Handbook of Environmental and Ecological Statistics

Alan Gelfand, Montse Fuentes, Jennifer A. Hoeting, Richard L. Smith

Time series methodology

Publication details
Peter F. Craigmile
Published online on: 15 Sep 2017

How to cite :- Peter F. Craigmile. 15 Sep 2017, Time series methodology from: Handbook of Environmental and Ecological Statistics CRC Press
Accessed on: 03 Oct 2020

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
3

Time series methodology

Peter F. Craigmile

Department of Statistics, The Ohio State University, Columbus, Ohio, USA.

CONTENTS

3.1 Introduction ................................................................. 33
3.2 Time series processes ...................................................... 35
3.3 Stationary processes ....................................................... 35
  3.3.1 Filtering preserves stationarity ..................................... 36
  3.3.2 Classes of stationary processes .................................... 36
    3.3.2.1 IID noise and white noise .................................. 37
    3.3.2.2 Linear processes ............................................ 37
    3.3.2.3 Autoregressive moving average processes ............... 37
3.4 Statistical inference for stationary series ............................ 39
  3.4.1 Estimating the process mean ....................................... 39
  3.4.2 Estimating the ACVF and ACF .................................... 40
  3.4.3 Prediction and forecasting ....................................... 41
  3.4.4 Using measures of correlation for ARMA model identification .... 42
  3.4.5 Parameter estimation ............................................ 43
  3.4.6 Model assessment and comparison ................................ 45
  3.4.7 Statistical inference for the Canadian lynx series ........... 46
3.5 Nonstationary time series ............................................... 48
  3.5.1 A classical decomposition for nonstationary processes ....... 48
  3.5.2 Stochastic representations of nonstationarity ................ 49
3.6 Long memory processes .................................................. 50
3.7 Changepoint methods ..................................................... 50
3.8 Discussion and conclusions ............................................. 51
Bibliography ................................................................. 52

3.1 Introduction

In environmental science and ecology we often make statistical inferences based on data collected sequentially in time. To perform accurate inferences we need to build \textit{time series models or processes} to precisely represent the dependencies observed in \textit{time series data}. Key to this modeling exercise is being able to specify different model components that account for all the effects that we observe in the data over time. For example, we may choose to model a time series in terms of the long term changes in the mean (the trend), regular periodic components (the seasonality), and the noise. This is called the classical decomposition (see Section 3.5.1). Typically for most \textit{time series analyses} the noise cannot
be assumed to be independent and identically distributed, and we need to find a model to describe the dependence that we observe in the noise process.

One application of this idea is to regression modeling. When the response variable is a time series, the noise, once we account for covariates, is unlikely to be independent over time. We specify a time series model for the noise, to perform accurate inference for the regression parameters. In other situations, we could also fit a time series model to remove certain features from the data. For example, we may want to account for, and remove seasonality in the data that is not of direct interest (we deseasonalize). We can also detrend to remove long-term mean effects.

An important application of time series analysis is to the prediction of future, past, or missing values. When predicting into the future we call this forecasting; when predicting the past, hindcasting (Hindcasting is useful, for example in paleoclimate when the interest is in studying past climate). We are interested not just in obtaining a point prediction, but in producing measures of prediction uncertainty.

There has also been attention to using time series models as simulators for physical phenomena. For example, time series models can be used as statistical emulators of turbulence [16, 40, 41] or hydrologic systems [32].

In this chapter we introduce useful classes of time series processes and models, and outline their theoretical properties. We take care to differentiate between so-called stationary and nonstationary processes. We then provide commonly used methods of statistical analysis for time series data. First, we introduce a motivating time series that we will explore as we progress through the chapter. [22] studied an ecological time series of the annual number of Canadian lynx trapped around the MacKenzie River between 1821 and 1934 (See [13] for an early review of statistical methods of analysis for the Canadian lynx dataset). Time series plots of the original and log-transformed series (Figure 3.1) indicate that the data have a seasonal pattern with a period of around 10 years, but that the amplitude of the seasonality are not constant in time. The log transformation makes the series more symmetrically-distributed, but the seasonal patterns are still not constant in time. We will demonstrate that these quasi-periodicities can be modeled using time series processes.
3.2 Time series processes

Formally a time series process is a stochastic process, a family of random variables (RVs) indexed in time,
\[ \{X_t : t \in T\}, \]
defined on a probability space. When \( T \subset \mathbb{Z} \) we have a discrete-time time series process; when \( T \subset \mathbb{R} \), a continuous-time time series process. In this chapter we focus on discrete-time processes. (See [10] for a review of continuous-time processes, including an exposition of continuous autoregressive moving average (CARMA) processes.) If \( X_t \) is a univariate RV for all \( t \) then \( \{X_t : t \in T\} \) is a univariate time series process, and when \( X_t \) is a random vector, we have a multivariate time series process. We focus here on univariate processes, although the methods extend straightforwardly to multivariate series [11, 44].

The functions \( \{X_t(\omega) : \omega \in \Omega, t \in T\} \) are known as the realizations or sample paths of the process \( \{X_t\} \). For this chapter we informally call these realizations the time series data or just the time series, denoted \( \{x_t\} \). Typically we want to perform inference upon the time series process on the basis of a finite collection of \( n > 1 \) observations, \( \{x_t : t = 1, \ldots, n\} \).

Mathematically, Kolmogorov’s theorem (e.g., [11], Theorem 1.2.1) can be used to check for existence of a time series process by verifying conditions on the finite dimensional distributions of the RVs over time. More commonly we demonstrate existence by building a time series process from a measurable function of another time series process; for example, in Section 3.3.1 we generate processes by filtering. An interesting subclass are the Gaussian processes: time series processes, for which all finite-dimensional distributions are multivariate normal.

3.3 Stationary processes

Stationary processes are time series processes for which some characteristic of the distribution of the series does not depend on the time points, only on the distance between the points. (A distance in time is often called the time lag or just the lag.) While stationarity may only be an approximation in practice, the assumption that a process is stationary (or the noise in a regression model is stationary, for example) permits a parsimonious representation of dependence. This allows, for example, for efficient estimation and prediction with time series.

We contrast between strictly and weakly stationary processes. For strictly stationary processes the joint distribution of the process at any collection of time points are unaffected by time shifts; formally \( \{X_t : t \in T\} \) is strictly stationary (also called narrow-sense) when \( (X_t, \ldots, X_{t+n}) \) is equal in distribution to \( (X_{t+h}, \ldots, X_{t+n+h}) \), for all integers \( h \) and \( n \geq 1 \), and time points \( \{t_k \in T\} \).

On the other hand, weakly stationary processes are defined only in terms of the mean and covariance of the process. A time series process \( \{X_t : t \in T\} \) is (weakly) stationary if

1. \( E(X_t) = \mu_X \) for some constant \( \mu_X \) which does not depend on \( t \); and
2. \( \text{cov}(X_t, X_{t+h}) = \gamma_X(h) \), a finite constant that can depend on the lag \( h \) but not on \( t \).
Other terms for weakly stationary include second-order, covariance, wide-sense. In the above definition $\gamma_X(\cdot)$ is called the autocovariance function (ACVF) of the stationary process. It can also be useful to work with the unitless autocorrelation function (ACF) of the stationary process,

$$\rho_X(h) = \text{corr}(X_t, X_{t+h}) = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad h \in \mathbb{Z}.$$ 

A strictly stationary process $\{X_t\}$ is also weakly stationary as long as $E(X_t)$ and $\text{var}(X_t)$ are finite for all $t$. Weak stationarity does not imply strict stationarity. However, a Gaussian weakly stationary process is strictly stationary.

Any ACVF of a stationary process has the following properties:

1. The ACVF is even: $\gamma_X(h) = \gamma_X(-h)$ for all $h$;
2. The ACVF is nonnegative definite (nnd): $\sum_{j=1}^{n} \sum_{k=1}^{n} a_j \gamma_X(j-k) a_k \geq 0$ for all positive integers $n$ and real-valued constants $\{a_j : j = 1, \ldots, n\}$.

More formally, Bochner’s theorem [5, 6] states that a real-valued function $\gamma_X(\cdot)$ defined on the integers is the ACVF of a stationary process if and only if it is even and nnd. Typically we do not use this theorem to construct ACVF functions for modeling stationary time series from even and nnd functions. (There are exceptions; e.g., [24] shows that the Cauchy function is even and nnd, and demonstrates the utility of using this ACVF to model long-range dependence – for a definition of long-range dependence, see Section 3.6). Instead we typically create new processes from measurable functions of other processes. The most popular way to create stationary processes is via linear filtering.

### 3.3.1 Filtering preserves stationarity

In the context of discrete-time processes, a linear time invariant (LTI) filter is an operation that takes a time series process $\{W_t : t \in T\}$ and creates a new process $\{X_t\}$ via

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \quad t \in T,$$

where $\{\psi_j : j \in \mathbb{Z}\}$ is a sequence of filter coefficients that do not depend on time and that are absolutely summable: $\sum_j |\psi_j| < \infty$.

Then, if $\{W_t\}$ is a mean zero stationary process with ACVF $\gamma_W(h)$, $\{X_t\}$ is a mean zero stationary process with ACVF

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_W(j-k+h), \quad h \in \mathbb{Z}.$$ 

This is the linear filtering preserves stationarity result. We now use this result to define useful classes of stationary processes.

### 3.3.2 Classes of stationary processes

Here follows some popular classes of stationary processes that are often used for modeling time series. (We also discuss other useful classes of time series processes, some of which are stationary, later in the chapter.)
3.3.2.1 IID noise and white noise

The simplest strictly stationary processes is IID noise: \( \{ X_t : t \in T \} \) is a set of independent and identically distributed RVs. There are no dependencies across time for this process.

The analogous “no correlation” model that is weakly stationary is white noise: \( \{ X_t \} \) in this case satisfies \( E(X_t) = \mu \), a constant independent of time for all \( t \), with

\[
\gamma_X(h) = \begin{cases} 
\sigma^2, & h = 0; \\
0, & \text{otherwise}. 
\end{cases}
\]

We write WN(\( \mu \), \( \sigma^2 \)) for a white noise process with mean \( \mu \) and variance \( \sigma^2 > 0 \).

3.3.2.2 Linear processes

Suppose that \( \{ \psi_j : j \in \mathbb{Z} \} \) is an absolutely summable sequence that is independent of time and \( \{ Z_t : t \in T \} \) is a WN(0, \( \sigma^2 \)) process. Then by the filtering preserves stationary result, the linear process \( \{ X_t : t \in T \} \), defined by a filtering of the white noise process,

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad t \in T,
\]

is a stationary process with zero mean and an ACVF given by

\[
\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}, \quad h \in \mathbb{Z}.
\]

This class of models is useful for theoretical studies of linear stationary processes, but in practice for statistical modeling we use linear filters of finite time support (i.e., \( \psi_j \neq 0 \) for a finite collection of indexes \( j \)) as we now describe.

3.3.2.3 Autoregressive moving average processes

Autoregressive moving average processes, arguably the most famous time series model, are defined in terms of two filtering operations – one for the process \( \{ X_t : t \in T \} \) itself, and one for the white noise \( \{ Z_t : t \in T \} \). These two filtering operations produce a flexible class for modeling linear dependence.

Suppose \( \{ Z_t \} \) is a WN(0, \( \sigma^2 \)) process throughout this section. We say that \( \{ X_t \} \) is an autoregressive moving average process of autoregressive order \( p \) and moving average order \( q \), or an ARMA(\( p, q \)) process for short, if there exists a stationary solution to the following equation:

\[
X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}, \quad t \in T,
\]

assuming that \( \phi_p \neq 0 \) and \( \theta_q \neq 0 \). By “there exists a stationary solution” we mean that we can represent \( \{ X_t \} \) as a (stationary) linear process, as defined by Section 3.3.2.2. The above ARMA process is a mean zero process; more generally \( \{ X_t \} \) is an ARMA(\( p, q \)) process with mean \( \mu \) if

\[
(X_t - \mu) - \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu) = Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}, \quad t \in T.
\]
There are two important simplifications of ARMA models. When \( q = 0 \) we obtain the autoregressive process of order \( p \) (the AR(\( p \)) process; [65, 66])

\[
X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t, \quad t \in T.
\]  
(3.1)

With \( p = 0 \), the moving average process of order \( q \) (the MA(\( q \)) process) satisfies

\[
X_t = Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}, \quad t \in T.
\]  
(3.2)

For ARMA processes the AR coefficients \( \{\phi_j : j = 1, \ldots, p\} \), the MA coefficients \( \{\theta_j : j = 1, \ldots, q\} \), and the innovation variance \( \sigma^2 \) determine the statistical properties of the process \( \{X_t\} \), such as the ACVF. Further, examining the AR coefficients allow us to test for stationarity and causality. (For a causal process \( X_t \) can be written as an MA process of infinite order; i.e., we set \( q = \infty \) in (3.2); an example of an AR processes that is not causal is a linear process for which \( \psi_{-1} \neq 0 \). Assuming causality is useful for prediction and calculating the ACVF.) The MA coefficients determine if we can represent \( X_t \) as an AR process of infinite order (such invertible processes allow us to find unique parameter estimates for ARMA processes). See e.g., [11] and [64] for properties of ARMA processes.

We now provide simple examples of ARMA processes. For any given application there is no a priori reason that these models will fit adequately, but they do illustrate simple models of time series dependence. The MA(1) process \( \{X_t : t \in T\} \), defined by

\[
X_t = Z_t + \theta Z_{t-1}, \quad t \in T,
\]

is a mean zero stationary process for all values of \( \theta \). The ACVF is

\[
\gamma_X(h) = \begin{cases} 
\sigma^2(1 + \theta^2), & h = 0; \\
\sigma^2 \theta, & h = \pm 1; \\
0, & \text{otherwise},
\end{cases}
\]

and the ACF is

\[
\rho_X(h) = \begin{cases} 
1, & h = 0; \\
\theta/(1 + \theta^2), & h = \pm 1; \\
0, & \text{otherwise}.
\end{cases}
\]

The MA(1) process is a model for any stationary time series in which we only have a lag-one correlation (this is also called a 1-correlated process). As this type of dependence is uncommon, in many applications an AR(1) process is more reasonable than an MA(1) process. The AR(1) process, defined by

\[
X_t = \phi X_{t-1} + Z_t, \quad t \in T,
\]  
(3.3)

is stationary, has mean zero, and is causal for \(-1 < \phi < 1\). The ACVF,

\[
\gamma_X(h) = \sigma^2 \frac{|\phi|}{1 - \phi^2}, \quad h \in \mathbb{Z}
\]

and ACF \( \rho_X(h) = |\phi|^h \), both decay exponentially in magnitude to zero. In many applications, we might imagine that we observe an AR(1) process with independent additive measurement error; this can be represented as a special case of an ARMA(1,1) process:

\[
X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}, \quad t \in T.
\]
This process is again stationary, mean zero, and causal for $-1 < \phi < 1$. The ACVF, 
\[
\gamma_X(h) = \begin{cases} 
\sigma^2 \left[ 1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right], & h = 0; \\
\sigma^2 \left[ \theta + \phi + \frac{(\theta + \phi)^2 \phi}{1 - \phi^2} \right], & h = \pm 1; \\
\phi^{\lvert h \rvert - 1} \gamma_X(1), & \text{otherwise},
\end{cases}
\]

exhibits features in common with both the MA(1) and AR(1) processes.

### 3.4 Statistical inference for stationary series

Suppose throughout this section that $\{X_t : t \in T\}$ is a stationary process with mean $\mu_X$ and ACVF $\gamma_X(h)$. Let $\rho_X(h)$ denote the associated ACF. We now outline methods of estimation for the mean, ACVF, and ACF based on time series data $x = (x_1, \ldots, x_n)^T$; let $X = (X_1, \ldots, X_n)^T$ denote the associated random vector. For a stationary process we note that the $n \times n$ covariance matrix of $X$, $\Sigma$, is a symmetric Toeplitz matrix with $(j,k)$ element given by $\Sigma_{jk} = \gamma_X(j - k)$.

#### 3.4.1 Estimating the process mean

The most commonly used estimator of the process mean $\mu_X$ is the sample mean,
\[
\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t.
\]

Except when $\{X_t\}$ is WN or IID noise, this unbiased estimator does not have the smallest variance among all linear unbiased estimators of $\mu_X$ (i.e., it is not the best linear unbiased estimator; BLUE). In its favor, it does not require knowledge of the ACVF to calculate it.

The BLUE estimator of $\mu_X$, $\hat{\mu}_n$, does require knowledge of the ACVF:
\[
\hat{\mu}_n = \left[ I^T \Sigma^{-1} - 1 \right] \left[ I^T \Sigma^{-1} X \right],
\]

where $1$ is a vector of ones.

The sampling distribution of $\overline{X}_n$ and $\hat{\mu}_n$ both depend on the (typically unknown) ACVF. With the sample mean, for example, var$(\overline{X}_n) = v_n/n$, where
\[
v_n = \sum_{h=-n+1}^{n-1} \left( 1 - \frac{|h|}{n} \right) \gamma_X(h).
\]

When $\{X_t\}$ is a stationary Gaussian process,
\[
\frac{\overline{X}_n - \mu_X}{\sqrt{v_n/n}}
\]

follows a standard normal distribution, which forms the basis for statistical inference for $\mu_X$. As for the independent case, there are central limit theorems for stationary process (e.g., [11], Section 7.1; [59], Section 1.3). These theorems allow us to carry out approximate
inference on the process mean for certain non-Gaussian stationary processes. Often a large sample approximation is made for $v_n$.

To calculate $v_n$ we need to estimate the ACVF from the data (see Section 3.4.2 – typically we truncate the number of lags to include in the estimator for $v_n$; see e.g., [12], p.59), or we can propose a time series model for our data and calculate $v_n$ using the ACVF from the estimated model (Section 3.4.5). For this reason we postpone inference for the mean in the lynx example until we have a time series model for the data.

3.4.2 Estimating the ACVF and ACF

Since the ACVF for a stationary process is even and nnd, we look for estimators of the ACVF and ACF that are even and nnd. Thus, to estimate

$$\gamma_X(h) = \text{cov}(X_t, X_{t+h}) = E[(X_t - \mu_X)(X_{t+h} - \mu_X)],$$

we use the sample ACVF, defined by

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \bar{X})(X_{t+|h|} - \bar{X}), \quad |h| < n.$$  

The sample ACF is

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}, \quad |h| < n.$$  

Clearly, given the definition of the sample ACVF and ACF, these estimators are more untrustworthy for larger lags $h$, relative to the sample size $n$ (since there is less information in the estimator). [8] suggests using these estimators to select a time series model only when $n \geq 50$ which is too limiting in practice (they suggest using “experience and past information to derive a preliminary model” when $n < 50$). As for how many lags to interpret, [8] suggests $h \leq n/4$. For univariate series, R ([53]) displays estimates at lags $|h| \leq 10\log_{10}n$, although this does not suggest how many lags to actually interpret. For complicated dependence structures we may need to investigate the ACVF and ACF at longer lags.

Both the sample ACVF and ACF are biased estimators at finite sample sizes $n$. At a fixed lag they become less biased as $n$ increases. Bartlett [2] provides the large sample distribution of the sample ACF for certain stationary processes, which can be useful to test the fit of time series models. In particular Bartlett’s result can be used as a basis for testing whether or not a time series process is IID. When $\{X_t\}$ is an IID process with finite variance, the sample ACFs $\{\hat{\rho}_X(h) : h = 1, 2, \ldots\}$ are approximately IID normal RVs each with mean 0 and variance $1/n$.

As a demonstration, Figure 3.2(a) displays a plot of the sample ACF versus the lag for the log-transformed Canadian lynx series (we discuss panel (b) after we define the PACF below). According to Bartlett’s result, if the process is IID, then approximately 95% of the non-zero lags of the ACF should be within the bands $\pm 1.96/\sqrt{n}$ (denoted by the dashed horizontal lines in the figure). The lynx series is clearly not a sample of IID noise; there is a strong seasonal pattern to the sample ACF, that possibly decays in amplitude at longer lags. The sample ACF at lags $h$ and $h + 10$ are similar, which suggests a periodicity in the log lynx numbers (as claimed by [22]).
Interpreting an ACF plot with pointwise bands leads to multiple testing concerns. This can be alleviated by building a test based on sums of squares of the ACF. The Box-Pierce test [7] tests the null hypothesis that \( \{X_t\} \) is IID (versus not IID). Under the null hypothesis, the test statistic

\[
Q = n \sum_{h=1}^{k} \hat{\rho}^2_X(h)
\]

is approximately chi-squared on \( k \) degrees of freedom. [66] suggested a revised statistic with improved power, although the advice is mixed on how many lags \( k \) to base any IID noise test on. For the log lynx series using 20 lags (the same number as presented in Figure 3.2), the Box-Pierce test statistic is 459.4 on 20 degrees of freedom – the corresponding P-value is very close to zero, which is strong evidence against the time series being generated from an IID process. This hypothesis test is not particularly useful here since a glance of the time series plot for the Canadian lynx series indicates that the data are not IID, and that a time series analysis is necessary. IID noise tests are more commonly used to check the fit of various models for time series – see Section 3.4.6 below. Also consult, e.g., [12] and [64] for other IID and WN tests.

### 3.4.3 Prediction and forecasting

Earlier we pointed out that we are often interested in predicting or forecasting values, along with some measure of uncertainty. (Similar methods apply to predicting missing values or values in the past.) In time series analysis, prediction is also a key ingredient in identifying, fitting, diagnosing, and comparing models.

Again suppose that \( \{X_t : t \in T\} \) is our stationary process with mean \( \mu_X \) and ACVF \( \gamma_X(h) \). Based on \( X = (X_1, \ldots, X_n)^T \) we want to forecast \( X_{n+h} \) at some integer lag in the future, \( h > 0 \). The best linear unbiased predictor (BLUP) of \( X_{n+h} \) based on a linear combination of \( X_1, X_2, \ldots, X_n \) is given by the following conditional expectation:

\[
E(X_{n+h} | X_n, \ldots, X_1) = \mu_X + \sum_{j=1}^{n} \phi_{n,j}(X_{n+1-j} - \mu_X),
\]
where $\phi_{n,j}$ is the solution of the following system of linear equations:

$$
\left\{ \gamma_X(h + 1 - k) = \sum_{j=1}^{n} \phi_{n,j} \gamma_X(k - j) \ : \ k = 1, \ldots, n \right\}.
$$

(3.4)

We often write $P_n X_{n+h}$ as a shorthand for $E(X_{n+h}|X_n, \ldots, X_1)$. In matrix form we have

$$
\phi_n = \Sigma^{-1} \gamma_{n,h},
$$

(3.5)

where $\phi_n = (\phi_{n,1}, \ldots, \phi_{n,n})^T$, $\Sigma$ is the $n \times n$ covariance matrix for $X$ introduced in Section 3.4.1, and $\gamma_{n,h} = (\gamma_X(h), \ldots, \gamma_X(h+n-1))^T$. It can be shown that the prediction coefficients $\phi_n$ only depend on the ACF for the stationary process. In practice when the ACF or ACVF is unknown, we often substitute their estimates into (3.5).

Formally the predictor $P_n X_{n+h}$ minimizes the mean squared prediction error given by

$$
MSPE = E[(X_{n+h} - P_n X_{n+h})^2] = E[U_{n+h}^2],
$$

where $U_{n+h} = X_{n+h} - P_n X_{n+h}$ is called the prediction error for $X_{n+h}$. In practice we solve (3.4) recursively using the Levinson-Durbin algorithm or the Innovation Algorithm (see, e.g., [11] for a full description of both algorithms). The Levinson-Durbin algorithm recursively calculates the coefficients of $P_n X_{n+1}$ given the coefficients of the previous prediction equation for $P_{n-1} X_n$, whereas the Innovations Algorithm calculates recursively the coefficients for calculating the next prediction error $U_{n+1}$ on the basis of coefficients from past prediction errors. The minimum value of the MSPE is also calculated recursively in each algorithm.

This minimum value of the MSPE, given by

$$
\nu_{n,h}^2 = \gamma_X(0) - \phi_n^T \gamma_{n,h},
$$

(3.6)

provides an estimate of the variance for the predictor $P_n X_{n+h}$. For a stationary Gaussian process $\{X_t\}$, a 100(1 - $\alpha$)% prediction interval for $X_{n+h}$ is

$$
P_n X_{n+h} \pm z_{1-\alpha/2} \sqrt{\nu_{n,h}^2},
$$

where $z_q$ is the $q$th quantile of the standard normal distribution. This interval is approximate for non-Gaussian processes.

### 3.4.4 Using measures of correlation for ARMA model identification

A traditional method of identifying plausible time series models is to examine the ACF and the partial ACF (PACF). The ACF looks at the correlation a certain number of lags apart, whereas the PACF looks at the correlation at a certain lag apart, while ignoring the effect of variables at intervening lags.

More formally the PACF for lag $h \geq 1$ is the correlation between the prediction error after predicting $X_{h+1}$ using $X_2, \ldots, X_h$, and the prediction error after predicting $X_1$ using $X_2, \ldots, X_h$ (For $h = 1$ we interpret this to mean simply the correlation between $X_2$ and $X_1$). We calculate the lag $h$ PACF, $\alpha_X(h)$, from the prediction equations (3.5):

$$
\alpha_X(h) = \phi_{h,h}, \quad h = 1, 2, \ldots
$$

To calculate the sample PACF we calculate the prediction coefficients using the sample ACF.
We can now summarize how to identify ARMA processes:

1. The PACF is ideally suited to the identification of AR\((p)\) processes, because we always have that the theoretical PACF is zero at all lags greater than \(p\). Moreover Quenouille [52] proved that the sample PACF are approximately IID normal with zero mean and variance \(1/n\), at all lags greater than \(h\). For AR\((p)\) processes, the ACF decays exponentially in magnitude after lag \(p\).

2. For MA\((q)\) processes, the ACF is zero after lag \(q\), but the PACF decays exponentially in magnitude.

3. For ARMA\((p, q)\) processes neither the ACF and PACF is zero at longer lags; instead, at longer lags, they both decay exponentially in magnitude.

The sample ACF and PACF will be noisy versions of this result. Often there can be several models that are plausible based on these guidelines; we should be cautious not to choose models that are too complicated (i.e., we should not choose \(p\) and \(q\) to be too large – we discuss the issue of overfitting in Section 3.4.6 below).

Returning to Figure 3.2, we examine the sample ACF and PACF plots produced using R. The PACF follows the pattern of an AR\((2)\) process: there is a non-zero lag 2 PACF followed by smaller PACF values after lag 2 – in reference to the Quenouille-based confidence intervals, these longer lags of the PACF may not be significantly different from zero; also, the ACF exponentially decays in magnitude. On the other hand, we could argue that the larger values at longer lags of the PACF require us to consider more involved AR or ARMA process alternatives. The pattern in the ACF rules out an MA process.

### 3.4.5 Parameter estimation

Once we have selected a time series model for our data \(x = (x_1, \ldots, x_n)^T\), we need to estimate the model parameters \(\omega\) (which we will assume is an \(m\)-dimensional vector). When \(\{X_t : t \in T\}\) is stationary, for example, we estimate the mean \(\mu_X\) and the model parameters that determine the ACVF \(\gamma_X(h)\).

There are numerous estimation methods for time series models (see, e.g., [77], [12], and [64]). Here we discuss two commonly used methods: (i) the method of moments Yule-Walker scheme for AR processes; and (ii) the maximum likelihood scheme which applies more generally, but requires us to specify the joint distribution of \(X = (X_1, \ldots, X_n)^T\).

For AR\((p)\) processes as defined by (3.1), we estimate \(\omega = (\phi_1, \ldots, \phi_p, \sigma^2)^T\), where \(\phi_p = (\phi_1, \ldots, \phi_p)^T\) are the \(p\) unknown AR coefficients. The Yule-Walker (Y-W) method sets the sample ACVF equal to the theoretical ACVF, and solves for the unknown parameter values (under the assumption that the AR process is causal). It can be shown that the Y-W estimates for \(\phi_p\) exactly correspond to solving the prediction equations given in Section 3.4.3, but substituting in the sample ACVF for the theoretical ACVF. The estimate for \(\sigma^2\) is the minimum value of the MSPE given by (3.6), substituting in the sample ACVF for the theoretical ACVF. Reusing notation from that section, with hat notation to denote sample-based quantities, the Y-W estimates for an AR\((p)\) model are

\[
\hat{\phi}_p = \Sigma_p^{-1} \hat{\gamma}_{n,p}, \quad \text{with} \quad \hat{\sigma}^2 = \hat{\gamma}_X(0) - \hat{\phi}_p^T \hat{\gamma}_{n,p}.
\]

In practice we use the Levinson-Durbin algorithm to estimate AR models recursively: first we obtain the estimates for the AR\((1)\) model, and then for each \(k = 2, \ldots, p\) we find the AR\((k)\) model estimates in terms of the AR\((k - 1)\) estimates. This recursive estimation method is computationally efficient, and can help speed up our calculations when we try to choose the order \(p\) of the AR model that fits well to the data (See Section 3.4.6 below).
terms of statistical inference for the AR model parameters, under certain conditions the Y-W estimators have the same sampling distribution as the maximum likelihood estimates [11], which we now discuss.

Maximum likelihood for time series models in its most basic form is the same as likelihood estimation for other statistical models. We write down the likelihood $L_n(\omega)$, the joint distribution of the data $x$, evaluated at the model parameters $\omega$. A Gaussian process is the most commonly made assumption for estimation in time series analysis (of course this assumption needs to be verified in practice; in other situations the process may not be Gaussian but the Gaussian likelihood is often used as the basis for an approximate estimation scheme). For a Gaussian mean zero stationary process with ACVF $\gamma_X(h)$, we have that $\mathbf{X}$ follows a $n$-variate normal distribution with zero mean, and covariance $\Sigma$ (remember that the $(j,k)$ element of $\Sigma$ is equal to $\gamma_X(j-k)$, which is now a function of $\omega$). Thus the likelihood function is

$$L_n(\omega) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right).$$

(3.7)

The maximum likelihood (ML) estimate of $\omega$, when it exists, is equal to

$$\hat{\omega}_{ML} = \arg \max_{\omega} L_n(\omega),$$

or equivalently $\hat{\omega}_{ML} = \arg \max_{\omega} l_n(\omega)$, where $l_n(\omega) = \log L_n(\omega)$ is the log-likelihood function.

Calculating the likelihood and log-likelihood functions for a general multivariate normal RV is an $O(n^3)$ calculation because of the matrix inversion. What becomes more useful for stationary processes is when we rewrite the log-likelihood function using one-step-ahead predictions, calculated using a recursive prediction algorithm. This turns the likelihood into a $O(n^2)$ calculation for the most general stationary process – we can speed the likelihood calculation up further for simpler time series models such as AR($p$) or MA($q$). More precisely, letting $U_t = X_t - P_{t-1}X_t$ denote the innovations (the one-step-ahead prediction errors), the log-likelihood is

$$l_n(\omega) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \log(\nu_{t-1,1}^2) - \frac{1}{2} \sum_{t=1}^{n} \left( \frac{\nu_{t-1,1}^2}{\nu_{t-1,1}^2} \right),$$

where $\nu_t$ is the observed innovation and $\nu_{t-1,1}^2$ is the MSPE for $P_{t-1}X_t$ defined by (3.6) at time $t = 1, \ldots, n$.

Typically the ML estimates for time series models are not available in a closed form; we use numerical optimization algorithms to find the ML estimates (e.g., [11, 37, 64]). This is the approach taken by the arima function in R. While there are large sample estimates of the covariance of the ML estimator, it is more common to estimate the covariance using the Hessian matrix. Let $H(\omega)$ denote the matrix of second derivatives of the log-likelihood with respect to the parameters $\omega$. Then statistical inference for the model parameters is based on assuming that $\hat{\omega}_{ML}$ is approximately a $m$-variate normal distribution with mean $\omega$ and covariance matrix $[-H(\omega)]^{-1}$ (e.g., [59]). In practice the true value of the parameter $\omega$ is unknown and we evaluate the Hessian at the ML estimates. This typically results in a loss of statistical efficiency. We should be wary of inferring time series parameters marginally, as usually there exists collinearity among the parameter estimates.

We can also use Bayesian methods of inference for time series models; see e.g., [15, 29, 31, 49].
3.4.6 Model assessment and comparison

Overfitting in time series models leads to highly uncertain and correlated parameter estimates which degrades inference and negatively impacts predictive performance. Thus, it is important to assess the fit of our model, being able to identify when to simplify the model, or to make the dependence structure more complicated when the situation warrants it. In Section 3.4.4 we suggested that we should try not to identify plausible time series models that are too complicated for the dependence that we observe in the data. But, once a model is fit there are a number of ways to assess its fit.

First we can test for significance of the model parameters, cognizant of the warnings above about marginal testing of the model parameters. Based on our inferences, we can drop insignificant parameters from the model, next investigating the fit of the simplified model.

Secondly we can check the fit of our models by examining the time series residuals, \( \{ R_t : t = 1, \ldots, n \} \), defined for each \( t \) by

\[
R_t = \frac{U_t}{\sqrt{\nu^2_{t-1, t}}}
\]

again \( U_t \) is the innovation (one-step-ahead predictor error) and \( \nu^2_{t-1, t} \) is the MSPE for the prediction – these quantities are calculated from the data and the estimated model parameters. Depending on what we assume about the time series model, the properties of \( \{ R_t \} \) can vary. With a Gaussian process assumption, the time series residuals are approximately IID normal with zero mean and constant variance; we often verify assumptions using time series plots, the sample ACF and PACF, and normal Q-Q plots of the time series residuals. We can back up our graphical assessments with IID noise tests, for example.

Thirdly, we can compare measures of model fit. Mimicking what happens in regression models, we should not choose the time series model that minimizes the innovation variance as this often leads to overfitting. Commonly we penalize the fit of the statistical model using a term that is related to number of parameters, \( m \), in the model.

Many measures of model fit are an approximation to the Kullbach-Leibler information, a likelihood-based measure that compares our estimated model to the true, but unknown, model. This includes the Akaike information criterion (AIC) [1],

\[
\text{AIC} = -2l_n(\hat{\omega}) + 2m,
\]

the bias corrected version,

\[
\text{AICC} = -2l_n(\hat{\omega}) + 2m \left[ \frac{n}{n-1-m} \right]
\]

(AIC tends to overestimate \( p \) for AR(\( p \)) models; [33, 56]), and the Bayesian information criterion (BIC), also known as the Schwarz criterion [55],

\[
\text{BIC} = -2l_n(\hat{\omega}) + m\log n.
\]

(Throughout \( \hat{\omega} \) is our estimator of \( \omega \).)

These criteria are often used for order selection, where we compare a number of time series models fit to the data, choosing those models that minimize the chosen criteria. A commonly used rule-of-thumb with AIC, AICC, and BIC is to consider models are equivalent (with respect to the model fit) if the criterion lies within 2 units.

The use of model selection criteria works best in combination with the other assessment methods described above, to fully understand how a given statistical model fits. Another
useful strategy is to compare simulated realizations from the estimated model with the observed data to see if we are capturing all the interesting features of the time series sufficiently.

### 3.4.7 Statistical inference for the Canadian lynx series

We now return to our analysis of the Canadian lynx counts around the MacKenzie river from 1821–1934. We previously claimed that an AR(2) model was reasonable for the log_{10}-transformed data \{x_t : t = 1, \ldots, 114\}. Fitting this model using the Y-W method with the \texttt{ar} function in R, we get $\hat{\mu}_X = 2.90$, $\hat{\phi}_1 = 1.35$, $\hat{\phi}_2 = -0.72$ with $\hat{\sigma}^2 = 0.058$. This is similar to the ML estimates calculated using the \texttt{arima} function: $\hat{\mu}_X = 2.90$ (0.06), $\hat{\phi}_1 = 1.35$ (0.06), $\hat{\phi}_2 = -0.74$ (0.06) and $\hat{\sigma}^2 = 0.051$ (the numbers in parentheses are the standard errors estimated for the ML estimates). Each parameter in the AR(2) model is significantly different from zero, indicating that we should not consider a simplified model. (Figure 3.3 confirms that a 95% confidence region for the AR(2) model parameters does not contain zero for either AR coefficient; see, e.g., [12], p.142, for how to calculate joint confidence regions for model parameters).

We next assess the ML fit of the AR(2) model by examining plots of the time series residuals – see Figure 3.4. We see that the series residuals are centered around zero with fairly constant spread. The normal Q-Q plot indicates that assuming Gaussianity is reasonable. The only issue is that we may not fully capture the time series dependence (the lag 10 ACF is outside the IID noise 95% confidence bands; the lags 10, 12 and 13 are outside the Quenouille–based 95% confidence bands for the PACF). This is confirmed by examining the Box-Pierce test: using 20 lags of the ACF, our test statistic is $Q = 35.0$. With a P-value of 0.056 we fail to reject the IID noise assumption with significance level $\alpha = 0.05$, but we could imagine that there may be dependence still left to model.

We calculate the AICC criterion for various model orders, $p$, of AR($p$) model to see if it is worth making the statistical model more complicated (see Figure 3.5; we omit presenting the AICC of 84.2 for the AR(1) model since this model fits so poorly). The AR(10) and AR(11) models fit best according to the AICC, and time series residual plots are indistinguishable from the AR(2) model.
FIGURE 3.4
From left to right, top to bottom, a time series plot, a normal Q-Q plot, and the sample ACF and PACF of the time series residuals, for the AR(2) model fit.

between these two models. Examining, in this case, pointwise CIs for the AR(11) model parameters, we see that \( \phi_3, \ldots, \phi_9 \) are not significantly different from zero (again, we should be wary of making multiple comparisons). We then fit a subset model with most of the model parameters set to zero to obtain \( \hat{\mu}_X = 2.90 \) (0.07), \( \hat{\phi}_1 = 1.19 \) (0.07), \( \hat{\phi}_2 = -0.47 \) (0.07), \( \hat{\phi}_{10} = 0.39 \) (0.07), \( \hat{\phi}_{11} = -0.39 \) (0.07), \( \hat{\sigma}^2 = 0.04 \), with the lowest AICC found so far of \(-25.2\).

In a similar fashion we could examine ARMA(\( p, q \)) model for the log-lynx numbers. But instead we investigate two interesting questions on the basis of our best fitting statistical model found so far. First, we infer upon the center of the distribution for the lynx series. Using the subset model, a 95\% CI for the process mean for the log series is

\[ 2.90 \pm 1.96 \times 0.07; \]

that is between 1.53 and 3.03. With the assumption of a symmetric distribution for the log lynx counts, the mean is equal to the median of the log series, and transforming back to the original scale, a 95\% CI for the median number of lynx captured on the MacKenzie river from 1821–1934 is between \( 10^{1.53} = 33.8 \) and \( 10^{3.03} = 1071.5 \).

Using the subset model, Figure 3.6 shows pointwise 95\% prediction intervals for the period 1935–1949. (We transform back the endpoints of the prediction intervals on the log\(_{10}\) scale to obtain the prediction limits on the original scale.) These predictions indicate that the subset AR time series model we have fit is able to capture the quasi-periodicity of around 10 years observed in the lynx numbers, and clearly demonstrate higher uncertainty around the peaks, as one would expect.
3.5 Nonstationary time series

3.5.1 A classical decomposition for nonstationary processes

In practice it may be unlikely that a given time series is stationary. There are often components that we observe that are time-varying. One simple representation that separates nonstationary and stationary components observed in a time series \{X_t : t \in T\} is the classical decomposition given by the additive model:

\[ X_t = \mu_t + s_t + \eta_t, \quad t \in T. \]

In the additive model \{\mu_t\} is a trend component that captures (usually) smooth changes in the mean, \{s_t\} is as seasonal or periodic component that captures changes in the mean (over and above the trend) that repeats regularly with some period \(d\) say, and \{\eta_t\} captures the (random or irregular) noise or error. For given data, we may not need a trend and/or seasonal component.
The definition of trend, seasonality, and noise is not unique for any given dataset, which leads to interesting modeling challenges. Regression models and filtering methods are commonly used to estimate the trend and seasonal components in the additive model (e.g., [12]). We often use a three-step strategy. For example: (i) We estimate the trend and seasonality using ordinary least squares regression (this does not account for possible time series dependence in the errors); (ii) Using the regression residuals we find plausible time series models for the error process \( \{ \eta_t \} \); (iii) We refit the regression model while accounting for time series dependence in the errors. (See the climate trends chapter for more information and a further discussion of the lack on uniqueness of the different model components.)

In other scenarios, statistical methods are used to manipulate the series to isolate specific components of the process while trying to eliminate others. For example in estimating monthly fish stocks, inference may be on isolating the deseasonalized count series; that is the original counts minus the estimated seasonal component.

Some methods of removing components such as the trend may alter the statistical properties of the remaining components. For example, differencing is a commonly used to remove a trend: differencing the series \( \{ X_t : t \in T \} \), yields the series \( \{ Y_t \} \) defined by

\[
Y_t = X_t - X_{t-1}, \quad t \in T.
\]

Differencing removes linear and locally linear trends, but can introduce more dependence in the differenced errors. Lag-differencing of order \( d \), defined by \( Y_t = X_t - X_{t-d} \), can be used to remove a seasonality of period \( d \). Both versions of differencing presented here are a form of linear filtering – this fact can be used to derive the statistical properties of differenced time series.

### 3.5.2 Stochastic representations of nonstationarity

Typically in the additive model (3.8) we assume that the trend and seasonality are deterministic functions of time or other covariates. Alternatively we can model the trend (and possibly the seasonality) stochastically. The most basic nonstationary process is the random walk process \( \{ X_t : t = 1, 2, \ldots \} \), defined by

\[
X_t = \sum_{s=1}^{t} Z_s, \quad t = 1, 2, \ldots, \tag{3.9}
\]

where \( \{ Z_t \} \) is a WN(0, \( \sigma^2 \)) process. This process is the discrete approximation to Brownian motion. Sample paths for this series exhibit local trends. This is a coarse model for a system which exhibits increasing variance as time progresses. We note that the differenced process, \( X_t - X_{t-1} \) (which equals \( Z_t \) in this case) is stationary.

Replacing \( \{ Z_t \} \) in (3.9) with a correlated stationary time series process leads to more interesting nonstationary time series models. A classical example is the autoregressive integrated moving average (ARIMA) model [8]. We say that \( \{ X_t \} \) is an ARIMA(\( p, d, q \)) process if it is an ARMA(\( p, q \)) process after the series is differenced \( d \) times. Extending the ARIMA model to allow for random seasonal components we get the seasonal ARIMA model.

Methods of estimation and prediction for ARIMA and seasonal ARIMA are widely available (see e.g., [12] and [64] for a discussion of the methods; software for fitting these models are available in R and SAS).
3.6 Long memory processes

Long memory (LM) or long-range dependent processes are models in which the ACVF decays slowly with increasing lags. Stationary LM processes exhibit significant local deviations. Looking over short periods of time, there is evidence of trends and seasonality in such series, but they disappear from view over longer time periods. [4] and [46] are good general references for LM processes.

A reason for choosing LM models for environmental applications is that they can arise from the slow mixing of many short-range time series (such as the AR(1) process) [25]; there are famous examples in hydrology [32, 44], turbulence [41], thermodynamics [14], and in climate science [84]. For statistical modeling, LM processes may fit better to time series with long range correlations, as compared to using high order AR, MA and ARMA models. Additionally, we can develop LM models that build on the ARIMA model framework given above to simultaneously capture long and short range dependence.

The autoregressive fractionally integrated moving average, ARFIMA($p$, $d$, $q$), process [26, 30] is defined from a simple idea: we replace the nonnegative integer differencing parameter $d$ in the ARIMA process with a real-valued parameter $d$ called the fractionally differencing or LM parameter. An ARFIMA($p$, $d$, $q$) process $\{X_t : t \in T\}$ is an ARMA($p$, $q$) process $\{Y_t\}$ after fractional differencing of order $d$, as defined by the following filtering operation:

$$Y_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1) \Gamma(d)} X_{t-j}, \quad t \in T;$$

(The definition of the filter follows from an application of the binomial theorem.) Any ARFIMA(0, $d$, 0) process is also called a fractionally differenced, FD($d$), process [30]. An ARFIMA process is stationary as long as the ARMA process is stationary and $d < 1/2$.

Notwithstanding that the theory of estimation and prediction is more challenging for LM processes (e.g., [4]), there are also practical problems. Gaussian ML can be used for parameter estimation in ARFIMA processes, but ML is computer intensive, can be numerically instable, and the estimation can be sensitive to the choice of the ARMA model components. [18] provides the theoretical properties of ML estimates (and the Whittle estimator, a large sample approximation to Gaussian ML based on a Fourier transform of the time series). There are many approximate, computationally efficient, methods for estimating the parameters of LM processes; e.g., approximate AR schemes, Fourier methods, and wavelet methods. These methods have less numerical computing issues, and can be robust to assumptions made about the model. For further details about estimation and prediction with LM processes, read [4] and [46]; see [23] for a comparison of Fourier and wavelet methods.

3.7 Changepoint methods

In many environmental and ecological problems interest lies in determining whether or not the statistical properties of a time series process changes at a given time point or time points. The point at which the statistical properties change is called a changepoint.

When a changepoint occurs at time $\tau$, the statistical properties of the time series process before $\tau$, $\{X_t : t < \tau\}$, is different from the statistical properties of the process starting at $\tau$, $\{X_t : t \geq \tau\}$. Extending the definition to multiple changepoints follows naturally. There is
an extensive literature in building test statistics for detecting changes in the distribution of a time series process. For normally distributed observations, [28] derived a likelihood ratio test for detecting changes in the mean and [36] detected changes in the variance. There is also a body of work on building, e.g., segmented regression models for which the properties of the model change at a given set of time points (see e.g., [45]).

For an early review of the theory and methods for time series, see [17]. [35] provides a review of changepoint methodology in environmental science, [3, 54] for climate science, and [60] for identifying so-called thresholds in ecology. [39] provides a review of changepoint methods in the context of introducing the changepoint R package (they also review other R packages for detecting changepoints).

3.8 Discussion and conclusions

Time series analysis is being used to analyze datasets in many areas, such as environmental science and ecology. One chapter cannot do the entire discipline justice.

This chapter focused heavily on measures of dependence that are defined over time, typically using lags of the ACVF or ACF of a stationary process. Even for the nonstationary processes of Section 3.5, the nonstationarity is driven by some usually hidden (“latent”) stationary process. In the search for more interesting time series processes, the use of latent time series models have become popular. Hierarchical statistical time series models, such as state space models (e.g., [17, 38, 49]) or dynamic linear models (e.g., [45]), combine hierarchies of time series models to create more complicated structures. The different hierarchies can have practical interpretation (e.g., the first level of the hierarchy could indicate how we observe some latent process with measurement error, and the second level could be a model for how this latent process evolves over time). Another reason for their popularity is the increasing availability of accessible and computationally efficient methods to fit hierarchical models. More complicated time series models, such as nonlinear processes (e.g., [61, 97]) and non-Gaussian processes (e.g., [17, 20]), are often are created from latent variable models. (See [62] for an analysis of the Canadian lynx series using a class of nonlinear processes called threshold models; also read the dynamic models chapter of this book for further details on latent variable time series models.)

We close by noting that dependence need not be defined in terms of the unit of time. There is a rich history of taking Fourier transforms of time series, re-interpreting dependence in terms of the contributions of random sinusoids of different frequencies. The leads to the spectral analysis of time series; see e.g., [9, 50, 51, 77]). More recently other transformations have been developed to model dependence. For example, wavelet analysis [64, 78] examines dependence both over time and scale (approximately frequency). Modeling processes that vary over time drive us to investigate how we can efficiently model (both statistically and computationally) time-varying time series dependence. An interesting compromise assumes the process is locally stationary, with a dependence structure that slowly evolves over time; see e.g., [19].

Acknowledgments

Craigmile is supported in part by the US National Science Foundation under grants DMS-1407604 and SES-1424481. My thanks to Peter Guttorp, Donald Percival, Alexandra Schmidt, and Marian Scott for providing comments that greatly improved this chapter.
Bibliography


