The long-memory of air quality data

Jeh-Nan Pan and Su-Tsu Chen

Outline

- Introduction
- Air Quality and Health
- Long-memory Data
- ARFIMA Model for Long Memory Data
- Conclusion

Introduction

- Pollutants in the air : carbon monoxide, hydrocarbons, nitrogen oxides, sulfur oxides, and particles.
- Motor vehicles, factories, and fuel burning are responsible for emitting most of these harmful substances into the air.
- Environmental performance of processes is notoriously hard to control.

Introduction

Regulations to help air quality:

1. The first version of Air Pollution Control Act in Taiwan is promulgated by presidential order in 1975.

2. The Standard of Air Quality sets specifications for all the important pollution.

Introduction

- total suspended particulate matter (TSP) : tiny airborne particles or aerosols less than 100 micrometers.

- PM10 : particles between 2.5 and 10 micrometers.

- PM2.5 : particles smaller than 2.5 micrometers.

Introduction

- PM10 is the most important pollutant that deteriorates the air quality in Taiwan(Chan and Hwang,1996).
- The main goal of this paper is to identify an appropriate time series model for PM10 data with long memory.

Air quality data

- Usually autocorrelated.
- Pan and Chen (2004), PM10 and O3 of air quality in the Taipei city follow ARIMA models.
  - **PM10~ARIMA(1, 1, 1)**
  - **O3 ~ARIMA(2, 1, 1)**

Air quality and health

- Among all the harmful pollutants, there are particles. These particles can cause severe health effects, especially to respiratory diseases.

Air quality and health

- A substantial literature of epidemiologic studies demonstrates an association between mortality and exposure to particulate matter in the air (Welty and Zeger, 2005).
- Much of their evidence comes from time-series studies by comparing daily variations in mortality with daily variations in particulate matter while controlling for confounding factors that also vary over time, such as weather and season.

Air quality and health

- Schwartz and Marcus (1990) conclude that particulate matters are strongly associated with mortality rates in London.
- Téllez-Rojo et al. (2000, 2001) discuss relationship between daily respiratory mortality and PM10 pollution in Mexico City. Their study confirms that there is an important impact of PM10 on respiratory morbidity among elderly subjects.

Air quality and health

- In the previous papers, a regression model for these data and other variables are considered.
- Time series model of PM10 data of air quality is considered in this paper.

ARIMA model

Definition 2.1 Let Z_t be a process such that

$$\Phi_p(B)(1-B)^d Z_t = \theta_0 + \Theta_q(B)a_t \qquad (1)$$

where d is positive integer or zero, B is the backshift operator ($BZ_t=Z_{t-1}$), ε_t is white noise, and

$\Phi_p(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$ and

$\Theta_q(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$ share no common factors.

Then Z_t is called an ARIMA(p, d, q) process.

ARFIMA model

- ARFIMA models was proposed by Granger and Joyeux (1980) and Hosking (1981) to fit long-memory data.

- Similar to Equation (1) except

  $d \in (-0.5, 0.5)$

Long-memory

- Long memory :
  - $\lim_{n \to \infty} \sum_{h=-n}^{n} |\rho(h)|$ is nonfinite
  - The spectral density $f(\omega)$ will be unbounded at low frequencies.

Long- Memory

- Long memory:

  its ACF has a hyperbolic decay,

  $$\rho(h) \sim C\, h^{2d-1} \quad \text{as } h \to \infty.$$

  $$(C \neq 0\ ,\ d < 0.5)$$

- Short memory:

  its ACF is geometrically bounded

  $$|\rho(h)| \leq C\, r^{|h|}$$

  $$(\ C > 0,\ 0 < r < 1)$$

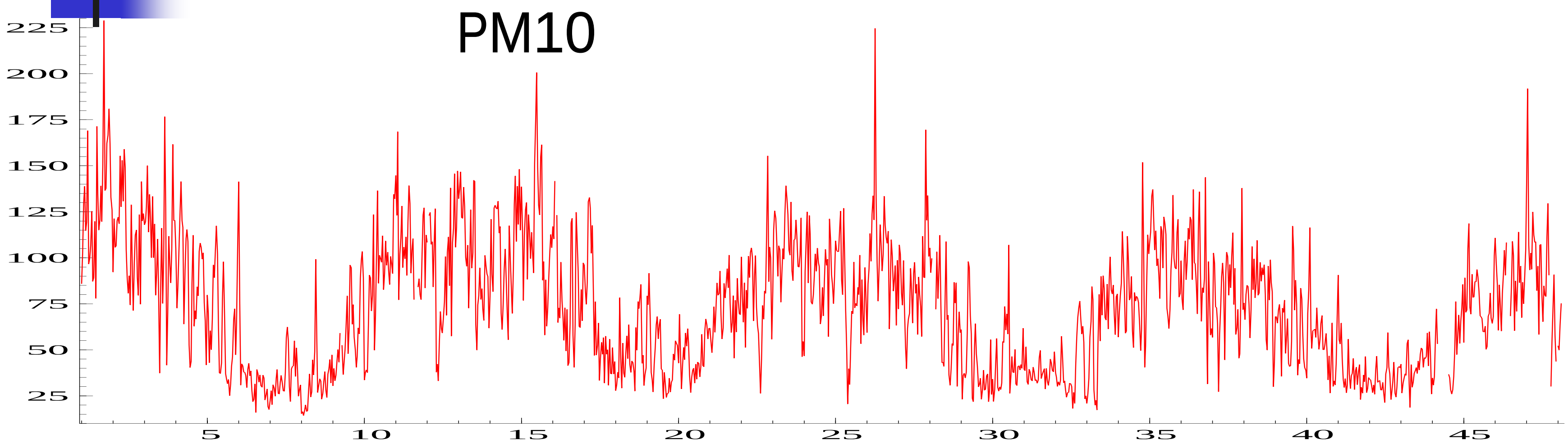
Empirical example 1

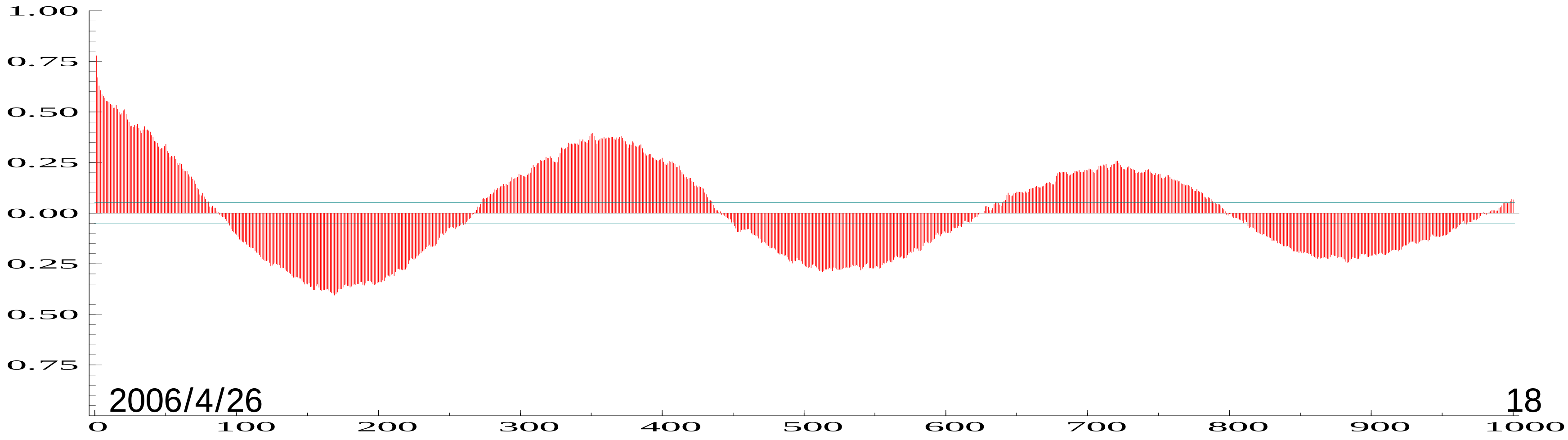
- Nantsz station is a surveillance stations located in the southern Taiwan where is known for a long history of public protest for pollution.

Empirical example 1

- The hourly air quality data of PM10 collected by Nantsz station between 1999 and 2002 is discussed.
- Missing data are replaced with linear interpolations.

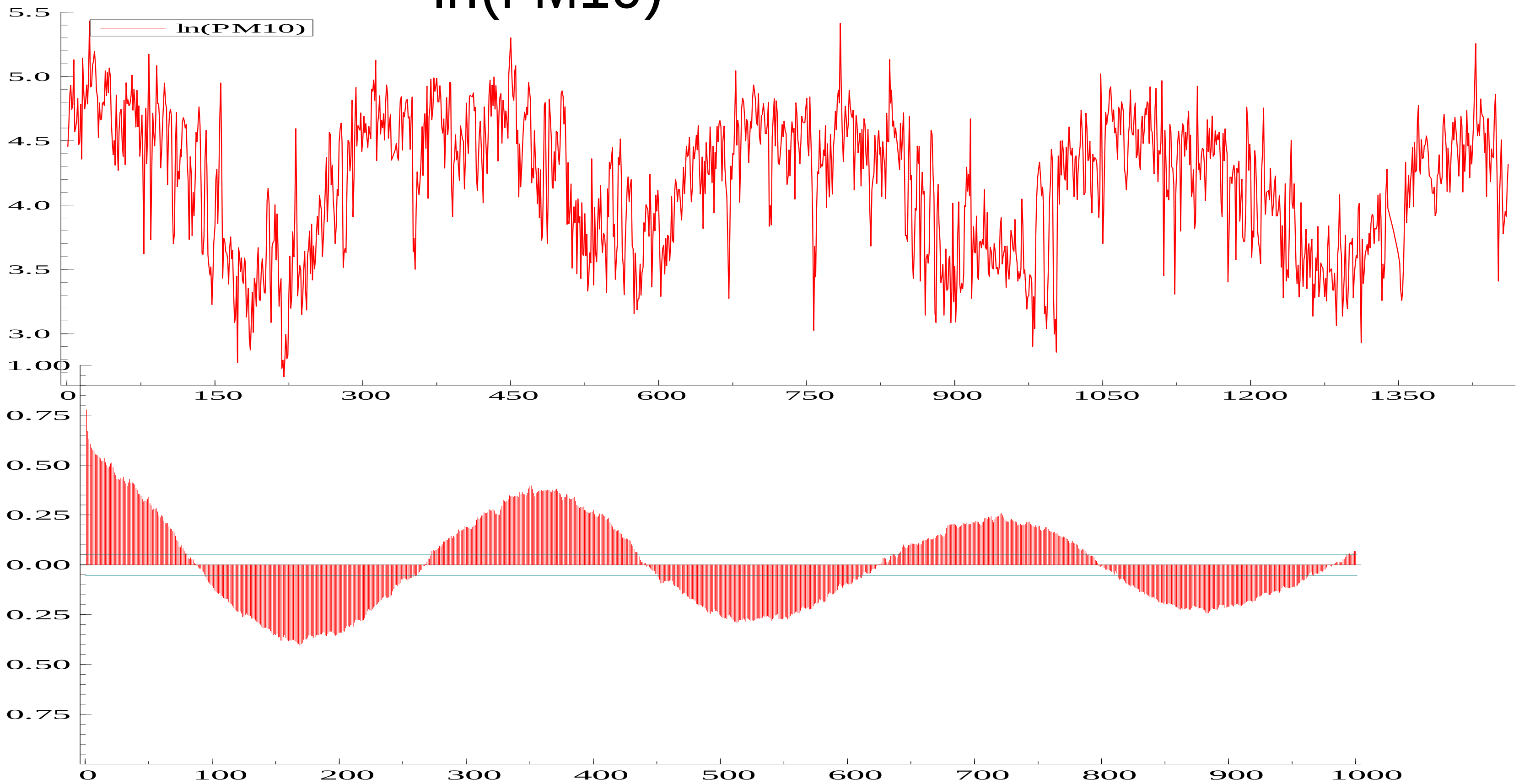
Empirical example 1





Empirical example 1

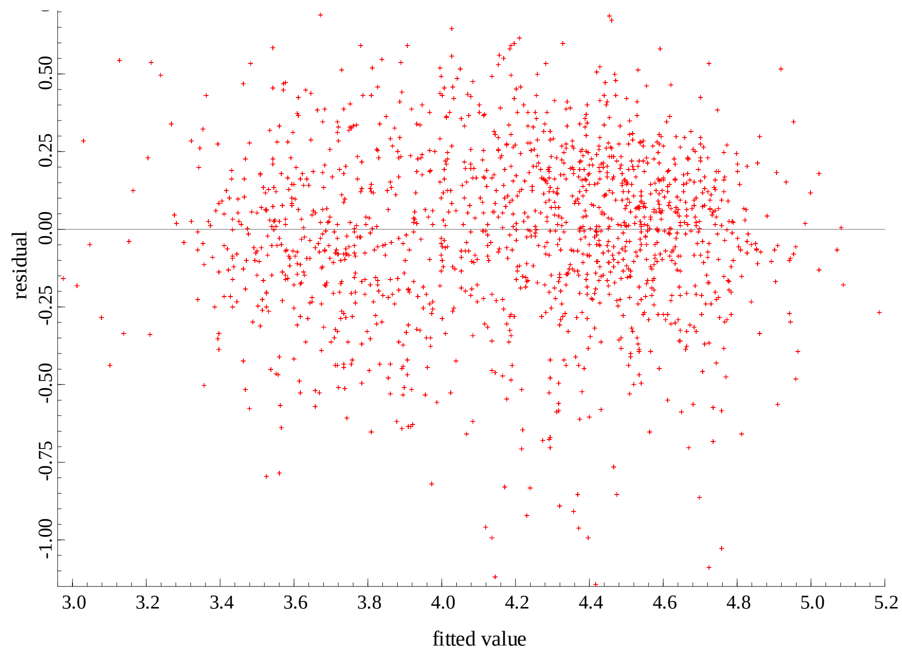
ln(PM10)

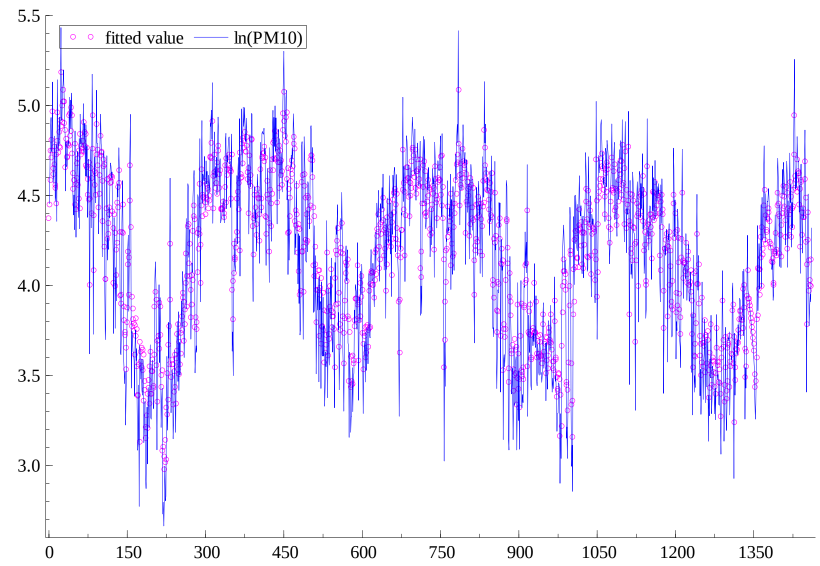


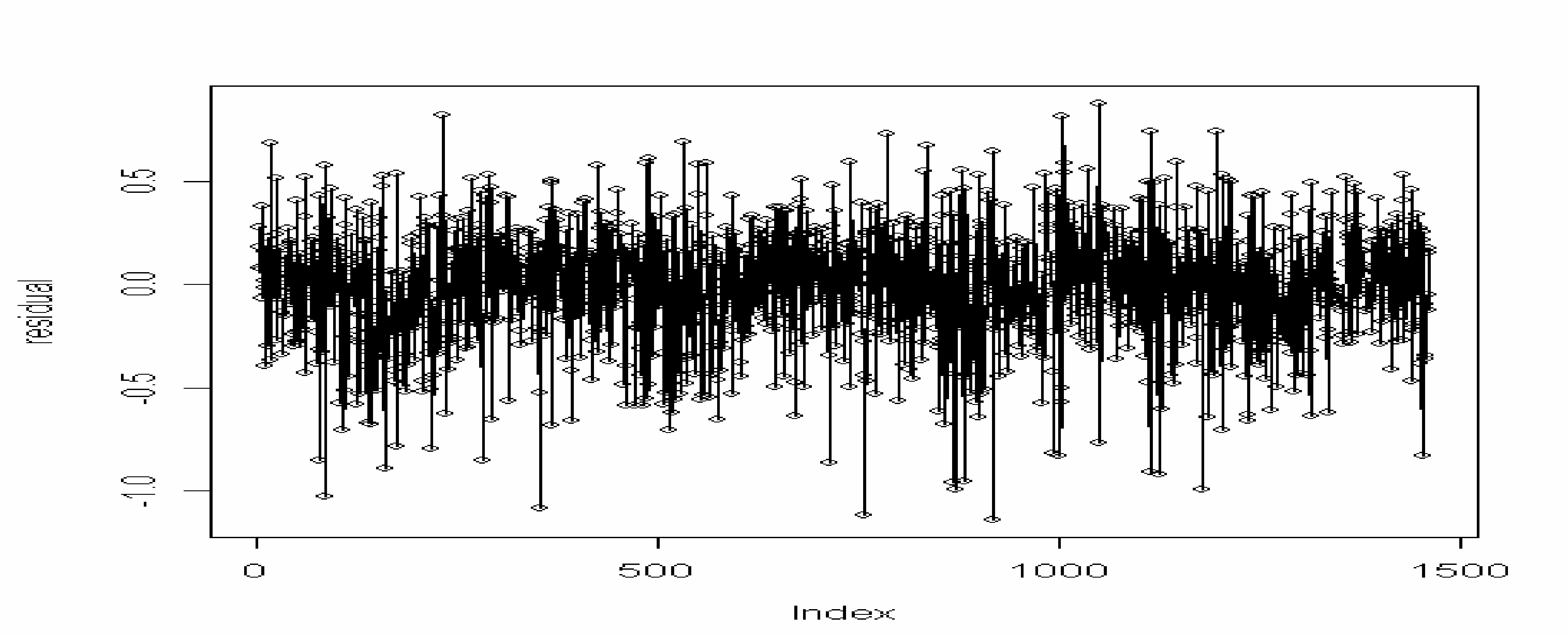
Empirical example 1

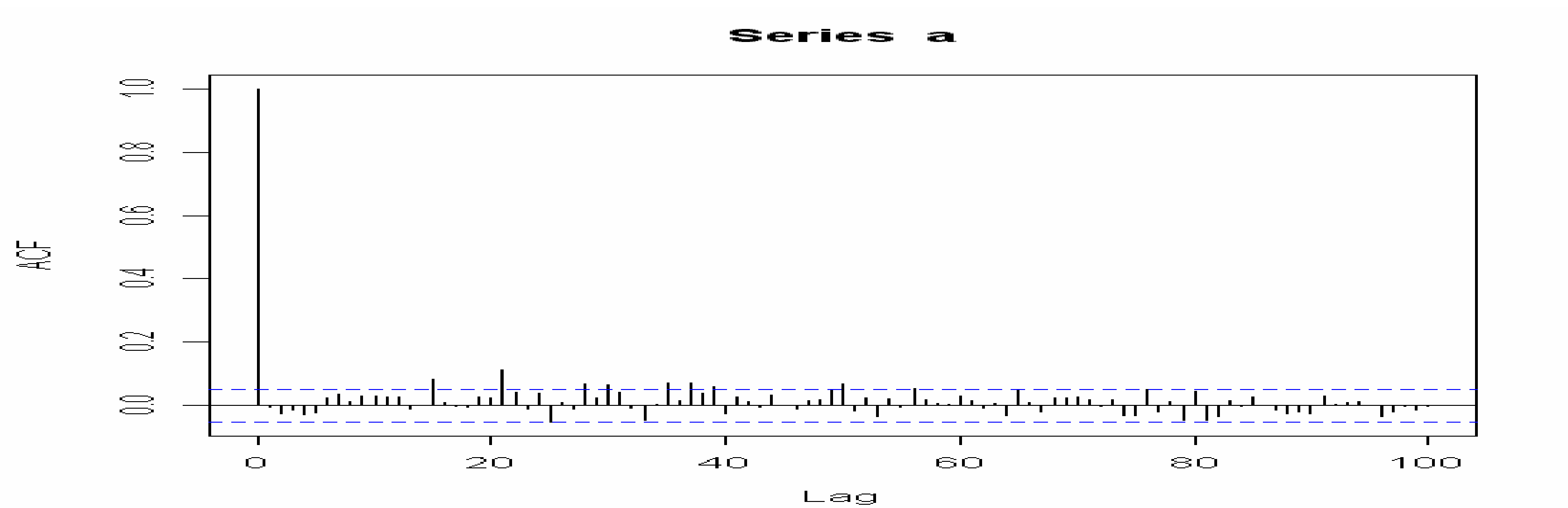
- The most appropriate model :

$$(1-B)^{0.47}(\ln(PM10_t)-4.23)=(1-0.16B)\quad \varepsilon_t \qquad (2)$$


residual
fitted value
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
5.0
5.2


fitted value
ln(PM10)
5.5
5.0
4.5
4.0
3.5
3.0
0
150
300
450
600
750
900
1050
1200
1350


residual
0.5
0.0
-0.5
-1.0
0
500
1000
1500
Index


Series a
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Lag

Empirical example 1

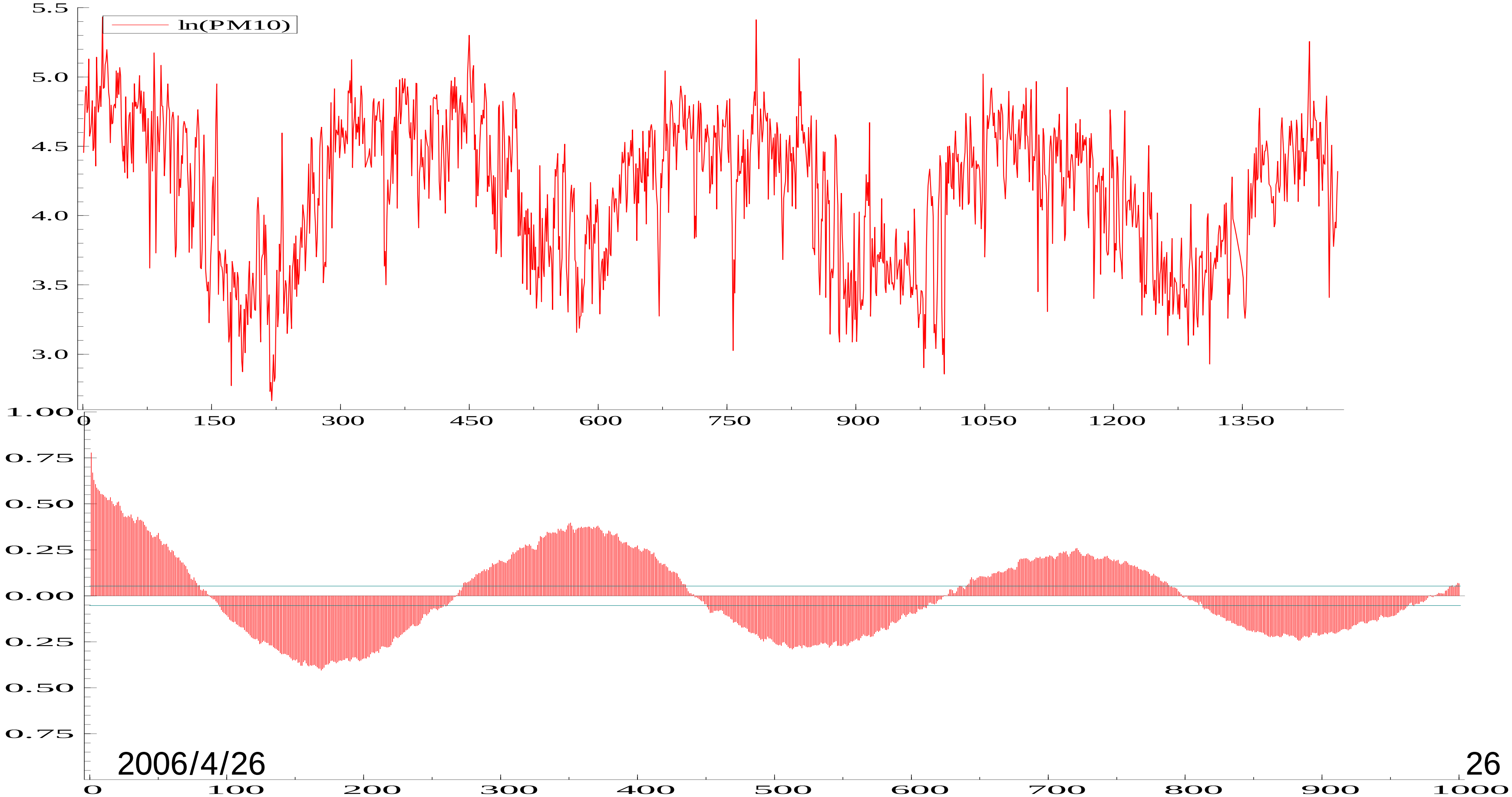
- If the long-term autocorrelations were ignore and ARIMA models are used to model the data, the residuals may be autocorrelated or have variance that is not constant, therefore no model passes the diagnostic of the residuals completely.

- The ACF of autocorrelated data should be calculated to large lag in order to detect the property of long memory.

Empirical example 2

- The daily PM10 data collected at Tsoying station between 1999 and 2002 is used to compare with Nantsz’s data.

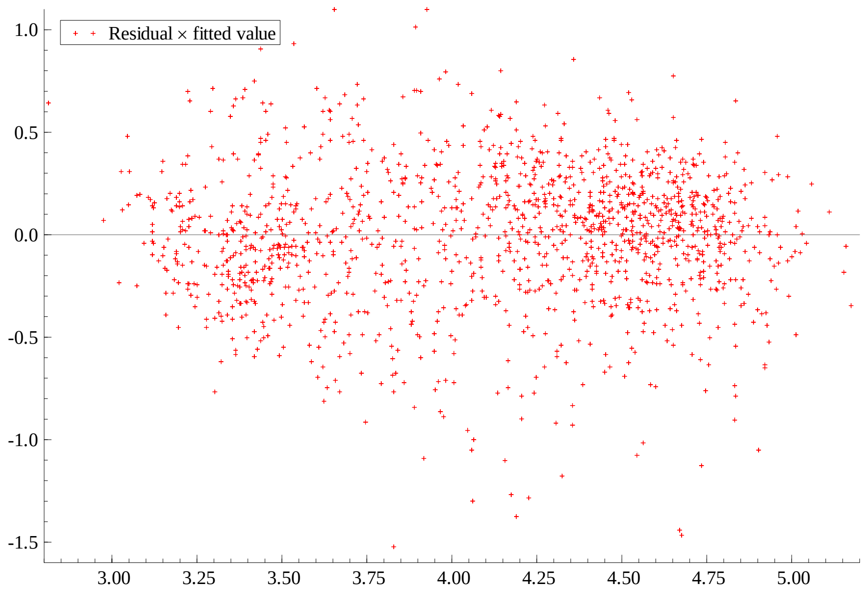
Empirical example 2

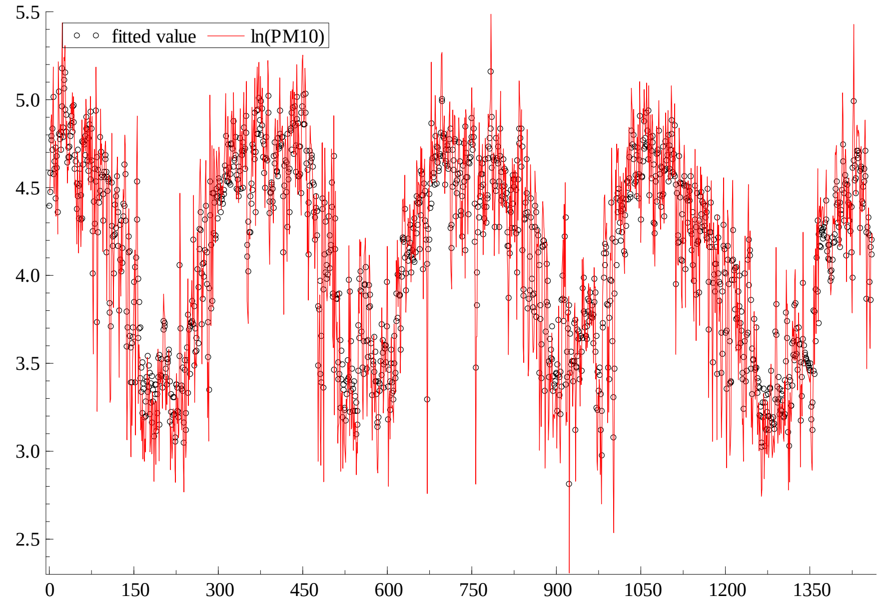


- The appropriate model :

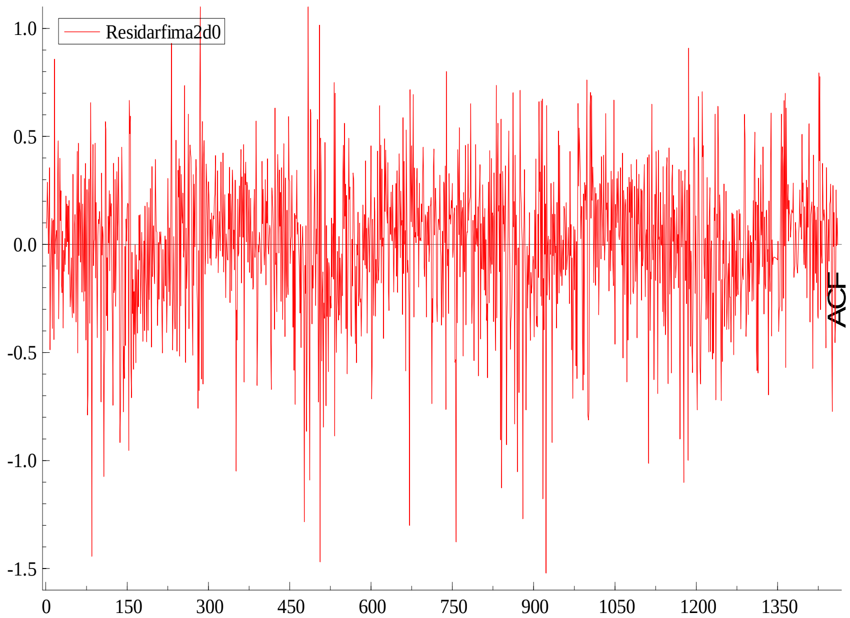
$$(1-0.16B+0.11B^2)(1-B)^{0.49}(\ln(PM10_t)-4.17)=\varepsilon_t$$

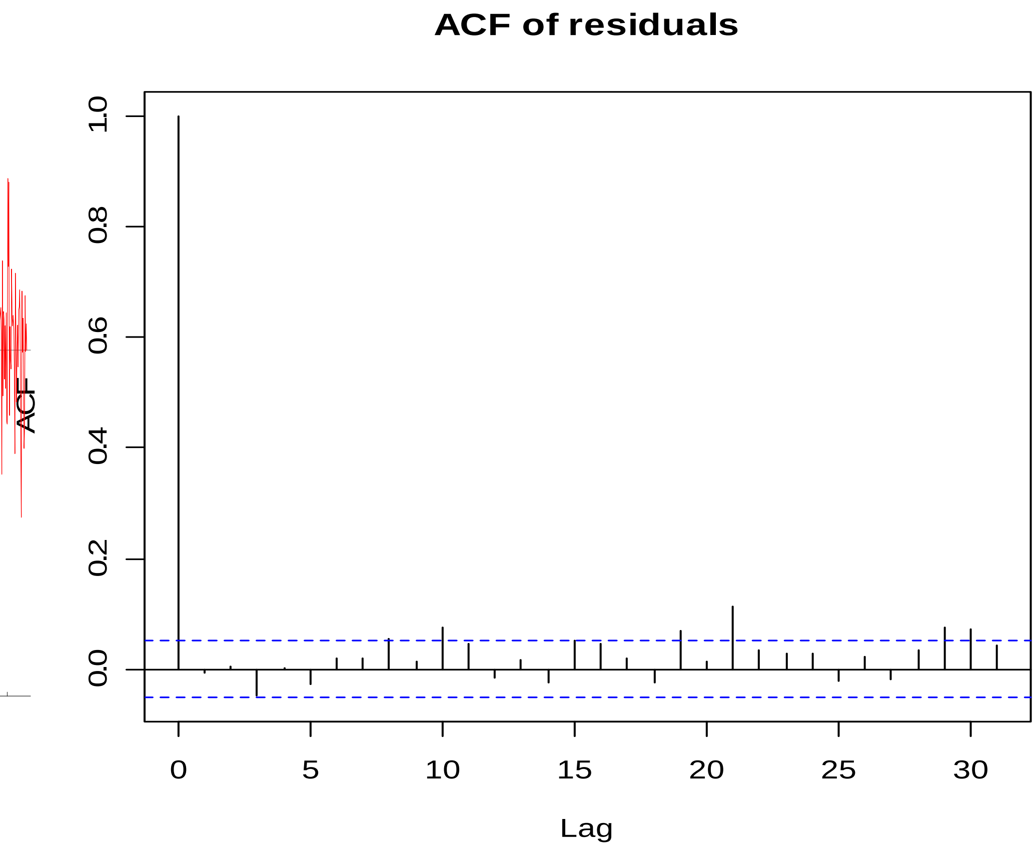
Empirical example 2





Empirical example 2





Empirical example 2

- no ARIMA model has residuals that could pass the diagnostic of the residuals.

Conclusion

- To detect the property of long memory, the ACF with large lag and the spectral density are helpful tools.
- To model long memory process, several models are selected as candidates which would be chosen according to some model selection criterion, which is the Akaike Information Criterion in this study.

Conclusion

- After model-fitting, a diagnostic of the residuals should be performed.
- When the property of long memory is ignored, ARIMA models are usually used to fit long memory data, but diagnostic checking of the adequacy of fitted models would show the model is not suitable.

The end