# Time Series Analysis

Klaus Nordhausen
Department of Mathematics and Statistics, University of Jyväskylä, Finland
klaus.k.nordhausen@jyu.fi

## Definition

A time series is a collection of data taken sequentially over time and has therefore a natural one-way order where it can be assumed that the current value is more influenced by the recent past than by observations from a long time ago. Time series analysis comprises methods for the analysis of time series which take this structure into account when extracting meaningful statistics from the data. Time series forecasting concerns the prediction of future values based on an observed time series.

## 1 Time Series Data

In geosciences collecting data that depend on time has been measured for centuries. For example, temperatures have been observed in standardized ways since approximately 1860. But also recordings of river discharge have a long tradition. Records of such phenomena varying irregularly in time and when sequentially observed are called time series. If the measurements are recorded continuously, like from a seismograph, the time series is called continuous time series. On the other hand if the measurements are taken at certain, usually regular, intervals like hourly, daily, monthly, or yearly they are denoted discrete time series. Discrete time series with equally spaced intervals are the most common form of time series data and also the focus of this article.

At a given time point $t$ it is possible to observe a single observation like temperature, a vector of observations like temperature, precipitation, and wind speed, or a curve like the temperature curve at day $t$. Then one distinguishes between univariate, multivariate and functional time series respectively.
However, independent of the type of time series, the irregularly varying nature of the recordings is addressed by assuming the observed time series is a realization of a stochastic model. And the most convenient assumption is that the stochastic model has some time invariant structures in which case it is called stationary. If however the stochastic model itself depends on time it is called nonstationary.
The last differentiation between types of time series made here is between linear and nonlinear time series. If the stochastic model can be expressed as the output of a linear model, it is called a linear time series and if not a nonlinear time series.

## 2 Objectives of Time Series Analysis

Time Series Analysis has three main objectives:

1. **Description**: Methods that describe or summarize characteristics of the time series. Thus this contains tools for visualization or appropriate summary statistics which help to capture the essential features of the time series.

2. **Modeling**: To represent the stochastic nature of the observed time series many different models have been suggested which help understanding the nature of the process. An adequate model needs to be selected and its parameters need to be estimated. The choice of a model is often based on the descriptive measures.

3. **Prediction**: Due to the serial nature of the time series future values can be predicted based on the past and present values. To predict the future it is important to have a model which represents the data well.

### 2.1 Descriptive Tools for Time Series

The first step of all time series analysis is its visualization. A typical time series plot is given in Figure 1 which gives the monthly average temperatures in centigrade (C) at the airport of Jyväskylä in central Finland from 1960 to 2005.
Usually in such a figure for a time series $x_t, t = 1, \ldots, T$ one evaluates if the mean over time has a tendency, which is usually denoted as trend $m_t$ and if there are

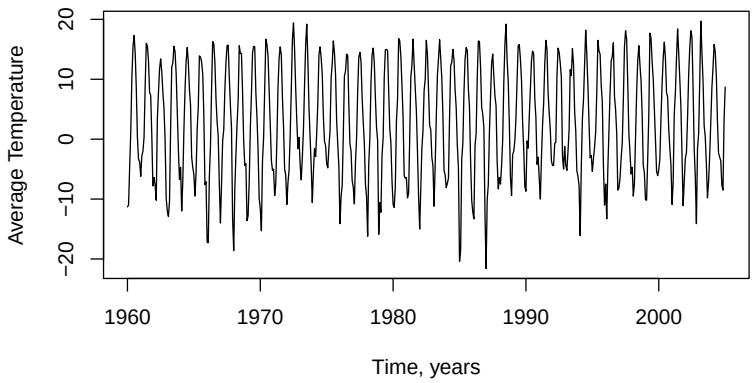


Figure 1: Average monthly temperature at the airport in Jyväskylä in C from 1960 to 2005.

regularly appearing patterns known as seasonalities $s_t$. If both are present one can consider an additive decomposition of the time series

$$x_t = m_t + s_t + r_t, \quad t = 1, \dots, T,$$

where $r_t$ represents the stochastic feature of data.
To obtain an idea of the trend, usually moving averages of order $n$ are used, which are defined as

$$m_t^n = \frac{1}{n} \sum_{j=t-k}^{t+k} x_j,$$

where $n = 2k + 1$. The larger $n$ the smoother the trend estimate will be.
The seasonal component is assumed to reoccur regularly. For example, for our monthly temperature data from Figure 1 visual inspection indicates no trend, but the values every $l = 12$ months appear to be similar.
To obtain an estimate for a seasonal component of length $l$ one usually estimates first the trend with $n = l$ if $l$ is odd like for weekly data or $n = 2l$ if $n$ is even as for monthly data. The estimate of the monthly effect is then the average of each month of the detrended series standardized so that their sum is zero.
The random reminder is then simply $r_t = x_t - m_t - s_t$.
For simplicity, we consider for now the observed time series $x_t$ as a realization of a stochastic process $X_t$ $t = 0, \pm 1, \dots, \pm\infty$ without a trend or seasonal component, like for example the reminder $r_t$ from above.
The mean value function of $X_t$ is $\mu_t = E(X_t)$ and the autocovariance function

at times $t$ and $t-\tau$ is defined as $Cov(X_t, X_{t-\tau}) = E((X_t - \mu_t)(X_{t-\tau} - \mu_{t-\tau}))$. For $\tau = 0$ this reduces to the variance at time $t$. Thus the serial dependence of time series data is captured especially in the autocovariance function. However, when all these characteristics depend on $t$ they are not very tractable and therefore a common practical assumption is that

$$E(X_t) = E(X_{t+\tau}) = \mu, \quad Var(X_t) = Var(X_{t+\tau}) = \sigma^2$$

and

$$Cov(X_t, X_{t-\tau}) = Cov(X_{t-l}, X_{t-l-\tau}) = C_\tau,$$

for all $\tau$ and all $l$. Thus the mean and variance do not depend on the time and the autocovariance depends only on the distance between the observations. Such a time series is called (weakly) stationary.
For an observed univariate time series $x_t, t = 1, \ldots, T$ then summary statistics to describe these time series features are under the stationarity assumption:

*Empirical mean:* $\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_t$,

*Empirical variance:* $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T}(x_t - \hat{\mu})^2$ and

*Empirical autocovariance:* $\hat{C}_\tau = \frac{1}{T}\sum_{i=\tau+1}^{T}(x_t - \hat{\mu})(x_{t-\tau} - \hat{\mu})$.

In addition, also the empirical autocorrelation $\hat{\gamma}_\tau = \hat{C}_\tau/\hat{\sigma}^2$ is often considered.
When the analysis of the time series is based mainly on the autocorrelation structure one says one does the analysis in the time domain. Another approach is to analyse the data in the frequency domain. Then by emplyoing spectral analysis the time series is decomposed into trigonometric functions at each frequency and its features are represented with the weights given to the periodic components.
If the autocovariance function $C_\tau$ rapidly decreases with increasing lag $\tau$ and satisfies

$$\sum_{\tau=-\infty}^{\tau=\infty} |C_\tau| < \infty,$$

one can define the Fourier transform of $C_\tau$. For frequencies $-0.5 \leq w \leq 0.5$ the power spectral density function is then

$$p(w) = \sum_{\tau=-\infty}^{\tau=\infty} C_\tau e^{-2\pi i \tau w}.$$

For an observed time series $x_1, \dots, x_T$ one uses then the natural frequencies $w_j = j/T$, $j = 0, \dots, \lfloor N/2 \rfloor$ to obtain the periodogram

$$p_j = \sum_{\tau=-T+1}^{T-1} \hat{C}_\tau e^{-2\pi i \tau w_j} = \hat{C}_0 + 2 \sum_{\tau=1}^{T-1} \hat{C}_\tau \cos(2\pi\tau w_j),$$

which can be used to obtain the sample spectrum

$$\hat{p}(w) = \sum_{\tau=-T+1}^{T-1} \hat{C}_\tau e^{-2\pi i \tau w_j}$$

by extending the domain to the continuous interval $[0, 0.5]$. Thus, there is a direct relation between the sample spectrum and the sample autocovariance

$$\hat{C}_\tau = \int_{-0.5}^{0.5} \hat{p}(w) e^{2\pi i \tau w} dw, \quad \tau = 0, \dots, T-1.$$

Notice that in practice the periodogram is usually computed with the Fast Fourier Transform (FFT) and in the time domain plots of the autocorrelation function are often inspected while in the frequency domain the periodogram is considered. Which approach is preferable often depends on the concrete application.
For more details on these topics see for example Gilge (2006); Kitagawa (2010); Shumway & Stoffer (2017).

## 2.2 Time Series Modelling

Exploring the autocorrelation structure and/or the periodogram of a time series provides the starting point for modelling.
A highly popular class of time series models are Autoregressive Moving Average (ARMA) models. For this purpose, we define first a white noise process. A process $\epsilon_t$ is called white noise process if $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma^2$ and $Cov(\epsilon_t, \epsilon_{t-\tau}) = 0$ for all $t$ and all $\tau > 0$. Thus a white noise process is centered with constant variance and serially uncorrelated. In the following, we will also assume that $\epsilon_t$ is Gaussian.
An ARMA model expresses then a time series $x_t$ as a linear combination of its past values $x_{t-i}$ and a white noise process $\epsilon_t$:

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \epsilon_t - \sum_{i=1}^{q} \beta_i \epsilon_{t-i}.$$

In this model, $p$ is the autoregressive order with autoregressive coefficients $\alpha_1, \ldots, \alpha_p$ while $q$ is the moving average order with moving average coefficients $\beta_1, \ldots, \beta_q$. A times series $x_t$ which follows an ARMA model is called an ARMA process and is usually denoted ARMA(p,q). The special case ARMA(p,0) is called simply an autoregressive process AR(p) and ARMA(0,q) a moving average process MA(q). It is meanwhile well established that almost all stationary time series can be modelled using autoregressive processes.
Interesting is also that, under certain conditions, all ARMA(p,q) models can be expressed as MA($\infty$) models, i.e. as linear combinations of present and past realizations of a white noise series

$$x_t = \sum_{i=0}^{\infty} \phi_i \epsilon_{t-i}.$$

The coefficients $\phi_i$ in this representation are called the impulse response function and are given by the recursive expression

$$\phi_0 = 1 \quad \text{and} \quad \phi_i = \sum_{j=1}^{i} \alpha_j \phi_{i-j} - \beta_i, \; i = 1, 2, \ldots,$$

where $\alpha_j = 0$ for $j > p$ and $\beta_j = 0$ for $j > q$.
The impulse response functions can then be used to express the autocovariance function of an ARMA(p,q) model

$$C_0 = \sum_{i=1}^{p} \alpha_i C_i + \sigma^2 \left(1 - \sum_{i=1}^{q} \beta_i \phi_i\right)$$

and

$$C_\tau = \sum_{i=1}^{p} \alpha_i C_{\tau-i} - \sigma^2 \left(1 - \sum_{i=1}^{q} \beta_i \phi_{i-\tau}\right), \; \tau = 1, 2, \ldots.$$

The power spectrum of the same model is analogously

$$p(f) = \sigma^2 \frac{\left|1 - \sum_{j=1}^{q} \beta_j e^{-2\pi i j f}\right|^2}{\left|1 - \sum_{j=1}^{p} \alpha_j e^{-2\pi i j f}\right|^2}.$$

Useful in this context is also the partial autocorrelation $\gamma_{\tau\tau}$ of a stationary time series $x_t$ which gives the correlation between $x_t$ and $x_{t-\tau}$ where the linear depen-

dence on $\{x_{t-1}, \dots, x_{t-(\tau-1)}\}$ has been removed, i.e.

$$\hat{\gamma}_{\tau\tau} = cor(x_t, x_{t-\tau} | x_{t-1}, \dots, x_{t-(\tau-1)}).$$

Then plots of the autocorrelation and partial autocorrelations can give an indication of the AR and MA orders:

- For a stationary AR(p) process the autocorrelations decay exponentially and the partial autocorrelations are zero for $\tau > p$.
- For a stationary MA(q) process the autocorrelations are zero for $\tau > q$ and the partial autocorrelations decay exponentially.
- For a stationary ARMA(p,q) process both, autocorrelations and partial autocorrelations, decay exponentially.

Figure 2 contains three samples of length 500 from an AR(1), MA(1) and ARMA(1,1) process with $\alpha_1 = \beta_1 = 0.6$ together with the corresponding estimates autocorrelation and partial autocorrelation functions. Values within the blue lines in the autocorrelation and partial autocorrelation plots are usually considered not differ significantly from zero. Thus, the figure clearly shows how these functions can assist in order selection. Such visualizations of the (partial) autocorrelation function (acf) are often refereed to as (partial) autocorrelogram.
As the periodogram contains the same information also its visualization can be used to get an idea of the order of ARMA processes, here the peaks and through of $\log(\hat{p}(f))$ are of interest. See Kitagawa (2010) for details. Rather than a graphical selection of the orders $p$ and $q$ more common is to use model selection criteria such as AIC or BIC to select the order, and for example R (R Core Team 2020) has many tools to select ARMA orders and to estimate the coefficients.
Two concepts for AR and MA processes are still relevant. For an AR(p) process

$$1 - \alpha_1 x - \alpha_2 x^2 - \dots - \alpha_p x^p = 0$$

is called the AR characteristic equation. Similarly, for an MA(q) process

$$1 - \beta_1 x - \beta_2 x^2 - \dots - \beta_q x^p = 0$$

denotes the MA characteristic equation. An MA process is called invertible if all roots of the MA characteristic equation have an absolute value larger than 1. An AR process is only stationary when all roots of the AR characteristic equation also have absolute values exceeding 1. Thus in practical ARMA modelling to work

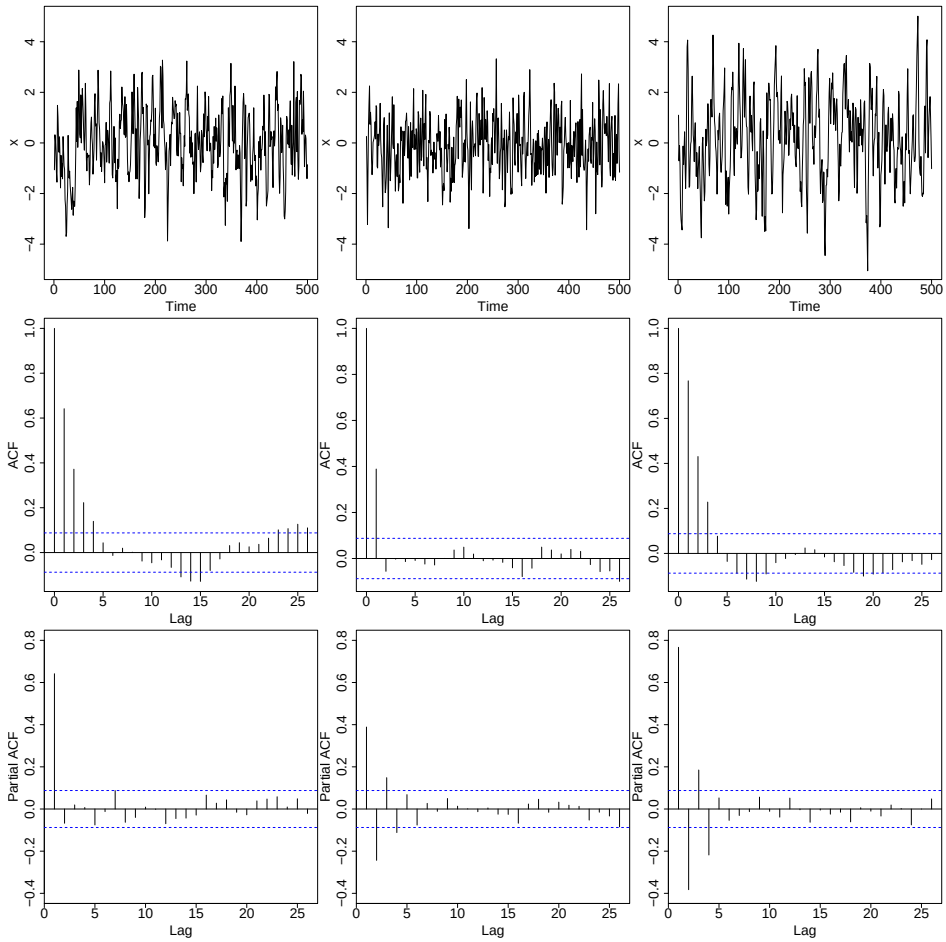


Figure 2: The top row shows from left to right a sample of T=500 observations from an AR(1) process, a MA(1) process and an ARMA(1,1) process where $\alpha_1 = \beta_1 = 0.6$. The second row shows the corresponding sample autocorrelation functions and the third the partial sample autocorrelation functions.

in a stationary framework it is required that the AR part is stationarity and the MA part invertible. Invertability for a time series means that we can transform it into an AR process, which means the time series can be expressed as a function of its past values. This is a key requirement when one is interested in predicting future values.

When we started out the modelling part we assumed that the mean of the time series is constant and does not contain a trend. In many situations, however, where the non-stationary of the time series comes from the non-constant mean, taking differences of the time series makes it more stationary. The first-order difference is defined as $\nabla x_t = x_t - x_{t-1}$ and the $d$th-order difference repeats this $d$ times and is denoted $\nabla^d x_t$. Using this approach to yield stationary ARMA processes was incorporated in the so called integrated autoregressive moving average (ARIMA) models of order (p,d,q) which means that $d$th difference $\nabla^d x_t$ follows a ARMA(p,q) model. In most practical situations $d$ is either 1 or 2.

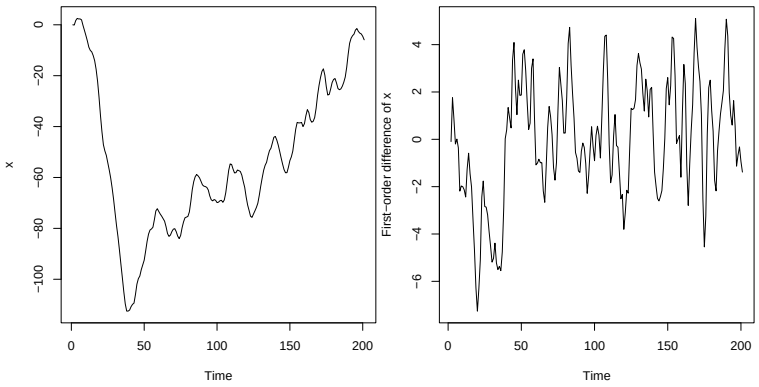


Figure 3: The left panel shows a sample of size 200 from an ARIMA(1,1,1) process with $\alpha_1 = \beta_1 = 0.6$. The right panel shows the first-order difference of the same process.

Figure 3 shows a sample of size 200 from an ARIMA(1,1,1) process with $\alpha_1 = \beta_1 = 0.6$ together with the first-order difference of the time series. While the original time series seems highly non-stationary the differences look much more stable. ARIMA models nowadays the standard tool when starting time series models and can also extended to allow for seasonal components. For further details about parameter estimation, model selection and inference in the context of ARIMA models see Brockwell & Davis (2020); Shumway & Stoffer (2017); Hassler (2019).

## 2.3 Time Series Forecasting

One of the primary reasons for time series analysis is the prediction of future values, which is usually called forecasting. Based on the available history of the process, $x_1, \ldots, x_t$, the goal is thus to forecast the value $x_{t+l}$, where $t$ is denotes the forecast origin and $l$ the lead time. The forecast itself is then denoted $\hat{x}_{t+l}$ and should be preferably the minimum square error forecast given by:

$$\hat{x}_{t+l} = E(x_{t+l}|x_1, \ldots, x_t).$$

Clearly, without strong assumptions such a prediction is difficult. However if the process is Gaussian it is known that the best predictor is a linear function of the available history. Assuming in addition an ARIMA process yields practical expressions for the predictions and the possibility to evaluate the uncertainties in the predictions by providing confidence intervals.
For stationary ARMA processes with $E(x_t) = \mu$ it holds that $\hat{x}_{t+l} - \mu$ decays to zero as lead $l$ increases. Which means that the further one wants to predict the future the more naive the forecast, approaching simply the mean value. This is however natural as the dependence in an ARMA process dies out and only for shorter leads we can improve the naive forecast and benefit from the available data. Details about ARIMA predictions are found in Brockwell & Davis (2020); Shumway & Stoffer (2017); Hassler (2019).

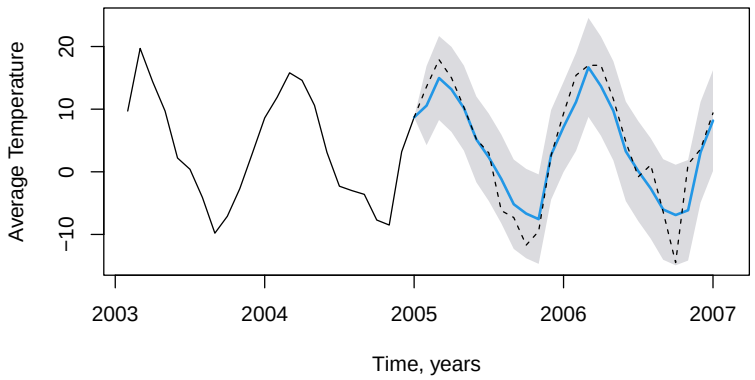


Figure 4: The predicted average monthly temperature for the airport in Jyväskylä in C for 2006 and 2007 (blue line). The shaded area corresponds to the 95% prediction interval and the dashed black line are the observed average temperatures.

Using the average monthly temperature data from Figure 1 a seasonal ARIMA

prediction for the years 2006 and 2007 is shown in Figure 4 as the blue line. The shaded grey area reflects the uncertainty of the predictions. The actual observed average temperatures are given as the dashed black line and shows that the predictions quite well, just forecasted less cold winters, but the true values are still within the grey areas.

## 2.4 Various extensions for univariate time series

The ARIMA model with $d \in \{0, 1, 2\}$ with a possible seasonal adjustment is one of the standard tools in time series analysis. They are however not always suitable and in the following some different time series settings and models are discussed.

**Time Series with Long Memory** In the ARIMA model as discussed above it assumed that the time series has a short memory which means that autocorrelation drops to zero after a certain lag or has an exponential rate of decay. For hydrological or meteorological time series like annual average rainfall, temperature or river flow data it can be however often observed that the autocorrelations are persistent over many lags. This persistence is then often denoted long memory. The extend of long memory is often quantified with the Hurst coefficient $H \in (0, 1)$ where 0.5 indicates the absence of long memory and a value larger than 0.5 indicates a persistent behaviour. Long memory time series modeling can be embedded in an ARIMA framework by allowing the parameter $d$ to take also fractional values, i.e. $d \in (0, 1)$ which is in detail discussed in Hassler (2019) and often denoted as fractional ARIMA modelling.

**Nonlinear time series** The ARMA model introduced above is linear in its terms which is mathematically convenient and proved to be able to approximate many natural phenomena, however not all. Some processes in nature are just not linear. The literature on nonlinear time series models is quite large and an overview of different applications of nonlinear time series models in geosciences is for example given in Donner & Barbosa (2008). In the following only two popular nonlinear time series models are introduced.

The simplest class of nonlinear models are piecewise linear models which are known as threshold models. In this class of models it is thought that there is a set of $K$ linear submodels and a mechanism that switches between the different submodels. A popular variant in this context is the threshold autoregressive

(TAR) model which has the form

$$x_t = \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} x_{t-i} + \beta^{(j)} \epsilon_t,$$

where $j \in \{1, \dots, K\}$ is an indicator for the active submodel and the submodels have the parameters $\alpha_0^{(j)}, \dots, \alpha_p^{(j)}$ and $\beta^{(j)}$ and $\epsilon_t$ is a white noise process. Which submodel is active is indicated by the series $z_t$ which takes values in $\{1, \dots, K\}$. If the state of $Z_t$ depends only on the past values of $x_t$, i.e., $z_t = j$ if $x_{t-\tau} \in R_j$ for some fixed $\tau > 0$ with $\bigcup_{j=1}^{K} R_j = \mathbb{R}$ then is a self-exciting TAR model. If however $z_t = j$ if and only if $y_{t-\tau} \in R_j$ where $y_t$ is an observable or unobservable covariate time series one says the TAR model is excited by $y_t$.

Another popular nonlinear model is the generalized autoregressive conditional heteroscedastic (GARCH) model. It can be formulated as

$$x_t = \sigma_t \epsilon_t,$$

where $\epsilon_t$ is a white noise process and

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

with model parameters $\alpha_i \geq 0$, $i = 0, \dots, p$, and $\beta_i \geq 0$, $i = 1, \dots, q$. Thus an ARMA(p,q) process is assumed for the variance of $x_t$ and for the series $x_t$ small values and large values should cluster.
Tests if nonlinear models are needed are well-established but have usually special nonlinear models as the alternative in mind. Similarly, parameter estimation and model selection for the above models are well investigated together with appropriate forecasting tools. For details about the above non-linear models as well as many other methods see for example Fan & Yao (2003).

Univariate time series methods from a geoscience perspective are for example also discussed in Gilge (2006).

## 3 Multivariate Time Series

So far we have assumed that only one variable is measured at time $t$. For multivariate time series at the time $t$ a $m$-variate vector $\boldsymbol{x}_t = (x_1, \dots, x_m)^\top$ is observed. Thus, besides the serial dependence, also the dependence between the different variables need to be considered. The characteristics of the underlying stochastic process are then described using the expected value $E(\boldsymbol{X}_t)$, the covariance matrix $\text{Cov}(\boldsymbol{X}_t) = E((\boldsymbol{X}_t - E(\boldsymbol{X}_t))(\boldsymbol{X}_t - E(\boldsymbol{X}_t))^\top)$ and the autocovariance matrix $\boldsymbol{C}_\tau(\boldsymbol{X}_t) = E((\boldsymbol{X}_t - E(\boldsymbol{X}_t))(\boldsymbol{X}_{t-\tau} - E(\boldsymbol{X}_{t-\tau}))\top)$. A key assumption is again that these quantities do not depend on the time and multivariate (weak) stationarity is given when

$$E(\boldsymbol{X}_t) = \boldsymbol{\mu}, \quad \text{Cov}(\boldsymbol{X}_t) = \boldsymbol{\Sigma}$$

and

$$\boldsymbol{C}_\tau(\boldsymbol{X}_t) = E((\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_{t-\tau} - \boldsymbol{\mu})^\top) = E((\boldsymbol{X}_{t-l} - \boldsymbol{\mu})(\boldsymbol{X}_{t-l-\tau} - \boldsymbol{\mu})^\top) = \boldsymbol{C}_\tau,$$

for all $\tau$ and $l$. For an observed time series $\boldsymbol{x}_t$, $t = 1, \dots, T$ the corresponding sample statistics are:

*Empirical mean:* $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{i=1}^{T} \boldsymbol{x}_t$,

*Empirical coariance:* $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})^\top$ and

*Empirical sutocovariance matrix:* $\hat{\boldsymbol{C}}_\tau = \frac{1}{T}\sum_{i=\tau+1}^{T}(\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_{t-\tau} - \hat{\boldsymbol{\mu}})^\top$.

The empirical autocorrelation matrix is obtained as $\hat{\boldsymbol{\Gamma}}_\tau = \boldsymbol{D}^{-1/2}\hat{\boldsymbol{C}}_\tau\boldsymbol{D}^{-1/2}$ where the diagonal matrix $\boldsymbol{D}$ contains the diagonal elements of $\hat{\boldsymbol{\Sigma}}$.

The univariate ARMA model has been extended to the multivariate case and is usually referred to as vector autoregressive moving avervage (VARMA) model

$$\boldsymbol{x}_t = \sum_{i=1}^{p} \boldsymbol{A}_i\boldsymbol{x}_{t-i} + \boldsymbol{\epsilon}_t - \sum_{i=1}^{q} \boldsymbol{B}_i\boldsymbol{\epsilon}_{t-i},$$

where $\boldsymbol{\epsilon}_t$ is an $m$-variate white noise process and the $m \times m$ matrices $\boldsymbol{A}_1, \dots, \boldsymbol{A}_p$ and $\boldsymbol{B}_1, \dots, \boldsymbol{B}_q$ are autoregressive and moving average coefficients, respectively. The special cases VARMA(p,0) is correspondingly referred to as vector autorgressive process VAR(p) and VARMA(0,q) as vector moving average process VMA(q). The conditions for stationarity and invertibility translate similarly to the multivariate

case as do order determination and one can choose again to consider the problem from a time or frequency point of view. For details see for example Wei (2019) and references therein. A big challenge in multivariate time series is however that already for small values of $m$ the number of parameters to estimate is quite considerable.

Therefore often dimension reduction methods are of interest in a time series context. On the one side is the dynamic factor model which is especially popular for financial data and for details we refer to Hallin et al. (2020). The approach presented here is based on blind source separation (BSS). In BSS it is assumed that the observable $m$-variate time series $\boldsymbol{x}_t$ is a linear mixture of $m$ latent unobservable processes $\boldsymbol{s}_t = (s_1, \ldots, s_m)^\top$. The model can thus be written as

$$\boldsymbol{x}_t = \boldsymbol{\Omega}\boldsymbol{s}_t + \boldsymbol{\mu},$$

where the full rank $m \times m$ matrix $\boldsymbol{\Omega}$ is the mixing matrix and $\boldsymbol{\mu}$ is the $m$-variate location vector. The goal in BSS is to estimate $\boldsymbol{s}_t$ based on $\boldsymbol{x}_t$ alone. Clearly, for that more assumptions are needed. The two basic assumptions are $E(\boldsymbol{s}_t) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{s}_t) = \boldsymbol{I}_m$. However, a third assumption is required and this third assumption distinguishes between different BSS approaches. A review of BSS methods in the context of time series is for example Pan et al. (2021) and in the following only the main principles are introduced.

In the first approach it is assumed that all $m$ latent processes are weakly stationary linear time series which have different autocovariance functions, i.e. $\boldsymbol{C}_\tau(\boldsymbol{s}_t) = \boldsymbol{D}_\tau$ where $\boldsymbol{D}_\tau$ is a diagonal matrix where the diagonal elements depend on $\tau$. This model is called the second order source separation (SOS) model as the separation can be based on second moments alone. For example AMUSE finds an unmixing matrix $\boldsymbol{\Gamma}$ as the matrix which simultaneously diagonalizes the covariance matrix and one autocovariance $\boldsymbol{C}_\tau$ in the sense that

$$\boldsymbol{\Gamma}\,\mathrm{Cov}(\boldsymbol{x}_t)\boldsymbol{\Gamma}^\top = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{\Gamma}\boldsymbol{C}_\tau(\boldsymbol{x}_t)\boldsymbol{\Gamma}^\top = \boldsymbol{\Lambda}_\tau,$$

where $\boldsymbol{\Lambda}_\tau$ is a diagonal matrix depending on $\tau$. Such simultaneous diagonalization of two matrices can be solved via a generalized eigenvector-eigenvalue decomposition. The decomposition is however only unique if the autocovariances of $\boldsymbol{s}_t$ are all distinct at the chosen lag $\tau$ and therefore the performance of AMUSE depends a lot on its choice. To be less dependent on this choice, the SOBI method suggests to jointly diagonalize $K$ autocovariance matrices plus the covariance matrix. This is nowadays considered a better approach and the added computational complexity is then negligible.

However, for example, in the case of GARCH processes there is no information in the second moments and higher order moments are of interest. The third assumption is then generalized in that way that the $m$ latent processes are uncorrelated at all second and fourth (cross)moments, where cross-moments mean computing these moments between observations measured at different time points. In that case methods like vSOBI and gSOBI extend SOBI by taking higher order moments into account in the joint diagonalization process.
The third approach assumes that all $m$ latent components have a constant mean but their variances change over time and the constraint of identity holds only on average for the observed time span $1, \dots, T$. In that case the model is called nonstationary source separation (NSS) and the idea for an unmixing matrix is to divide the observed time series into $K \geq 2$ non-overlapping intervals. The method NSS-JD then jointly diagonalizes the $K$ covariance matrices obtained from the intervals under the constraint that the total covariance matrix must be the identity. Similarly, NSS-JD-TD chooses, in addition to the $K$ covariance matrices, from the intervals additional autocovariance matrices for selected lags computed on the separate intervals to be jointly diagonalized under the same constraint.
Given an unmixing matrix estimate, the components $\hat{s}_t = \hat{\boldsymbol{\Gamma}}(\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})$ are computed. Note that none of the BSS methods can recover actually the order of the components and their signs but this is not relevant for the purpose of BSS which is three fold:

1. The estimated components might be easier to interpret and might give an better impression of the different underlying dynamics. For interpretations in relation to $\boldsymbol{x}_t$ the loadings from the unmixing matrix can be interpreted similarly as the loadings for example in PCA.

2. As the components are assumed uncorrelated, and often even the stronger assumption of independence is made, it is possible to model the $m$ latent components separately from the others. Thus, no multivariate models need to be fitted but one fits $m$ univariate models. This is often considered easier and more flexible.

3. It is often the case that not all $m$ latent components are of interest and some might be simply noise. Thus dimension reduction can be performed by concentrating in the remaining analysis on the interesting components.

One type of data that is by definition multivariate is compositional data. Naturally, compositional data can also be observed over time which is usually denoted then as a compositional time series (CTS). To address the special nature of

compositional observations, usually the compositions are first transformed into a coordinate system that follows a Euclidean representation, and then standard multivariate methods can be applied. The same strategy is usually also used for CTS. Kynclova et al. (2015) discuss vector autoregressive models in the context of CTS while Nordhausen et al. (2021) investigate the use of blind source separation models in that context.

## 4 Functional Time Series

If an almost continuous time series can be divided into natural non-overlapping consecutive intervals one can consider it as a series of ordered curves. Such a series of curves is known then as functional time series. Consider for example data coming from a magnetometer. Such a device measures almost continuously the magnetic field which depends heavily on the rotation of the earth. Hence it seems natural to look at the shape of these values based on the time of the day, i.e. the position of the sun, and therefor one has for each day one curve. Such a curve is then denoted $\mathcal{X}_t(u)$, where $t$ indicates the day and $u$ the time of the day. Usually, it is assumed that the $\mathcal{X}_t$'s are elements of Hilbert spaces and descriptive statistics are the functional mean, the covariance operator, and the autocovariance operator which can often be obtained by pointwise evaluations. While the standard ARMA model has been extended to the functional setting, research mainly focuses on functional autoregressive (FAR) models for which order estimation and forecasting tools are well established. For further details, we refer to Kokoszka & Reimherr (2017) and references therein.

## 5 Software

As time series analysis is a common task in many fields most statistical software packages provide tools to visualize, summarize time series. Similarly estimation and forecasting tools are usually also available. All the methods described here are for example available in R R Core Team (2020) and there is an R TASK VIEW dedicated to time series analysis listing the most important functions and packages, see https://CRAN.R-project.org/view=TimeSeries.

## Summary

In geosciences many measurements are taken at regular intervals and of interest is to understand the dynamics in the underlying processes and to have an idea of what will happen in the future. Depending on the nature of the data there is a large selection of time series methods available. For simplicity usually the modelling starts with stationary linear models. If these models are then not adequately describing the phenomena under consideration, different possibilities exist to address possible flaws like relaxing the requirements of stationarity and or linearity.

## Cross References

Autoregressive Moving Average Process, Stationarity, Stochastic Process, Time Series Analysis, Power Spectral Density, Spectral Analysis, Fourier Transform, Fast Fourier Transform, Compositional Data
*

## Acknowledgements

## References

Brockwell, Peter J. & Richard A. Davis. 2020. *Time series: theory and methods*. 2nd Edition. New York: Springer.

Donner, Reik V. & Susana M. Barbosa. 2008. *Nonlinear time series analysis in the geosciences*. Berlin: Springer Verlag.

Fan, Jianqing & Qiwei Yao. 2003. *Nonlinear time series. nonparametric and parametric methods*. New York: Springer.

Gilge, Hans. 2006. *Univariate time series in geosciences*. Berlin: Springer Verlag.

Hallin, Marc, Marco Lippi, Matteo Barigozzi, Mario Forni & Paolo Zaffaroni. 2020. *Time series in high dimensions*. World Scientific.

Hassler, Uwe. 2019. *Time series analysis with long memory in view*. Hobocken: Wiley.

Kitagawa, Genshiro. 2010. *Introduction to time series modeling*. Boca Raton: CRC Press.

Kokoszka, Piotr & Matthew Reimherr. 2017. *Introduction to functional data analysis*. Boca Raton: CRC Press.

Kynclova, Petra, Peter Filzmoser & Karel Hron. 2015. Modeling compositional time series with vector autoregressive models. *Journal of Forecasting* 34(4). 303–314.

Nordhausen, Klaus, Gregor Fischer & Peter Filzmoser. 2021. Blind source separation for compositional time series. *Mathematical Geosciences*. 905–924.

Pan, Yan, Markus Matilainen, Sara Taskinen & Klaus Nordhausen. 2021. A review of second-order blind identification methods. *WIREs Computational Statistics*. e1550.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Shumway, Robert H. & David S. Stoffer. 2017. *Time series analysis and its applications*. Cham: Springer Verlag.

Wei, William W. S. 2019. *Multivariate time series analysis and applications*. Hobocken: Wiley.