4-23-2013 12:00 AM

# Seasonal Decomposition for Geographical Time Series using Nonparametric Regression

Hyukjun Gweon, *The University of Western Ontario*

Supervisor: Dr. A. Ian McLeod, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

# Seasonal Decomposition for Geographical Time Series using Nonparametric Regression

(Thesis format: Monograph)

by

**Hyukjun Gweon**

Graduate Program
in
Statistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

# ABSTRACT

A time series often contains various systematic effects such as trends and seasonality. These different components can be determined and separated by decomposition methods. In this thesis, we discuss time series decomposition process using nonparametric regression. A method based on both loess and harmonic regression is suggested and an optimal model selection method is discussed. We then compare the process with seasonal-trend decomposition by loess $STL$ (Cleveland, 1979). While $STL$ works well when that proper parameters are used, the method we introduce is also competitive: it makes parameter choice more automatic and less complex.

The decomposition process often requires that time series be evenly spaced; any missing value is therefore necessarily estimated. For time series with seasonality, it is preferable to use the seasonal information to estimate missing observations. The seasonal adjustment algorithm (McLeod et al., 1983) can be used for monthly time series. In this thesis, we examine the algorithm and revise it to cover daily data cases.

KEY WORDS: Time series decomposition, Loess, Harmonic regression, Model selection, Missing data estimation.

# ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Dr. A. Ian McLeod, for all the support and persistent guidance throughout my graduate studies at Western University. His excellent advice and encouragement have been truly indispensable to my research work as well as writing of this thesis. I am also grateful to the faculty members at the Department of Statistical and Actuarial Sciences for providing me with a valuable opportunity and assistance for the research. My sincere thanks also go to my thesis committee, Dr. Matt Davison, Dr. Hao Yu and Dr. Tim Conley, for their insightful comments and valuable discussions.

Last but not least, I would like to thank my parents, brothers and friends. Their love and encouragements have enabled me to complete this thesis.

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1

# Introduction

When one analyzes a time series, it often shows different systematic patterns. These patterns include trends and seasonality and are the results of various causes. Generally, a trend component represents the long-term effects of a time series, while a seasonal component logically explains effects that results from the changing seasons in a year. Regarding a time series as a combined output of different components, we can build a statistical model for each component and decompose the series into different parts. In this thesis, we introduce time series decomposition methods using nonparametric estimations where a proper model form is determined by the model selection process. Decomposition of time series into long term trends, cyclical components including stochastic cyclical components as with business cycles or deterministic cyclical components due to seasonality, has a long history in economic time series. Klein (1997, Ch. 9) provides a review of some of these time series decompositions methods. The focus in this thesis is on environmental time series. Useful research monographs on existing methods for such time series include Mudelsee (2010); Chandler and Scott (2011). The objective of this thesis is to extend the seasonal decomposition methods discussed by Hipel and McLeod (1994); McLeod et al. (1983) to the case of daily time series and allowing for irregular spacing and missing values.

In Chapter Two, we introduce locally weighted scatterplot smoothing or loess (Cleveland, 1979) and harmonic regression, including their features. Loess, a widely used nonparametric regression method, has sufficient flexibility to be useful to fit a

model for either a long-term or a seasonal effect. However, for a seasonal pattern that has a one-year period, we can also use harmonic regression where the predictors consist of sinusoids. Using sinusoids as predictors highlights several important characteristics of harmonic regression, including aliasing, orthogonality and Fourier frequency. We introduce these features because they are essential for building a proper model.

In nonparametric model estimation, we compare various model structures and use one that balances both conformity to data and the parsimony principle. This approach can be formulated by analogy with mathematical programming, viz. the model with the fewest parameters that adequately fits the data is preferred (McLeod, 1993). We can achieve the work based on model selection criteria, such as AIC and BIC. In Chapter Three, we present Akaike information criterion AIC (Akaike, 1974) and Bayesian information criterion BIC (Schwarz, 1978). Using these criteria, we can determine suitable models for fitting time series components.

In Chapter Four, we consider time series where a seasonal pattern exists. Using graphical methods such as a seasonal subseries plot (Cleveland, 1993) and multiple boxplots, we visually confirm the existence of a seasonality. Then loess or harmonic regression can be employed to model the component. If the variations are seasonal, scaling is preferred for variance correction.

Chapter Five presents the time series decomposition assuming an additive or multiplicative model. Since a multiplicative model can be transformed into an additive model by taking logarithms, we can only consider additive model situations. For decomposition, we introduce the seasonal-trend decomposition process by loess STL (Cleveland et al., 1990), which uses loess to fit any component present in a time series. We then present another method, where loess and harmonic regression are used for long-term and seasonal components, respectively. By the backfitting algorithm (Friedman and Stuetzle, 1981), the components are separated without having effects

from any other components.

When we build models for seasonal patterns in a time series using loess or harmonic regression, we are often required to accommodate a condition for the time series to be evenly spaced. In reality, however, we commonly encounter situations where some observations are missing. In Chapter Six, we specify the missing data problem in a seasonal time series. For an unevenly spaced time series where a seasonal pattern is present, we can measure the seasonal effects that repeat and use them to estimate of missing observations. As a measuring tool, we consequently introduce the seasonal adjustment algorithm (McLeod et al., 1983). Using this algorithm, missing values in a seasonal time series can be estimated, which makes an unevenly spaced time series an evenly spaced one. The algorithm can be used not only on monthly data but also on daily time series.

# Chapter 2

# Smoothing Methods

## 2.1 Introduction

Various regression methods are used for analyzing time series. Our decision to employ a particular method generally depends on our assumptions and the nature of the data. One of the popular methods is linear regression because it is simple to use and easy to model. However, it is not an appropriate choice if the underlying pattern of data is not linear. An underlying critical assumption is also that the parametric form of the true model is known. These features are often considered unrealistic and a better strategy may be to let the data choose a proper model. In other words we may prefer nonparametric models such as loess (Cleveland, 1979) and spline regression. In particular, if a seasonal component is present in a time series, loess or harmonic regression may be useful for estimating the component. Loess has sufficient flexibility that it may provide a good fit for cyclical fluctuation of the series. On the other hand, because a seasonal pattern repeats with a one year period, harmonic regression, where the predictors consist of sinusoids may also be employed.

In Section 2.1, we start with the local regression principle and introduce the robust loess method. Although this is a nonparametric regression method, both smoothing and the polynomial degree parameters must be given.

In Section 2.2, we present harmonic regression and its properties. We consider a few fundamental features of harmonic regression such as aliasing, orthogonality and Fourier frequency (Bloomfield, 1976). These properties are essential for building

proper harmonic models and for deriving simple formulas for least square estimates of parameters.

## 2.2 Loess

### 2.2.1 The Model

Loess (Cleveland, 1979), also known as local regression, is a nonparametric smoothing method of a dependent variable $y$ given covariates $x$. Although the number of observation is finite, the smoothing curve can be defined everywhere as well as at the observed points $x_i$, $i$=0,1,...,$n$, where $n$ is the sample size.

We assume a model of the form

$$y = g(x) + \epsilon$$

where $g$ is an unknown regression function and $\epsilon$ is an error term. The errors are IID with mean 0 and a constant variance.

In order to fit a loess model, we must choose the smoothing parameter (also known as bandwidth parameter) and polynomial degree, denoted as $\alpha$ and $p$, respectively. Usually, $p$ has a value of 1 or 2, where $p = 1$ corresponds to local linear fits and $p = 2$ to local quadratic polynomials. $\alpha$ indicates the proportion of the data used for local regression and thus should be a positive value between 0 and 1. $\alpha$ is called the smoothing parameter because it controls the smoothness of a regression fit. For example, if $\alpha = 0.2$, only twenty percent of the data is used for local regression and, in effect, the resulting model reflects drastic variations and is less smooth. On the other hand, if $\alpha = 0.7$, seventy percent of the data is used for a local fit, giving a much smoother model. That is, a large $\alpha$ makes the model less sensitive to data variations.

The first step for fitting a loess model at any point $x$ is to select $n\alpha$ number of observations whose explanatory variable outcomes are closest to $x$. If $n\alpha$ is not an integer, it would be rounded to the next largest integer. Let $q$ be $n\alpha$ truncated to be an integer and $x_{(k)}$ be the $k$th closest point ($k$=1,2,...,$q$).

Before we fit a local polynomial with degree $p$ using the subset observations, we must give a weight to each subset point based on its horizontal distance to $x$. Generally, we use the tricube weight function $W(x)$ written as

$$W(x) = \begin{cases} (1-|x|^3)^3 & |x|\leq 1 \\ 0 & otherwise. \end{cases}$$

Then, for any $x$, the weight given to an observed point ($x_i$, $y_i$) is determined as $W(\frac{x_i - x}{x_{(q)} - x})$. Note that the weight at $x_{(k)}$ $(k = 1, 2, ..., q)$ is maximized when $k$=1 and decreases as $k$ increases. In other words, according to the weight function, points have large weights if their explanatory variable value outcomes are near to $x$, while small weights are given to observations that are far from $x$. Zero weights are assigned to observations whose horizontal distances are equal to or further than the distance between $x$ and $x_q$. Using the weighted subset data, we build a local fit at any $x$. The value on the polynomial is the loess fit.

At any point $x$, a local polynomial with degree $p$ is fitted by minimizing the weighted sum of squares

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - ... - \beta_p x_i^p)^2 W(\frac{x_i - x}{x_{(q)} - x})$$

with respect to $\beta_0, \beta_1, ..., \beta_p$. The loess fit at $x$ is then

$$\hat{g}(x) = \hat{\beta_0} + \hat{\beta_1} x + ... + \hat{\beta_p} x^p.$$

### 2.2.2 Robust Loess Method

If there appears to be outliers in the data, it may be preferable to use a robust version of loess to reduce model sensitivity to the outliers. The robust loess method (Cleveland, 1979) starts with calculating the residuals from a loess model. Let $e_i$ be the residual at $x_i$. That is,

$$e_i = y_i - \hat{g}(x_i)$$

where $\hat{g}(x_i)$ is the loess fit at $x_i$. Using the bisquare weight function which is defined as

$$B(x) = \begin{cases} (1-|x|^2)^2, & |x| \leq 1 \\ 0 & otherwise \end{cases},$$

the robustness weight $\delta_i$ is determined by

$$\delta_i = B(\frac{e_i}{6M})$$

where $M$ is the median of absolute values of the residuals.

With the robustness weights, the local regression model is revised by minimizing

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - ... - \beta_p x_i^p)^2 \delta_i W(\frac{x_i - x}{x_{(q)} - x}).$$

Then new loess fits are obtained and so are new residuals. We repeat the procedure several times until it results in the final robust estimate.

## 2.3 Harmonic Regression

### 2.3.1 The Model

The harmonic regression uses harmonics as its explanatory variables. The model is especially useful when analyzing a time series where the seasonality exists. For a periodic time series $z_t$, a simple model with a single sinusoid may be written as

$$z_t = \mu + R\cos 2\pi(ft + \phi) + \epsilon_t \qquad (1)$$

where $\mu$, $R$, $f$, $\phi$ and $\epsilon_t$ are the constant, amplitude, frequency, phase and white noise, respectively. Since frequency is known and is a fixed number, the model has three parameters - the constant, amplitude and phase. The least square method may be used for estimating the parameters. Computation is complicated, however, in this equation form, so instead the following version is preferable.

$$\begin{aligned} z_t &= \mu + R\cos 2\pi(ft + \phi) + \epsilon_t \\ &= \mu + R(\cos 2\pi\phi \cos 2\pi ft - \sin 2\pi\phi \sin 2\pi ft) + \epsilon_t \\ &= \mu + R\cos 2\pi\phi \cos 2\pi ft - R\sin 2\pi\phi \sin 2\pi ft + \epsilon_t \\ &= \mu + A\cos 2\pi ft + B\sin 2\pi ft + \epsilon_t \end{aligned}$$

where $A$ is $R\cos 2\pi\phi$ and $B$ is $-R\sin 2\pi\phi$. Now $\mu$, $A$ and $B$ are model parameters and estimating these parameters is equivalent to estimating the parameters in (1).

A time series often shows a complicated periodic pattern. In this case, building a proper model is barely possible using a single sinusoid because the residuals of the model would still exhibit parts of seasonality. This problem may be solved by adding other harmonic terms with different frequencies. That is, the modeling strategy is to

keep adding new sinusoids until the residuals no longer show periodic behaviour. For instance, a harmonic model with two sinusoids may be described as

$$\begin{aligned} z_t &= \mu + R_1 \cos 2\pi(f_1 t + \phi_1) + R_2 \cos 2\pi(f_2 t + \phi_2) + \epsilon_t \\ &= \mu + A_1 \cos 2\pi f_1 t + B_1 \sin 2\pi f_1 t + A_2 \cos 2\pi f_2 t + B_2 \sin 2\pi f_2 t + \epsilon_t \end{aligned}$$

### 2.3.2 Aliasing

The use of harmonic regression requires that the frequency $f$ be within the range $0 < f \leq \frac{1}{2\Delta}$, where $\Delta$ is the sampling interval. If $1/2\Delta < f < 1/\Delta$, there exists an 'alias', denoted as $\lambda$, such that $\lambda = 1/\Delta - f$ and thus $0 < \lambda \leq 1/2\Delta$. Then,

$$\begin{aligned} z_t &= \mu + A \cos 2\pi f t + B \sin 2\pi f t + \epsilon_t \\ &= \mu + A \cos 2\pi(1/\Delta - \lambda)t + B \sin 2\pi(1/\Delta - \lambda)t + \epsilon_t \\ &= \mu + A \cos\left(2\pi t/\Delta - 2\pi\lambda t\right) + B \sin\left(2\pi t/\Delta - 2\pi\lambda t\right) + \epsilon_t \\ &= \mu + A \cos 2\pi\lambda t - B \sin 2\pi\lambda t + \epsilon_t \end{aligned}$$

Thus, $f$ and $\lambda$ are indistinguishable in a harmonic model. As a solution of this indeterminacy, the range of $f$ is restricted to be between 0 and $1/2\Delta$. In this paper, we assume that $\Delta = 1$.

### 2.3.3 Fourier frequency

A frequency $f$ is a Fourier frequency if the corresponding sinusoid executes an integer number of cycles in the span of the series. In other words, a Fourier frequency $f_j$ is

defined as

$$f_j = \frac{j}{n} \quad j = 1, 2, ..., \frac{n}{2}$$

where $n$ is the length of the time series. For instance, a sinusoid with frequency $f_3 = 3/n$ completes exact 3 cycles in the span of the series. Note that $f_{\frac{n}{2}} = \frac{1}{2}$ and this is the highest Fourier frequency due to the aliasing property.

Fourier frequency is crucial to hold the orthogonality between predictor variables, as will be introduced in the next section.

### 2.3.4 Orthogonality

Our focus is to evaluate the orthogonality between predictor variables (i.e. harmonic terms) at Fourier frequencies. In order to do so, we use the Euler relation

$$e^{ix} = \cos x + i \sin x$$

and its inverse forms

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}) \qquad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix}).$$

First,

$$\sum_{t=1}^{n} \cos 2\pi f t = \frac{1}{2} \sum_{t=1}^{n} (e^{2\pi i f t} + e^{-2\pi i f t}).$$

$$\begin{aligned} \sum_{t=1}^{n} e^{2\pi i f t} &= \frac{e^{2\pi i f}(e^{2\pi i f n} - 1)}{e^{2\pi i f} - 1} \\ &= \frac{e^{2\pi i f} e^{\pi i f n}(e^{\pi i f n} - e^{-\pi i f n})}{e^{\pi i f}(e^{\pi i f} - e^{-\pi i f})} \end{aligned}$$

$$
\begin{aligned}
&= e^{\pi i f(n+1)}\left(\frac{e^{\pi i f n} - e^{-\pi i f n}}{e^{\pi i f} - e^{-\pi i f}}\right) \\
&= (\cos \pi f(n+1) + i \sin \pi f(n+1)) \frac{2i \sin \pi f n}{2i \sin \pi f} \\
&= (\cos \pi f(n+1) + i \sin \pi f(n+1)) \frac{\sin \pi f n}{\sin \pi f}
\end{aligned}
$$

and similarly,

$$
\sum_{t=1}^{n} e^{-2\pi i f t} = (\cos \pi f(n+1) - i \sin \pi f(n+1)) \frac{\sin \pi f n}{\sin \pi f}.
$$

$$
\begin{aligned}
\sum_{t=1}^{n} \cos 2\pi f t &= \frac{1}{2} \sum_{t=1}^{n} (e^{2\pi i f t} + e^{-2\pi i f t}) \\
&= \frac{1}{2}((\cos \pi f(n+1) + i \sin \pi f(n+1)) \frac{\sin \pi f n}{\sin \pi f} \\
&+ (\cos \pi f(n+1) - i \sin \pi f(n+1)) \frac{\sin \pi f n}{\sin \pi f}) \\
&= \cos \pi f(n+1) \frac{\sin \pi f n}{\sin \pi f}
\end{aligned}
$$

and similarly,

$$
\sum_{t=1}^{n} \sin 2\pi f t = \sin \pi f(n+1) \frac{\sin \pi f n}{\sin \pi f}.
$$

If $f$ is a Fourier frequency, $fn$ becomes an integer and thus $\sin \pi f n = 0$, which results in $\sum_{t=1}^{n} \cos 2\pi f t = \sum_{t=1}^{n} \sin 2\pi f t = 0$. In other words, any cosine or sine term at a Fourier frequency is orthogonal to the constant term. The relations below also demonstrate that any cosine or sine term at a Fourier frequency is orthogonal to another cosine or sine term at another Fourier frequency.

Let $f_k$ and $f_j$ be Fourier frequencies and $f_k > f_j$. Then both $f_k + f_j$ and $f_k - f_j$ are also Fourier frequencies and thus

$$\begin{aligned}\sum_{t=1}^{n} \cos(2\pi f_k t)\sin(2\pi f_j t) &= \sum_{t=1}^{n} \frac{1}{2}(\sin 2\pi(f_k + f_j)t + \sin 2\pi(f_k - f_j)t) \\ &= \frac{1}{2}\sum_{t=1}^{n} \sin 2\pi(f_k + f_j)t + \frac{1}{2}\sum_{t=1}^{n} \sin 2\pi(f_k - f_j)t \\ &= 0 + 0 = 0\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{n} \cos(2\pi f_k t)\cos(2\pi f_j t) &= \sum_{t=1}^{n} \frac{1}{2}(\cos 2\pi(f_k + f_j)t + \cos 2\pi(f_k - f_j)t) \\ &= 0\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{n} \sin(2\pi f_k t)\sin(2\pi f_j t) &= -\sum_{t=1}^{n} \frac{1}{2}(\cos 2\pi(f_k + f_j)t - \cos 2\pi(f_k - f_j)t) \\ &= 0\end{aligned}$$

### 2.3.5 Least Square Estimates with orthogonality

Based on the relations introduced in the previous section, we can easily obtain least square estimators of the model parameters. Consider a general regression model below. We use cosines and sines with all possible Fourier frequencies as predictors.

At this moment we assume that $n$ is even.

$$z_t = \mu + \sum_{0<j<\frac{n}{2}} (A_j \cos 2\pi f_j t + B_j \sin 2\pi f_j t) + A_{\frac{n}{2}} \cos 2\pi f_{\frac{n}{2}} t + \epsilon_t$$

Note that $j = \frac{n}{2}$ has no sine term, since $\sin \pi t = 0$, $\forall t = 1, 2, ..., n$. Let $X$ be the design matrix; then $X$ is a ($n$ x $n$) matrix. That is,

$$X = \begin{bmatrix} 1 & \cos 2\pi f_1 t_1 & \sin 2\pi f_1 t_1 & \cos 2\pi f_2 t_1 & \cdots & \cos 2\pi f_{\frac{n}{2}} t_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos 2\pi f_1 t_n & \sin 2\pi f_1 t_n & \cos 2\pi f_2 t_n & \cdots & \cos 2\pi f_{\frac{n}{2}} t_n \end{bmatrix}$$

By the orthogonal property,

$$\begin{aligned} \sum_{t=1}^{n} \cos^2 2\pi f_k t &= \sum_{t=1}^{n} \frac{1}{2}(\cos 4\pi f_k t + 1) \\ &= \begin{cases} n & k = 0, \frac{n}{2} \\ \frac{n}{2} & otherwise \end{cases} \end{aligned}$$

$$\sum_{t=1}^{n} \sin^2 2\pi f_k t = \begin{cases} 0 & k = 0, \frac{n}{2} \\ \frac{n}{2} & otherwise \end{cases}$$

Then the least square estimates of $\mu, A_1, B_1, ..., A_{\frac{n}{2}}$ are obtained as follows:

$$X^T X = \begin{bmatrix} \cdots & 1 & \cdots \\ \cdots & \cos 2\pi f_1 t & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & \cos 2\pi f_{\frac{n}{2}} t & \cdots \end{bmatrix} \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & \cos 2\pi f_1 t & \cdots & \cos 2\pi f_{\frac{n}{2}} t \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$= \begin{bmatrix} n & 0 & \cdots & \cdots & 0 \\ 0 & \frac{n}{2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{n}{2} & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{B}_1 \\ \vdots \\ \hat{A}_{\frac{n}{2}} \end{bmatrix} = (X^T X)^{-1} X^T z_t$$

$$= \begin{bmatrix} \frac{1}{n} & 0 & \cdots & \cdots & 0 \\ 0 & \frac{2}{n} & 0 & \cdots & 0 \\ \vdots & & \frac{2}{n} & \vdots & \vdots \\ & & & \ddots & \\ 0 & \cdots & \cdots & \cdots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} \cdots & 1 & \cdots \\ \cdots & \cos 2\pi f_1 t & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & \cos 2\pi f_{\frac{n}{2}} t & \cdots \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} z_t \\ \frac{2}{n} \sum_{t=1}^{n} z_t \cos 2\pi f_1 t \\ \frac{2}{n} \sum_{t=1}^{n} z_t \sin 2\pi f_1 t \\ \vdots \\ \frac{1}{n} \sum_{t=1}^{n} z_t (-1)^t \end{bmatrix}$$

When $n$ is odd, the design matrix $X$ is also a ($n$ x $n$) matrix. In this case, the model becomes

$$z_t = \mu + \sum_{0<j<\frac{n}{2}} (A_j \cos 2\pi f_j t + B_j \sin 2\pi f_j t) + \epsilon_t$$

and

$$X = \begin{bmatrix} 1 & \cos 2\pi f_1 t_1 & \cdots & \cos 2\pi f_{\frac{n-1}{2}} t_1 & \sin 2\pi f_{\frac{n-1}{2}} t_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos 2\pi f_1 t_n & \cdots & \cos 2\pi f_{\frac{n-1}{2}} t_n & \sin 2\pi f_{\frac{n-1}{2}} t_n \end{bmatrix}.$$

Thus, the least square estimates are computed by

$$\begin{bmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{B}_1 \\ \vdots \\ \hat{A}_{\frac{n-1}{2}} \\ \hat{B}_{\frac{n-1}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} z_t \\ \frac{2}{n}\sum_{t=1}^{n} z_t \cos 2\pi f_1 t \\ \frac{2}{n}\sum_{t=1}^{n} z_t \sin 2\pi f_1 t \\ \vdots \\ \frac{2}{n}\sum_{t=1}^{n} z_t \cos 2\pi f_{\frac{n-1}{2}} t \\ \frac{2}{n}\sum_{t=1}^{n} z_t \sin 2\pi f_{\frac{n-1}{2}} t \end{bmatrix}$$

The results demonstrate that, for a given time series, the coefficient of each cosine and sine term is determined only by its frequency. It means that, in a harmonic regression model, neither adding nor omitting predictors affects least square estimates provided that the Fourier frequency holds for all sinusoid terms.

## 2.4 Illustrative Examples

### 2.4.1 Loess Modelling with Lake Huron Data

The built-in R data LakeHuron consists of the annual water levels, in feet, of Lake Huron. The data were measured from 1875 to 1972. The original time series is shown in figure 2.1. Loess models are fitted with different combinations of smoothing parameters and polynomial degrees. Figure 2.2 contains 4 subplots with parameters

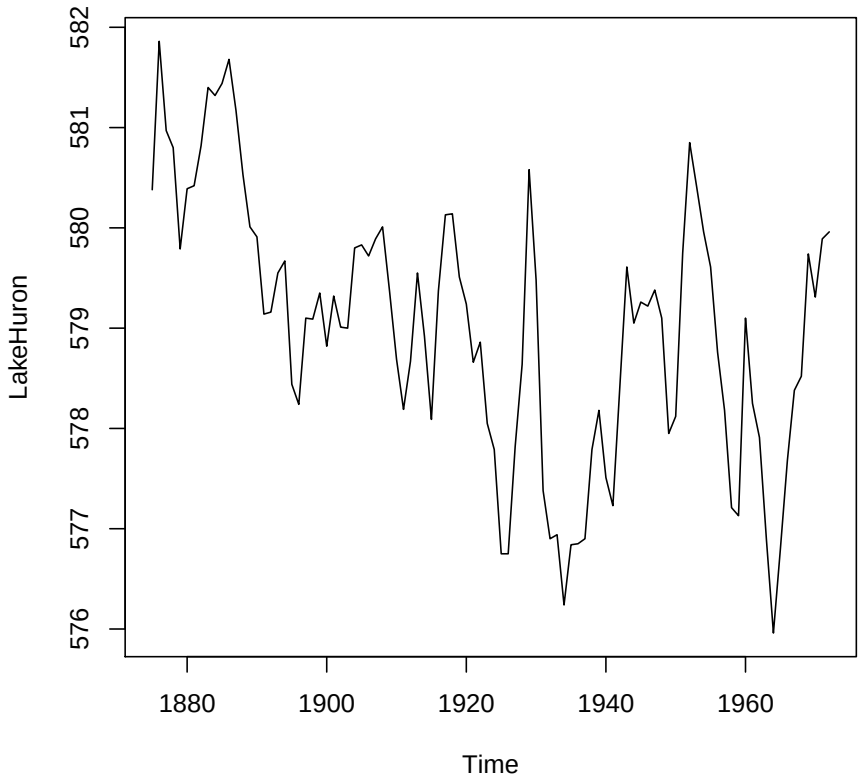


Figure 2.1: Lake Huron Data during 1875 - 1972

($\alpha$, $p$) = (0.1, 1), (0.3, 1), (0.5, 1) and (0.7, 1). Four other subplots with ($\alpha$, $p$) = (0.1, 2), (0.3, 2), (0.5, 2) and (0.7, 2) are drawn in figure 2.3

When $\alpha$ is very small (e.g. 0.1 in our exercise), the resulting model overfits the data, while the model becomes too smooth with an $\alpha$ close to 1 (e.g. 0.7 in our exercise). Among the subplots, we may conclude that the model with ($\alpha$, $p$) = (0.3, 1) and (0.5, 2) fit reasonably well. Note that this conclusion is fairly subjective and a better choice may exist. One of the methods for selecting the best smoothing parameter is cross-validation (CV). We introduce cross validation in later chapters.

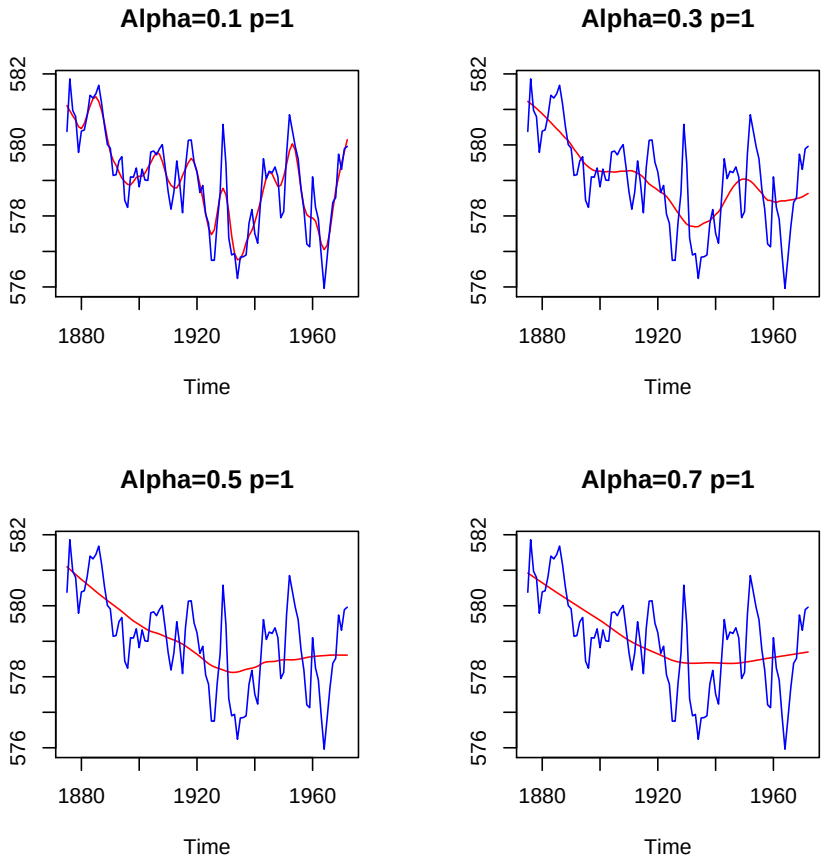


Figure 2.2: Loess models with $p = 1$ and different smoothing parameters

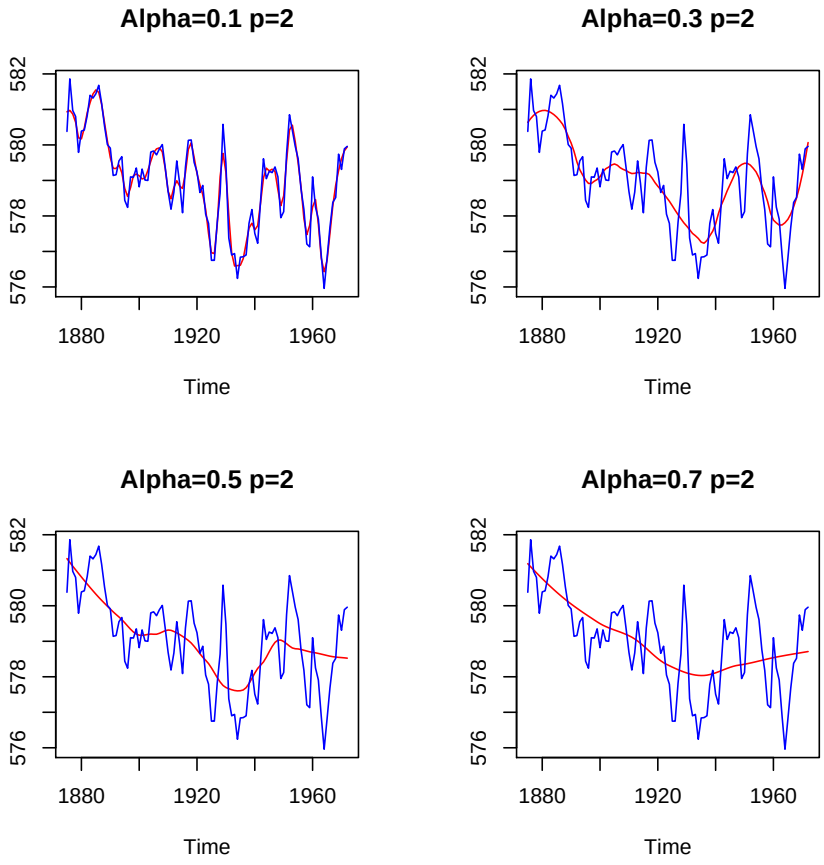


Figure 2.3: Loess models with $p = 2$ and different smoothing parameters

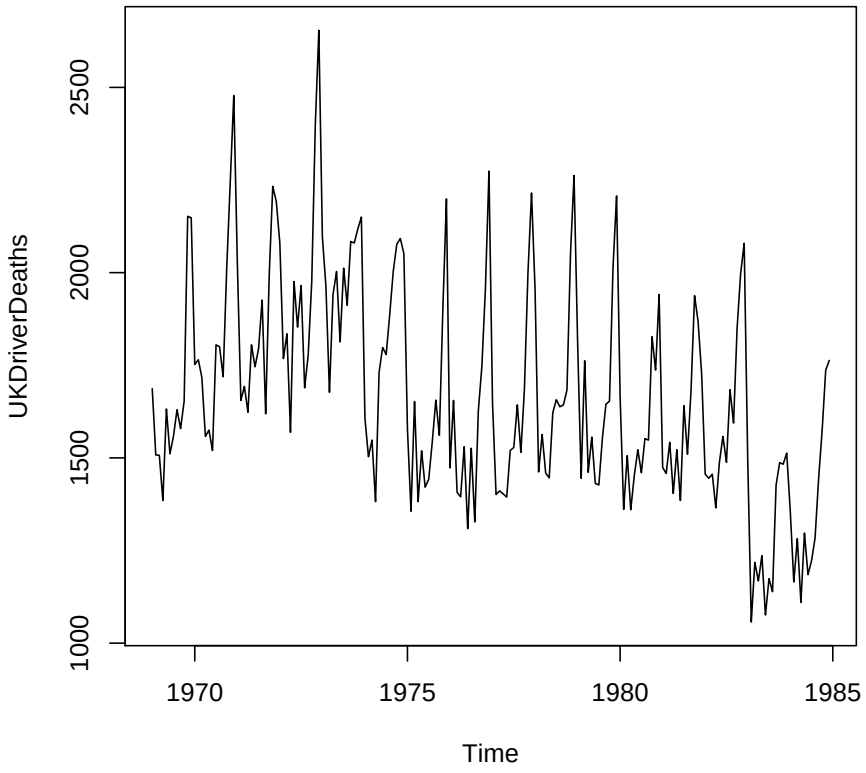


Figure 2.4: UKDriverDeaths

### 2.4.2 UKDriverDeaths

We consider the monthly time series UKDriverDeaths. The data contain the monthly totals from January 1969 to December 1984 for people in the United Kingdom who were killed or seriously injured while driving.

As seen in figure 2.4, the original series exhibits seasonal patterns; the number of deaths or serious injures by car accidents in winter is huge compared to the number of accidents occurred during summer. We employ harmonic regression to fit the seasonal data.

As the first step, we try to fit a harmonic model using a single sinusoid. The fitted model and the residuals are shown in figure 2.5. Since only one sinusoid with frequency $f = 1/12$ is used, the fitted model exhibits a monthly seasonal wave. However, the semi-monthly pattern is still present in the residuals (figure 2.6). The semi-monthly seasonal pattern can be captured by fitting another model - by adding a new harmonic term with $f = 1/6$ to the previous model. The revised fitted curve and the corresponding residuals are graphed in figure 2.7. No periodic behaviour is now present in the residuals. Note that the original data show long-term oscillation with no certain period and that this oscillation still remains in the residuals. This feature can be captured by a decomposition process that will be introduced in later chapters.

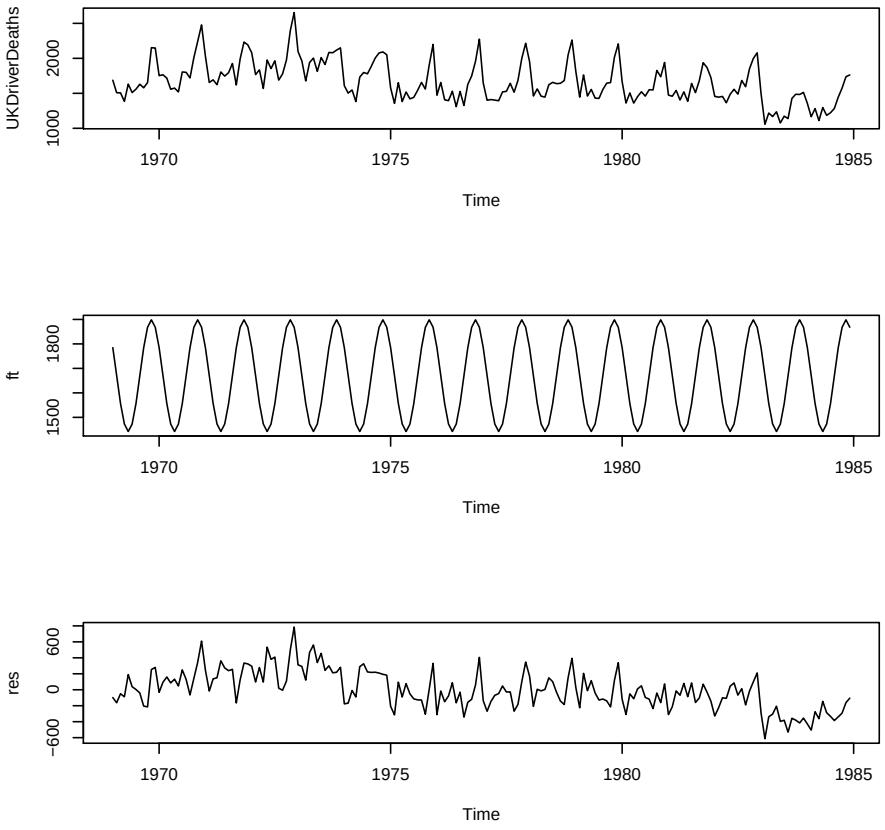


Figure 2.5: Fitted values of the harmonic model and residuals

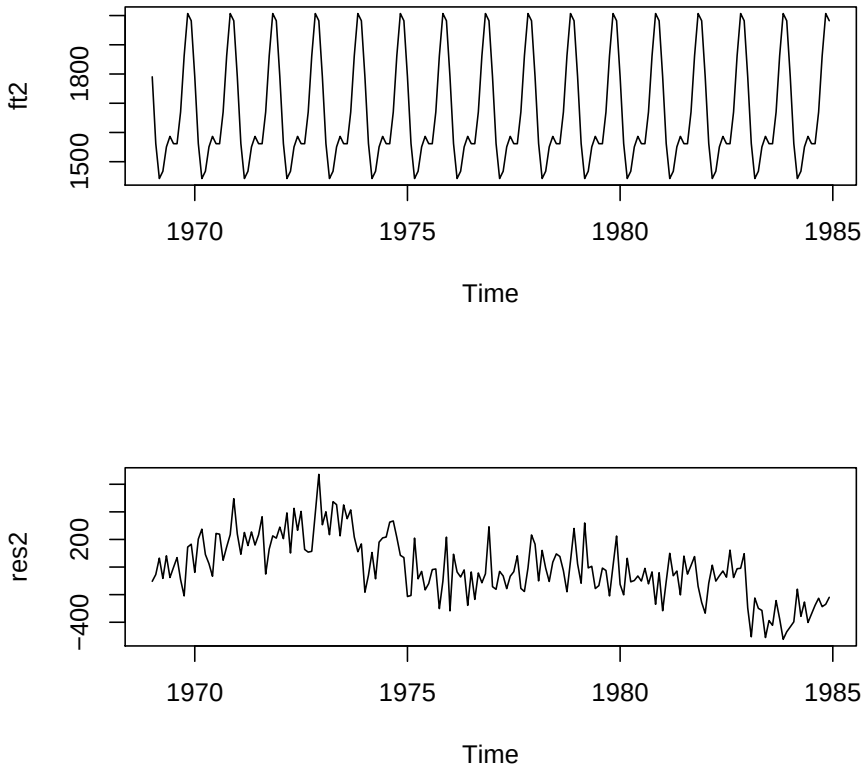


Figure 2.6: Revised fitted values of the harmonic model and residuals

# Chapter 3

# Model Selection Criteria

## 3.1 Introduction

Suppose that we consider a nonparametric model defined as

$$y = g(x) + \epsilon$$

where $g$ is an unknown regression function and $\epsilon$ is an error term. In nonparametric regression, we take into account several possible models with different numbers of parameters. In choosing among our candidate models, our goal is to determine one that is the closest to the true model $g$. That is, we let the data choose the best model. A proper model must not only sufficiently explain the data but must also have as small a number of parameters as possible. This idea is referred to as the parsimony principle.

Generally, the goodness of fit of a model increases as new parameters are added. However, once the parameter dimension becomes sufficiently large, new parameters added afterwards may create an overfitting problem. In other words, after the model achieves enough parameters, adding other parameters may yield models that do not approach the true function $g$ but approach the observed values $y$. One way to avoid this overfitting issue is to penalize the number of parameters in a model. In this chapter, we introduce several model selection methods most of which are based on the penalizing principle. Each method assesses how well a model balances both conformity to the data and parsimony.

In Section 3.2, we present model selecting criteria such as $AIC$ (Akaike, 1974) and $BIC$ (Schwarz, 1978). We use these criteria to compare models that have different parameters and we determine which model is best.

In Section 3.3, we introduce an $R$ package named $FitAR$ (McLeod et al., 2011). Assuming that an $AR(p)$ model properly fits to a given time series, the package calculates the values of a specified criterion as $p$ changes. We then select the optimal $AR(p)$ which minimizes the criterion value.

## 3.2 Selection Criteria

### 3.2.1 Akaike Information Criterion AIC

Akaike Information Criterion $AIC$ (Akaike, 1974) is defined as

$$AIC = -2 \log L + 2p,$$

where $L$ is the maximized likelihood function and $p$ is the number of effective parameters. The best model is the one with the smallest $AIC$. The likelihood function part reflects the goodness of fit of the model to the data, while $2p$ is described as a penalty. Since $L$ generally increases with $p$, $AIC$ reaches the minimum at a certain $p$.

$AIC$ is based on the information theory. The fundamental idea is to measure information loss by computing Kullback-Leibler information between the true model and a given candidate model. Although the true model is unknown, $AIC$ makes it possible to compare estimated information losses from different models. The model with the minimum $AIC$ also minimizes information loss. $AIC$ is asymptotically efficient but is not consistent.

### 3.2.2 Bayesian Information Criterion BIC

Bayesian Information Criterion $BIC$ (Schwarz, 1978) is defined as

$$BIC = -2 \log L + p \log n,$$

where $L$ denotes the maximized value of the likelihood function, $p$ represents the number of parameters and $n$ is the sample size. The model with the minimum value of $BIC$ is preferred. As indicated, $BIC$ strongly penalizes effective parameters compared to the $AIC$ where $\log n > 2$.

## 3.3 R package FitAR

We may determine proper models for a time series using information criterion. An $R$ package $FitAR$ (McLeod et al., 2011) may be employed to construct the best $AR(p)$ model for given data. Assuming that a given time series is adequately explained by an $AR(p)$ model, the package selects the best $p$ in which the information criteria are minimized. One of the criteria such as $AIC$ and $BIC$ may be used.

## 3.4 Illustrative Examples

### 3.4.1 Diagnostic Test of Autoregressive models with Information Criterion

The aim of this simulation is to test information criterion by comparing the true model and a model suggested by model selection process. In order to do so, we use the $FitAR$ package by which the values of a specified information criterion for $AR(p)$ models are automatically calculated with various $p$. In each simulation we employ

either $AIC$ or $BIC$ and the number of parameters $p$ is restricted to be no greater than 20.

1. $AR(1)$ with $\phi = 0.8$, $n = 100$

For the first exercise we use an $AR(1)$ model defined as

$$\begin{aligned} z_t &= \phi_1 z_{t-1} + \epsilon_t \\ &= 0.8 z_{t-1} + \epsilon_t \end{aligned}$$

From the model we obtain sample data with size 100. Tables 3.1 and 3.2 display the two smallest $AIC$ and $BIC$ with corresponding $p$, respectively.

In the first table, $AIC$ is minimized to -7.68 when the model has one parameter. Thus, $AIC$ agrees with the true model. Another simulation is made using $BIC$ and it yields the same conclusion as the first test.

| | p | AIC-Exact | AIC-Approx |
|---|---|---|---|
| 1 | 1 | -7.6819 | -113.2866 |
| 2 | 2 | -6.7473 | -112.2936 |

Table 3.1: Optimal number of parameters with $AIC$ when $AR(1)$

| | p | BIC-Exact | BIC-Approx |
|---|---|---|---|
| 1 | 1 | -2.4716 | -105.4711 |
| 2 | 2 | 1.0682 | -101.8729 |

Table 3.2: Optimal number of parameters with $BIC$ when $AR(1)$

2. $AR(3)$ with ($\phi_1$,$\phi_2$,$\phi_3$) = (0.4,0.7,-0.3). $n = 1000$

An $AR(3)$ model desrcibed as

$$
\begin{aligned}
z_t &= \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \epsilon_t \\
&= 0.8 z_{t-1} + 0.7 z_{t-2} - 0.3 z_{t-3} + \epsilon_t
\end{aligned}
$$

It is used as the true model. In this simulation, $AIC$ disagrees with the true model by selecting $p = 5$. On the other hand $BIC$ still works well with $p = 3$.

| | p | AIC-Exact | AIC-Approx |
|---|---|---|---|
| 1 | 5 | 68.8821 | -935.1556 |
| 2 | 3 | 69.8660 | -934.1150 |

Table 3.3: Optimal number of parameters with $AIC$ when $AR(3)$

| | p | BIC-Exact | BIC-Approx |
|---|---|---|---|
| 1 | 3 | 89.4970 | -909.5762 |
| 2 | 4 | 96.2828 | -907.5547 |

Table 3.4: Optimal number of parameters with $BIC$ when $AR(3)$

3. $AR(4)$ with ($\phi_1$,$\phi_2$,$\phi_3$,$\phi_4$) = (-0.5,-0.7,0.1,0.2), $n = 1000$

We test another simulation with the true model $AR(4)$ with ($\phi_1$,$\phi_2$,$\phi_3$,$\phi_4$) = (-0.5,-0.7,0.1,0.2). For this exercise, both $AIC$ and $BIC$ choose the true parameter space.

| | p | AIC-Exact | AIC-Approx |
|---|---|---|---|
| 1 | 4 | -15.8468 | -1654.1622 |
| 2 | 5 | -14.4951 | -1653.1081 |

Table 3.5: Optimal number of parameters with $AIC$ when $AR(4)$

| | p | BIC-Exact | BIC-Approx |
|---|---|---|---|
| 1 | 4 | 8.6919 | -1624.7157 |
| 2 | 5 | 14.9515 | -1618.7538 |

Table 3.6: Optimal number of parameters with $BIC$ when $AR(4)$

### 3.4.2 Optimal smoothing parameter seletion in loess

We consider the LakeHuron data again. For loess modeling, rather than selecting parameters based on several graphs, we find the optimal choice of parameters using the generalized cross validation (Craven and Wahba, 1979). In $R$, we employ a package named $fANCOVA$, which automatically selects the parameters based on $GCV$. For the LakeHuron time series, the best smoothing and degree parameters are $(\alpha,\ p) = (0.25,\ 1)$ with the residual standard error 0.494 or $(0.50,\ 2)$ with the residual standard error 0.329. Note that the results are somewhat different from those determined in the previous chapter.

Deseasonalizaion

# Chapter 4

# Deseasonalizaion

## 4.1 Introduction

Many time series exhibit seasonal patterns. We may identify a seasonal effect using several graphical methods. Since a seasonal pattern repeats over and over again in the span of the data, we may determine the pattern using statistical methods and use it for analyzing or forecasting. Generally, nonparametric models such as loess, spline regression and harmonic regression are used for estimation.

In Section 4.2, we introduce graphical ways to detect a seasonal pattern in a time series. A seasonal subseries plot gives us much visual information that we cannot easily obtain from the original data plot. It enables us to identify not only relationships between subseries but also patterns within subseries. Multiple boxplots are also especially useful for a time series with a large sample size.

In Section 4.3, we present loess to deal with seasonality. This is a part of seasonal-trend decomposition by loess $STL$ (Cleveland et al., 1990) which will be introduced in Chapter 5.

A repeating seasonal pattern may be estimated by a model using sinusoids. In other words, we may use harmonic regression to determine a seasonal pattern. In Section 4.4, we introduce deseasonalization by harmonic regression (Hipel and McLeod, 1994). In cases where data variations vary with time, scaling is used for variance correction.

Harmonic regression requires specification of the parameter dimension prior to estimating regression coefficients for seasonal means or standard deviations. In Section 4.5, we present how to determine the optimal numbers of harmonics used for estimating seasonal means and standard deviations.

## 4.2 Graphical Methods for Detecting Seasonality

### 4.2.1 Seasonal Subseries Plot

Seasonal patterns in a time series can often be conveniently detected using a seasonal subseries plot (Cleveland, 1993). Assuming that the period of the seasonality is known as $p$, a subseries plot includes $p$ number of season plots, each of which indicates the seasonal mean. With a subseries plot, one can easily detect different features between subseries. It also gives visual inferences between or within season plots.

As an example, we consider a monthly time series of temperatures in Nottingham during 1920-1939. Figures 4.1 and 4.2 show the original series and its seasonal subseries plot, respectively. The original data plot shows a clear seasonal pattern repeating over time. However determining how, on average, the temperature changes over months is not easy. In figure 4.2, the first plot segment consists of January observations ordered in time. The mean of the January values is marked as a short horizontal line. The second subseries corresponds to the February data, and so on. Compared to figure 4.1, figure 4.2 gives much more information regarding seasonality. First, the temperature in Nottingham obviously reaches its highest value in July and its lowest in February, on average. The temperature gap between two consecutive months also tends to be huge in the fall and relatively moderate in the winter season. Lastly, we see some upwards trends in June, August and November. That is, the data

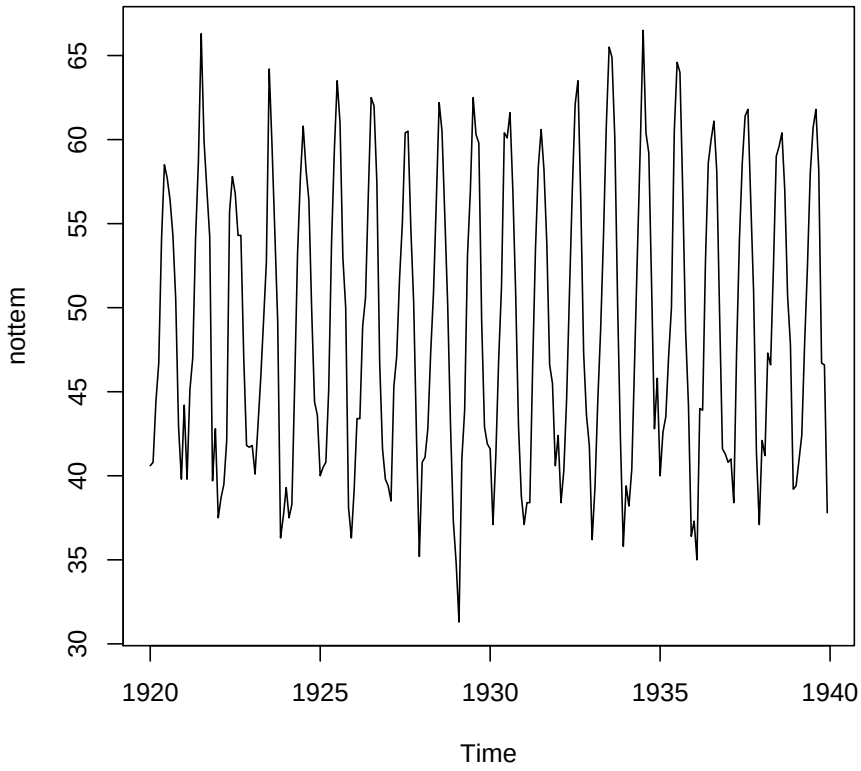


Figure 4.1: Nottem : monthly time series of temperatures in Nottingham during 1920 - 1939

indicate that the temperature during these months tends to increase over years, while the other seasons show no noticeable features.

### 4.2.2 Multiple Box Plots

When a time series has a large number of observations, we may prefer to use multiple box plots to obtain any available clues for the existence of seasonality. This consists of $p$ number of box plots, where $p$ represents the number of seasons in a year. We create multiple boxplots for the nottem data that are shown in figure 4.3. Similar

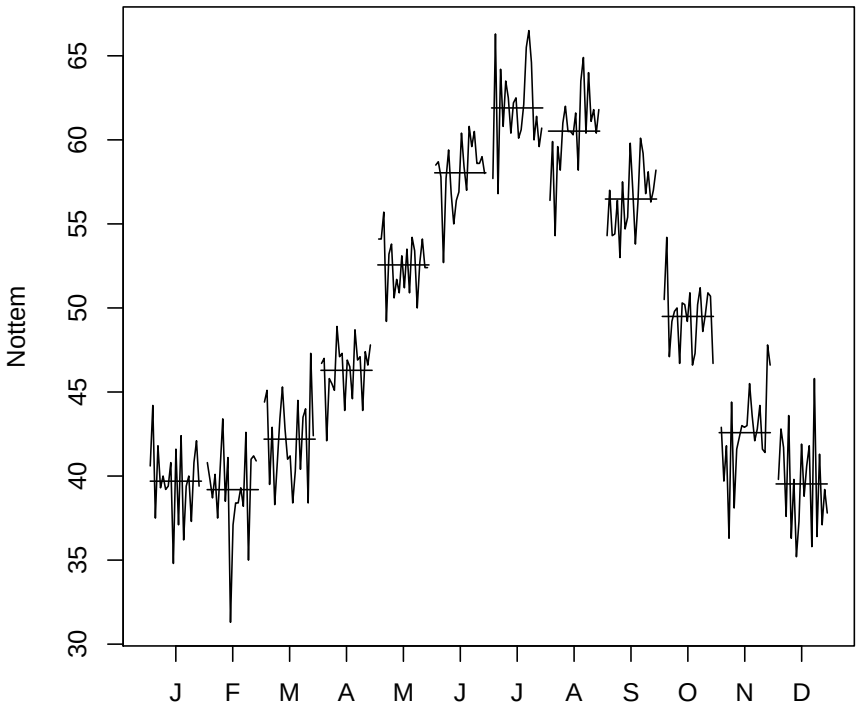


Figure 4.2: Seasonal Subseries Plot for nottem data

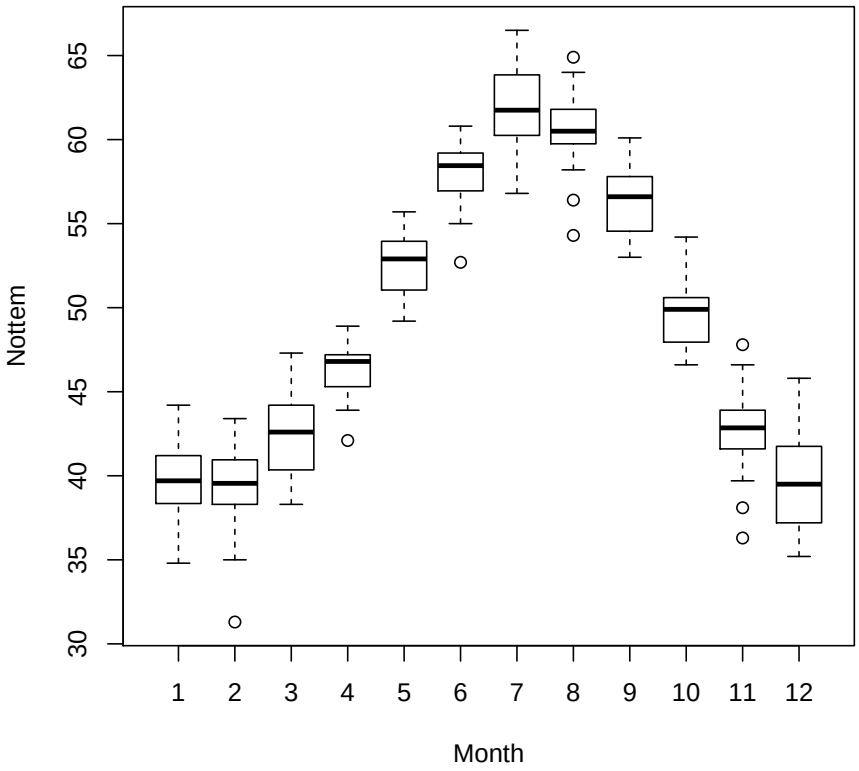


Figure 4.3: Monthly boxplots for nottem data

to a seasonal subseries plot the boxplots exhibit a strong seasonal pattern between months. Each boxplot gives a specific location and variation information for each month. However, the boxplots provide no trend or pattern information for each month.

## 4.3 Deseasonalization by Loess

One can obtain a deseasonailzed time series by employing loess. Note that two positive integers, $k$ and $p$ must be chosen to use loess. Let $x$ be any value of the independent

variable. Then $k$ nearest neighbours of $x$ are selected and the neighbourhood weights are given to each neighbour. The neighbourhood function is introduced in Chapter 2. After that, a local polynomial with degree $p$ is fitted to the weighted values. The local polynomial value at $x$ is the loess fit. Once a local regression fits to the data, a deseasonalized series is easily determined by subtracting the fitted values from the original data.

## 4.4 Deseasonalization by Harmonic Regression

For a time series $z_t$, we start with an assumption that the series after removing seasonal effect is stationary. A deseasonalized series, $w_t$, may be written as

$$w_t = \frac{z_t - \mu_t}{\sigma_t}$$

where $\mu_t$ are the seasonal means and $\sigma_t$ represent the seasonal standard deviations. Scaling is used to make a deseasonalized series stationary if $\sigma_t$ change over time. On the other hand, if a reasonable model for seasonal standard deviations is a constant model, it may be evidence of constant seasonal variance and thus scaling is unnecessary.

In reality, $\mu_t$ and $\sigma_t$ are unknown and thus we need to estimate them using harmonic regression. For a harmonic model, we may use as many predictors (i.e. sinusoids) with different frequencies as we want. Often, however, using too many predictors is unnecessary and even brings on an overfitting problem. Fortunately, based on several information criteria such as $AIC$ and $BIC$, we obtain the best parameter dimensions for the seasonal mean and standard deviation models. The process is discussed in section 4.5.

In harmonic regression, the true model for $z_t$ is written as

$$z_t = A_\mu^{(0)} + \sum_{i=1}^{Fm}(A_\mu^{(i)} \cos(\frac{2\pi it}{s}) + B_\mu^{(i)} \sin(\frac{2\pi it}{s})) + \epsilon_t$$

where $A_\mu^{(0)}$ is the overall mean, $Fm$ is the number of sinusoids used, $A_\mu^{(i)}$, $B_\mu^{(i)}$ where i = 1, 2, ... , $Fm$ denote the sinusoid parameters, $s$ is the seasonal period with $s = 12$ for monthly data and 365.25 for daily data and $\epsilon_t$ is the error term. Similarly, we may also define a regression model for $e_t^2 = (z_t - \hat{\mu}_t)^2$ as

$$e_t^2 = A_\sigma^{(0)} + \sum_{i=1}^{Fs}(A_\sigma^{(i)} \cos(\frac{2\pi it}{s}) + B_\sigma^{(i)} \sin(\frac{2\pi it}{s})) + v_t$$

The numbers of sinusoids, $Fm$ and $Fs$, must be chosen. A strategy for optimal $Fm$ and $Fs$ is explained in Section 4.5.

Once $Fm$ and $Fs$ are determined, the parameters of each model may be estimated by the least square method. Note that the sample mean is used if $Fm = 0$ and the sample standard deviation if $Fs = 0$. After the least square estimates are obtained, $\mu_t$, $\sigma_t^2$ and the deseasonalized series $w_t$ may be respectively described as

$$\hat{\mu}_t = \hat{A}_\mu^{(0)} + \sum_{i=1}^{Fm}(\hat{A}_\mu^{(i)} \cos(\frac{2\pi it}{s}) + \hat{B}_\mu^{(i)} \sin(\frac{2\pi it}{s}))$$

$$\hat{\sigma}_t^2 = \hat{A}_\sigma^{(0)} + \sum_{i=1}^{Fs}(\hat{A}_\sigma^{(i)} \cos(\frac{2\pi it}{s}) + \hat{B}_\sigma^{(i)} \sin(\frac{2\pi it}{s}))$$

and

$$\hat{w}_t = \frac{z_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

If no scaling is desired, we use $\hat{w}_t = z_t - \hat{\mu}_t$ for all $t$.

## 4.5 Selecting the Optimal Model

Assuming that a deseasonalized series is stationary and is adequately explained by an $AR$ model, then the optimal $Fm$ and $Fs$ can be calculated based on $AIC$ or $BIC$ (McLeod and Gweon, 2013). As mentioned earlier, $Fm$ and $Fs$ are the numbers of sinusoids used in harmonic models for estimating seasonal means and standard deviations, respectively. For each model, the range of $Fm$ and $Fs$ should be restricted. For a monthly time series, since its frequency is fixed as 12, if we consider harmonics with Fourier frequencies we have at most 6 sinusoids to avoid aliasing. In other words, for monthly data $0 \leq Fm$, $Fs \leq 6$. For a daily time series, since the seasonal behaviour is generally not too complicated it is practically reasonable to set $0 \leq Fm$, $Fs \leq 6$.

On the other hand, it is convenient to use generalized $AIC$, denoted as $GIC_\alpha$. It is written as

$$GIC_\alpha = -2 \log L + \alpha k$$

where $L$ is the maximized log-likelihood, $\alpha$ is 2 for the $AIC$ and $\log(n)$ for the $BIC$ and $k$ denotes the number of free parameters. We assume that a deseasonalized series $w_t$ is stationary and there exists an $AR(p)$ that can adequately explain $w_t$. For each pair of $Fm$ and $Fs$, where $0 \leq Fm \leq 6$ and $0 \leq Fs \leq 6$, $w_t$ is obtained by least sqaure estimation. We then find $p$ that minimizes $GIC_\alpha$. Note that each pair of $Fm$ and $Fs$ gives the $p$ corresponding to the minimum of $GIC_\alpha$. Among all the combinations, we choose the $Fm$ and $Fs$ by which $GIC_\alpha$ becomes the minimum.

On the other hand, the scales of the deseasonalized series change as $Fs$ varies. Comparison of $GIC_\alpha$ requires consideration of the scale effect. The scale effect can be adjusted by subtracting $2 \log J$ from $GIC_\alpha$, where $J$ is the Jacobian of the transformation $w_t$ into $z_t$. The Jacobian is obtained as follows.

For a time series $z_t$, the deseasonalized series $w_t$ is written as

$$\hat{w}_t = \frac{z_t - \hat{\mu}_t}{\hat{\sigma}_t}.$$

Thus the corresponding Jacobian matrix is written as

$$\begin{bmatrix} \frac{\partial w_1}{\partial z_1} & \frac{\partial w_1}{\partial z_2} & \cdots & \frac{\partial w_1}{\partial z_n} \\ \frac{\partial w_2}{\partial z_1} & \frac{\partial w_2}{\partial z_2} & \cdots & \frac{\partial w_2}{\partial z_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial w_n}{\partial z_1} & \frac{\partial w_n}{\partial z_2} & \cdots & \frac{\partial w_n}{\partial z_n} \end{bmatrix}.$$

From the relationship between $z_t$ and $w_t$, the following is easily obtained.

$$\frac{\partial w_i}{\partial z_j} = \begin{cases} \hat{\sigma}_j^{-1} & i = j \\ 0 & otherwise \end{cases}$$

The Jacobian $J$ is described as

$$J = \begin{vmatrix} \frac{\partial w_1}{\partial z_1} & \frac{\partial w_1}{\partial z_2} & \cdots & \frac{\partial w_1}{\partial z_n} \\ \frac{\partial w_2}{\partial z_1} & \frac{\partial w_2}{\partial z_2} & \cdots & \frac{\partial w_2}{\partial z_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial w_n}{\partial z_1} & \frac{\partial w_n}{\partial z_2} & \cdots & \frac{\partial w_n}{\partial z_n} \end{vmatrix} = \begin{vmatrix} \hat{\sigma}_1^{-1} & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}_n^{-1} \end{vmatrix} = \prod_{t=1}^{n} \hat{\sigma}_t^{-1}.$$

Thus the adjusted generalized $AIC$, denoted as $AGIC_\alpha$, can be written as

$$\begin{aligned} AGIC_\alpha &= GIC_\alpha - 2\log J \\ &= GIC_\alpha - 2\log \prod_{t=1}^{n} \hat{\sigma}_t^{-1} \\ &= GIC_\alpha + 2\sum_{t=1}^{n} \log \hat{\sigma}_t \end{aligned}$$

.

Note that $AGIC_\alpha = GIC_\alpha$ if $\hat{\sigma}_t$=1 for all $t$, as no scale is applied to the original series. The optimal pair of $Fm$ and $Fs$ is the one that minimizes $AGIC_\alpha$.

## 4.6 Illustrative Examples

### 4.6.1 Monthly Saugeen River Flow, 1915 - 1979

We start with monthly Saugeen River flow data ($m^3/sec$) that were measured at Walkerton from 1915 to 1979. Figure 4.4 shows 744 observations and due to the large number of observations, we use the cut-and-stack method. Multiple box plots for monthly subseries are shown in figure 4.5. The graphs clearly show that a seasonal pattern is present in the Saugeen data and many of the monthly distributions are positively skewed. Generally, positive skewness can be handled by taking a log-transformation. The time series plot and multiple box plots of log-transformed data are shown in figure 4.6 and 4.7. Compared to the previous plots, the monthly distributions become more symmetrical.

| Fm | Fs | p | AIC |
|---|---|---|---|
| 5.000 | 4.000 | 3.000 | -1171.936 |

In $R$, we use a *deseasonalize* package (McLeod and Gweon, 2012). Using $AIC$, the optimal choice of ($Fm$,$Fs$) is (5,4) with the lowest $AIC$ = -1171.936. Note that $Fs = 4$ indicates that the monthly standard deviations are not constant.

With the suggested parameter dimensions, the regression model for $\mu_t$ is described as

$$\hat{\mu}_t = \hat{A}_\mu^{(0)} + \sum_{i=1}^{5} (\hat{A}_\mu^{(i)} \cos(\frac{\pi i t}{6}) + \hat{B}_\mu^{(i)} \sin(\frac{\pi i t}{6}))$$

where

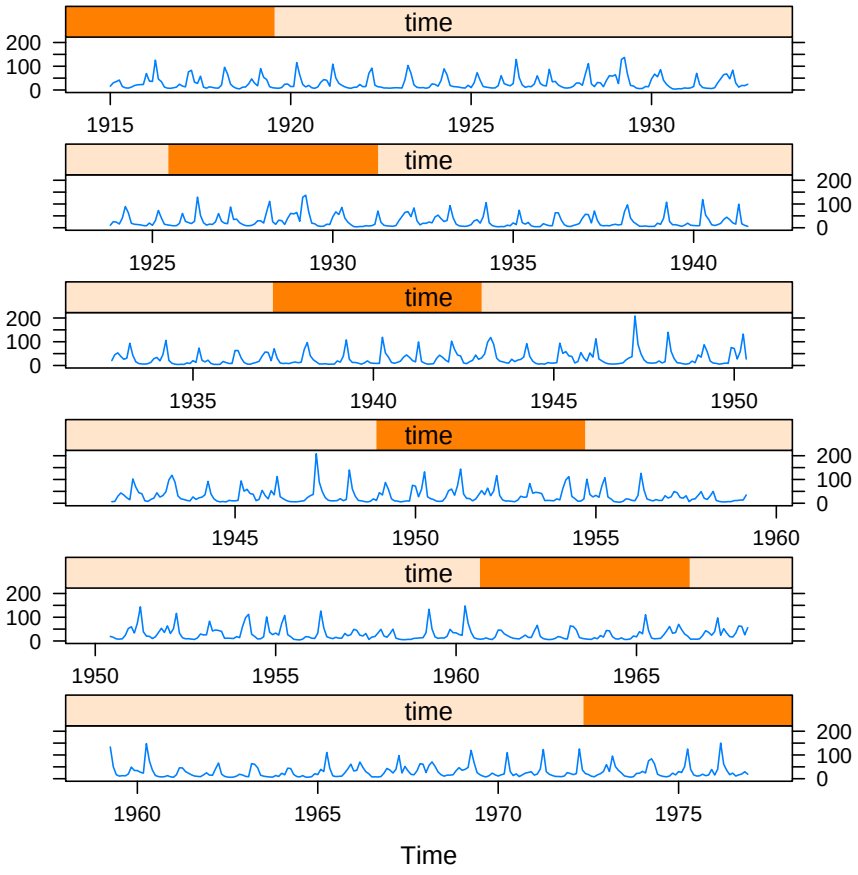


Figure 4.4: Monthly Saugeen River Flow during 1915 - 1979

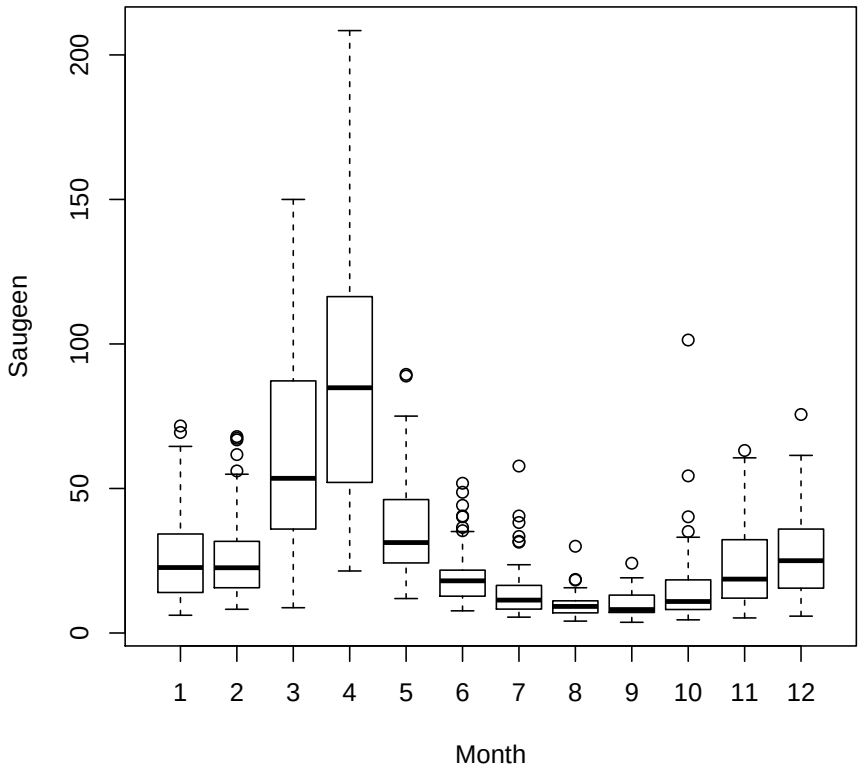


Figure 4.5: Boxplots of monthly subseries of Saugeen

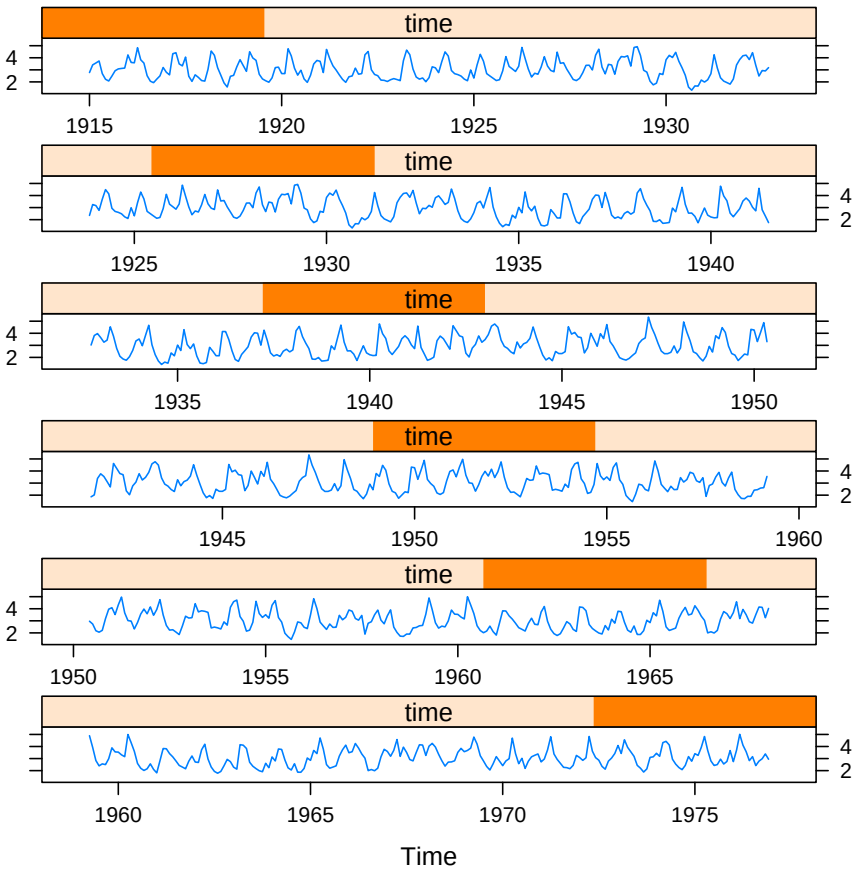


Figure 4.6: log(Saugeen) during 1915 - 1979

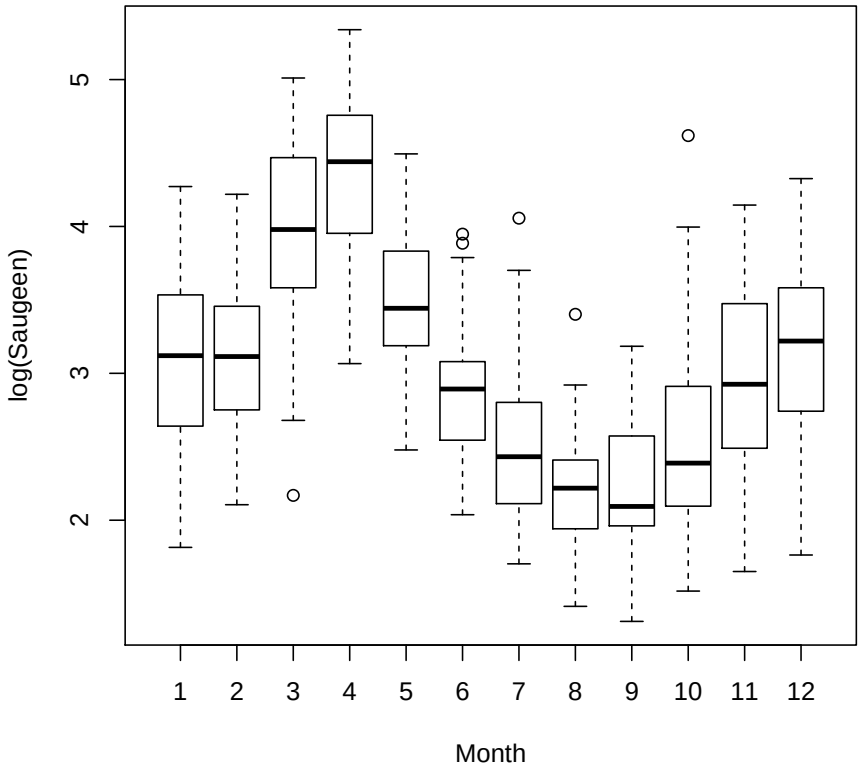


Figure 4.7: Boxplots of log(Saugeen)

```
        A0          A1          B1          A2          B2          A3
3.04178556  0.78534867 -0.02395692 -0.34463950 -0.03273296 -0.10124492
        B3          A4          B4          A5          B5
0.19832998  0.10235134  0.01009820 -0.01381673 -0.04652837
```

Similarly, the harmonic model for $\sigma_t$ is

$$\hat{\sigma}_t^2 = \hat{A}_\sigma^{(0)} + \sum_{i=1}^{4} (\hat{A}_\sigma^{(i)} \cos(\frac{\pi i t}{6}) + \hat{B}_\sigma^{(i)} \sin(\frac{\pi i t}{6}))$$

where

```
        A0          A1          B1          A2          B2
0.279625538  0.028551266  0.097286343 -0.011296493  0.025479986
        A3          B3          A4          B4
-0.033501226 -0.022538696  0.033870862 -0.006264534
```

Finally, the deseasonalized series is determined by

$$\hat{w}_t = \frac{z_t - \hat{\mu}_t}{\hat{\sigma}_t}.$$

For the our example, the deseasonalized series is shown in figure 4.8. No seasonal pattern is evident.

### 4.6.2 Daily maximum temperatures in Melbourne, Australia, 1981-1990

We try another example with daily time series. The data consist of minimum temperatures in Melbourne during 1981-1990 and are shown in figure 4.9. The series has a seasonal pattern that repeats 10 times for 10 years. Based on $BIC$ criteria, the optimal numbers of parameters for seasonal means and standard deviations are determined as 2 and 0. $Fs = 0$ is an evidence of a constant standard deviation and thus no

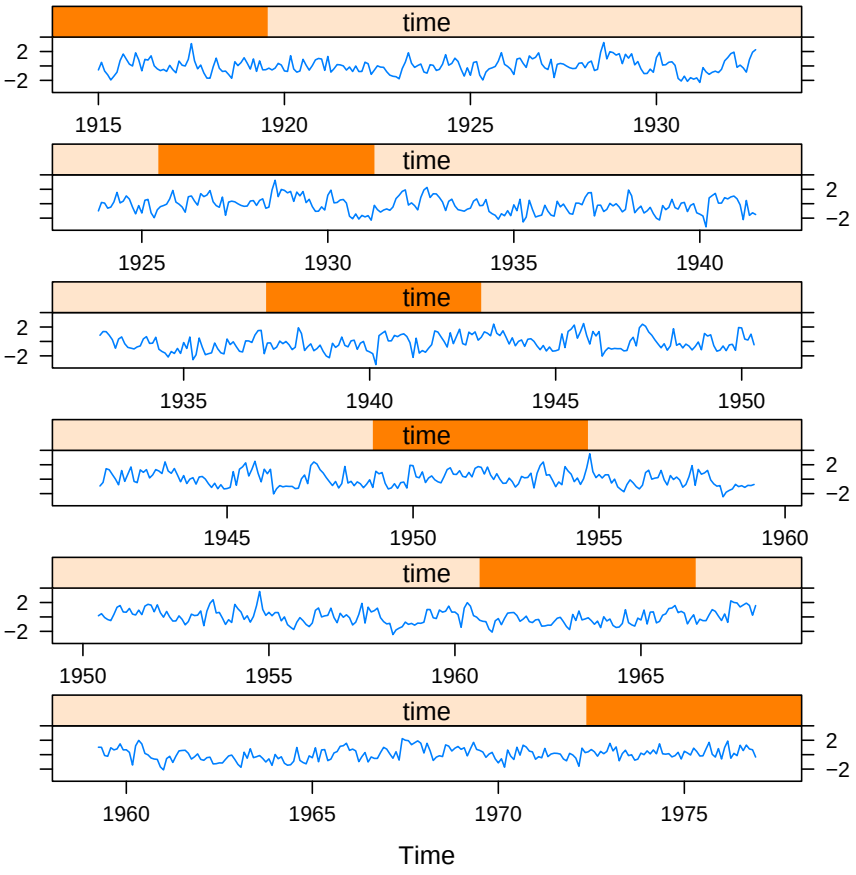


Figure 4.8: Deseasonalized series

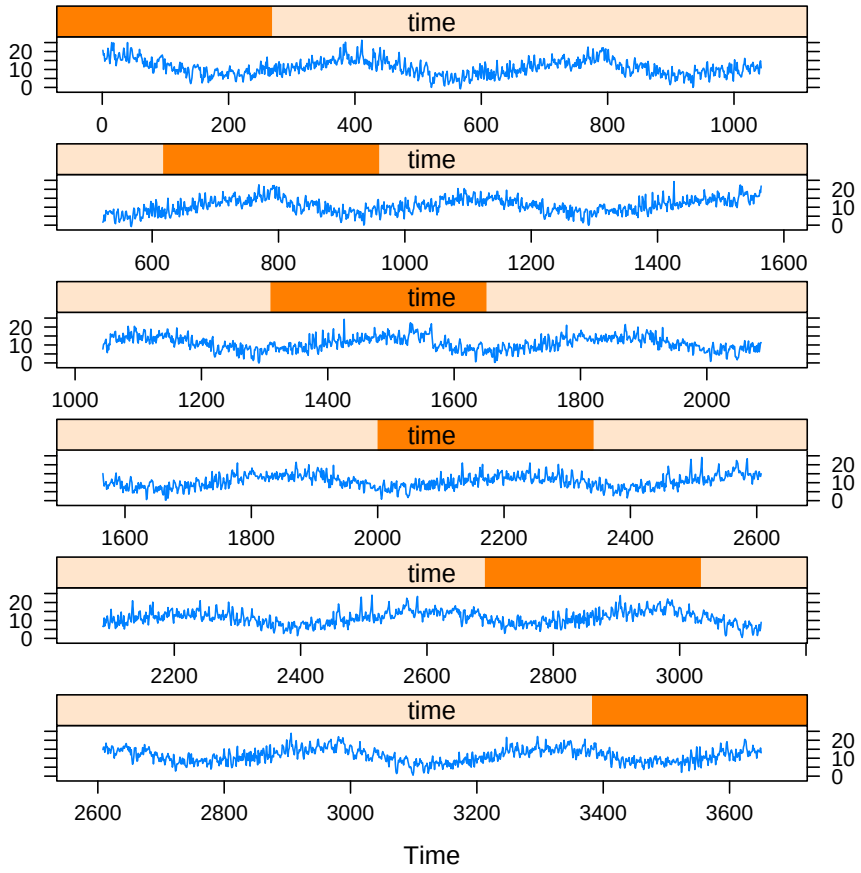


Figure 4.9: Daily maximum temperatures in Melbourne during 1981-1990

scaling is needed. In this case the deseasonalized series is simply obtained by subtracting seasonal means from the original series. Figure 4.10 displays the deseasonalized series. No pattern or variation change is present.

| Fm | Fs | p | BIC |
|---|---|---|---|
| 2.000 | 0.000 | 2.000 | 6328.215 |

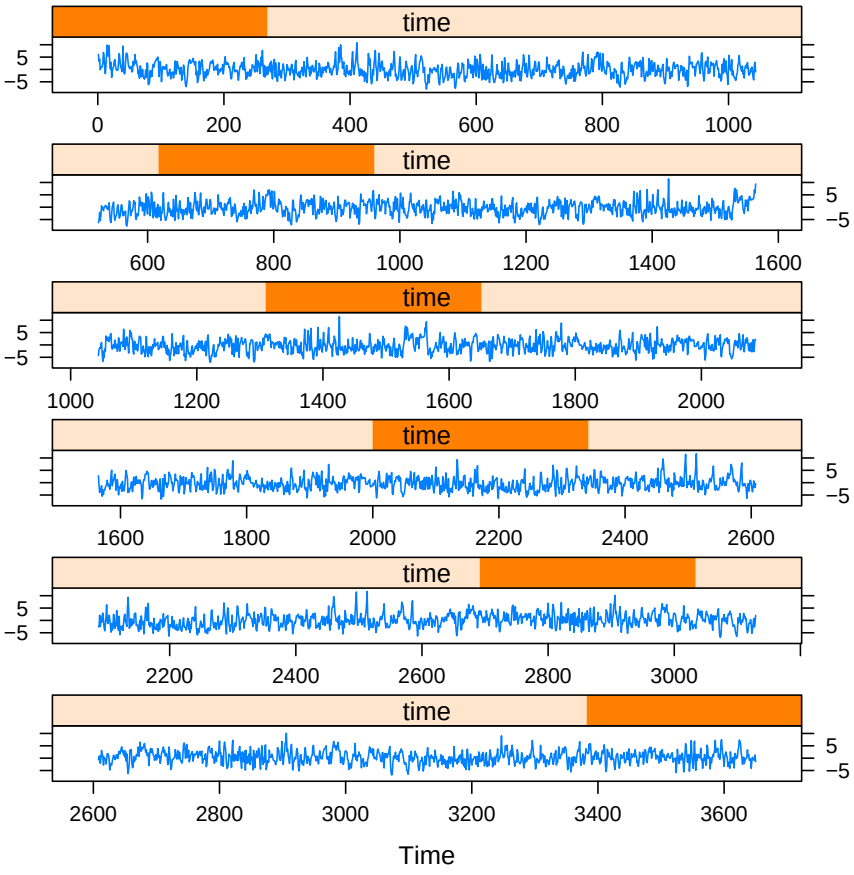


Figure 4.10: Deseasonalized maximum temperature series

# Chapter 5

# Decomposition

## 5.1 Introduction

A time series will often show systematic patterns that results from the nature of the data. Generally, several components, such as trend and seasonal terms, may be underlying in a time series. Knowing the underlying components is useful for analyzing and forecasting. Fortunately these components may be determinable under certain conditions. The decomposition process is a method used to determine systematic components separately.

For a time series decomposition, we generally assume either an additive or multiplicative model. For example, under the additive model assumption a time series $z_t$ may be described as $z_t = T_t + S_t + I_t$, where $T_t$ is long-term trend and $S_t$ and $I_t$ represent seasonal and irregular components, respectively. If we employ a multiplicative model, then $z_t = T_t S_t I_t$. Fortunately, a multiplicative model is easily transformed into an additive model by taking logarithms on both sides of the model. In Section 5.2, we start by introducing the concept of additive models. The additive effect assumption is fairly strong.

Generally, a component is determined after the other systematic terms are removed from the original time series. For instance, a model for seasonality is fitted to a detrended series. Similarly, we determine a trend component using a deseasonalized series. Thus, the decomposition process is basically an iterative process. For additive models, we use the backfitting algorithm, which is introduced in Section 5.3.

In Section 5.4, we review the $STL$ that is widely used for decomposition. Each seasonal and trend component must have a smoothing parameter and polynomial degree provided. Proper parameters may be suggested by cross validation.

In Section 5.5, we propose another decomposition method that uses both loess and harmonic regression. We employ loess for long-term components and harmonic regression for seasonality.

When we use an additive model $z_t = T_t + S_t + I_t$, the irregular term $I_t$ is defined as $z_t - T_t - S_t$. That is, it represents remainder effects that neither the trend nor the seasonal component can explain. It is desired that an irregular component exhibits no pattern. Occasionally, however, $I_t$ illustrates a cyclical pattern that is evidence of the existence of another systematic component. In that case, the cyclical information may also be determined (Cleveland, 1993). In Section 5.6, we introduce an algorithm for the additional cyclical component case.

## 5.2 Additive Models

For a response variable $Y$ and covariates $X_1$, ... , $X_p$, the general nonparametric model may be written as

$$Y = f(X_1, X_2, ..., X_p) + \epsilon$$

where $\epsilon$ is IID error term with zero mean and a constant variance. If we assume an additive model then

$$f(X_1, X_2, ..., X_p) = f_1(X_1) + f_2(X_2) + ... + f_p(X_p).$$

That is, the additive model with $p$ number of smooths is a special case of the smooth function with $p$ covariates. In an additive model, the impact of each covariate is additive.

## 5.3 Backfitting Algorithm

The backfitting algorithm (Friedman and Stuetzle, 1981) is an iterative approximation process used in order to fit a generalized additive model (GAM). When $X_1, ..., X_p$ are predictor variables and $Y$ is the dependent variable, a GAM may be described as

$$E(Y|X_1, X_2, ..., X_p) = f_0 + \sum_{j=1}^{p} f_j(X_j)$$

where $f_j$, $j = 1, 2, ...p$ are smoothing functions. First the mean of the whole data is subtracted from each observation and a regression is fitted to the deviations. The regression surface is then also removed from the data and we fit another regression to the deviations, and so on until all $p$ functions are estimated.

For the next iteration, we update each coefficient of the regression model using the adjusted data in which the global mean and all other recently-updated-regression surfaces are subtracted. This process continues until all coefficients do not change substantially. The following explains the algorithm.

1. Initializing: $f_0 = n^{-1} \sum_{i=1}^{n} y_i$ and $f_j \equiv 0$, $j = 1, 2, ..., p$

2. At each iteration, $f_j$ is updated as follows from $j$=1 to $p$.

$$Y^{(j)} = Y - f_0 - \sum_{k \neq j} f_k(X_k)$$

$$f_j = E(Y^{(j)}|X_j)$$

3. Repeat step 2 until $f_j$ converges for all $j$.

## 5.4 Seasonal Trend Decomposition by Loess

Seasonal trend decomposition by loess (Cleveland, 1979) is a decomposition method widely used to determine the trend, seasonal and remainder components. Since loess models are fitted for both trend and seasonal terms, two smoothing parameters ($\alpha_t$, $\alpha_s$) and two polynomial degrees ($p_t$, $p_s$) must be chosen.

In stl, we assume an additive model. That is, the original time series , $z_t$, can be written as

$$z_t = T_t + S_t + I_t$$

where $T_t$, $S_t$ and $I_t$ represent the trend, seasonal and irregular components, respectively. We decompose the series by applying the backfitting algorithm. At the first stage, a trend component $T_{t,1}$ is first determined by fitting loess to $z_t$ with $\alpha_t$ and $p_t$. A seasonality $S_{t,1}$ is then fitted to the residuals with $\alpha_s$ and $p_s$. The irregular term $I_{t,1}$ is obtained by subtracting both the trend and seasonality from the original series.

During the second iteration, we update the trend component $T_{t,1}$ to $T_{t,2}$ by fitting loess to $z_t - S_{t,1}$. Similarly, we replace the seasonal term to $S_{t,2}$, which is computed using $z_t - T_{t,2}$. The irregular component at the second iteration is then written as $I_{t,2} = z_t - T_{t,2} - S_{t,2}$.

We continue the iterative process until all parts converge. This beckfitting process results in a trend term that contains no seasonal effects and no long-term impact is present in the seasonal component after decomposition.

## 5.5 Decomposition using Loess and Harmonic Regression

In the previous section only local regression is used for decomposition. However, harmonic regression is also useful to explain seasonal patterns of time series. Where $T_t$ represent the trend component, we assume that the detrended series $z_t - T_t$ is described as

$$z_t - T_t = A^{(0)} + \sum_{i=1}^{Fm} (A^{(i)} \cos(\frac{2\pi it}{s}) + B^{(i)} \sin(\frac{2\pi it}{s})) + \epsilon_t$$

where $A^{(0)}$ is the overall mean, $Fm$ is the number of sinusoids used, $A^{(i)}$ and $B^{(i)}$, $i$=1,2,...,$Fm$ denote the sinusoid parameters, $s$ is the seasonal period with $s$=12 for monthly data and 365.25 for daily data and $\epsilon_t$ is the error term. Using the least square estimation, we obtain the seasonal component estimates $\hat{S}_t$. That is,

$$\hat{S}_t = \hat{A}^{(0)} + \sum_{i=1}^{Fm} (\hat{A}^{(i)} \cos(\frac{2\pi it}{s}) + \hat{B}^{(i)} \sin(\frac{2\pi it}{s})).$$

Instead of $z_t - T_t$, we use $z_t - \hat{T}_t$ where $\hat{T}_t$ is obtained by loess, as the actual trend term is unknown. After $\hat{T}_t$ and $\hat{S}_t$ are determined, $I_t$ is estimated as $\hat{I}_t = z_t - \hat{T}_t - \hat{S}_t$. Assuming an additive model, the components are updated during the backfitting process until they converge.

## 5.6 More than one trend component case

It is desirable that the irregular component remaining after decomposition shows no oscillatory movements. Often, however, the remainder series exhibits an cyclical behaviour even after the trend and seasonality are removed from the original series. Actually the underlying behaviour should also be captured and extracted from the irregular component.

In the previous sections, $z_t$ is described as an additive model $z_t = T_t + S_t + I_t$. It is desirable that the irregular component $I_t$ shows no oscillatory movements. Often, however, the condition may be violated if there is another underlying pattern apart from the trend and seasonal components. Generally, we want to determine all components separately that underlie a time series and thus any pattern present in $I_t$ should also be captured and extracted. One simple way is to revise the original additive model by adding an oscillatory component, denoted as $O_t$, to the model. That is, if a cyclical component is present in $I_t$, then a revised model is written as

$$z_t = T_t + S_t + O_t + I_t$$

so that $I_t$ shows no systematic movements. We may use loess to fit a model for the movements (Cleveland,1993?). To determine $O_t$, we use the backfitting algorithm with the deseasonalized series $z_t - S_t$ where $S_t$ is obtained assuming $z_t = T_t + S_t + I_t$. That is, $T_t$ and $O_t$ are revised in each iteration and the remainder term $I_t$ is written as

$$I_t = z_t - S_t - T_t - O_t.$$

The iteration repeats until the components become stable.

## 5.7 Illustrative Examples

### 5.7.1 Monthly CO2 from Mauna Loa

The time series consists of 468 observations of monthly atmospheric concentrations of carbon dioxide values from Mauna Loa. Each is expressed in parts per million (ppm). The data are displayed in figure 5.1. The time series $z_t$ exhibits an upward

linear trend and seasonal cycles that repeat with a 12-month period. We assume an additive model and try the two different methods introduced in the previous section.

For $STL$ decomposition, there is a default function in $R$ named *stl*. Use of the function requires that two smoothing parameters and two degrees of polynomials be given. For this exercise, we choose smoothing parameters 0.215 and 0.075 for the trend and seasonality respectively. Local linear models are used for both terms. The decomposition result is shown in figure 5.2. The CO2 time series is now separated into trend, seasonal and remainder components. As expected, the trend term exhibits a linear-like curve and the seasonal component shows periodic fluctuations with slightly increasing peak amplitude over time. The remainder term is determined by subtracting both the trend and seasonal terms from the original series. It also exhibits systematic cycles even though there is no certain period. Note that a decomposition result depends on parameter selection. Had the model used different parameters for components, the process might have given different results.

We compare stl with the method that uses both loess and harmonic regression by decompose CO2 data once again (figure 5.3). The result is not very different from that of stl; a clear upward trend, stable seasonal cycles and a remainder term with oscillations. One difference is that the peak amplitude of the seasonal component is fixed, while it may vary in stl.

We now consider finding the cyclical pattern present in the remainder component. The oscillation repeats 9 times through the span of the series. We revise the model to $z_t = T_t + S_t + O_t + I_t$ and estimate the components. We fit the oscillatory term by using local quadratic curves with smoothing parameter 0.075. As shown in figure 5.4, the oscillatory term is removed and the remainder component has no systematic behaviours.

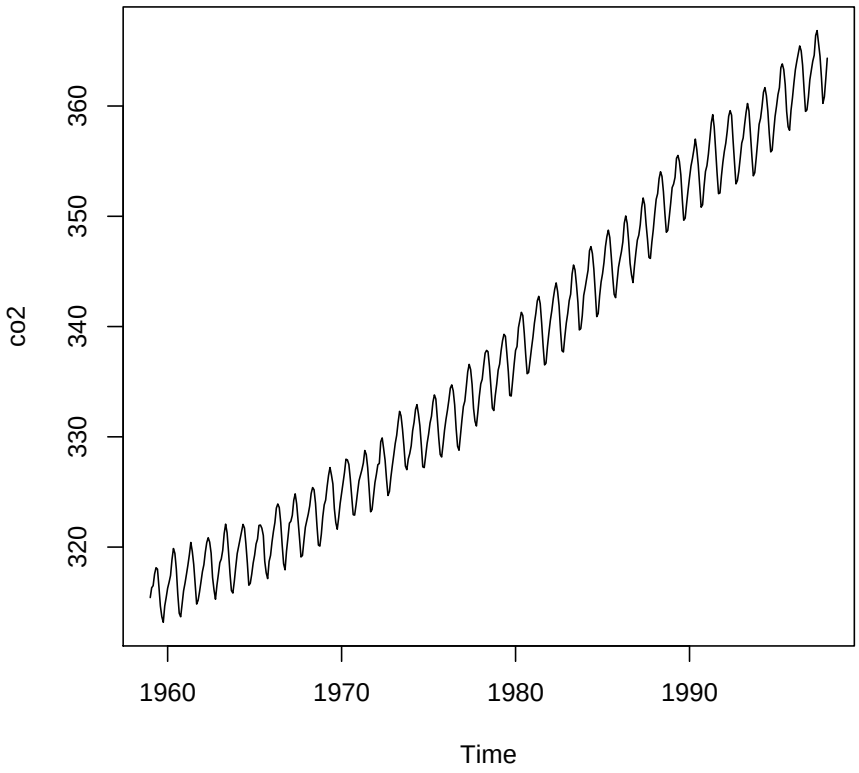


Figure 5.1: $CO2$ data time series

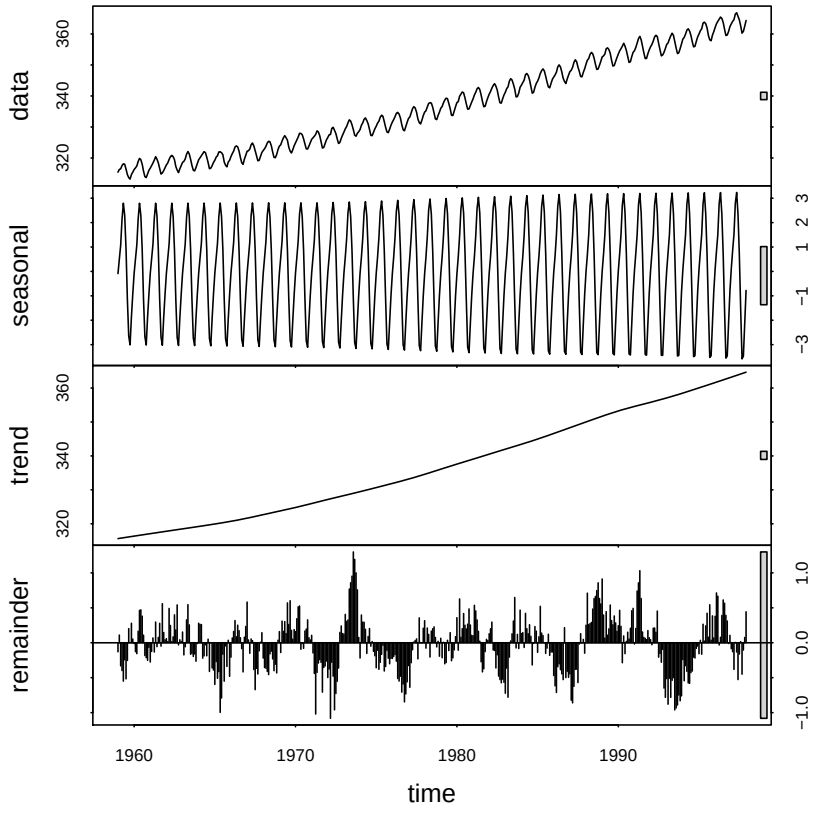


Figure 5.2: Decomposition of the $CO2$ series using stl

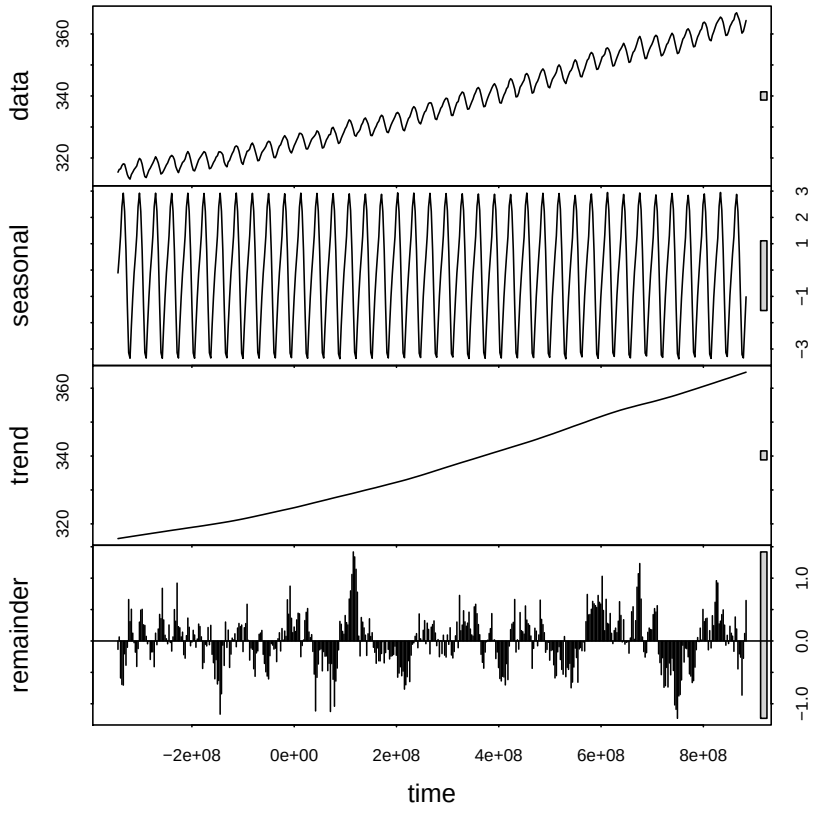


Figure 5.3: Decomposition of the $CO2$ series using sltl

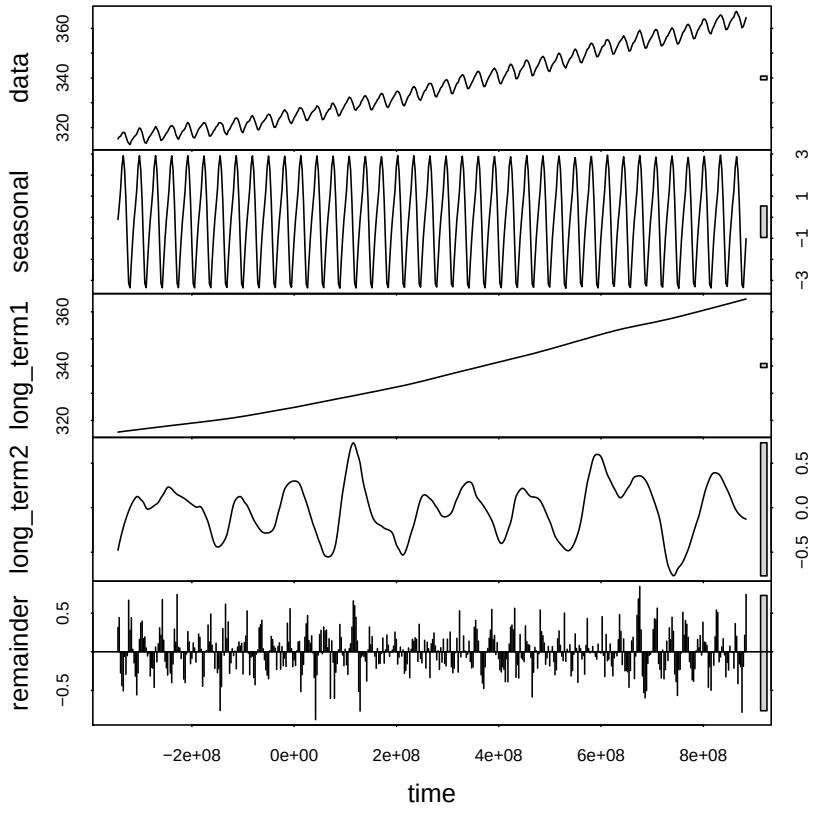


Figure 5.4: Decomposition of the $CO2$ series including the oscillatory component

# Chapter 6

# Missing Data Problem

## 6.1 Introduction

When we analyze time series data, the data should be carefully collected and evenly spaced. In reality, however, such conditions rarely occur. A common issue that we may encounter when analyzing time series is missing observations.

There are many causes of data observation failure: for example, results of non-response. In some fields, it is possible that data are systematically missing. For instance, daily stock price data of a company are measured only during weekdays in many countries. In this case, all values for annual holidays will be missing. Based on the nature of missing data, we may create three categories of problems: MCAR(Missing Completely At Random), MAR (Missing At Random) and MNAR (Missing Not At Random).

Data elements are classified as MCAR if the probability that an observation is missing is unrelated to the observed or missing values. If the probability depends on observed variables but not on values of the missing variable, then that data are classified as MAR. On the other hand, data are classified as MNAR if they are neither MCAR nor MAR. Generally, missing data are ignorable if they are MCAR or MAR. Otherwise, missing data are non-ignorable.

To handle missing data problems, we may use strategies such as deletion, direct analysis and multiple imputation. For time series analysis, it is often desired that missing observations be estimated using a proper methodology. In particular, if the

series exhibits seasonal cycles repeating over time, we may measure the seasonal information and use it for an estimation of the missing data. In Section 6.2, we discuss the missing data problem in seasonal time series.

In Section 6.3, we introduce a useful algorithm (McLeod et al., 1983), by which we can measure the seasonality or trends of unevenly spaced time series. The algorithm may be used not only on monthly data but also daily time series.

## 6.2 The Missing Data Problem in Seasonal Time Series

An unevenly spaced time series makes analysis difficult. In order to make inferences for the irregularly spaced time series, we must often estimate missing values. There are many deterministic or stochastic models for doing so. In time series data, the specific method depends on the nature of the data, as we want to use as much information for estimation as possible. For example, if a monthly time series exhibits a seasonal pattern, we may believe that the seasonal effect for each month repeats every year as a part of the monthly value. In this case, estimating missing values based on the repeating intra-year pattern determined by a proper algorithm may be a superior method than estimating using other methods such as least square approximation and spline interpolation.

## 6.3 Seasonal Adjustment Algorithm

An equal time interval condition of a time series is essential for any stochastic modeling. Various seasonal adjustment procedures may be used to build evenly spaced time series. One available procedure is the seasonal adjustment algorithm (McLeod et al., 1983) which can be used even when a large proportion of data are missing. For

the original time series $z_t$, we assume an additive seasonal adjustment model defined as

$$z_t = C_t + S_t + I_t = C_y + S_m + I_t$$

where $t$ denotes the Julian day number, $y$ is the year and $m$ is either the month for monthly data or the day for daily data. $C$ is the year-based long-term component, $S$ is the seasonal factor that is affected not by years but by seasons, and $I$ is the irregular term that contains the remainder of effects, except for $C$ and $S$. The seasonal adjustment algorithm is as follows.

(1) Set the initial estimates of each component. Let $k$ be either 12 for monthly data or 366 for daily data. The leap year case is taken into account and, thus, the stable seasonal factor $S_m$ should be considered 366 for the number of seasons for a daily time series. Let $k$ be 12 for monthly data or 366 for daily data. Then, we initialize the components as follows:

$$\begin{aligned} C_y &= C \\ &= median(z_t) \\ S_m &= S'_m - \frac{1}{k}\sum_{m=1}^{k} S'_m \\ I_t &= z_t - C - S_m \end{aligned}$$

where $S'_m$ consists of monthly medians of $z_t - C$ when $k = 12$. For a daily time series, $S'_m$ consists of daily medians of $z_t - C$.

(2) For $I_t$, calculate the interquantile range ($IQR$), the distance between the first

and third quantiles. Let an outer fence be 3 times $IQR$ outside either the first or third quantile. Values beyond outer fences are then called far-out values. If a far-out value is present in $I_t$, replace it with the nearest outer fence.

(3) Define a deseasonalized series $D_t$ as

$$D_t = C + I_t$$

(4) Let $C_y$ be the mean of $D_t$ in year $y$. If no $D_t$ values are available during that year, the mean of $D_t$ for the surrounding years is given to $C_y$.

(5) Revise the seasonal and irregular terms as

$$\begin{aligned} S_m &= S"_m - \frac{1}{k}\sum_{m=1}^{k} S"_m \\ I_t &= z_t - S_m - C_y \end{aligned}$$

where $S"_m$ consists of the monthly (or daily) medians of $z_t - C_t$.

(6) Any far-out value of $I_t$ is replaced by the nearest outer fences.

(7) The trend factor $C_t$ is adjusted by adding the mean of the irregular series for the whole span of time. Then, the estimated value for the $m$th month (or day) of year $y$, for which no data were given, is computed as

$$z_{y,m} = C_y + S_m$$

where $z_{y,m}$ denotes the estimated monthly (daily) value for the year $y$ and season $m$.

After the algorithm is completed, the resulting time series is now evenly spaced. All missing values are estimated in the last step.

## 6.4 Illustrative Examples

### 6.4.1 Water Usage in Ontario, 1966-1988

The following example is a time series that reports monthly water usage in Ontario between 1966 and 1988. Figure 6.1 illustrates two plots. The upper panel shows the original time series, and its monthly subseries is shown in the lower panel. From both graphs, it is clear that the water usage data has an upward trend as well as a seasonal pattern, which tends to peak during the summer and decline until the winter.

Assuming that missing values are randomly occurring, we can randomly choose and remove about 10 percent of the observations. The missing values are then estimated by the seasonal adjustment algorithm. In Figure 6.2, the original water usage data is described by the solid black lines and the dashed red lines represent the estimated values. For a closer look, we focus on a certain period between 1969 and 1975 (in the lower panel). The original series is indicated as solid lines. Thin line sections represent actual values that have been removed at the beginning for simulation, while thick line sections are the remaining observations. The estimated values are described by dashed lines. As can be seen, the estimated values are reasonably close to the actual values.

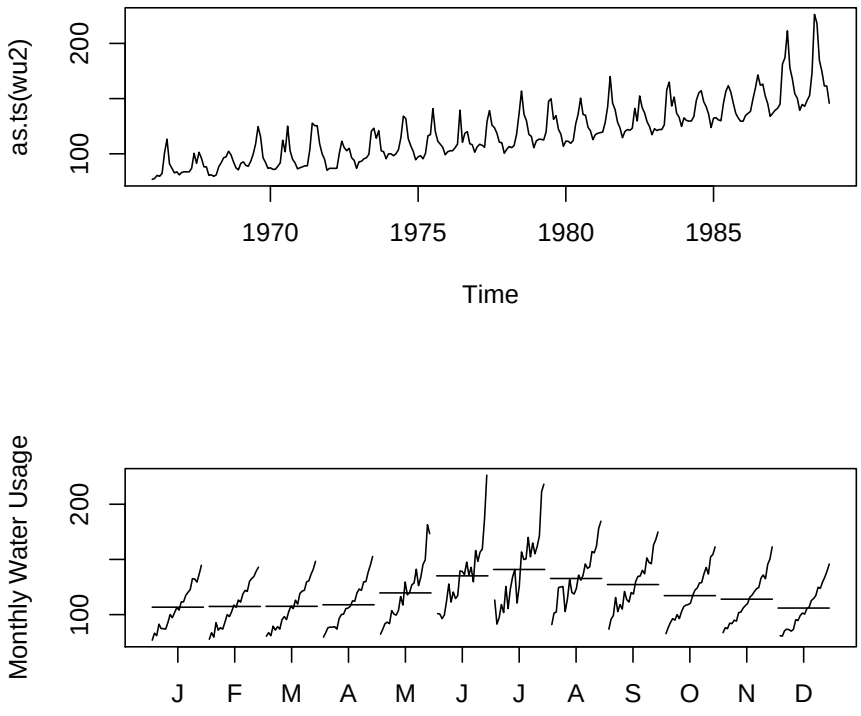


Figure 6.1: Monthly Water Usage data (upper panel) and Seasonal Subseries Plot (lower panel)

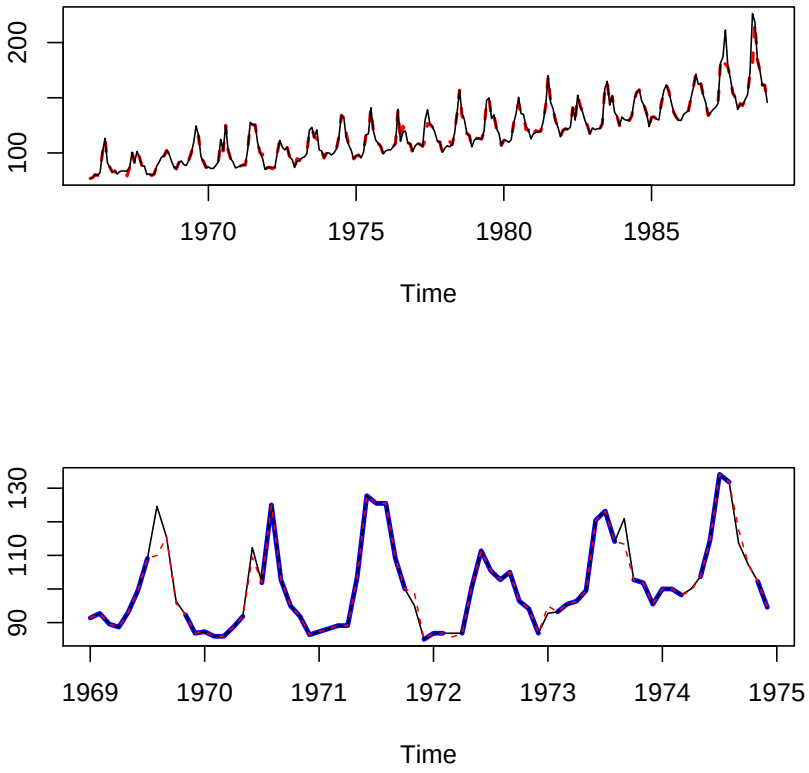


Figure 6.2: Estimation results for Missing data

### 6.4.2 Daily Maximum Temperature in Toronto, 2008-2012

We examine the algorithm for a daily case. We use a daily time series that contains maximum temperatures in Toronto from 2008 to 2012. The data is obtained from the Environment Canada website. As the first step, about ten percent of observations are randomly chosen and removed. We then estimate the missing values using the seasonal adjustment algorithm. Figure 6.3 shows time series revised by the algorithm (dashed lines) as well as the original observed data overlaid (solid lines). For a closer look, results of year 2010 are isolated and shown in figure 6.4. As seen from the figure, estimated values are reasonably close to the observed data.

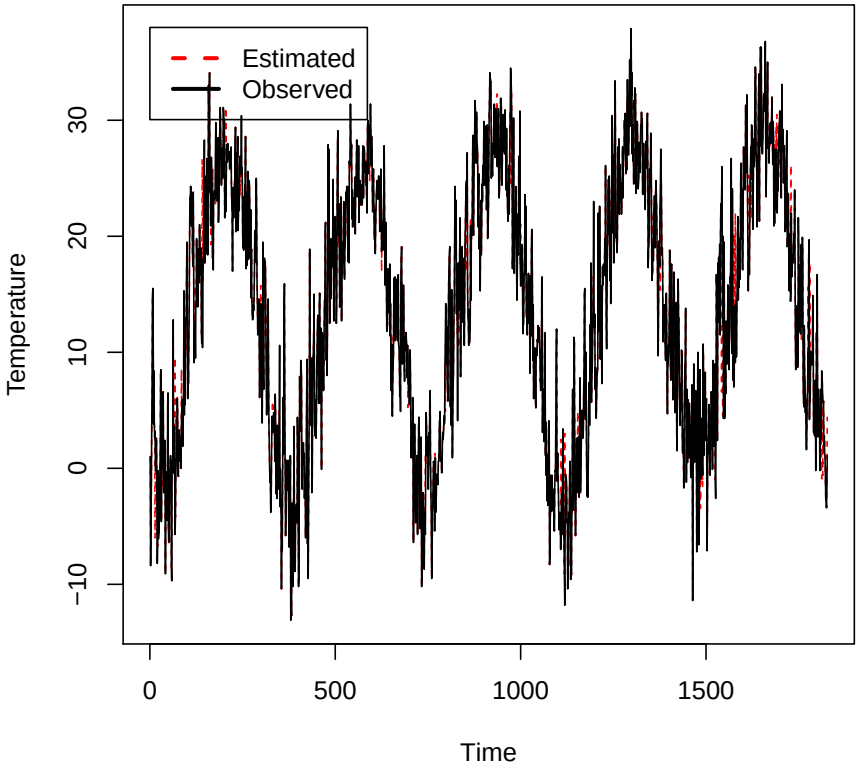


Figure 6.3: Daily Maximum Temperature in Toronto, 2008-2012

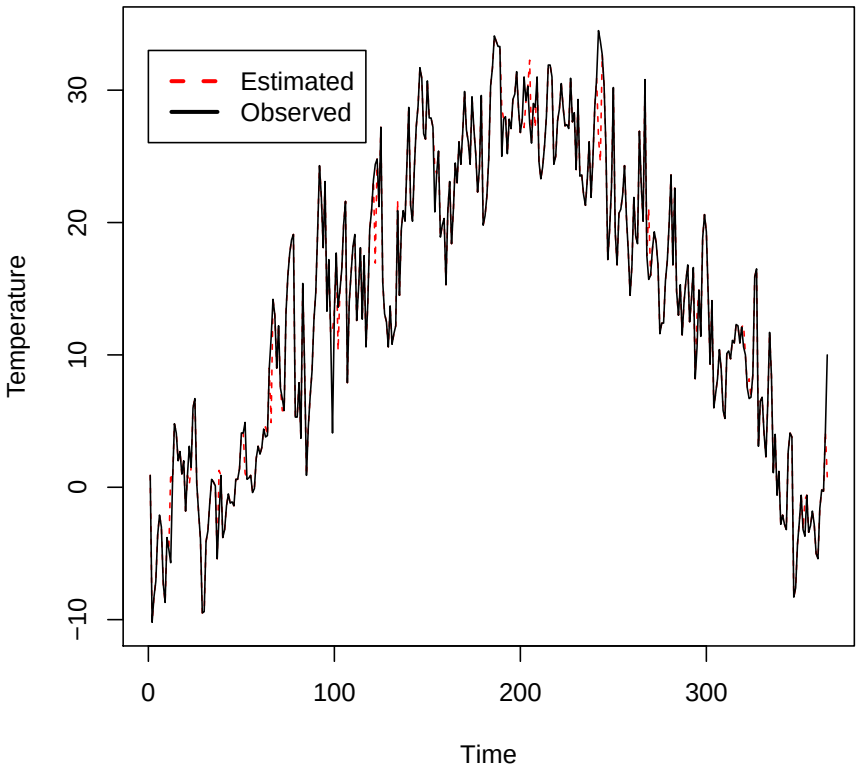


Figure 6.4: Daily Maximum Temperature in Toronto, 2010

# Chapter 7

# Conclusion

Several systematic patterns, such as trends and seasonal terms, may underlie a time series. Analyzing the underlying components is often required, and a decomposition process can be used to do so. In this thesis, we have discussed time series decomposition with nonparametric regression. A decomposition process using loess (Cleveland, 1979) and harmonic regression has been suggested. Based on model selection criteria, we can determine the optimal harmonic regression form for seasonality. We then compared the process with STL (Cleveland et al., 1990). Given that the model forms or parameters for any model that will be used are properly chosen, the method we introduce works as well as STL. Using harmonic regression for a seasonal component, however, is more automatic and less complex than using loess when it comes to building proper model forms. On the other hand, as a solution for an unevenly spaced time series, we introduced the seasonal adjustment algorithm (McLeod et al., 1983) and revised the algorithm to handle daily time series data. The revised algorithm works fairly well .

In Chapter Two, we reviewed loess (Cleveland, 1979) and harmonic regression, which are widely used in time series analysis. Loess is useful to fit a model for either a long-term or seasonal effect, while we can employ harmonic regression for a seasonal pattern that has a one-year period. Since harmonic regression uses sinusoids as predictors, we have also reviewed several important features, such as aliasing, orthogonality, and Fourier frequency.

In considering nonparametric model estimation, model selection is an important aspect of this. In Chapter Three, we introduced several criteria, such as $AIC$ (Akaike, 1974) and $BIC$ (Schwarz, 1978). Based on these criteria, we can determine suitable models for fitting time series components.

In Chapter Four, we examined the deseasonalization process using harmonic regression. Scaling is preferred if the variations are seasonal. The optimal model can be determined using model selection methods. Seasonal subseries plots (Cleveland, 1993) and multiple boxplots are used to visually confirm the existence of a seasonality.

In Chapter Five, we started with the additive model assumption. We then introduced STL as a widely used decomposition method. We tried another method, using loess and harmonic regression for long-term and seasonal components, respectively. The backfitting algorithm (Friedman and Stuetzle, 1981) enables the components to be separated without being affected by any other components.

In Chapter Six, we considered the missing data problem in a seasonal time series. The main idea was that we would first measure the repeating seasonal effects and use them to estimate missing observations. We introduced the seasonal adjustment algorithm, which works well for monthly time series. We then extend the algorithm to cover daily time series data.

# REFERENCES


H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

P. Bloomfield. *Fourier Analysis of Time Series: An Introduction*. Wiley, New York, 1976.

R. Chandler and M. Scott. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. Wiley, 2011.

R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6: 3–73, 1990.

W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

W. S. Cleveland. *Visualizing Data*. Hobart Press, 1993.

P. Craven and G. Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics*, 31:377–403, 1979.

Environment Canada. National Climate Data and Information Archive. URL `http://www.climate.weatheroffice.ec.gc.ca/climateData`. [Retrieved March 2013].

J. H. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American Statistical Association*, 76:817–823, 1981.

K. W. Hipel and A. I. McLeod. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam, 1994. URL `http://www.stats.uwo.ca/faculty/aim/1994Book/`.

J. L. Klein. *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*. Cambridge University Press, 1997.

A. McLeod and H. Gweon. Optimal deseasonalization for geophysical time series. *Journal of Environmental Statistics*, 4(11), 2013. URL `http://www.jenvstat.org/`.

A. McLeod, K. Hipel, and F. Camacho. Trend assessment of water quality time series. *Water Resources Bulletin*, 19:537–547, 1983.

A. I. McLeod. Parsimony, model adequacy and periodic correlation in forecasting time series. *International Statistical Review*, 61:387–393, 1993.

A. I. McLeod and H. Gweon. *Optimal deseasonalization for geophysical time series using AR fitting*, 2012. URL `http://cran.r-project.org/web/packages/deseasonalize`.

A. I. McLeod, Y. Zhang, and C. Xu. *FitAR: Subset AR Model Fitting*, 2011. URL `http://CRAN.R-project.org/package=FitAR`. R package version 1.92. Accessed 22-December-2011.

M. Mudelsee. *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. Springer, 2010.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.

## Curriculum Vitae

| | |
|---|---|
| **Name:** | Hyukjun Gweon |
| **Post-Secondary Education and Degrees:** | University of Western Ontario<br>London, ON<br>2011 - 2013 M.Sc. |
| | Korea University<br>Seoul, Korea<br>2006 - 2010 B.Sc. |
| **Honours and Awards:** | Western Graduate Research Scholarship<br>2011-2013 |
| | National Science and Technology Scholarship<br>2006 - 2010 |
| **Related Work Experience:** | Teaching Assistant<br>University of Western Ontario<br>2011- 2013 |