Análisis Estadístico
de
Datos Climáticos

Análisis exploratorio de datos univariados

2015

DATOS UNIVARIADOS

Robustez y Resistencia

Medidas numéricas de localización, dispersión, asimetría ….

Técnicas gráficas

Distribuciones

……..

Transformaciones

Anomalías

Hace 8 ºC afuera, ¿está frío?

Dónde, a qué hora, en qué estación del año, etc.

$$Z = X - \bar{X}$$

En general se definen respecto del ciclo anual
(o eventualmente el ciclo diario si fuera el caso)

$\bar{X}$ Es entonces el promedio de la variable para cada mes.

---

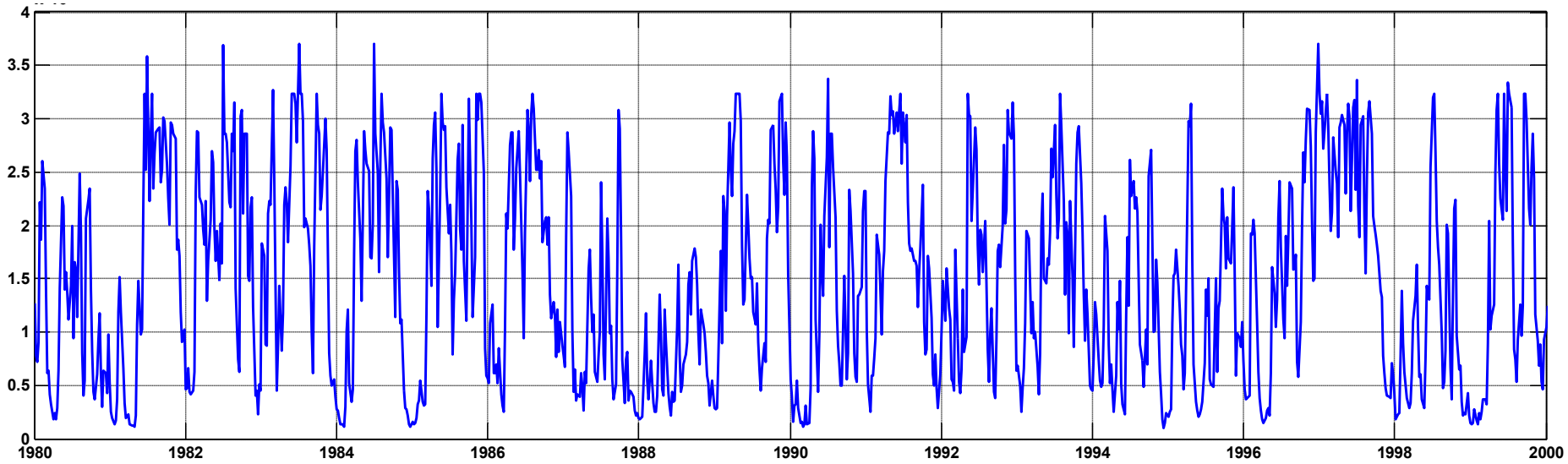
z=x-mean(x)

[lx,ax]=size(x)

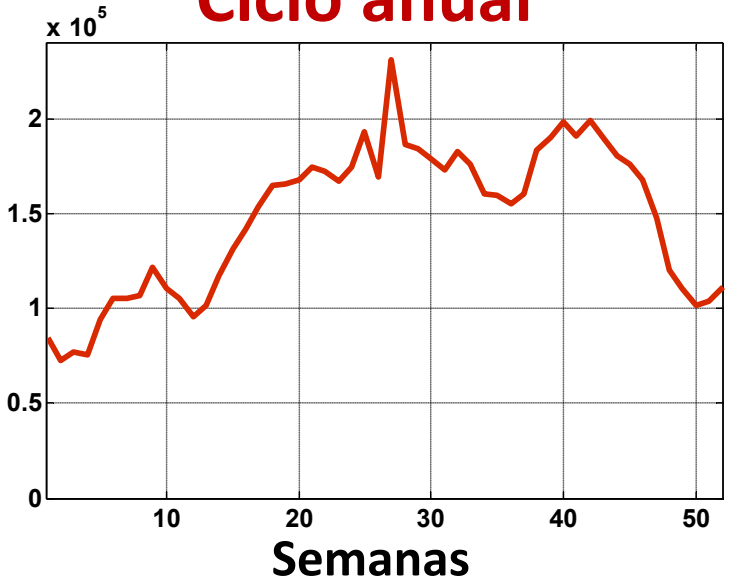
↓ Dim

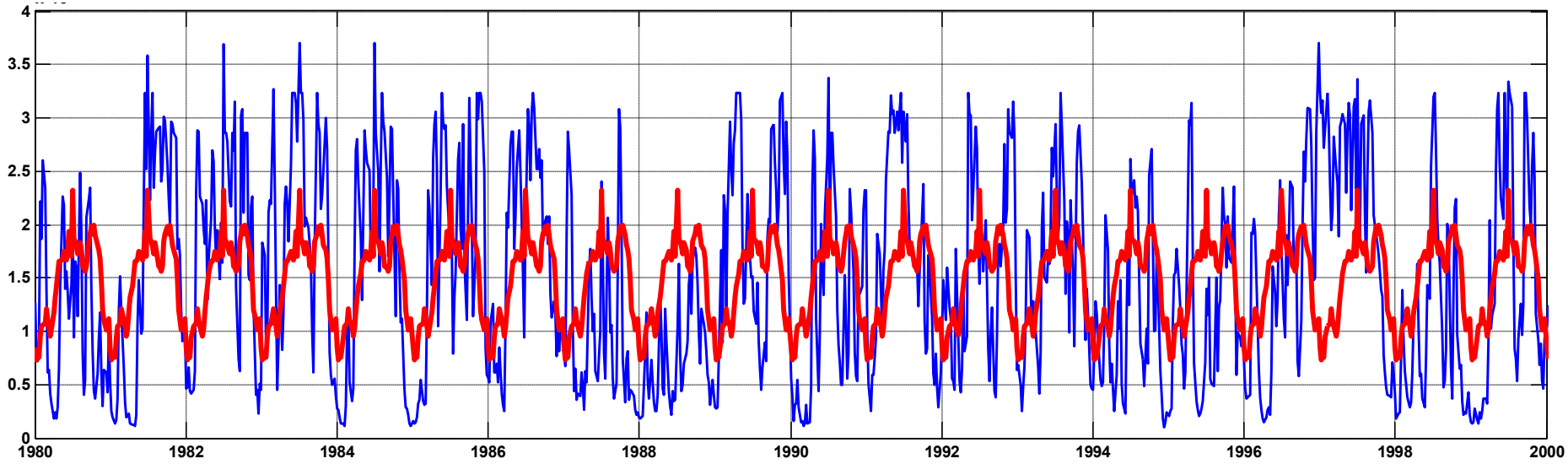
z=x-mean(x,2)*ones(1,ax)

Caudales

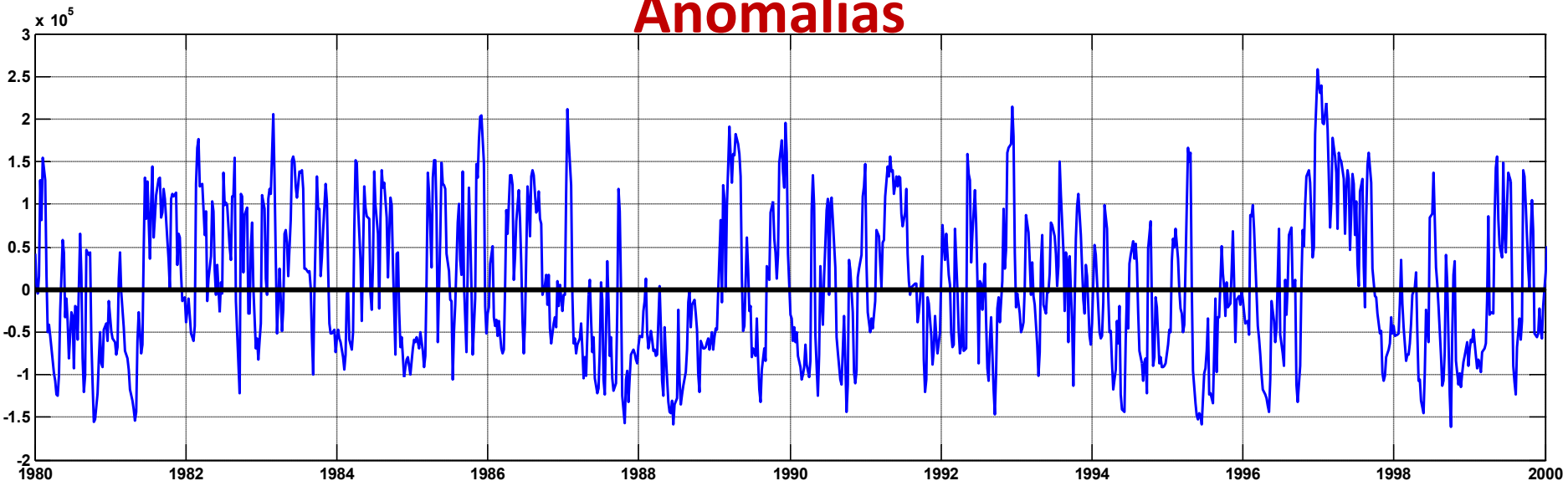


Ciclo anual

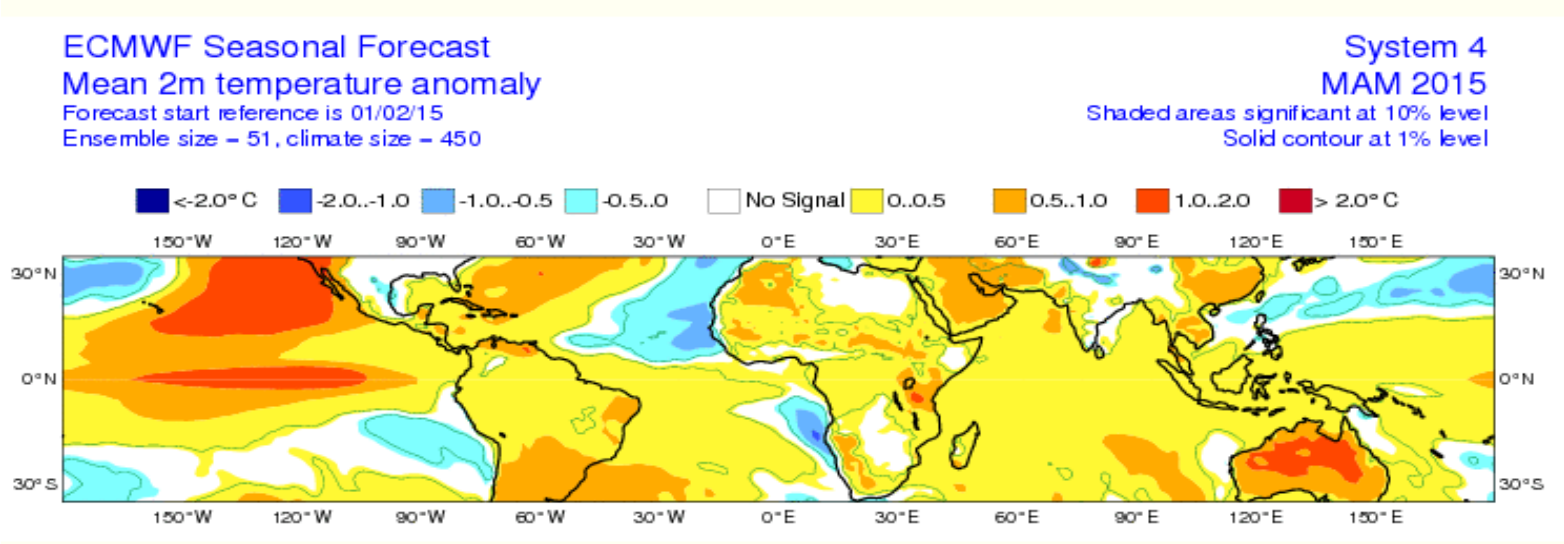


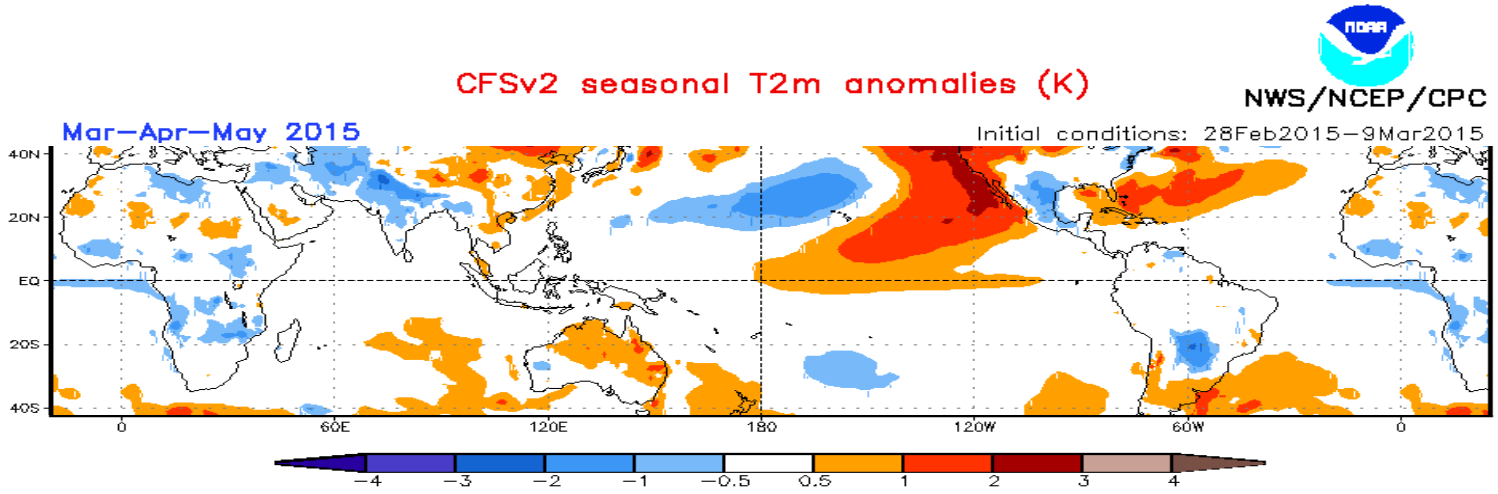

4
3.5
3
2.5
2
1.5
1
0.5
0
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000

Anomalías


x 10^5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000

¿Y cuando ya te presentan las anomalías?





¡ Hay que preguntar por la climatología de base !

Anomalías estandarizadas

$$Z = \frac{X - \bar{X}}{S_x}$$

Z no tiene unidades

Z tiene media 0 y desviación estándar 1

Esto facilita compara variables diferentes

---

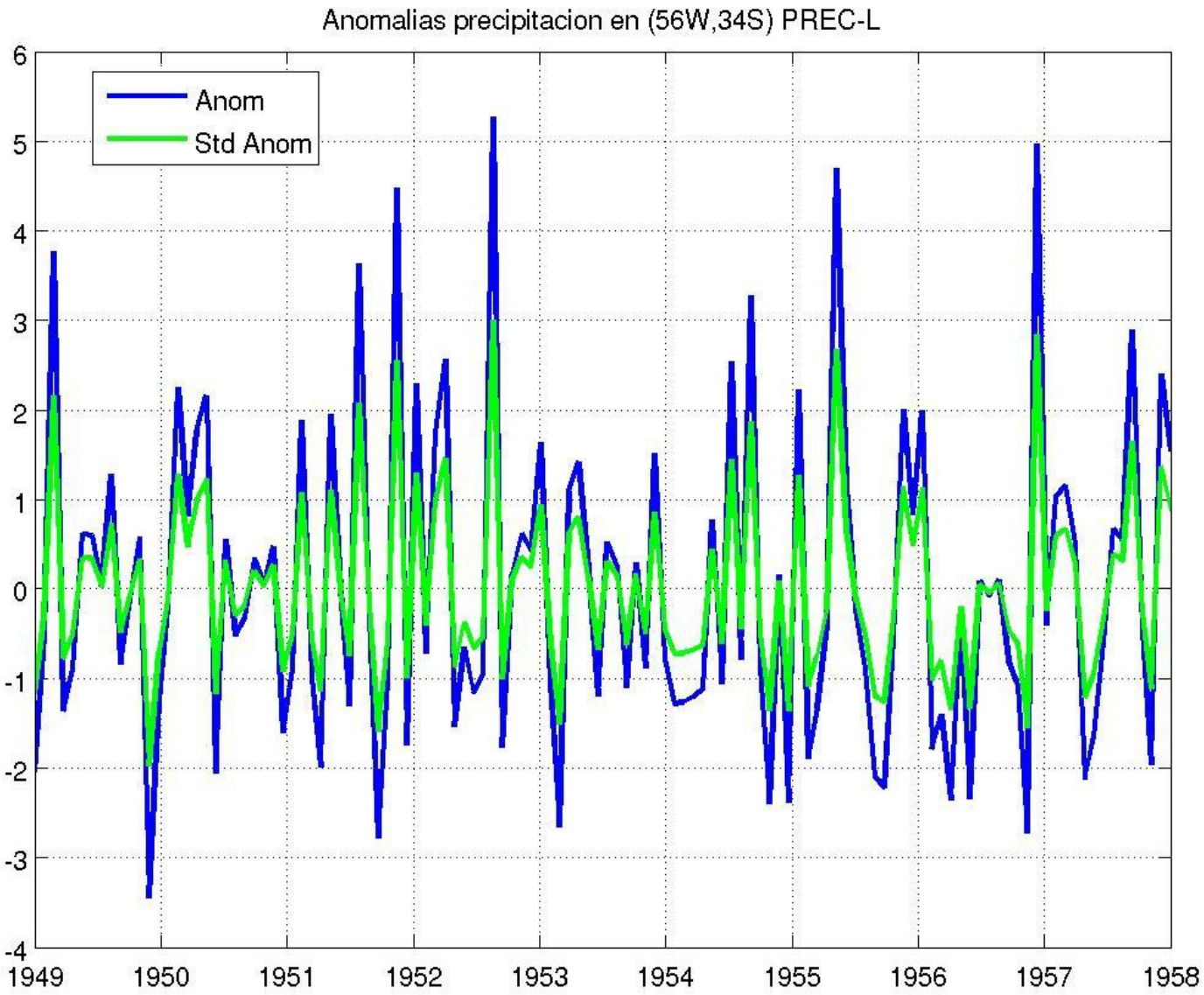
z=(x-mean(x))/std(x)

[lx,ax]=size(x)

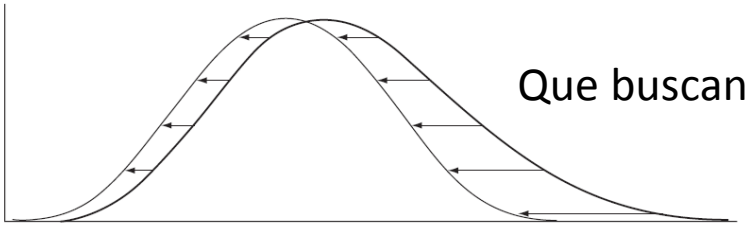
Dim

z=(x-mean(x,2)*ones(1,ax))./(std(x,0,2)*ones(1,ax))

Flag (N-1 o N)


Anomalias precipitacion en (56W,34S) PREC-L
Anom
Std Anom
6
5
4
3
2
1
0
-1
-2
-3
-4
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958

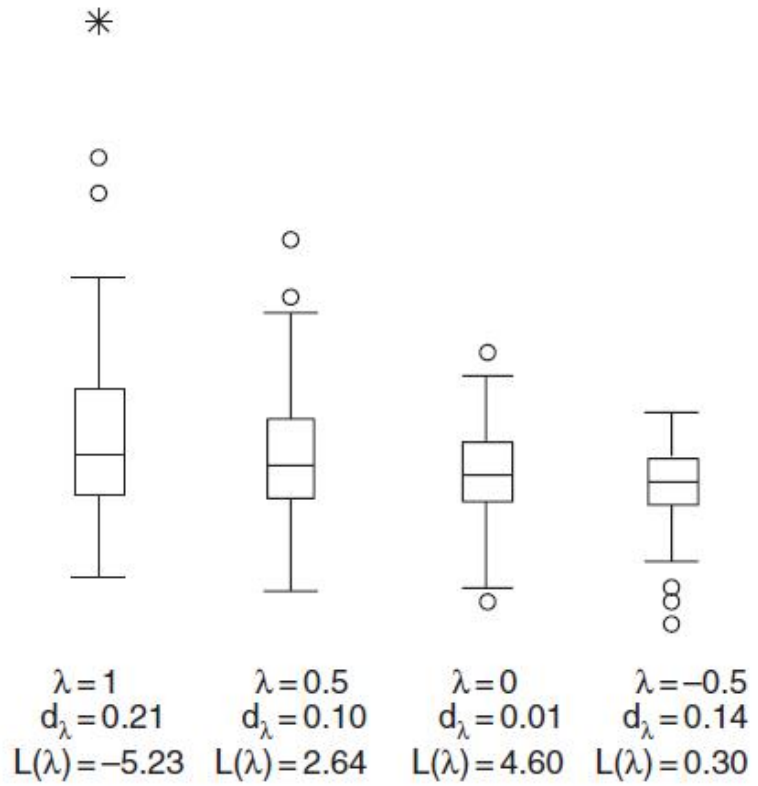
Otras transformaciones



Que buscan “simetrizar” los datos

$$T_1(x) = \begin{cases} x^\lambda, & \lambda > 0 \\ \ln(x), & \lambda = 0, \\ -(x^\lambda), & \lambda < 0 \end{cases} \qquad (3.18a)$$

$$T_2(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(x), & \lambda = 0. \end{cases} \qquad (3.18b)$$



Precipitación

Wilks, Pg.43

DATOS APAREADOS

Dos series de datos simultáneas
(misma variable en distintos puntos o diferente variable en mismo punto)
de los que quiere explorar si hay alguna relación

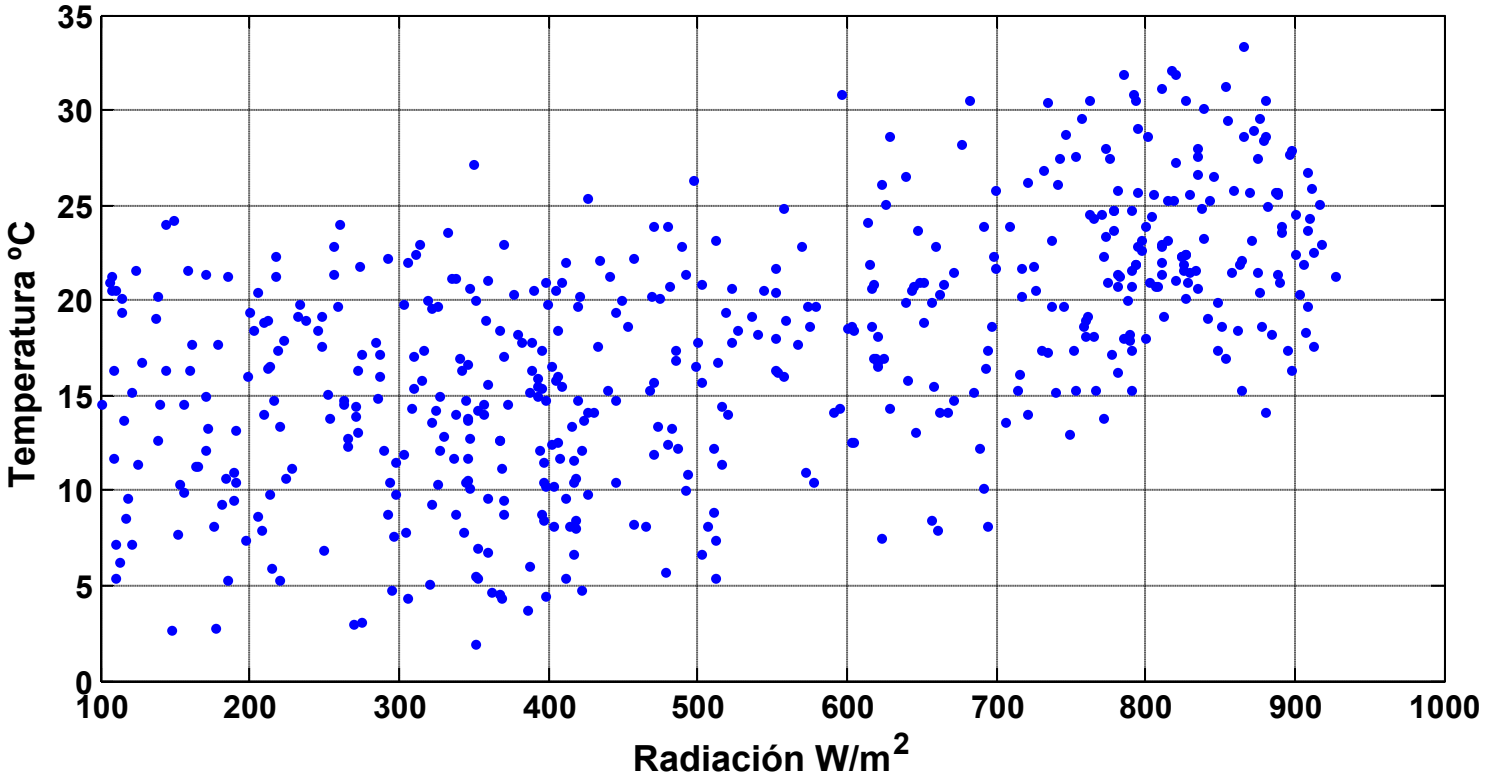
Diagramas de dispersión (“scatter plots”)

Correlaciones de Pearson

Correlaciones de Spearman (rango)

Diagrama de dispersión

Un punto por instante de tiempo, una variable en cada eje: plot(x,y,'+')
Da una primera idea de la relación entre las variables, o su ausencia



Correlación de Pearson

Cociente de las covarianzas entre el producto de las desviaciones estándar, adimensionado

$$r_{xy} = \frac{\text{Cov}\ (x, y)}{s_x s_y} = \frac{\dfrac{1}{n-1}\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{\left[\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}\left[\dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}}$$

$$= \frac{\sum_{i=1}^{n}(x_i' y_i')}{\left[\sum_{i=1}^{n}(x_i')^2\right]^{1/2}\left[\sum_{i=1}^{n}(y_i')^2\right]^{1/2}},$$

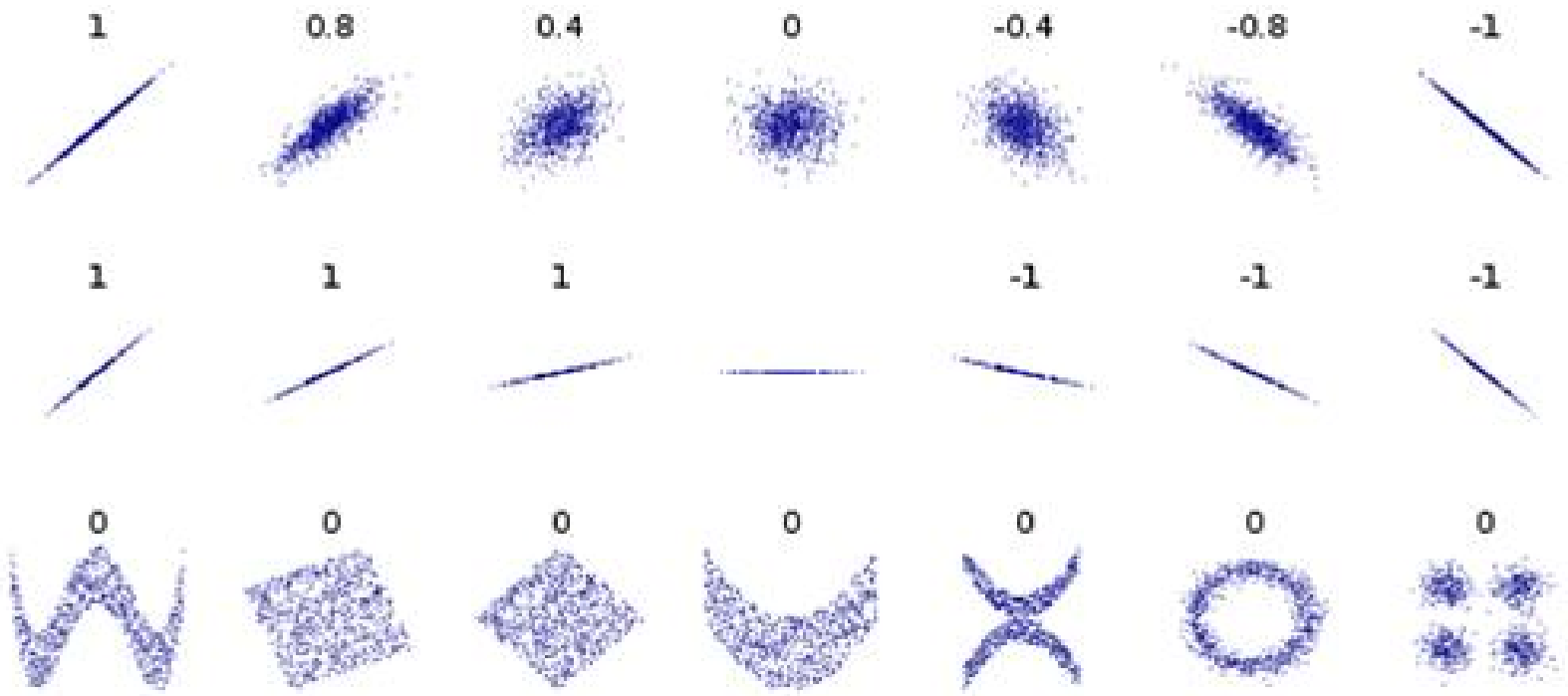
Operando, queda el promedio de las anomalías estandarizadas

$$r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{(x_i - \bar{x})}{s_x}\frac{(y_i - \bar{y})}{s_y}\right] = \frac{1}{n-1}\sum_{i=1}^{n} z_{x_i} z_{y_i}$$

corr(x,y)

Correlación de Pearson

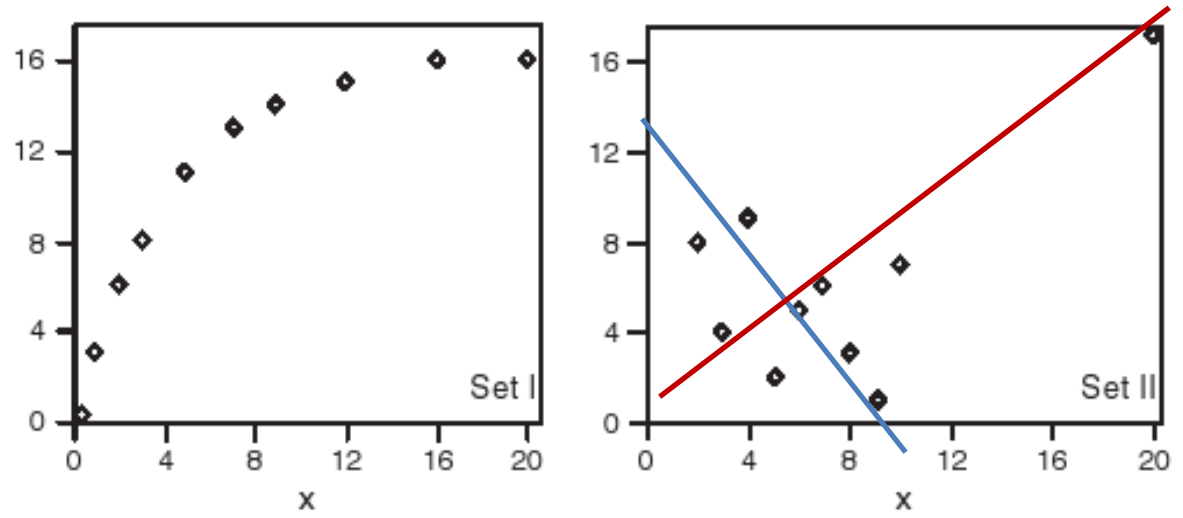
$$-1 \leq r_{xy} \leq 1$$



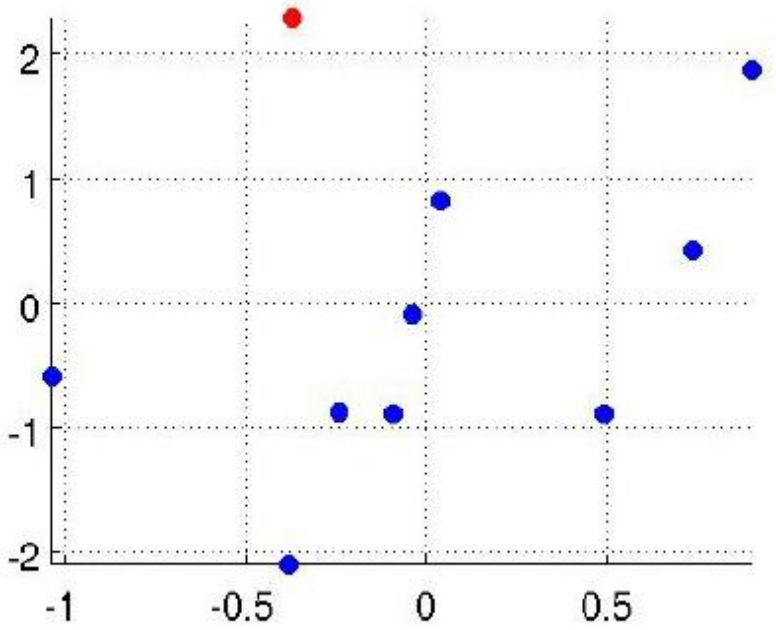
Correlación de Pearson

r_{xy}^2 Indica la fracción de la varianza de una de las dos variables que está descrita **linealmente** por la otra

No es **ni robusta** (caracteriza relaciones lineales)
ni resistente (sensible a outliers)



Correlación de Pearson



Poca resistencia a outliers

Correlación con punto rojo

r=0.34

Correlacion sin dato rojo

r=0.61

Correlación de Spearman (de rango)

Consiste en hacer lo mismo que en Pearson pero a los rangos de los datos.

Como los números son naturales, haciendo cuentas, se puede expresar la correlación como:

$$r_{rank} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$$

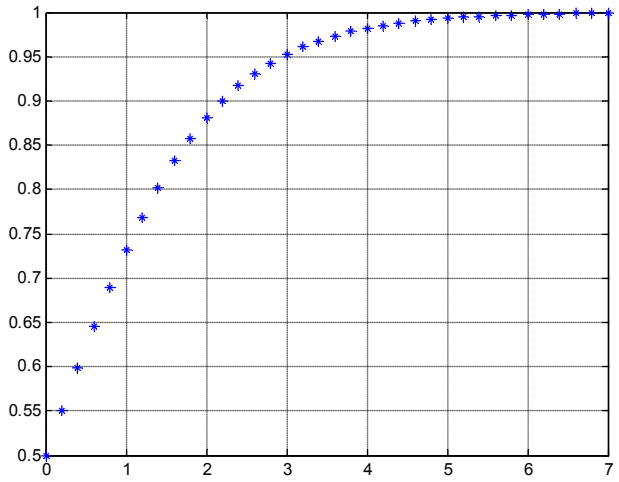
Donde $D_i = rango(x_i) - rango(y_i)$

Es robusta y resistente

corr(x,y,'Type','Spearman')

Correlación de Spearman (de rango)

Así como Pearson captura la relación lineal, Spearman captura una relación monotónica



r_Pearson=0.85

r_Spearman=1

DATOS APAREADOS CONSIGO MISMO

AUTOCORRELACIONES desfasados

(Podrían estudiarse también correlaciones desfasadas -"lagged"- entre variables distintas)

Se calcula:

$$r_1 = \frac{\sum_{i=1}^{n-1}[(x_i - \bar{x}_-)(x_{i+1} - \bar{x}_+)]}{\left[\sum_{i=1}^{n-1}(x_i - \bar{x}_-)^2 \sum_{i=2}^{n}(x_i - \bar{x}_+)^2\right]^{1/2}}$$

Pero es simplemente la correlación (usualmente, pero no necesariamente, de Pearson) entre la serie y sí misma con un corrimiento de un lugar.

Caso de $e^{-\frac{t_i}{\tau}}$ **i=1, …. n**

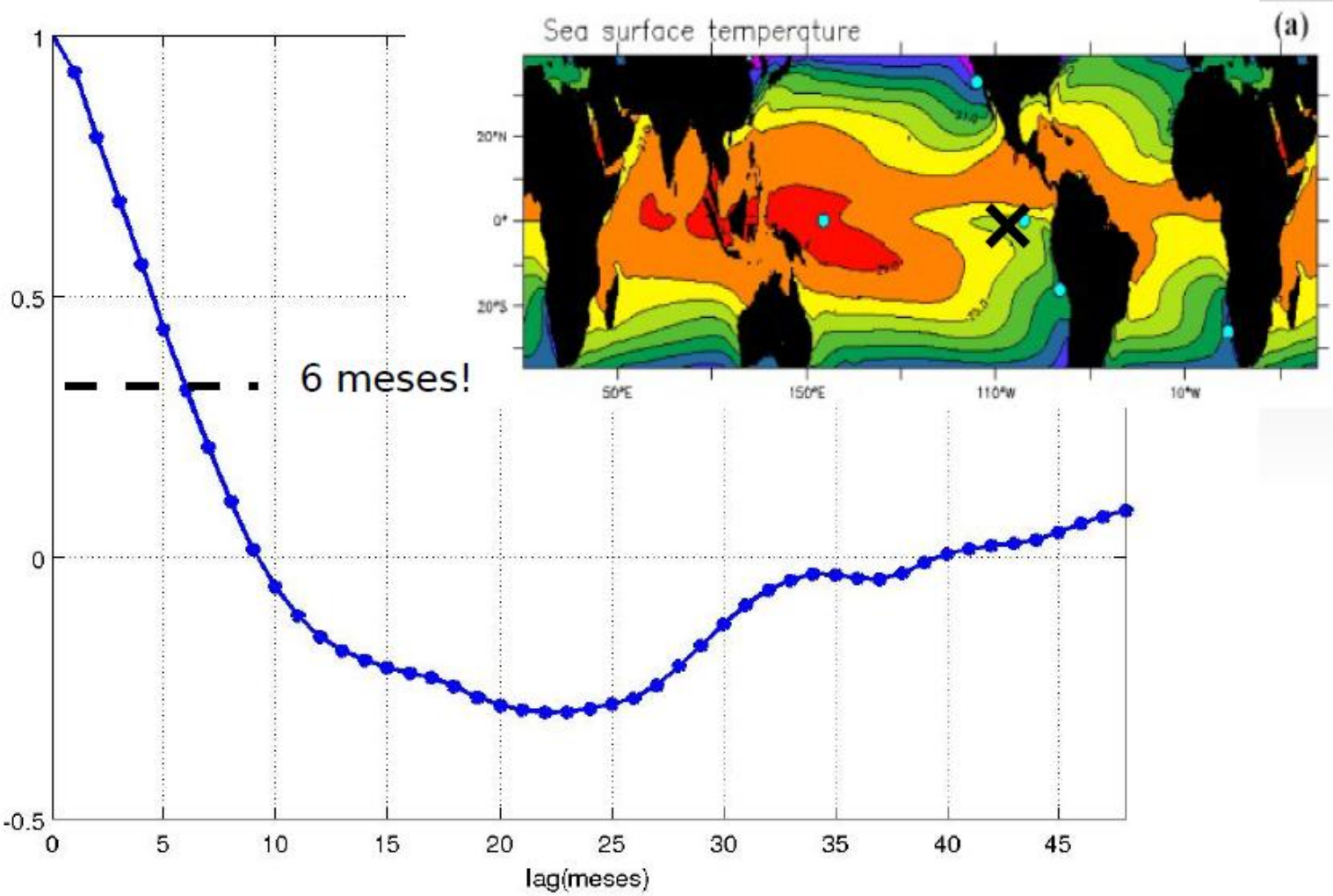
Funciones de Auto-correlación

Lo mismo que antes pero en función del desfasaje, por lo que obtengo una función

Para cualquier k, puedo calcular la auto-correlación con desfasaje k,
ojo que a medida que k aumenta cada vez tengo menos datos.

$$r_k = \frac{\sum_{i=1}^{n-k} [(x_i - \bar{x}_-)(x_{i+k} - \bar{x}_+)]}{\left[\sum_{i=1}^{n-k} (x_i - \bar{x}_-)^2 \sum_{i=k+1}^{n} (x_i - \bar{x}_+)^2\right]^{1/2}}.$$

Autocorrelación TSM lengua fria del Pacífico

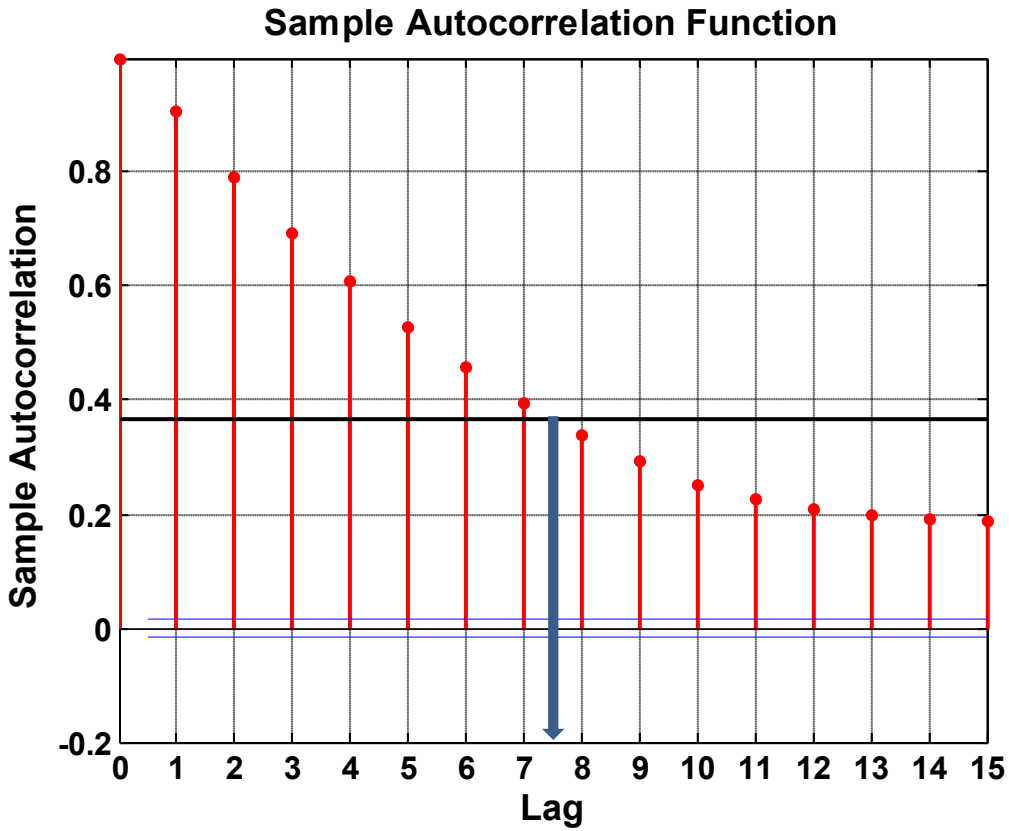


La persistencia de las anomalías de temperatura de superficie de mar es del órden de 3 meses dependiendo de la región. Eso permite pronosticar el estado del océano con cierta antelación.

Funciones de Auto-correlación

autocorr(x)

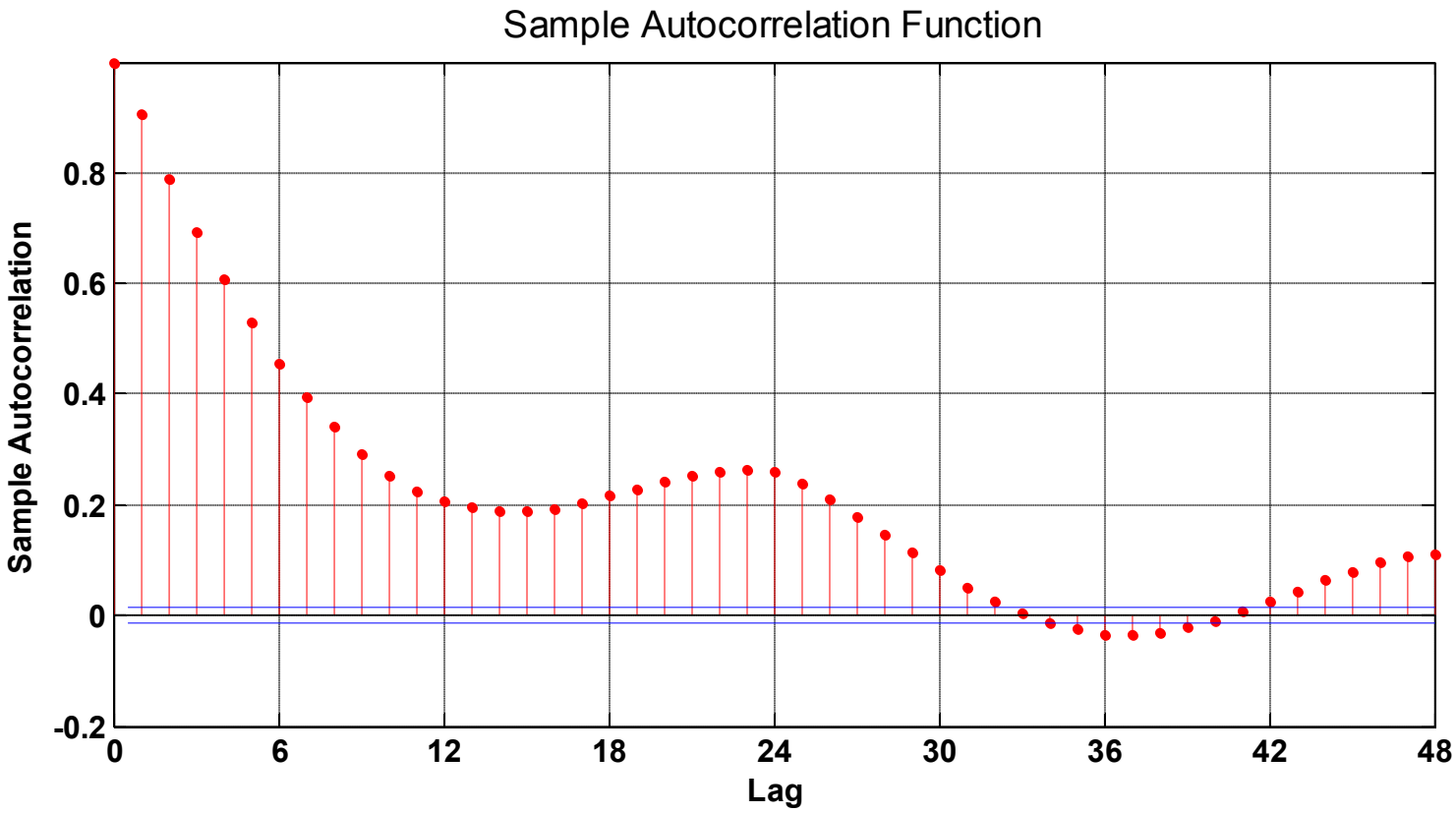
Módulo Viento



Horas

Tiempo de persistencia (tal que autocorrelación = e^{-1}): 7.5 horas

Funciones de Auto-correlación



Lo mismo pero a 48 horas

Funciones de Auto-correlación

Caso de $\sin(t_i)$ i=1, …. N

Hago integral en períodos completos para simplificar

$$f(\tau) = \frac{1}{n\pi}\int_0^{2\pi n} \sin(t)\,.\,\sin(t+\tau)\,.\,ds$$

$$f(\tau) = \cos(\tau)$$

¿Caso de $\cos(t_i)$ i=1, …. N ?

Funciones de Auto-correlación

¿Y si primero le saco el ciclo diario al viento y trabajo con anomalías?

La exploración no tiene límites …..