Introduction to Statistical Analysis of Time Series

Richard A. Davis
Department of Statistics

Outline

- Modeling objectives in time series
- General features of ecological/environmental time series
- Components of a time series
- Frequency domain analysis-the spectrum
- Estimating and removing seasonal components
- Other cyclical components
- Putting it all together

Time Series: A collection of observations x_t, each one being recorded at time t. (Time could be discrete, $t = 1,2,3,\ldots$, or continuous $t > 0$.)

Objective of Time Series Analaysis

- Data compression

  -provide compact description of the data.

- Explanatory

  -seasonal factors

  -relationships with other variables (temperature, humidity, pollution, etc)

- Signal processing

  -extracting a signal in the presence of noise

- Prediction

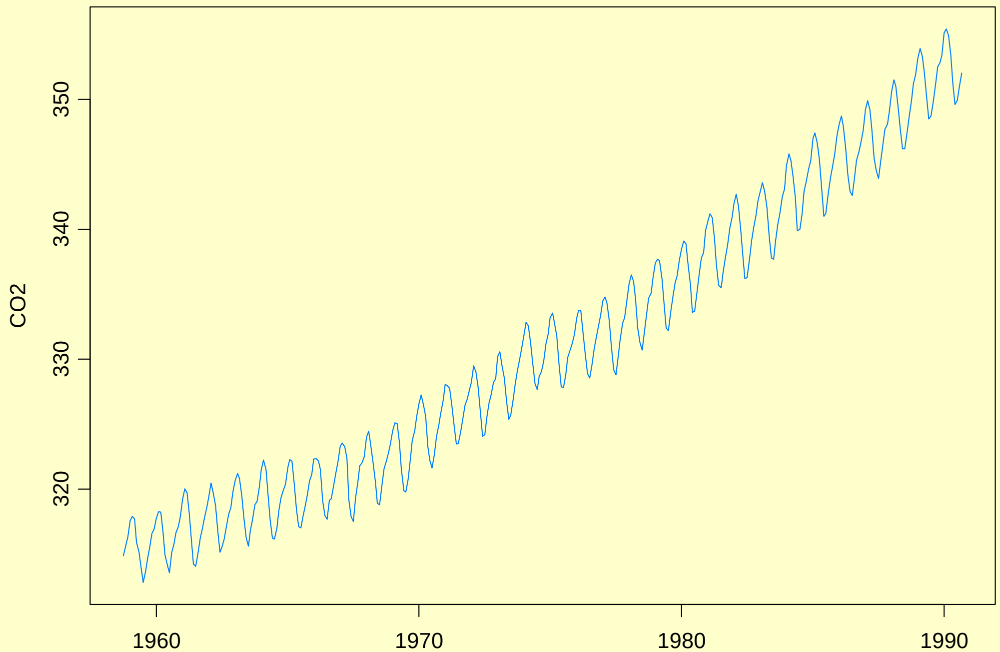
  -use the model to predict future values of the time series

Colorado State University
Knowledge to Go Places

General features of ecological/environmental time series

Examples.

1. Mauna Loa (CO_2, Oct `58-Sept `90)

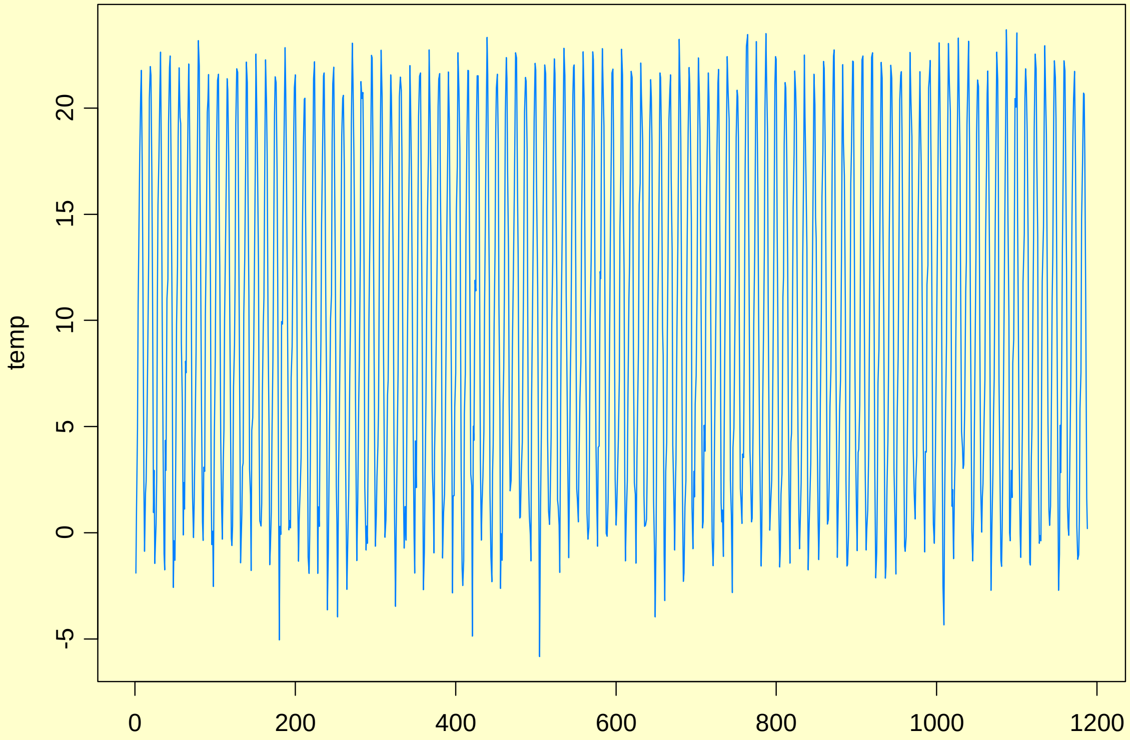


Features

- increasing trend (linear, quadratic?)
- seasonal (monthly) effect.

PRIMES

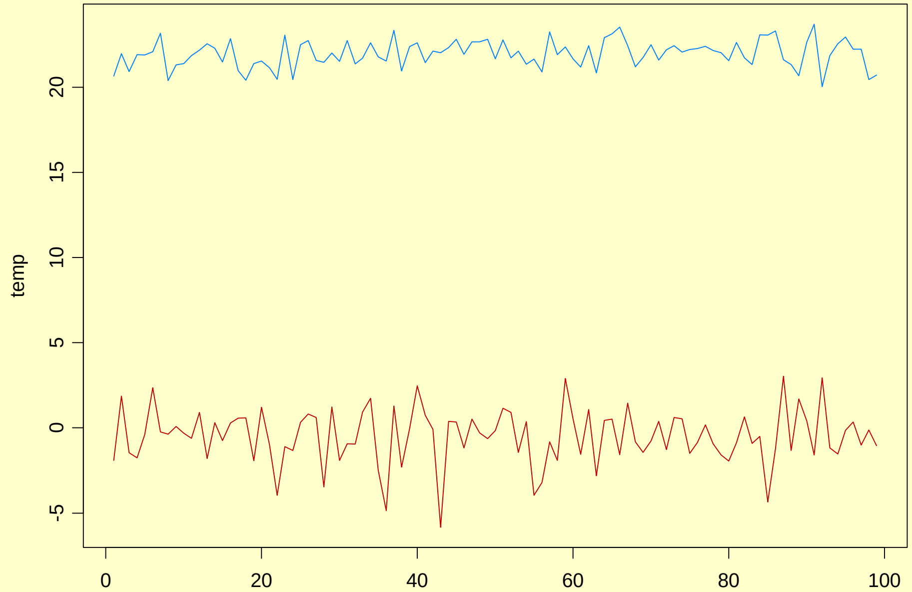
2. Ave-max monthly temp (vegetation=tundra, 1895-1993)



Features

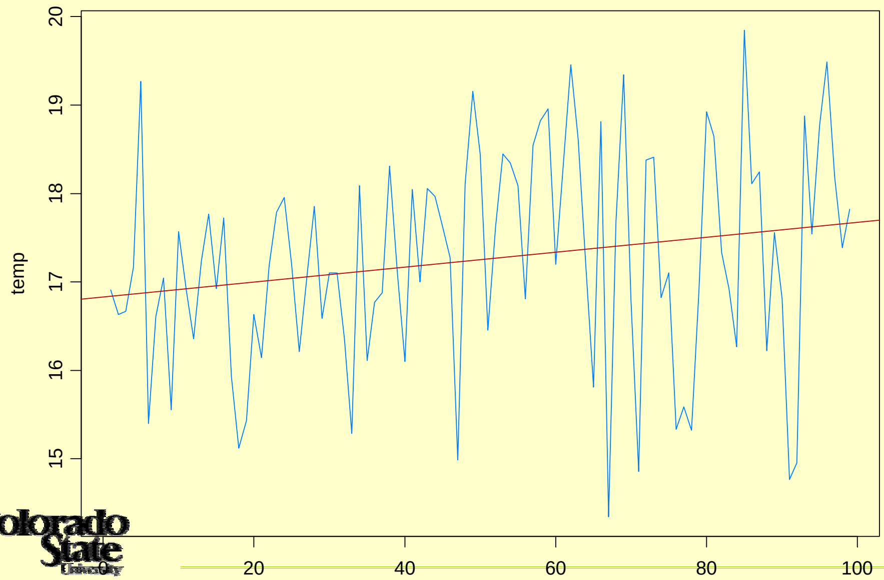
- seasonal (monthly effect)
- more variability in Jan than in July

Colorado State University
Knowledge to Go Places



July: mean = 21.95, var = .6305

Jan : mean = -.486, var =2.637



Sept : mean = 17.25, var =1.466

Line: 16.83 + .00845 t

Components of a time series

Classical decomposition

$$X_t = m_t + s_t + Y_t$$

- m_t = trend component (slowly changing in time)
- s_t = seasonal component (known period d=24(hourly), d=12(monthly))
- Y_t = random noise component (might contain irregular cyclical components of unknown frequency + other stuff).

Go to ITSM Demo

Estimation of the components.

$$X_t = m_t + s_t + Y_t$$

Trend m_t

- filtering. E.g., for monthly data use

$$\hat{m}_t = (.5x_{t-6} + x_{t-5} + \cdots + x_{t+5} + .5x_{t+6})/12$$

- polynomial fitting

$$\hat{m}_t = a_0 + a_1 t + \cdots + a_k t^k$$

Colorado State University
Knowledge to Go Places

PRIMES

Estimation of the components (cont).

$$X_t = m_t + s_t + Y_t$$

Seasonal s_t

- Use seasonal (monthly) averages after detrending.

(standardize so that s_t sums to 0 across the year.

$$\hat{s}_t = (x_t + x_{t+12} + x_{t+24} \cdots)/N, \; N = \text{number of years}$$

- harmonic components fit to the time series using least squares.

$$\hat{s}_t = A\cos(\frac{2\pi}{12}t) + B\sin(\frac{2\pi}{12}t)$$

Irregular component Y_t

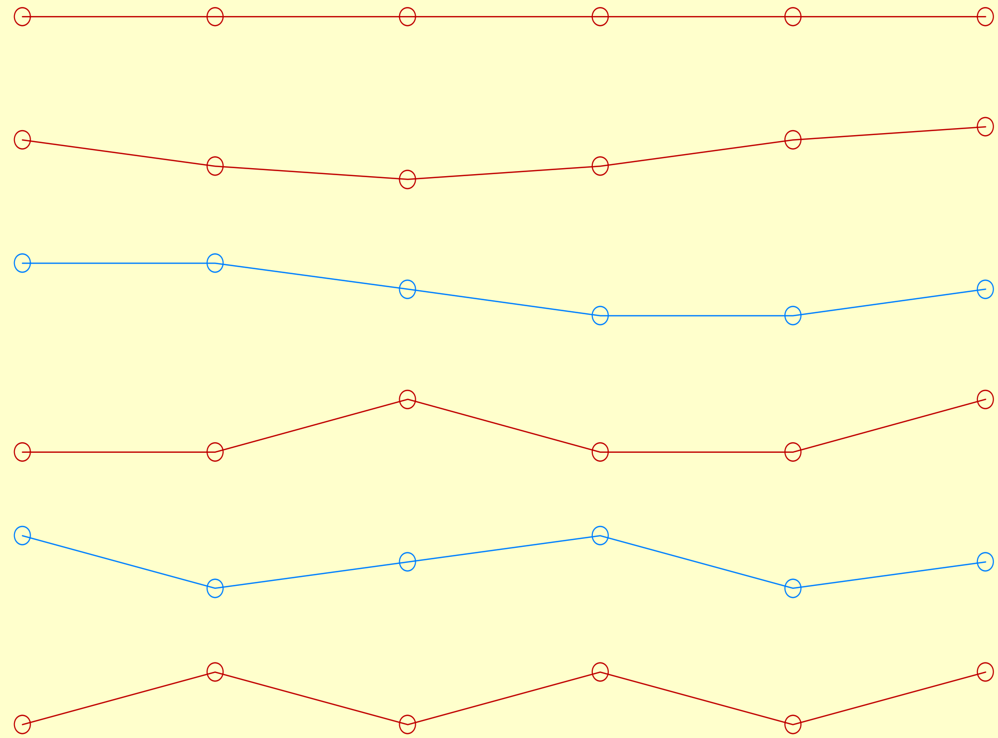
$$\hat{Y}_t = X_t - \hat{m}_t - \hat{s}_t$$

Colorado State University
Knowledge to Go Places

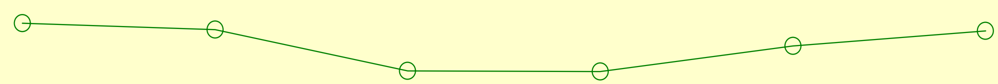
The spectrum and frequency domain analysis

Toy example. (n=6)

$\mathbf{c}_0 = (\cos(\frac{2\pi 0}{6}), \ldots, \cos(\frac{2\pi 0}{6} 6))' / \text{sqrt}(6)$
$\mathbf{c}_1 = (\cos(\frac{2\pi}{6}), \ldots, \cos(\frac{2\pi}{6} 6))' / \text{sqrt}(3)$
$\mathbf{s}_1 = (\sin(\frac{2\pi}{6}), \ldots, \sin(\frac{2\pi}{6} 6))' / \text{sqrt}(3)$
$\mathbf{c}_2 = (\cos(\frac{2\pi 2}{6}), \ldots, \cos(\frac{2\pi 2}{6} 6))' / \text{sqrt}(3)$
$\mathbf{s}_2 = (\sin(\frac{2\pi 2}{6}), \ldots, \sin(\frac{2\pi 2}{6} 6))' / \text{sqrt}(3)$
$\mathbf{c}_3 = (\cos(\frac{2\pi}{2}), \ldots, \cos(\frac{2\pi}{2} 6))' / \text{sqrt}(6$



$\mathbf{X}$=(4.24, 3.26, -3.14, -3.24, 0.739, 3.04)’ = $2\mathbf{c}_0+5(\mathbf{c}_1+\mathbf{s}_1)-1.5(\mathbf{c}_2+\mathbf{s}_2)+.5\mathbf{c}_3$



PRIMES

Fact: Any vector of 6 numbers, $\mathbf{x} = (x_1, \ldots, x_6)'$ can be written as a linear combination of the vectors $\mathbf{c}_0$, $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{c}_3$.

More generally, any time series $\mathbf{x} = (x_1, \ldots, x_n)'$ of length n (assume n is odd) can be written as a linear combination of the basis (orthonormal) vectors $\mathbf{c}_0$, $\mathbf{c}_1$, $\mathbf{c}_2$, …, $\mathbf{c}_{[n/2]}$, $\mathbf{s}_1$, $\mathbf{s}_2$, …, $\mathbf{s}_{[n/2]}$. That is,

$$\mathbf{x} = a_0\mathbf{c}_0 + a_1\mathbf{c}_1 + b_1\mathbf{s}_1 + \cdots a_m\mathbf{c}_m + b_m\mathbf{s}_m, \ \ m = [n/2]$$

$$\mathbf{c}_0 = \left(\frac{1}{n}\right)^{1/2}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}, \ \mathbf{c}_j = \left(\frac{2}{n}\right)^{1/2}\begin{bmatrix}\cos(\omega_j)\\\cos(2\omega_j)\\\vdots\\\cos(n\omega_j)\end{bmatrix}, \ \mathbf{s}_j = \left(\frac{2}{n}\right)^{1/2}\begin{bmatrix}\sin(\omega_j)\\\sin(2\omega_j)\\\vdots\\\sin(n\omega_j)\end{bmatrix}$$

$$\mathbf{x} = a_0\mathbf{c}_0 + a_1\mathbf{c}_1 + b_1\mathbf{s}_1 + \cdots a_m\mathbf{c}_m + b_m\mathbf{s}_m, \;\; m = [n/2]$$

Properties:

1. The set of coefficients {a_0, a_1, b_1, ... } is called the **discrete Fourier transform**

$$a_0 = (\mathbf{x}, \mathbf{c}_0) = \frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t$$

$$a_j = (\mathbf{x}, \mathbf{c}_j) = \frac{2^{1/2}}{n^{1/2}} \sum_{t=1}^{n} x_t \cos(\omega_j t)$$

$$b_j = (\mathbf{x}, \mathbf{s}_j) = \frac{2^{1/2}}{n^{1/2}} \sum_{t=1}^{n} x_t \sin(\omega_j t)$$

2. Sum of squares.

$$\sum_{t=1}^{n} x_t^2 = a_0^2 + \sum_{j=1}^{m} \left(a_j^2 + b_j^2\right)$$

3. ANOVA (analysis of variance table)

Source	DF	Sum of Squares	Periodgram
ω_0	1	a_0^2	$I(\omega_0)$
$\omega_1 = 2\pi/n$	2	$a_1^2 + b_1^2$	$2\,I(\omega_1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\omega_m = 2\pi m/n$	2	$a_m^2 + b_m^2$	$2\,I(\omega_m)$
	n	$\sum_t x_t^2$	

Applied to toy example

Source	DF	Sum of Squares
$\omega_0=0$ (period 0)	1	$a_0^2 = 4.0$
$\omega_1=2\pi/6$ (period 6)	2	$a_1^2 + b_1^2 = 50.0$
$\omega_2=2\pi 2/6$ (period 3)	2	$a_2^2 + b_2^2 = 4.5$
$\omega_3=2\pi 3/6$ (period 2)	1	$a_3^2 = 0.25$
	6	$\sum_t x_t^2 = 58.75$

Test that period 6 is significant

H_0: $X_t = \mu + \varepsilon_t$, $\{\varepsilon_t\}$ ~ independent noise

H_1: $X_t = \mu + A \cos(t2\pi/6) + B \sin(t2\pi/6) + \varepsilon_t$

Test Statistic: $(n-3)I(\omega_1)/(\Sigma_t x_t^2 - I(0) - 2I(\omega_1)) \sim F(2,n-3)$

$(6-3)(50/2)/(58.75-4-50)=15.79 \Rightarrow$ p-value = .003

The spectrum and frequency domain analysis

Ex. Sinusoid with period 12.

$$x_t = 5\cos\left(\frac{2\pi}{12}t\right) + 3\sin\left(\frac{2\pi}{12}t\right), \quad t = 1,2,\ldots,120.$$

Ex. Sinusoid with periods 4 and 12.

Ex. Mauna Loa

ITSM DEMO

Differencing at lag 12

Sometimes, a seasonal component with period 12 in the time series can be removed by differencing at lag 12. That is the differenced series is

$$y_t = x_t - x_{t-12}$$

Now suppose x_t is the sinusoid with period 12 + noise.

$$x_t = 5\cos(\frac{2\pi}{12}t) + 3\sin(\frac{2\pi}{12}t) + \varepsilon_t, \quad t = 1,2,\ldots,120.$$

Then

$$y_t = x_t - x_{t-12} = \varepsilon_t - \varepsilon_{t-12}$$

which has correlation at lag 12.

Other cyclical components; searching for hidden cycles

Ex. Sunspots.

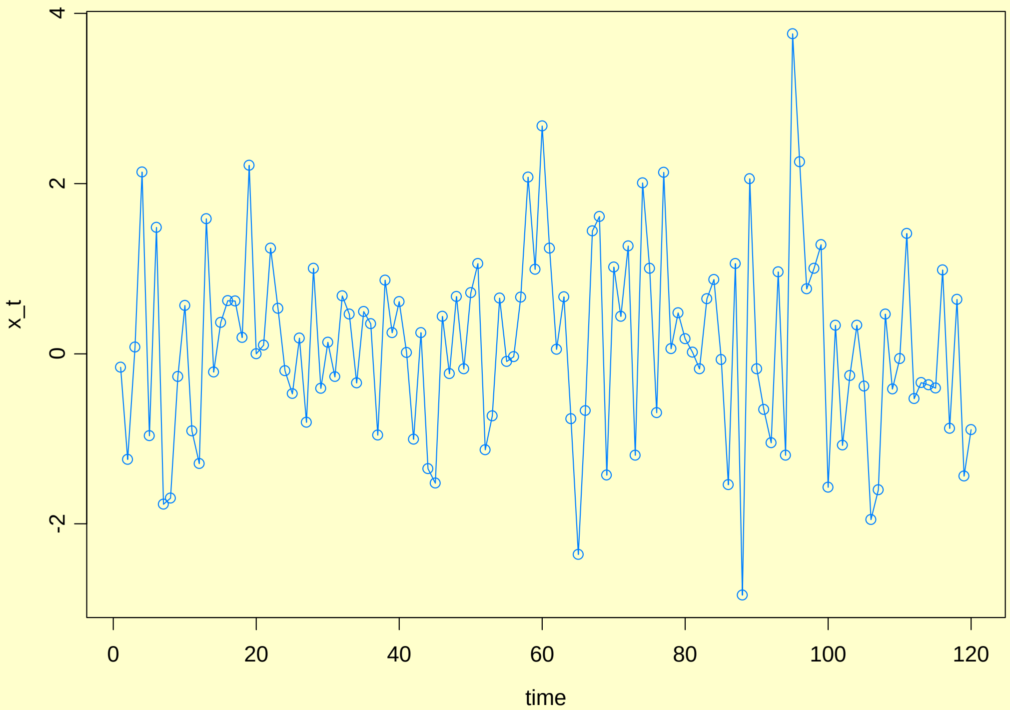
- period ~ $2\pi/.62684=10.02$ years
- Fisher's test $\Rightarrow$ significance

What model should we use?

ITSM DEMO

Noise.

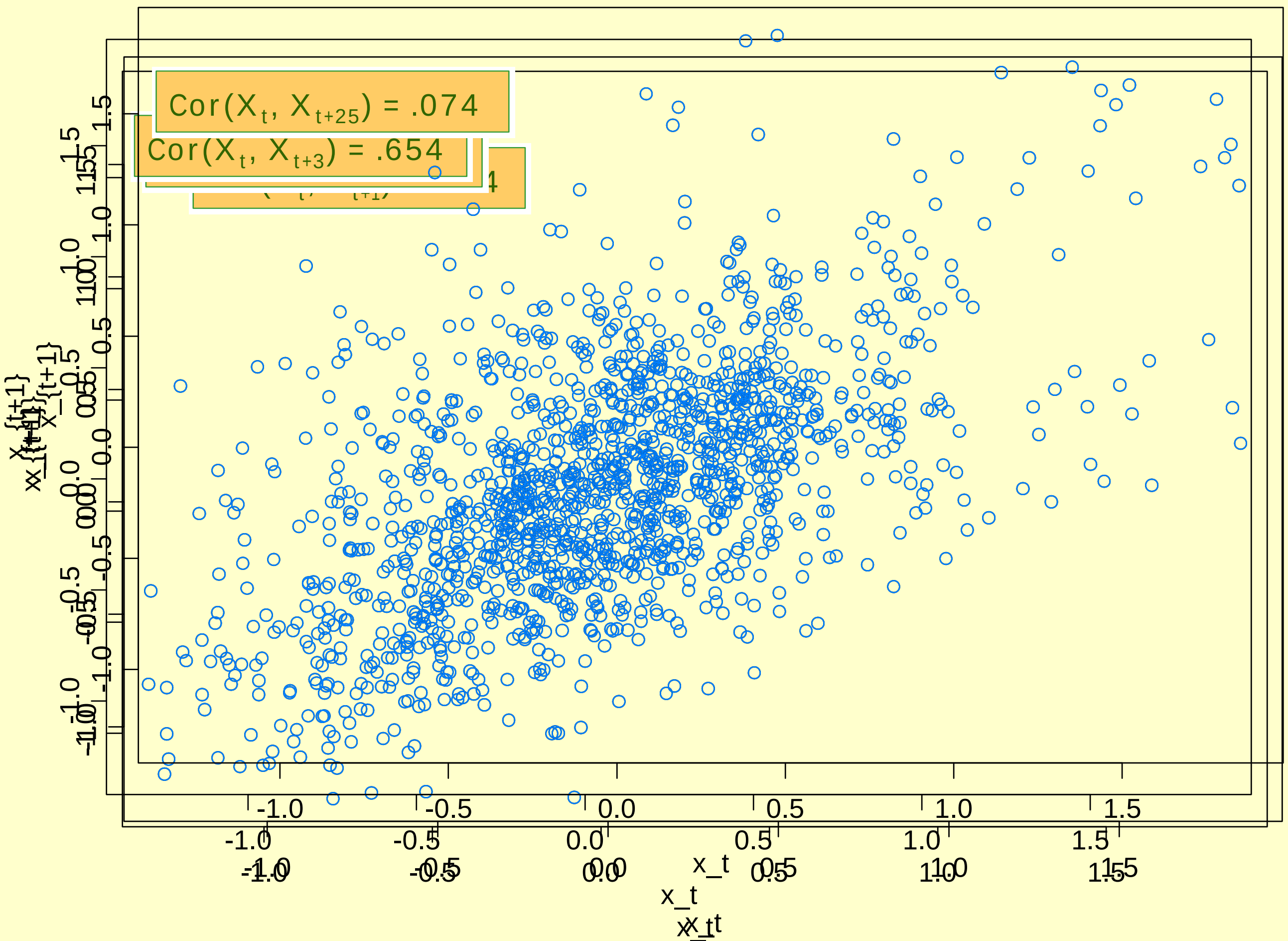
The time series $\{X_t\}$ is **white** or **independent noise** if the sequence of random variables is independent and identically distributed.



Battery of tests for checking whiteness.

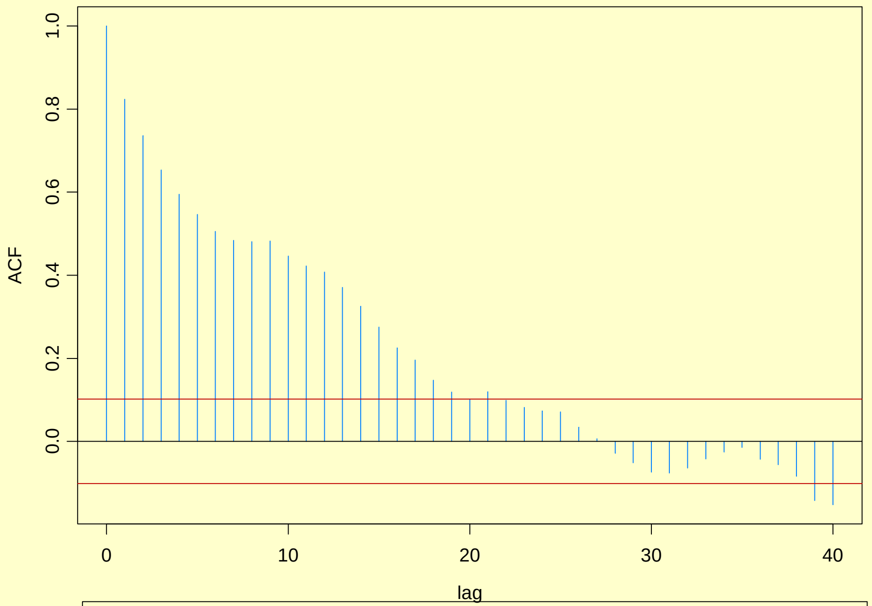
In ITSM, choose statistics => residual analysis => Tests of Randomness

Residuals from Mauna Loa data.



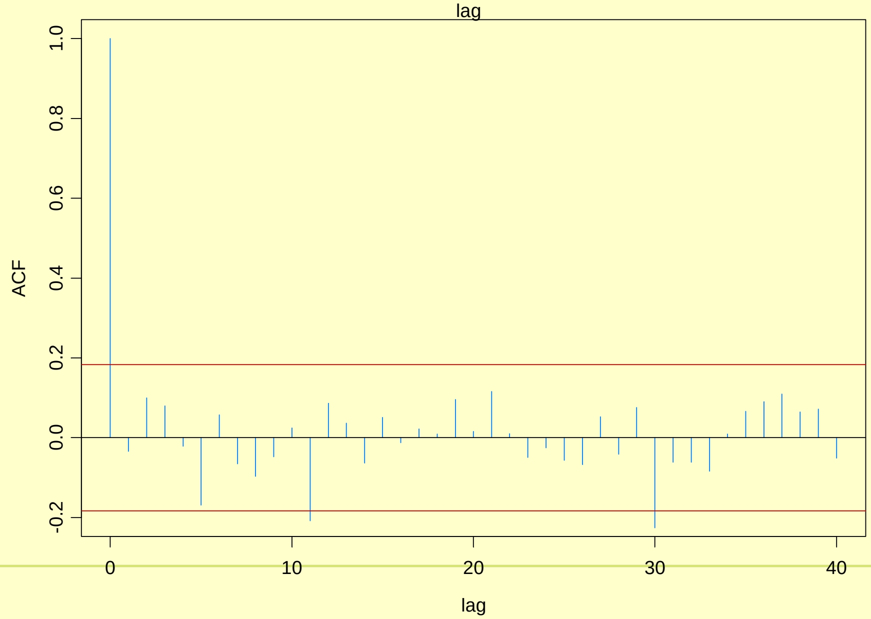
t	r_t	r_{t+25}
1	-.19	.13
2	-.14	.04
3	-.25	.20
4	-.13	.47
	…	

Autocorrelation function (ACF):



Mauna Loa residuals

Conf Bds: ± 1.96/sqrt(n)

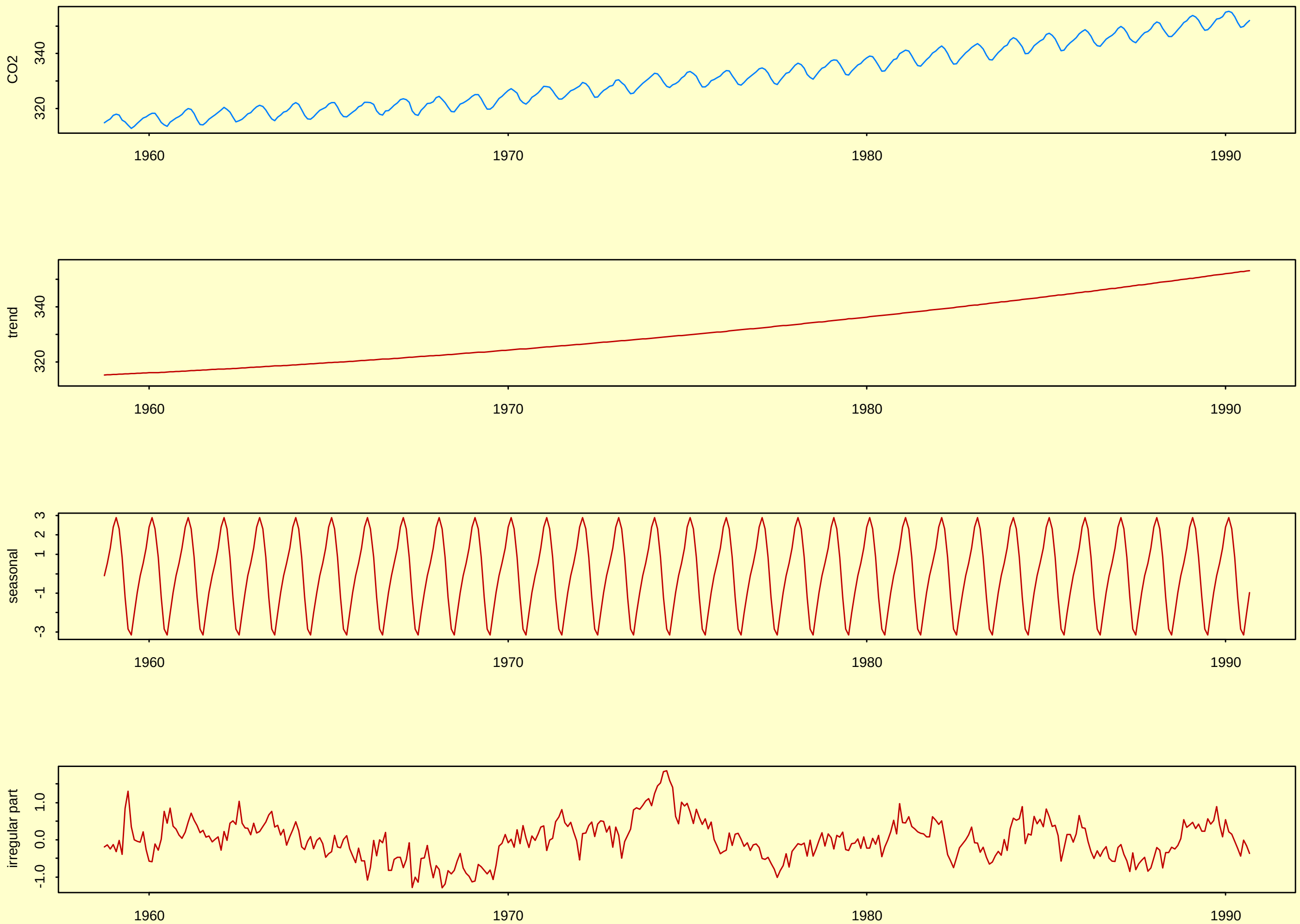


white noise

Colorado State University
Knowledge to Go Places

Putting it all together

Example: Mauna Loa



Strategies for modeling the irregular part $\{Y_t\}$.

- Fit an autoregressive process
- Fit a moving average process
- Fit an ARMA (autoregressive-moving average) process

In ITSM, choose the best fitting AR or ARMA using the menu option

Model => Estimation => Preliminary => AR estimation

or

Model => Estimation => Autofit

PRIMES

How well does the model fit the data?

1. Inspection of residuals.

   Are they compatible with white (independent) noise?

   - no discernible trend
   - no seasonal component
   - variability does not change in time.
   - no correlation in residuals or **squares** of residuals

   Are they normally distributed?

2. How well does the model predict.

   - values within the series (in-sample forecasting)
   - future values

3. How well do the simulated values from the model capture the characteristics in the observed data?

ITSM DEMO with Mauna Loa

Model refinement and Simulation

- Residual analysis can often lead to model refinement
- Do simulated realizations reflect the key features present in the original data

Two examples

- Sunspots
- NEE (Net ecosystem exchange).

Limitations of existing models

- Seasonal components are fixed from year to year.
- Stationary through the seasons
- Add intervention components (forest fires, volcanic eruptions, etc.)

Other directions

- Structural model formulation for trend and seasonal components
  - Local level model

    $$m_t = m_{t-1} + noise_t$$

  - Seasonal component with noise

    $$s_t = -s_{t-1} - s_{t-2} - \ldots - s_{t-11} + noise_t$$

  - $X_t = m_t + s_t + Y_t + \varepsilon_t$

- Easy to add intervention terms in the above formulation.
- Periodic models (allows more flexibility in modeling transitions from one season to the next).