



available at www.sciencedirect.com



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



Gibbs sampling of climatic trends and periodicities

N.T. Kottegoda *, L. Natale, E. Raiteri

Dipartimento di Ingegneria Idraulica e Ambientale, Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy

Received 1 March 2006; received in revised form 23 February 2007; accepted 16 March 2007

KEYWORDS

Climatic time series;
Trend;
Periodicity;
Bayesian methods;
Markov Chain Monte
Carlo;
Gibbs sampling

Summary Trends and periodic movements in climatic series are treated as non-stationary components. A time series model and Bayesian statistics are combined through a Markov chain Monte Carlo procedure. Gibbs sampling is used in the Monte Carlo application. Monthly series of river flow, rainfall and temperature from northern Italy are used. Some late temperature rises are noted, otherwise there are no systematic increases or decreases in the series. Changes in periodicity are also of a random nature. From the results it is also possible to compare these properties between different locations and climatic indicators.

© 2007 Elsevier B.V. All rights reserved.

Introduction

Numerous papers have been published on climatic change in hydrology and the atmospheric sciences in the recent past (e.g., Kite, 1989; Nash and Gleick, 1991; Grotch and MacCracken, 1991; Wilby et al., 1999; Mann et al., 2003; Soon et al., 2003). The most commonly used methods for the study of trend, in particular, and seasonal effects are devised to find global patterns of these properties in a time series. The general objective is to establish a mean trend line and an averaged periodicity component over the period of the record. Nevertheless, by using moving averages, moving windows, filters and the like it is possible to see local effects of the trend or other component. The Holt-Winters forecasting procedure, reviewed and compared with Box Jenkins forecasts by Chatfield (1978), is one of the earliest;

basically it generalizes simple exponential smoothing to model trend and periodicity.

New tools for smoothing climatic time series such as advanced spectral methods reviewed by Ghil et al. (2002) are adopted, for example, by Mann (2004). Although these methods are advantageous they have some limitations because by fixing the length of the window one cannot properly visualize temporal variations in properties such as trend in a time series. Long windows tend to cut off slow movements and short windows provide insufficient smoothing.

Of the sophisticated methods that have emerged during the past 20 years, wavelet analysis, devised for investigating localized variations of power (with reference to signals or other nonrandom behaviour) over a time series, has had much attention (see, e.g., the introduction by Boggess and Narcovich, 2001). Thousands of articles have been written in science and engineering and numerous conferences held on the subject; there are several web sites and software is available with diverse climatic data sets such as

* Corresponding author.

E-mail address: nathabandu.kottegoda@polimi.it (N.T. Kottegoda).

sea surface temperatures, the Southern Oscillation Index and monsoon rainfalls. On the other hand, [Baliunas et al. \(1997\)](#), applied standard wavelet and adaptive wavelet transform algorithms to Central England temperatures but found that estimates of a data trend are sensitive to the cut off time scale of the filter. Besides, problems arise such as those concerning the probability distribution of the data.

An alternative approach is to model trend and periodicity as non-stationary components in a time series formulation. Accordingly, one can apply a stochastic dynamic system model to the time series with the state functions incorporating trend and periodicity; here the term 'dynamic' refers to changes in a process with respect to time. Because the statistical properties such as the variances of the trend and periodic components are unknown, these aspects should be incorporated in the algorithm. Such a representative time series model can be adopted advantageously, that is by making the method general and usable, through a Bayesian estimation procedure. Because there are usually computational problems to surmount in a direct application, it seems that the only tractable way to implement this type of procedure is through a Markov chain Monte Carlo approach, referred to by the acronym MCMC. In this way a random walk model is applied to model the state variables dynamically using available data.

Bayes' theorem and MCMC sampling

Through the Bayesian framework one can combine initial knowledge about the parameters, such as trend and periodicity, of a model with observations and the model output. Hence the posterior distribution on the parameter space can be obtained.

For example, consider a time series of a climatic variable such as temperature. This can be expressed in the classical additive form by

$$y_i = t_i + p_i + \eta_i \quad (1)$$

in which at time i , y_i represents the observed value of the variable, t_i and p_i are the parameters of trend and periodicity, respectively, and η_i is a random component.

Our next objective is to update the current estimates of the parameters on the basis of new information, using Bayes' theorem. This can be summarized as follows. Let θ represent the unknown parameter space such as trend and periodicity in our model, the prior probability of which is denoted as $P(\theta)$ based on current knowledge of the process. After receipt of additional data y , the posterior distribution $P(\theta|y)$ is found from

$$P(\theta|y) = \frac{P(\theta)P(y|\theta)}{P(y)}$$

in which $P(y|\theta)$ is the likelihood function and $P(y)$ is a normalising quantity. In applications, however, problems usually arise. These happen because it is difficult to estimate the moments of the posterior distribution as shown here. Such computational difficulties can be overcome by means of Markov chain Monte Carlo (MCMC) sampling which is a simulation procedure of recent origin (see [Gilks et al., 1996](#)). The Metropolis–Hastings algorithm provides a general form of MCMC simulation developed by [Hastings](#)

(1970) following [Metropolis et al. \(1953\)](#). This is an iterative procedure, some practicalities of which are described by [Geyer \(1992\)](#) and [Brooks \(1998\)](#); [Chib and Greenberg \(1995\)](#) provide a clear description. The idea is to construct a Markov chain whose stationary and ergodic distribution is the posterior distribution from Bayes' theorem (usually intractable by analytical means). However, repeated simulations are required to reach the desired states of stationarity and ergodicity.

In the hydrological literature, initial applications of the Metropolis–Hastings algorithm were limited to assessing parameter uncertainty in conceptual rainfall-runoff modelling ([Kuczera and Parent, 1998](#); [Bates and Campbell, 2001](#); and [Marshall et al., 2004](#)). It is implied in all cases that the model is correct and the source of error arises from the estimation of parameters. See also subsequent work of [Efendiev et al. \(2005\)](#) and [Reis and Stedinger \(2005\)](#).

A particular type of MCMC algorithms is the Gibbs sampler. This is devised to enable one to solve practical statistical problems. The advantages are its simplicity and convergence to the stationary and ergodic states. The procedure was originated by [Geman and Geman \(1984\)](#) who used it in applying the Gibbs distribution in lattices, although it is now applied using various other distributions; a similar method was already in use in statistical physics under the name "the heat bath algorithm". The Gibbs sampler has been used in medicine, archaeology, and many other fields (see for example [Gelfand et al., 1990](#); [Gilks et al., 1993](#); and [Smith and Roberts, 1993](#)). However, let us be warned that some types of probability density functions or criteria or the fact that the model is not tractable enough warrant the use of the wide and general class of Metropolis–Hastings algorithms.

The procedure is to construct for a variable, say, x a Markov chain with stationary distribution $\pi(x)$. Under certain conditions the sample output from the chain can be used asymptotically to estimate the expected value with respect to $\pi(x)$ of a function, say $f(x)$ of interest; see [Casella and George \(1992\)](#), [Smith and Roberts \(1993\)](#), [Besag and Green \(1993\)](#) and [Tierney \(1994\)](#). Let $\pi(x) = \pi(x_1, \dots, x_n)$ denote a joint density and let $\pi(x_i | x_{-i})$ denote the conditional densities of each of the components x_i using values of the other components and let $x_{-i} = (x_j, j \neq i), i = 1, \dots, n$. We commence, at stage 0, by taking arbitrary starting values, $x^0 = (x_1^0, \dots, x_n^0)$. For the next stage, we draw randomly from conditional distributions $\pi(x_i | x_{-i}), i = 1, \dots, n$, as follows:

$$\begin{aligned} x_1^1 & \text{ from } \pi(x_1 | x_2^0, x_3^0, \dots, x_n^0) \\ x_2^1 & \text{ from } \pi(x_2 | x_1^1, x_3^0, \dots, x_n^0) \\ & \vdots \\ x_n^1 & \text{ from } \pi(x_n | x_1^1, x_2^1, \dots, x_{n-1}^1). \end{aligned}$$

The conditionality in the distributions is the key feature. Thus after the first iteration we transit to $x^1 = (x_1^1, \dots, x_n^1)$ from $x^0 = (x_1^0, \dots, x_n^0)$. On completion of t and more such iterations we have a sequence $x^0, x^1, \dots, x^t, \dots$ which is a realization of a Markov chain with the following transition probability from x^t to x^{t+1} :

$$G(x^t, x^{t+1}) = \prod_{l=1}^n \pi(x_l^{t+1} | x_k^t, k > l, \text{ and } x_k^{t+1}, k < l).$$

This gives the basic idea. When all distributions are taken as conditional on the data, the marginal distributions give the marginal posteriors as in Bayes' theorem, as demonstrated in the next section.

Hydrological applications of the Gibbs sampling scheme have been made by Adamson et al. (1999) for simulating flood hydrographs in the Mekong River in Vientiane, Laos; Sansò and Guenni (1999) who simulated rainfalls in Venezuela; Perreault et al. (2000) for modelling annual energy inflows to two large hydropower dams in Quebec, Canada for change-point analysis in time series; and Onibon et al. (2004) who used the procedure to simulate Sahelian rainfields in West Africa.

The working algorithm

The trend component of Eq. (1) can be written as

$$t_{i+1} = t_i + s_i \quad (2)$$

using an auxiliary variable s_i . The uncertainty in the trend is modelled as

$$s_{i+1} = s_i + \omega_{1i} \quad (3)$$

in which ω_{1i} is a 'latent' variable (in Bayesian terminology) that describes the random fluctuations in this component. With regard to the periodicity component, let us consider Eq. (1) as a monthly time series and simplify this component so that it is modeled as a truncated Fourier series, in the usual way, but including only the fundamental frequency, $f = 1/12$ (per month, corresponding to the annual cycle). This is closely representative for monthly mean temperatures and generally provides a good approximation for some other monthly climatic series. More about this follows. Accordingly, the periodicity component is modelled, following West and Harrison (1997) as

$$p_{i+1} = p_i \cos(2\pi f) + q_i \sin(2\pi f) \quad (4)$$

and

$$q_{i+1} = -p_i \sin(2\pi f) + q_i \cos(2\pi f) + \omega_{2i} \quad (5)$$

using an auxiliary variable q_i , and another 'latent' variable ω_{2i} that describes the random fluctuations in this component. Thus the trend component is treated as a simple linear process whereas the periodicity component is a linear sum of sine and cosine curves. The advantage in this type of non-stationary modelling is that one can incorporate differences in the cyclical structure and also changes in the mean level from year to year. The dynamic linear model of the system, also called the state space model, is then written as follows:

$$x_{i+1} = Fx_i + G\omega_i \quad (6)$$

$$y_i = Hx_i + \eta_i \quad (7)$$

where

$$x_i = [t_i, s_i, p_i, q_i]^T$$

and

$$\omega_i = [\omega_{1i}, \omega_{2i}]^T.$$

Also,

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \quad (8)$$

$$G = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \quad (9)$$

and

$$F = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{bmatrix} \quad (10)$$

such that Eq. (6) represents Eqs. (2)–(5) and (7) represents Eq. (1). Because of the lack of prior knowledge we arbitrarily specify the initial conditions $x_0 = [t_0, s_0, p_0, q_0]$, of the variables in Eqs. (2)–(5) related to trend and periodicity.

Eqs. (6) and (7) are similar to the well known Kalman filter (see, for example, Brown and Hwang, 1992). However, we adopt a generalized approach by modelling the state variables as random walks in which the variances are treated as parameters within a Bayesian estimation framework, in the manner of West and Harrison (1997).

Accordingly, we write the stochastic dynamic model of Eqs. (6) and (7) in a static form which facilitates application of the Gibbs sampler. For the stochastic representation we follow Magni and Bellazzi (2004):

$$y = Az + \eta \quad (11)$$

$$x = Bz \quad (12)$$

Here y represents n observations, as in Eq. (1), corresponding to which η signifies n measurement errors and x denotes the $2n$ values of trend t and periodicity p to be evaluated Eqs. (2), (4) and z is a variable incorporating the four initial conditions of t , s , p and q Eqs. (2)–(5) followed by the $2(n-1)$ error terms of Eqs. (3) and (5) of the trend and periodicity components.

Thus

$$x = [t_0 \ p_0 \ \dots \ t_{n-1} \ p_{n-1}]^T$$

$$y = [y_0 \ \dots \ y_{n-1}]^T$$

$$\eta = [\eta_0 \ \dots \ \eta_{n-1}]^T$$

and

$$z = [t_0 \ s_0 \ p_0 \ q_0 \ \omega_{10} \ \omega_{20} \ \dots \ \omega_{1n-2} \ \omega_{2n-2}]^T$$

Also, to conform with Eqs. (6)–(10) the A and B matrices of constants in Eqs. (11) and (12) are defined as

$$A = \begin{bmatrix} H & 0 & \dots & 0 \\ HF & HG & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ HF^{n-1} & HF^{n-2}G & \dots & HG \end{bmatrix} \quad (13)$$

and

$$B = \begin{bmatrix} M & 0 & \dots & 0 \\ MF & MG & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ MF^{n-1} & MF^{n-2}G & \dots & MG \end{bmatrix} \quad (14)$$

in which

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (15)$$

A preliminary schematic diagram of the model of Eqs. (11) and (12) is given in Fig. 1.

To implement the scheme, one needs to specify the probability distributions of the error terms. This also applies to the initial states. Let us assume that

$$P(\omega_{k_i}) = N(0, \sigma_{\omega_k}^2), \quad k = 1, 2, \quad (16)$$

$$P(\eta_i) = N(0, \sigma_{\eta_i}^2), \quad (17)$$

$$P(x_0) = N\left(0, \text{diag}\left[\sigma_{x_{0_1}}^2 \quad \sigma_{x_{0_2}}^2 \quad \sigma_{x_{0_3}}^2 \quad \sigma_{x_{0_4}}^2\right]\right), \quad (18)$$

where $N(\cdot, \cdot)$ denotes the multinormal distribution and $\text{diag}[\cdot]$ signifies a diagonal matrix.

This generalization of the algorithm is made in order to apply the Gibbs sampler, as previously stated. In particular, $\sigma_{\omega_1}^2$ and $\sigma_{\omega_2}^2$ are treated as parameters to be estimated from the data within the Bayesian framework. As in Gilks et al. (1996) we assume that the unknown variances in Eq. (16) pertaining to trend and periodicity have inverse gamma distributions. For the prior state these distributions are vaguely defined, that is they are taken to be somewhat like uniform so that the final estimates do not have a large bearing on these distributions. As regards Eq. (17), it is assumed for simplicity that the variances in the measurement error term are fixed; however; this can be easily generalized by suitable modifications of the ensuing steps. As regards Eq. (18), since the initial values of the t , s , p and q variables of Eqs. (2)–(5) are assumed arbitrarily, as stated, their variances are taken sufficiently large so that the data will dominate the posterior distribution.

In order to update trend and periodicity through Bayes' theorem we compute the first two moments of the joint posterior probability

$$P(t, p, \sigma_{\omega_1}^2, \sigma_{\omega_2}^2 | y), \quad (19)$$

where for n values of data y , $t = [t_0 \dots t_{n-1}]^T$ and $p = [p_0 \dots p_{n-1}]^T$ are the trend and periodicity components respectively. To implement the MCMC procedure through Gibbs sampling, we partition the random parameters into three groups or subsets: $\sigma_{\omega_1}^2$, $\sigma_{\omega_2}^2$ and z . It means that, the dependent structure shown in Fig. 1 should accommodate a scheme that extracts a sample in each iteration from each of the following three conditioned distributions (i.e. the distribution of a variable conditioned on the other variables and sample data, y) based on Eq. (11) and the previously stated strategy for Gibbs sampling:

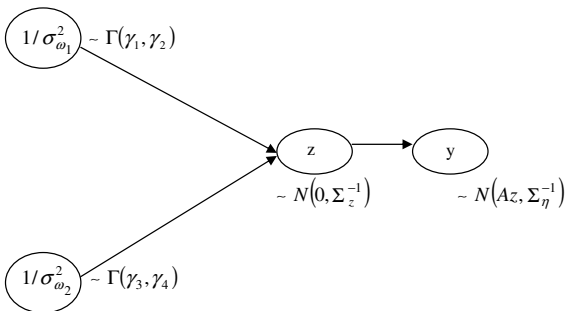


Figure 1 Schematic diagram for implementing the Gibbs sampling scheme.

$$P\left(\frac{1}{\sigma_{\omega_1}^2} \middle| \sigma_{\omega_2}^2, z, y\right) = \Gamma(n/2 + \gamma_1, (\omega_1^T \omega_1)/2 + \gamma_2) \quad (20)$$

$$P\left(\frac{1}{\sigma_{\omega_2}^2} \middle| \sigma_{\omega_1}^2, z, y\right) = \Gamma(n/2 + \gamma_3, (\omega_2^T \omega_2)/2 + \gamma_4) \quad (21)$$

$$P(z | \sigma_{\omega_1}^2, \sigma_{\omega_2}^2, y) = N(D^{-1} A^T \sum_{\eta=1}^{-1} y, D^{-1}) \quad (22)$$

$$\text{with } D = A^T \sum_{\eta}^{-1} A + \sum_z^{-1}$$

in which $\Gamma(\cdot, \cdot)$ is the gamma distribution and $N(\cdot, \cdot)$ is the multinormal distribution, σ_{ω}^2 is a vector containing the sequence $\{\sigma_{\omega_1}^2, \sigma_{\omega_2}^2\}$ repeated $(n-1)$ times, $\omega_i = [\omega_{i_0} \dots \omega_{i_{n-2}}]^T$ with $i = 1, 2$ and $(\gamma_1, \dots, \gamma_4)$ pertain to the parameters of the *prior* gamma distributions of $1/\sigma_{\omega_1}^2$, $1/\sigma_{\omega_2}^2$. The sample data are given by y as defined previously.

Also, $\sum_{\eta} = \text{diag}([\sigma_{\eta_0}^2 \dots \sigma_{\eta_{n-1}}^2])$ and $\sum_z = \text{diag}([\sigma_{x_{0_1}}^2 \quad \sigma_{x_{0_2}}^2 \quad \dots \quad \sigma_{x_{0_4}}^2 \quad \sigma_{\omega}^2])$.

The iterative procedure

The iterative procedure can be summarized as follows:

- (1) The samples of $P(\sigma_{\omega_1}^2, \sigma_{\omega_2}^2, z | y)$ are drawn iteratively by the Bayesian estimator and Gibbs sampler using Eqs. (20)–(22). This is in the main loop of the computer programme.
- (2) We commence with prior values of the variances of ω_1 and ω_2 as in Eqs. (20) and (21) (regarding the parameters of their inverse gamma distributions).
- (3) The vector z is simulated through the multinormal distribution of Eq. (22), based on Eq. (11), conditioned on sample data y and the two variances of ω_1 and ω_2 . Recall that the variable z contains, except for the four initial conditions of t, s, p and q at the beginning, alternating values of ω_1 and ω_2 . The vector z forms a row in the matrix Z , at each iteration.
- (4) New values of ω_1 and ω_2 are abstracted from z and taken to Eqs. (20) and (21).
- (5) Eqs. (20) and (21) provide the new variances of ω_1 and ω_2 .
- (6) The new variances are brought forward each time to Eq. (22).
- (7) From these iterations k series (runs) are obtained for Z .
- (8) Then the initial si series are discarded and the mean values of a z row vector are formed by averaging the $sr = k - si$ remaining series for each point in time.
- (9) We draw through Eq. (12) the sampling posterior distribution of the trend and periodicity components by a direct transformation of z to x .

In the next section we discuss procedures to determine k and si of steps 7 and 8 of the procedure.

Results and discussion

Fig. 2 gives the monthly mean temperature at Torino for the period 1753 to 1993 with the data grouped by years.

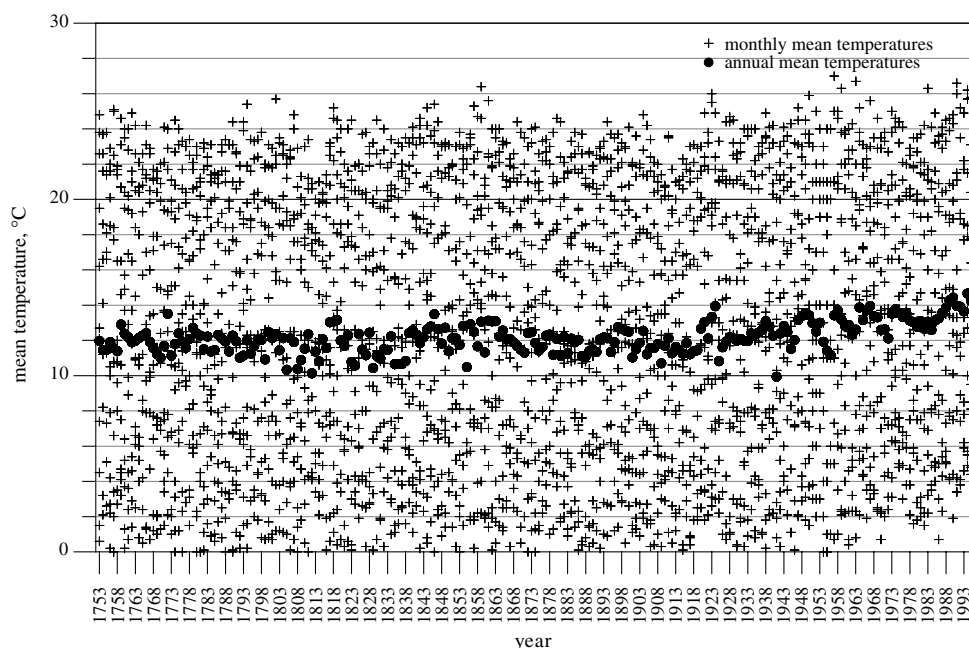


Figure 2 Monthly mean temperature at Torino for the period 1753–1993 — data grouped by years.

Also shown are annual mean temperatures. As expected, the variability in the annual mean temperatures is seen to be much smaller than in the monthly data. The annual temperature data does not show a significant trend except for a rise after the first quarter of the century, or thereabouts. In Fig. 3 the same data are grouped by months. The averages of the monthly mean temperatures are also shown with a close harmonic fit using only the first harmonic. This justifies the assumption made previously. More about this follows. Fig. 4 gives the following annual time series of temperature in Torino: summer mean temperatures, annual mean temperatures, and winter mean tem-

peratures. In each case a main trend line is drawn and compared with 25-year nonoverlapping (secondary) linear trend lines. This shows random changes in trend over the 25-year periods. The tendency for a rise in temperature in the final 25-year period is also seen in this figure. Such a period of temperature rise may extend back to a longer period, as seen in Fig. 2.

The Bayesian estimation and Gibbs sampling scheme is then applied to the monthly temperatures in Torino from 1971 to 1993. First, Fig. 5 shows the trend component. This does not indicate any general movement. Second, Fig. 6 shows the periodicity component up to 1993. This has sharp

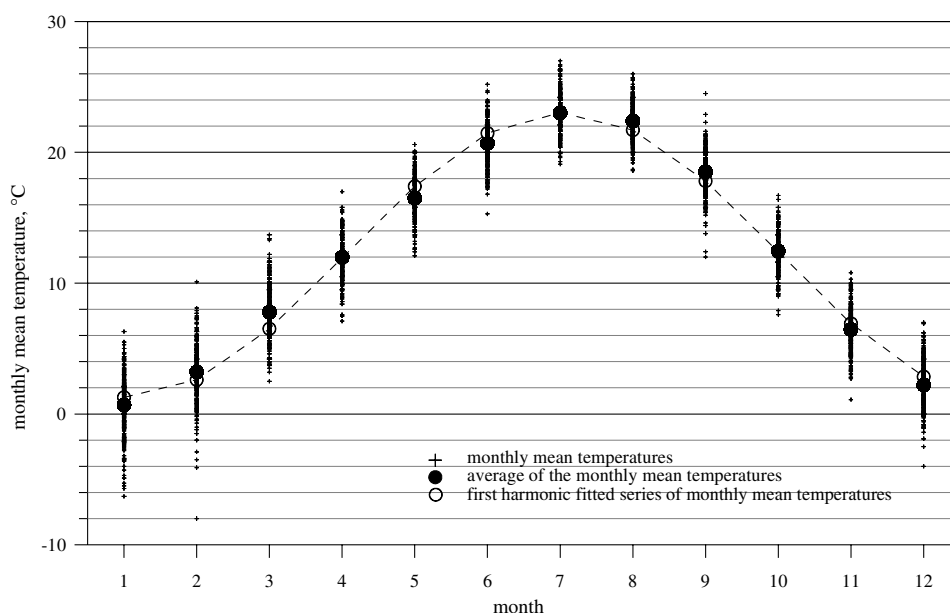


Figure 3 Monthly mean temperature at Torino — data grouped by months with a first harmonic fit to the averages.

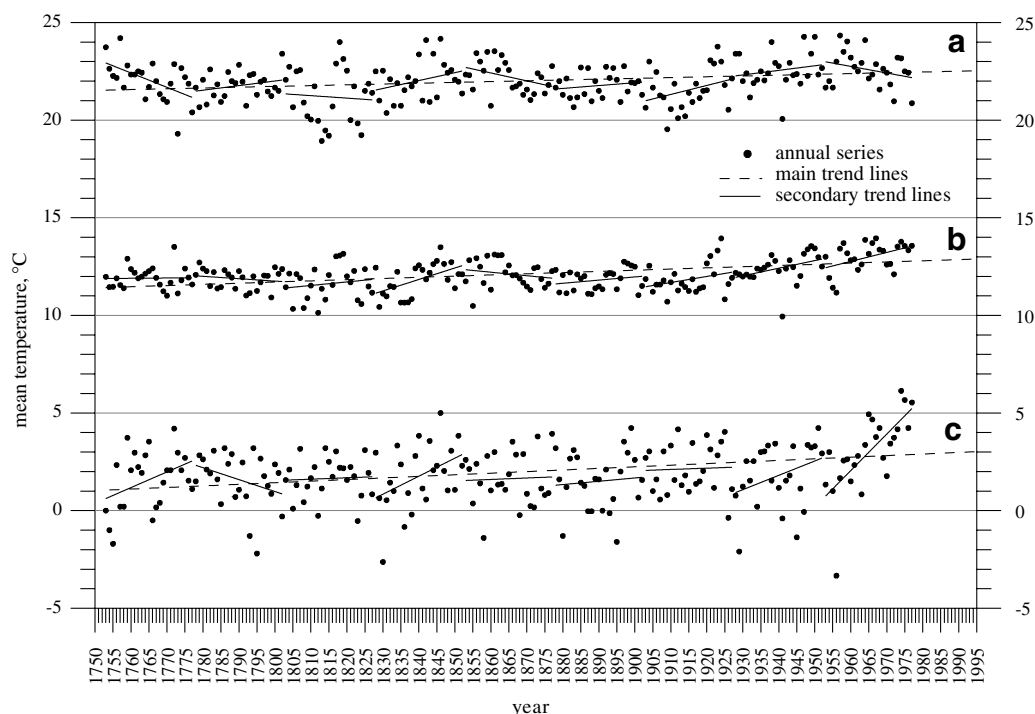


Figure 4 Annual series of temperature in Torino: 25-year nonoverlapping (secondary) linear trend lines compared with the main linear trend line for (a) summer mean temperatures, (b) annual mean temperatures and (c) winter mean temperatures.

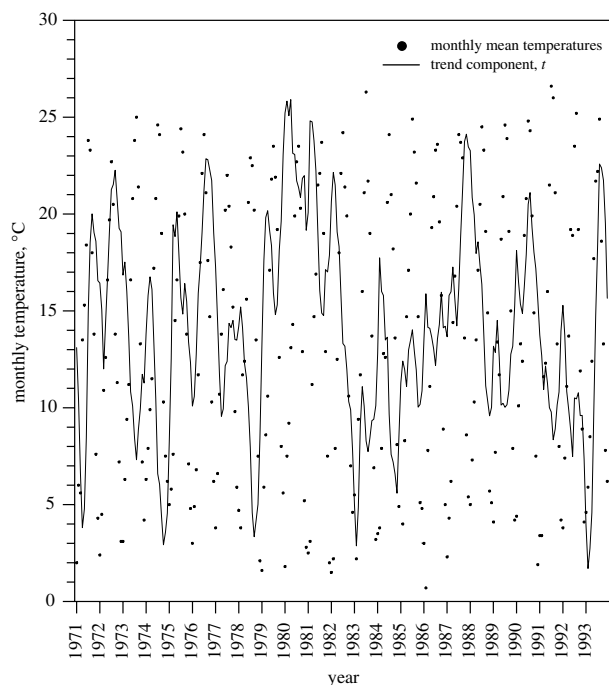


Figure 5 Monthly mean temperature in Torino for the period 1971–1993 with the trend component t from Gibbs sampling.

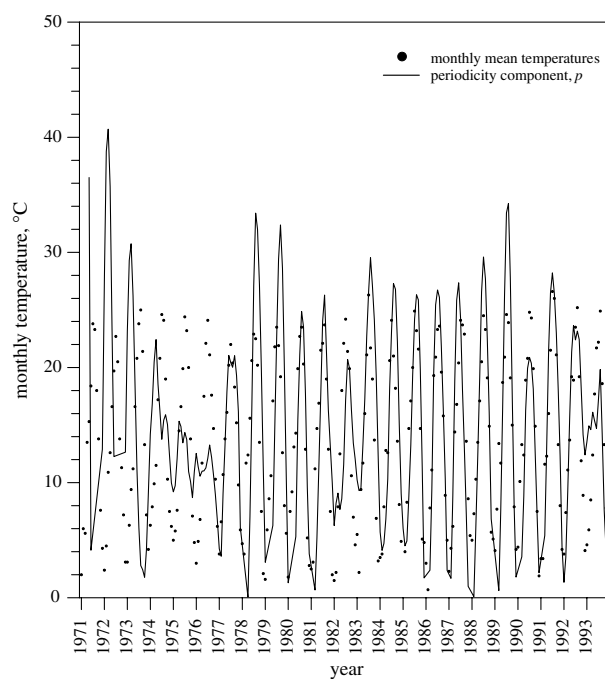


Figure 6 Monthly mean temperature in Torino for the period 1971–1993 with the periodicity component p from Gibbs sampling.

irregular movements within periods of 12 months indicating the nonstationarity of the periodicity component.

Fig. 7 shows monthly total rainfall in Milano for the period 1840–2003 with the data grouped by years. Also shown are annual values represented by mean monthly rainfalls

for the year. The variabilities in the rainfalls are much larger than those in the temperatures in Fig. 2, and show stronger characteristics of nonstationarity because of the various processes involved. During this period there is no indication of a definite trend.

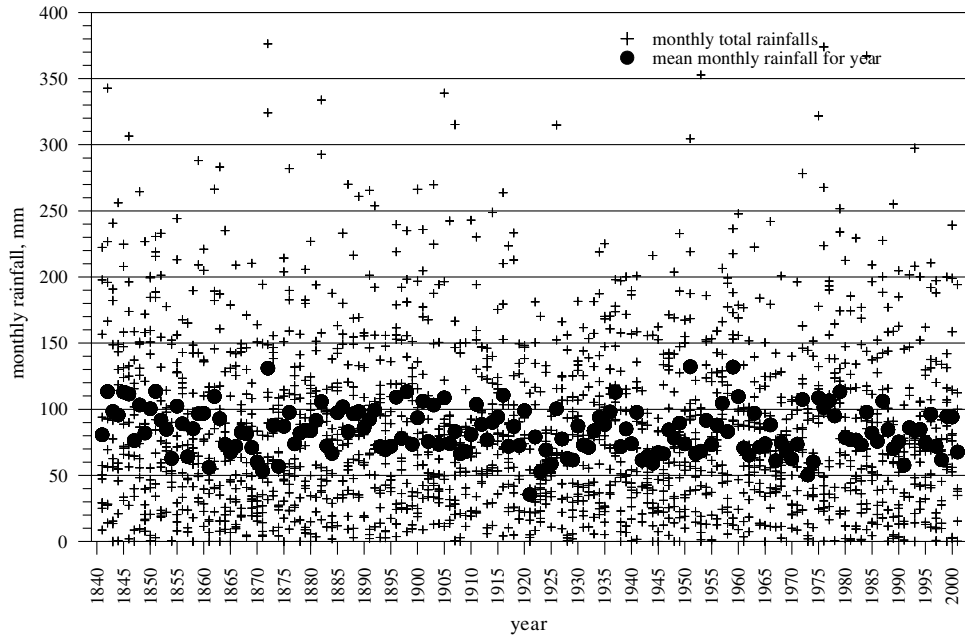


Figure 7 Monthly total rainfall in Milano for the period 1840–2000 – data grouped by years.

Let us clarify, with respect to the Milano rainfalls, the exploratory procedure used to determine the number of series sr from the simulations to be averaged for the determination of the trend, say, (steps 7 and 8 of section ‘The iterative procedure’). We experimented with different appropriate combinations of sr and initial series si , that need to be discarded because they do not satisfy the necessary criteria. The differences d_i between the averaged trend components $t_{i,sr/1.2}$ and $t_{i,sr}$ were squared and summed for all the points $i = 1, 2, \dots, n$ where $n = 276$ and $sr = 90, \dots, 250$. We denote the square root of the average as the ‘error’ in the trend:

$$e_{sr} = \sqrt{\sum_{i=1}^n d_i^2 / n}.$$

A typical plot of e_{sr} vs sr is shown in Fig. 8. We adopt $sr = 200$ in the simulations as the effect of further increases in sr is minimal.

We also decided to investigate the choice of the number of series si from the simulations to be discarded prior to the determination of the trend component. For this purpose we found the differences d'_i between the averaged trend components $t_{i,sr}$, as previously defined, and $t_{i,si}$, based only on the series to be discarded – for these calculations we ignored the initial 10 series which caused an excessive distortion in the results. The differences d'_i were squared and summed for $i = 1, 2, \dots, n$ where $n = 276$ and $si = 10, 20, \dots, 90$. The ‘error’ in the trend is written as

$$e_{si} = \sqrt{\sum_{i=1}^n d_i'^2 / n}.$$

A typical plot of e_{si} vs si is shown in Fig. 9. We decided to adopt $si = 50$ in the simulations.

Figs. 10 and 11 give monthly total rainfalls in Milano for the 24-year period from 1982 to 2001 with the trend and

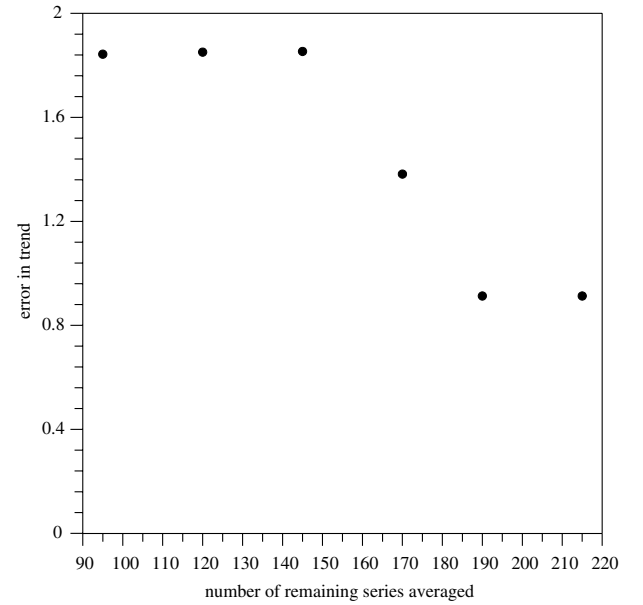


Figure 8 Exploratory analysis of the number of series sr (step 8 of section ‘The iterative procedure’) to be averaged for the determination of the trend component t from Gibbs sampling in Milano rainfall data for the period 1982–2001.

periodicity components, respectively, via the Bayesian estimation and Gibbs sampling scheme. The trend component does not have any clear long-term movements. The periodicity component has oscillations that are much sharper than for the trend component with movements within the year due to its intrinsic nature. The 95% and 5% credibility limits are also shown. For this purpose we rank the x values at each point in time i through Eq. (12) (prior to averaging of z in step 8 of section ‘The iterative procedure’) and obtain

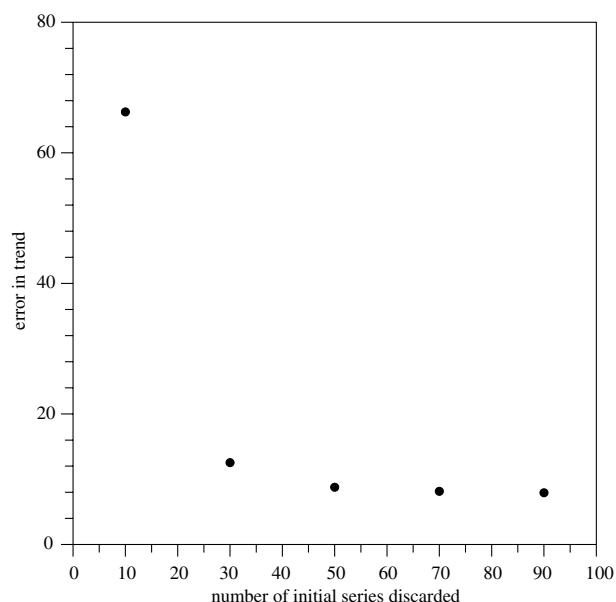


Figure 9 Exploratory analysis of the number of initially simulated series si (step 8 of section 'The iterative procedure') to be discarded prior to the determination of the trend component t from Gibbs sampling in Milano rainfall data for the period 1982–2001.

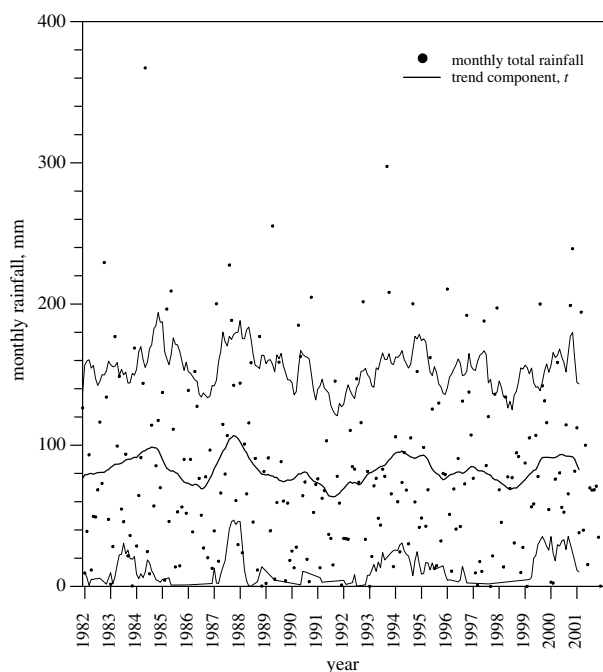


Figure 10 Monthly total rainfall in Milan for the period 1982–2001 with the trend component t from Gibbs sampling – also shown are the 95% and 5% posterior credibility limits.

the rounded x values with 0.95 and 0.05 probabilities of exceedance.

Fig. 12 shows the monthly mean flow in the river Po at Pontelagoscuro for the period 1920 to 1991 with the data grouped by years. As in the case of Milano rainfalls, there

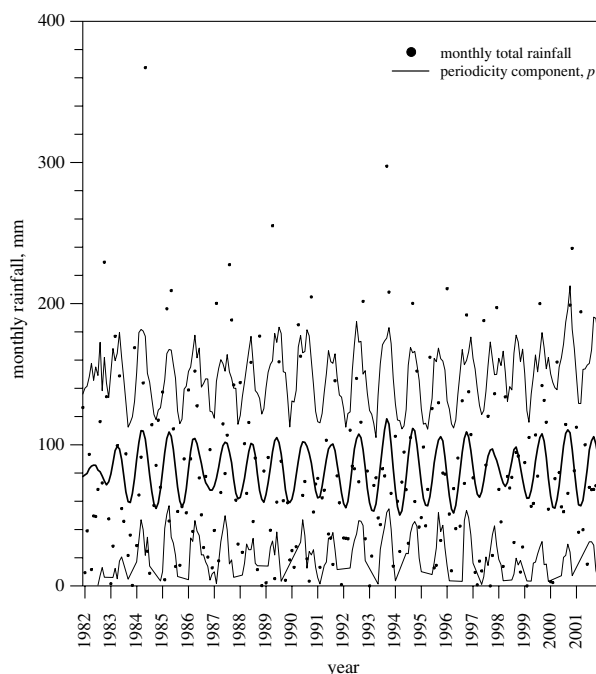


Figure 11 Monthly total rainfall in Milan for the period 1982–2001 with the periodicity component p from Gibbs sampling – also shown are the 95% and 5% posterior credibility limits.

is no steady movement up or down. This is the largest river in Italy and at Pontelagoscuro the catchment area is 70,091 km². Figs. 13 and 14 show monthly mean flows in the Po at Pontelagoscuro for the 24-year period from 1970 to 1991 with the trend and periodicity components respectively, via Bayesian estimation and Gibbs sampling. The trend component indicates local variations. These variations follow the general patterns in monthly flows. The periodicity component is further accentuated by the movements in the annual cycle but is not well defined as in Fig. 11 for the Milano rainfall data, this seems to indicate an increase in the non-stationary behaviour but the scales are different. Both these figures do not show any definite movements. The 95% and 5% posterior credibility limits are also shown.

Fig. 15 shows a comparison of the trend component of the monthly total rainfall for the period 1951 to 1973 at three stations in Lombardia, northern Italy: Milano, Mantova and Tirano. The trend lines follow the general characteristics of the monthly rainfalls. There are similarities in the trend components at the three stations with a slight down-trend towards the end of the period.

Fig. 16 shows the variances of the 'latent' variables associated with trend (ω_1) and periodicity (ω_2) in Eqs. (3) and (5) of the monthly total rainfalls in Milan. In Fig. 16a the variances of ω_1 are given for the period 1982 to 2001. These are not constant and have a non-symmetrical distribution which is expected because of the gamma distribution used. The variances of ω_2 are given in Fig. 16b for the same period. They are much higher than for ω_1 . This is expected from Figs. 10 and 11; because by their nature the oscillations in periodicity exceed in amplitude those for trend.

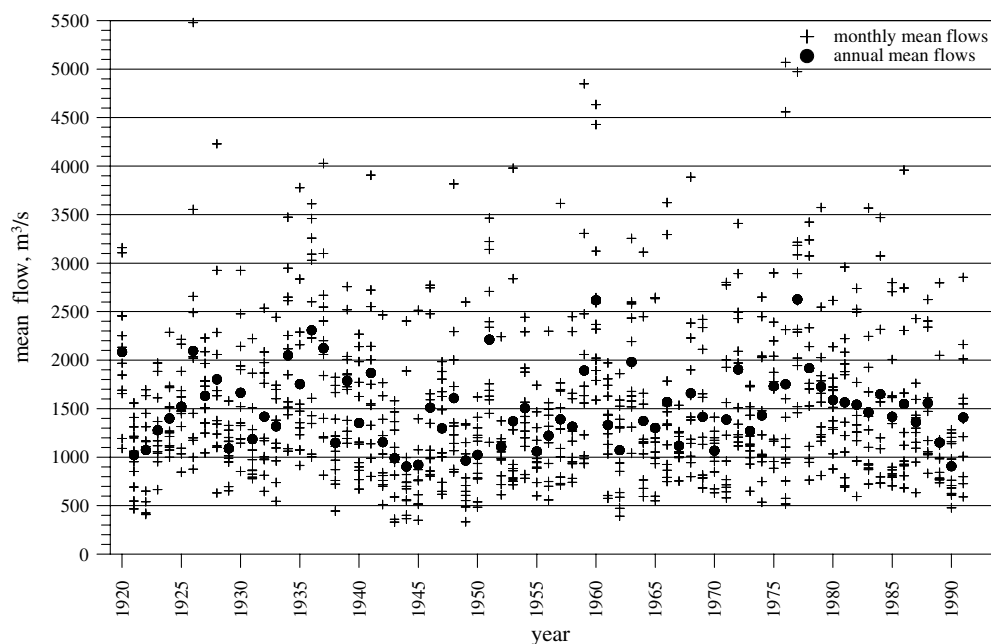


Figure 12 Mean flow in the Po at Pontelagoscuro for the period 1920–1991 – data grouped by years.

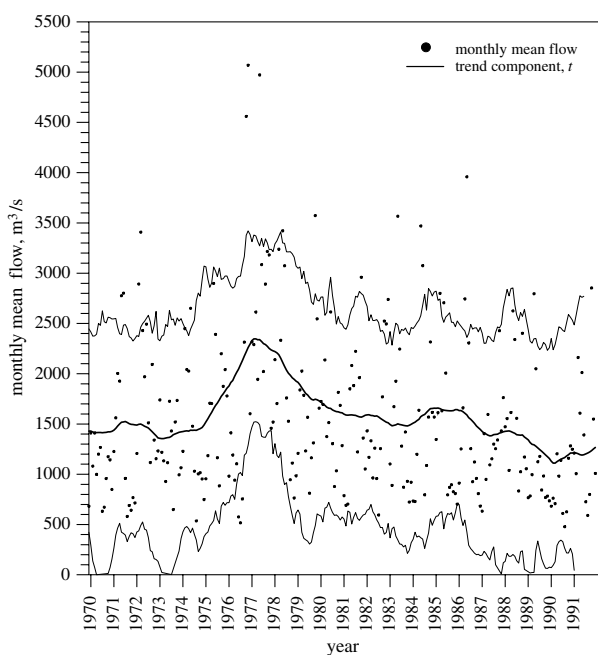


Figure 13 Monthly mean flow in the Po at Pontelagoscuro for the period 1970–1991 with the trend component t from Gibbs sampling – also shown are the 95% and 5% posterior credibility limits.

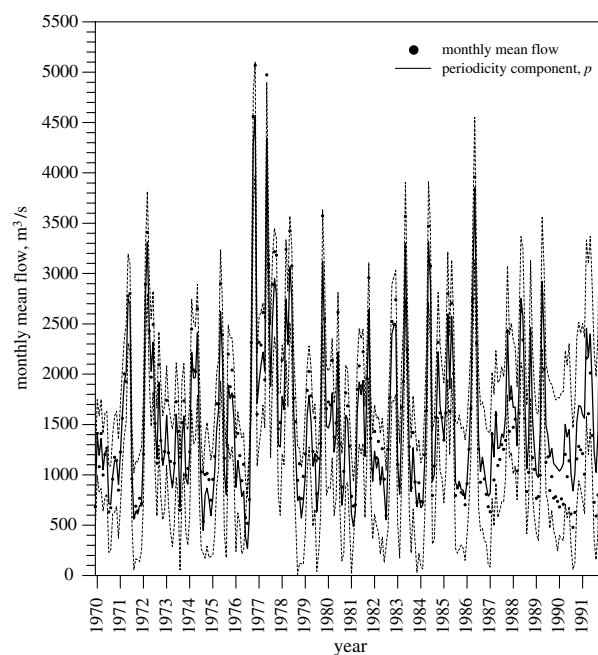


Figure 14 Monthly mean flow in the Po at Pontelagoscuro for the period 1970–1991 with the periodicity component p from Gibbs sampling – also shown are the 95% and 5% posterior credibility limits.

Conclusions and further comments

Seasonality is represented in the monthly data of this study. For some series Eqs. (4)–(10) need to be extended to incorporate a larger number of harmonics, than only the first harmonic applied here (Kottegoda et al., 2004). This also requires modifications or extensions to several equations

from Eqs. (4)–(18), (20)–(22). Also, the assumption of normality made in Eqs. (16)–(18), as usually made in this type of general algorithm, requires further scrutiny although Figs. 2 and 3 do not seem to contradict the assumption. Alternatively, more advanced general simulation models may be used (see West and Harrison, 1997; Carlin et al., 1992; Carter and Kohn, 1994).

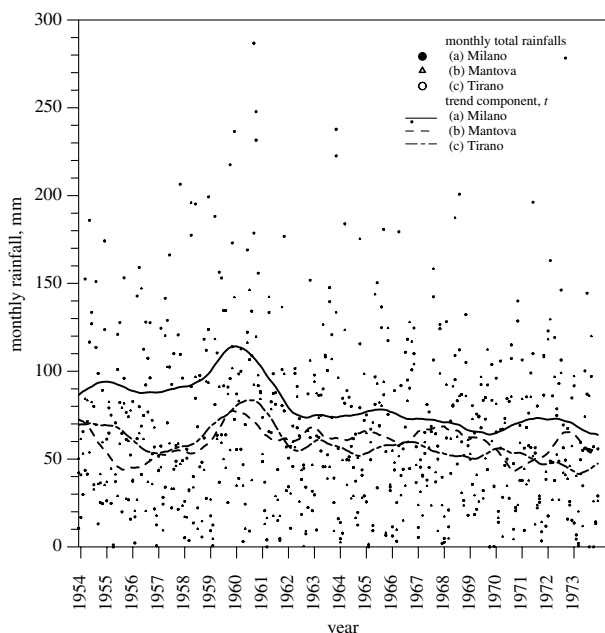


Figure 15 Trend components of monthly total rainfalls for the period 1954–1973 at three stations in Lombardia: (a) Milano, (b) Mantova and (c) Tirano.

Differently from other analyses of climatic time series, trend and periodicity are treated as random components by us. Elsewhere, smoothing and filtering seem to be the norm. We are motivated by considerations that such techniques may sometimes lead to biased results arising mainly from the width of the chosen window. That is, we have not been confined to a prior definition of the width of the moving window.

Three types of climatic time series are studied. In the trend components, they do not seem to show any general movement indicating a definitive rise or decay except that some late temperature rises are noted. There are similarities in behaviour between different types. The time varying periodicities seem to be more pronounced in the flow series but the scales are different. The calculated serial variances in the latent variables associated with trend and periodicity (Eqs. 3 and 5) show that the variable for periodicity has a much higher series of variances.

In the context of climatic change, trends are of greater interest than periodicities. Overall there are illusory aspects in all trend lines as seen in the secondary lines of Fig. 4. Trends only show what has happened in the past. That too is subject to the reliability of the data and their correct interpretation. Extrapolations of trend lines to the future are irrelevant if they are based on assumptions that do not hold. For instance, down trending of populations in some highly industrialized countries to absurd levels takes no account of incentives that governments will offer and the changing attitudes of the younger generations in the future. Another downtrend might show world oil supplies completely depleted before 2050; yet, the current high sale price of oil has justified reaching unexplored vast subterranean sources; besides, alternative sources of energy are being increasingly researched and utilized. Likewise, before

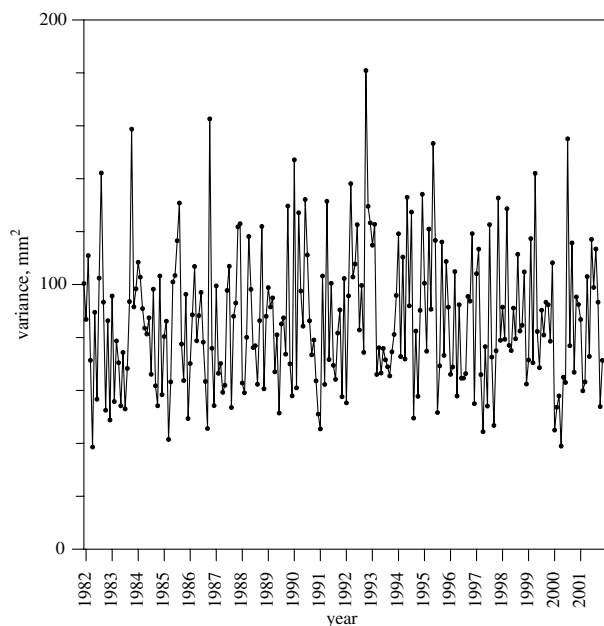


Figure 16a Variance of the random component ω_1 of trend t for monthly total rainfall in Milano during the period 1982–2001 as simulated by Gibbs sampler.

extrapolating a climatic series such as temperature to the future, on the basis of a physical model, variations in solar intensity and the possible effects of other cyclical phenomena in the decades to come should be considered; there are also likely errors associated with the highly complex thermodynamics of the earth's atmosphere.

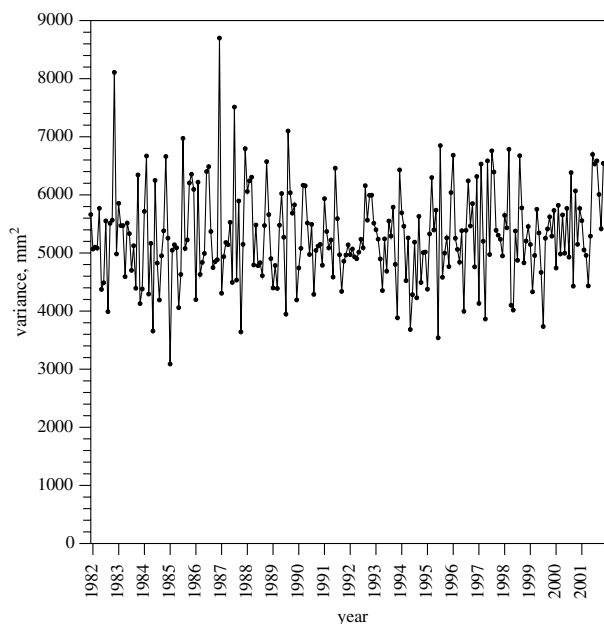


Figure 16b Variance of the random component ω_2 of periodicity p for monthly total rainfall in Milano during the period 1982–2001 as simulated by Gibbs sampler.

Acknowledgements

Financial support provided by Fondazione Lombardia per l'Ambiente is gratefully acknowledged. We thank Dr. Paolo Magni, Department of Computing and Systems, University of Pavia for help with the computer program for Gibbs sampling. We thank the anonymous reviewers for their useful comments. The assistance provided by Gabriele Bria and Matteo Pagliardi in the computations is acknowledged.

References

- Adamson, P.T., Metcalfe, A.V., Parmentier, B., 1999. Bivariate extreme values distributions: an application of the Gibbs sample to the analysis of floods. *Water Resour. Res.* 35, 2825–2832.
- Baliunas, S., Frick, P., Sokoloff, D., Soon, W., 1997. Time scales and trends in the Central England temperature data (1659–1990): A wavelet analysis. *Geophys. Res. Lett.* 24 (11), 1351–1354.
- Bates, B.C., Campbell, E.P., 2001. A Markov chain Monte Carlo scheme for parameters estimation and inference in conceptual rainfall-runoff modeling. *Water Resour. Res.* 37, 937–947.
- Besag, J., Green, P.J., 1993. Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. B* 55, 25–37.
- Boggess, A., Narcovich, F.J., 2001. *A First Course in Wavelets and Fourier Analysis*. Prentice Hall, Upper Saddle River, NJ.
- Brooks, S.P., 1998. Markov chain Monte Carlo method and its application. *The Statistician* 47, 69–100.
- Brown, R.G., Hwang, P.Y.C., 1992. *Introduction to Random Signals and Applied Kalman Filtering*. 2nd ed.. Wiley, New York.
- Carlin, B.P., Polson, N.G., Stoffer, D.S., 1992. Monte Carlo approach to nonnormal and nonlinear state - space. *J. Am. Statist. Assoc.* 87, 493–500.
- Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. *Am. Statist.* 46, 167–174.
- Chatfield, C., 1978. The Holt - Winters forecasting procedure. *Appl. Statist.* 27 (3), 264–275.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings distribution. *Am. Statist.* 40, 327–335.
- Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X., Mallick, B., 2005. An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resour. Res.* 41, W12423, doi:10.1029/2004WR003764.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Statist. Assoc.* 85, 972–985.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Med. Intell.* 6, 721–741.
- Geyer, C.J., 1992. Practical Markov chain Monte Carlo. *Statist. Sci.* 7, 473–511.
- Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A.W., Tian, Y., Varadi, F., Yiou, P., 2002. Advanced spectral methods for climatic time series. *Rev. Geophys.* 40 (1 100). doi:10.1029/2000RG000092.
- Gilks, W.R., Clayton, D.G., Spiegelhalter, D.J., Best, N.G., McNeil, A.J., Sharples, L.D., Kirby, A.J., 1993. Model complexity : applications of Gibbs sampling in medicine. *J. Roy. Statist. Soc. B* 55, 39–52.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, New York.
- Grotch, S.L., MacCracken, M.C., 1991. The use of general circulation models to predict regional climatic change. *J. Climate* 4 (3), 286–303.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Kite, G., 1989. Use of time series analysis to detect climatic change. *J. Hydrol.* 111, 259–279.
- Kottegoda, N.T., Natale, L., Raiteri, E., 2004. Some considerations of periodicity and persistence in daily rainfalls. *J. Hydrol.* 296, 23–37.
- Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* 211, 69–85.
- Magni, P., Bellazzi, R., 2004. Analysing Italian voluntary abortion data using a Bayesian approach to the time series decomposition. *Statist. Med.* 23, 105–123.
- Mann, M.E., 2004. On smoothing potentially non-stationary climate time series. *Geophys. Res. Lett.* 31, L07214. doi:10.1029/2004GL019569.
- Mann, M.E., Ammann, C.M., Bradley, R.S., Briffa, K.R., Crowley, T.J., Hughes, M.K., Jones, P.D., Oppenheimer, M., Osborn, T.J., Overpeck, J.T., Rutherford, S., Trenberth, K.E., Wigley, T.M.L., 2003. On past temperatures and anomalous late - 20th century warmth. *Eos Trans. AGU* 84, 256–258.
- Marshall, L., Nott, D., Sharma, A., 2004. A conceptual study of Markov chain Monte Carlo methods for conceptual rainfall - runoff modelling. *Water Resour. Res.* 40, W02501. doi:10.1029/2003WR002378.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092.
- Nash, L.L., Gleick, P.H., 1991. Sensitivity of streamflow in the Colorado basin to climatic changes. *J. Hydrol.* 125, 221–241.
- Onibon, H., Lebel, T., Afouda, A., Guillot, G., 2004. Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resour. Res.* 40, W08401. doi:10.1029/2003WR002029.
- Reis Jr., D.S., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. *J. Hydrol.* 313, 97–116.
- Perreault, L., Bernier, J., Bobée, B., Parent, E., 2000. Bayesian change-point analysis in hydrometeorological time series, Part 1. The normal model revisited. *J. Hydrol.* 235, 221–241.
- Sansò, B., Guenni, L., 1999. A stochastic model for tropical rainfall at a single location. *J. Hydrol.* 214, 64–73.
- Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B* 55, 3–23.
- Soon, W., Baliunas, S., Idso, C., Idso, S., Legates, D.R., 2003. Reconstructing climatic and environmental changes of the past 1000 years: A reappraisal. *Energy Environ.* 14, 233–296.
- Tierney, L., 1994. Markov chains for exploring posterior distributions. *Ann. Statist.* 22, 1701–1762.
- West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models*. second ed.. Springer, New York.
- Wilby, R.L., Hay, L.E., Leavesley, G.H., 1999. A comparison of downscaled and raw GCM output: implications for climatic change scenarios in the San Juan River basin, Colorado. *J. Hydrol.* 225, 67–93.