y) = \int p(\psi, \lambda|y)\,\mathrm{d}\lambda,
$$

contains the marginal information in the model and prior concerning ψ. It can be further summarized to moments, quantiles, or the mode of $p(\psi|y)$.

The mean and mode of the posterior density are point summaries of $p(\theta|y)$, but confidence regions or intervals are often required. The Bayesian analogue of a $100(1-\alpha)\%$ confidence interval is a $100(1-\alpha)\%$ credible set, defined to be a set, C, of values of θ, whose posterior probability content is at least $1 - \alpha$. When θ is continuous this is

$$
1 - \alpha = p(\theta \in C|y) = \int_C p(\theta|y)\,\mathrm{d}\theta.
$$

When θ is discrete, the integral is replaced by $\sum_{\theta \in C} p(\theta|y)$. For scalar θ, such a set is equi-tailed if it has form (θ_L, θ_U), where θ_L and θ_U are the posterior $\alpha/2$ and $1 - \alpha/2$ quantiles of θ, that is, $p(\theta < \theta_L|y) = p(\theta > \theta_U|y) = \alpha/2$. Often C is chosen so that the posterior density for any θ in C is higher than for any θ not in C. Such a region is called a highest posterior density (HPD) credible set.

A credible set may contain the same values of θ as a confidence interval, but its interpretation is different. In the Bayesian framework the data are regarded as fixed and the parameter as random, so the endpoints of the credible set are fixed and the probability statement concerns the parameter, regarded as a random variable. The frequentist approach treats the parameter as an unknown constant and the confidence interval endpoints as random variables; the probability statement concerns their behaviour in repeated sampling from the model.

The goal of Bayesian data analysis is posterior inference for quantities of interest, and this involves integration over one or more of the parameters. Usually the integrals cannot be obtained in closed form and numerical approximations must be used. Deterministic integration procedures

can sometimes be applied, but they are typically useful only for low-dimensional integrals. The most powerful tool for approximate calculation of posterior densities is numerical integration by Monte Carlo simulation.

A.2   Gibbs sampler and Metropolis-Hastings algorithms

Monte Carlo simulation draws values from the required distribution and approximates expectations through sample averages. When the values are independent, laws of large numbers ensure that the approximation can be made as accurate as desired by increasing the sample size. Sometimes drawing values independently is not feasible, and it can be advisable to use a sample of dependent values generated from a Markov chain. The idea of Markov chain Monte Carlo simulation is to construct a Markov chain that will, if run for an infinitely long period, generate samples from the required distribution. A Markov chain is a special kind of stochastic process, that is a set of indexed random variables,

$$\left\{ X^{(t)} \right\}, \; t = 0, 1, 2, \dots,$$

who has the Markov property that the next state depends only on the current state and not on the history of the chain:

$$p(X^{(t+1)} \in A | x^{(0)}, \dots, x^{(t)}) = p(X^{(t+1)} \in A | x^{(t)}).$$

The distribution $p(\cdot|\cdot)$ is called the transition kernel of the chain. If it does not depend on t, the chain is said to be time-homogeneous. Subject to regularity conditions, the chain will eventually converge to a unique stationary or invariant distribution. A Markov chain Monte Carlo (MCMC) method for the simulation of a distribution f is any method producing an ergodic Markov chain $(X^{(t)})$ whose stationary distribution is f. The ergodic theorem is the analogous of the law of large numbers for Markov chains, and it guarantees the almost sure convergence of the empirical average

$$\mathcal{I}_T = \frac{1}{T} \sum_{t=1}^{T} h(X^{(t)})$$

to the quantity $E_f(h(X))$, so that a sequence $(X^{(t)})$ produced by a MCMC algorithm can be used like an iid sample (Gilks et al., 1996).

A particular Markov chain algorithm that is considered useful in many problems is the Gibbs sampler, that we outline following Gamerman and Lopes (2006). Let $\theta = (\theta_1, \dots, \theta_k)$ be a parameter of dimension k whose joint density $p(\theta|y)$ is unknown. Although $p(\theta|y)$ itself is unknown, we suppose that we can simulate observations from the full conditional densities $p(\theta_i|\theta_{-i}, y)$, where $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$. Often in practice the constant normalizing $p(\theta|y)$ is unknown, but as it does not appear in the $p(\theta_i|\theta_{-i}, y)$, this causes no difficulty. If $p(\theta|y)$ is proper, then the Hammersley-Clifford theorem (see, for example, Davison (2003)) implies that under mild conditions $p(\theta|y)$ is determined by these densities; this does not imply that any set of full conditional densities determines a proper joint density. Gibbs sampling is successive simulation from the $p(\theta_i|\theta_{-i}, y)$ according to the algorithm:

1. initialize by taking arbitrary values of $\theta_1^{(0)}, \dots, \theta_k^{(0)}$.

2. Then for $i = 1, \dots, I$,

   (a) generate $\theta_1^{(i)}$ from $p(\theta_1|\theta_2 = \theta_2^{(i-1)}), \dots, \theta_k = \theta_k^{(i-1)}, y)$,

(b) generate $\theta_2^{(i)}$ from $p(\theta_2|\theta_1 = \theta_1^{(i)}, \theta_3 = \theta_3^{(i-1)}, \ldots, \theta_k = \theta_k^{(i-1)}, y)$,

(c) generate $\theta_3^{(i)}$ from $p(\theta_3|\theta_1 = \theta_1^{(i)}, \theta_2 = \theta_2^{(i)}, \theta_4 = \theta_4^{(i-1)}, \ldots, \theta_k = \theta_k^{(i-1)}, y)$,

$\vdots$

(d) generate $\theta_k^{(i)}$ from $p(\theta_k|\theta_1 = \theta_1^{(i)}, \ldots, \theta_{k-1} = \theta_{k-1}^{(i)}, y)$.

Here we update each of the θ_j in turn, conditional on the value of all the others. This gives a stream of random variables $\theta_1^{(1)}, \ldots, \theta_k^{(1)}, \ldots, \theta_1^{(I)}, \ldots, \theta_k^{(I)}$, so for the jth component of θ we have a sequence $\theta_j^{(1)}, \ldots, \theta_j^{(I)}$ and the vector $(\theta_1^{(I)}, \ldots, \theta_k^{(I)})$ is approximately a sample from $p(\theta)$.

When the full conditional distributions have a non-standard form, the Gibbs sampler cannot be used. The Metropolis-Hastings algorithm is a useful alternative in this case, and it is based on proposing values sampled from an instrumental distribution, which are then accepted with a certain probability that reflects how likely it is that they are from the target distribution. We start considering a particular case named Metropolis algorithm, that will be employed in the data analysis. It works as follows (Hoff, 2009):

1. Set a starting point $\theta^{(0)}$.

2. For $i = 1, \ldots, I$:

   (a) sample a value θ^* from a proposal distribution $J(\theta^*|\theta^{(i-1)})$, that must be symmetric, i.e. satisfying the condition $J(\theta_a|\theta_b) = J(\theta_b|\theta_a)$. In the application we will use $J(\theta^*|\theta^{(i-1)}) = N(\theta^{(i-1)}, \delta^2)$, tuning δ^2 in such a way that proposed values goes a reasonable distance in the parameter space, but are not rejected too frequently.

   (b) Compute the acceptance ratio

   $$r = \frac{p(\theta^*|y)}{p(\theta^{(i-1)}|y)} = \frac{p(\theta^*)p(y|\theta^*)}{p(\theta^{(i-1)})p(y|\theta^{(i-1)})}.$$

   (c) Let

   $$\theta^{(i)} = \begin{cases} \theta^* \text{ with probability } \min(r, 1), \\ \theta^{(t-1)}, \text{otherwise.} \end{cases}$$

The Metropolis-Hastings algorithm generalizes the Metropolis algorithm allowing the proposal distribution to be not symmetric, and replacing the ratio r with

$$r = \frac{p(\theta^*|y)/J(\theta^*|\theta^{(i-1)})}{p(\theta^{(i-1)}|y)/J(\theta^{(i-1)}|\theta^*)},$$

to correct for the asymmetry in the proposal.

The Gibbs sampler can be seen as a particular case of the Metropolis-Hastings algorithm, where the $i-$th step is divided in k sub-steps, in each of which the full conditional distribution is used as proposal density and the acceptance ratio is always 1.

It can be shown (Gelman et al., 2013) that the values generated by the Metropolis-Hastings algorithm are a sample from a Markov chain that converges to the stationary distribution of interest, so they can be used to make inference on it.

If some of the conditional posterior distributions can be sampled from directly and some cannot, then the parameters can be updated one at a time, with the Gibbs sampler used where possible and Metropolis otherwise.

One of the main issues is to know how long the algorithm should be run to get convergence, because if the iterations have run not enough, simulated values may be unrepresentative of the target distribution. In practice there are some convergence diagnostics that can at best detect non-convergence. To diminish the influence of the starting values, generally the first bit of the simulated values is discarded, the so-called burn-in or warm-up period. To lower the correlation among the sampled values, it can be useful to thin the sequence by keeping every $k-$th simulation draw, and discarding the rest.

A first graphic tool is the plot of the sample path, that should fluctuate rapidly without showing particular patterns. It can reveal if the chain has not reached the stationary distribution and how the support of the target distribution has been explored, even if it is not possible to know if the entire support has been explored. To get further ideas about the distribution of interest, one can examine histograms, box-plot or quantile-plot of the sample values. Another tool is the cumulative averages plot, to check if the ergodic mean can give a good approximation of the mean of the target distribution.

A.3   State-space models and the Kalman filter

The state space form is very widely used to handling a broad range of time series models (Harvey, 1990).

The general state space form (SSF) applies to a multivariate time series, $y(t)$, containing N elements. These observable variables are related to an $m \times 1$ vector, $x(t)$, known as the *state vector*, via a *measurement* or *observation* equation

$$y(t) = \beta(t)x(t) + d(t) + \varepsilon(t), \ t = 1, \ldots, n,$$

where $\beta(t)$ is an $N \times m$ matrix, $d(t)$ is an $N \times 1$ vector and $\varepsilon(t)$ is an $N \times 1$ vector of serially uncorrected disturbances with mean zero and covariance matrix $H(t)$, that is

$$E(\varepsilon(t)) = 0, \quad \text{Var}(\varepsilon(t)) = H(t).$$

In general the elements of $x(t)$ are not observable. However, they are known to be generated by a first-order Markov process, expressed through the *transition* or *state* equation

$$x(t) = \Phi(t)x(t-1) + c(t) + R(t)\eta(t), \ t = 1, \ldots, n,$$

where $\Phi(t)$ is an $m \times m$ matrix, $c(t)$ is an $m \times 1$ vector, $R(t)$ is an $m \times g$ matrix and $\eta(t)$ is a $g \times 1$ vector of serially uncorrected disturbances with mean zero and covariance matrix, $\Sigma(t)$, that is

$$E(\eta(t)) = 0, \ \text{Var}(\eta(t)) = \Sigma(t).$$

The specification of the state space system is completed by further assuming that the initial state vector, $x(0)$, has a mean of a_0 and a variance-covariance matrix P_0, that is

$$E(x(0)) = a_0, \ \text{Var}(x(0)) = P_0,$$

and that the disturbances $\varepsilon(t)$ and $\eta(t)$ are uncorrelated with each other in all time periods, and uncorrelated with the initial state, that is

$$E(\varepsilon(t)\eta(s)^\top) = 0 \ \forall s, t = 1, \ldots, n$$

and

$$E(\varepsilon(t)X(0)^{\top}) = 0, \ E(\eta(t)X(0)^{\top}) = 0 \text{ for } t = 1, \ldots, n.$$

The matrices $\beta(t)$, $d(t)$ and $H(t)$ in the measurement equation and the matrices $\Phi(t)$, $c(t)$, $R(t)$ and $\Sigma(t)$ in the transition equation will be referred to as the system matrices and in our exposition they are non-stochastic. As a result the system is linear and for any value of t, $y(t)$ can be expressed as a linear combination of present and past $\varepsilon(t)$'s and $\eta(t)$'s and the initial state vector, $x(0)$. If the system matrices $\beta(t)$, $d(t)$, $H(t)$, $\Phi(t)$, $c(t)$, $R(t)$ and $\Sigma(t)$ do not change over time, the model is said to be time-invariant or time-homogeneous, and this will be the case in our work.

Once a model has been put in a state space form, the Kalman filter can be applied for computing the optimal estimator of the state vector at time t, based on the information available at time t. This information consists of the observations up to and including $y(t)$. The system matrices together with a_0 and P_0 are assumed to be known in all time periods.

In certain engineering applications the Kalman filter is important because of on-line estimation. The current value of the state vector is of prime interest (for example, it may represent the co-ordinates of a rocket in space) and the Kalman filter enables the estimate of the state vector to be continually updated as new observations become available. At first sight, the value of such a procedure in environmental applications would appear to be limited: in our case the emphasis is on estimating the state vector in the past, on the basis of a given sample, using all the information available. This problem is known as *smoothing* and a solution can be based on Kalman filter. The derivation of the Kalman filter given below rests on the assumption that the disturbances and initial state vector are normally distributed. A standard result on the multivariate normal distribution can then be used to show how it is possible to calculate recursively the distribution of $X(t)$, conditional on the information set at time t, for all t from 1 to n. These conditional distributions are themselves normal and hence are completely specified by their means and covariance matrices. It is these quantities which the Kalman filter computes. It can be shown that the mean of the conditional distribution of $x(t)$, is an optimal estimator of $x(t)$, in the sense that it minimises the mean square error (MSE). When the normality assumption is dropped, there is no longer any guarantee that the Kalman filter will give the conditional mean of the state vector. However, it is still an optimal estimator in the sense that it minimises the mean square error within the class of all linear estimators.

Under the normality assumption, the initial state vector, $x(0)$, has a multivariate normal distribution with mean a_0 and covariance matrix P_0. The disturbances $\eta(t)$ and $\varepsilon(t)$ also have multivariate normal distributions for $t = 1, \ldots, n$ and are distributed independently of each other and of $x(0)$. The state vector at $t = 1$ is given by

$$x(1) = \Phi(1)x(0) + c(1) + R(1)\eta(1).$$

Thus $x(1)$ is a linear combination of two vectors of random variables, both with multivariate normal distributions, and a vector of constants. Hence it is itself multivariate normal with a mean of

$$a_{1|0} = \Phi(1)a_0 + c(1)$$

and a variance-covariance matrix

$$P_{1|0} = \Phi(1)P_0\Phi(1)^{\top} + R_1\Sigma(1)R_1^{\top}.$$

The notation $a_{1|0}$ indicates the mean of the distribution of $x(1)$ conditional on the information at time $t = 0$.

On its hand, $y(1)$ is multivariate Gaussian with mean

$$E(y(1)) = \beta(1)a_{1|0} + d(1),$$

and variance-covariance matrix

$$\text{Cov}(y(1)) = \beta(1)P_{1|0}\beta(1)^\top + H(1).$$

The covariance between $y(1)$ and $x(1)$ is:

$$\text{Cov}(y(1), x(1)) = \text{Cov}(\beta(1)x(1) + d(1) + \varepsilon(1), x(1))$$
$$= \beta(1)P_{1|0}.$$

From above it can be seen that the vector $(x(1)^\top y(1)^\top)^\top$ has a multivariate normal distribution with mean $(a_{1|0}^\top \ (\beta(1)a_{1|0} + d(1))^\top)^\top$ and variance-covariance matrix

$$\begin{pmatrix} P_{1|0} & P_{1|0}\beta(1)^\top \\ \beta(1)P_{1|0} & \beta(1)P_{1|0}\beta(1)^\top + H(1) \end{pmatrix}.$$

We now report a standard result about multivariate Gaussian conditional distribution (see for example Anderson (2003)) that will be useful here and later on:

Lemma A.3.1 (Multivariate Gaussian distribution). *Let X be distributed according to $N_p(\mu, V)$ with V nonsingular. Let us partition $X = (X_{(1)}^\top \ X_{(2)}^\top)^\top$ into q- and $(p-q)$-component subvectors, respectively. The density $f(x_{(1)}|x_{(2)})$ is a q-variate Gaussian density with mean*

$$E(X_{(1)}|x_{(2)}) = \mu_{(1)} + V_{12}V_{22}^{-1}(x_{(2)} - \mu_{(2)}),$$

and covariance matrix

$$\text{Cov}(X_{(1)}|x_{(2)}) = V_{11} - V_{12}V_{22}^{-1}V_{21}.$$

$\square$

Applying Lemma A.3.1 we have that the distribution of $x(1)$, conditional on a particular value of $y(1)$, is multivariate Gaussian with mean

$$a_1 = a_{1|0} + P_{1|0}\beta(1)^\top F_1^{-1}(y(1) - \beta(1)a_{1|0} - d_1)$$

and covariance matrix

$$P_1 = P_{1|0} - P_{1|0}\beta(1)^\top F_1^{-1}\beta(1)P_{1|0}$$

where

$$F_1 = \beta(1)P_{1|0}\beta(1)^\top + H(1)$$

Repeating the previous steps for $t = 2, \ldots, n$ yields a set of equations which are the Kalman filter recursions:

$$a_{t|t-1} = \Phi(t)a_{t-1} + c(t)$$
$$P_{t|t-1} = \Phi(t)P_{t-1}\Phi(t)^\top + R(t)\Sigma(t)R(t)^\top;$$
$$v_t = y(t) - \beta(t)a_{t|t-1} - d(t)$$
$$F_t = \beta(t)P_{t|t-1}\beta(t)^\top + H(t)$$
$$a_t = a_{t|t-1} + P_{t|t-1}\beta(t)^\top F_t^{-1}v_t$$
$$P_t = P_{t|t-1} - P_{t|t-1}\beta(t)^\top F_t^{-1}\beta(t)P_{t|t-1}.$$

Given a_{t-1} and P_{t-1}, the optimal estimator of $x(t)$ is given by

$$a_{t|t-1} = \Phi(t)a_{t-1} + c(t),$$

while the covariance matrix of the estimation error is

$$P_{t|t-1} = \Phi(t)P_{t-1}\Phi(t)^\top + R(t)\Sigma(t)R(t)^\top, t = 1, \ldots, n.$$

These two equations are known as the prediction equations. Once the new observation, $y(t)$ becomes available, the estimator of $x(t)$, $a_{t|t-1}$, can be updated. The updating equations are

$$a_t = a_{t|t-1} + P_{t|t-1}\beta(t)^\top F_t^{-1}(y(t) - \beta(t)a_{t|t-1} - d(t)$$

and

$$P_t = P_{t|t-1} - P_{t|t-1}\beta(t)^\top F_t^{-1}\beta(t)P_{t|t-1}$$

where

$$F_t = \beta(t)P_{t|t-1}\beta(t)^\top + H(t), \ t = 1, \ldots, n,$$

where it is assumed that F_t is invertible, which is the case for the models in this work.

The derivation given previously enables us to interpret a_t and P_t as the mean and covariance matrix of the conditional distribution of $x(t)$. In setting out the Kalman filter a_t was described as an optimal estimator of $x(t)$ based on the information available at time t, while P_t was described as the covariance matrix of the estimation error.

A.4   Gibbs sampling for state-space models

In this Section we outline an algorithm that will be a fundamental part of the climate reconstruction procedure.

Let us consider the time-homogeneous linear state-space model

$$y(t) = \beta_0 + \beta^\top x(t) + \varepsilon(t), \ \varepsilon(t) \sim N(0, \psi^2); \tag{A.2}$$

$$x(t + 1) = \Phi x(t) + \eta(t + 1), \ \eta(t + 1) \sim N(0, \Sigma), \ t = 1, \ldots, n, \tag{A.3}$$

where $y(t)$ is an observed scalar variable at time t and $x(t)$ is the corresponding vector of dimension m of non-observable states. Let $Y^j = \{y(1), \ldots, y(j)\}$ the $(j \times 1)$ vector of observations from time 1 to time j and $X = \left\{x(1)^\top, \ldots, x(n)^\top\right\}^\top$ the $(nm \times 1)$ total state vector.

We are interested in doing inference on X using the whole information available, so we are interested in the distribution $p(X|Y^n)$. According to Carter and Kohn (1994) we can write:

$$
\begin{aligned}
p(X|Y^n) &= p(X, Y^n)/p(Y^n) \\
&= p(x(1), \ldots, x(n), Y^n)/p(Y^n) \\
&= p(x(1)|x(2), \ldots, x(n), Y^n)p(x(2), \ldots, x(n), Y^n)/p(Y^n) \\
&= p(x(1)|x(2), \ldots, x(n), Y^n)p(x(2), \ldots, x(n)|Y^n) \\
&= p(x(1)|x(2), y(1))p(x(2), \ldots, x(n)|Y^n),
\end{aligned}
$$

because information on $x(3), \ldots, x(n)$ or $y(2), \ldots, y(n)$ has no influence on the distribution of $x(1)$.

Again, we can write

$$
\begin{aligned}
p(x(2), \ldots, x(n)|Y^n) &= p(x(2), \ldots, x(n), Y^n)/p(Y^n) \\
&= p(x(2)|x(3), \ldots, x(n), Y^n)p(x(3), \ldots, x(n), Y^n)/p(Y^n) \\
&= p(x(2)|x(3), \ldots, x(n), Y^n)p(x(3), \ldots, x(n)|Y^n) \\
&= p(x(2)|x(3), Y^2)p(x(3), \ldots, x(n)|Y^n),
\end{aligned}
$$

because knowing $x(4), \ldots, x(n)$ or $y(3), \ldots, y(n)$ has no influence on the distribution of $x(2)$. We need $y(1)$ along with $y(2)$ because it contains information about $x(1)$, so indirectly about $x(2)$. Proceeding in the same way, we have

$$
p(X|Y^n) = p(x(n)|Y^n) \prod_{t=1}^{n-1} p(x(t)|x(t+1), Y^t). \tag{A.4}
$$

The first step is to generate a value for $x(n)$ from $P(x(n)|Y^n)$. We assume a Gaussian distribution for the initial state vector, $x(0)$: $P(x(0)) = N_m(a_0, P_0)$, where a_0 and P_0 are considered known. From the state equation (A.3) and using the notation of the previous Section we have:

$$
a_{1|0} = E(x(1)) = \Phi E(x(0)) + E(\eta(1)) = \Phi a_0
$$

$$
\begin{aligned}
P_{1|0} = \operatorname{Var}(x(1)) &= \Phi \operatorname{Var}(x(0))\Phi^\top + \operatorname{Var}(\eta(1)) + \operatorname{Cov}(x(0), \eta(1)) + \operatorname{Cov}(\eta(1), x(0)) \\
&= \Phi P_0 \Phi^\top + \Sigma.
\end{aligned}
$$

This distribution has to be updated conditionally on the observation $y(1)$. From the observation equation (A.2) and equations above, the distribution of $y(1)$ is Gaussian, with mean

$$
E(y(1)) = E(\beta_0) + E(\beta^\top x(1)) = \beta_0 + \beta^\top a_{1|0},
$$

and variance

$$
\operatorname{Var}(y(1)) = \operatorname{Var}(\beta_0 + \beta^\top x(1) + \varepsilon(1)) = \beta^\top P_{1|0}\beta + \psi^2.
$$

The covariance between $x(1)$ and $y(1)$ is:

$$
\begin{aligned}
\operatorname{Cov}(x(1), y(1)) &= \operatorname{Cov}(x(1), \beta_0 + \beta^\top x(1) + \varepsilon(1)) \\
&= \operatorname{Var}(x(1))\beta = P_{1|0}\beta.
\end{aligned}
$$

The joint distribution of $x(1)$ and $y(1)$ then is:

$$
(x(1)^\top, y(1))^\top \sim N\left((a_{1|0}, \beta_0 + \beta^\top a_{1|0}), \begin{pmatrix} P_{1|0} & P_{1|0}\beta \\ \beta^\top P_{1|0} & \beta^\top P_{1|0}\beta + \psi^2 \end{pmatrix} \right).
$$

The conditional distribution $x(1)|y(1)$ is Gaussian with mean a_1 and variance P_1 computed from Lemma A.3.1:

$$
a_1 = a_{1|0} + P_{1|0}\beta(\beta^\top P_{1|0}\beta + \psi^2)^{-1}(y(1) - \beta_0 - \beta^\top a_{1|0}),
$$

$$
P_1 = P_{1|0} - P_{1|0}\beta(\beta^\top P_{1|0}\beta + \psi^2)^{-1}\beta^\top P_{1|0}.
$$

In order to compute the distribution of $x(2)|Y^2$, we start from $P(x(2)|y(1))$, that is Gaussian with mean

$$a_{2|1} = E(x(2)|y(1)) = \Phi E(x(1)|y(1)) + E(\eta(2)|y(1)) = \Phi a_1,$$

and variance

$$P_{2|1} = \text{Var}(x(2)|y(1)) = \Phi \text{Var}(x(1)|y(1))\Phi^\top + \text{Var}(\eta(2)|y(1))$$
$$+ \text{Cov}(x(2)|y(1), \eta(2)|y(1)) + \text{Cov}(\eta(2)|y(1), x(2)|y(1))$$
$$= \Phi P_1 \Phi^\top + \Sigma.$$

Let us compute the marginal distribution of $y(2)|y(1)$:

$$E(y(2)|y(1)) = \beta_0 + \beta^\top E(x(2)|y(1)) = \beta_0 + \beta^\top a_{2|1}.$$

$$\text{Var}(y(2)|y(1)) = \beta^\top \text{Var}(x(2)|y(1))\beta + \psi^2 = \beta^\top P_{2|1}\beta + \psi^2.$$

The covariance between $x(2)|y(1)$ and $y(2)|y(1)$ is:

$$\text{Cov}(x(2)|y(1), y(2)|y(1)) = \text{Var}(x(2)|y(1))\beta = P_{2|1}\beta.$$

So we can write the joint distribution of $x(2)|y(1)$ and $y(2)|y(1)$ as:

$$(x(2)|y(1)^\top, y(2)|y(1))^\top \sim N\left((a_{2|1}, \beta_0 + \beta^\top a_{2|1}), \begin{pmatrix} P_{2|1} & P_{2|1}\beta \\ \beta^\top P_{2|1} & \beta^\top P_{2|1}\beta + \psi^2 \end{pmatrix}\right).$$

From Lemma A.3.1 the conditional distribution of $x(2)|Y^2$ is Gaussian with mean

$$a_2 = a_{2|1} + P_{2|1}\beta(\beta^\top P_{2|1}\beta + \psi^2)^{-1}(y(2) - \beta^\top a_{2|1})$$

and variance

$$P_2 = P_{2|1} - P_{2|1}\beta(\beta^\top P_{2|1}\beta + \psi^2)^{-1}\beta^\top P_{2|1}.$$

Summarizing, we start from a_0 and P_0, considered known, and we compute, for $t = 1, \ldots, n$:

$$a_{t|t-1} = \Phi a_{t-1},$$
$$P_{t|t-1} = \Phi P_{t-1}\Phi^\top + \Sigma,$$
$$v_t = y(t) - \beta_0 - \beta^\top a_{t|t-1},$$
$$F_t = \beta^\top P_{t|t-1}\beta + \psi^2,$$
$$a_t = a_{t|t-1} + P_{t|t-1}\beta F_t^{-1}v_t,$$
$$P_t = P_{t|t-1} - P_{t|t-1}\beta F_t^{-1}\beta^\top P_{t|t-1}.$$

At the end we get the distribution of $x(n)|Y^n$, that is $N(a_n, P_n)$.

Once we have generated a value from $x(n)|Y^n$, the next step is to generate from $P(x(t)|Y^t, x(t+1)$, for $t = n-1, \ldots, 1$. Let us start from $P(x(n-1)|Y^{n-1}, x(n))$. From the state equation we have

$$x(n) = \Phi x(n-1) + \eta(n), \ \eta(n) \sim N(0, \Sigma). \tag{A.5}$$

Since we work conditionally on $x(n)$, equation (A.5) can be considered as a set of m observation equations, which the Kalman filter can be applied to. In order to exploit the m observations, we apply the Cholesky decomposition to Σ, the covariance matrix of $\eta(n)$:

$$\Sigma = L\Delta L^\top,$$

and we compute

$$\tilde{x}(n) = L^{-1}x(n), \ \tilde{\Phi} = L^{-1}\Phi, \ \tilde{u}(n) = L^{-1}u(n).$$

Then, we can write

$$\tilde{x}(n) = \tilde{\Phi}x(n-1) + \tilde{u}(n).$$

Denoting with $\tilde{\Phi}_i$ the $i-$th row of $\tilde{\Phi}$ and Δ_i the $i-$th element on the diagonal of Δ, we have, for $i = 1, \ldots, m$, the 'observation' equations:

$$\tilde{x}_i(n) = \tilde{\Phi}_i^\top x(n-1) + \tilde{u}_i(n), \ \tilde{u}_i(n) \sim N(0, \Delta_i).$$

We use as initial distribution for the state $x(n-1)$, $p(x(n-1)|Y^{n-1})$, that from previous computation has mean a_{n-1} and covariance matrix P_{n-1}. It is convenient to set $x(n-1|n-1, 0) = a_{n-1}$ and $S(n-1|n-1, 0) = P_{n-1}$. For $i = 1, \ldots, m$, let

$$x(n-1|n-1, i) = E(x(n-1)|Y^{n-1}, x_1(n), \ldots, x_i(n)),$$
$$S(n-1|n-1, i) = \text{Var}(x(n-1)|Y^{n-1}, x_1(n), \ldots, x_i(n)).$$

First we update the initial distribution of $x(n-1)$ conditioning on the value of $x_1(n)$ or $\tilde{x}_1(n)$, if Σ is not diagonal. From equations above we have:

$$E(\tilde{x}_1(n)|Y^{n-1}) = \tilde{\Phi}_1^\top E(x(n-1)|Y^{n-1}) + E(\tilde{\eta}_1(n)) = \tilde{\Phi}_1^\top x(n-1|n-1, 0),$$

$$\text{Var}(\tilde{x}_1(n)|Y^{n-1}) = \tilde{\Phi}_1^\top \text{Var}(x(n-1)|Y^{n-1})\tilde{\Phi}_1 + \text{Var}(\tilde{\eta}_1(n)) + 2\,\text{Cov}(\tilde{\Phi}_1^\top x(n-1)|Y^{n-1}, \tilde{\eta}_1(n))$$
$$= \tilde{\Phi}_1^\top S(n-1|n-1, 0)\tilde{\Phi}_1 + \Delta_1,$$

$$\text{Cov}(x(n-1), \tilde{x}_1(n)|Y^{n-1}) = \text{Cov}(x(n-1), \tilde{\Phi}_1^\top x(n-1) + u_1(n))$$
$$= \text{Var}(\tilde{x}_1(n)|Y^{n-1})\tilde{\Phi}_1$$
$$= S(n-1|n-1, 0)\tilde{\Phi}_1.$$

Let us define

$$\epsilon(n-1, 1) = \tilde{x}_1(n) - E(\tilde{x}_1(n)|Y^{n-1}) = \tilde{x}_1(n) - \tilde{\Phi}_1^\top x(n-1|n-1, 0)$$

and

$$R(n-1, 1) = \text{Var}(\tilde{x}_1(n)|Y^{n-1}) = \tilde{\Phi}_1^\top S(n-1|n-1, 0)\tilde{\Phi}_1 + \Delta_1.$$

From Lemma A.3.1 the conditional distribution $P(x(n-1)|Y^{n-1}, \tilde{x}_1(n))$ is Gaussian with mean

$$x(n-1|n-1, 1) = x(n-1|n-1, 0) + S(n-1|n-1, 0)\tilde{\Phi}_1^\top \epsilon(n-1, 1)/R(n-1, 1),$$

and variance

$$S(n-1|n-1,1) = S(n-1|n-1,0) - S(n-1|n-1,0)\tilde{\Phi}_1\tilde{\Phi}_1^\top S(n-1|n-1,0)/R(n-1,1).$$

For $i = 2, \ldots, m$ we repeat the reasoning by updating the distribution of the state with $\tilde{x}_i(n)$. We compute:

$$\epsilon(n-1,i) = \tilde{x}_i(n) - \tilde{\Phi}_i^\top x(n-1|n-1,i-1),$$

$$R(n-1,i) = \tilde{\Phi}_i^\top S(n-1|n-1,i)\tilde{\Phi}_i + \Delta_i,$$

$$x(n-1|n-1,i) = x(n-1|n-1,i-1) + S(n-1|n-1,i-1)\tilde{\Phi}_i^\top \epsilon(n-1,i)/R(n-1,i),$$

$$S(n-1|n-1,i) = S(n-1|n-1,i-1) - S(n-1|n-1,i-1)\tilde{\Phi}_i\tilde{\Phi}_i^\top S(n-1|n-1,i-1)/R(n-1,i).$$

At the end we have obtained

$$x(n-1|n-1,m) = E(x(n-1)|Y^{n-1}, x(n)), \text{ and}$$
$$S(n-1|n-1,m) = \text{Var}(x(n-1)|Y^{n-1}, x(n)).$$

By repeating the procedure for t from $n-2$ to 1, we can extract values from $P(X|Y^n)$.

Appendix B

Graphics

This Appendix collects graphics mentioned in other parts of the thesis, to keep the text fluid. Figures B.1–B.7 refer to the simulation study in Section 3.3.2; figures 4.2– B.24 illustrate temperature and moisture data analysed in Section 4.1; figures 4.3–B.41 illustrate temperature and PDSI data analysed in Section 4.2; figures B.42 – B.62 illustrate model comparisons done in Section 4.3.
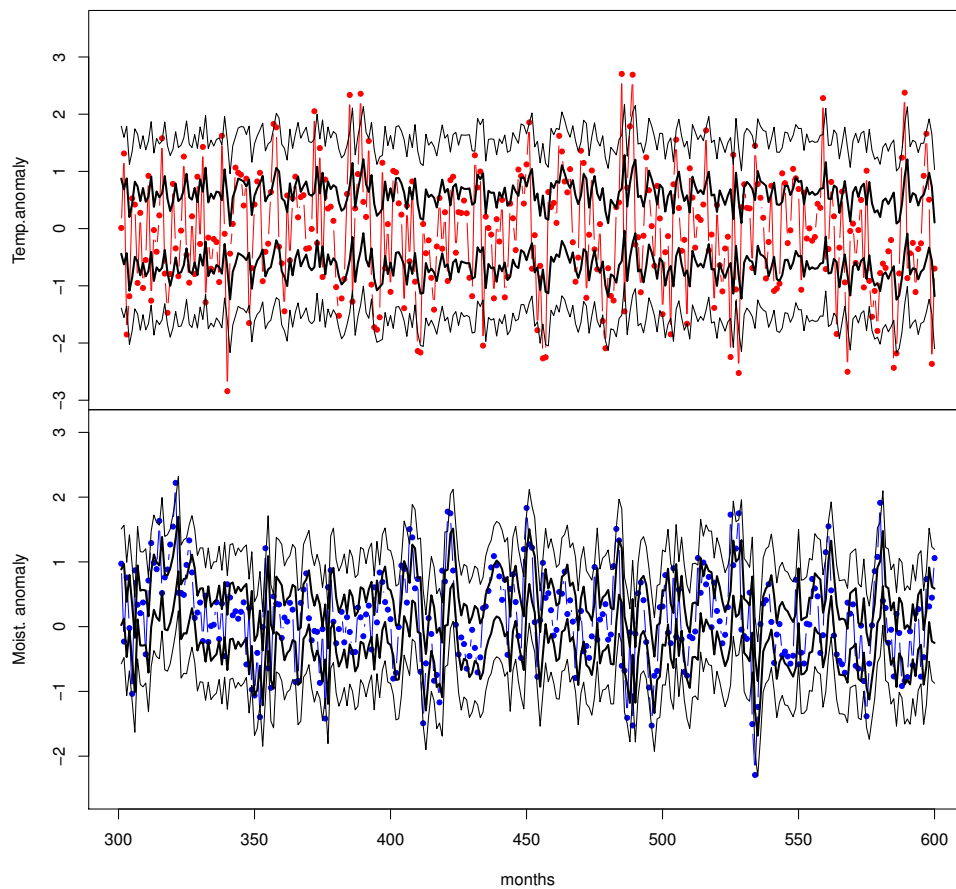
Figure B.1: Simulated series in the calibration period, months from 301 to 600. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
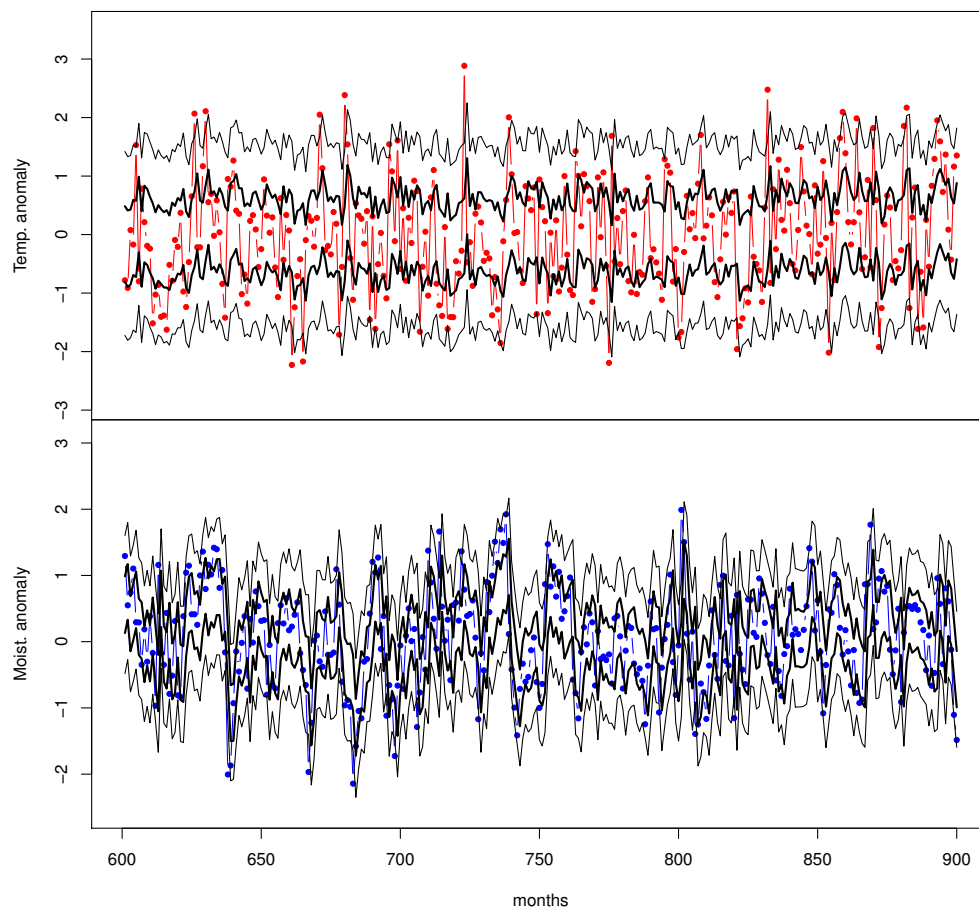
Figure B.2: Simulated series in the calibration period, months from 601 to 900. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
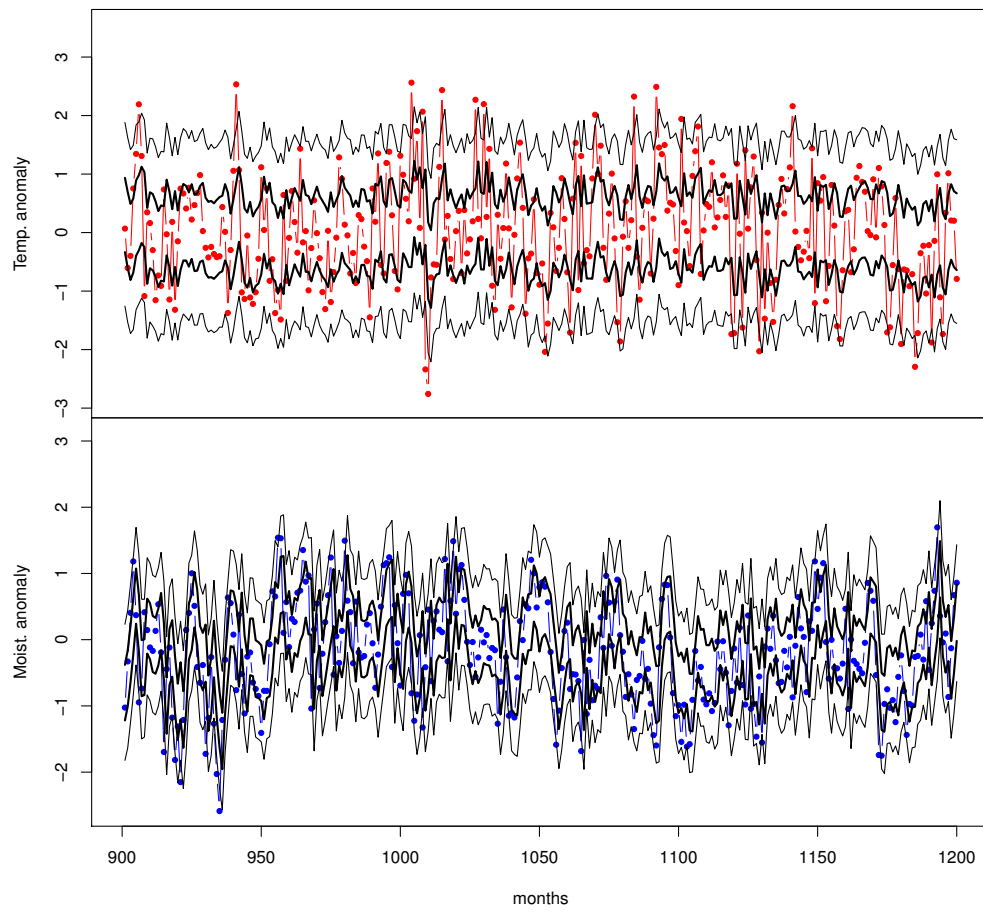
Figure B.3: Simulated series in the calibration period, months from 901 to 1200. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
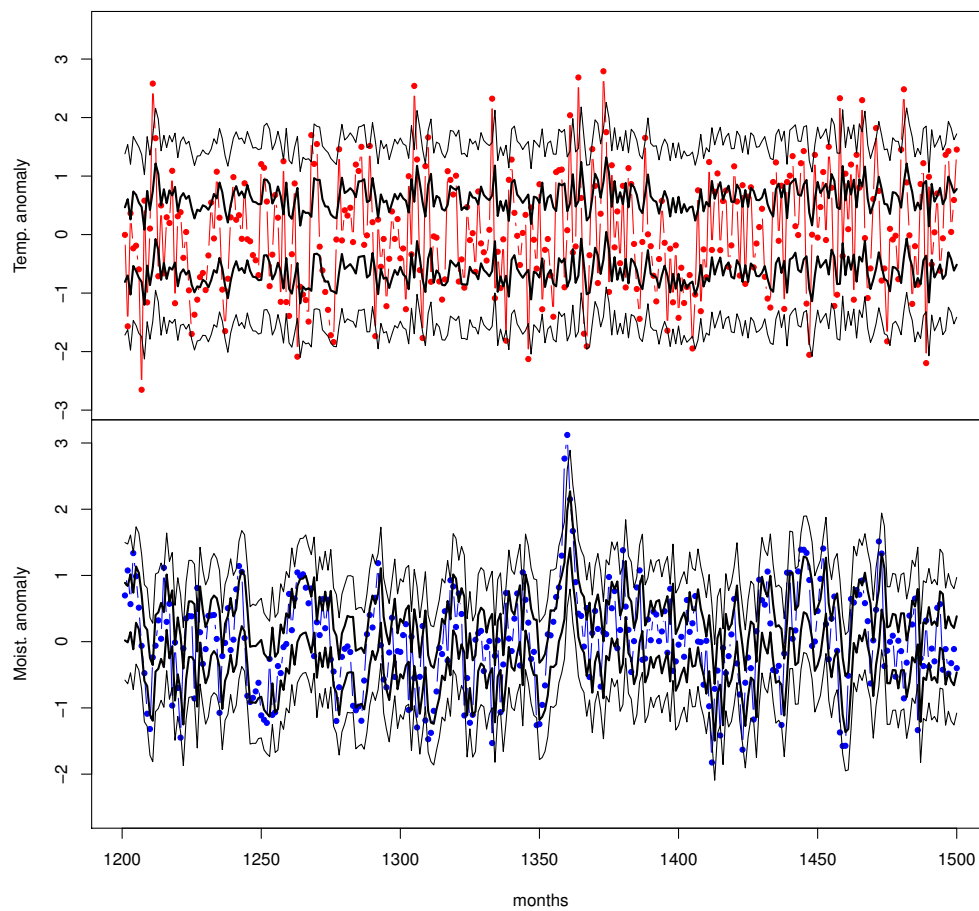
Figure B.4: Simulated series in the calibration period, months from 1201 to 1500. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
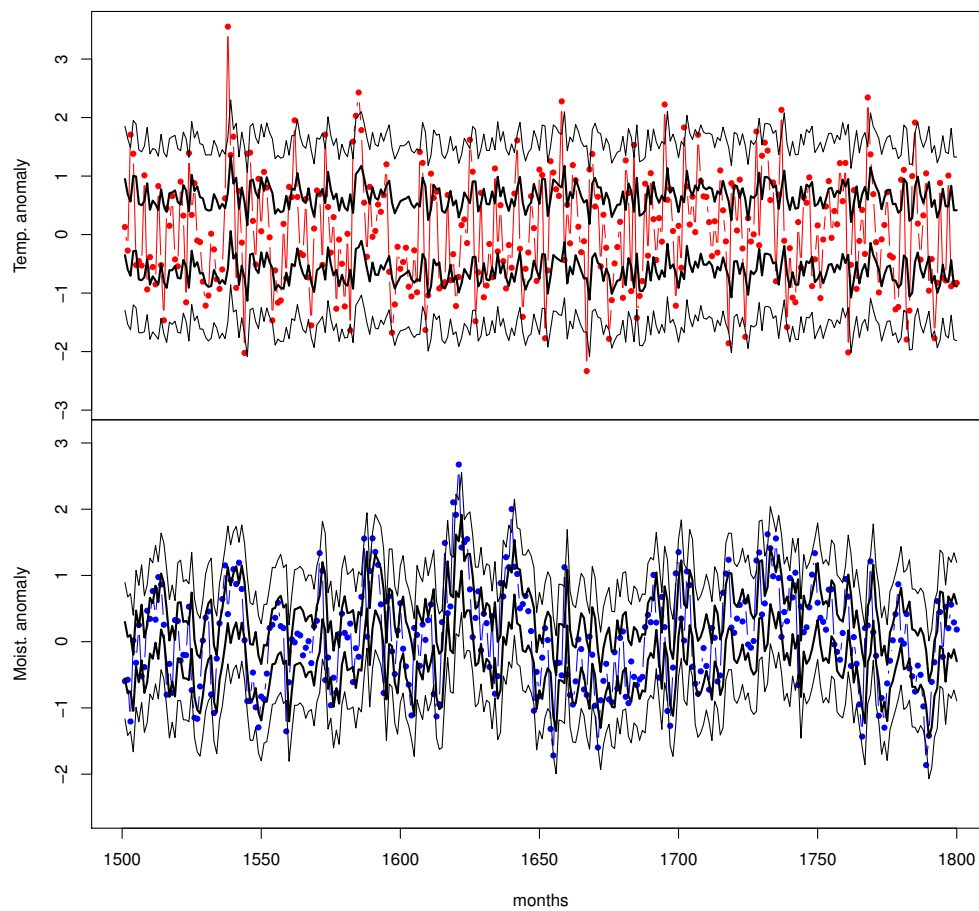
Figure B.5: Simulated series in the calibration period, months from 1501 to 1800. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
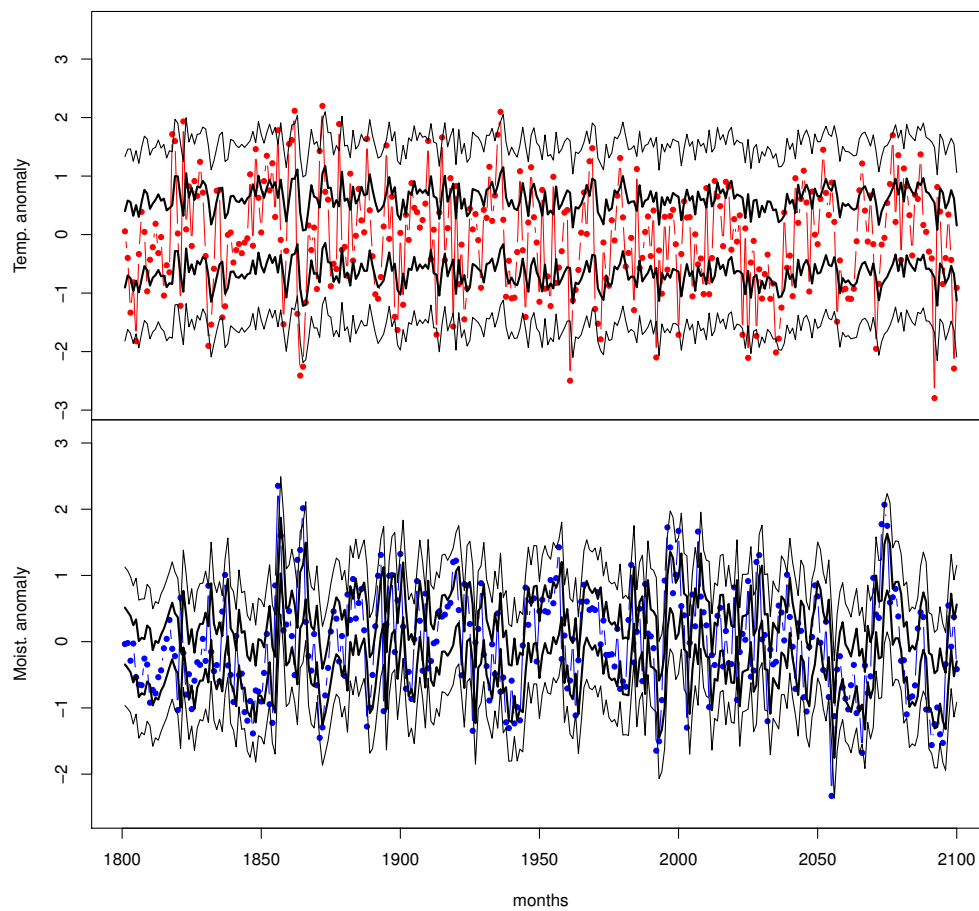
Figure B.6: Simulated series in the calibration period, months from 1801 to 2100. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
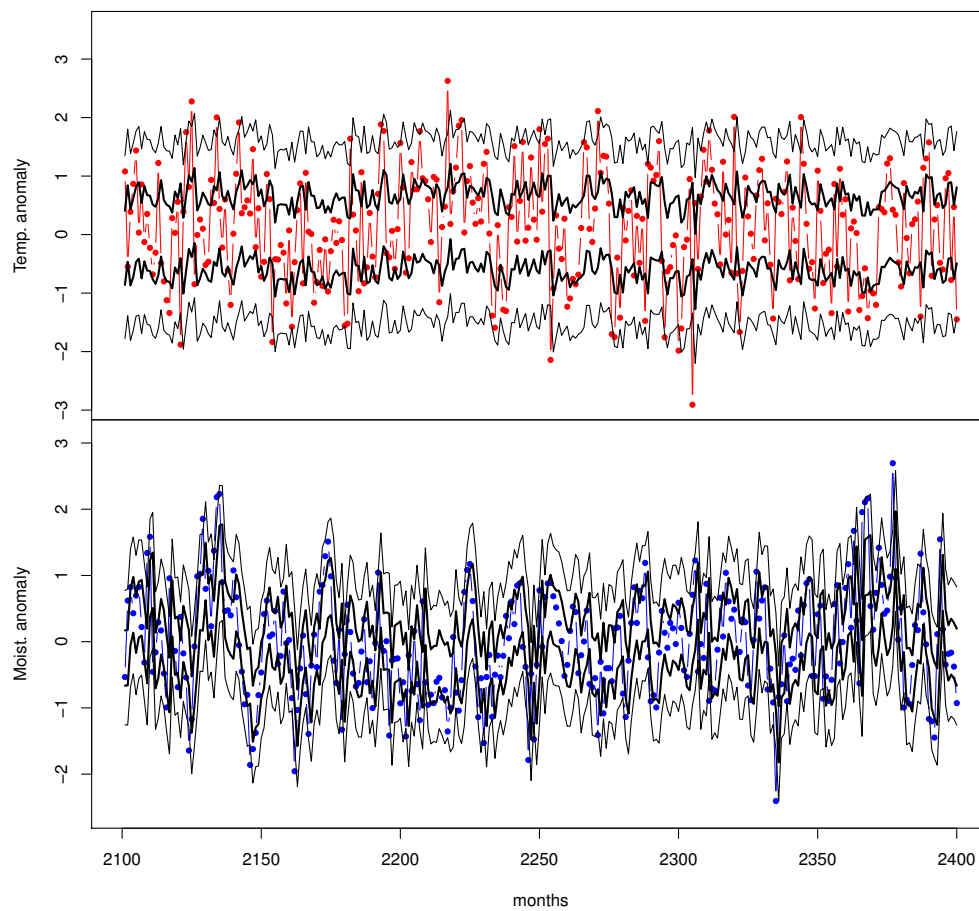
Figure B.7: Simulated series in the calibration period, months from 2101 to 2400. Colored lines are simulated anomalies (red for temperature and blue for moisture), black lines are 90% and 50% credible sets.
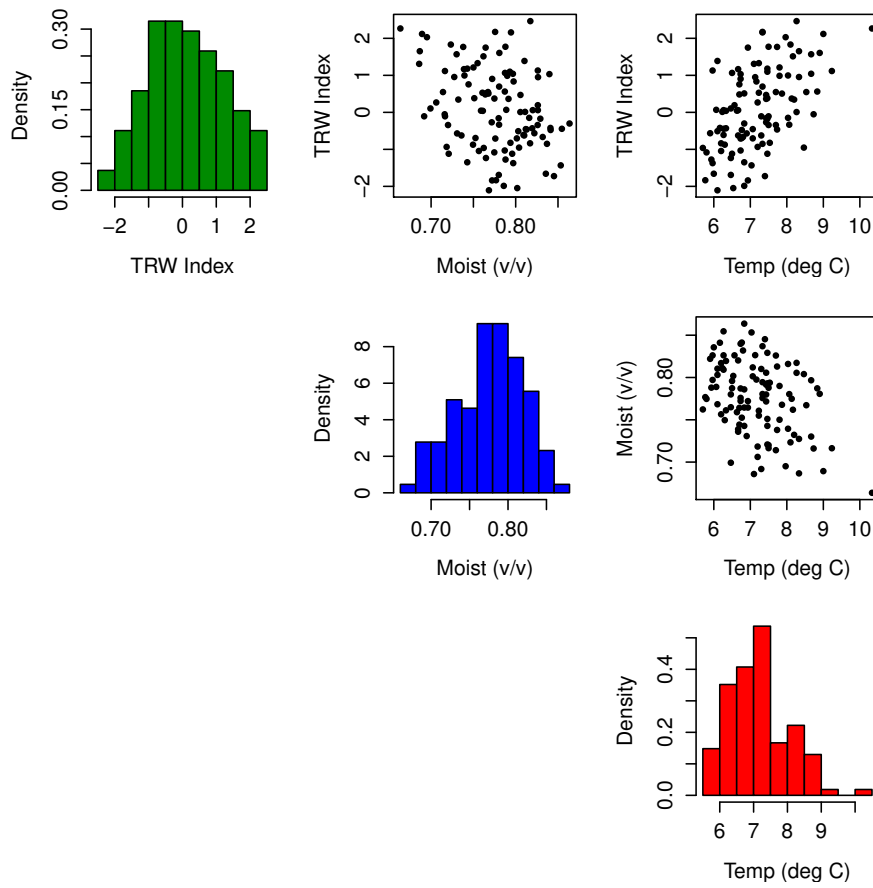
Figure B.8: Site A. Histograms and scatter-plots for tree-ring widths index, moisture and temperature in the period 1901–2008.
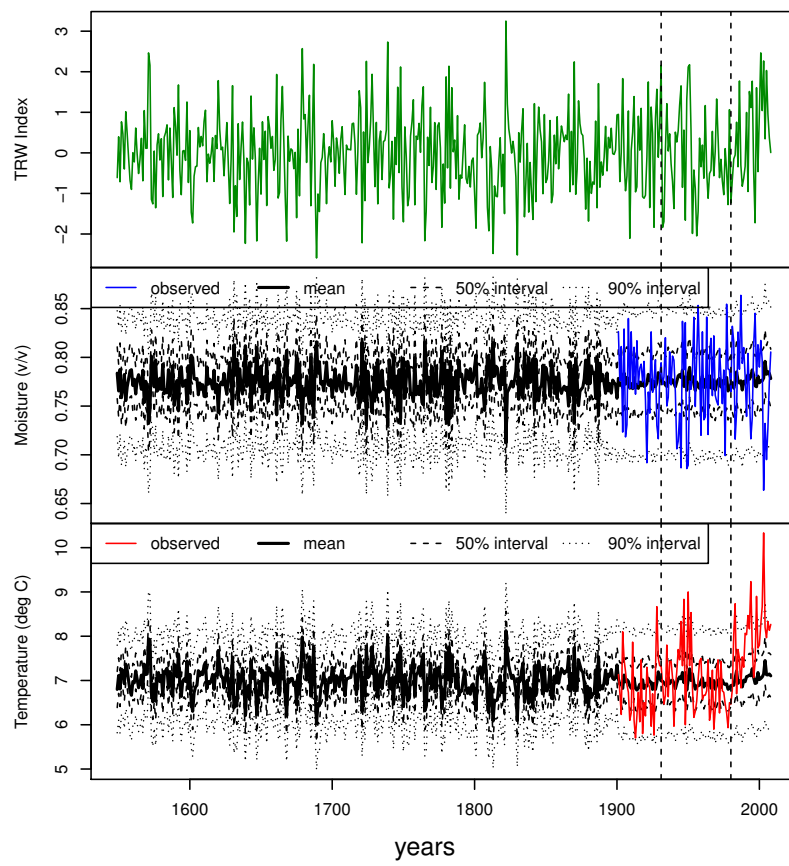
Figure B.9: Site A: estimated summer series (temperature/moisture model).
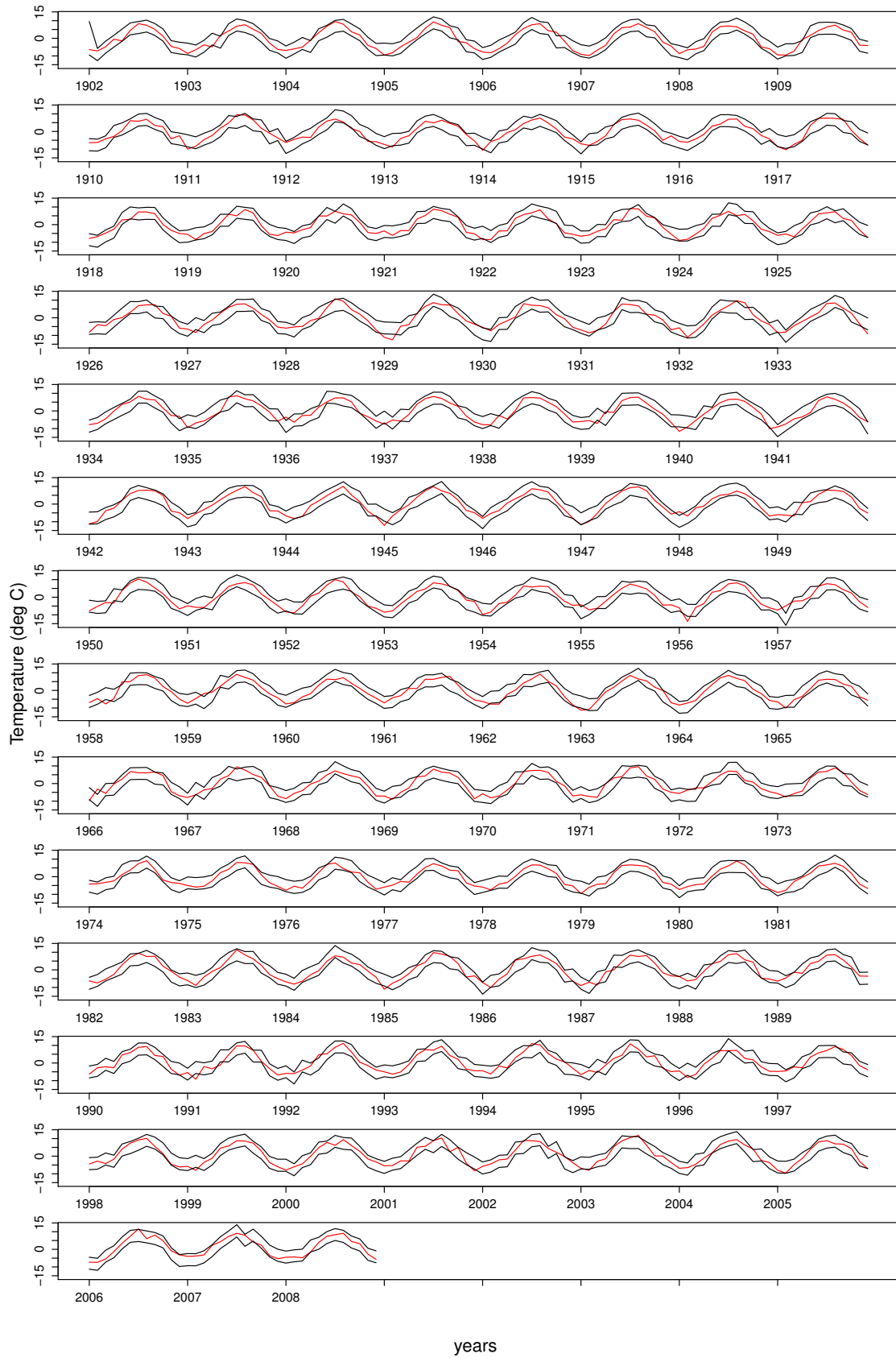
Figure B.10: Site A: monthly temperature series in 1902 – 2008. In red the observed series, in black 90% credibility interval (temperature/moisture model).
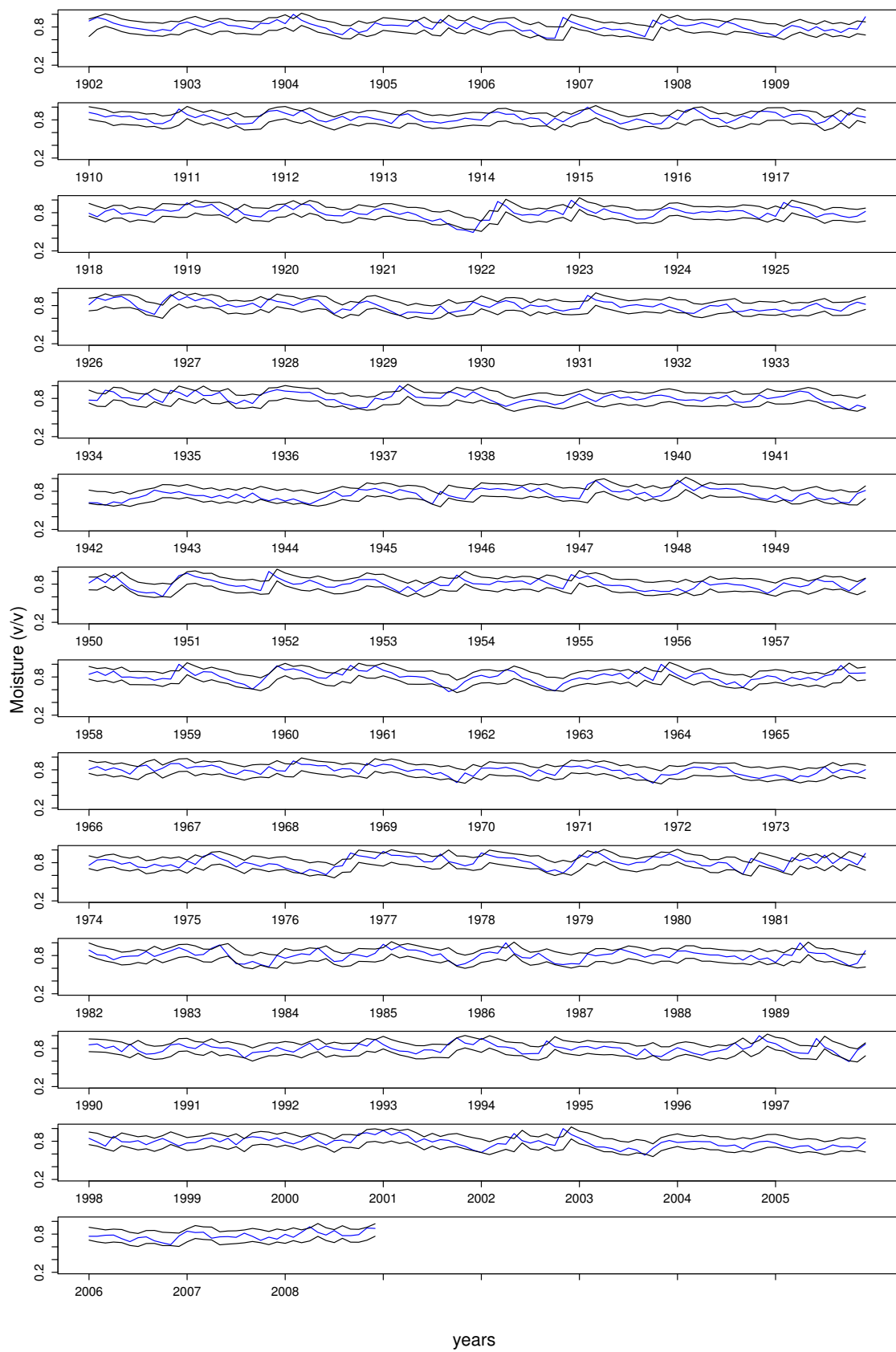
Figure B.11: Site A: monthly moisture series in 1902 – 2008. In blue the observed series, in black 90% credibility interval (temperature/moisture model).
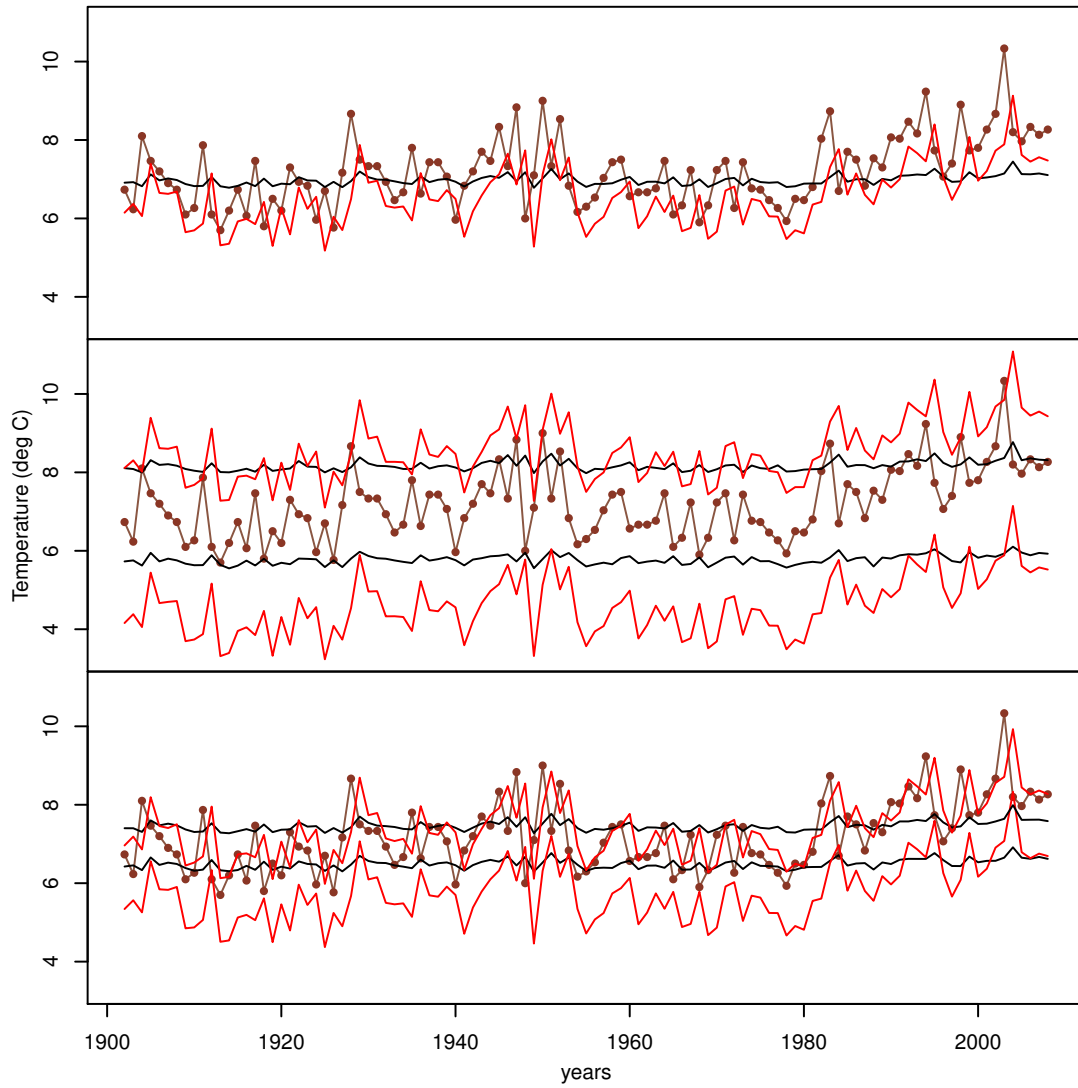
Figure B.12: Site A: comparison between temperature in the observation period: in red the result from the monthly algorithm, in black from the annual one. Brown series is the observed one. From top to bottom: mean, 90% interval, 50% interval (temperature/moisture model).
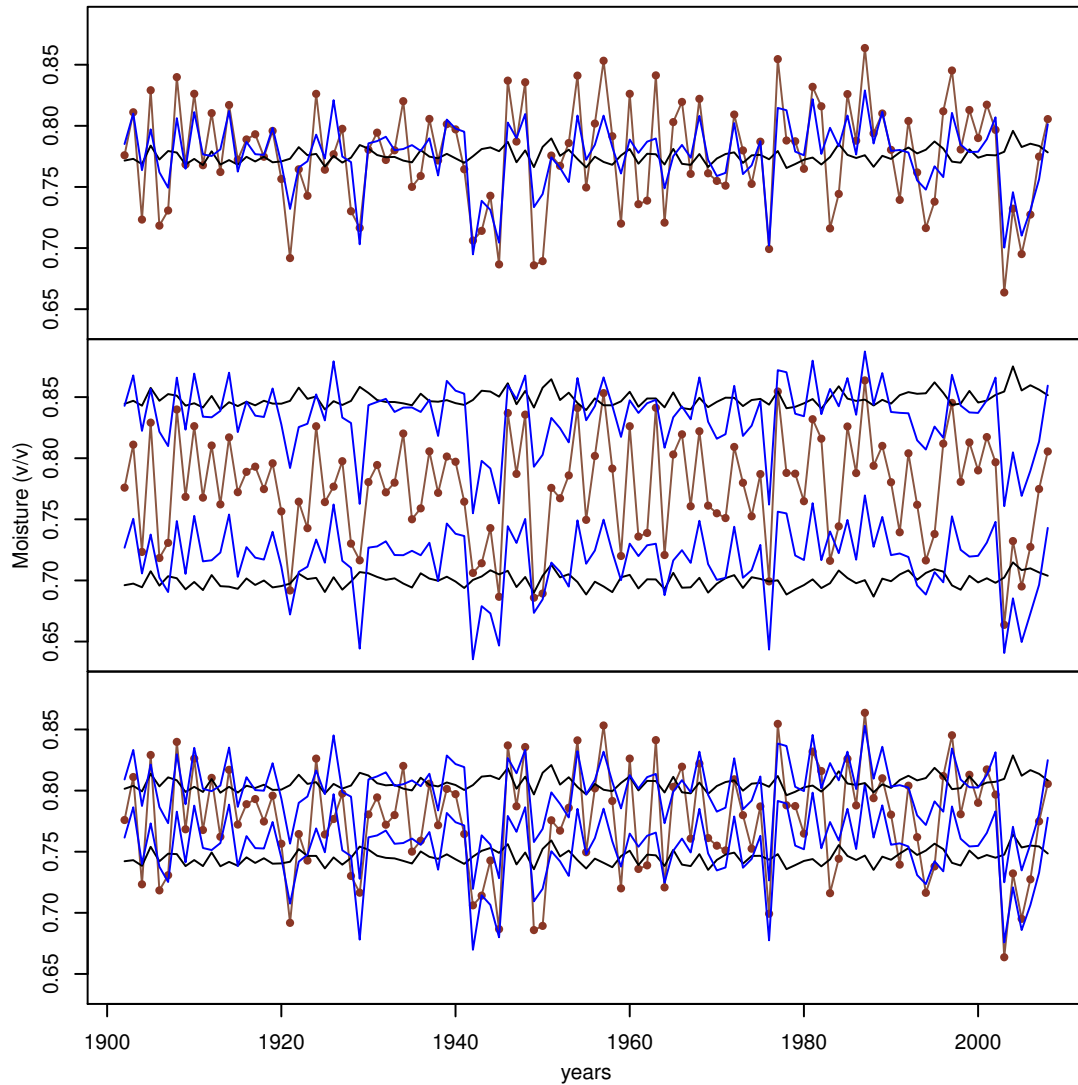
Figure B.13: Site A: comparison between moisture in the observation period: in blue the result from the monthly algorithm, in black from the annual one. Brown series is the observed one. From top to bottom: mean, 90% interval, 50% interval (temperature/moisture model).
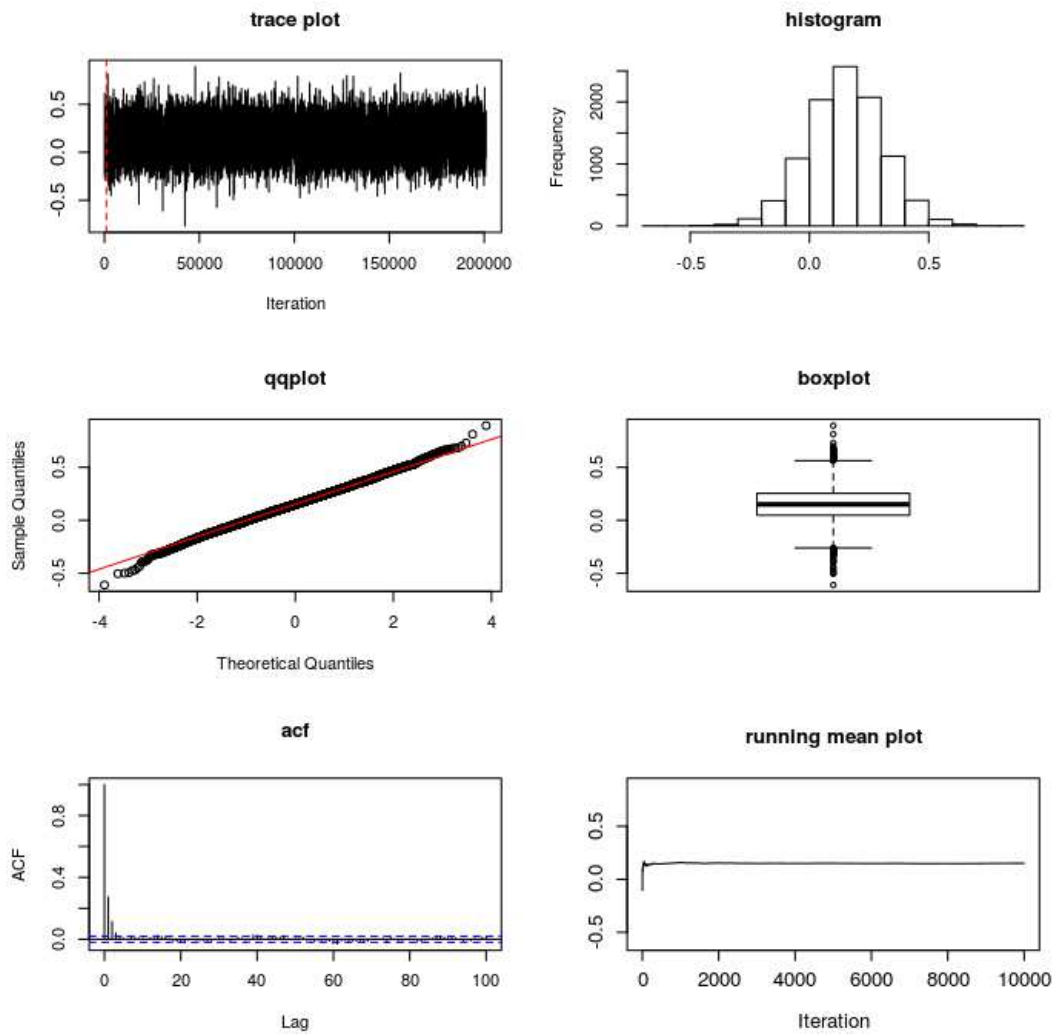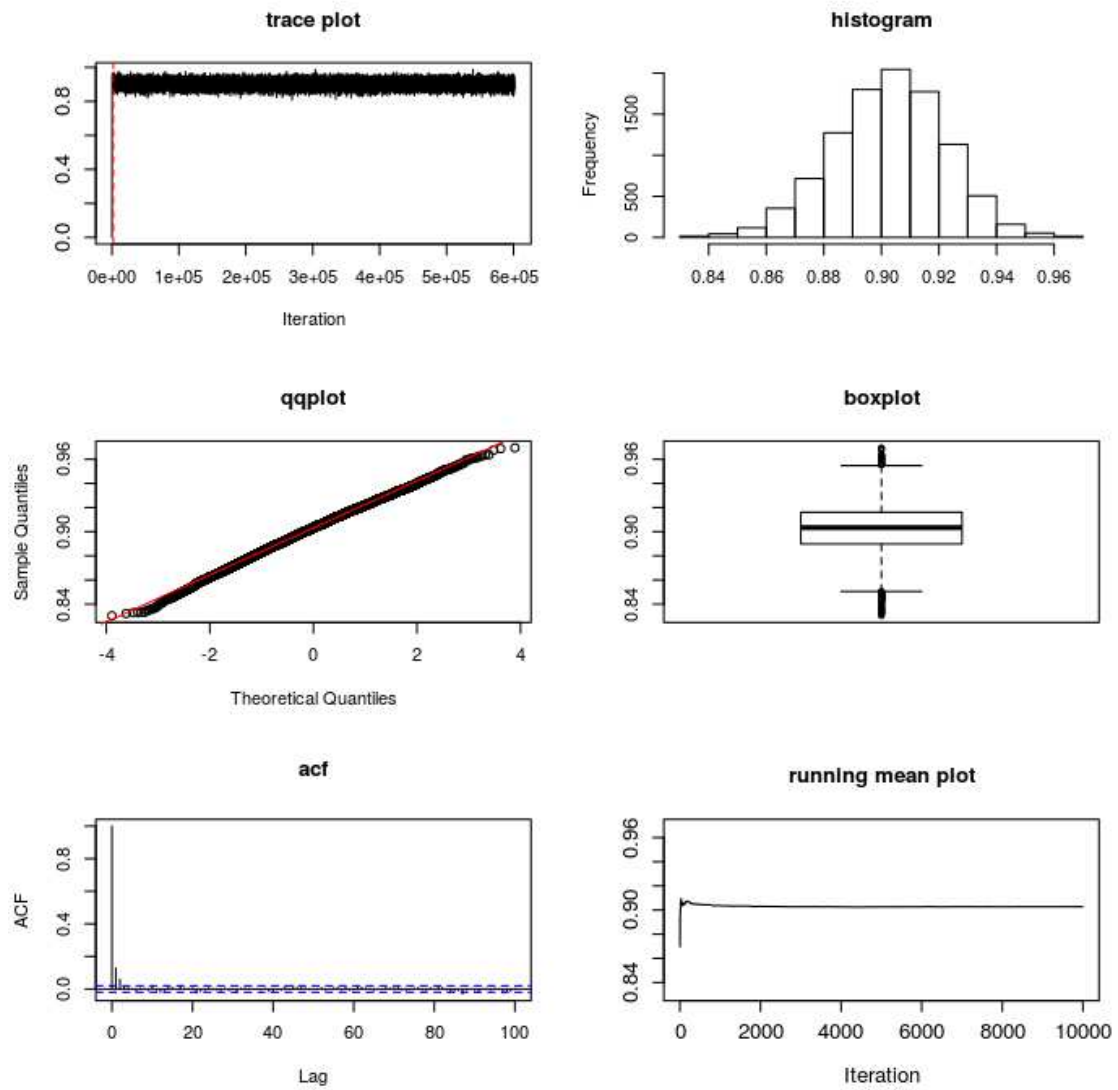
Figure B.14: Site A: diagnostic on the estimation of $\Phi(1,1)$ (temperature/moisture model).
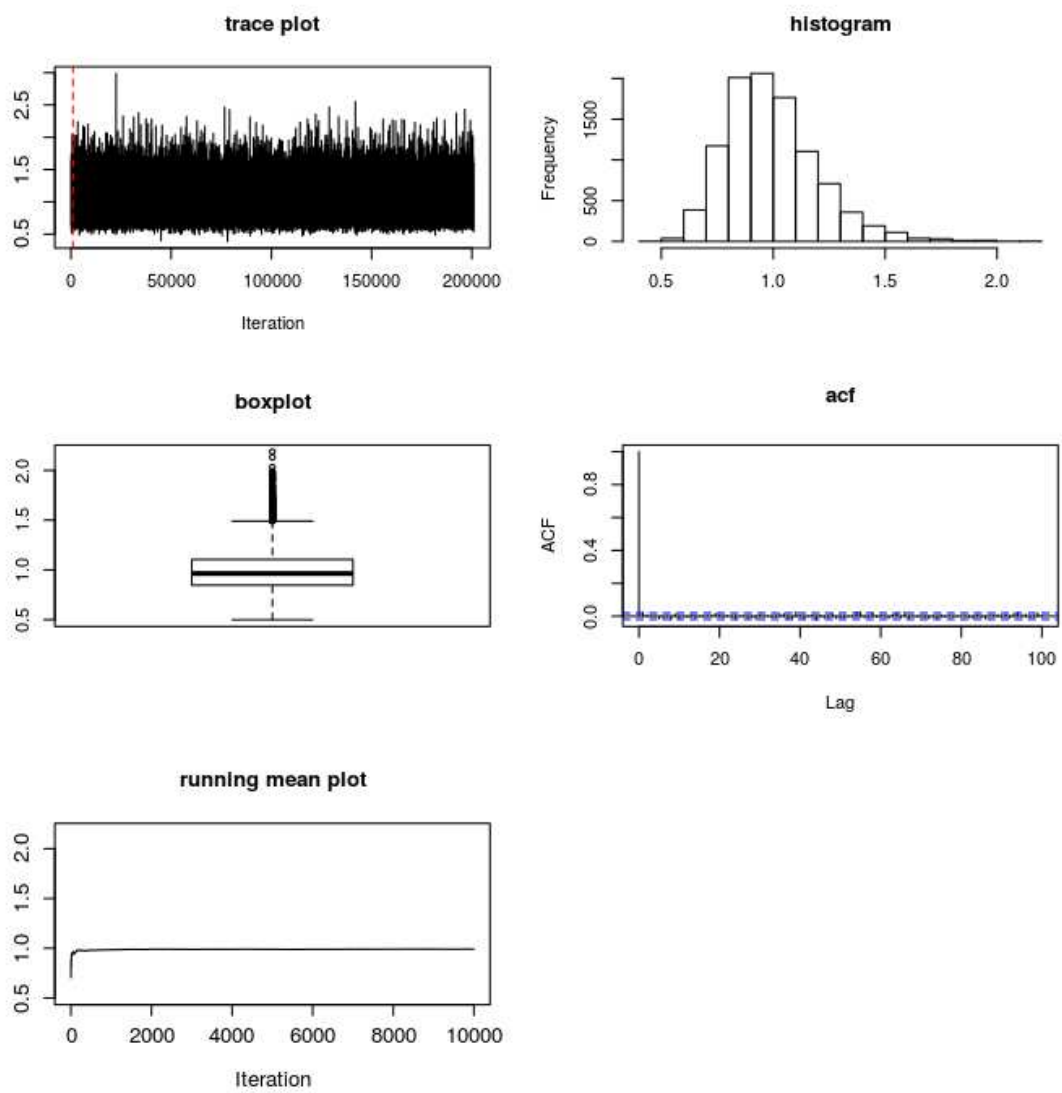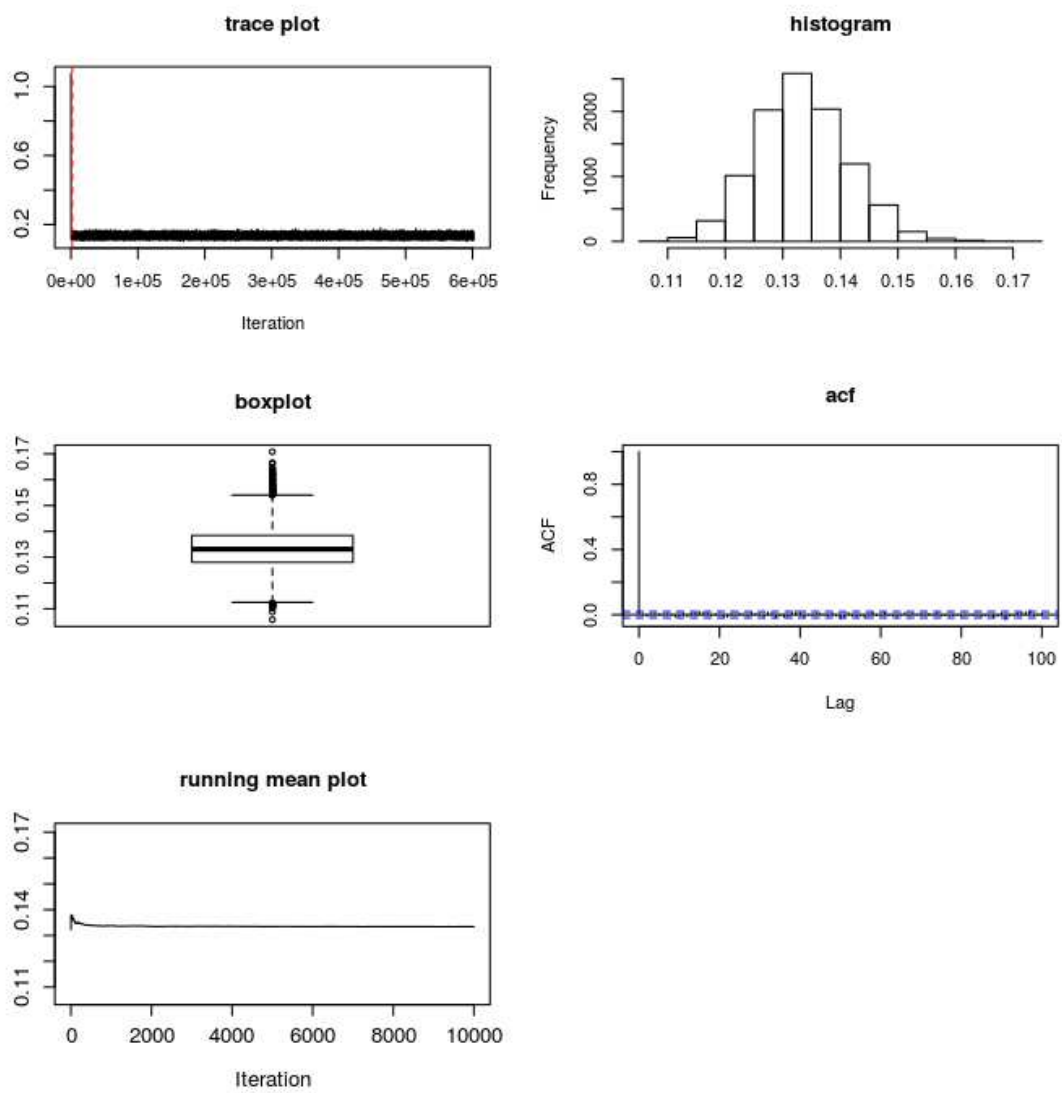
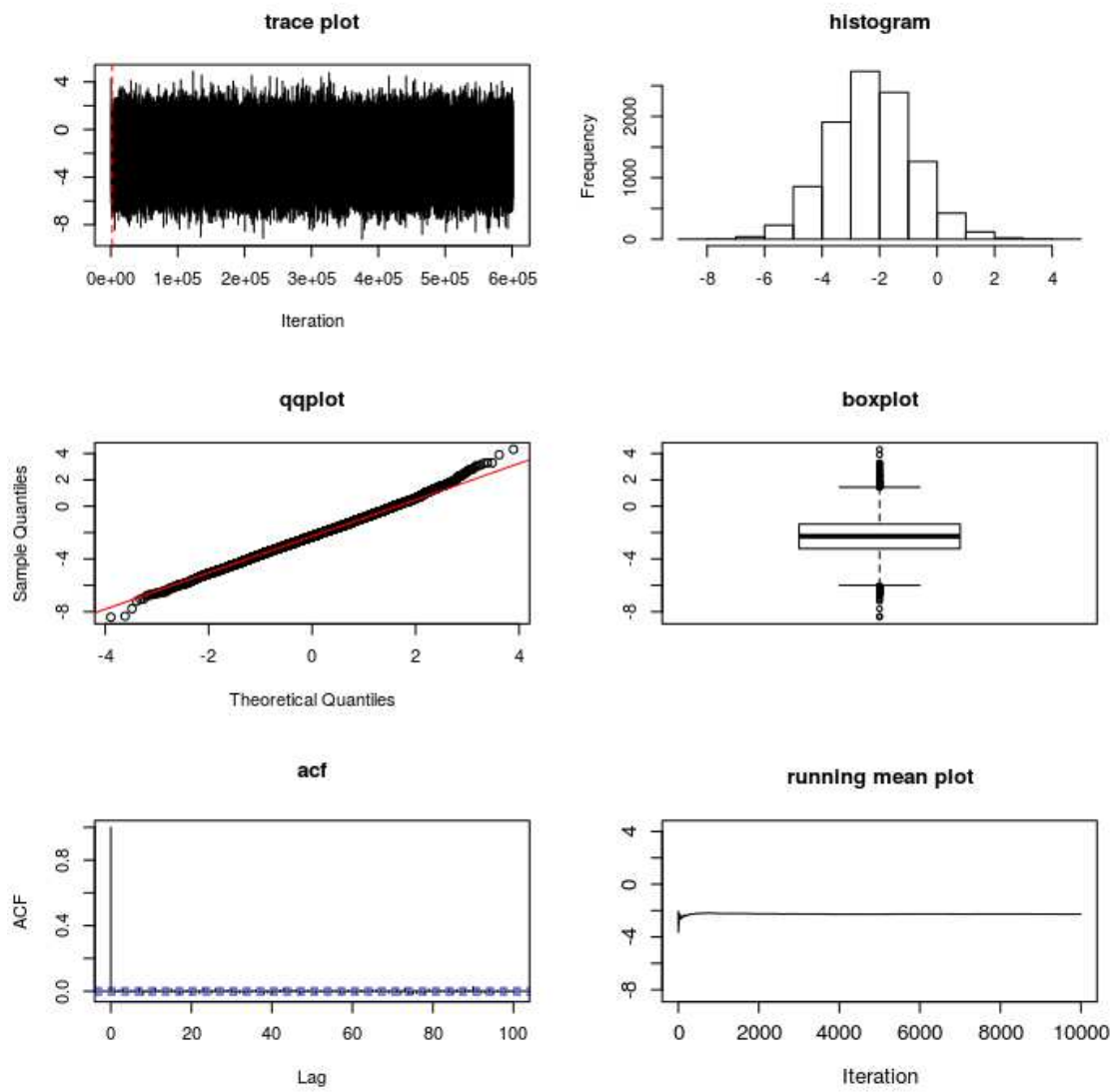Figure B.15: Site A: diagnostic on the estimation of $\Sigma(1,1)$ (temperature/moisture model, annual algorithm).
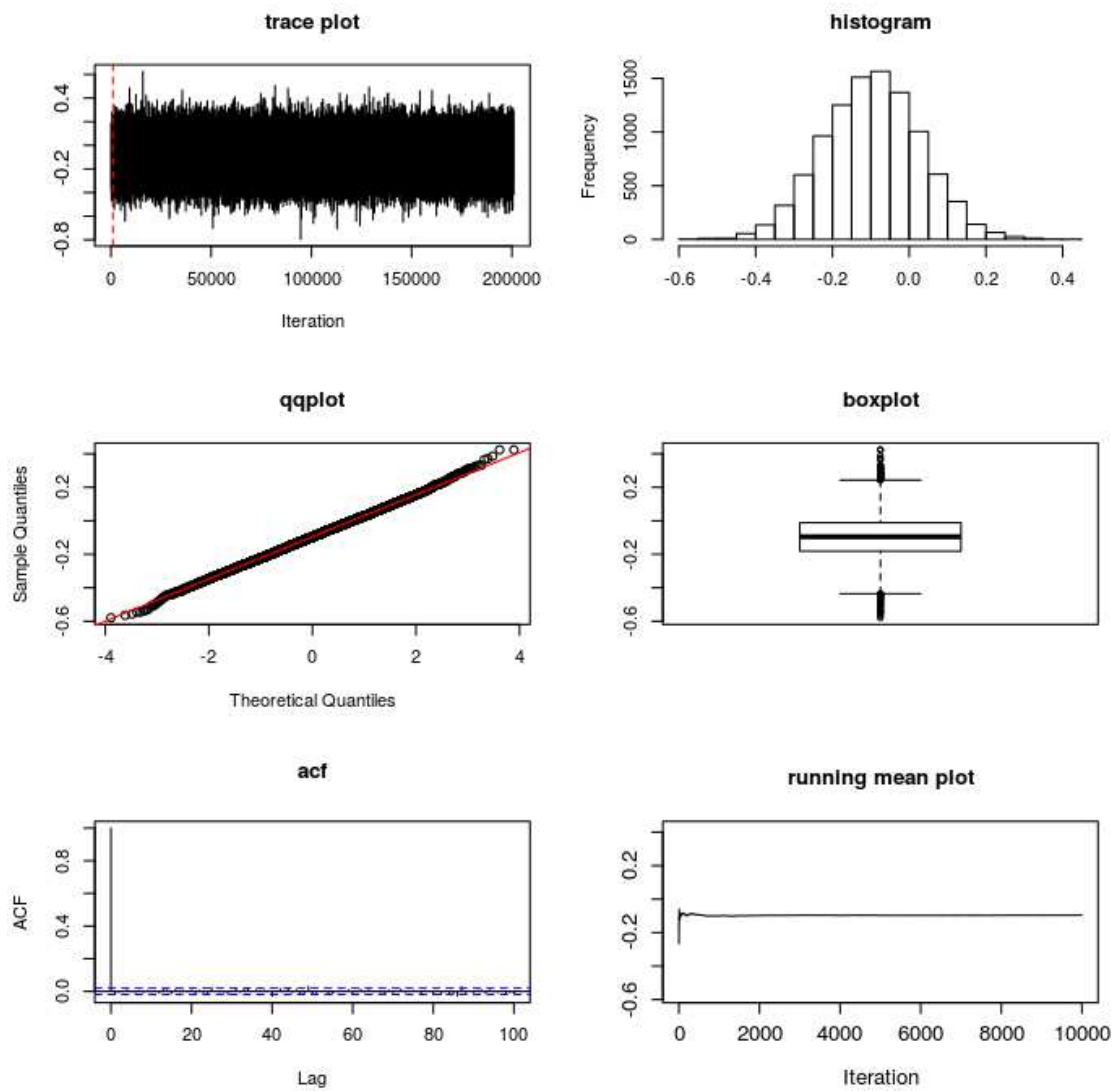
Figure B.16: Site A: diagnostic on the estimation of β_0 (temperature/moisture model, annual algorithm).
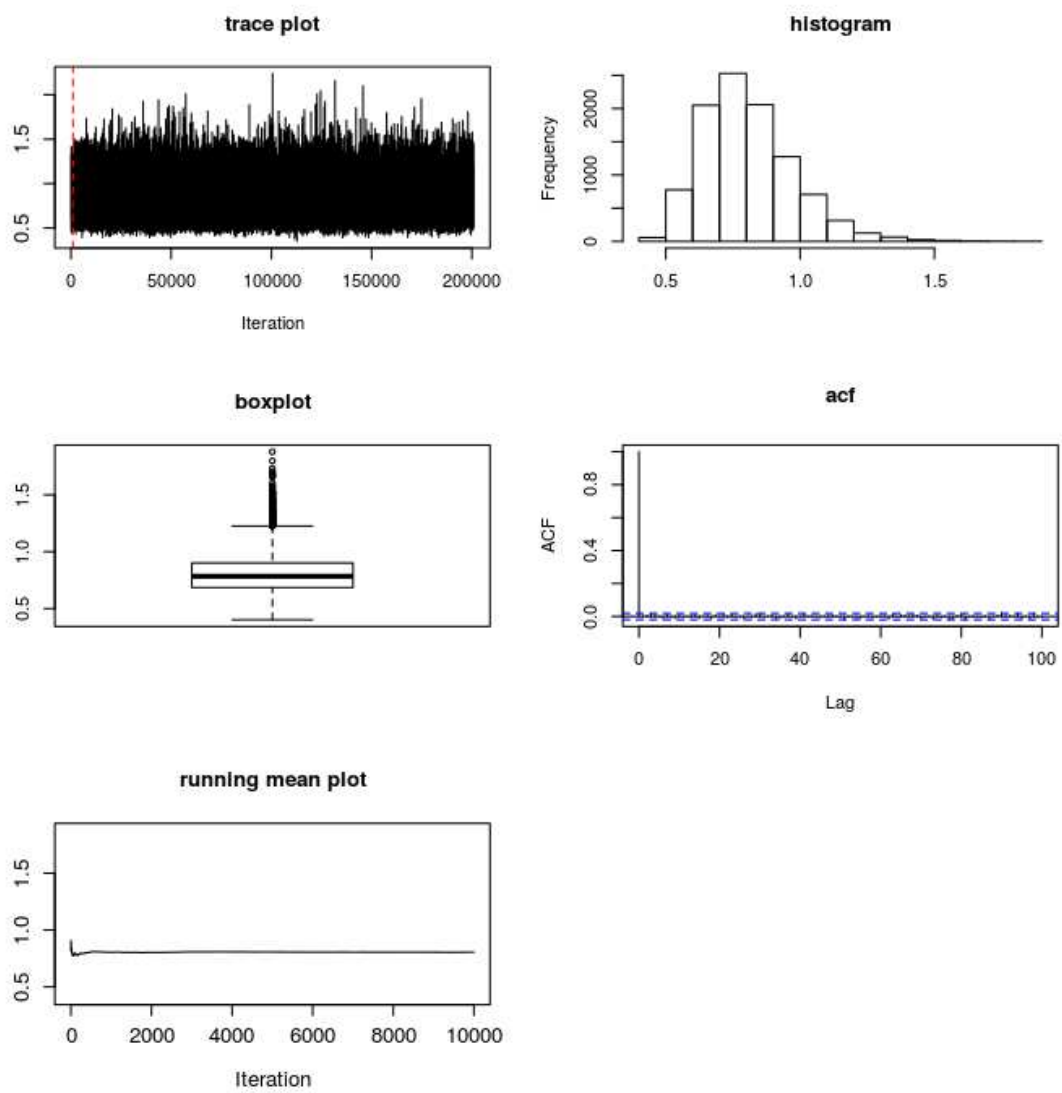
Figure B.17: Site A: diagnostic on the estimation of ψ^2 (temperature/moisture model, annual algorithm).
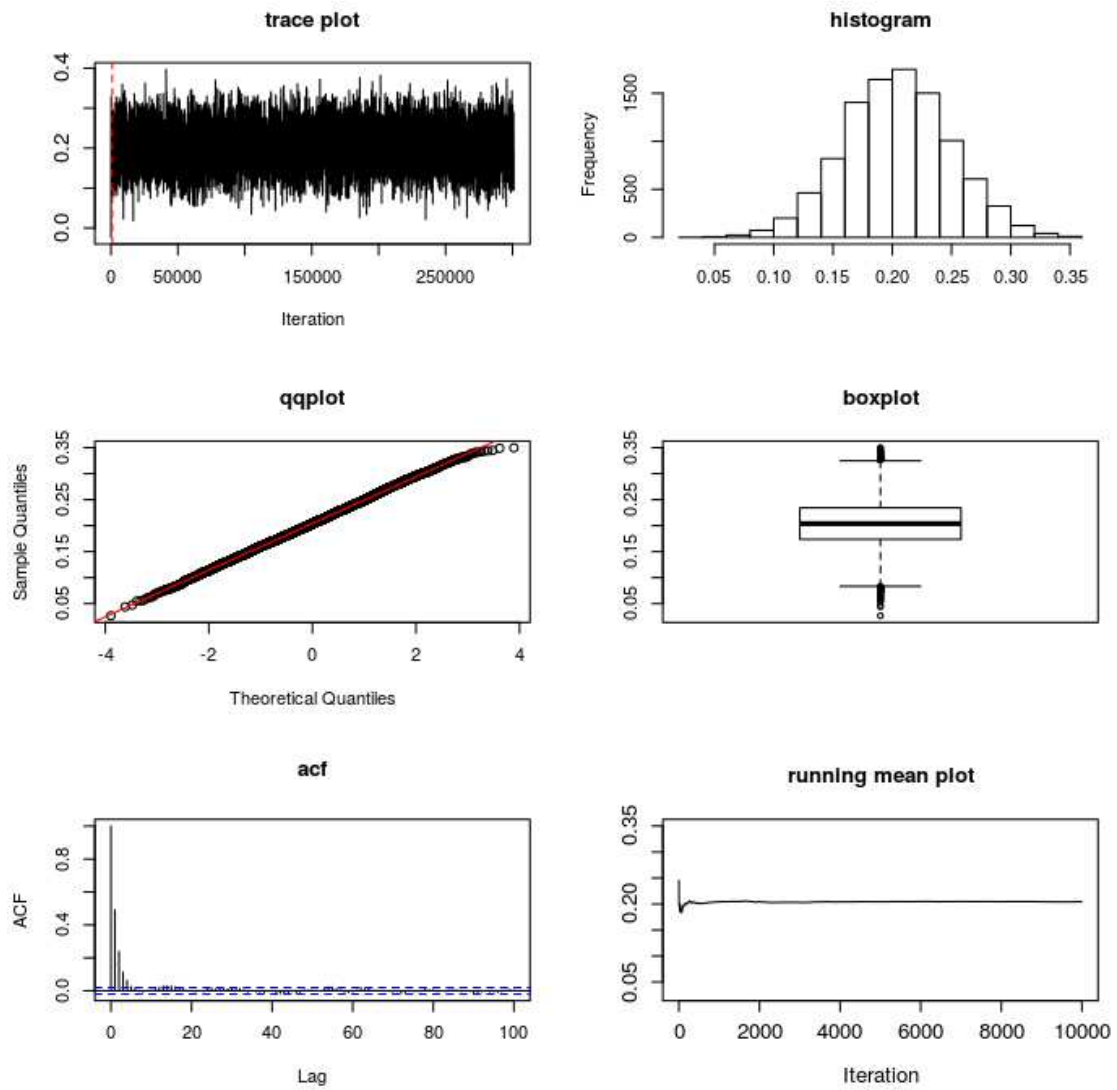
Figure B.18: Site A: diagnostic on the estimation of $\Phi_1(1,1)$ (temperature/moisture model).
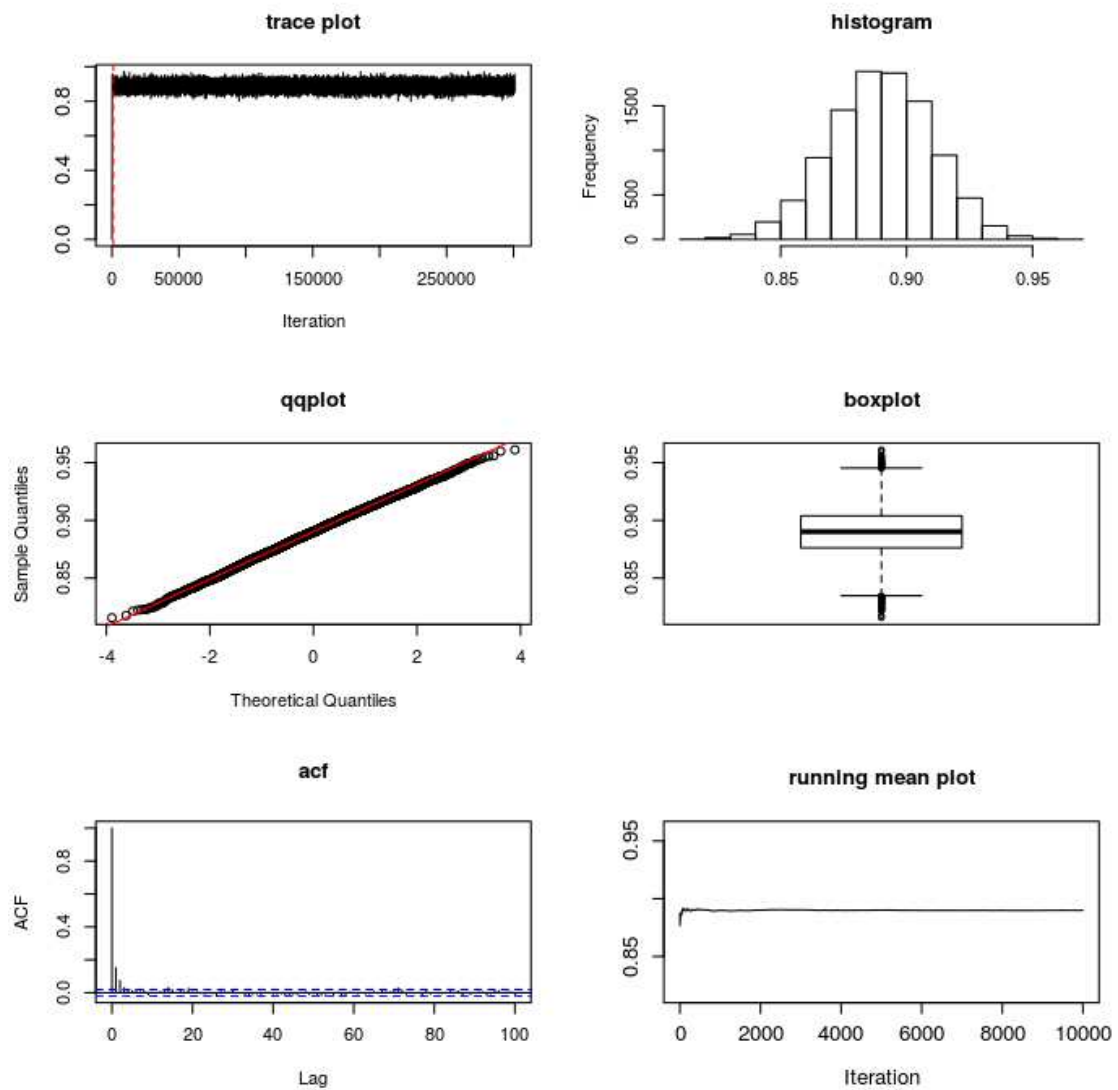
Figure B.19: Site A: diagnostic on the estimation of $\Phi_S(1,1)$ (temperature/moisture model).
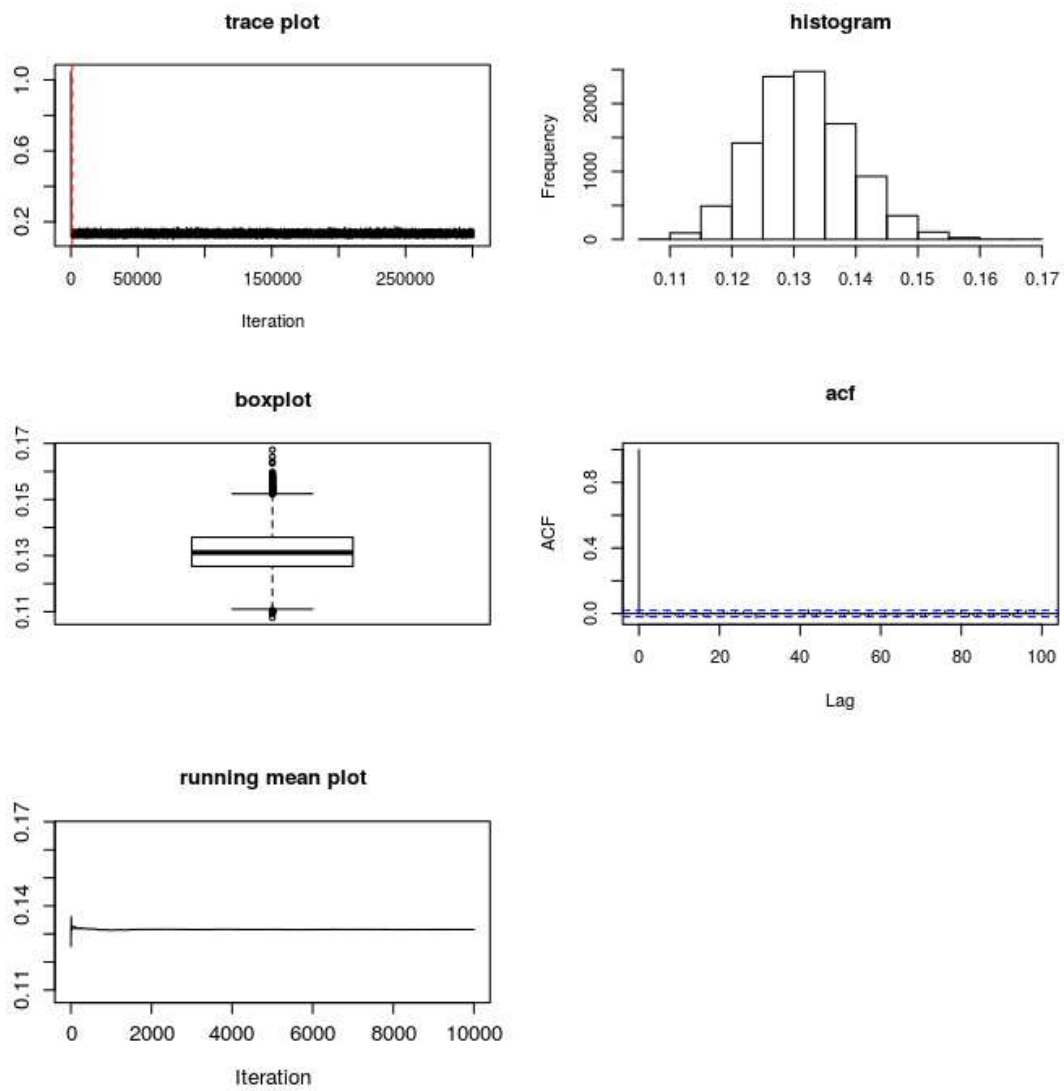
Figure B.20: Site A: diagnostic on the estimation of $\Sigma(1,1)$ (temperature/moisture model, monthly algorithm).
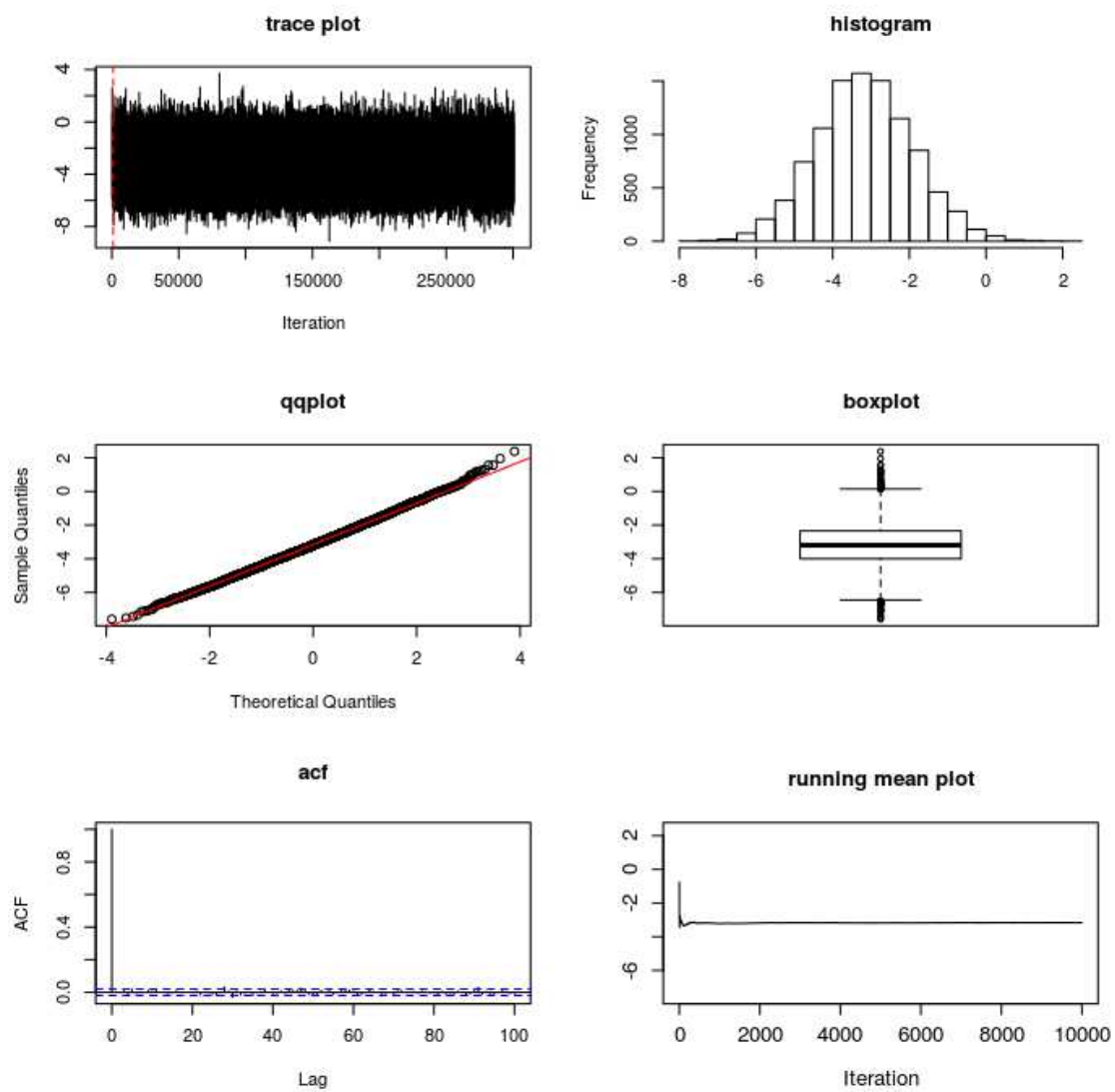
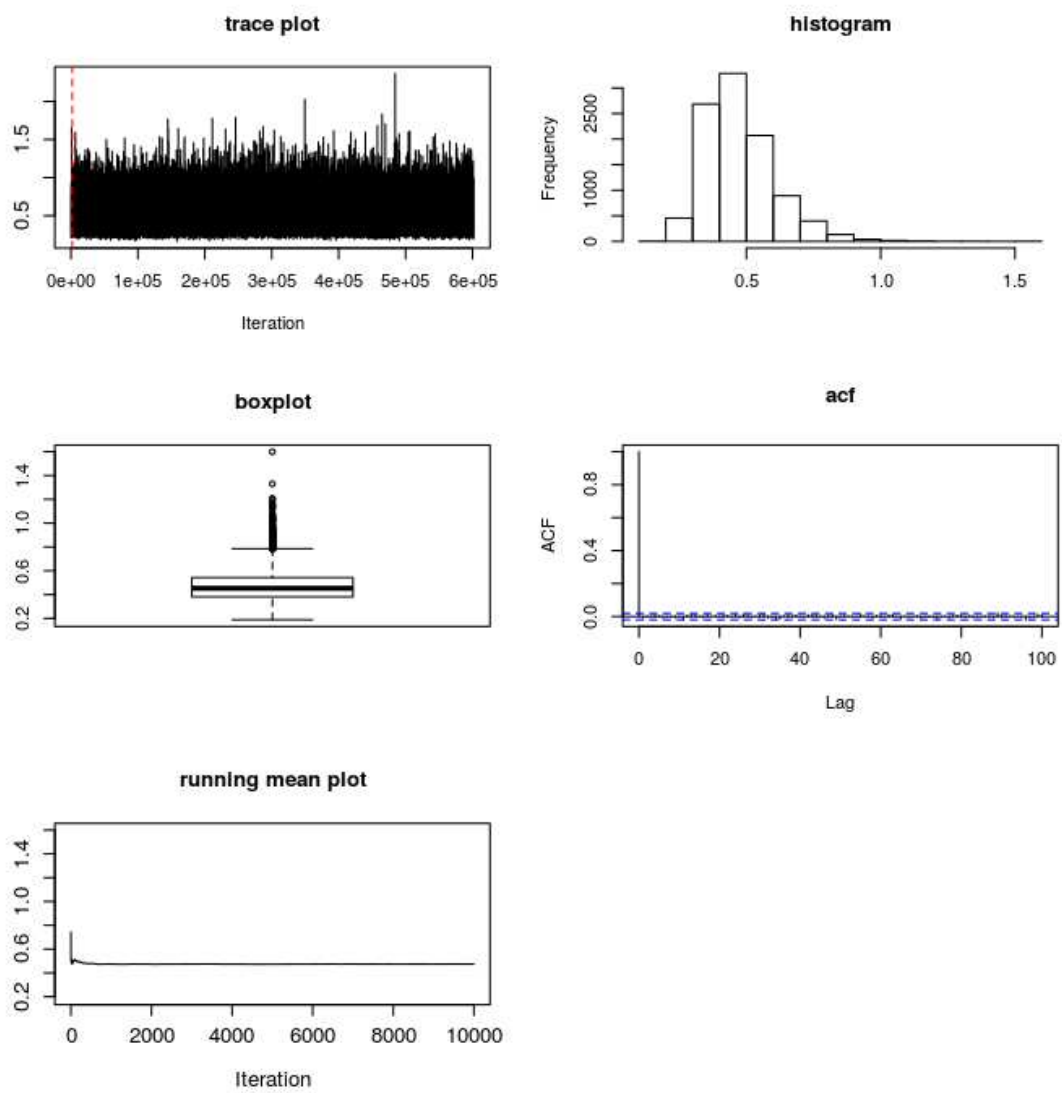Figure B.21: Site A: diagnostic on the estimation of β_0 (temperature/moisture model, monthly algorithm).

trace plot

histogram

boxplot

acf

running mean plot
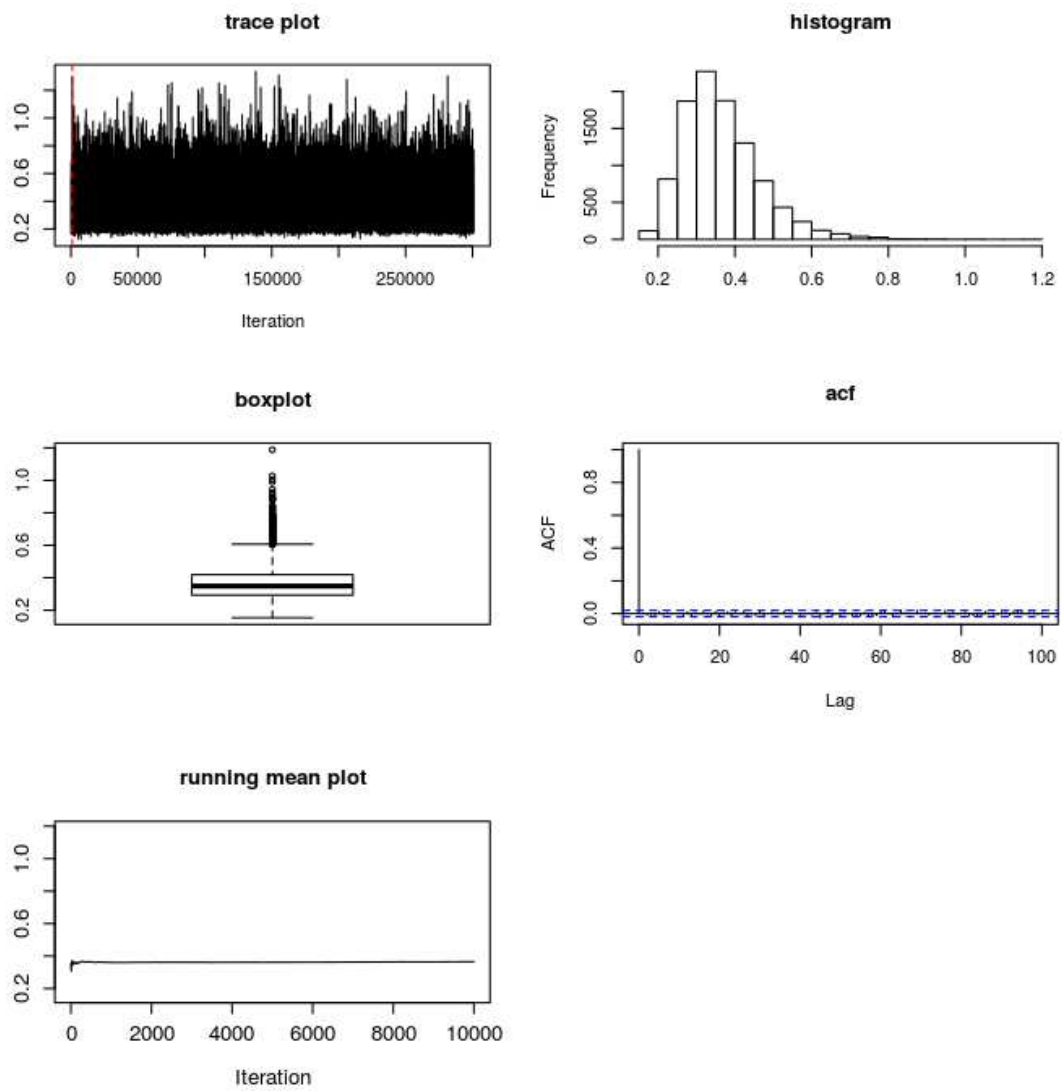
Figure B.22: Site A: diagnostic on the estimation of ψ^2 (temperature/moisture model, monthly algorithm).
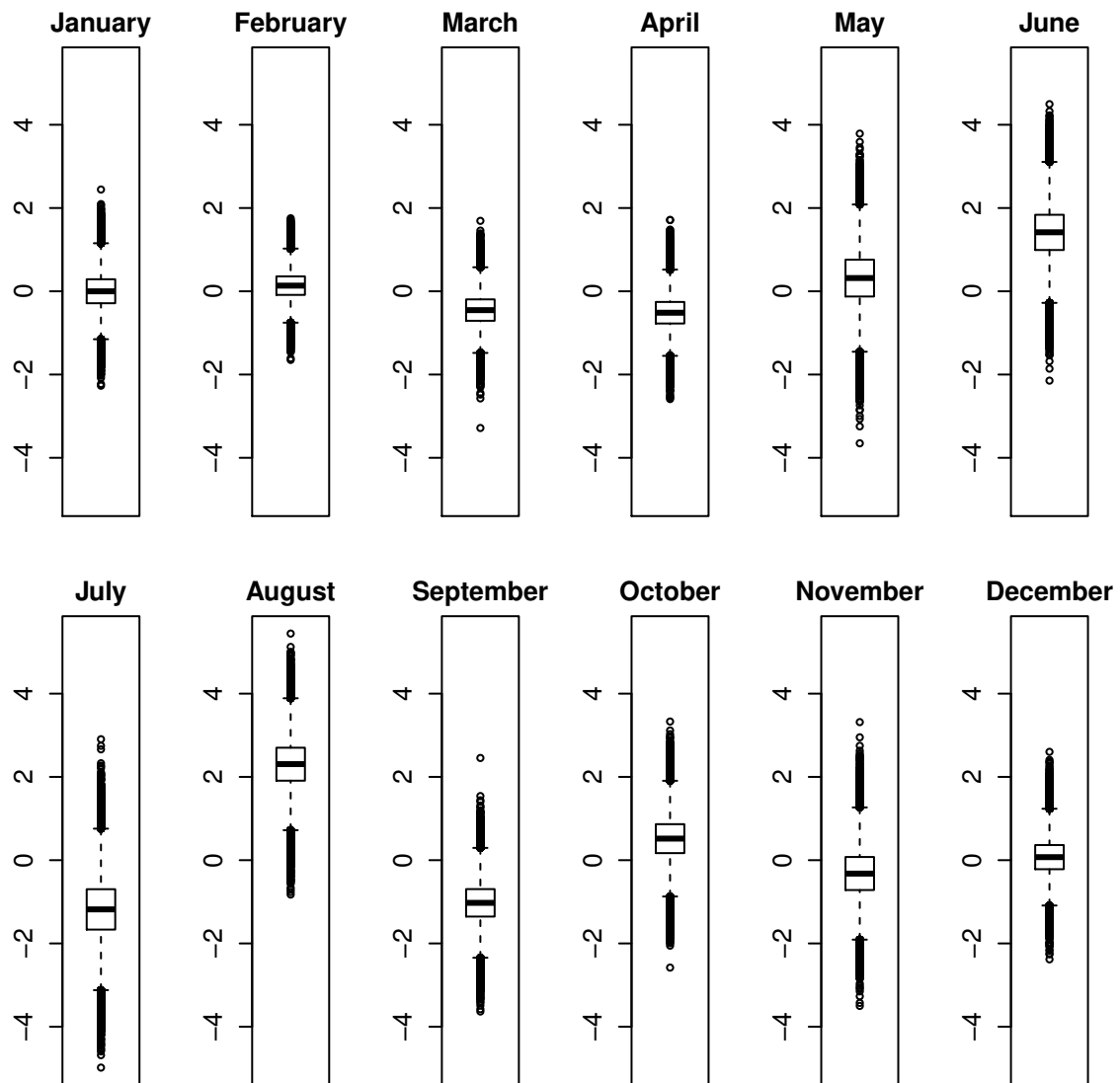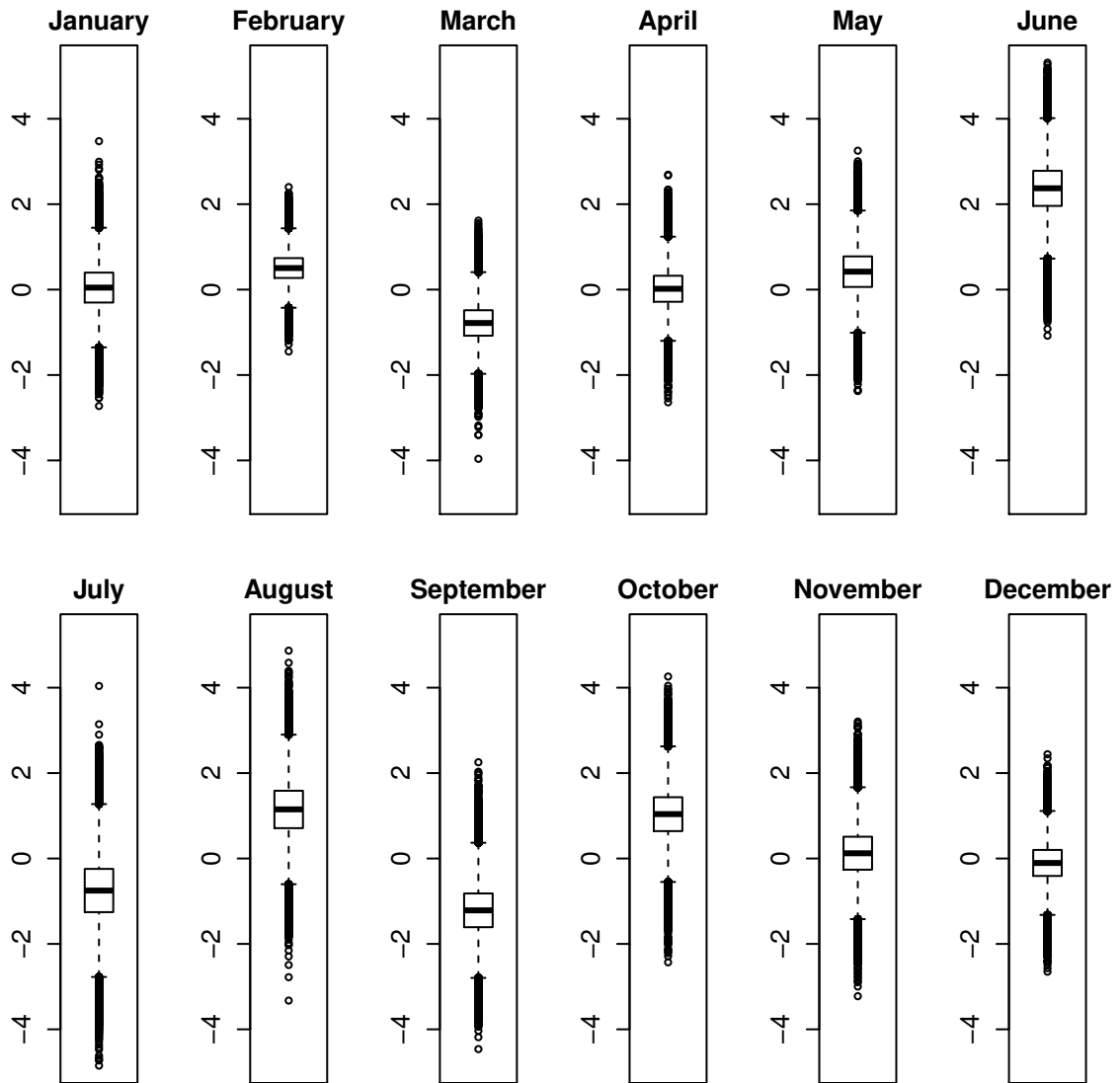
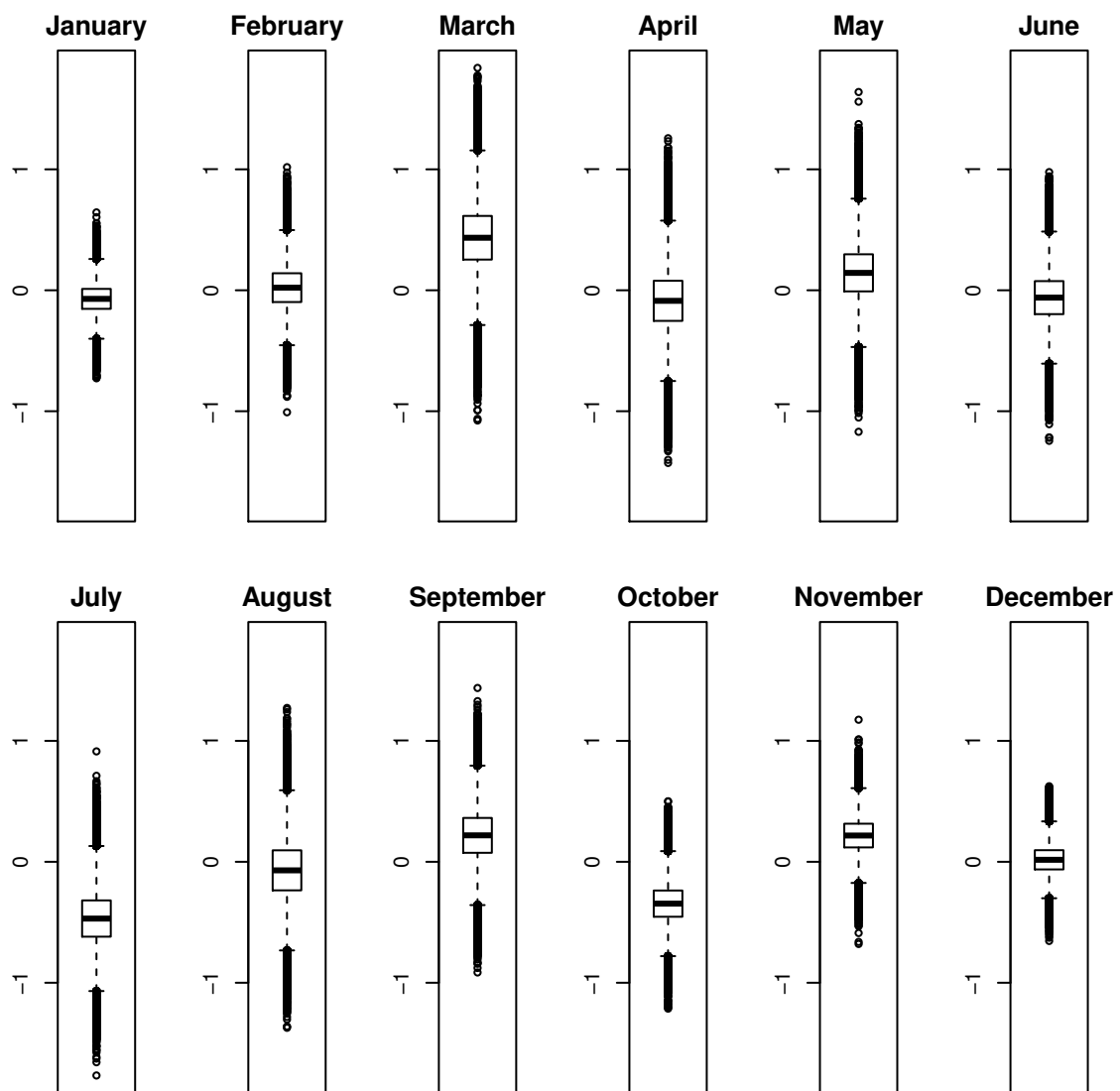Figure B.23: Site A: Temperature anomalies' coefficients (temperature/moisture model, monthly algorithm).
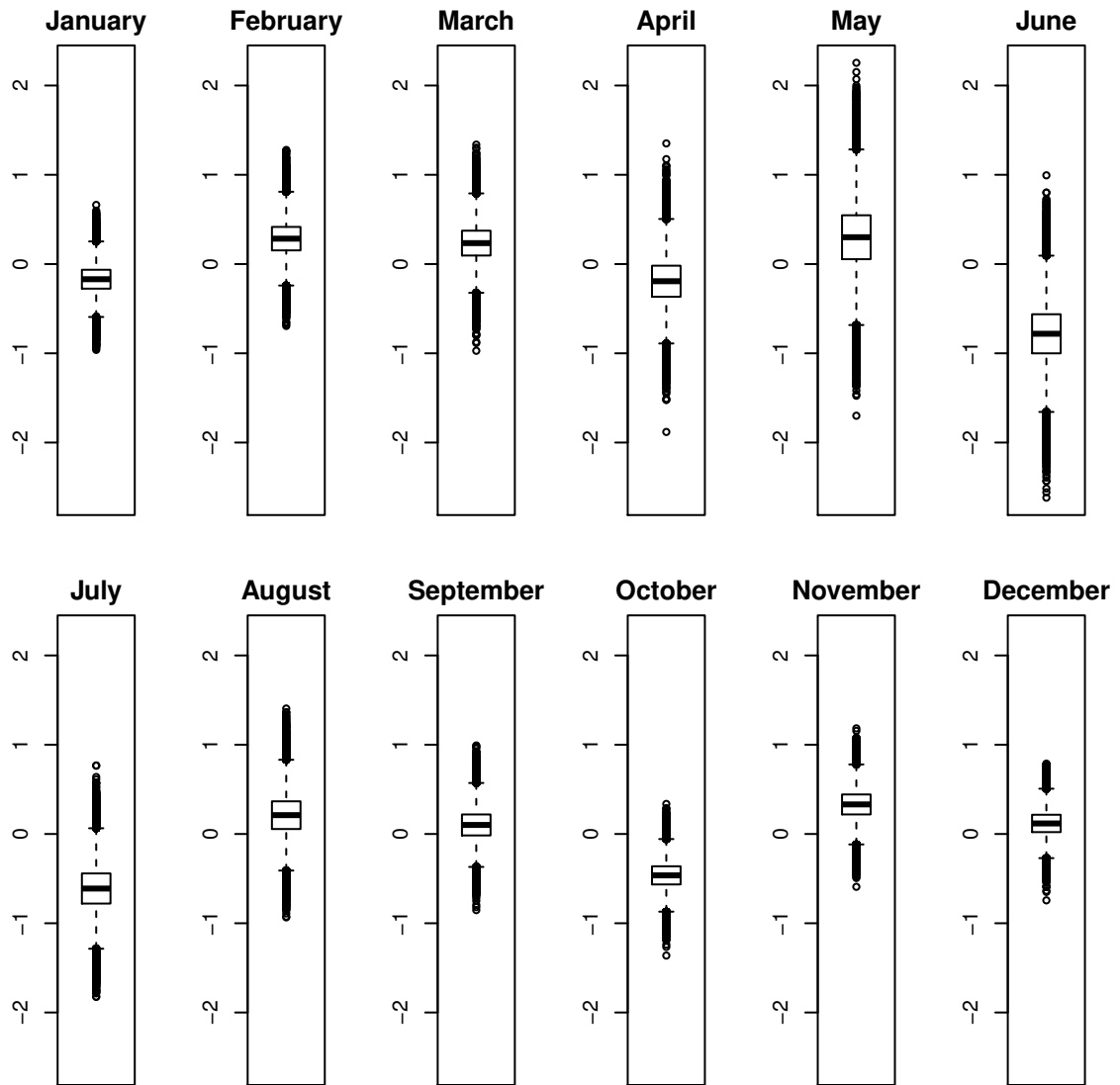
Figure B.24: Site A: Moisture anomalies' coefficients (temperature/moisture model, monthly algorithm).
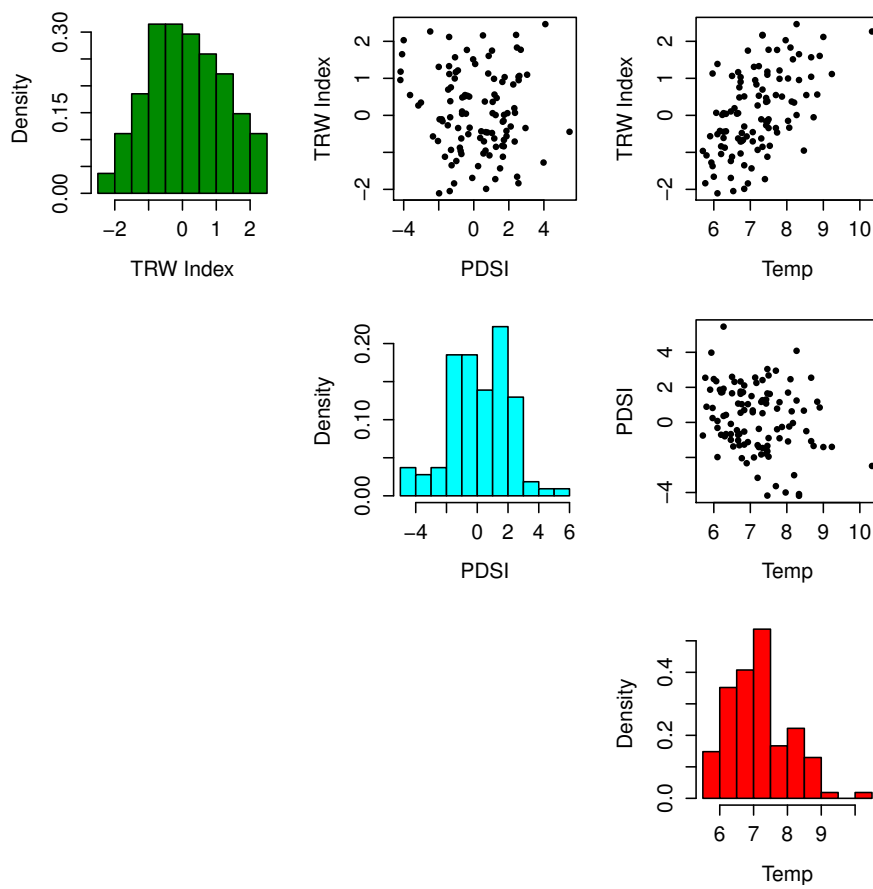
Figure B.25: Site A. Histograms and scatter-plots for tree-ring widhts index, PDSI and temperature in the period 1901–2008.
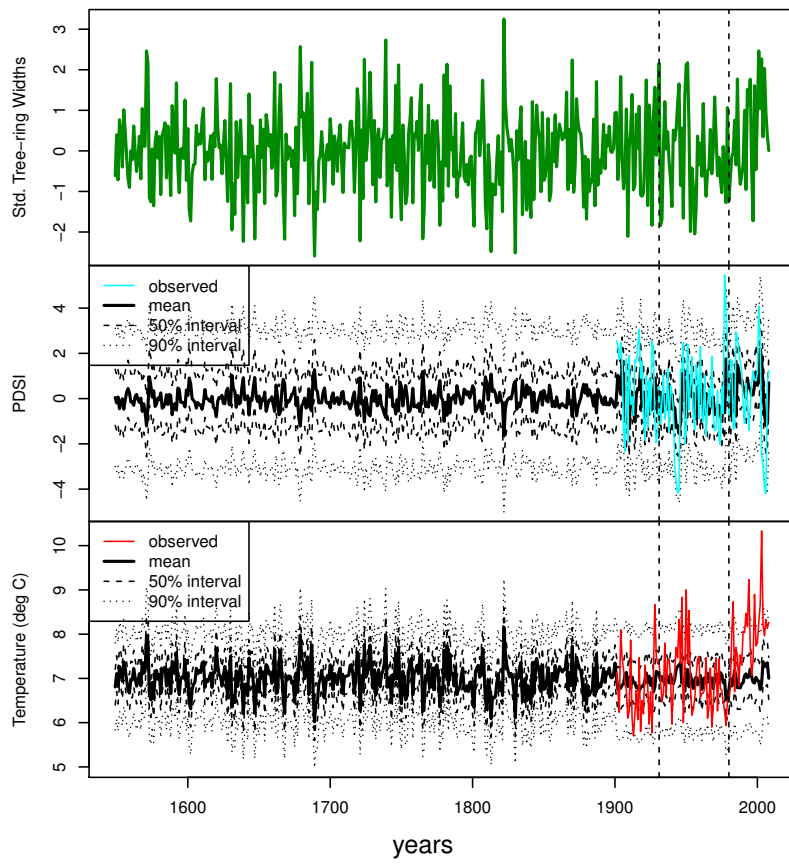
Figure B.26: Site A: estimated annual series (temperature/PDSI model).
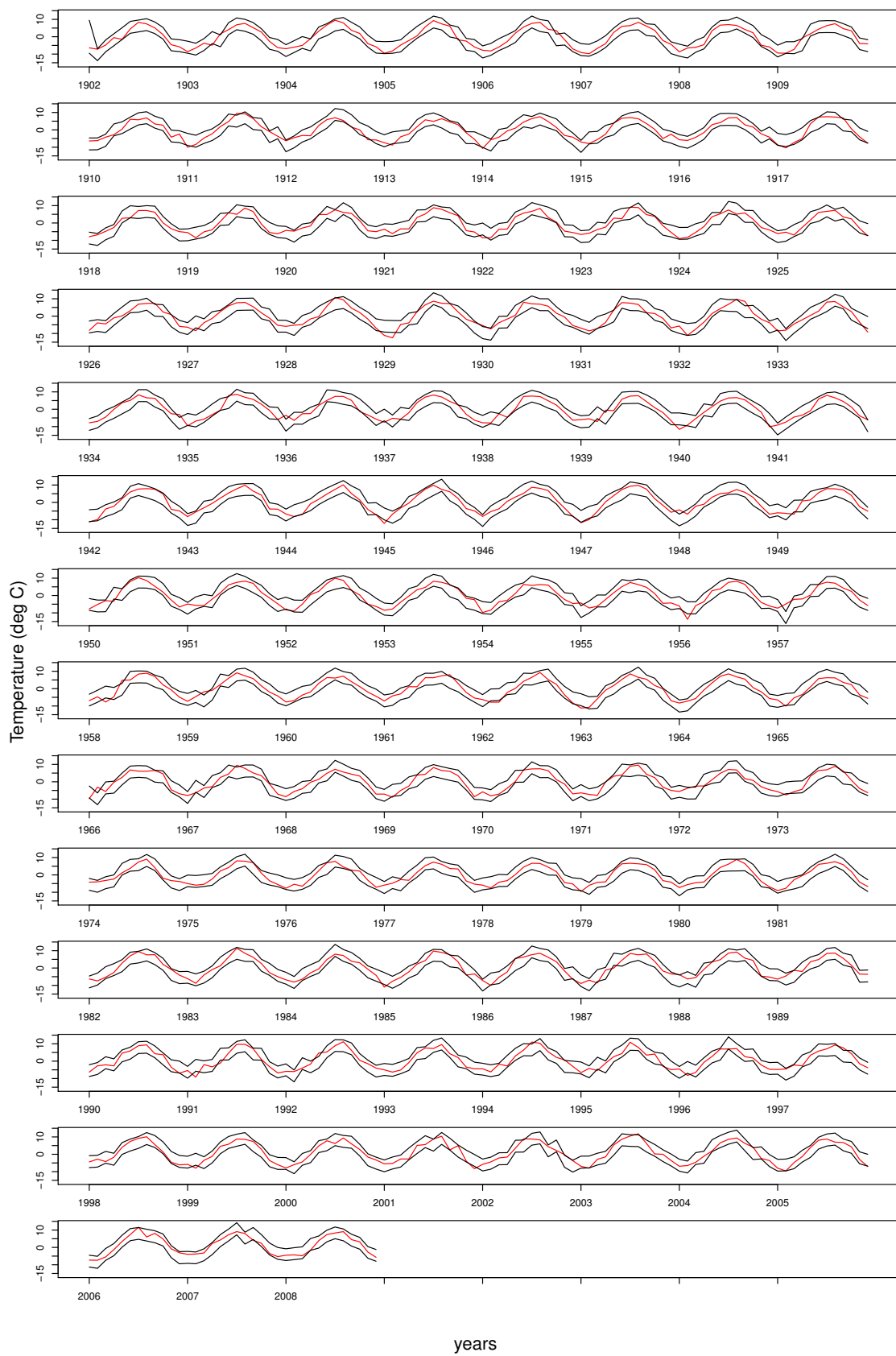
Figure B.27: Site A: monthly temperature series in 1902 – 2008. In red the observed series, in black 90% credibility interval (temperature/PDSI model).
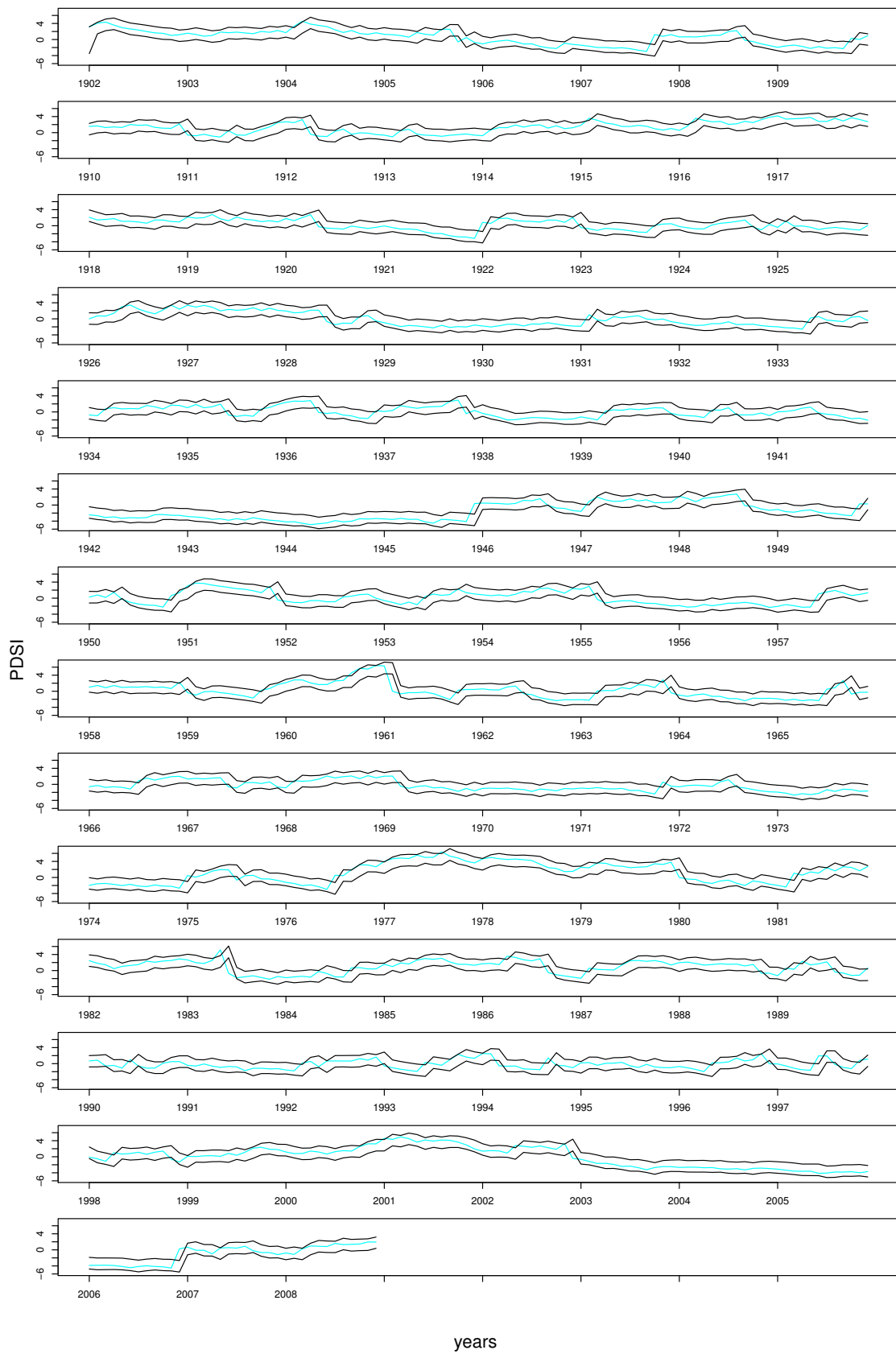
Figure B.28: Site A: monthly PDSI series in 1902 – 2008. In light-blue the observed series, in black 90% credibility interval (temperature/PDSI model).
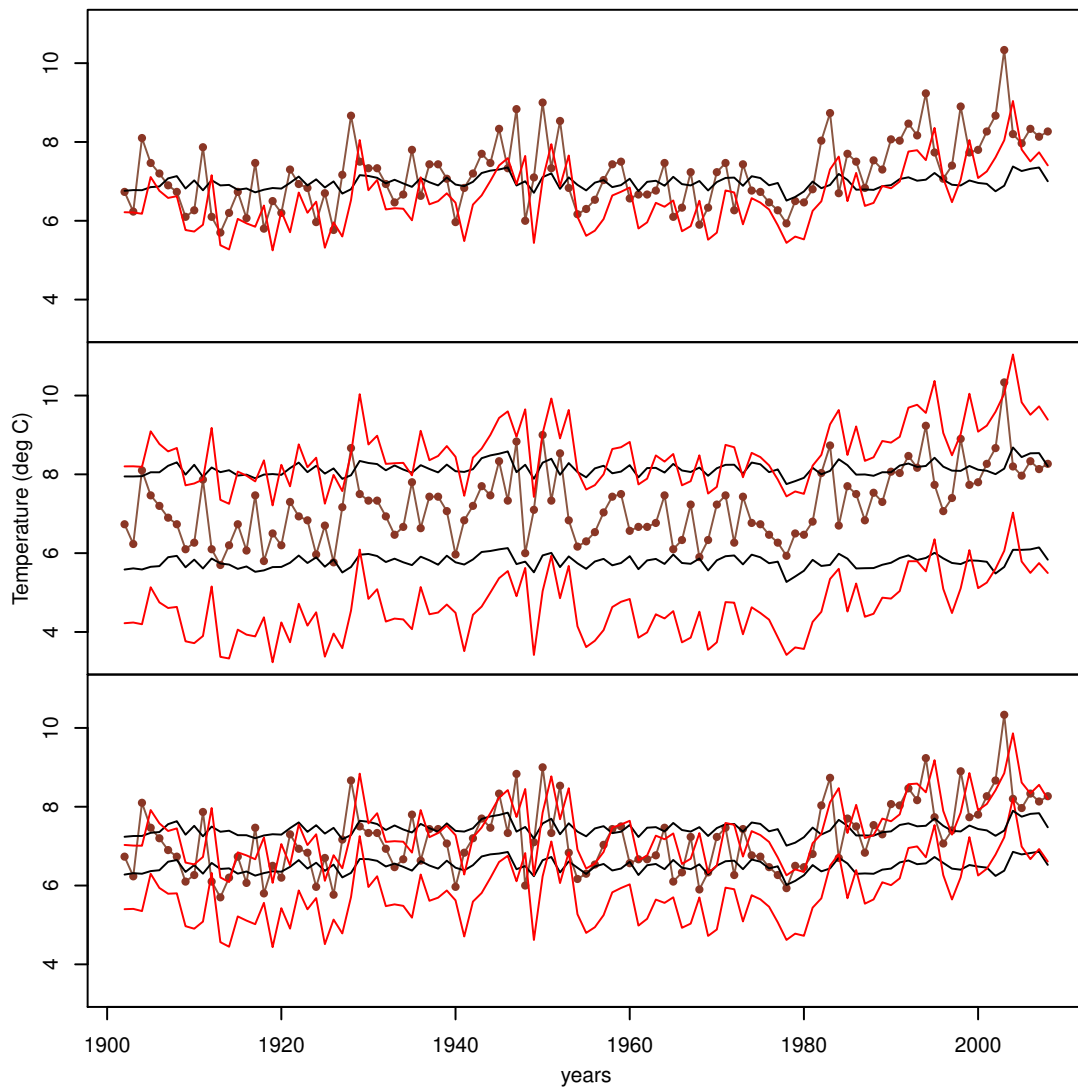
Figure B.29: Site A: comparison between temperature in the observation period: in red the result from the monthly algorithm, in black from the annual one. Brown series is the observed one. From top to bottom: mean, 90% interval, 50% interval (temperature/PDSI model).
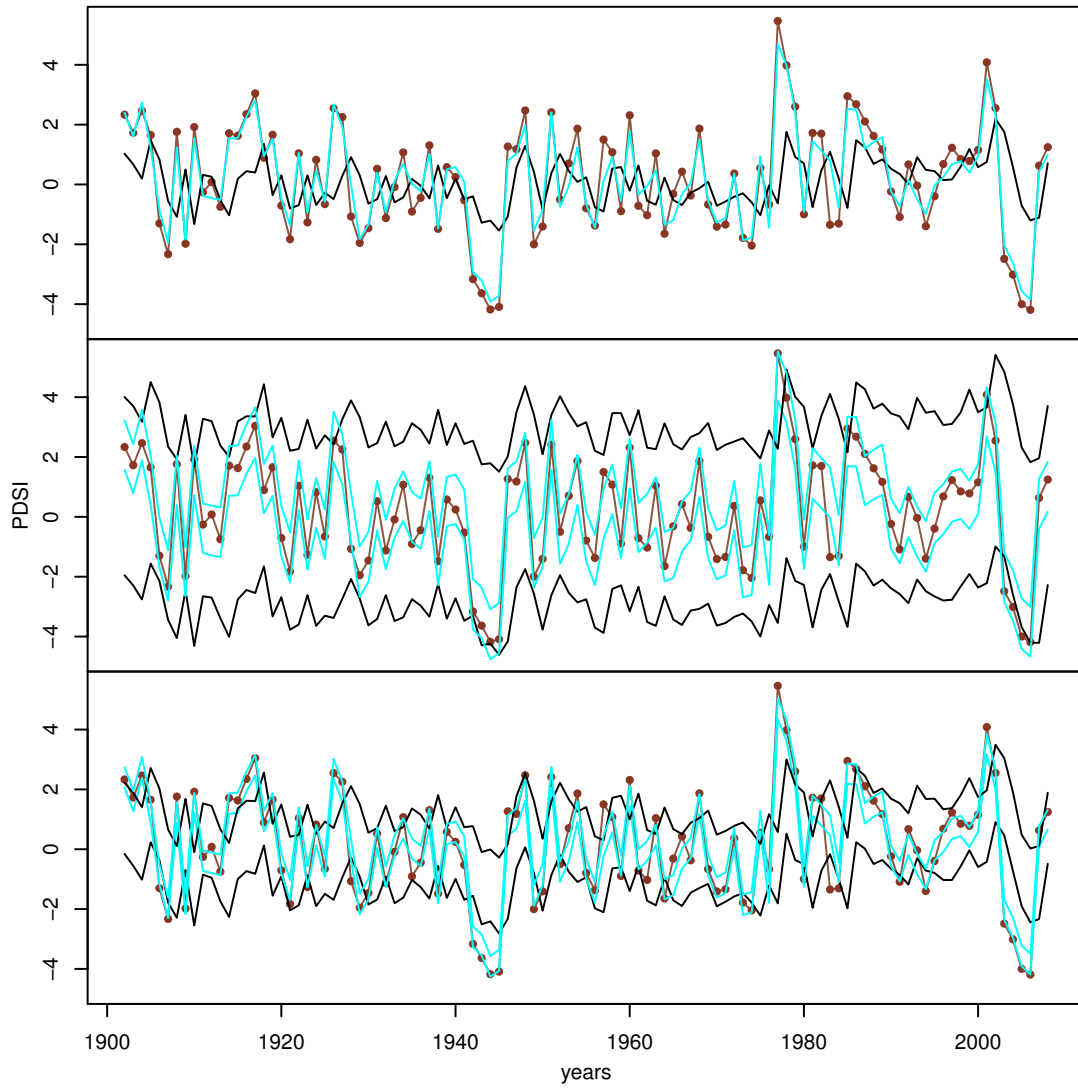
Figure B.30: Site A: comparison between PDSI in the observation period: in light-blue the result from the monthly algorithm, in black from the annual one. Brown series is the observed one. From top to bottom: mean, 90% interval, 50% interval (temperature/PDSI model).
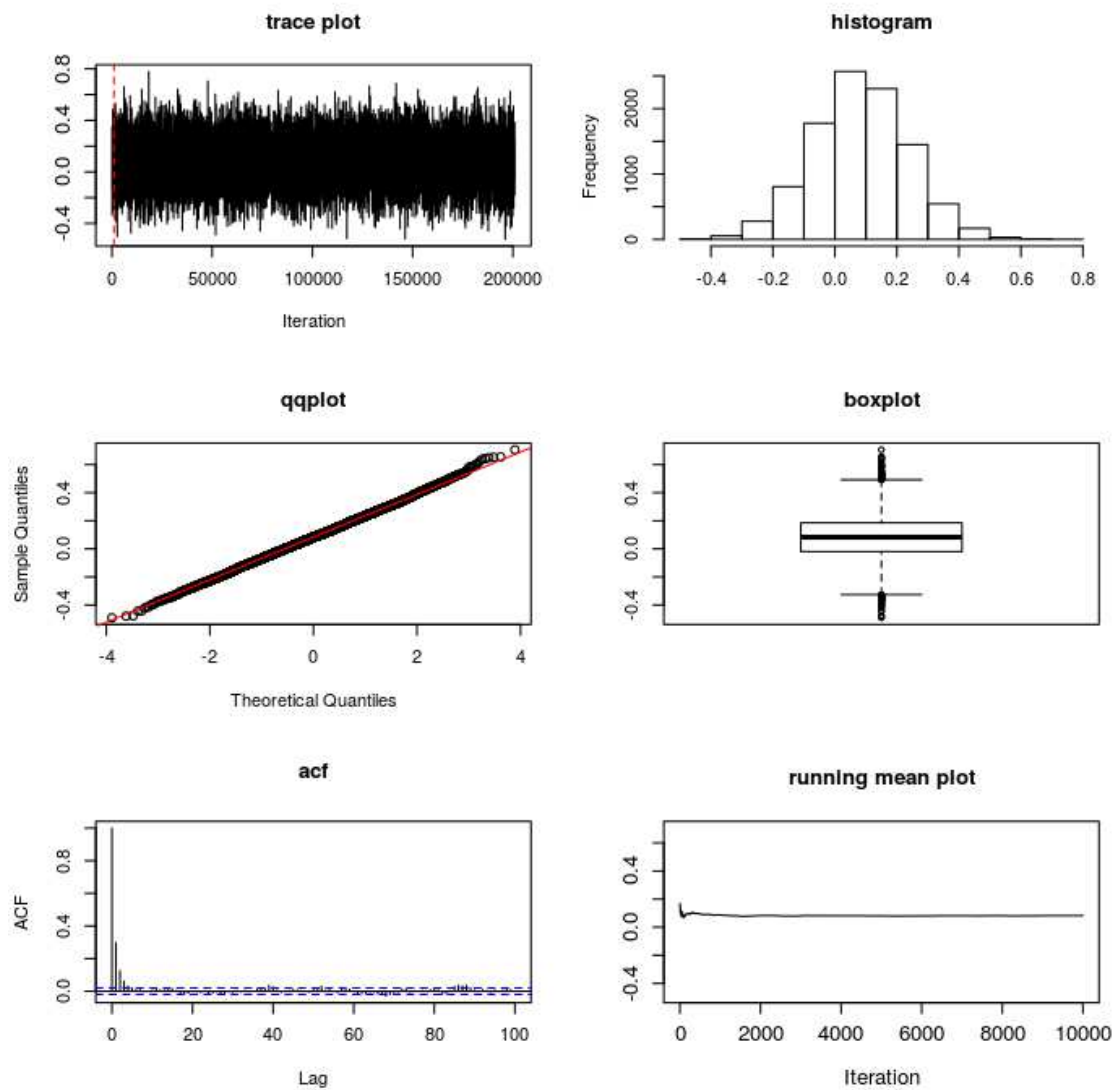
Figure B.31: Site A: diagnostic on the estimation of $\Phi(1,1)$ (temperature/PDSI model).
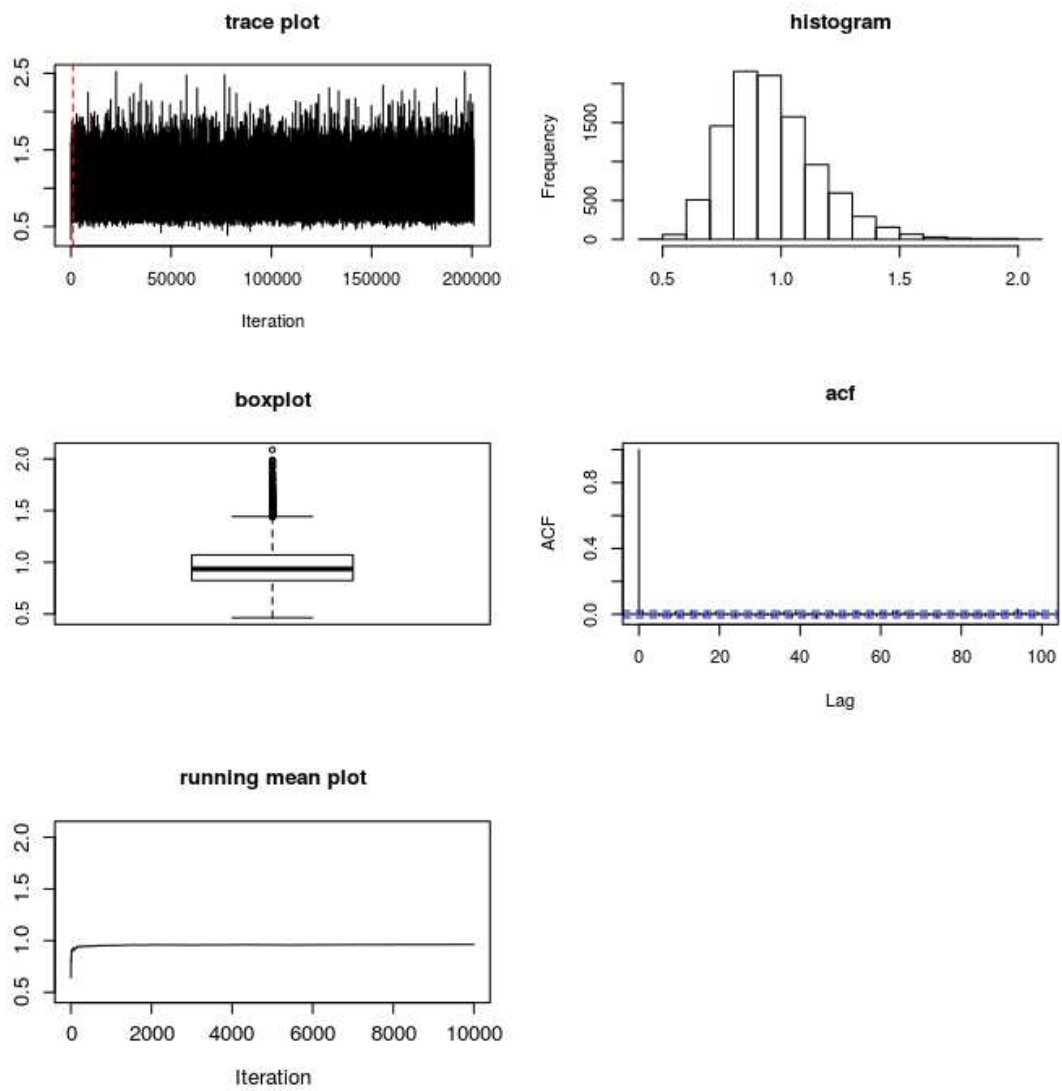
Figure B.32: Site A: diagnostic on the estimation of $\Sigma(1,1)$ (temperature/PDSI model, annual algorithm).
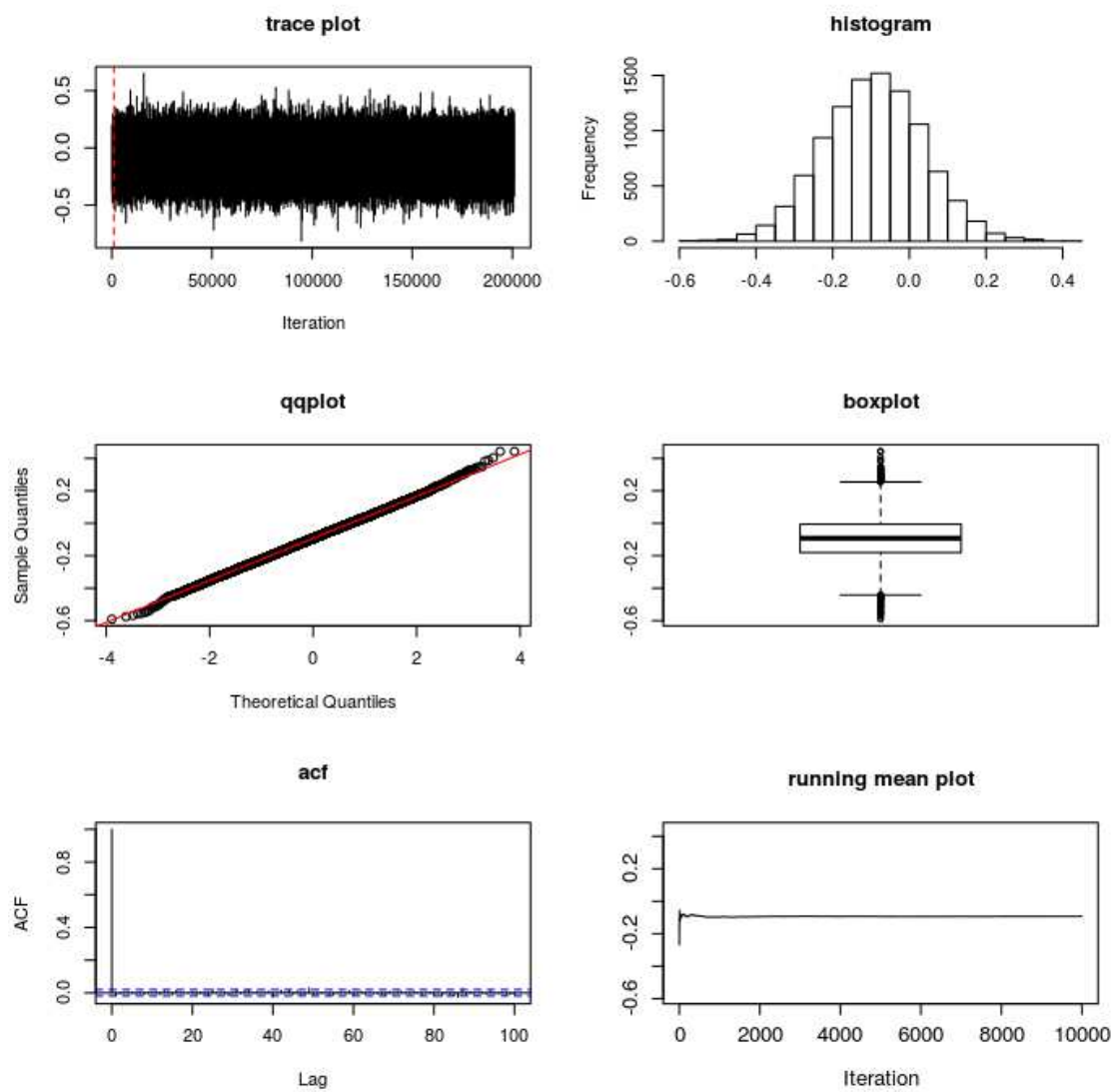
Figure B.33: Site A: diagnostic on the estimation of β_0 (temperature/PDSI model, annual algorithm).
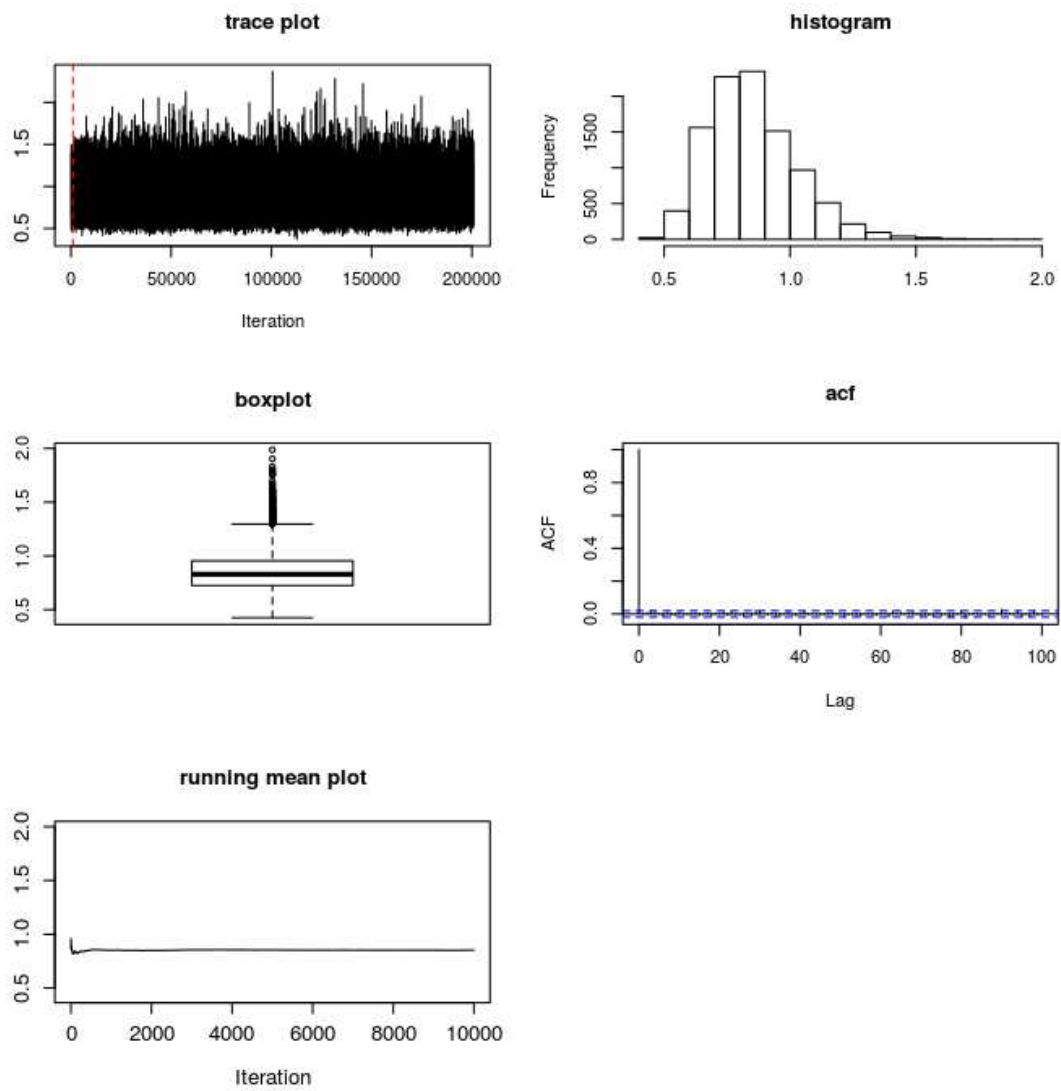
Figure B.34: Site A: diagnostic on the estimation of ψ^2 (temperature/PDSI model, annual algorithm).
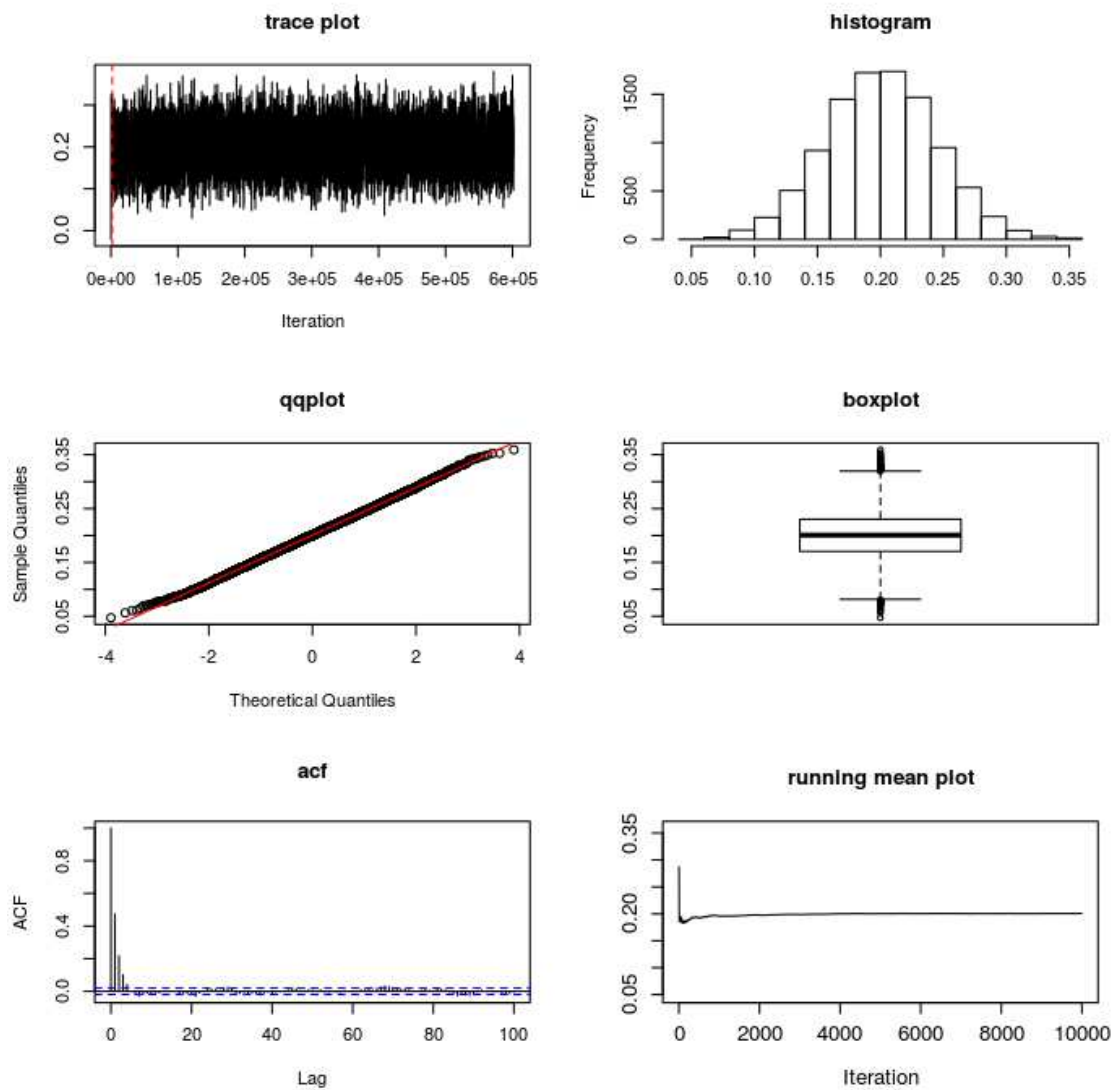
Figure B.35: Site A: diagnostic on the estimation of $\Phi_1(1,1)$ (temperature/PDSI model).

Figure B.36: Site A: diagnostic on the estimation of $\Phi_S(1, 1)$ (temperature/PDSI model).

Figure B.37: Site A: diagnostic on the estimation of $\Sigma(1,1)$ (temperature/PDSI model, monthly algorithm).

Figure B.38: Site A: diagnostic on the estimation of β_0 (temperature/PDSI model, monthly algorithm).

Figure B.39: Site A: diagnostic on the estimation of ψ^2 (temperature/PDSI model, monthly algorithm).

Figure B.40: Site A: Temperature anomalies' coefficients (temperature/PDSI model, monthly algorithm).

Figure B.41: Site A: PDSI anomalies' coefficients (temperature/PDSI model, monthly algorithm).
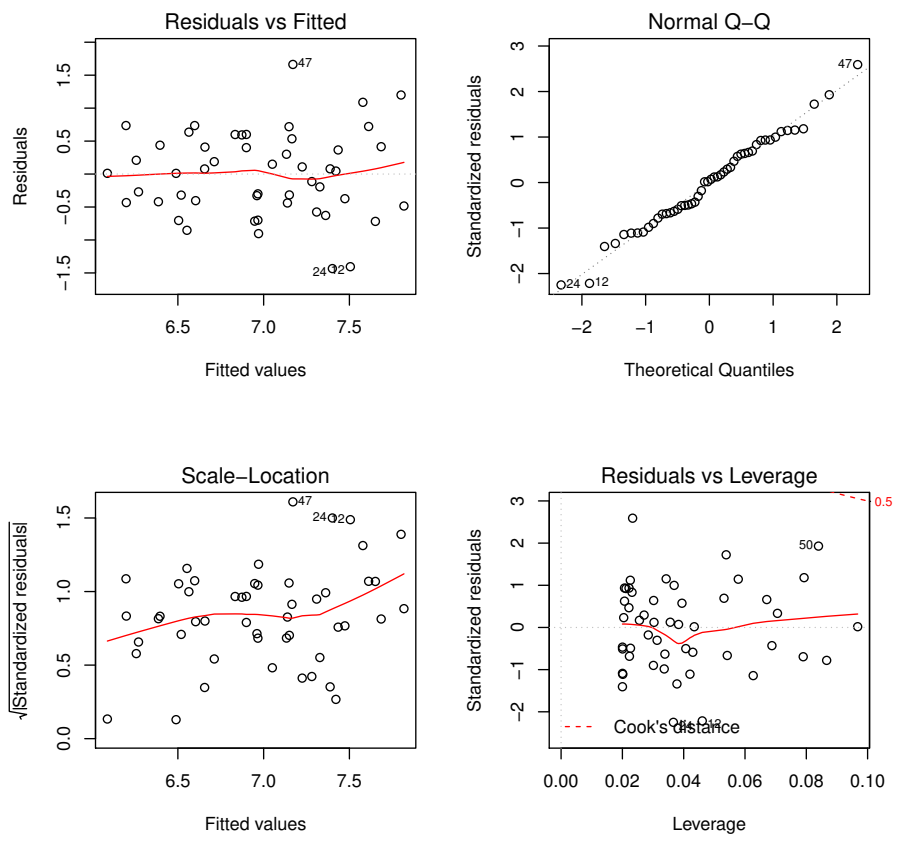
Figure B.42: Site A: temperature series. Diagnostics for the regression-based traditional approach.
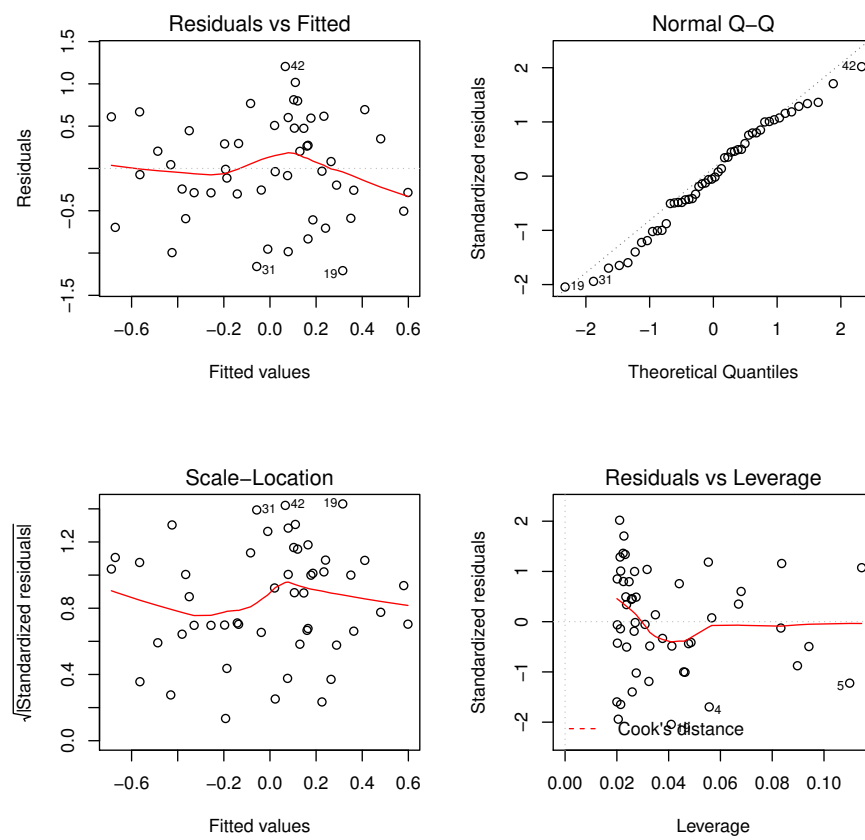
Figure B.43: Site A: moisture anomaly series. Diagnostics for the regression-based traditional approach.
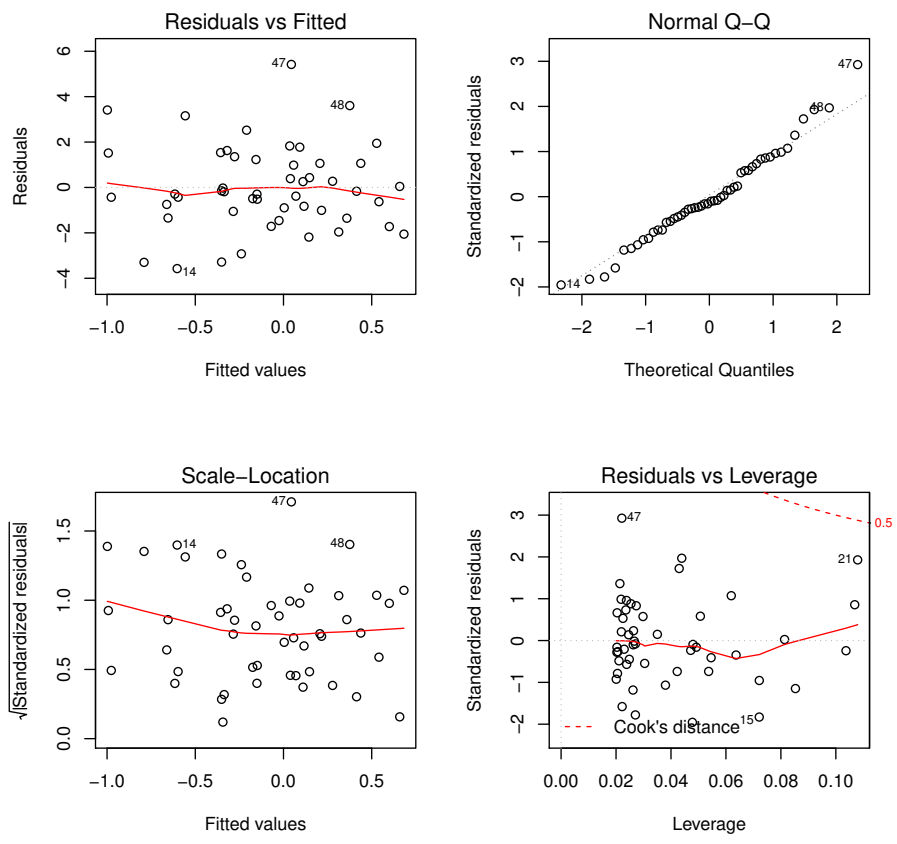
Figure B.44: Site A: PDSI series. Diagnostics for the regression-based traditional approach.
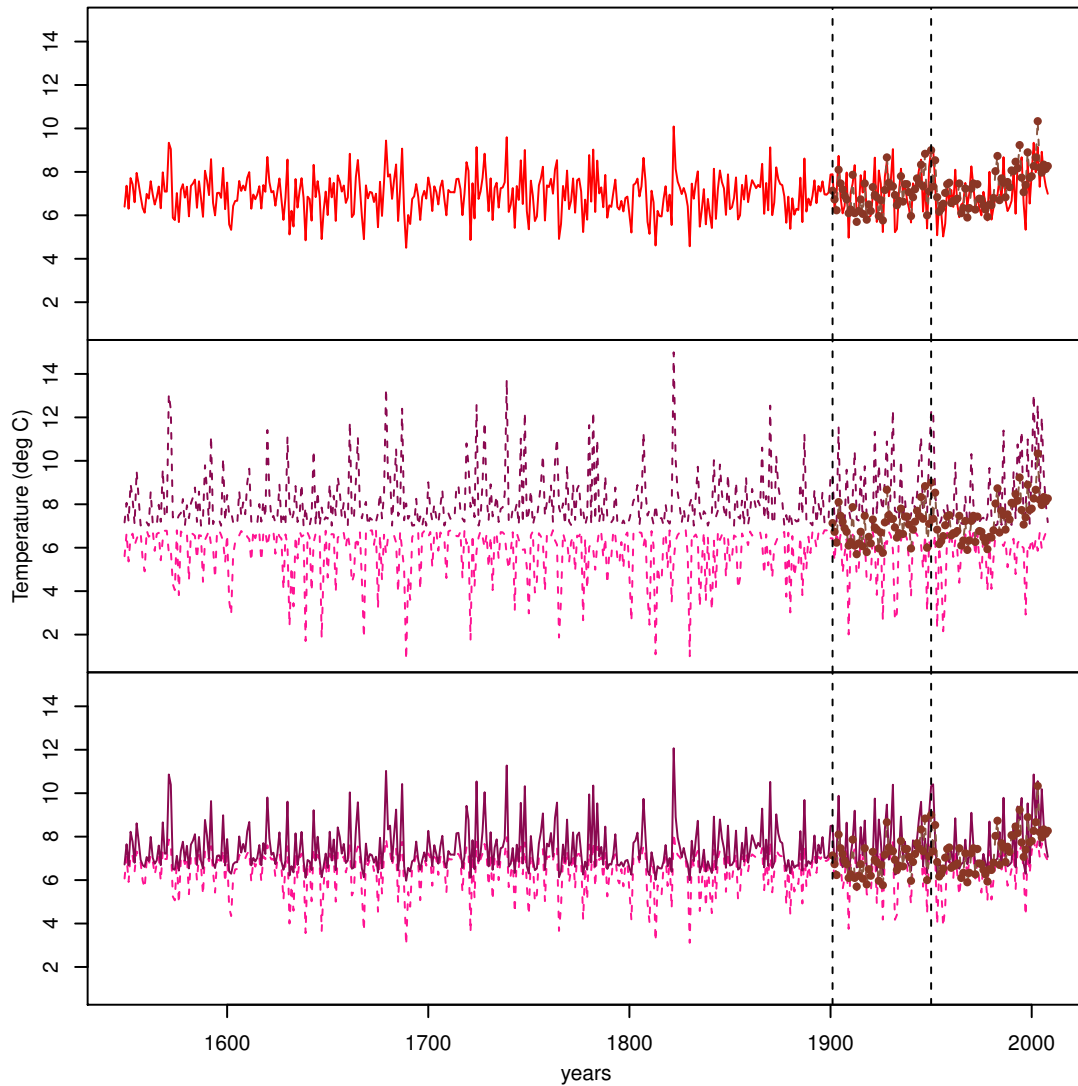
Figure B.45: Site A: temperature series. Estimates obtained by the direct regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of purple. The vertical lines delimit the calibration period.
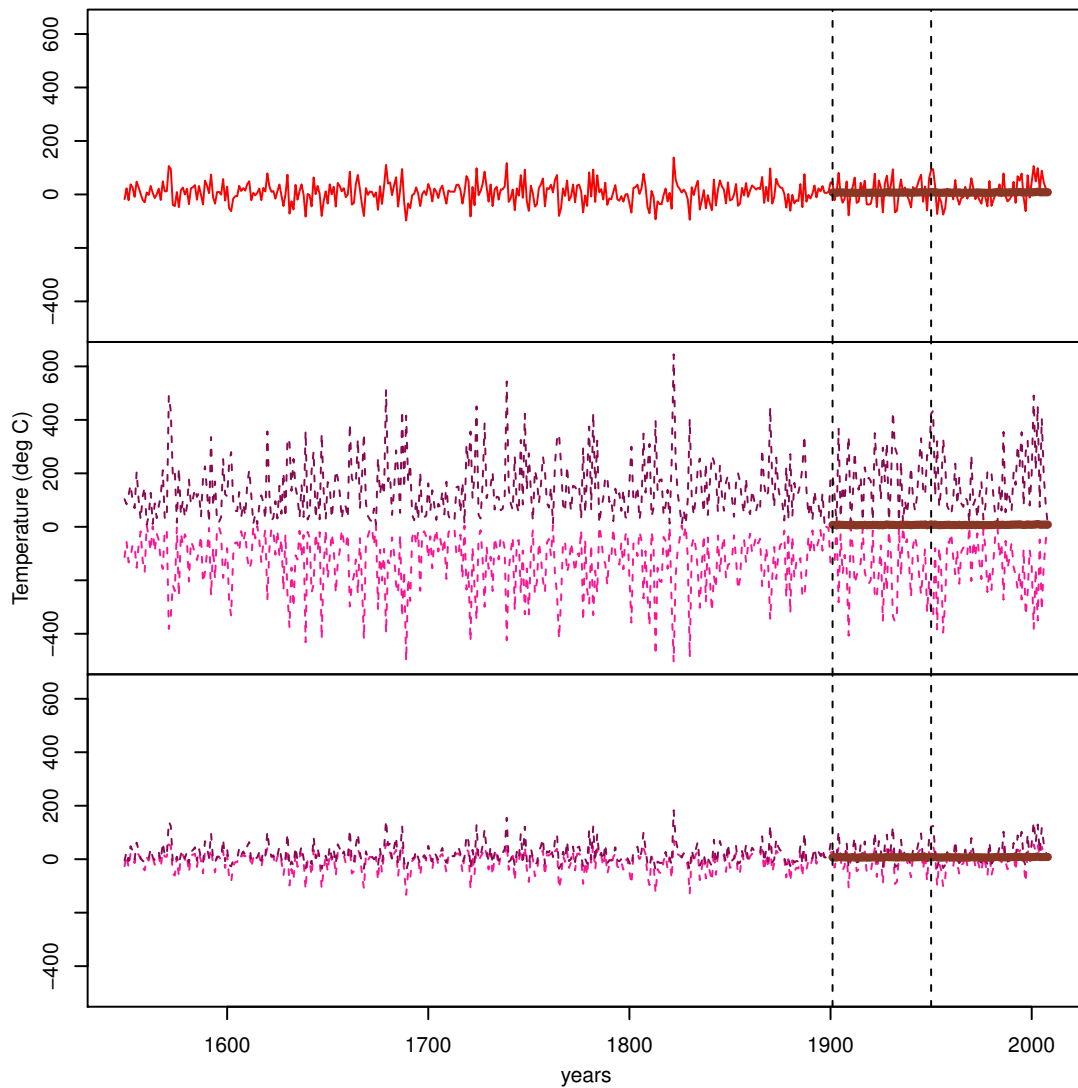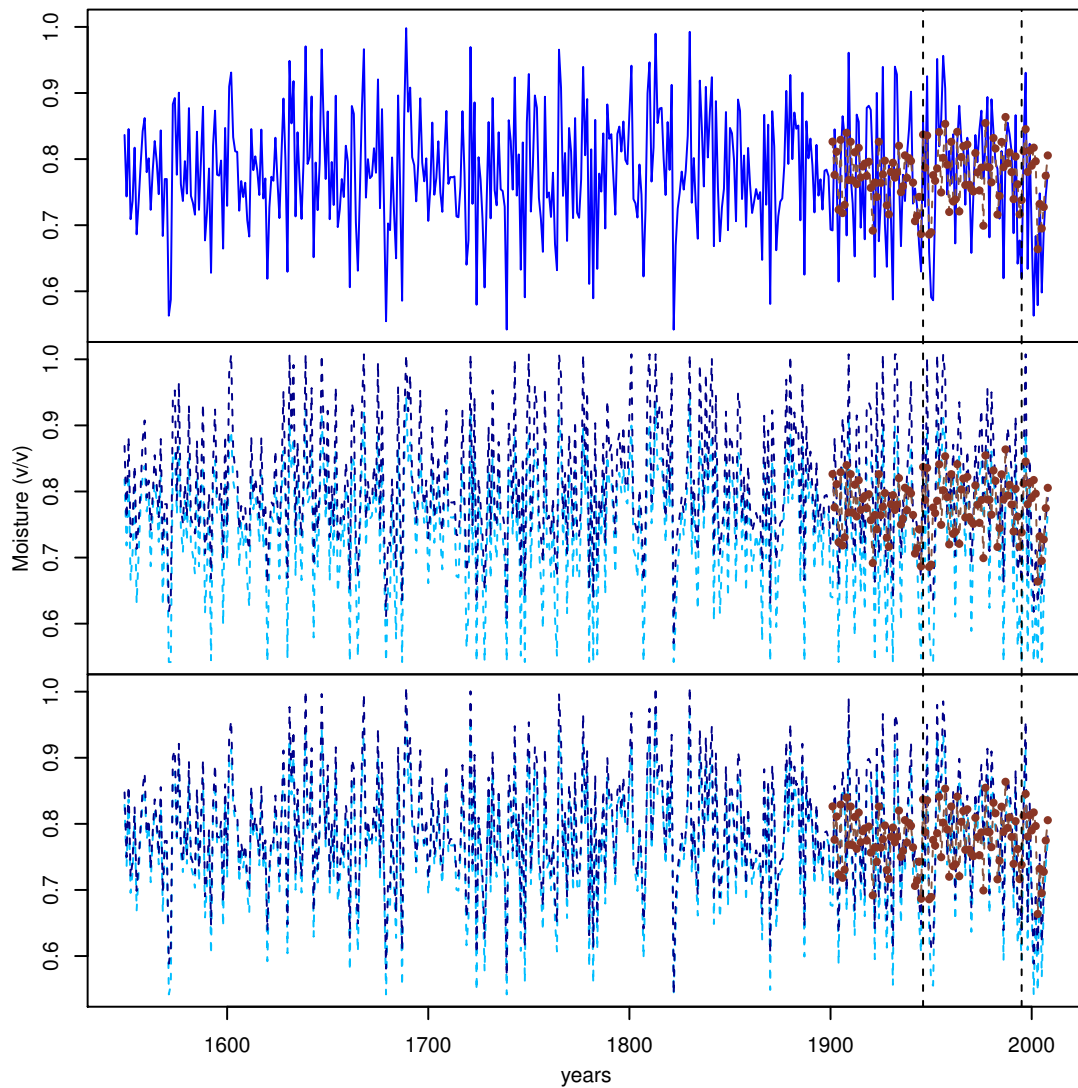
Figure B.46: Site A: temperature series. Estimates obtained by the inverse regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of purple. The vertical lines delimit the calibration period.
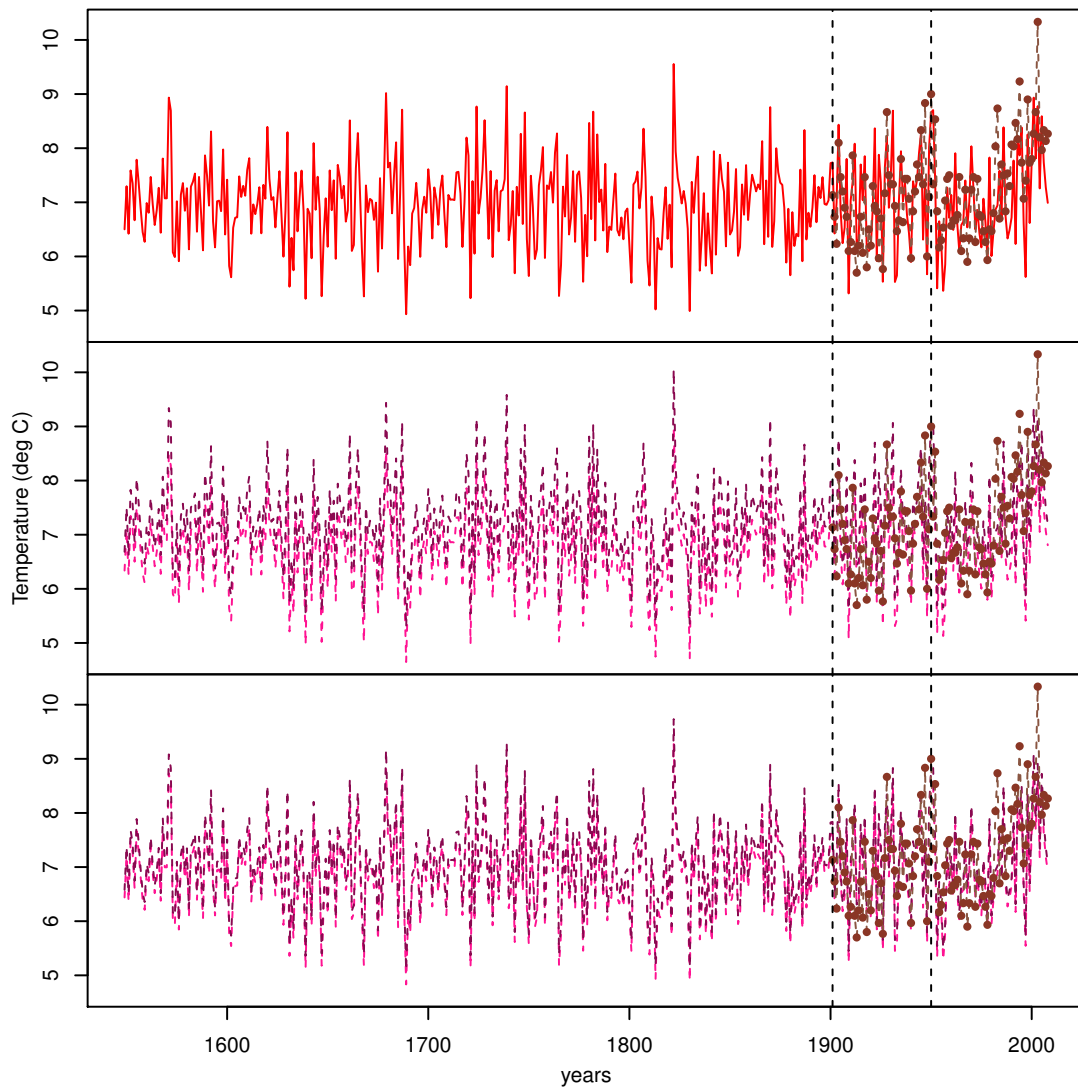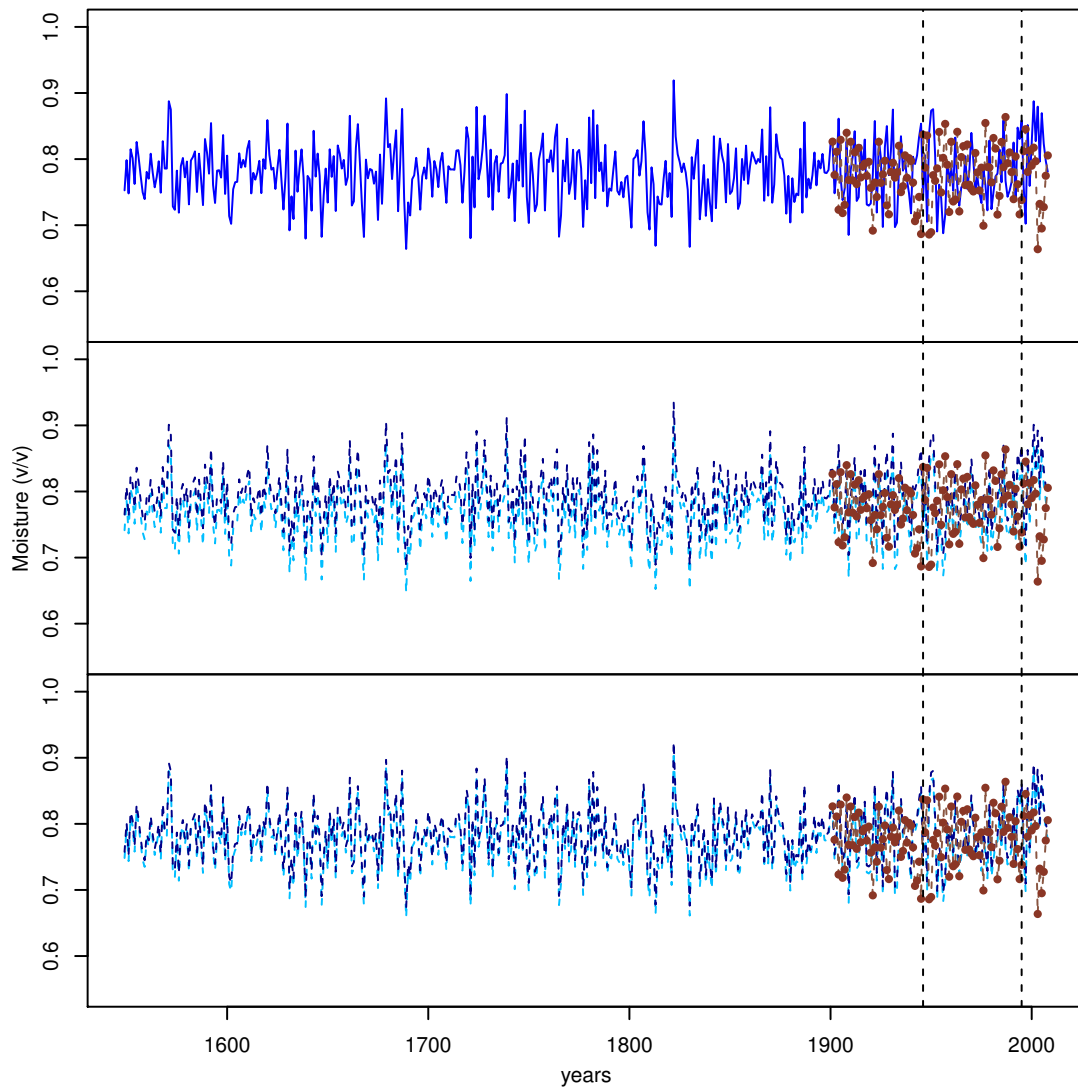
Figure B.47: Site A: temperature series. Estimates obtained by the variance matching algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of purple. The vertical lines delimit the calibration period.
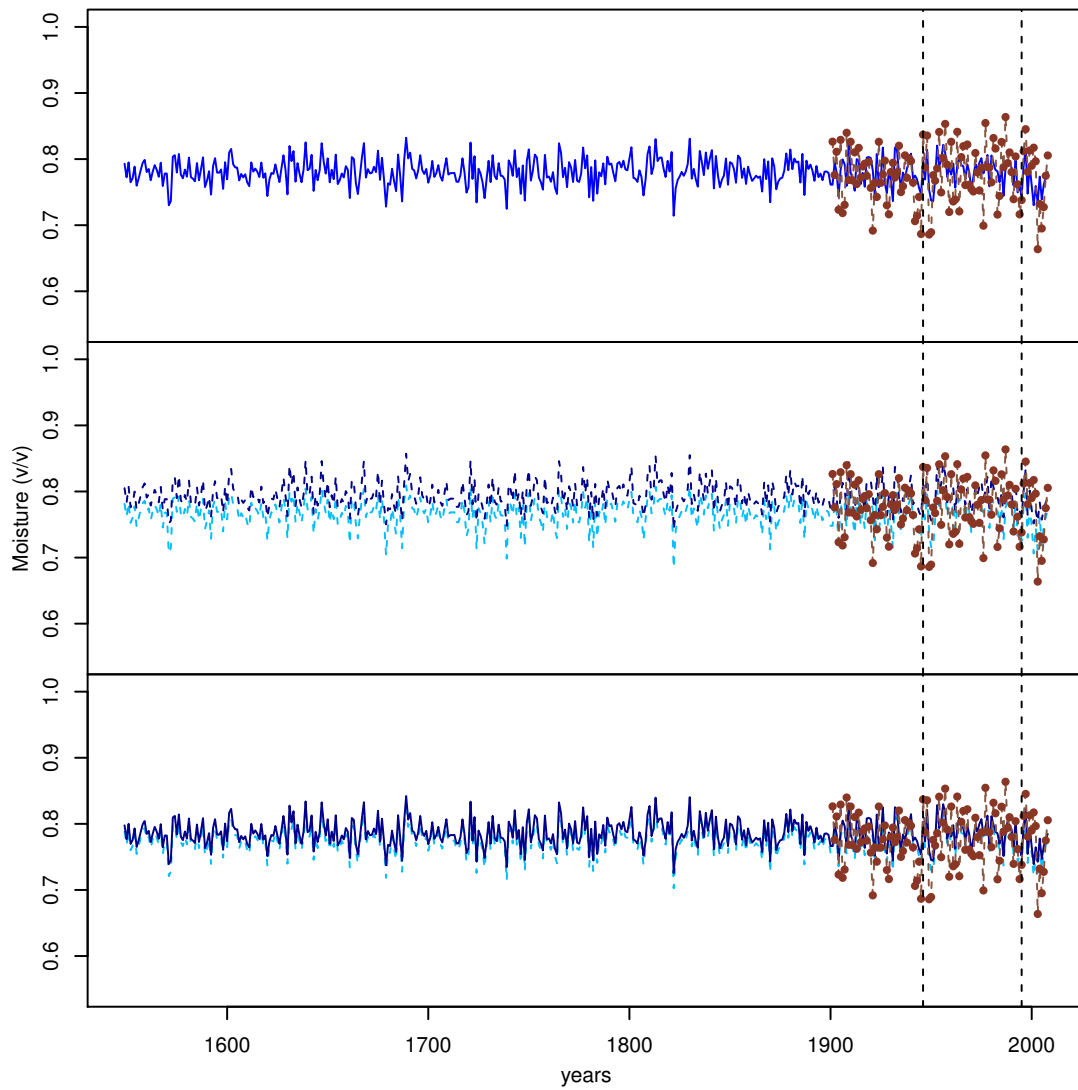
Figure B.48: Site A: moisture series. Estimates obtained by the direct regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of blue. The vertical lines delimit the calibration period.
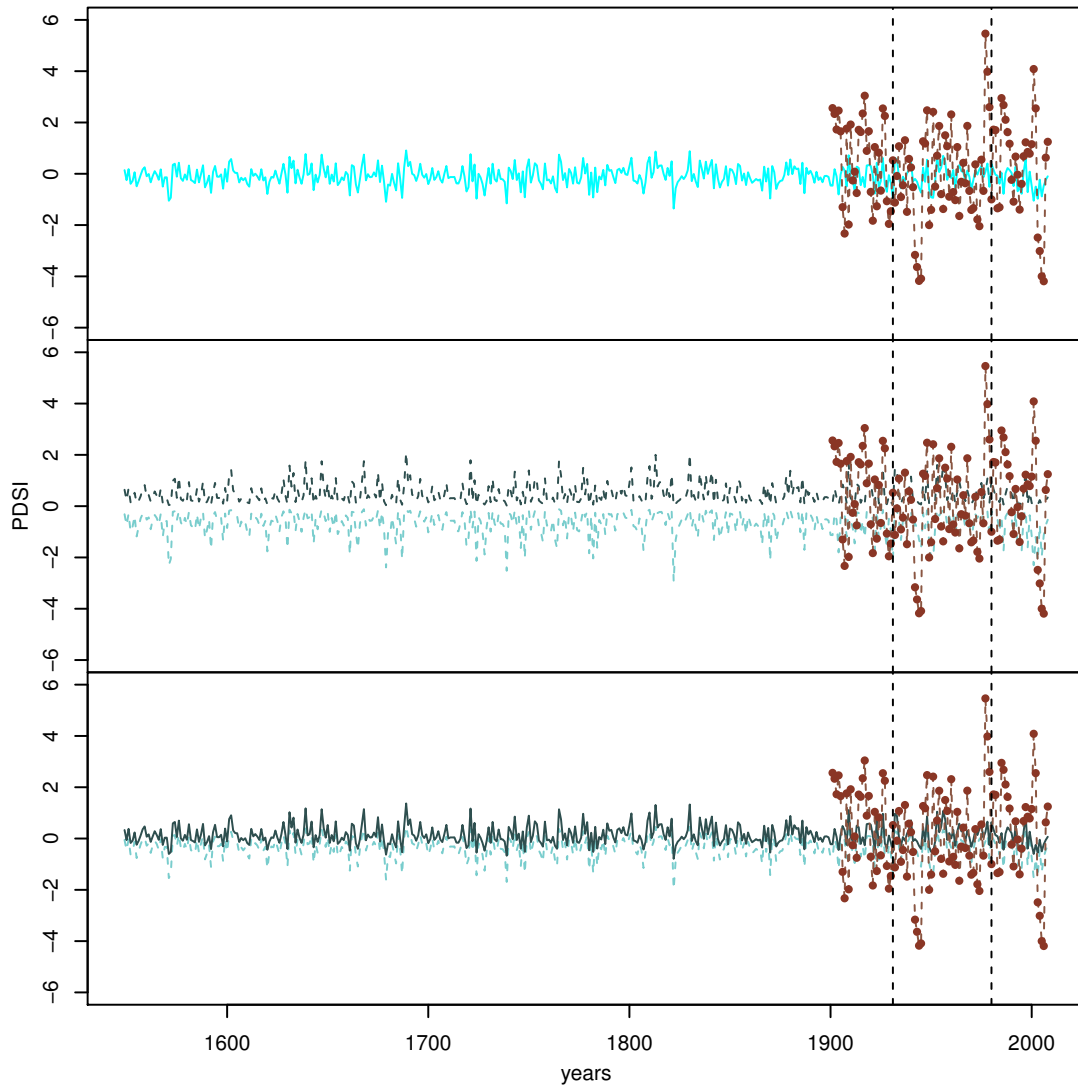
Figure B.49: Site A: moisture series. Estimates obtained by the inverse regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of blue. The vertical lines delimit the calibration period.
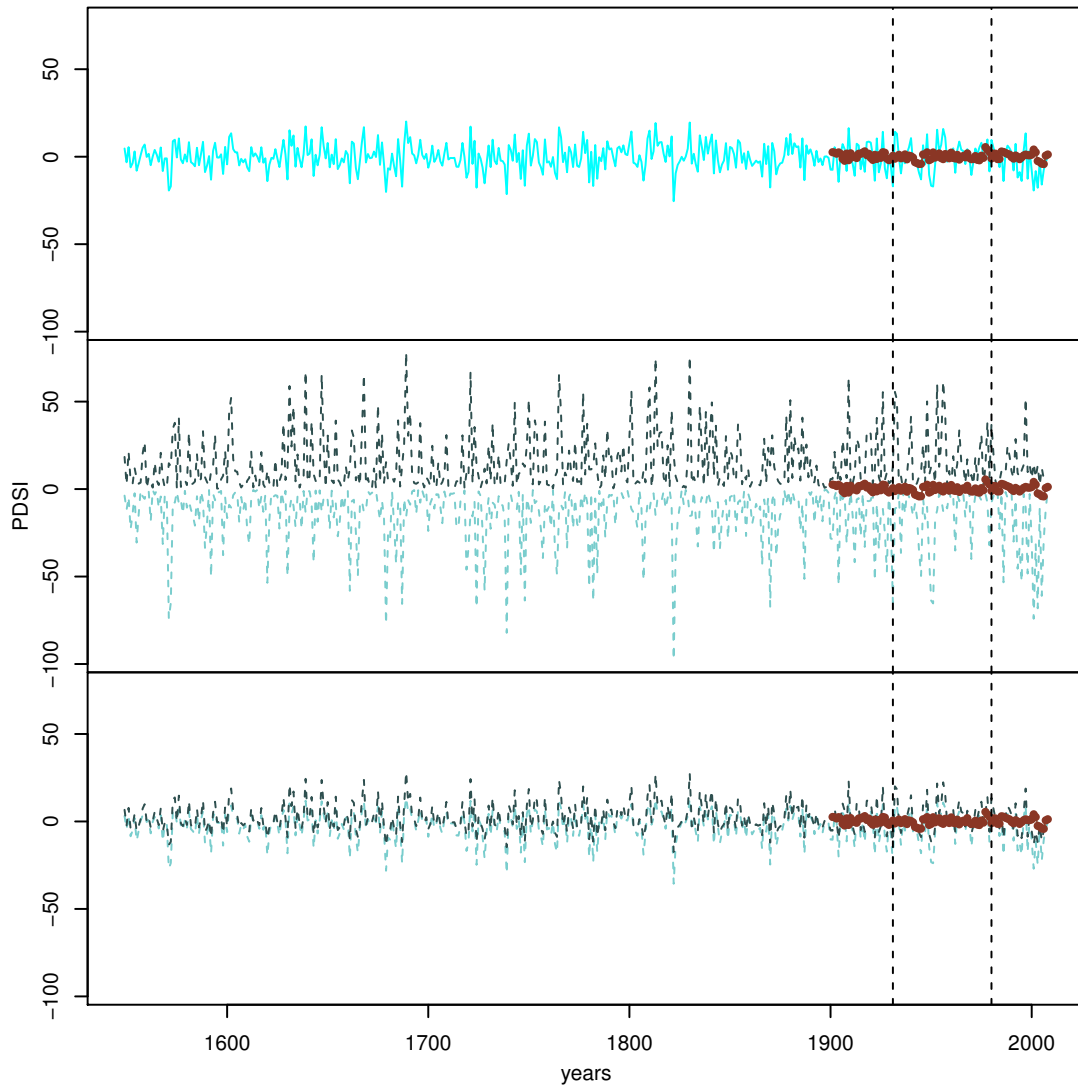
Figure B.50: Site A: moisture series. Estimates obtained by the variance matching algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of blue. The vertical lines delimit the calibration period.

Figure B.51: Site A: PDSI series. Estimates obtained by the direct regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of cyan. The vertical lines delimit the calibration period.
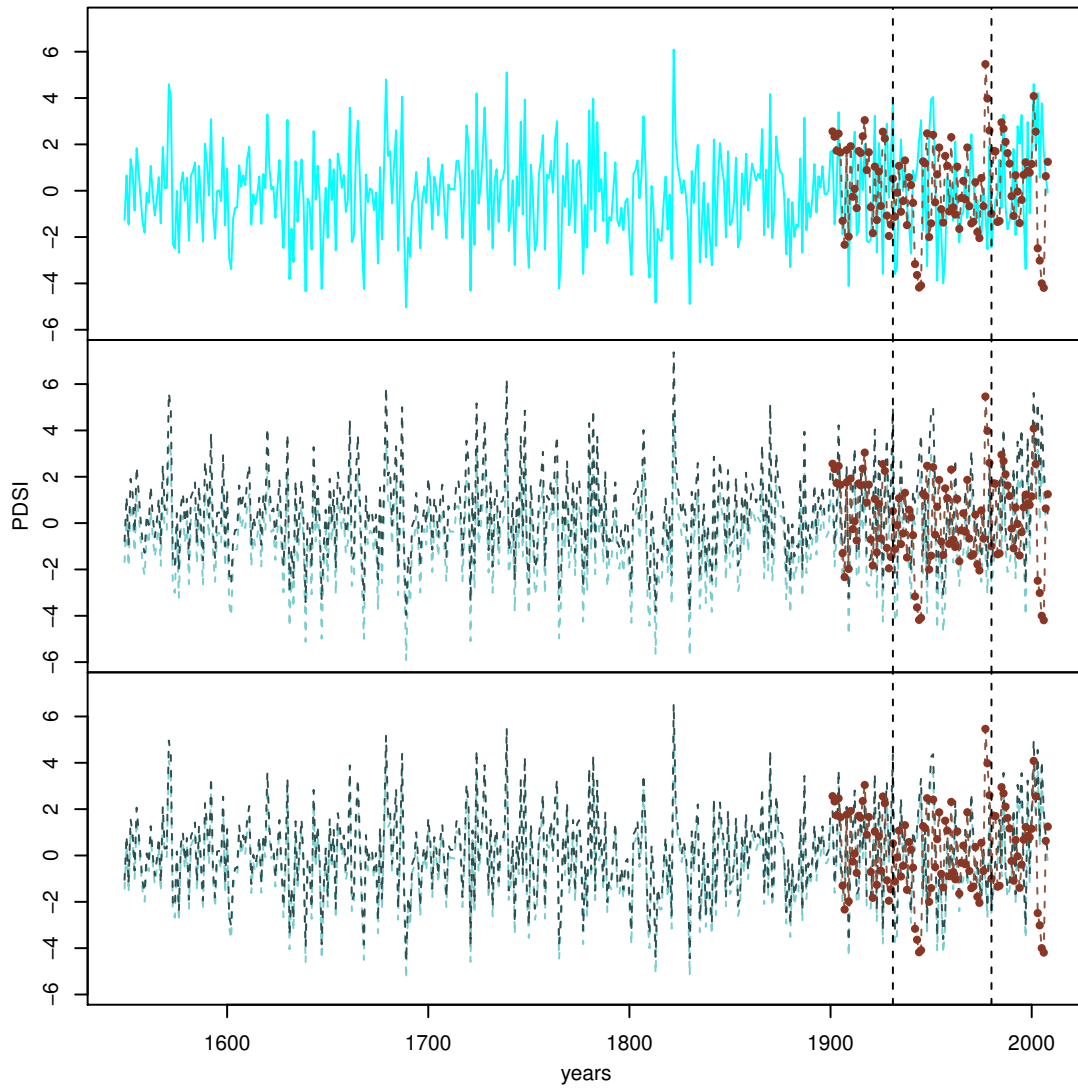
Figure B.52: Site A: PDSI series. Estimates obtained by the inverse regression algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of cyan. The vertical lines delimit the calibration period.
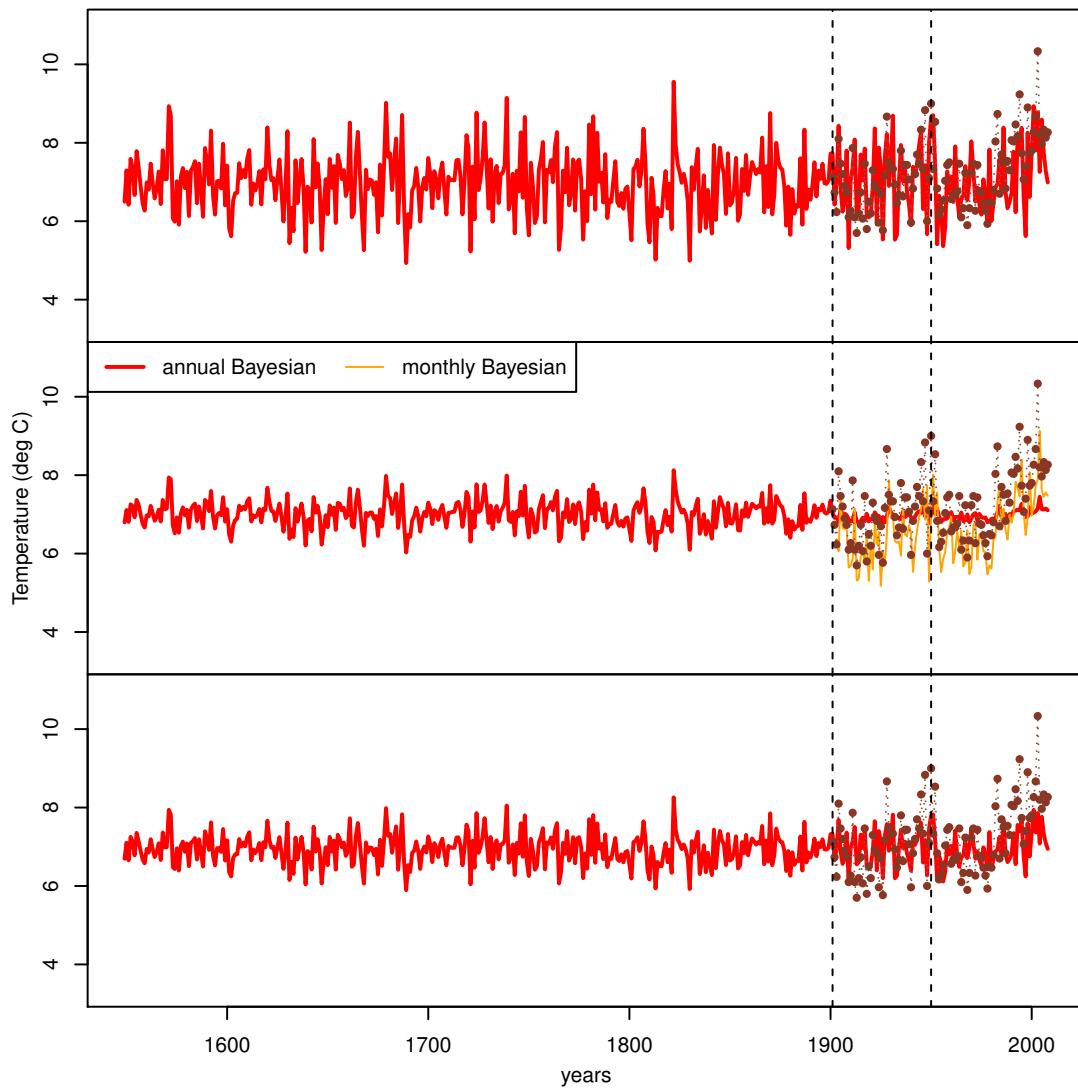
Figure B.53: Site A: PDSI series. Estimates obtained by the variance matching algorithm in terms of mean (top panel) 90% interval (middle panel) 50% interval (bottom panel). The observed series is plotted in brown while the intervals' boundaries are in different shades of cyan. The vertical lines delimit the calibration period.

Figure B.54: Site A: temperature series. Comparison between the estimates obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
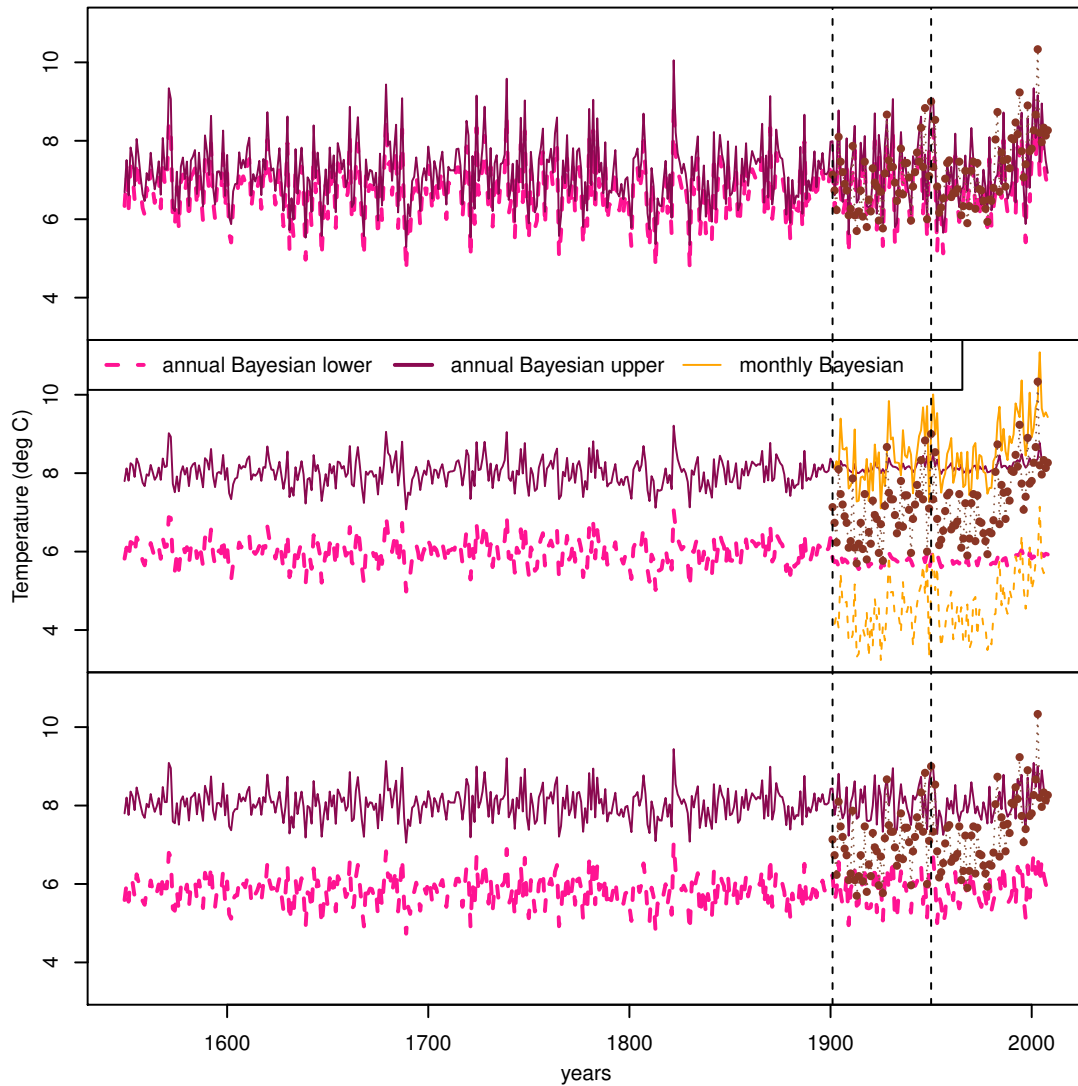
Figure B.55: Site A: temperature series. Comparison between the 90% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
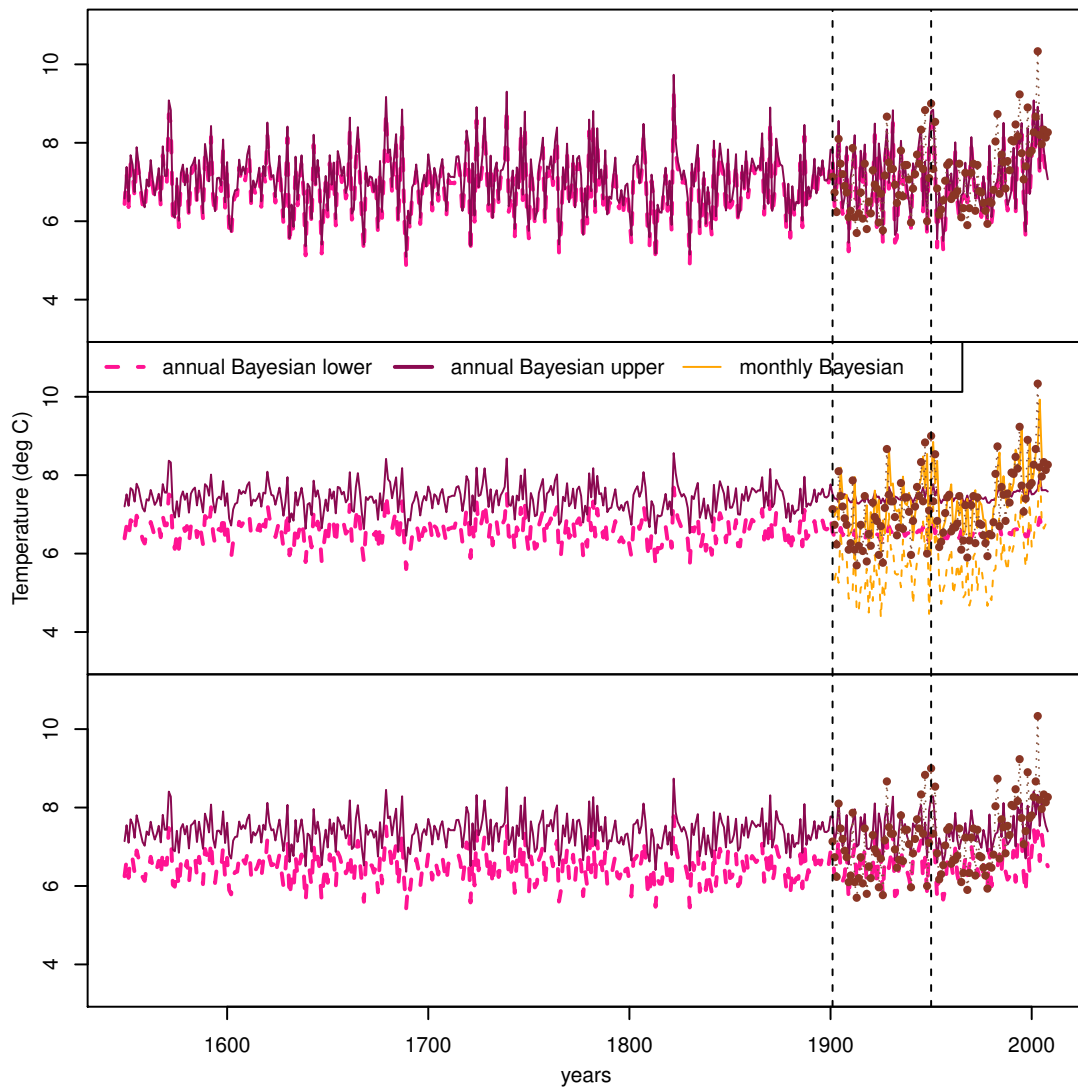
Figure B.56: Site A: temperature series. Comparison between the 50% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
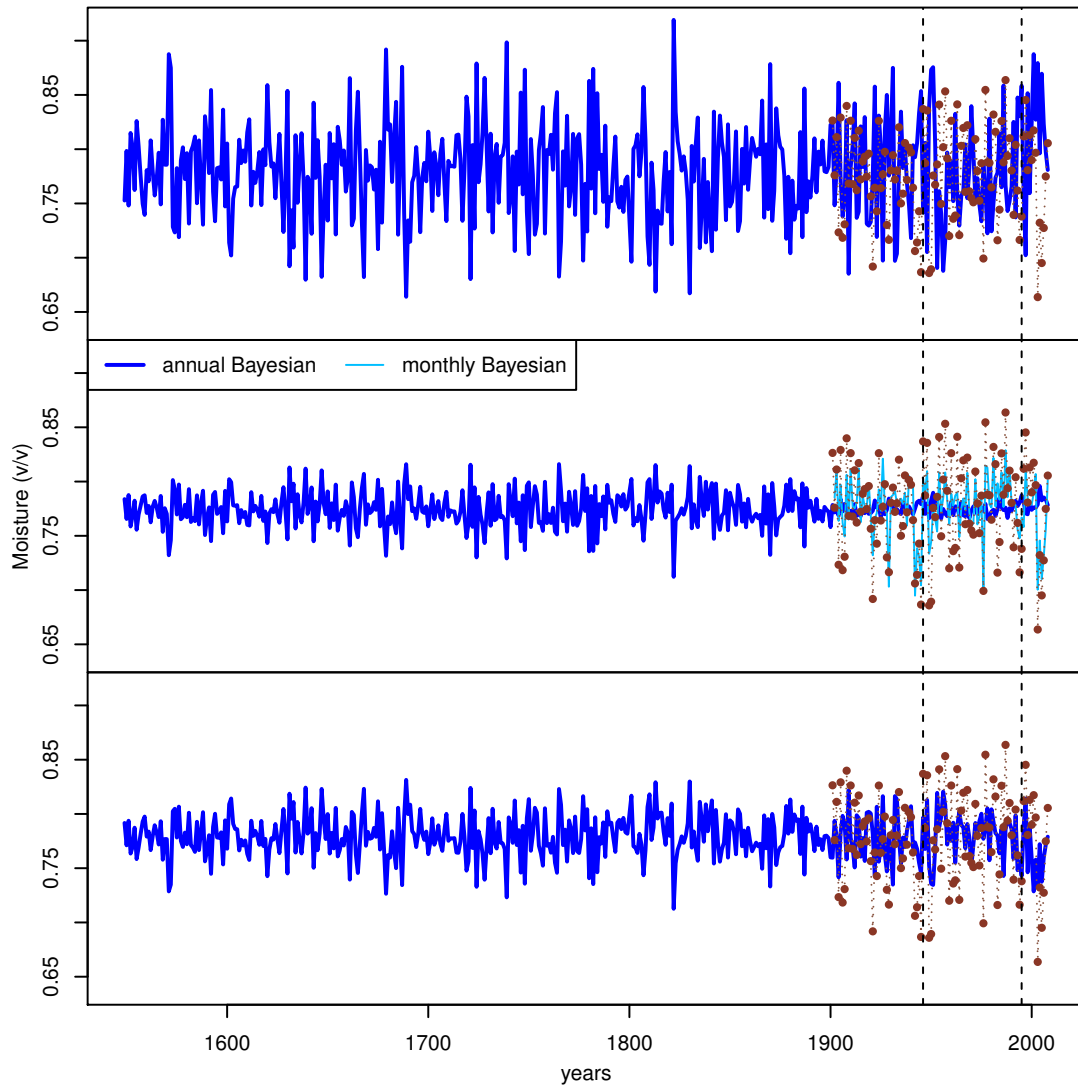
Figure B.57: Site A: moisture series. Comparison between the estimates obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
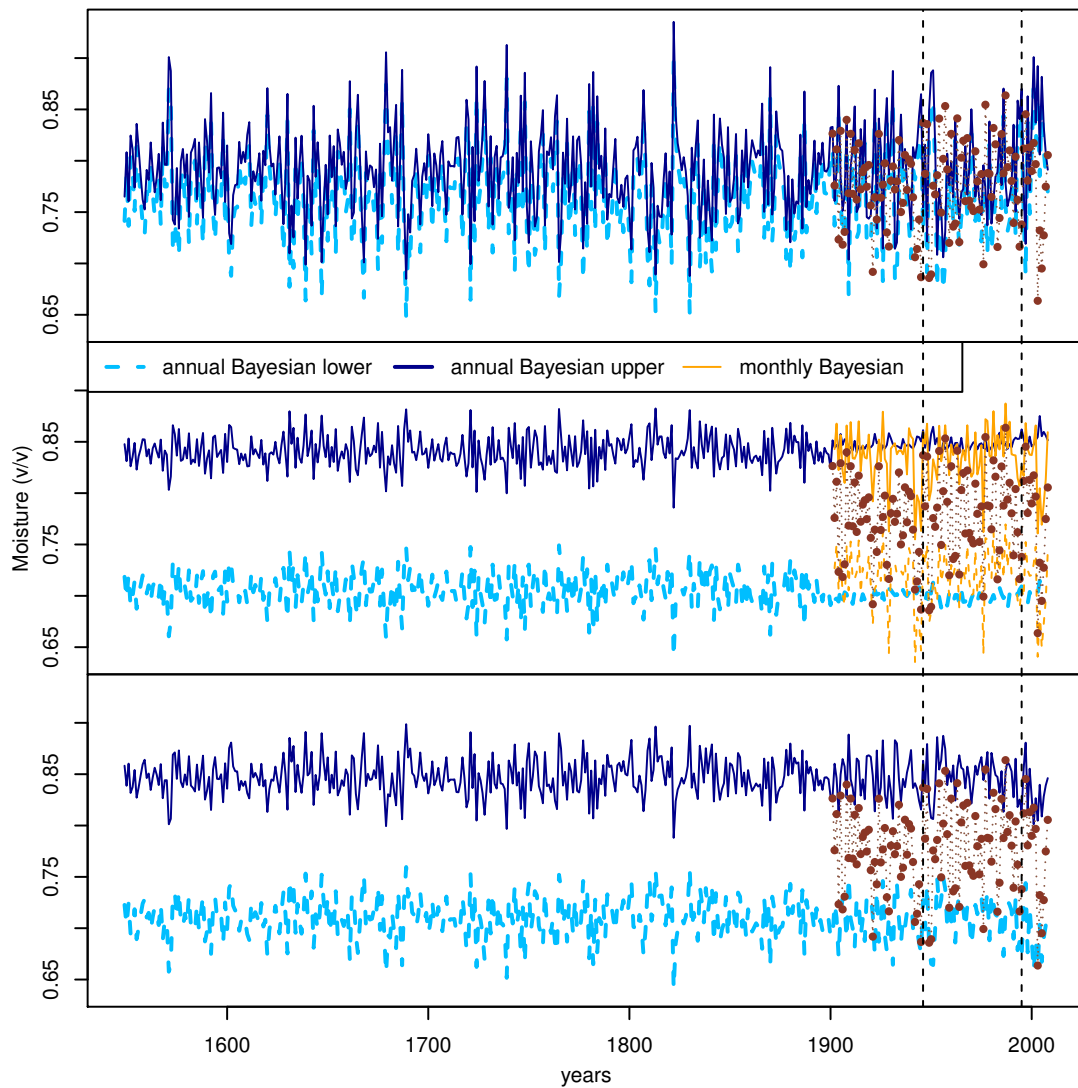
Figure B.58: Site A: moisture series. Comparison between the 90% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
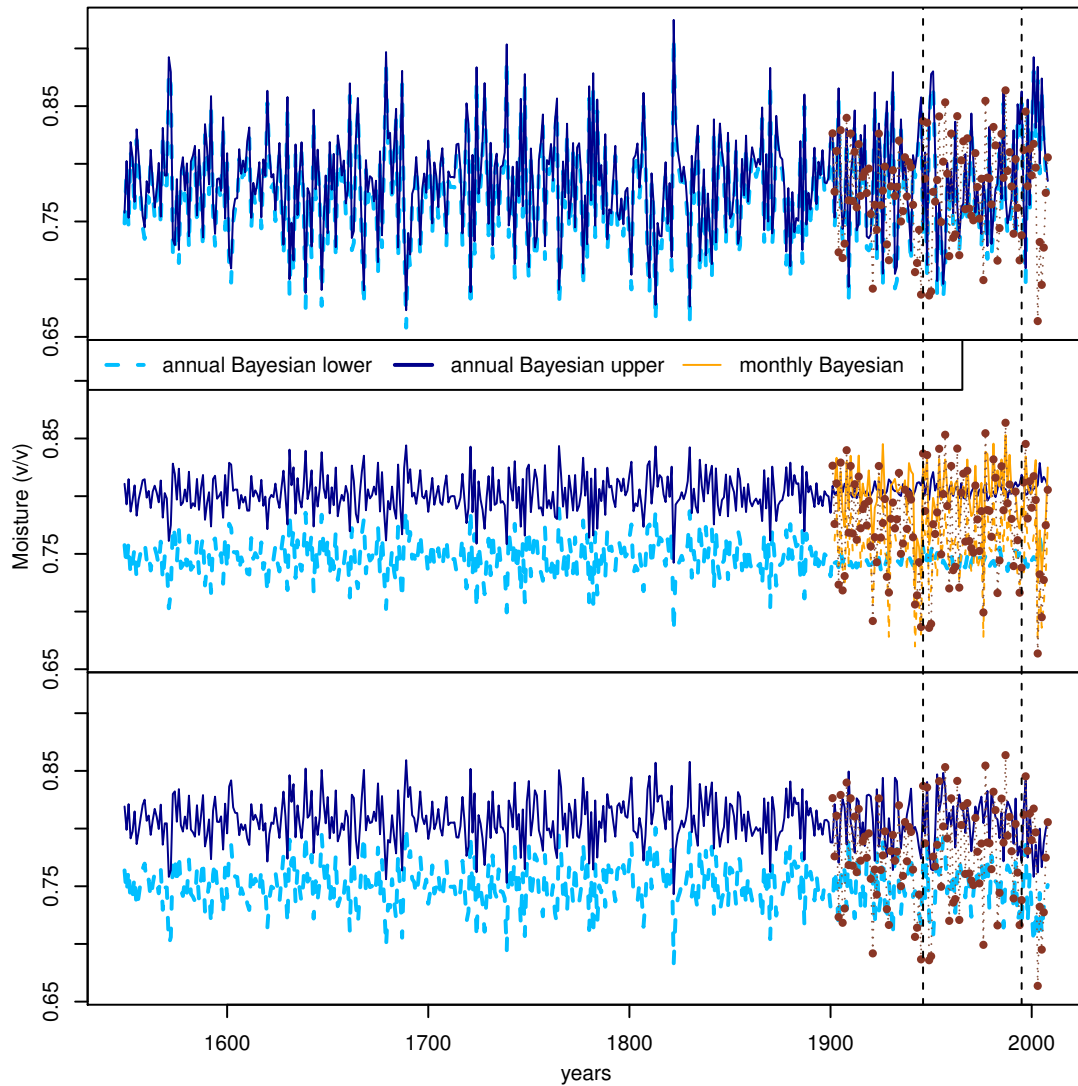
Figure B.59: Site A: moisture series. Comparison between the 50% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
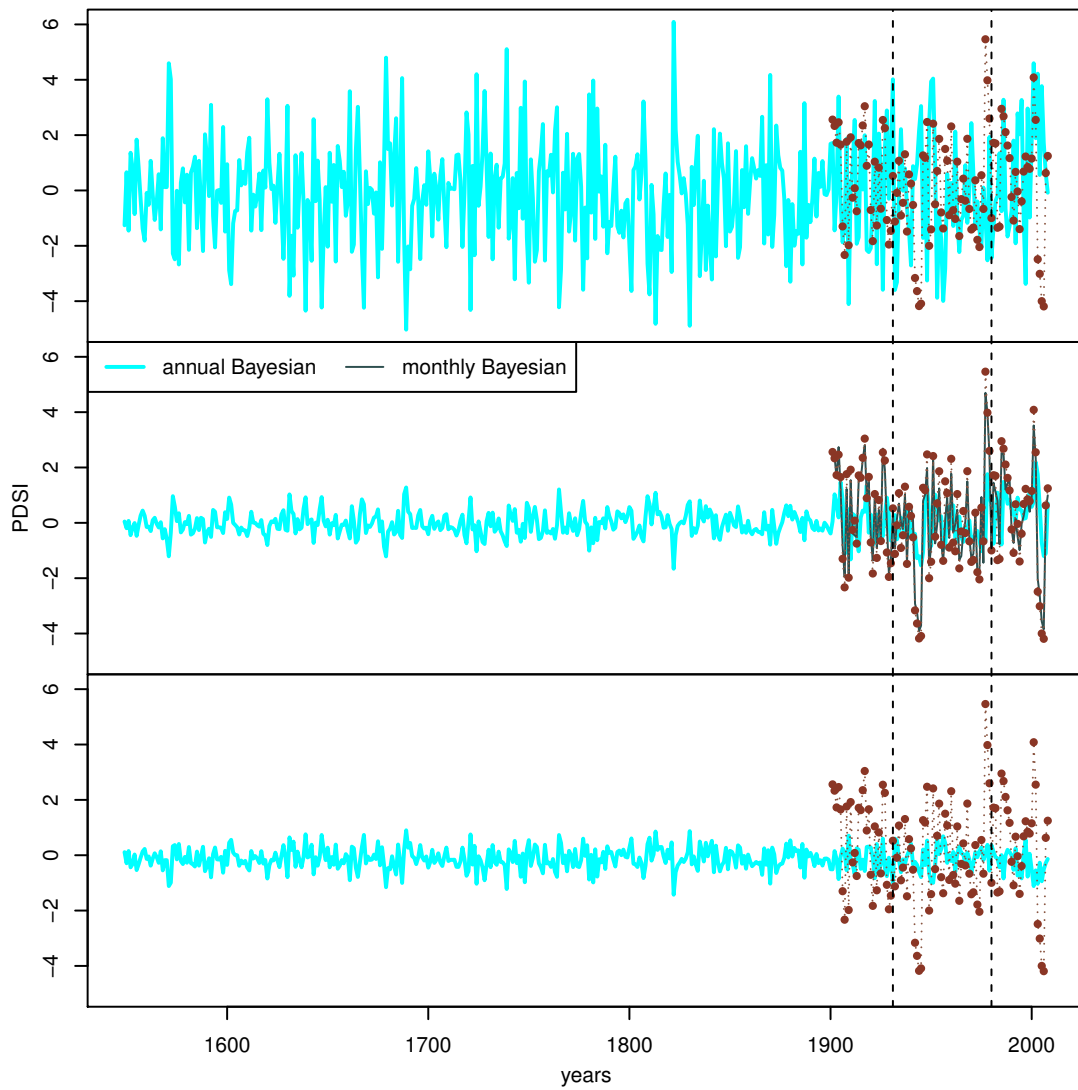
Figure B.60: Site A: PDSI series. Comparison between the estimates obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
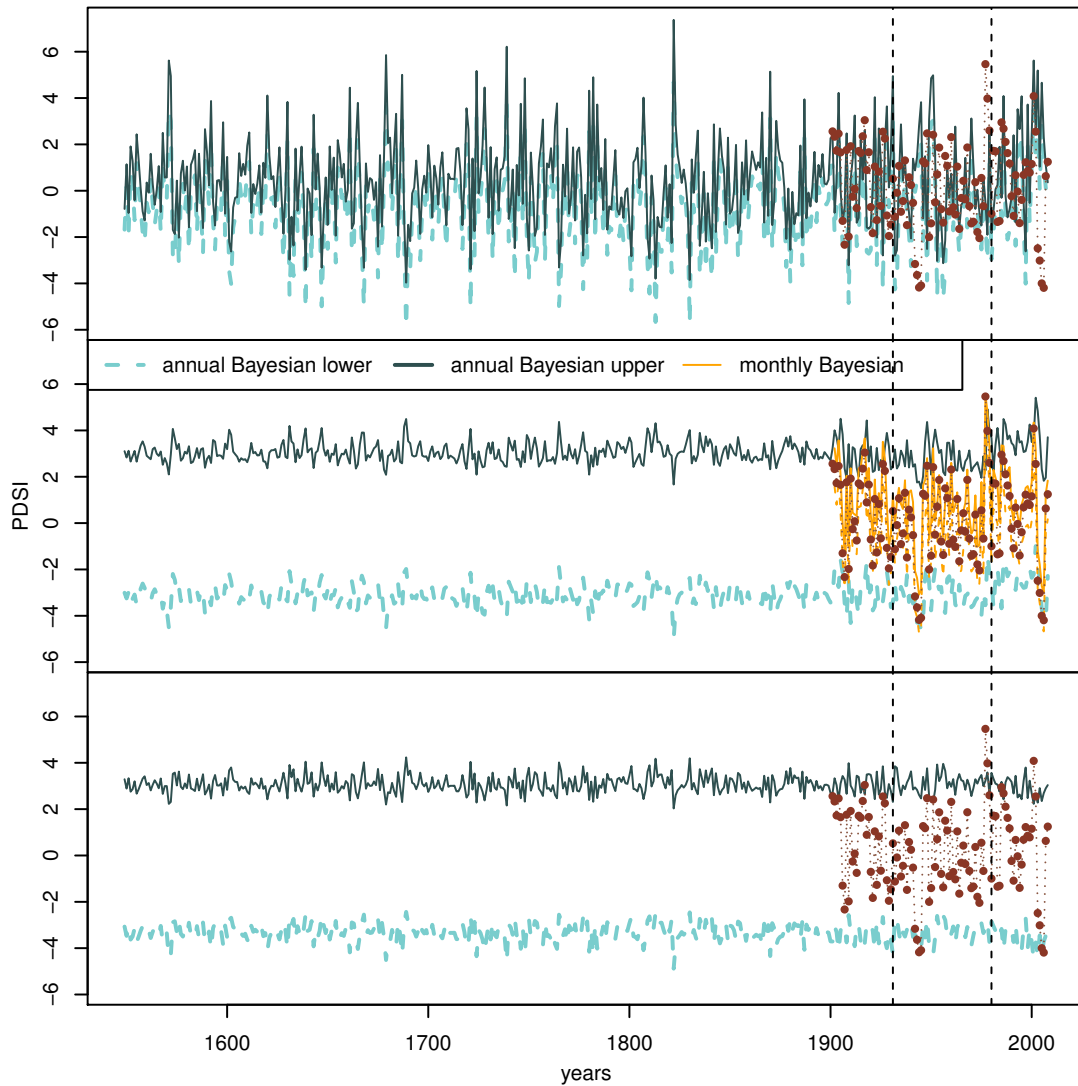
Figure B.61: Site A: PDSI series. Comparison between the 90% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.

Figure B.62: Site A: PDSI series. Comparison between the 50% intervals obtained by the variance matching algorithm (top panel), the Bayesian algorithms (middle panel) and by linear regression (bottom panel). The observed series is plotted in brown and the vertical lines delimit the calibration period.
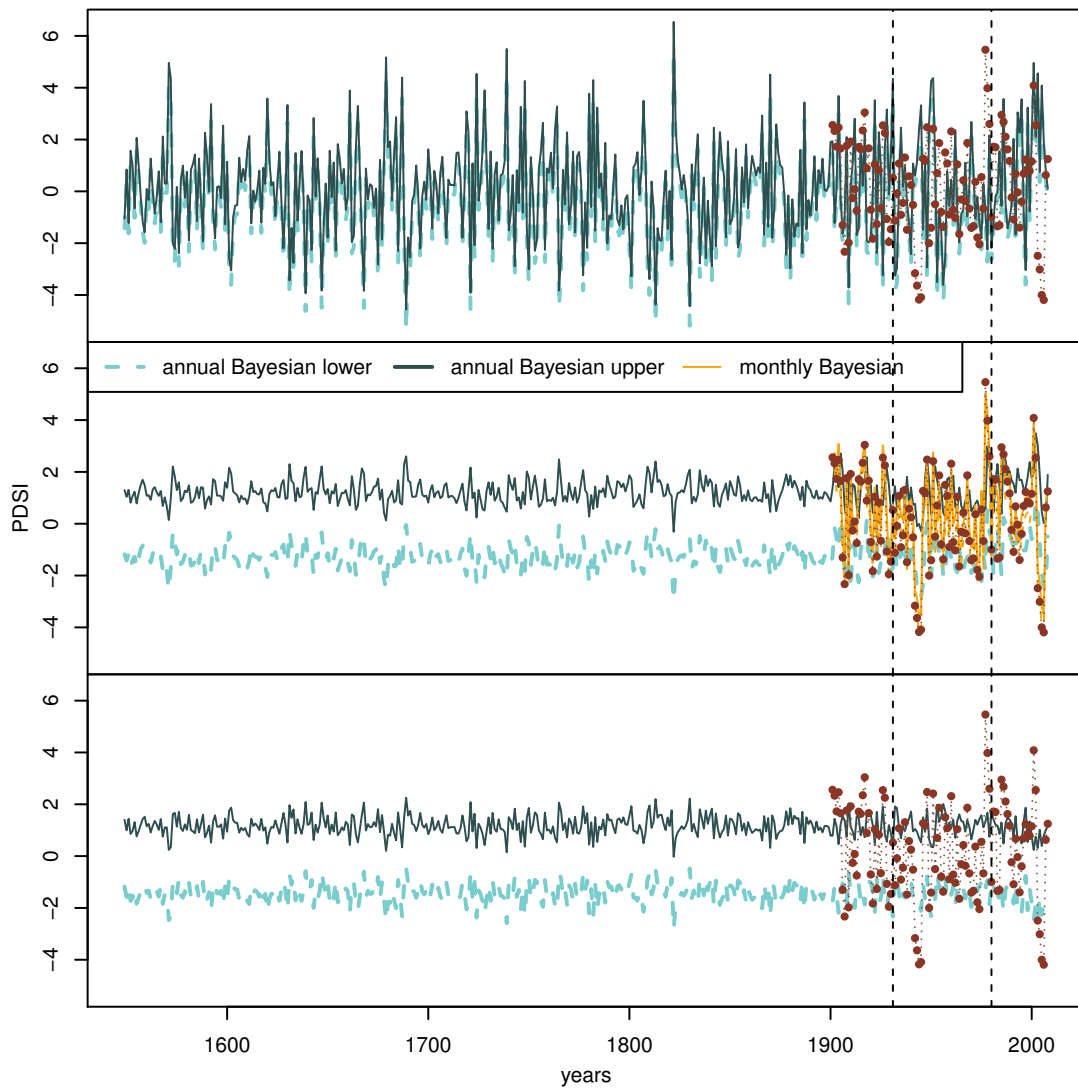
References

Ahmed, M., Anchukaitis, K. J., Asrat, A., Borgaonkar, H. P., Braida, M., Buckley, B. M., Buntgen, U., Chase, B. M., Christie, D. A., Cook, E. R., Curran, M. A. J., Diaz, H. F., Esper, J., Fan, Z.-X., Gaire, N. P., Ge, Q., Gergis, J., Gonzalez-Rouco, J. F., Goosse, H., Grab, S. W., Graham, N., Graham, R., Grosjean, M., Hanhijarvi, S. T., Kaufman, D. S., Kiefer, T., Kimura, K., Korhola, A. A., Krusic, P. J., Lara, A., Lzine, A.-M., Ljungqvist, F. C., Lorrey, A. M., Luterbacher, J., Masson-Delmotte, V., McCarroll, D., McConnell, J. R., McKay, N. P., Morales, M. S., Moy, A. D., Mulvaney, R., Mundo, I. A., Nakatsuka, T., Nash, D. J., Neukom, R., Nicholson, S. E., Oerter, H., Palmer, J. G., Phipps, S. J., Prieto, M. R., Rivera, A., Sano, M., Severi, M., Shanahan, T. M., Shao, X., Shi, F., Sigl, M., Smerdon, J. E., Solomina, O. N., Steig, E. J., Stenni, B., Thamban, M., Trouet, V., Turney, C. S., Umer, M., van Ommen, T., Verschuren, D., Viau, A. E., Villalba, R., Vinther, B. M., von Gunten, L., Wagner, S., Wahl, E. R., Wanner, H., Werner, J. P., White, J. W., Yasue, K., and Zorita, E. (2013). Continental-scale temperature variability during the past two millennia. *Nature Geoscience*, 6(5):339–346.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis.* Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 3. ed edition.

Barboza, L., Li, B., Tingley, M. P., and Viens, F. G. (2014). Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models. *The Annals of Applied Statistics*, 8(4):1966–2001.

Bradley, R. S. (2013). *Paleoclimatology: Reconstructing Climates of the Quaternary.* Academic Press.

Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541.

Coppola, A., Leonelli, G., Salvatore, M. C., Pelfini, M., and Baroni, C. (2013). Tree-ring-based summer mean temperature variations in the Adamello-Presanella Group (Italian Central Alps), 1610-2008 AD. *Climate of the Past*, 9(1):211–221.

Dai, A., Trenberth, K. E., and Qian, T. (2004). A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology*, 5(6):1117–1130.

Davison, A. C. (2003). *Statistical Models.* Cambridge University Press, Leiden.

Fritts, H. C. (1976). *Tree rings and climate.* Academic Press, London ; New York.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference.* Number 68 in Texts in statistical science series. Chapman and Hall/CRC, Boca Raton, second edition.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis.* Statistical Science. Chapman and Hall/CRC, third edition.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo In Practice.* (eds Gilks W. R., Richardson S. and Spiegelhalter D. J.) Chapman & Hall.

Harris, I., Jones, P., Osborn, T., and Lister, D. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset: updated high-resolution grids of monthly climatic observations. *International Journal of Climatology*, 34(3):623–642.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer Texts in Statistics. Springer New York, New York, NY.

Huang, J., van den Dool, H. M., and Georgakakos, K. P. (1996). Analysis of Model Calculated Soil Moisture over the United States (1931–1993) and Applications to Long-Range Temperature Forecasts. *Journal of Climate*, 9:1350–1362.

IPCC (2014). *Climate change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), IPCC, Geneva, Switzerland, 151 pp.

Lee, T. C. K., Zwiers, F. W., and Tsao, M. (2008). Evaluation of proxy-based millennial reconstruction methods. *Climate Dynamics*, 31(2-3):263–281.

Li, B., Nychka, D. W., and Ammann, C. M. (2010). The Value of Multiproxy Reconstruction of Past Climate. *Journal of the American Statistical Association*, 105(491):883–895.

Mann, M. E., Bradley, R. S., and Hughes, M. K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392(6678):779–787.

McShane, B. B. and Wyner, A. J. (2011). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *The Annals of Applied Statistics*, 5(1):5–44.

Schofield, M. R., Barker, R. J., Gelman, A., Cook, E. R., and Briffa, K. R. (2016). A Model-Based Approach to Climate Reconstruction Using Tree-Ring Data. *Journal of the American Statistical Association*, 111(513):93–106.

Tingley, M. P., Craigmile, P. F., 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.

Tolwinski-Ward, S. E. (2015). VSLiteR package, https://github.com/suztolwinskiward/VSLiteR.

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J. (2011). An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dynamics*, 36(11-12):2419–2439.

Tolwinski-Ward, S. E., Tingley, M. P., Evans, M. N., Hughes, M. K., and Nychka, D. W. (2015). Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response. *Climate Dynamics*, 44(3-4):791–806.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

Elisa Carraro

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4174
e-mail: carraro@stat.unipd.it

Current Position

Since January 2014; (expected completion: September 2017)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Tree-ring based palaeoclimate reconstruction using a hierarchical Bayesian model.
Supervisor: Prof. Claudio Agostinelli, University of Trento.
Co-supervisor: Dr. Luigi Spezia, Biomathematics & Statistics Scotland, The James Hutton Institute, Aberdeen, Scotland.

Research interests

- Bayesian statistics
- Time series analysis
- Application of statistics to palaeoclimatology

Education

December 2005 – July 2008
Master (*laurea specialistica/magistrale*) degree in Statistica e sistemi informativi aziendali.
University of Venice, Faculty of Economics
Title of dissertation: "Classificazione di dati funzionali con applicazione al monitoraggio della qualita' dell'acqua nella laguna di Venezia"
Supervisor: Prof. Stefano F. Tonellato
Final mark: 110/110 e lode

September 2002 – November 2005
Bachelor degree (*laurea triennale*) in Statistica e informatica per la gestione delle imprese.
University of Venice, Faculty of Economics
Title of dissertation: "La classe dei modelli lineari generalizzati: aspetti teorici e contesti applicativi"
Supervisor: Prof. Stefano F. Tonellato
Final mark: 110/110 e lode.

Visiting periods

December 2015 – February 2016

Biomathematics and Statistics Scotland,
Aberdeen, Scotland .
Supervisor: Dr. Luigi Spezia

March 2015 – June 2015
University College Dublin School of Mathematical Sciences,
Dublin, Ireland.
Supervisor: Prof. Andrew Parnell

Awards and Scholarship

2014
PhD scolarship granted by Fondazione CARIPARO.

Computer skills

- Programming: R, Fortran77, HTML
- Operating systems: Linux, Microsoft Windows.
- Software: LaTeX 2_ε, MS Office, OpenOffice, ECDL Start licence.

Language skills

Italian: native; English: good (written) good (spoken); C1 level in listening and reading skills certified by University of Padova.

Publications

Conference abstracts
Carraro, E. (2016). Tree-ring based climate reconstruction using a hierarchical Bayesian model. In S. Cabras, M. Guindani, *Book of Abstracts*, page 269. International Society for Bayesian Analysis World Meeting, Cagliari, 13–17 June 2016. Cooperativa Universitaria Editrice Cagliaritana. ISBN 9-788884-679833.

Conference presentations

Carraro, E. (2016). Tree-ring based Italian palaeoclimate reconstructions using a hierarchical Bayesian model. (oral) , Venice, Italy, 7th October.

Carraro, E. (2016). Tree-ring based palaeoclimate reconstruction using a hierarchical Bayesian model. (oral) *Applied Bayesian statistics school*, Como, Italy, 30th August.

Carraro, E. (2016). Tree-ring based palaeoclimate reconstruction using a hierarchical Bayesian model. (poster) *International society for Bayesian Analysis world meeting*, Cagliari, Italy, 12th-17 June.

Carraro, E. (2016). Tree-ring based climate reconstruction using a hierarchical Bayesian model. (oral) , Aberdeen, Scotland, 15th January.

Carraro, E. (2015). Reconstructing temperatures for the past centuries. A review of some proposals. (oral) , Venice, Italy, 13th February.

Teaching experience

November 2013
Course name: Probability
Degree: PhD
Teaching task (exercises evaluation), total number of hours 10
Institution: University of Venice Ca' Foscari
Instructor: Prof. Stefano F. Tonellato

Other Interests

Piano playing, choral singing.
Volunteering.

References

Prof. Claudio Agostinelli
University of Trento
Department of Mathematics
via Sommarive 14, 38123 Povo (TN)
Phone: +39 0461 283900
e-mail: claudio.agostinelli@unitn.it

Dr. Luigi Spezia
Biomathematics & Statistics Scotland
The James Hutton Institute, Craigiebuckler
Aberdeen, AB15 8QH, Scotland, UK
Phone: +44 (0)1224 395130
e-mail: luigi.spezia@bioss.ac.uk