Chapter 2
Persistence Models

Abstract Climatic noise often exhibits persistence (Sect. 1.3). Chapter 3 presents bootstrap methods as resampling techniques aimed at providing realistic confidence intervals or error bars for the various estimation problems treated in the subsequent chapters. The bootstrap works with artificially produced (by means of a random number generator) resamples of the noise process. Accurate bootstrap results need therefore the resamples to preserve the persistence of $X_{\text{noise}}(i)$. To achieve this requires a model of the noise process or a quantification of the size of the dependence. Model fits to the noise data inform about the "memory" of the climate fluctuations, the span of the persistence. The fitted models and their estimated parameters can then be directly used for the bootstrap resampling procedure.

It turns out that for climate time series with discrete times and uneven spacing, the class of persistence models with a unique correspondence to continuous-time models is rather limited. This "embedding" is necessary because it guarantees that our persistence description has a foundation on physics. The first-order autoregressive or AR(1) process has this desirable property.

Keywords Persistence • Memory • Autocorrelation • Serial dependence • First-order autoregressive model • AR(1) process Hurst phenomenon

2.1 First-Order Autoregressive Model

The AR(1) process is a simple persistence model, where a realization of the noise process, $X_{\text{noise}}(i)$, depends on just the value at one time step earlier, $X_{\text{noise}}(i-1)$. We analyse even and uneven spacing separately.

M. Mudelsee, *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*, Atmospheric and Oceanographic Sciences Library 51,
DOI 10.1007/978-3-319-04450-7_2,

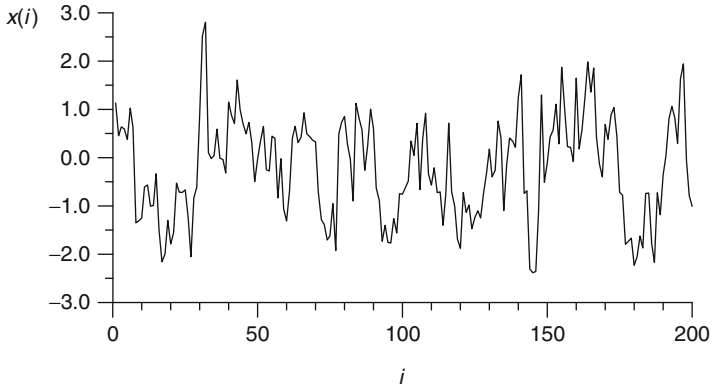

Fig. 2.1 Realization of an AR(1) process (Eq. 2.1); $n = 200$ and $a = 0.7$

2.1.1 Even Spacing

In Eq. (1.2) we let the time increase with constant spacing $d(i) = d > 0$ and write the discrete-time Gaussian AR(1) noise model:

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\text{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\text{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.1}$$

Herein, $-1 < a < 1$ is a constant and $\mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(\cdot)$ is a Gaussian random process with mean μ, variance σ^2 and no serial dependence, that is, $E\left[\mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(i) \cdot \mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(j)\right] = 0$ for $i \neq j$. It readily follows that $X_{\text{noise}}(i)$ has zero mean and unity variance, as assumed in our decomposition (Eq. 1.2). Figure 2.1 shows a realization of an AR(1) process.

The autocorrelation function,

$$\begin{aligned} \rho(h) &= \frac{E\Big[\big\{X_{\text{noise}}(i+h) - E\big[X_{\text{noise}}(i+h)\big]\big\} \cdot \big\{X_{\text{noise}}(i) - E\big[X_{\text{noise}}(i)\big]\big\}\Big]}{\Big\{VAR\big[X_{\text{noise}}(i+h)\big] \cdot VAR\big[X_{\text{noise}}(i)\big]\Big\}^{1/2}} \\ &= E\Big[X_{\text{noise}}(i+h) \cdot X_{\text{noise}}(i)\Big], \end{aligned} \tag{2.2}$$

where h is the time lag, E is the expectation operator and VAR is the variance operator, is given by (Priestley 1981: Sect. 3.5 therein)

$$\rho(h) = a^{|h|}, \qquad h = 0, \pm 1, \pm 2, \ldots. \tag{2.3}$$

For $a > 0$, this behaviour may be referred to as "exponentially decreasing memory" (Fig. 2.2).

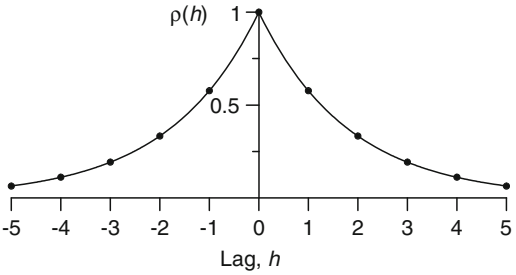

Fig. 2.2 Autocorrelation function of the AR(1) process, $a > 0$. In the case of even spacing (Sect. 2.1.1), $\rho(h)$ is given by $a^{|h|} = \exp[-|h| \cdot d/\tau]$, and in the case of uneven spacing (Sect. 2.1.2), by $\exp[-|T(i+h) - T(i)|/\tau]$. In both cases, the decrease is exponential with decay constant τ

Note that the assumptions in Eq. (1.2), namely, $E[X_{\text{noise}}(i)] = 0$ and $VAR[X_{\text{noise}}(i)] = 1$, required the formulation of the AR(1) model as in Eq. (2.1), which is non-standard. See Sect. 2.6 for the standard formulation.

Persistence estimation for the AR(1) model means estimation of the autocorrelation parameter, a. To illustrate autocorrelation estimation, assume that from the time series data, $\{x(i)\}_{i=1}^{n}$, the outliers have been removed and the trend and variability properties (Eq. 1.2) determined and used (as in Fig. 1.12) to extract $\{x_{\text{noise}}(i)\}_{i=1}^{n}$, realizations of the noise process. An estimator of the autocorrelation parameter, that is, a recipe how to calculate a from $\{x_{\text{noise}}(i)\}_{i=1}^{n}$, is given by

$$\hat{a} = \sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-1) \Big/ \sum_{i=2}^{n} x_{\text{noise}}(i)^2. \tag{2.4}$$

(Chapter 3 introduces estimators and the "hat notation".) Note that estimator $\hat{a}$ is biased, that is, if $\{X_{\text{noise}}(i)\}$ is an AR(1) process with parameter a, then $E(\hat{a}) \neq a$. Only approximation formulas exist for the bias in general autocorrelation estimation. Such formulas can be used for bias correction. Similarly, also the estimation variance, $VAR(\hat{a})$, is only approximately known. In general, bias and variance decrease with n. The background material (Sect. 2.6) gives various bias and variance formulas, informs about bias correction and lists other autocorrelation estimators.

The suitability of the AR(1) model can be assessed using the estimation residuals:

$$\epsilon(i) = x_{\text{noise}}(i) - \hat{a} \cdot x_{\text{noise}}(i-1), \qquad i = 2, \ldots, n. \tag{2.5}$$

As realizations of a standard normal random process, the residuals should not exhibit patterns in the lag-1 scatterplot (Fig. 1.13).

Effective Data Size

Persistence ($a > 0$) means a reduced information content of a time series compared to a situation without positive serial dependence. In a statistical estimation, more

data have then to be available to achieve a confidence interval (Chap. 3) of same width. An effective data size, n', can be defined for estimators of parameters of processes with persistence via the estimation variance. Consider the mean estimator, $\bar{X} = \sum_{i=1}^{n} X(i)/n$, and the AR(1) process Eq. (2.1) for two cases: $a > 0$ and $a = 0$. Then

$$VAR\left(\bar{X}\right) = VAR[X(i)] \Big/ n'_{\mu} \qquad (a > 0)$$

(the index refers to mean estimation) is set equal to

$$VAR\left(\bar{X}\right) = VAR[X(i)]/n \qquad (a = 0).$$

Bayley and Hammersley (1946) show that

$$n'_{\mu} = n\left[1 + 2\sum_{i=1}^{n-1}(1 - i/n)\,\rho(i)\right]^{-1}, \tag{2.6}$$

which for the AR(1) process with the autocorrelation given in Eq. (2.3) can be readily solved using the geometric series as well as the arithmetic–geometric series:

$$n'_{\mu} = n\left\{1 + \frac{2}{n}\,\frac{1}{1-a}\left[a\left(n - \frac{1}{1-a}\right) - a^{n}\left(1 - \frac{1}{1-a}\right)\right]\right\}^{-1}. \tag{2.7}$$

von Storch and Zwiers (1999: Sect. 17.1 therein) define a related quantity, the decorrelation time as

$$\tau_D = \lim_{n\to\infty} \frac{n}{n'_{\mu}}. \tag{2.8}$$

An AR(1) process thus has $\tau_D = (1 + a)/(1 - a)$.

Even for moderate values of n ($\gtrapprox$ 50) and a ($\gtrapprox$ 0.5), the influence of persistence on n'_{μ} can be considerable (Sect. 2.6). Equation (2.7) is valid only for the mean estimator. Because the definition of n' depends of the type of estimation (von Storch and Zwiers 1999), such formulas have limited practical relevance. Section 2.6 gives n' for variance and correlation estimation.

2.1.2 *Uneven Spacing*

In Eq. (1.2), we let the time increase with an uneven spacing $d(i) > 0$ and write the discrete-time Gaussian AR(1) noise model:

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\text{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= \exp\left\{-\left[T(i) - T(i-1)\right]/\tau\right\} \cdot X_{\text{noise}}(i-1) \\
&\quad + \mathcal{E}_{\text{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau\})}(i), \qquad i = 2, \ldots, n.
\end{aligned} \tag{2.9}
$$

The "loss of memory" increases with the time difference scaled by the persistence time, τ (Fig. 2.2). The random innovation, $\mathcal{E}(\cdot)$, is now heteroscedastic instead of homoscedastic as in the case of even spacing. It follows that this noise model for uneven spacing has zero mean, unity variance and autocorrelation:

$$
E\left[X_{\text{noise}}(i+h) \cdot X_{\text{noise}}(i)\right] = \exp\left[-|T(i+h) - T(i)|/\tau\right]. \tag{2.10}
$$

Estimation of the persistence time using noise data $\{t(i), x_{\text{noise}}(i)\}_{i=1}^{n}$ is more complex than in the case of even spacing. A least-squares estimation uses the sum of squares,

$$
S(\tilde{\tau}) = \sum_{i=2}^{n} \left[x_{\text{noise}}(i) - \exp\left\{-\left[t(i) - t(i-1)\right]/\tilde{\tau}\right\} \cdot x_{\text{noise}}(i-1)\right]^2, \tag{2.11}
$$

and takes the minimizer as τ estimator, $\hat{\tau} = \text{argmin}\left[S(\tilde{\tau})\right]$. The minimization has to be carried out numerically (Sect. 2.7). In the case of equidistance, $t(i) - t(i-1) = d \;\forall\; i$, the least-squares estimator corresponds to the estimator given in Eq. (2.4), with $\hat{a} = \exp\left(-d/\hat{\tau}\right)$.

The bias in the estimation of τ for unevenly spaced data seems to defy an analytical derivation. Figure 2.3 shows the bias studied by means of Monte Carlo simulations. The simulations demonstrate that the bias is similar to the value in a situation with even spacing.

Suitability of the AR(1) model for uneven spacing can be assessed using the residuals

$$
\begin{aligned}
\epsilon(i) &= x_{\text{noise}}(i) - \exp\left\{-\left[t(i) - t(i-1)\right]/\hat{\tau}\right\} \cdot x_{\text{noise}}(i-1), \\
&\quad i = 2, \ldots, n.
\end{aligned} \tag{2.12}
$$

Embedding in Continuous Time

In continuous time, the AR(1) noise model is given in "differential notation" by

$$
dX_{\text{noise}}(T) = a \cdot X_{\text{noise}}(T)\, dT + dW(T), \tag{2.13}
$$

where a is the autocorrelation parameter, dT is a time increment and $dW(T)$ is an innovation term of the Wiener process (also called Brownian motion), $W(T)$. As the discrete-time model, the continuous-time AR(1) model has an exponentially decaying autocorrelation function (Priestley 1981).

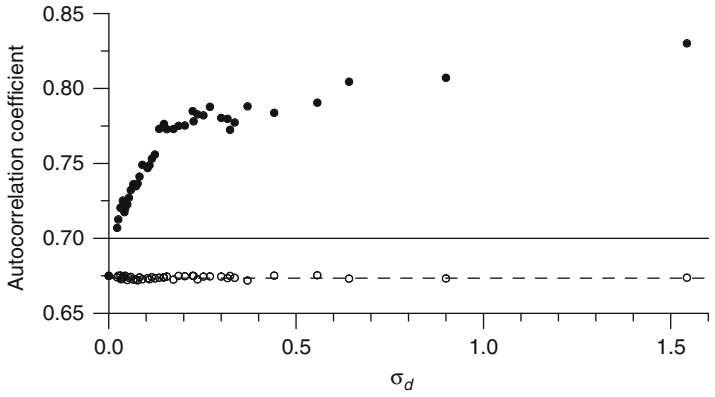

Fig. 2.3 Monte Carlo study of the bias in the autocorrelation estimation of an AR(1) process, known mean, uneven spacing. Time series were generated after Eq. (2.9) with $n = 50$ and $\tau = -1/\ln(0.7) \approx 2.804$ by means of a random number generator (Sect. 2.7). The start was set to $T(1) = t(1) = 1$; the spacing, $d(i)$, was drawn from a gamma distribution with a predefined order parameter (Sect. 2.7) and subsequently scaled such that $t(n) = 50$ or $\bar{d} = 1$. The "equivalent autocorrelation coefficient" is $\bar{a} = \exp(-\bar{d}/\tau) = 0.7$. The standard deviation of the spacing, σ_d, was used as a measure of the unevenness. For each time grid (σ_d fixed), a number ($n_{\text{sim}} = 10{,}000$) of time series were generated and $\hat{\tau}$ determined after Eq. (2.11). Shown (*open symbols*) is the average of the quantity $\exp(-\bar{d}/\hat{\tau})$ over the simulations. Also shown (*filled symbols*) is the average of the estimator $\hat{a}$ (Eq. 2.4) applied to the linearly interpolated, equidistant time series (same start and end, same data size). (The standard error ($\sim 1/\sqrt{n_{\text{sim}}}$) of the estimation averages is smaller than the symbol size.) The true autocorrelation value (*solid line*) is underestimated by the $\hat{\tau}$ estimator. This negative bias is excellently described by the bias approximation (*dashed line*) of White (1961) from the case of even spacing. The interpolation, on the other hand, leads to serious overestimation. This effect is owing to the serial dependence introduced by the interpolation (Fig. 1.14); it increases with σ_d. No formulas to correct for this bias exist

Let us consider the continuous-time noise model to be sampled at discrete times, which may be unevenly spaced, resulting in the discrete-time model, $\{T(i), X_{\text{noise}}(i)\}_{i=1}^{n}$. Robinson (1977) showed that for $a > 0$, this resulting model equals the discrete-time AR(1) model given in Eq. (2.9). The discrete-time AR(1) model is said to be embedded in continuous time; it determines uniquely the underlying continuous-time AR(1) model given in Eq. (2.13). This is an important property of the AR(1) model because the embedding allows a foundation on physics, which works in continuous time (differential equations).

2.2 Second-Order Autoregressive Model

We assume even spacing and write the discrete-time second-order autoregressive or AR(2) noise model:

$$X_{\text{noise}}(i) = a_1 \cdot X_{\text{noise}}(i-1) + a_2 \cdot X_{\text{noise}}(i-2) + \mathcal{E}(i), \tag{2.14}$$

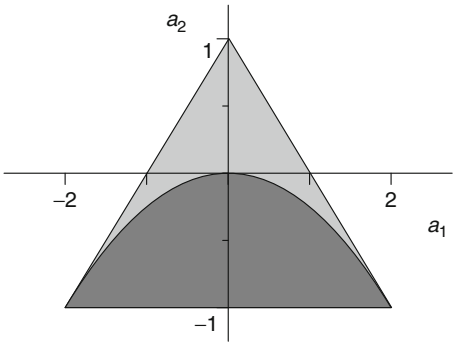

Fig. 2.4 Regions of asymptotic stationarity for the AR(2) process (Eq. 2.14) (*shaded*). The region for complex roots (*dark shaded*), which allows quasiperiodic behaviour, lies below the parabolic, $a_2 < -a_1^2/4$

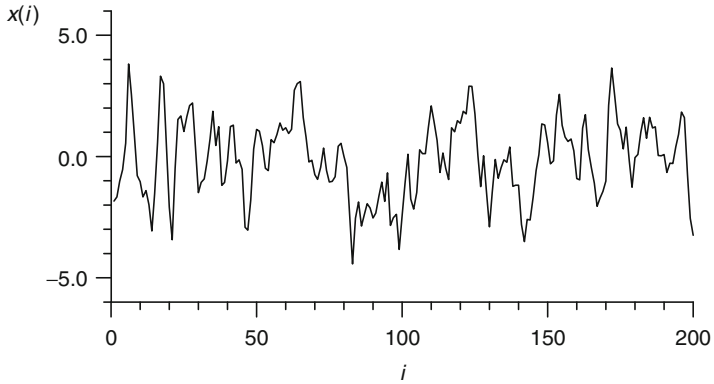

Fig. 2.5 Realization of an AR(2) process (Eq. 2.14); $n = 200$, $a_1 = 1.0$, $a_2 = -0.4$ and $\mathcal{E}(i) = \mathcal{E}_{\mathrm{N}(0,\,1)}(i)$. The first 5000 generated data points were discarded to approach asymptotic stationarity. The graph exhibits a quasi-cyclical behaviour with an approximate period of 9.5 time units

where $\mathcal{E}(i)$ is a stationary purely random process (no serial dependence). This AR(2) process is not strictly stationary but, conditional on a_1 and a_2, only asymptotically stationary (its moments approach saturation with $i \to \infty$); see Sect. 2.6. The regions of asymptotic stationarity in the a_1–a_2 plane are shown in Fig. 2.4. The behaviour of the autocorrelation function of the AR(2) process depends on where a_1 and a_2 lie. For $a_2 \geq -a_1^2/4$, $a_1 > 0$ and $a_2 < 0$, $\rho(h)$ decays smoothly to zero and for $a_2 \geq -a_1^2/4$, $a_1 < 0$ and $a_2 < 0$, $\rho(h)$ alternates its sign as it decays. In the connection with spectral analysis (Chap. 5), the case $a_2 < -a_1^2/4$ is interesting, because then $\rho(h)$ shows besides a decay a quasi-cyclical behaviour with a period of

$$T = 2\pi/\theta, \tag{2.15}$$

where $\cos(\theta) = a_1/(2\sqrt{-a_2})$. Figure 2.5 shows a realization of the process in Eq. (2.14).

Commonly used estimators of the AR(2) model include the Yule–Walker estimators

$$\begin{aligned} \widehat{a_1} &= \hat{\rho}(1) \cdot [1 - \hat{\rho}(2)] \big/ \left[1 - \hat{\rho}(1)^2\right], \\ \widehat{a_2} &= \left[\hat{\rho}(2) - \hat{\rho}(1)^2\right] \big/ \left[1 - \hat{\rho}(1)^2\right], \end{aligned} \tag{2.16}$$

where

$$\hat{\rho}(h) = \hat{R}(h) \Big/ \hat{R}(0) \,, \qquad h = 1, 2, \tag{2.17}$$

is the autocorrelation estimator and

$$\hat{R}(h) = \sum_{i=h+1}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-h), \qquad h = 0, 1, 2, \tag{2.18}$$

is the autocovariance estimator. Estimation bias occurs also in case of the AR(2) model; approximations were given by Tjøstheim and Paulsen (1983).

2.3 Mixed Autoregressive Moving Average Model

We assume even spacing and write the discrete-time Gaussian mixed autoregressive moving average or ARMA(p, q) noise model:

$$\begin{aligned} X_{\text{noise}}(i) = {} & a_1 \cdot X_{\text{noise}}(i-1) + \cdots + a_p \cdot X_{\text{noise}}(i-p) \\ & + b_0 \cdot \mathcal{E}_{\text{N}(0,\,\sigma^2)}(i) + \cdots + b_q \cdot \mathcal{E}_{\text{N}(0,\,\sigma^2)}(i-q), \\ & i = \max(p, q) + 1, \ldots, n. \end{aligned} \tag{2.19}$$

Note that the model is given only for a subset of $i = 1, \ldots, n$. Similar conditions as in the preceding sections may be formulated for $\{a_1, \ldots, a_p;\ \ b_0, \ldots, b_q\}$ to ensure stationarity. The AR(1) model (Sect. 2.1), AR(2) model (Sect. 2.2), and the autoregressive model of general order (AR(p))—these are special cases of the ARMA(p, q) model ($q = 0$). The moving average process of general order (MA(q)) arises from the ARMA(p, q) model with $p = 0$; it is not considered further in this book. Estimation techniques for the ARMA(p, q) model are mentioned in the background material (Sect. 2.6).

As regards the context of this book, the problem with the discrete-time ARMA(p, q) model under uneven time spacing is that no embedding in a continuous-time process can be proven. Indeed, already for a discrete-time, real-valued Gaussian AR(1) process with $a < 0$, it was shown (Chan and Tong 1987) that no embedding in a continuous-time, real-valued Gaussian AR(1) process exists. No embedding of $X_{\text{noise}}(i)$ means no foundation on physics. Suppose, for example, that physical laws governing the climate system to be analysed yield an ARMA(p_1, q_1) continuous-time noise model. Even with a "perfect" estimation (estimation bias and

variance both zero), it would then not be possible to determine the model parameters $\{a_1, \ldots, a_{p_1}; b_0, \ldots, b_{q_1}\}$ uniquely from an unevenly spaced sample time series $\{t(i), x_{\text{noise}}(i)\}_{i=1}^{n}$. For climate time series, a perfect time estimation would also be required, which is not usually possible. A further complication arises from "model aliasing". Bartlett (1946) showed that an evenly sampled continuous-time AR(p) process becomes a discrete-time ARMA($p, p-1$) process (the "alias"), which has implications already for the AR(2) model (Sect. 2.2). However, for a certain type of uneven spacing, namely, "missing observations" (Fig. 1.16e), where $\{t(i)\}_{i=1}^{n}$ is a subset of $\{t(j)\}_{j=1}^{m}$ with $d(j) = \text{const.}$ and $m - n \ll m$, the embedding problem vanishes and estimation techniques exist (Sect. 2.6).

The majority of sampled climate time series, at least within this book, exhibits uneven, irregular spacing (Fig. 1.16), for which only the simple AR(1) model ensures the embedding property. Fortunately, this is no serious limitation as climatic theory shows that climatic noise is to a first order of approximation well described by the AR(1) process (Sect. 2.5).

2.4 Other Models

The discrete-time Gaussian ARMA(p, q) process (Eq. 2.19) composes $X_{\text{noise}}(i)$ as a linear combination of past $X_{\text{noise}}(j)$, $j < i$, and innovations, $\mathcal{E}_{\text{N}(0,\sigma^2)}(j)$. We briefly review other processes that can be seen as extensions of the ARMA(p, q) process. These processes might provide somewhat more realistic models for $X_{\text{noise}}(i)$. However, usage of many of these models seems to be restricted to evenly spaced time series (perhaps with missing values) because of the embedding problem (Sect. 2.1.2) and lack of statistical theory.

2.4.1 Long-Memory Processes

The AR(1) process has an exponentially ("fast") decaying autocorrelation function (Fig. 2.2). Also the ARMA(p, q) process has a similar bound (Brockwell and Davis 1991),

$$|\rho(h)| \leq C\, r^{|h|}, \qquad h = 0, \pm 1, \pm 2, \ldots, \tag{2.20}$$

where $C > 0$ and $0 < r < 1$. A long-memory process is a stationary process (loosely speaking, with time-constant statistical properties such as mean and standard deviation) for which

$$\rho(h) \to C\, h^{2H-1} \qquad \text{as } h \to \infty, \tag{2.21}$$

where $C \neq 0$ and $H < 0.5$. This decrease is slower than in the case of ARMA(p, q); hence, it is said to exhibit long-range serial dependence or long memory.

Examples of long-memory processes are:

1. Fractional Gaussian noise
2. Fractional autoregressive integrated moving average models, denoted as ARFIMA(p, δ, q)

The relation between the ARFIMA(p, δ, q) and ARMA(p, q) models is as follows. δ defines (Sect. 2.6) a fractional difference operator, $(1 - B)^{\delta}$, where $|\delta| < 0.5$ and B is the backshift operator. The backshift operator shifts one step back in time, for example, $B\, X_{\text{noise}}(i) = X_{\text{noise}}(i - 1)$. The ARFIMA$(p, \delta, q)$ model is then an ARMA(p, q) model (Eq. 2.19), where $X_{\text{noise}}(j)$ is replaced by $(1 - B)^{\delta}\, X_{\text{noise}}(j)$. For the trivial case $\delta = 0$, the ARFIMA(p, δ, q) model reduces to the ARMA(p, q) model, which already shows the embedding problem (Sect. 2.3). (One can define a nonstationary ARIMA model by allowing $\delta = 1, 2, \ldots$.) The ARFIMA(p, δ, q) model has $H = \delta$ (Brockwell and Davis 1991). Although continuous-time ARFIMA(p, δ, q) models have been developed (Comte and Renault 1996), the embedding for $\delta \neq 0$ seems not yet to have been analysed.

In the special case of the ARFIMA$(0, \delta, 0)$ model, it has been shown (Hwang 2000) that for uneven spacing, the estimation of δ is biased, with the bias depending on the spacing. It appears that in the general case, the theory of long-memory processes for unevenly spaced data is not well developed enough to be applied in the context of the present book. Section 2.6 gives more details on estimation of long-memory models for evenly spaced data, while Sects. 2.5.2 and 2.5.3 present examples where long- and short-range models are fitted to climate series.

2.4.2 *Nonlinear and Non-Gaussian Models*

Stationary nonlinear models allow a richer structure to be given to the noise process, $X_{\text{noise}}(i)$. Of particular interest for climatology is the class of threshold autoregressive models (Tong and Lim 1980). Let the real line $\mathbf{R}$ be partitioned into l nonoverlapping, closed segments, $\mathbf{R} = \mathbf{R_1} \cup \mathbf{R_2} \cup \cdots \cup \mathbf{R_l}$. The discrete-time Gaussian self-exciting threshold autoregressive process of order $(l; k, \ldots, k)$ or SETAR$(l; k, \ldots, k)$ process, where k is repeated l times, is given by

$$X_{\text{noise}}(i) = a_0^{(m)} + \sum_{j=1}^{k} a_j^{(m)} \cdot X_{\text{noise}}(i - j) + \mathcal{E}_{\text{N}(0,\, \sigma^{2(m)})}(i), \tag{2.22}$$

conditional on $X_{\text{noise}}(i - j) \in \mathbf{R_m}; m = 1, 2, \ldots, l$. As an example, Fig. 2.6 shows a realization of the SETAR$(2; 1, 1)$ process,

$$X_{\text{noise}}(i) = \begin{cases} +2.0 + 0.8 X_{\text{noise}}(i - 1) + \mathcal{E}_{\text{N}(0,\, 1)}(i) & \text{if } X_{\text{noise}}(i - 1) \leq 0, \\ -1.0 + 0.4 X_{\text{noise}}(i - 1) + \mathcal{E}_{\text{N}(0,\, 2)}(i) & \text{if } X_{\text{noise}}(i - 1) > 0, \end{cases} \tag{2.23}$$

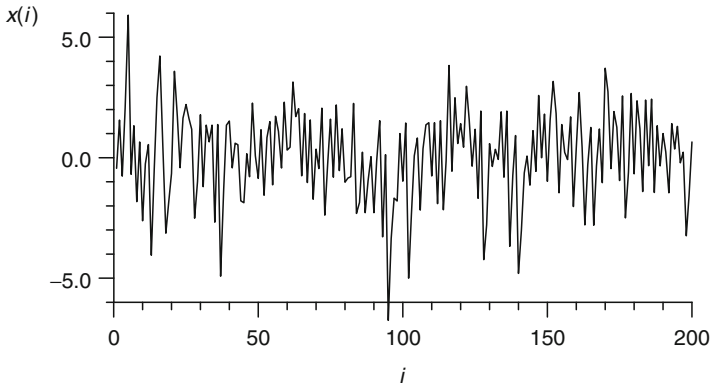

Fig. 2.6 Realization of a SETAR(2; 1, 1) process (Eq. 2.23); $n = 200$ (The first 5000 generated data points were discarded)

which may be a model of random fluctuations between two climate regimes with different mean values and persistence times. Also quasi-cyclical behaviour can be reproduced by threshold models. In practical applications, the number of regimes, l, is usually limited to a few. Estimation is carried out iteratively: guess of l, maximum likelihood estimation of parameters, calculation of a goodness-of-fit measure such as AIC or a normalized version (Tong and Yeung 1991), analysis of residuals and autocorrelation functions, improved guess of l, etc. Continuous-time threshold autoregressive models have been formulated (Tong and Yeung 1991) but it seems that the embedding problem for unevenly spaced time series has not yet been analysed. This would mainly concern the SETAR(2; 1, 1) case.

Many more types of nonlinear persistence models can be perceived (Sect. 2.6). It may for some climate data even be useful to consider in Eq. (1.2) the process $S(i) \cdot X_{\text{noise}}(i)$ as belonging to the class of stochastic volatility models, for which $S(i)$ depends on past $X_{\text{noise}}(i - j)$, $j > 0$, and/or past $S(i - j)$. This process of time-varying variability could model "burst" phenomena such as earthquakes and serve also as a formulation of the outlier process in Eq. (1.2). One common problem, however, with complex, nonlinear time series models, is the embedding of the discrete-time process in continuous time (Sect. 2.1.2). We have to concede that complex, unevenly observed climatic processes may not be accessible to a meaningful parametric estimation.

Also non-Gaussian random innovation terms can be used to construct ARMA(p, q) models. In such cases, however, formulas for the estimation bias are hardly available. One possibility is to introduce a transformation,

$$X_{\text{noise}}(i) = f\left(X'_{\text{noise}}(i)\right), \tag{2.24}$$

where $X'_{\text{noise}}(i)$ is a Gaussian process, and to infer f from a probability density estimation (Sect. 1.6) using $\{x_{\text{noise}}(i)\}_{i=1}^{n}$.

2.5 Climate Theory

A dynamical view of the climate system gives motivation that climatic persistence may to a first order of approximation be written as an AR(1) process. This was shown by an influential paper entitled "Stochastic climate models" (Hasselmann 1976).

The dynamical view seems to be challenged by a series of papers claiming a "universal power law" in temperature records. This law indicates long memory, not short as the AR(1) models suggest. Long memory of temperature fluctuations over timescales from days to decades should seriously impact the development of climate theory. It would further have enormous practical consequences as weather forecasts for intervals considerably longer than what is currently feasible (a few days) would become principally possible.

The reanalysis of some crucial data here (Sect. 2.5.2) makes it hard to accept the "universal power law". It should nevertheless be kept in mind that the AR(1) model need not be a good higher-order description of $X(T)$. However, after allowing nonlinear trends and outlier processes (Eq. 1.2), the AR(1) model is likely not a bad candidate to describe $X_{\text{noise}}(T)$.

2.5.1 Stochastic Climate Models

The derivation of the AR(1) model of climate persistence is based on three assumptions.

Assumption 1 (Timescale separability). The climate system as a whole (p. 3) is composed of a slowly varying component ("climate" in original sense), representing oceans, biosphere and cryosphere, and a fast varying component ("weather"), representing the atmosphere.

The differential equations governing the climate evolution may then be written as

$$\frac{dX(T)}{dT} = F(X(T), Y(T)), \qquad \text{timescale } \tau_X, \tag{2.25}$$

$$\frac{dY(T)}{dT} = G(X(T), Y(T)), \qquad \text{timescale } \tau_Y, \tag{2.26}$$

where $\tau_X \gg \tau_Y$, X and Y are the slowly and fast varying components (vectors of possibly high dimension), respectively, and F and G are some nonlinear system operators. τ_Y is of the order of a few days, τ_X of several months to years and more (Hasselmann 1976).

It was previously thought that the influence of Y on X, of weather on climate, could be accounted for by simply averaging

$$\frac{dX(T)}{dT} \simeq F^*(X(T)), \qquad \text{timescale } \tau_X, \tag{2.27}$$

where the modified climate system operator, F^*, is the time average of $F(X, Y)$.

Since the work of Hasselmann (1976) it is accepted that the weather noise cannot so easily be cancelled out. Consider $0 \leq T \leq \tau_X$. Then

$$\begin{aligned} \frac{dX(T)}{dT} &\simeq F(X(0), Y(T)), \\ &= W(T), \end{aligned} \tag{2.28}$$

where W is a stochastic (Wiener) process. Discretization yields

$$X(T+1) = X(T) + \mathcal{E}_{\mathrm{N}(0,\sigma^2)}(T). \tag{2.29}$$

Here we have made

Assumption 2. The unknown weather components $Y(T)$ add up to yield after the central limit theorem (Priestley 1981: Sect. 2.14 therein) a Gaussian purely random process $\mathcal{E}_{\mathrm{N}(0,\sigma^2)}(T)$.

Now let $T > \tau_X$. Then the time-dependence of $F(X(T), Y(T))$ has to be taken into account. Since the climate system trajectories have to be bounded, we must invoke a negative feedback mechanism. The simplest model for that is given by

Assumption 3. The negative feedback is proportional to the climate variable, $X(T)$, yielding $F(X(T), Y(T)) = -\beta \cdot X(T) + W(T)$.

Assumption 3 makes Eq. (2.29), which is a nonstationary random walk process (Sect. 2.6), to a stationary AR(1) process,

$$X(T+1) = a \cdot X(T) + \mathcal{E}_{\mathrm{N}(0,\sigma^2)}(T), \tag{2.30}$$

with $0 \leq a = 1 - \beta < 1$. (Strictly speaking, this is an "asymptotically stationary" AR(1) process; see "Background Material".) This explanation of Hasselmann's (1976) derivation of climate's AR(1) model is from von Storch and Zwiers (1999: Sect. 10.4.3 therein). Another account of the 1976 paper and its influence on climatology is given by Arnold (2001).

The suitability of the AR(1) noise model depends on how well Assumptions 1–3 are fulfilled. Assumption 1 (timescale separability) is generally thought to be well fulfilled. The root cause is that the atmosphere has a much smaller density and heat

capacity than most of the rest of the climate system, allowing weather processes to run faster. One caveat may be that some climatically relevant biological processes, such as algae growth (Lovelock and Kump 1994), may act also on short timescales. The validity of Assumption 2 is difficult to prove. It might well be that some weather influences do not add but rather multiply with each other, producing non-Gaussian distributional shape (Sect. 1.2; Sura et al. 2005). But Assumption 2 can be relaxed to recognize this, leading to non-Gaussian AR(1) models. Assumption 3 is certainly not exactly fulfilled but may be a good first-order approximation. More sophisticated feedback mechanisms would lead to higher-order ARMA(p, q) or nonlinear models. An interesting case would be a nonlinear dynamical climate system with several local attractors and the occurrence probability within the attracting regions depending on the weather noise (Hasselmann 1999; Arnold 2001). The present knowledge about feedback processes is, however, too limited to permit the theoretical derivation of the precise model form. A further point is that external climate forcing mechanisms (e.g. volcanic eruptions) have to be included for achieving a full set of climate equations. The size of such forcings is currently not well understood (Sect. 8.4).

The dynamical equations in this section are for the evolution of high-dimensional climate variables, $\boldsymbol{X}(T)$, and not just for the noise part of one variable, $X_{\text{noise}}(T)$. In Eq. (1.1), we composed a climate variable of trend, outliers and noise. By allowing nonlinear trends and outlier processes, effects of violations of the assumptions made above are reduced. This lends credence to the AR(1) noise model.

2.5.2 *Long Memory of Temperature Fluctuations?*

Peng et al. (1994) introduced Detrended Fluctuation Analysis (DFA) to measure persistence in DNA sequences. Peng et al. (1995) elaborated DFA in more detail and applied it to heartbeat time series. Koscielny-Bunde et al. (1996) introduced DFA to climate time series analysis and found a "universal power law governing atmospheric variability".

DFA uses a time series $\{t(i), x(i)\}_{i=1}^{n}$ with constant spacing $d > 0$ to calculate the so-called profile:

$$y(i) = \sum_{j=1}^{i} x(j), \qquad i = 1, \ldots, n. \tag{2.31}$$

The profile is divided into nonoverlapping, contiguous segments of length l (multiple of d), discarding the mod$(n, l/d)$ last points. The $y(i)$ series is detrended by segment-wise fitting and subtracting polynomials of order 0, 1, 2, etc. Most commonly used are mean and linear detrending. Koscielny-Bunde et al. (1998a,b) studied the influence of different detrending types, also other than polynomial

detrending. The fluctuation function, $F(l)$, is the standard deviation of detrended $y(i)$ within a segment, averaged over all segments. $F(l)$ is usually plotted on a double-logarithmic scale because power laws, $F(l) \propto l^{\alpha}$, appear in such plots as a straight line, with slope α.

For a polynomial of order 0 and $x(i)$ from the Gaussian AR(1) model in Eq. (2.1), it readily follows that $F(l) \propto l^{1/2}$. For data with long-range dependence (Eq. 2.21), the power law $F(l) \propto l^{H+1/2}$ results, that is, $\alpha = H + 1/2$ (Taqqu et al. 1995; Talkner and Weber 2000). Thus, DFA can be seen as a method to estimate long-range dependence.

Koscielny-Bunde et al. (1996, 1998a,b) analysed daily temperature series covering typically the past 100 a using DFA and found $\alpha \approx 0.65$ for many records. The claimed universal long-range power-law dependence would have serious theoretical and practical consequences. Govindan et al. (2002) went further and analysed temperature output from a number of AOGCMs, the currently most sophisticated mathematical tools for climate simulation. Since the AOGCMs did not produce a single, universal value but rather a scatter of α values, it was concluded that the models were not able to provide realistic climate forecasts. In particular, the predicted size of global warming would be overestimated. Later, Fraedrich and Blender (2003) used DFA to analyse monthly temperature series. They disputed the existence of the universal power law and suggested the following: $\alpha = 1$ for oceanic data, $\alpha = 0.5$ for inner continental data and $\alpha = 0.65$ for data from the transition regions. This led to an exchange of arguments (Bunde et al. 2004; Fraedrich and Blender 2004), where in particular the results from the temperature record from Siberian station Krasnoyarsk were assessed controversially. It appears that in reply to a criticism by Ritson (2004), the originators partly stepped back (Vyushin et al. 2004) from the claimed universality to a position with two memory laws, one for the ocean and the other for the continents.

Here we reanalyse the Krasnoyarsk temperature record and also one series of North Atlantic air temperatures to assess whether the power-law exponents are similar or not. It is also asked whether the power laws are actually good models of the serial dependence in the temperature data and whether simple AR(1) models are not already sufficient. A simulation study helps to quantify the uncertainty of the result coming from sampling variations. We use the same gridded raw data set as Fraedrich and Blender (2003); also the technique for removing the annual cycle, the orders of the DFA polynomials (0, 1) and the scaling range of l = 1–15 a for α determination are identical to what these authors employed. Two points may limit comparability of results. First, we restrict ourselves to those time intervals where the monthly series have no gaps. Fraedrich and Blender (2003) took longer series that start in 1900, without explaining how they adapted their methodology to the case of missing data. Second, Fraedrich and Blender (2003) gave the coordinates 30°W, 50°N for the North Atlantic, but the raw data for the grid cells around this point start clearly later than 1900. We take a grid cell from somewhat more south that starts earlier.

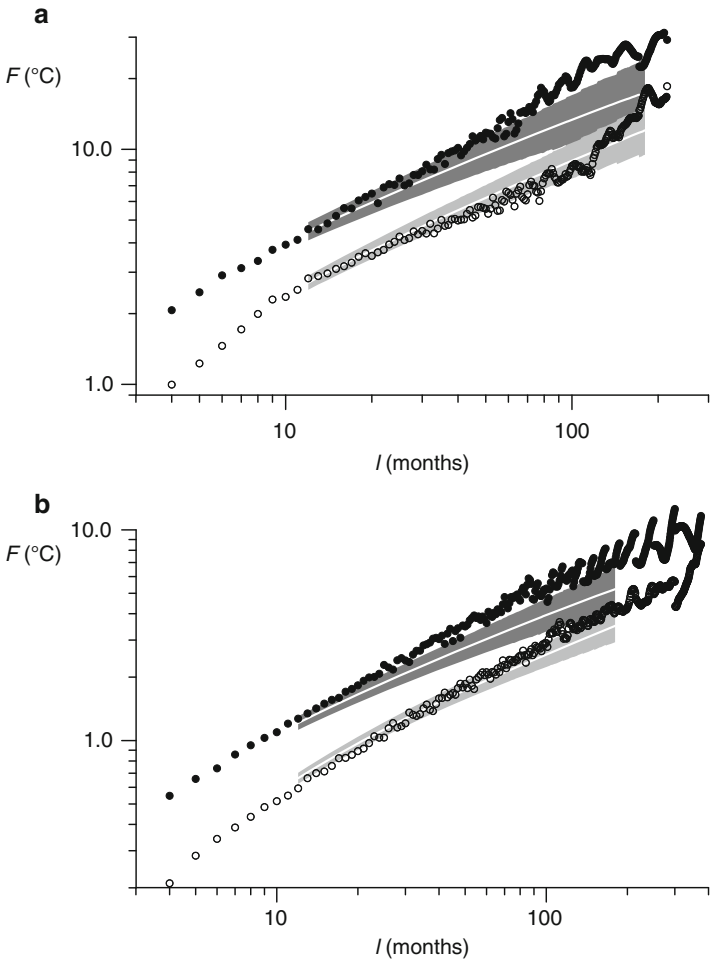

Fig. 2.7 Detrended Fluctuation Analysis for temperature records (Fig. 1.11) from Siberia (**a**) and North Atlantic (**b**). Shown as *filled* (*open*) *symbols* are fluctuation functions, F, against segment sizes, l [months] $= 4, \ldots, n/4$, for the mean-detrended (linearly detrended) DFA variants. Also shown as *shaded areas* are the 90 % confidence bands from simulation experiments based on AR(1) model fits (*dark*, mean-detrended; *light*, linearly detrended); the median (50 %) simulation results are drawn as *white lines*. Simulation results are plotted for the range $l =$ 1–15 a, for which power laws and the related question after the suitability of short- and long-memory models for the data are discussed in the main text

The results of the DFA are as follows (Fig. 2.7). The $F(l)$ curves show increases that resemble on first sight a power law. The α estimates in Table 2.1 were determined by fitting the power-law regression model to the $F(l)$ points inside the selected scaling range 1–15 a. Note that fits of a linear regression model to logarithmically transformed l and F values would likely lead to a biased estimation, because Gaussian additive (measurement) noise of $X(i)$ and $F(i)$ would be lost by

Table 2.1 Result of DFA study (Fig. 2.7), estimated power-law exponents α

	DFA		
Station (grid point)	Mean detrending	Linear detrending	ARFIMA$(1, \delta, 0)$
Krasnoyarsk (50–55°N, 90–95°E)	0.65 ± 0.02	0.79 ± 0.02	0.61
North Atlantic (35–40°N, 25–30°W)	0.58 ± 0.01	0.68 ± 0.01	0.73

DFA errors are standard errors from unweighted least-squares regression (see Chap. 4). ARFIMA models were fitted using Whittle's approximate maximum likelihood technique (Beran 1994: Chap. 5 therein) and α determined via the relation $\alpha = \delta + 1/2$

the transformation (Sect. 2.6). The resulting values exhibit a variation with the DFA detrending type that is considerably larger than the standard errors for α, indicating systematic estimation errors. The α estimations via DFA deviate also considerably from the values obtained via ARFIMA fits (Table 2.1).

A close inspection of the $F(l)$ curves (Fig. 2.7) reveals marked deviations from straight line in the double-logarithmic plots, especially for larger l. Such a behaviour might be referred to as "crossover" (Peng et al. 1995) and different scaling regions with different α values could be further investigated. Philosophy of science, however, says that this is a problematic step because it violates the principle of parsimony. A simpler explanation of the $F(l)$ curves is that the class of power-law models is not ideally suited to describe the data.

The DFA study therefore explores also how the simplest persistence model, AR(1), is suited to describe the data. The AR(1) model (Eq. 2.1) was fitted using the estimator Eq. (2.4) and bias correction (Eq. 2.45), yielding $\hat{a}' = 0.23$ (Siberia) and $\hat{a}' = 0.44$ (North Atlantic). For both cases, $n_{\mathrm{sim}} = 10{,}000$ AR(1) time series were generated (identical means, variances and autocorrelations as the data) and two DFA variants applied (mean detrending and linear detrending). The central 90 % of simulated $F(l)$ at each point l in the scaling region are shown as shaded area (Fig. 2.7). This is a percentile confidence band, that is, a set of percentile confidence intervals (Chap. 3) for F over a range of l values. The confidence bands contain large portions of the $F(l)$ curves from the data. This indicates that DFA is not an ideal method to discriminate between power-law and AR(1) models. The median simulation result illustrates this point, where the AR(1) model produces an almost perfect straight line, which could be misinterpreted as power-law behaviour (Fig. 2.7). However, systematic deviations exist between AR(1) and power-law models in the DFA plots for larger l. These could indicate some significant long-memory behaviour. This finding is supported by the ARFIMA fits, which have lower AICC values (Eq. 2.46) than simple AR(1) fits.

As regards the dispute about the universality of the power law in temperature series, on basis of the AR(1) and ARFIMA estimations (Table 2.1), we conclude that the oceanic data have a stronger memory than the land data. Because the difficulties associated with DFA in interpreting the double-logarithmic plots and selecting the suitable detrending method, ARFIMA models with their elaborated estimation

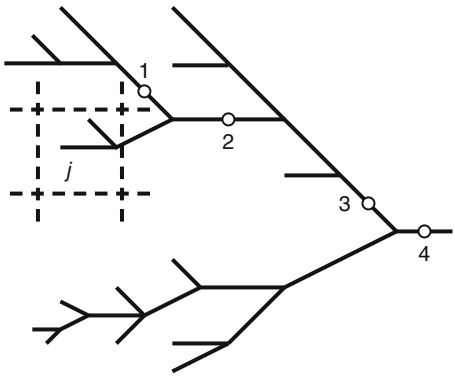

Fig. 2.8 River network. Runoff at a point (e.g. 4) is the spatial aggregation of runoff from upstream (to the left in the picture). Shown is also a hypothetical spatial unit j with area A_j. The basin size, A, for a point is given by the sum of the areas A_j upstream (From Mudelsee (2007), with permission from the publisher)

techniques (Beran 1994) are to be preferred for quantifying long memory. This could also be the reason why DFA is almost completely absent from the statistical literature.

Although the evidence for long-memory dependence in temperature time series seems yet not strong, more records should be analysed with ARFIMA estimation for achieving a better overview. However, analysing aggregated spatial averages, such as northern hemisphere temperature (Rybski et al. 2006), is likely unsuited for this purpose because the aggregation of short-memory AR(1) processes with distributed autocorrelation parameters yields a long-memory process (Granger 1980): the long memory may be a spurious effect of the aggregation. This has been noted also by Mills (2007). Aggregation is likely not a problem for the data analysed here, which come from one station (Siberia) and less than or equal to four stations (North Atlantic), respectively. Hemispheric averages may, however, result from processing several thousand individual records (Chap. 8).

2.5.3 Long Memory of River Runoff

Hurst found in an influential paper (Hurst 1951) evidence for long memory in runoff records from the Nile. The long-memory property of runoff time series has subsequently been confirmed for a number of rivers; see, for example, Mandelbrot and Wallis (1969), Hosking (1984), Mesa and Poveda (1993), Montanari et al. (1997), Montanari (2003), Pelletier and Turcotte (1997), Koutsoyiannis (2002), Bunde et al. (2005) and Koscielny-Bunde et al. (2006). Hurst's finding inspired the development of the theory of long-memory processes (Sect. 2.4.1) and of their estimation (Hosking 1984). Up to now, no widely accepted physical explanation of the "Hurst phenomenon" of long memory, that is, on the basis of the physical–hydrological system properties, has been found (Koutsoyiannis 2005a,b).

The paper by Mudelsee (2007) presents an explanation, which suggests that a river network aggregates short-memory precipitation and converts it into long-memory runoff. River basins (Fig. 2.8) form a network of tributaries, confluences

and reservoirs that has been geometrically described as a fractal object (Rodriguez-Iturbe and Rinaldo 1997). Consider a single, hydrologically homogeneous area A_j, that is, a reservoir with a linear input–output rule described by a dimensionless positive constant k_j. If the input to the reservoir, given by precipitation minus evaporation, is a purely random process, then it has been shown (Klemeš 1978) that the output, $X_j(i)$, is an AR(1) process (Sect. 2.1) with autocorrelation parameter $a_j = 1/(k_j + 1)$. This further implies that the runoff at a point in a river is not from a single reservoir but a cascade (Klemeš 1974) of reservoirs, one feeding the next (Fig. 2.8):

$$Q(i) = \sum_{j=1}^{m} X_j(i), \qquad i = 1, \ldots, n, \tag{2.32}$$

where $X_j(i)$ are mutually independent AR(1) processes with autocorrelation parameter a_j. It has been shown previously that if the $X_j(i)$ have identical means (zero), identical standard deviations (unity) and the a_j are either beta-distributed (Granger 1980) or uniformly distributed (Linden 1999), then for $m \to \infty$ the aggregated process $X_j(i)$ is a long-memory process. Monte Carlo simulations (Mudelsee 2007) reveal that the estimated long-memory parameter δ increases with m and that saturation of δ sets in already for $m \approx 100$. This leads to the suggestion (Mudelsee 2007) that long memory in river runoff results from spatial aggregation of many short-memory reservoir contributions.

To test the aggregation hypothesis, Mudelsee (2007) studied the long-memory parameter δ of fitted ARFIMA$(1, \delta, 0)$ models in dependence on the basin size, A. The idea behind the $\delta(A)$ estimation is that with increasing A also the number m of contributions $X_j(i)$ grows. Thereby should also δ increase, from zero ($m = 1$) to a saturation level below 0.5 (m large). The fact that the distribution of the a_j for real rivers is difficult to derive empirically or analytically can be ignored at this low level of sophistication of the hypothesis. The resulting $\delta(A)$ curve for one of the longest available runoff records, from the river Weser (Germany) (Fig. 2.9), basically confirms the aggregation explanation of the Hurst phenomenon.

Mudelsee (2007) estimated $\delta(A)$ curves also for other rivers, finding similar $\delta(A)$ increases (Elbe, Rhine, Colorado and Nile) but also in one case (Mississippi) a $\delta(A)$ decrease. This paper discusses the validity of the various assumptions made by the aggregation hypothesis. One particular criticism is that the linear input–output release rule may be violated for very large reservoirs. Another obstacle is the requirement of very long time series (above, say, 70 years) for obtaining sufficient accuracy. A major criticism is that the aggregation of AR(1) processes is not an ARFIMA process (Linden 1999) and that the result (Fig. 2.9) may therefore be affected by estimation bias. This paper (Mudelsee 2007) further finds little evidence for long memory in precipitation records from the same regions as the river basins. It thus appears appropriate to reserve the concept of the "Hurst phenomenon" for hydrological time series, and not for climate time series in general.

The aggregation hypothesis has been recently confirmed on instrumental runoff series from rivers Flint (Hirpa et al. 2010) and Po (Montanari 2012).

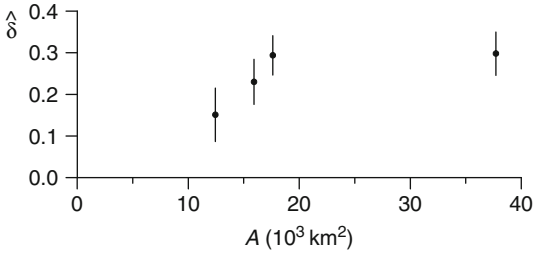

Fig. 2.9 Long-memory parameter in dependence on basin size, river Weser. The δ estimates (*dots*) are shown with bootstrap standard errors (Doornik and Ooms 2003). The time series are monthly runoff values from January 1857 to April 2002. δ was estimated using an ARFIMA$(1, \delta, 0)$ model and maximum likelihood (Doornik and Ooms 2003). Prior to the estimations, the data were logarithmically transformed, the annual cycle removed by subtracting the day-wise long-term averages and the linear trends removed. This ARFIMA model had for all four river stations better AIC values than the ARFIMA$(0, \delta, 0)$ model (After Mudelsee 2007)

2.6 Background Material

Textbooks on the theory and estimation of ARMA(p, q) processes were written by Priestley (1981), Brockwell and Davis (1991, 1996), Box et al. (1994) and Chatfield (2004). Nonlinear time series models are covered by Priestley (1988), Tong (1990) and Fan and Yao (2003). The latter two books have the notable aim to bridge the gap between statistics and nonlinear dynamics; see also Tong (1992, 1995). Long-memory processes are the topic of Beran (1994, 1997), Doukhan et al. (2003) and Robinson (2003). ARFIMA(p, δ, q) processes are reviewed by Beran (1998) and fractional Gaussian noise processes by Mandelbrot (1983). Identification and fitting of time series models are described from the perspective of hydrology (Hurst phenomenon), engineering and environmental system analysis by Hipel and McLeod (1994). Tables of series and other formulas can be found in the books by Abramowitz and Stegun (1965) and Gradshteyn and Ryzhik (2000).

The **AR(1) model standard formulation** is (Priestley 1981: Sect. 3.5.2 therein)

$$X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i-1) + \mathcal{E}(i), \tag{2.33}$$

where $\mathcal{E}(i)$ is a stationary purely random process with mean μ_ϵ and standard deviation σ_ϵ. In the general case, this noise model is not stationary. For $\mu_\epsilon \neq 0$ and $|a| < 1$, $E\,[X_{\text{noise}}(i)]$ is not constant but approaches with time a saturation value of $\mu_\epsilon/(1-a)$. For $\mu_\epsilon = 0$ and $|a| < 1$, $VAR\,[X_{\text{noise}}(i)]$ is not constant but approaches with time a saturation value of $\sigma_\epsilon^2/(1-a^2)$. The case $\mu_\epsilon = 0$ and $|a| = 1$ results in a random walk process (p. 55). The standard formulation in Eq. (2.33) describes an "asymptotically stationary" process, whereas Eq. (2.1) describes a strictly stationary process. In practical applications such as random number generation (Sect. 2.7), the standard model (Eq. 2.33) can be used if the transient sequence of numbers (say, the first 5000) is discarded.

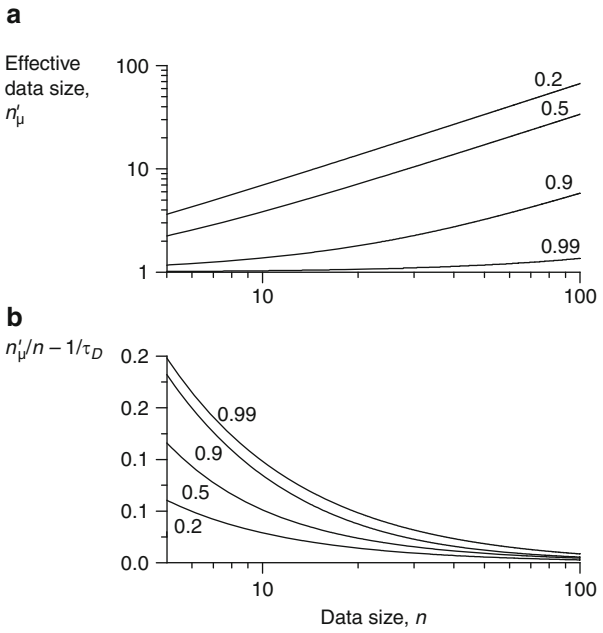

Fig. 2.10 Effective data size, mean estimation of an AR(1) process. (**a**) Dependence of n'_μ on n after Eq. (2.7), for various persistence values a (0.2, 0.5, 0.9 and 0.99). (**b**) Comparison of the exact expression (Eq. 2.7) with a simplified version based on the decorrelation time (Eq. 2.8)

The **covariance** between two random variables, X and Y, is

$$COV\,[X,Y] = E\Big[\big(X - E\,[X]\big)\cdot\big(Y - E\,[Y]\big)\Big]. \tag{2.34}$$

A special case is $COV\,[X,X] = VAR\,[X]$.

The **effective data size** is reduced for a persistent process. This is shown (Fig. 2.10a) for the case of mean estimation of an AR(1) process, where even for moderate values ($n \gtrapprox 50$ and $a \lessapprox 0.5$), n'_μ is considerably smaller than n. The data size reduction is quantified by Eq. (2.7). A simplified version based on the decorrelation time (Eq. 2.8) underestimates n'_μ by less than 5 % for $n \gtrapprox 50$ and $a \lessapprox 0.5$, as shown by Fig. 2.10b. Even for moderate values of n ($\gtrapprox 50$) and a ($\lessapprox 0.5$), the influence of persistence on n'_μ can be considerable. Eq. (2.3) is valid only for the mean estimator. The effective data size for variance estimation of an AR(1) process,

$$n'_{\sigma^2} = n\left[1 + 2\sum_{i=1}^{n-1}(1 - i/n)\,\rho(i)^2\right]^{-1} \tag{2.35}$$

(Bayley and Hammersley 1946), is given by

$$n'_{\sigma^2} = n \left\{ 1 + \frac{2}{n} \frac{1}{1-a^2} \left[a^2 \left(n - \frac{1}{1-a^2} \right) - a^{2n} \left(1 - \frac{1}{1-a^2} \right) \right] \right\}^{-1}. \quad (2.36)$$

Likewise, the effective data size for correlation estimation between two processes $X(i)$ and $Y(i)$ with autocorrelation functions $\rho_X(i)$ and $\rho_Y(i)$,

$$n'_\rho = n \left[1 + 2 \sum_{i=1}^{n-1} (1 - i/n)\, \rho_X(i)\, \rho_Y(i) \right]^{-1} \quad (2.37)$$

(von Storch and Zwiers 1999), is in the case of two AR(1) processes with persistence parameters a_X and a_Y given by

$$n'_\rho = n \left\{ 1 + \frac{2}{n} \frac{1}{1-a_X a_Y} \left[a_X a_Y \left(n - \frac{1}{1-a_X a_Y} \right) - (a_X a_Y)^n \left(1 - \frac{1}{1-a_X a_Y} \right) \right] \right\}^{-1}. \quad (2.38)$$

Early papers in climatology on effective data size and the influence of persistence on estimation variance include Matalas and Langbein (1962), Leith (1973), Laurmann and Gates (1977), Thiébaux and Zwiers (1984), Trenberth (1984a,b) and Zwiers and von Storch (1995).

Various **approximate bias** (and variance) formulas have been published for estimators of the autocorrelation parameter a in evenly spaced AR(1) models (Eqs. 2.1 and 2.33). Marriott and Pope (1954) analysed $\hat{a}$ (Eq. 2.4) and gave

$$E(\hat{a}) \simeq (1 - 2/n)\, a. \quad (2.39)$$

White (1961) gave an approximation of higher order in terms of powers of $(1/n)$:

$$E(\hat{a}) \simeq \left(1 - 2/n + 4/n^2 - 2/n^3\right) a + \left(2/n^2\right) a^3 + \left(2/n^2\right) a^5. \quad (2.40)$$

One drawback of these approximations is that they are not accurate for large a. For $a \to 1$, also $\hat{a} \to 1$ (Eq. 2.4) and the bias, $E(\hat{a}) - a$, approaches zero. This behaviour is not contained in Eqs. (2.39) or (2.40). It is, however, contained in the bias formula of Mudelsee (2001a):

$$E(\hat{a}) \simeq [1 - 2/(n-1)]\, a + \left[2/(n-1)^2\right] \left(a - a^{2n-1}\right) / \left(1 - a^2\right). \quad (2.41)$$

Mudelsee (2001a) showed that this approximation is more accurate than Eq. (2.40) for $a \gtrapprox 0.88$. The estimation variance of $\hat{a}$ is to a low approximation order (Bartlett 1946)

$$VAR(\hat{a}) \simeq \left(1 - a^2\right)/n \quad (2.42)$$

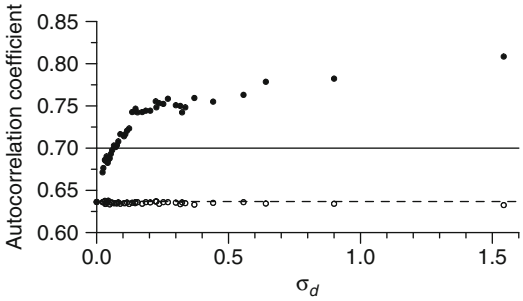

Fig. 2.11 Monte Carlo study of the bias in the autocorrelation estimation of an AR(1) process, unknown mean, uneven spacing. Identical Monte Carlo parameters and time series properties were used as in Fig. 2.3. The estimators are also the same, with the exception that the sample mean was removed from the time series prior to the estimations. The negative bias approximation (*dashed line*) is from the case of even spacing (Kendall 1954). See Fig. 2.3 for further explanation

and to a higher order (White 1961)

$$VAR\,(\hat{a}) \simeq \left(1/n - 1/n^2 + 5/n^3\right) - \left(1/n - 9/n^2 + 53/n^3\right) a^2 - \left(12/n^3\right) a^4. \tag{2.43}$$

Higher-order approximations of the first four moments of $\hat{a}$ are given by Shenton and Johnson (1965). From a practical point of view, it is more realistic to assume that the mean of $X_{\text{noise}}(i)$ is unknown and has to be subtracted. In case of the AR(1) process with unknown mean, the analogue of the estimator in Eq. (2.4) is

$$\hat{a} = \sum_{i=2}^{n} \left[x_{\text{noise}}(i) - \bar{x}_{\text{noise}}\right] \cdot \left[x_{\text{noise}}(i-1) - \bar{x}_{\text{noise}}\right] \Bigg/ \sum_{i=2}^{n} \left[x_{\text{noise}}(i) - \bar{x}_{\text{noise}}\right]^2, \tag{2.44}$$

where $\bar{x}_{\text{noise}} = \sum_{i=1}^{n} x_{\text{noise}}(i)/n$ is the sample mean. The approximate expectation of this estimator is (Kendall 1954)

$$E\,(\hat{a}) \simeq a - (1 + 3a)\,/\,(n-1)\,. \tag{2.45}$$

Monte Carlo simulations (Fig. 2.11) indicate that this approximation can be used for bias correction in situations with uneven spacing and moderate autocorrelation (a less than, say, 0.9). The bias-corrected autocorrelation coefficient, $\hat{a}'$, is obtained from $\hat{a}$ by inserting $\hat{a}'$ for a on the right-hand side and $\hat{a}$ for $E(\hat{a})$ on the left-hand side in one of the equations describing the bias, say Eq. (2.45), and solving this equation for $\hat{a}'$. Approximations for the bias of least-squares and Yule–Walker estimators of AR(p) processes with known/unknown mean have been given for $p \leq 6$ by Shaman and Stine (1988). Sample mean subtraction is a special case of detrending. It follows that if we obtain $\{x_{\text{noise}}(i)\}_{i=1}^{n}$ from the data $\{x(i)\}_{i=1}^{n}$ by removing an estimated trend function, $\{x_{\text{trend}}(i)\}_{i=1}^{n}$ (Eq. 1.2), in principle we have

to replace $\bar{x}_{\text{noise}}$ in Eq. (2.44) by $x_{\text{trend}}(i)$. For trends more complex than a constant function, bias properties of such estimators seem, however, to be analytically untractable.

Another AR(1) parameter estimator was introduced (Houseman 2005), based on an estimation function (which has zero expectation at the true parameter value). This author gave the estimation function robustness with respect to outliers (Chap. 3), included a linear regression term (Chap. 4), performed a joint estimation and presented an application to unevenly spaced water monitoring time series from Boston Harbor.

ARMA(p,q) estimation. A least-squares estimation (Brockwell and Davis 1991: Chap. 8 therein) can be used to fit even-spacing ARMA(p,q) models to data $\{x_{\text{noise}}(i)\}_{i=1}^{n}$. Besides least squares, statistical practice normally uses the maximum likelihood principle, which means to search for the ARMA(p,q) parameters $\{a_1,\ldots,a_p;b_0,\ldots,b_q\}$ that maximize the likelihood that the fitted model had produced the data (Brockwell and Davis 1991: Chap. 8 therein). This may be numerically difficult but is no fundamental restriction. Another important point is model identification, that is, selection of p and q. Guidance for that gives the sample autocorrelation function (Eq. 2.17), which can be compared with the autocorrelation function of the model candidate; analogously used are the partial autocorrelation functions (Brockwell and Davis 1991: Chap. 9 therein). Some quantitative measures of goodness of fit exist, such as Akaike's (1973) information criterion (AIC). A bias-corrected version of the AIC, referred to as the AICC (Hurvich and Tsai 1989), is calculated from the maximized likelihood, L, plus a penalty term for the number of parameters,

$$\text{AICC} = -2\ln(L) + 2(p+q+1)n/(n-p-q-2). \tag{2.46}$$

The penalty is a mathematical expression of Ockham's razor: It is easier to fit a model with more parameters, or alternatively, it is preferable to explain the data using a model with fewer parameters. The context of this book, uneven spacing and the embedding problem for ARMA(p,q) processes, forces us to take a rather parsimonious position by adopting the AR(1) model ($p=1, q=0$). It could therefore be that the parametric AR(1) model misses some properties, representable in higher-order continuous-time models, of the observed noise process. A milder case is when the spacing arises from equidistance with some missing observations (Fig. 1.16). Then a possible solution is to use a state-space representation, Kalman filtering and an adaption of the likelihood function. This approach has been pioneered by Jones (1981, 1985, 1986) and Jones and Tryon (1987).

A **maximum likelihood estimate for the AR(1) model** with uneven spacing is given by Robinson (1977). For even and uneven spacing, it is useful to visually check the fit residuals (predictions for $x_{\text{noise}}(i)$ by the fitted model minus the data $x_{\text{noise}}(i)$), which are for a proper fit realizations of a purely random process. For higher-order ARMA(p,q) models, it seems that no one embarked on the derivation of analytical approximations of the estimation bias.

ARFIMA(p, δ, q) models were introduced by Granger and Joyeux (1980) and Hosking (1981). For $|\delta| < 0.5$, the fractional difference operator $(1 - B)^{\delta}$ is defined by

$$\begin{aligned}(1-B)^{\delta} &= \sum_{k=0}^{\infty} \binom{\delta}{k} (-B)^k \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(\delta+1)}{\Gamma(k+1)\,\Gamma(\delta-k+1)} (-B)^k ,\end{aligned} \tag{2.47}$$

where B is the backshift operator and $\Gamma(\cdot)$ is the gamma function. Maximum likelihood and other estimation techniques are described by Beran (1994, 1997).

Hypothesis tests for long memory (see Sect. 3.6 for tests) and other inference tools, with an emphasis on finite-sample properties, are considered by Reisen and Lopes (1999), Couillard and Davison (2005), Hamed (2007), Ohanissian et al. (2008) and Magdziarz et al. (2013). A recent review is by Witt and Malamud (2013). The relation between DFA and spectral analysis (Chap. 5) is shown by Heneghan and McDarby (2000).

Nonlinear persistence models can be elegantly formulated in a generalized manner (Battaglia and Protopapas 2012a) to include thresholds not only in X (e.g. SETAR processes) but also in T (cf. Sect. 4.2). These authors applied that approach to instrumental temperature records from the Alpine region and detected an accelerated warming; their methods and results were discussed in a joint issue of the journal Statistical Methods and Applications from the viewpoints of statistical science (Giordano et al. 2012; Piccolo 2012; Tong 2012) and climatology (Böhm 2012; Mudelsee 2012a); upon which the authors responded (Battaglia and Protopapas 2012b).

Double-logarithmic transformations followed by linear regression are generally not suited to estimate power-law models when the original data have Gaussian error distributions. Although this has since long been known in various disciplines (Rützel 1976; Freund and Minton 1979; Miller 1984; Jansson 1985; Newman 1993; Mudelsee and Stattegger 1994; Xiao et al. 2011), the transformation is still frequently encountered in various applied sciences today. The low power of the double-logarithmic plot to discriminate between scaling and no scaling in noisy data has been criticized by Tsonis and Elsner (1995). These authors suggest a test for scaling, namely, to plot the slope with bootstrap confidence interval (Chap. 3) against the scale and look for a plateau (constant slope). However, for too small data sizes, no plateau behaviour might be found, despite the existence of scaling. Maraun et al. (2004) showed in a Monte Carlo study analysing artificial time series with DFA that well over 100,000 data points may be required. This result together with the other criticisms makes DFA irrelevant for climate time series analysis.

A **random walk** process arises from the case $a = 1$ in Eq. (2.33). For this process, $VAR[X_{\text{noise}}(i)]$ increases linearly with time, that is, the random walk is not stationary. In climatology, where the variables are within certain bounds, the

random walk has to be modified to serve as a noise model. In that manner, it has been applied to short-term temperature fluctuations (Kärner 2002). In case of Pleistocene timescales (Table 1.1), Wunsch (2003) suggested a random walk for explaining the 100-ka ice-age cycle. He put bounds to the ice-volume variable $X_{\text{noise}}(i)$; when the system attempts to leave the permitted range, it is thrown back. Mostly other fields than climatology, such as econometrics, apply tests of the hypothesis "$a = 1$" for the autoregressive process, or its generalization ("unit-root tests") for the ARMA(p, q) process (Fuller 1996: Chap. 10 therein). Several bootstrap hypothesis tests (Chap. 3) for unit roots were examined by means of Monte Carlo simulations (Palm et al. 2008). A climatological application is the paper by Stern and Kaufmann (2000), where unit roots were identified in tests of hemispheric temperature records, circa 1855–1995. Because these tests generally have poor power (loosely speaking, detection probability; see Eq. (3.41)) (Chatfield 2004: Sect. 13.4 therein), and because of the nonstationarity, we will not consider further random walk models for $X_{\text{noise}}(i)$.

Further **examples** of time series models fitted to climate data are the following. It is fair to say that the vast majority of such papers used the AR(1) process with $a > 0$ as a model of $X_{\text{noise}}(i)$. One classic, from meteorology, is Gilman et al. (1963). As an example from late Pleistocene climatology, we cite Roe and Steig (2004), who characterize Arctic and Antarctic climate by means of the AR(1) persistence time. Since this book adopts the AR(1) noise model, we will encounter various applications of it in the following chapters. It may be noted that in hydrology, cases of $a < 0$ (anti-persistence or blue noise) are found (Milly and Wetherald 2002). Annual layer thickness in ice cores may also exhibit blue noise on very short (annual) timescales (Fisher et al. 1985): if the true thickness of a layer is, say, larger than measured, then the true thickness of a neighbouring layer is likely smaller than measured (since the overall thickness is constrained). Yule (1927) fitted an AR(2) model to sunspot data, 1749–1924. These data (Fig. 2.12), which exhibit quasiperiodic behaviour with period ∼11 a, have since this pioneering work been the hobbyhorse of time series analysts. Tong and Lim (1980) took a SETAR$(2; 4, 12)$ process, which reproduces the sunspot cycles' asymmetry (rise and descent). Jones (1981) fitted various ARMA(p, q) models, and Priestley (1981: Sect. 11 therein) compared the fits of AR(p), ARMA(p, q) and threshold autoregressive models to the sunspot data. Seleshi et al. (1994) fitted a high-order autoregressive model to the sunspot data but found that AR(p) or ARMA(p, q) models gave no satisfactory fit to the rainfall series, 1900–1991, from Addis Ababa. This stimulated their search for a transformation of the rainfall data which could produce a better relation with solar activity variations. Matyasovszky (2001) fitted an AR(4) and threshold autoregressive models to the longest record of monthly instrumental observations, the central England temperature time series (Manley 1974), which starts in January 1659. He found four regimes, one of which has a limit cycle of about 2 a period and might correspond to the meteorological phenomenon of the quasi-biennial oscillation, which refers to zonal wind-speed variations in the tropical stratosphere. Stedinger and Crainiceanu (2001) considered ARMA(p, q) models for the logarithmically transformed maximum annual runoff of the river Missouri

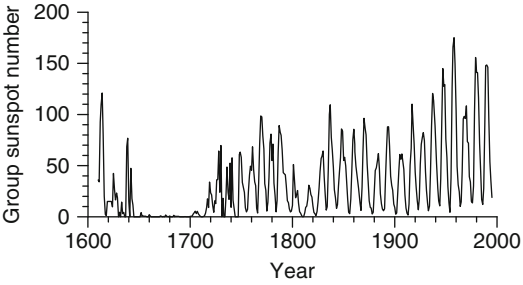

Fig. 2.12 Group sunspot number, 1610–1995. Sunspots are dark spots on the Sun's surface, visible from the Earth with a telescope. They present regions of reduced temperature. Satellite measurements, available since 1980, show that the solar activity correlates positively with the number of sunspots (Willson and Hudson 1988), that is, the solar constant is no constant. The group sunspot number is a way of counting the sunspots as groups and thought to give a more accurate picture over the previous centuries than using the individual sunspot data (Hoyt and Schatten 1998). The long lasting minimum during approximately 1645–1715 is the Maunder Minimum. Beer et al. (1998) demonstrate that this was not a period without solar activity variations (Data have $d(i) = 1$ a and are from Hoyt and Schatten (1998))

from 1898 to 1998. Koen and Lombard (1993) applied ARMA(p, q) modelling to astronomical time series. Stattegger (1986) used ARMA(p, q) processes to describe variations in the composition of heavy metals in sedimentary deposits of rivers in Austria. Newton et al. (1991) fitted ARMA(p, q) models to the Pleistocene SPECMAP δ^{18}O curve. Giese et al. (1999) searched for suitable ARMA(p, q) and ARIMA models for δ^{18}O variations in the Pliocene–Pleistocene deep-sea sediment core V28-239 without success. Stephenson et al. (2000) analysed the wintertime North Atlantic Oscillation (NAO) index from 1864 to 1998, that is, the difference between standardized December–March mean sea-level pressures measured at Lisbon and Iceland. (Other versions exist, which use Azores instead of Lisbon. A proxy record of the NAO that goes back to A.D. 1049 was constructed (Trouet et al. 2009) on basis of speleothem data from Scotland and tree-ring data from Morocco.) The NAO index is used as a measure to summarize the mean westerly atmospheric flow over the North Atlantic region, which in turn influences the weather in Europe (Hurrell 1995). It was found (Stephenson et al. 2000) that an ARFIMA(1, 0.13, 0) model describes the data as good as an AR(10) model and better than the asymptotic stationary AR(1) or random walk models. The more parsimonious long-memory model was preferred to make NAO forecasts. Divine et al. (2008) presented a model of piecewise AR(1) processes and a maximum likelihood technique for estimating the autocorrelation parameters and the change-point times. This model constitutes an interesting augmentation of the SETAR$(l; k, \ldots, k)$ model. Divine et al. (2008) applied their method to detect changes in records of the NAO, ENSO and ice core δ^{18}O. Kallache et al. (2005) fitted ARFIMA(p, δ, q) models to runoff records from small rivers in southern Germany and found via an AICC variant that the parsimonious AR(1) model, contained in the ARFIMA model, is often a suitable description but also that frequently a nonzero long-term parameter δ is

required. Vyushin et al. (2012) compared fitting short-memory AR(1) and long-memory ARFIMA processes to gridded monthly temperature reconstructions for the interval from 1957 to 2002. The reconstructions come from the Coupled Model Intercomparison Project 3 (CMIP3). These authors found that both persistence representations fit the detrended data equally well according to goodness-of-fit tests and that both are potentially useful for statistical applications. Lovejoy (2013), on the other hand, asserted that temperature series from various scales (instrumental period back to the late Pleistocene) exhibit different types of scaling. He suggested (Lovejoy 2013: p. 2 therein) that the "prevailing weather–climate dichotomy" (boundary at 30 a scale) "should be replaced by a weather–macroweather–climate trichotomy", with the new third entity in scale between 10–30 years and 5–10 days. Lovejoy (2013) did not consider trends. These examples illustrate that there is an interplay between what is put into the noise process and what is put into the trend. Taking more complex noise processes does reduce the need to consider complex trend functions. This book is devoted to using simpler noise processes and more complex trend functions. Besides the necessity to keep the noise simple because of the embedding problem (Sect. 2.3), this position allows to quantify trend parameters, for example, climate changes, using regressions (Chap. 4).

Conceptual climate models are simple mathematical renderings of the physical processes deemed essential for describing a climate subsystem. These models often have less spectral power (Chap. 5) than the observed processes they are intended to simulate. Stochastic processes, added to the mathematical formulation of conceptual climate models, can fill this "spectral gap", as explained by Wilks (2010). The model equations become stochastic differential equations.

2.7 Technical Issues

The **gamma distribution** in its standard form with order parameter a has the following PDF:

$$f(x) = x^{a-1} \exp(-x) / \Gamma(a) , \qquad x \geq 0. \tag{2.48}$$

Ahrens and Dieter (1974) devised an algorithm for generating gamma random variables. See Johnson et al. (1994: Chap. 17 therein) for more details on the gamma distribution.

The **gamma function** is defined by

$$\Gamma(z) = \int_0^\infty y^{z-1} \exp(-y) dy. \tag{2.49}$$

When z is an integer, $\Gamma(z+1) = z!$; otherwise, approximations have to be used for calculation. Lanczos (1964) devised an algorithm for approximating $\ln[\Gamma(z)]$.

Ratios of gamma functions can then be numerically advantageously evaluated as $\Gamma(a)/\Gamma(b) = \exp\{\ln[\Gamma(a)] - \ln[\Gamma(b)]\}$.

The **beta distribution** in its standard form with parameters $p > 0$ and $q > 0$ has the following PDF:

$$f(x) = x^{p-1}(1-x)^{q-1}/B(p,q)\,, \qquad 0 \le x \le 1. \tag{2.50}$$

The beta function, B, is given by $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$. See Johnson et al. (1995: Chap. 25 therein) for more details on the beta distribution.

TAUEST is a Fortran 77 program that estimates the persistence time of an AR(1) process and uneven spacing. The minimization of the least-squares sum (Eq. 2.11) is done using Brent's search (Press et al. 1992). The software includes residual analysis, bias correction and construction of a bootstrap percentile confidence interval (which is described in Chap. 3). TAUEST is described by Mudelsee (2002). The software is available at the web site for this book.

Jones (1981) gave Fortran 77 subroutines to fit data to a continuous-time autoregressive process. This program uses a state-space representation and a maximum likelihood principle.

ITSM 2000 is a Windows package for the estimation of ARMA, ARFIMA and other models under even spacing. It includes tools for transformations, regression, forecasting, smoothing and spectral estimation. A version for a limited data size is included in the book by Brockwell and Davis (1996), a full version can be obtained from B & D Enterprises, Inc. (`pjbrock@stat.colostate.edu`, email from 16 April 2004).

STAR is a DOS program for fitting threshold autoregressive models to evenly spaced time series. The book by Tong (1990) informs that the software could be obtained from Microstar Software (Canterbury, UK).

S-Plus routines for fitting long-memory models to data are listed in the monograph by Beran (1994). This book contains also an S-Plus routine for the simulation of ARFIMA processes. Hosking (1984) gives another algorithm for ARFIMA simulation.

Ox is a computer language, for which a package (Doornik and Ooms 2001) for maximum likelihood fitting of ARFIMA models is available (http://www.doornik.com, 9 November 2013).

OxMetrics is a commercial software package that includes fitting of persistence models to data (http://www.timberlake.co.uk/software/?id=64, 9 November 2013).

Random number generators are required to perform Monte Carlo experiments such as that from Fig. 2.3. Also bootstrap resampling (Chap. 3) uses such tools. (Strictly speaking, physical experiments with uncontrollable factors can be performed as well. It is interesting to note that the discoverer of the Hurst phenomenon used a pack of cards (Hurst 1957), a procedure that appears even for that time period somewhat anachronistic.) Almost exclusively employed today are pseudorandom numbers, which are generated by mathematical algorithms. One simple form is the multiplicative congruential generator, $Z_i = A\,Z_{i-1} \pmod M$, $i \ge 1$, where A, M and Z_i (pseudorandom numbers) are integers. Z_1 is primed (seeded). Z_i can be

mapped onto the interval $[0; 1]$ to produce a uniform distribution. The uniform serves also as basis to generate other types of distribution. For example, the Gaussian arises from the uniform by the transformation given by Box and Muller (1958). The success of Monte Carlo experiments and bootstrap resampling depends critically on whether the generator "supplies sequences of numbers from which arbitrarily selected nonoverlapping subsequences appear to behave like statistically independent sequences and where the variation in an arbitrarily chosen subsequence of length k (≥ 1) resembles that of a sample drawn from the uniform distribution on the k-dimensional unit hypercube" (Fishman 1996: p. 587 therein). Uniform filling of the hypercube can be assessed by inspecting whether regular patterns are absent in two- or three-dimensional hyperplanes. Park and Miller (1988) show that good random number generators "are hard to find". They give also a multiplicative congruential generator with $A = 16{,}807$ and $M = 2{,}147{,}483{,}647$, which may serve as minimal standard. Schrage (1979) lists a Fortran 77 code of this generator type. Other generators are described by Press et al. (1992: Chap. 7 therein), Fishman (1996: Chap. 7 therein) and Knuth (2001: Chap. 3 therein); the latter book (Sect. 3.4.2, Algorithm P therein) contains a recipe for producing random permutations. Multiprocessor machines can be advantageously used to produce parallel streams of random numbers. L'Ecuyer et al. (2002) describe such a package, which is available in C++ (http://www.iro.umontreal.ca/~lecuyer, 9 November 2013); a Fortran 90 implementation is included in the molecular dynamics package CP2K (http://www.cp2k.org, 9 November 2013). Also the web site for this book has a zipped Fortran 90 package (rng.7z), which is based on the random number generator by Marsaglia and Zaman (1994).

http://www.springer.com/978-3-319-04449-1

Climate Time Series Analysis
Classical Statistical and Bootstrap Methods
Mudelsee, M.
2014, XXXII, 454 p. 103 illus., Hardcover
ISBN: 978-3-319-04449-1