



# DECSAI

**Departamento de Ciencias de la Computación e I.A.**

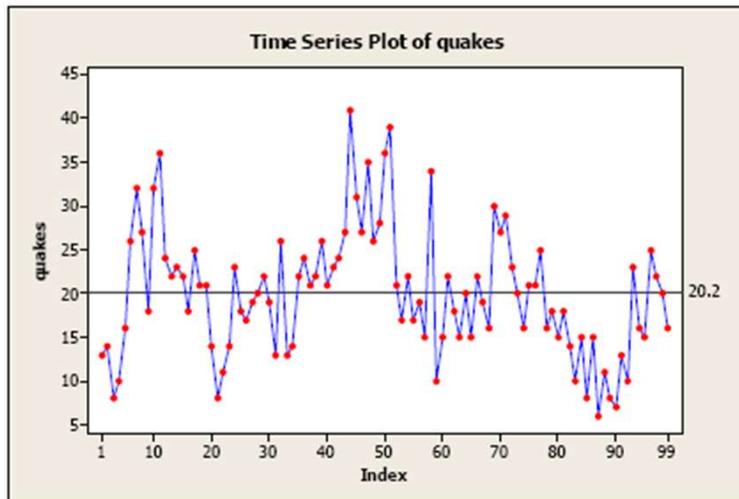
Universidad de Granada



## Series temporales

© Fernando Berzal, [berzal@acm.org](mailto:berzal@acm.org)

# Datos secuenciales



Start

Human G T T T T G A G G --- A T G T T C A A C A A A T G C T C C T T T C A T T C C T C T A T T T A C A G A C C T G C C G C A  
 Chimpanzee G T T T T G A G G --- A T G T T C A A T A A A T G C T G C T T T C A C T C G T C T A T T T A C A G A C C T G C C G C A  
 Macaque G T T T T G A G G --- A T G C T C A A T A A A T G C T C C T T T C A T T C C T C C A T T T A C A A A A C T T G C C G C A

Human G A C A A T T C T G C T A G C A G C C T T T G T G C T A T T A T C T G T T T T C T A A A C T T A G T A A T T G A G T G T  
 Chimpanzee G A C A A T T C T G C T A G C A G C C T T T G T G C T A T T A T C T G T T T T C T A A A C T T A G T A A T T G A G T G T  
 Macaque G A C A A T T C T G C T A G C A G C C T T T G T G C T A T T A T C T G T T T T C T A A A C T T A G T A A T T G A G T G T

Human G A T C T G G A G A C T A A - C T C T G A A A T A A A T A A G C T G A T T A T T T A T T T A T T T T C T C A A A A C A A  
 Chimpanzee G A T C T G G A G A C T A A A C T C T G A A A T A A A T A A G C T G A T T A T T T A T T T A T T T T C T C A A A A C A A  
 Macaque T A T C T G G A G A C T A A A C T C T G A A A T A A A T A A G C T G A T T A T T T A T T T A T T T T C T C A A A A C A A

Human C A G A A T A C G A T T T A G C A A A T T A C T T C T T A A G A T A T T A T T T T A C A T T T C T A T A T T C T C T A  
 Chimpanzee C A G A A T A C G A T T T A G C A A A T T A C T T C T T A A G A T A C T A T T T T A C A T T T C T A T A T T C T C T A  
 Macaque C A G A A T A T G A T T T A G C A A A T T A C C T C T T A A G A T A T T A T T T T G C A C T T C T A T A T T C T C T A

Human C C C T G A G T T G A T G T G T G A G C A A T A T G T C A C T T T C A T A A A G C C A G G T A T A C A - - - - T T A T G  
 Chimpanzee C C C T G A G T T G A T G T G T G A G C C G T A T G T C A C T T T C A T A A A G C C A G G T A T A C A - - - - T T A T G  
 Macaque C C C T G A G T T G A T G T G T G A G C A A T A T G T C A C T T C C A C A A A G C C A G G T A T A T A T A C A T T A C G

Human G A C A G G T A A G T A A A A A C A T A T T A T T T A T T C T A C G T T T T T G T C C A A A A T T T T A A A T T T C  
 Chimpanzee G A C A G G T A A G T A A A A A C A T A T T A T T T A T T C T A C G T T T T T G T C C A A A A T T T T A A A T T T C  
 Macaque G A C A G G T A A G T A A A A A - C A T A T T A T T T A T T C T A G G T T T T T G T C C A A G A G T T T T A A A T T T C

Human A A C T G T T G C G C G T G T G T T G G T A A - - - T G T A A A A C A A A C T C A G T A C A  
 Chimpanzee A A C T G T T G C G C G T G T G T T G G T A A - - - T G T A A A A C A A A C T C A G T A C A  
 Macaque A A C T G T T G T G C A T G T G T T G G T A A - - - C G T A A A A C A A A T T C A G T A C G



# Series temporales



- Características de las series temporales
- Visualización de series temporales
- Preprocesamiento
- Filtrado de series temporales
  - Medias móviles
  - Suavizado exponencial
- Técnicas de regresión
  - Regresión lineal
  - Coeficiente de correlación de Pearson
- Función de autocorrelación
- Caso práctico: Una sesión de análisis



# Series temporales



## Evolución de PIB en España. 1971 a 2020



# Características



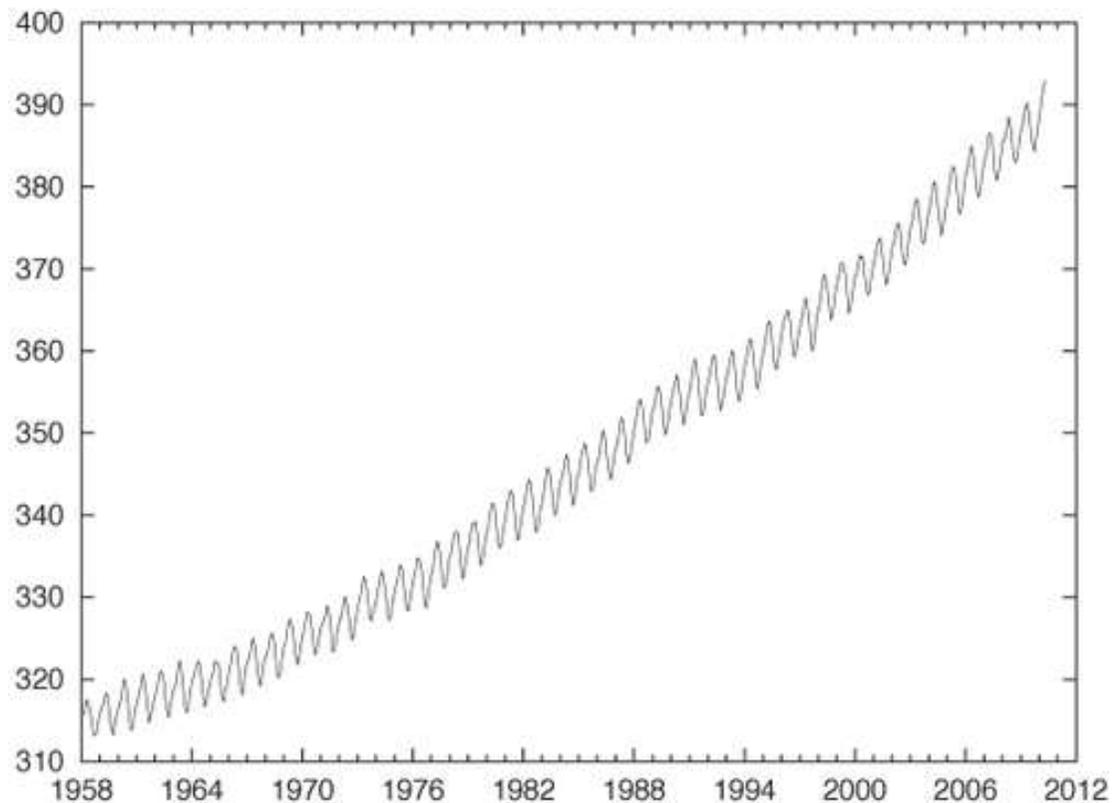
- Tendencias
- Estacionalidad (comportamientos periódicos)
- Ruido
- Otros, p.ej. cambios bruscos de comportamiento



# Ejemplos



## Tendencia y estacionalidad



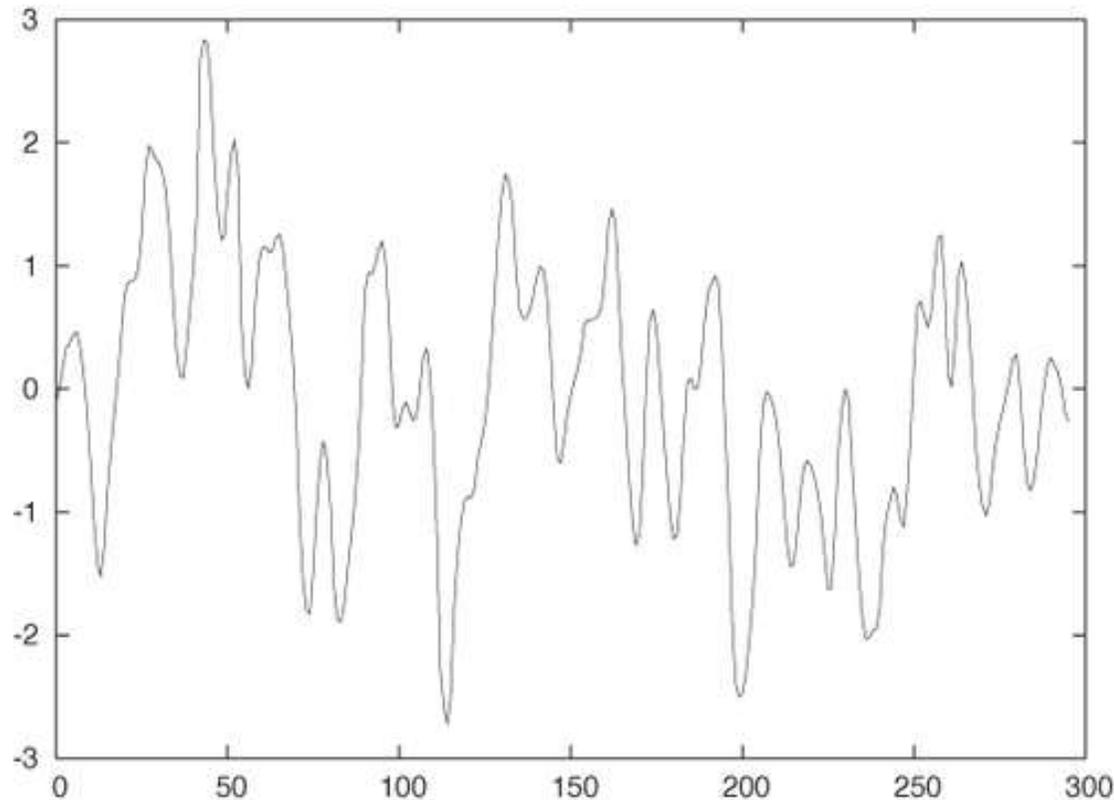
Concentración de CO<sub>2</sub>  
medida en el observatorio de Mauna Loa, Hawaii.



# Ejemplos



## Variación "suave" pero sin tendencia a largo plazo



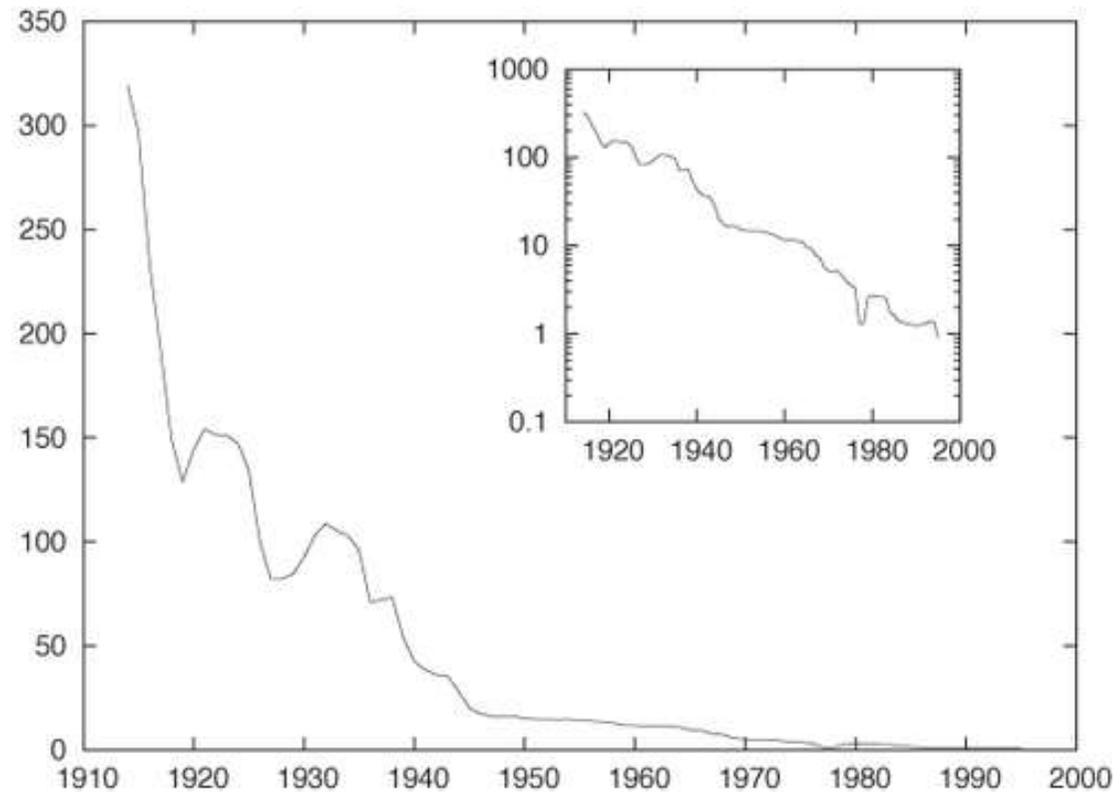
Concentración de gas a la salida de una caldera



# Ejemplos



## Tendencia no lineal



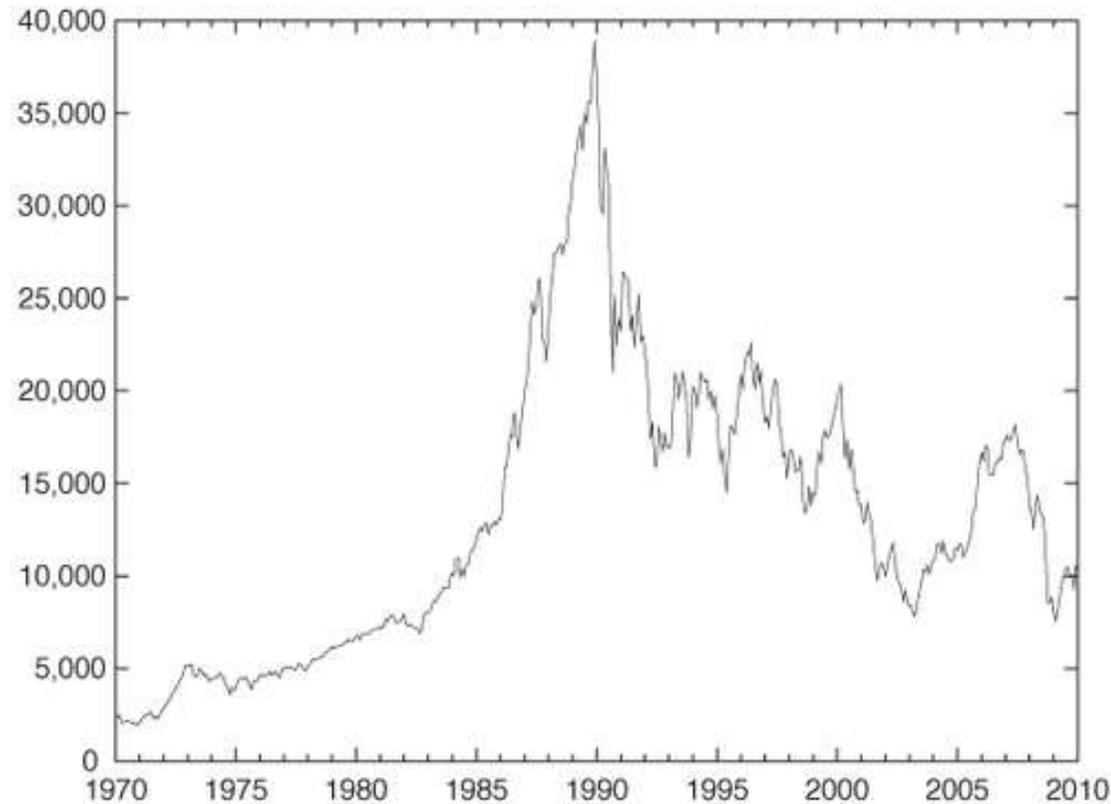
Coste de llamadas telefónicas de larga distancia (USA)



# Ejemplos



## Cambios “bruscos” de comportamiento



Índice Nikkei (Bolsa de Tokyo)



# Ejemplos

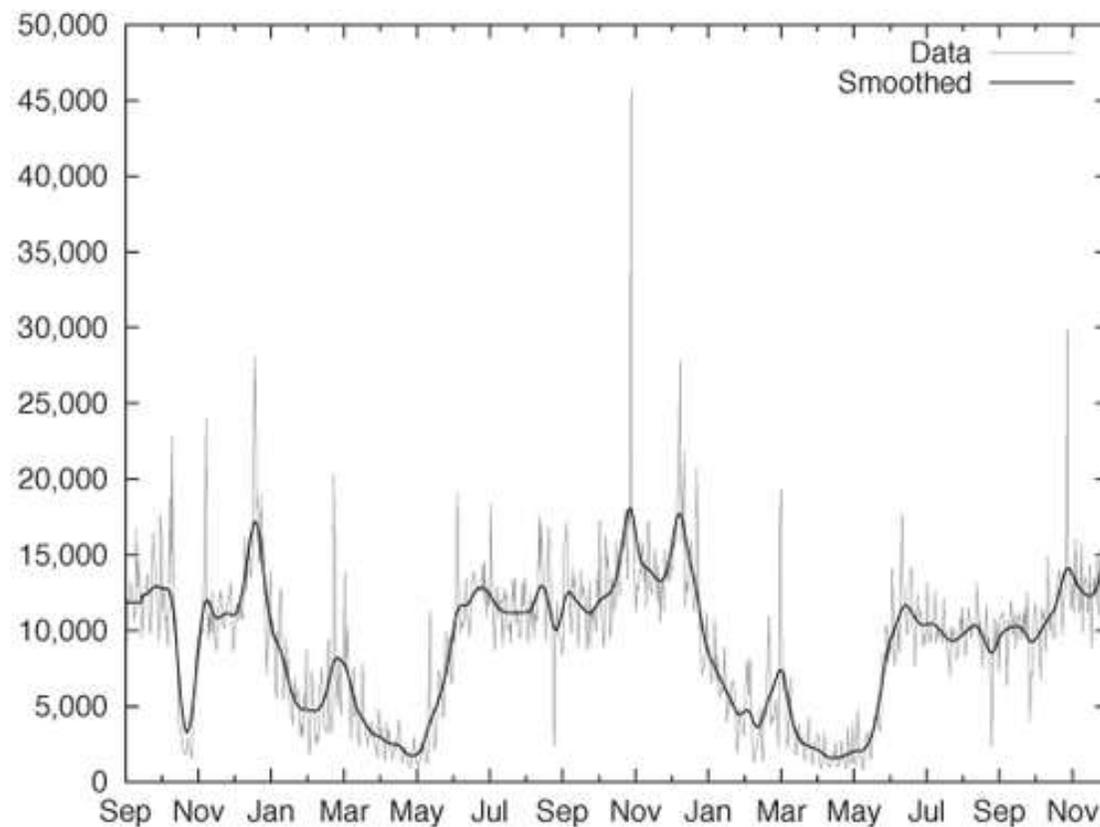


## Conjuntos de datos reales...

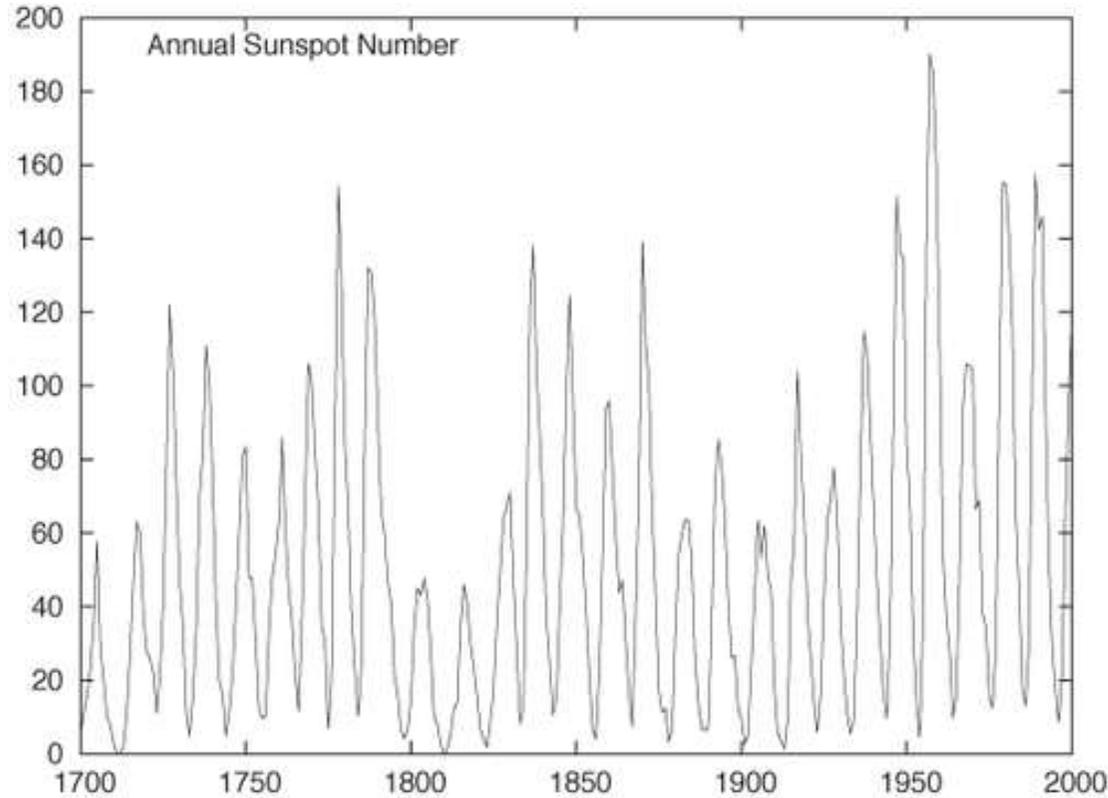
Estacionalidad a corto y largo plazo,  
posibles cambios de comportamiento

**y ruido**

Llamadas diarias  
a un call-center



# Visualización



Número anual de manchas solares durante 300 años

**Una relación de aspecto incorrecta  
hace difícil reconocer los detalles de cada ciclo.**

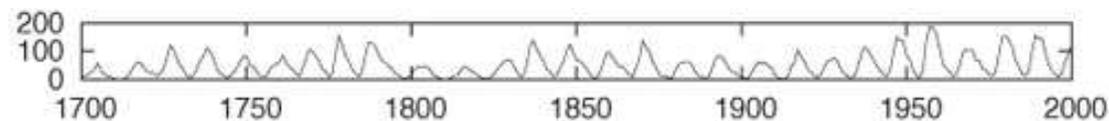


# Visualización



## **Banking** [Banking to 45 degrees]

Los cambios casi verticales de la figura anterior nos cuesta trabajo apreciarlos. Sin embargo, reconocemos mejor los cambios en una serie cuando se dibujan con un ángulo de  $45^\circ$ :



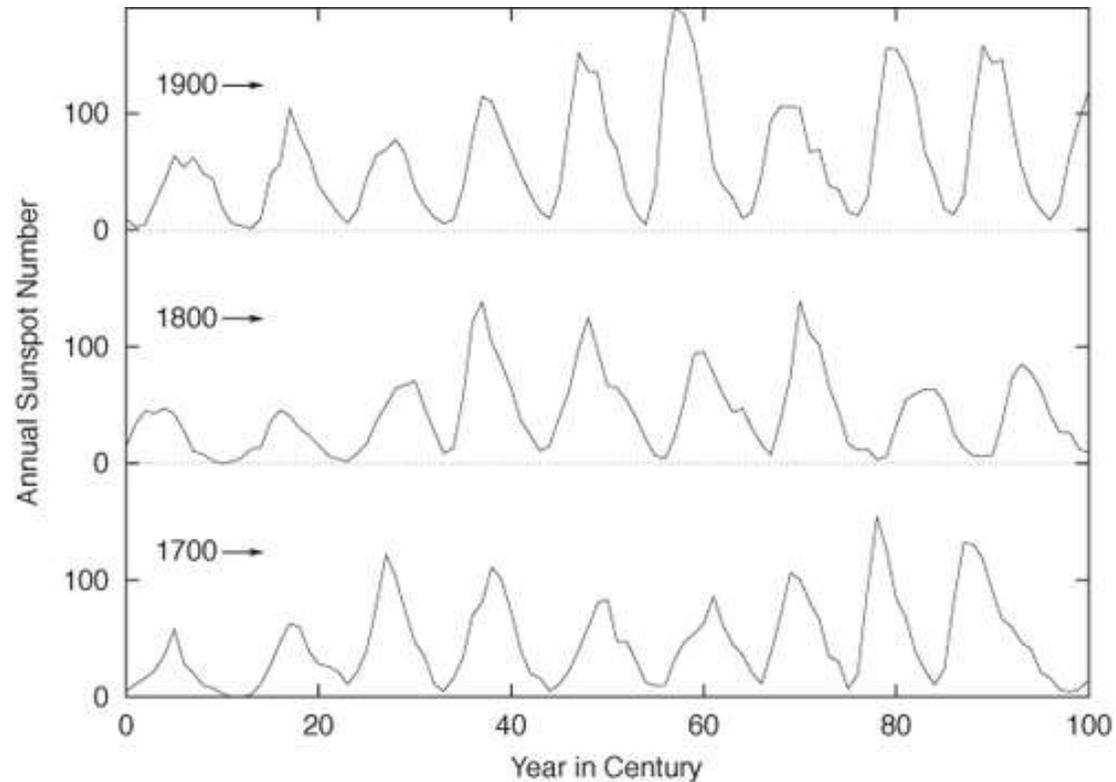
Ahora podemos apreciar que las “subidas” son más rápidas que las bajadas, aunque la figura es tan pequeña que apenas se pueden analizar detalles...



# Visualización



## Stacking



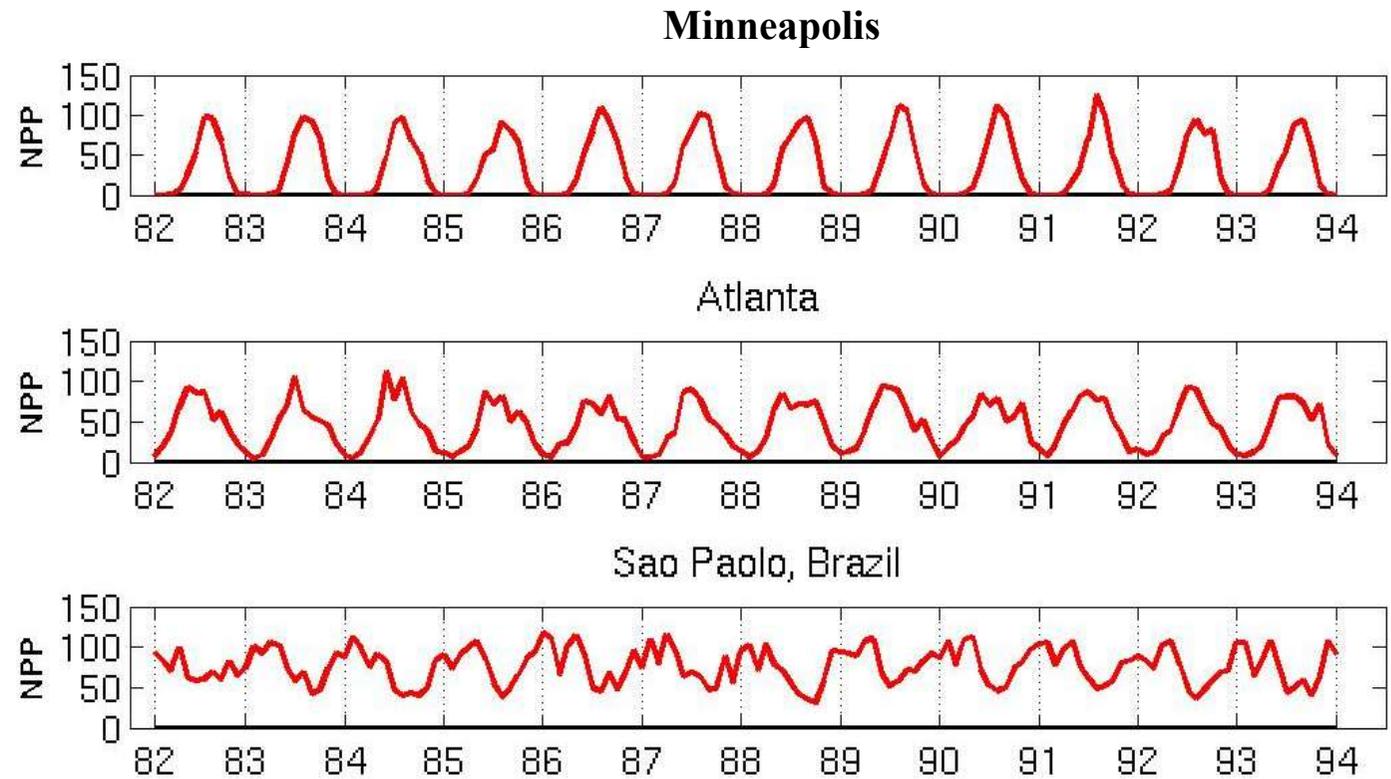
Dividiendo el eje temporal en 3 fragmentos, mantenemos el “banking” y generamos un gráfico con unas dimensiones más razonables (p.ej. 4:3).



# Preprocesamiento



## Datos originales



## Correlación entre las series temporales

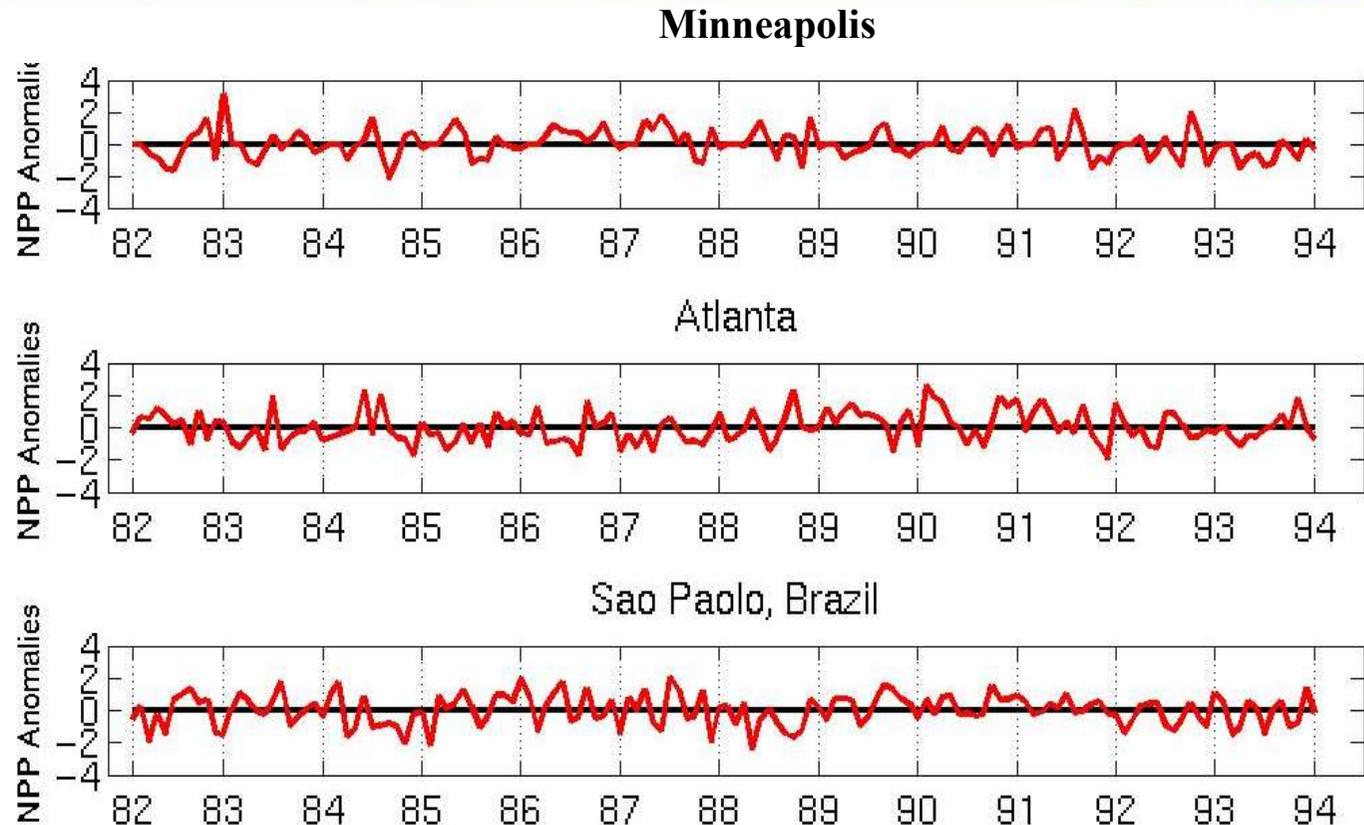
	Minneapolis	Atlanta	Sao Paulo
Minneapolis	1.0000	0.7591	-0.7581
Atlanta	0.7591	1.0000	-0.5739
Sao Paulo	-0.7581	-0.5739	1.0000



# Preprocesamiento



## Normalización z-score mensual



## Correlación entre las series temporales estacionalizadas

	Minneapolis	Atlanta	Sao Paulo
Minneapolis	1.0000	0.0492	0.0906
Atlanta	0.0492	1.0000	-0.0154
Sao Paulo	0.0906	-0.0154	1.0000

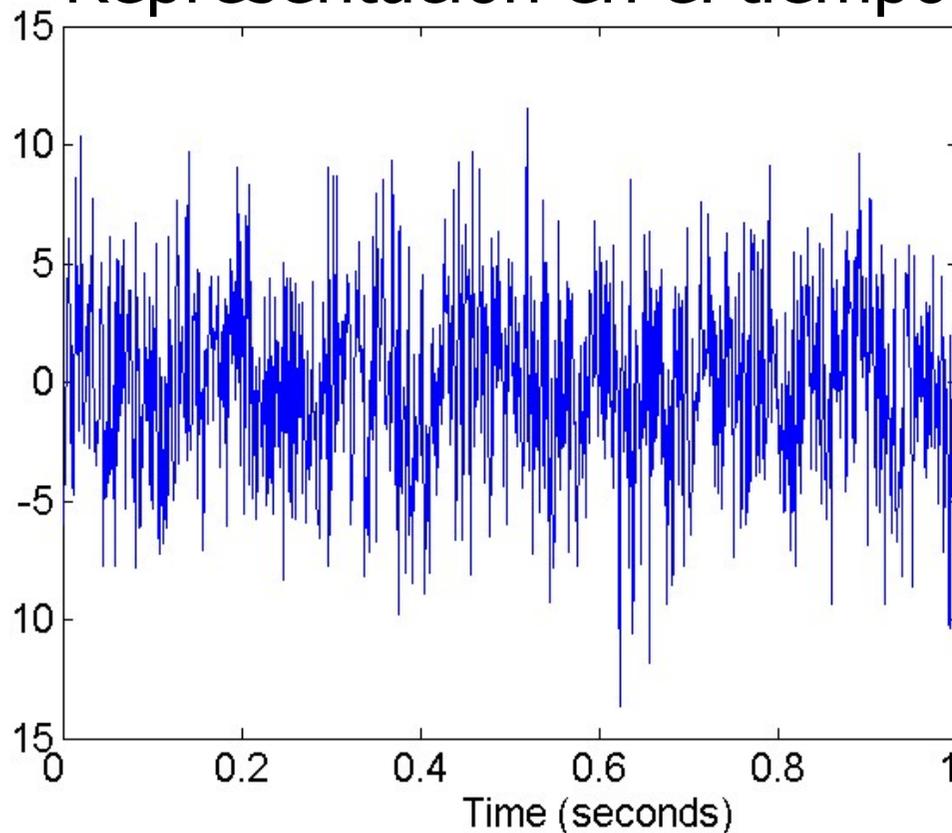


# Preprocesamiento

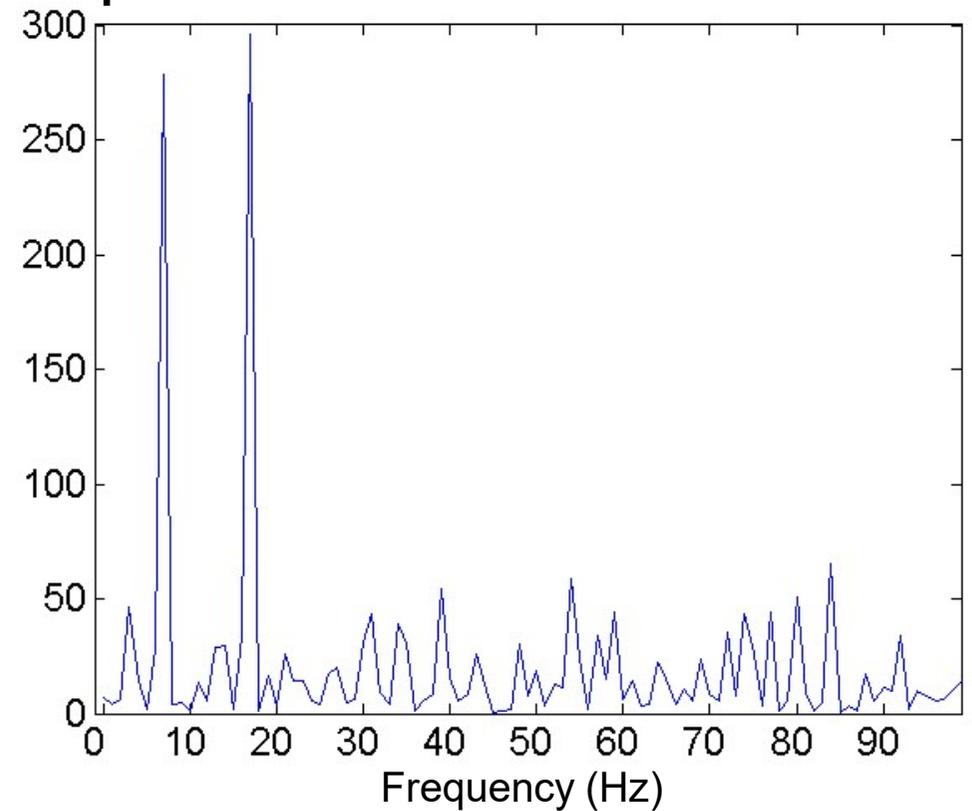


## Transformada de Fourier

Representación en el tiempo



Representación en frecuencia



# Filtrado de series temporales



## Medias móviles [moving averages]

IDEA: Reemplazar el punto central de una serie de un número impar de números consecutivos por su media aritmética (filtro “paso bajo”).

$$s_i = \frac{1}{2k + 1} \sum_{j=-k}^k x_{i+j}$$



# Filtrado de series temporales



## Medias móviles [moving averages]

PROBLEMA: La presencia de un pico en la ventana  $[i-k, i+k]$  distorsiona la media móvil.

POSIBLE SOLUCIÓN: Utilización de pesos (menores en los extremos de la ventana).

$$s_i = \sum_{j=-k}^k w_j x_{i+j} \quad \text{donde} \quad \sum_{j=-k}^k w_j = 1$$

Ejemplos: Gaussiana, ventana de Hamming...

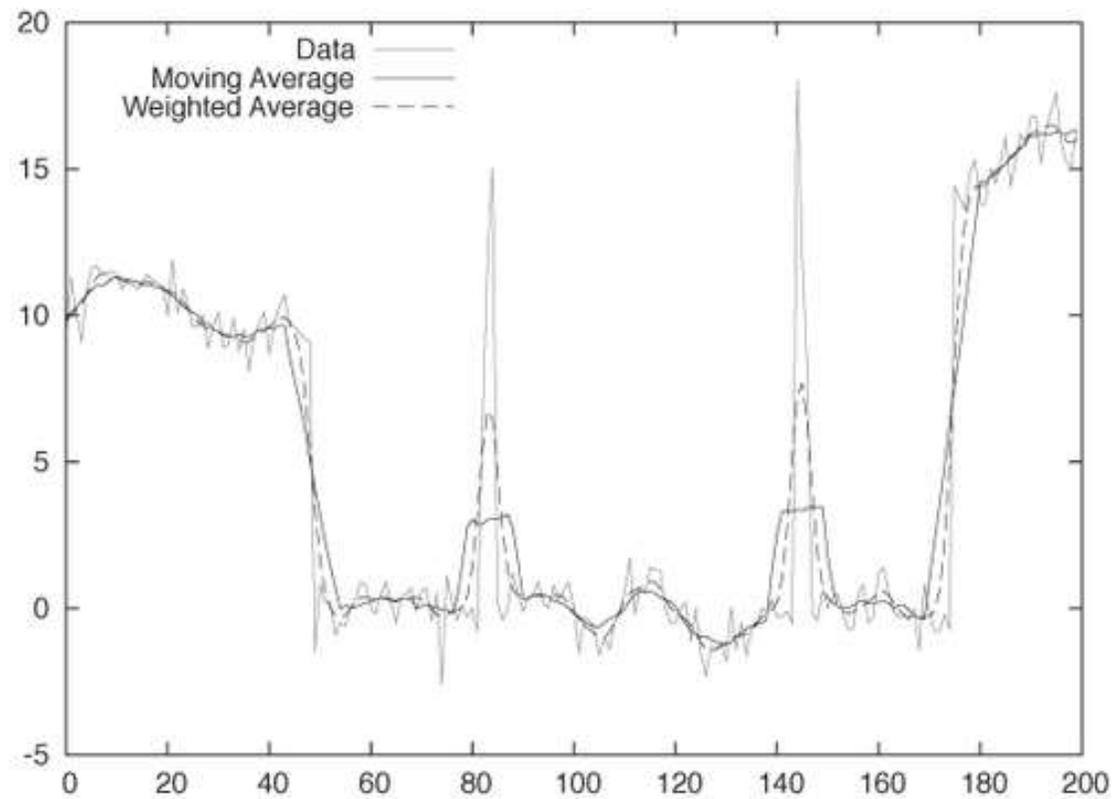
[http://en.wikipedia.org/wiki/Window\\_function#Hann\\_window](http://en.wikipedia.org/wiki/Window_function#Hann_window)



# Filtrado de series temporales



## Medias móviles [moving averages]



$k=5$



# Filtrado de series temporales



## **Medias móviles** [moving averages]

Limitaciones de las medias móviles:

- “Costosas” de calcular: Cuando se utilizan pesos, el cálculo hay que hacerlo desde cero para cada valor.
- Problemáticas en los extremos de las series de datos (dada la anchura de la ventana, no se pueden extender hasta el final de la serie, que suele ser lo más interesante).
- No se pueden definir fuera de la serie temporal, por lo que no se pueden utilizar para realizar predicciones.



# Filtrado de series temporales



## **Suavizado exponencial** [exponential smoothing]

Proporciona un filtrado fácil de calcular, además evita los problemas de las medias móviles:

- **Suavizado exponencial simple**  
(para series sin tendencia ni estacionalidad).
- **Suavizado exponencial doble**  
(para series con tendencia pero no estacionalidad).
- **Suavizado exponencial triple**  
(para series con tendencia y estacionalidad).



# Filtrado de series temporales



## Suavizado exponencial simple

$$s_i = \alpha x_i + (1 - \alpha)s_{i-1}$$

Los distintos métodos de suavizado exponencial actualizan el resultado del anterior valor con el último dato de la serie original (combinando la información ya disponible con la aportada por el nuevo dato mediante un parámetro,  $0 < \alpha < 1$ ).



# Filtrado de series temporales



## Suavizado exponencial simple

¿Por qué se llama suavizado exponencial?

Si expandimos la recurrencia, obtenemos:

$$s_i = \alpha \sum_{j=0}^i (1 - \alpha)^j x_{i-j}$$

Todas las observaciones previas contribuyen al valor suavizado, pero su contribución se suprime por el exponente creciente del parámetro  $\alpha$ .



# Filtrado de series temporales



## Suavizado exponencial simple

“Uso” en predicción: Si extendemos el suavizado más allá del final de los datos disponibles, la predicción es extremadamente simple :-)

$$x_{i+h} = S_i$$

Ante la presencia de tendencias, la señal suavizada tiene ir retrasada con respecto a los datos originales salvo que utilicemos un valor de  $\alpha$  cercano a 1.



# Filtrado de series temporales



## Suavizado exponencial doble

$$s_i = \alpha x_i + (1 - \alpha)(s_{i-1} + t_{i-1})$$

$$t_i = \beta (s_i - s_{i-1}) + (1 - \beta)t_{i-1}$$

El suavizado exponencial doble retiene información acerca de la tendencia: la señal suavizada  $s_i$  y la tendencia suavizada  $t_i$ .

El parámetro  $\beta$  se utiliza para realizar un suavizado exponencial sobre la tendencia.



# Filtrado de series temporales



## Suavizado exponencial doble

“Uso” en predicción:

Si extendemos el suavizado más allá del final de los datos disponibles, la predicción es la siguiente:

$$x_{i+h} = s_i + ht_i$$



# Filtrado de series temporales



## **Suavizado exponencial triple**

(a.k.a. método de Holt-Winters)

Una tercera cantidad se utiliza para describir la estacionalidad, que puede ser aditiva o multiplicativa según nos interese.

NOTA:

$p_i$  modela el componente periódico de la señal, donde  $k$  es el período observado.



# Filtrado de series temporales



## Suavizado exponencial triple

(a.k.a. método de Holt-Winters)

ESTACIONALIDAD ADITIVA

$$s_i = \alpha(x_i - p_{i-k}) + (1 - \alpha)(s_{i-1} + t_{i-1})$$

$$t_i = \beta(s_i - s_{i-1}) + (1 - \beta)t_{i-1}$$

$$p_i = \gamma(x_i - s_i) + (1 - \gamma)p_{i-k}$$

$$x_{i+h} = s_i + ht_i + p_{i-k+h}$$



# Filtrado de series temporales



## Suavizado exponencial triple

(a.k.a. método de Holt-Winters)

ESTACIONALIDAD MULTIPLICATIVA

$$s_i = \alpha \frac{x_i}{p_{i-k}} + (1 - \alpha)(s_{i-1} + t_{i-1})$$

$$t_i = \beta(s_i - s_{i-1}) + (1 - \beta)t_{i-1}$$

$$p_i = \gamma \frac{x_i}{s_i} + (1 - \gamma)p_{i-k}$$

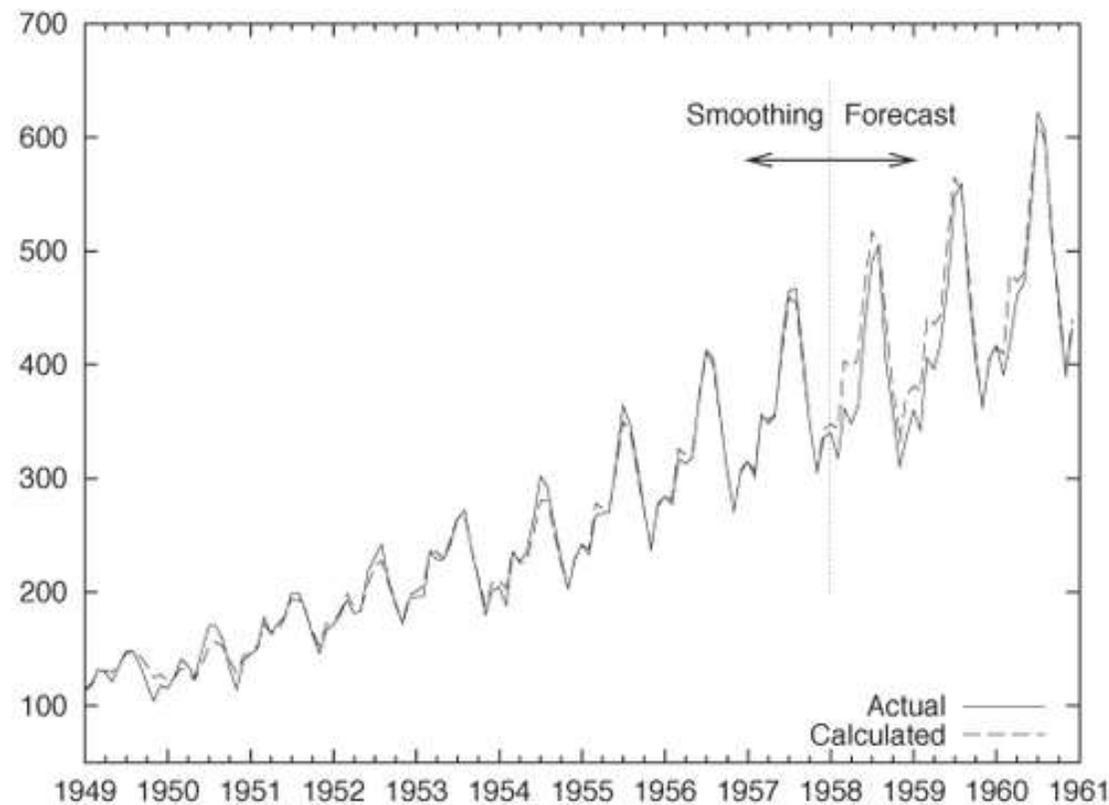
$$x_{i+h} = (s_i + ht_i)p_{i-k+h}$$



# Filtrado de series temporales



## Suavizado exponencial [exponential smoothing]



Número mensual de pasajeros (en miles).



# Filtrado de series temporales

## Suavizado exponencial [exponential smoothing]

For more competitions, visit [kaggle.com](#) [Sign in](#)

**kaggle** *in Class* [Sign Up](#) [About](#) [How it works](#) [Competitions](#) [Create a competition](#) [Blog](#) [Kaggle](#)

### First International Competition of Time Series Forecasting

PRIZE POOL: **Kudos** TEAMS: **16** COMPLETED: **6 days ago**

Information Data Forum **Results**

#### Private Leaderboard

This competition has completed, this leaderboard reflects the preliminary final standings. The results will become final after the competition organizers verify the results. <sup>\*</sup> in the money

#	Δ1w	Team Name	SMAPE	Entries	Last Submission UTC (Best Submission - Last)
1	↑2	NT *	13.82165	16	Fri, 10 Feb 2012 00:02:18 (-6d)
2	↓1	DuckTite	14.99062	32	Fri, 10 Feb 2012 15:58:28
3	↓1	woobe	15.59030	24	Sun, 05 Feb 2012 01:42:24
4	-	bobo	16.78477	26	Fri, 10 Feb 2012 21:19:10 (-14.8d)
7	↓1	<b>boder</b>	<b>21.91173</b>		
9	↓2	Forbin	23.09967	4	Thu, 26 Jan 2012 23:56:07 (-19.2d)
10	↓1	JCW	27.31540	3	Tue, 03 Jan 2012 08:38:39 (-23.8h)
11	↓1	Dennis Jaheruddin	33.68753	1	Mon, 05 Dec 2011 12:10:16
	↓1	Last Observed Value Benchmark	34.39842		
13	↓1	TT	42.16377	1	Mon, 05 Dec 2011 11:56:48
14	↓1	J O	128.83499	1	Mon, 19 Dec 2011 07:48:40
14	new	chwolfe	128.83499	2	Fri, 10 Feb 2012 15:27:40 (-21.1h)
	↓2	All zeros	200.00000		

16 teams as of Fri, 10 Feb 2012 23:59:00

Sólo ajustando 3 parámetros se obtiene un "buen" modelo...

**7º puesto**

International Competition on Time Series Forecasting ICTSF'2012

SMAPE = 21.91

vs. 13.82 (ganador)

vs. 34.40 (benchmark)



# Técnicas de regresión



La predicción (numérica) es...

- Similar a la clasificación:
  - Se construye un modelo a partir de un conjunto de entrenamiento.
  - Se utiliza el modelo para predecir el valor de una variable (continua u ordenada).
- Diferente a la clasificación:
  - El modelo define una función continua.

Método más empleado: **Regresión**



# Técnicas de regresión



Las técnicas de regresión modelan la relación entre una o más variables independiente (predictores) y una variable dependiente (variable de respuesta).

## Métodos de regresión

- Regresión lineal
- Regresión no lineal
- Árboles de regresión (p.ej. CART)
- ...



# Técnicas de regresión



## Regresión lineal simple

Una única variable independiente:

$$y = w_0 + w_1 x$$

donde  $w_0$  (desplazamiento) y  $w_1$  (pendiente) son los coeficientes de regresión.

### ■ Método de los mínimos cuadrados

(estima la línea recta que mejor se ajusta a los datos):

$$w_0 = \bar{y} - w_1 \bar{x} \quad w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2}$$

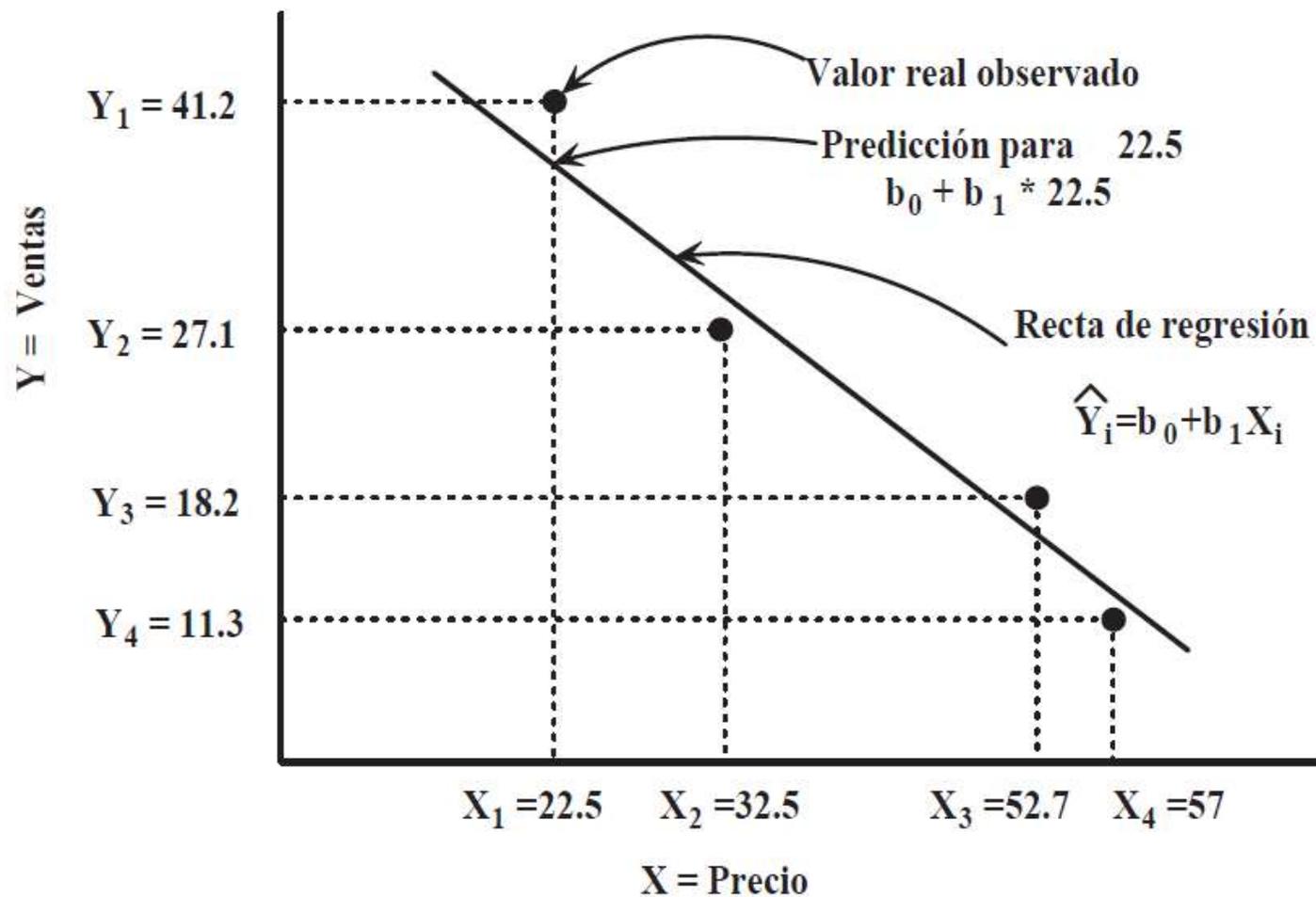


# Técnicas de regresión



## Regresión lineal simple

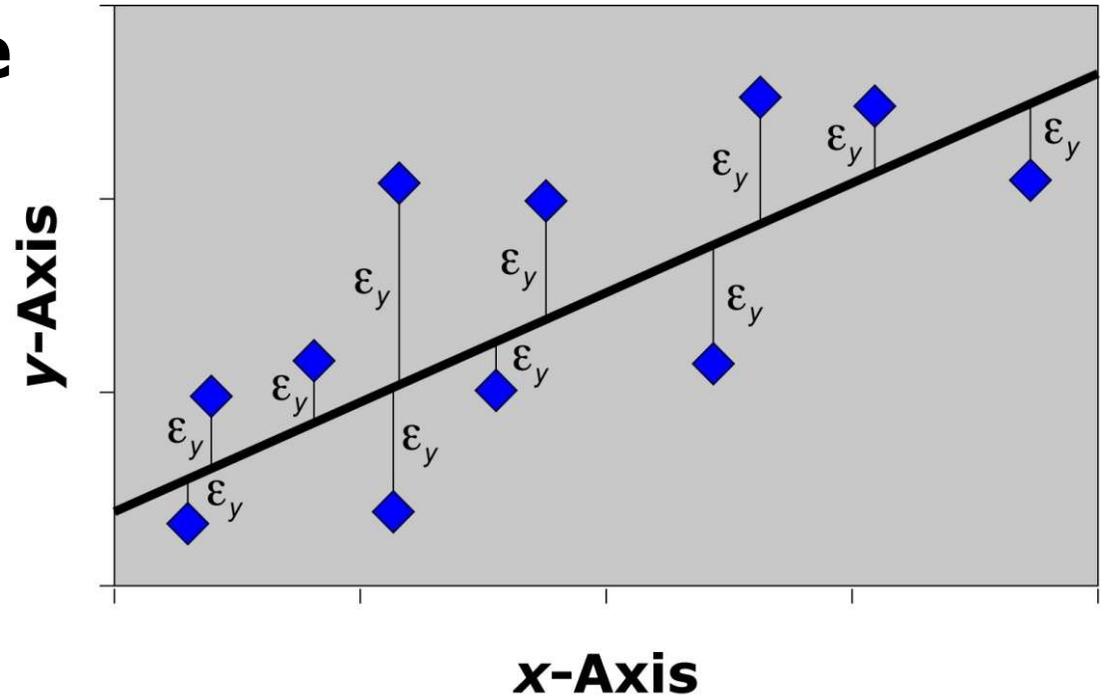
$$\hat{Y}_i = b_0 + b_1 X_i$$



# Técnicas de regresión



## Regresión lineal simple



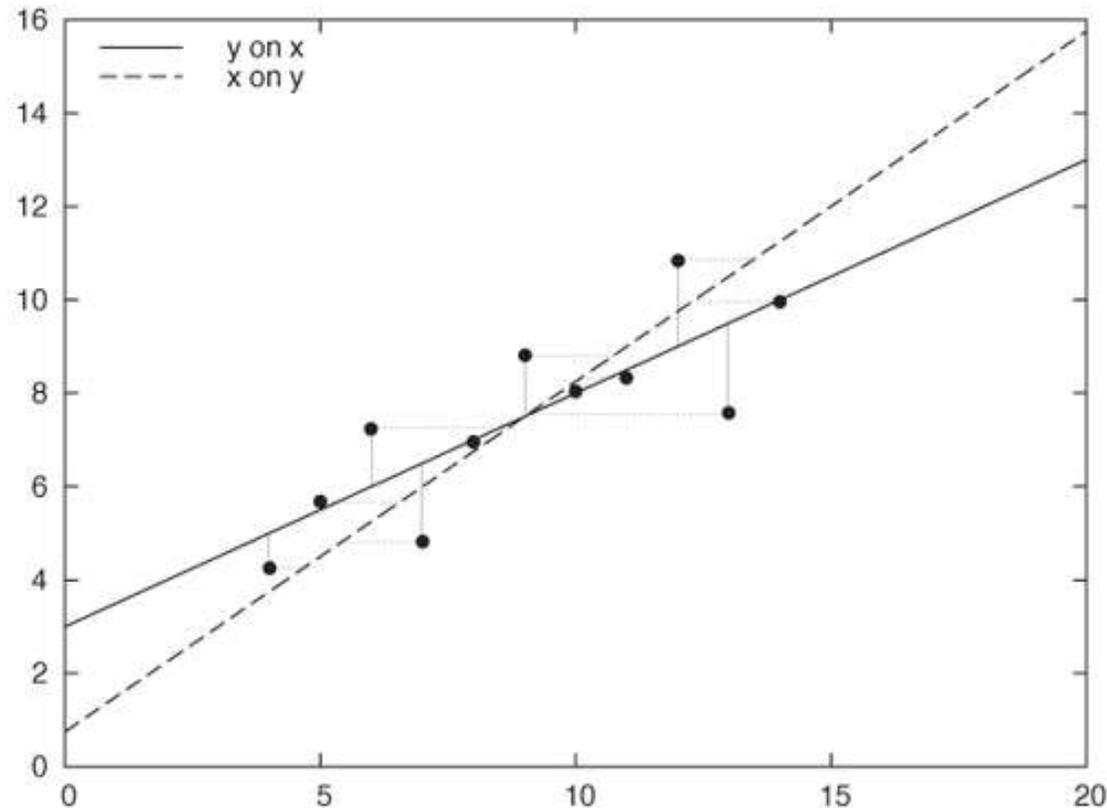
El método de los mínimos cuadrados minimiza la suma de los cuadrados de los residuos  $\varepsilon_i$  (las diferencias entre las predicciones y los valores observados).



# Técnicas de regresión



## Regresión lineal simple



¡OJO! Al utilizar regresión lineal, la recta  $y=f(x)$  que se obtiene es distinta a la que obtenemos si  $x=f(y)$ .



# Técnicas de regresión



## Regresión lineal múltiple

Varias variables independientes:

$$y = w_0 + w_1 x_1 + w_2 x_2 + \dots$$

- Resoluble por métodos numéricos de optimización.
- Muchas funciones no lineales pueden transformarse en una expresión lineal.

p.ej. Un modelo de regresión polinomial

$$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

puede transformarse en un modelo lineal definiendo las variables  $x_2 = x^2$ ,  $x_3 = x^3$ :

$$y = w_0 + w_1 x + w_2 x_2 + w_3 x_3$$



# Técnicas de regresión



## Regresión lineal

Condiciones necesarias para aplicar regresión lineal:

- Obviamente, la muestra ha de ser aleatoria.
- El tipo de dependencia descrita ha de ser lineal.
- Fijado un valor de la(s) variable(s) independiente(s), la variable dependiente se distribuye según una distribución normal.
- Los errores han de tener la misma varianza (nube de puntos homogénea).

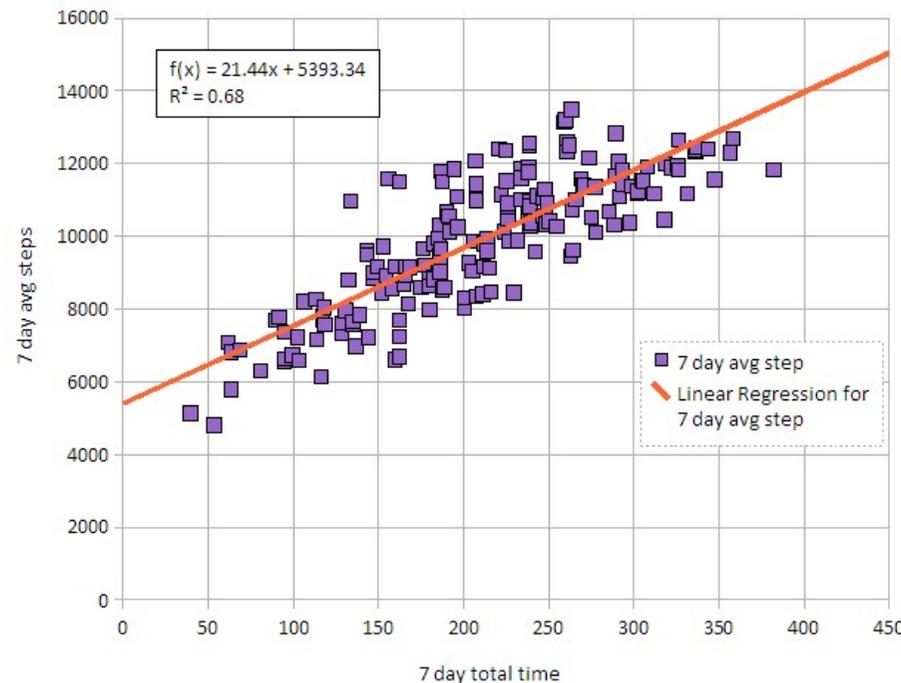


# Técnicas de regresión



## Regresión lineal simple

1. Mediante un diagrama de dispersión, comprobamos visualmente si existe una relación lineal entre las variables X (predictor) e Y (respuesta):



# Técnicas de regresión



## Regresión lineal simple

2. Cuantificamos la relación construyendo la recta que resume la dependencia y damos una medida de cómo se ajusta la recta a los datos (correlación):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \in [-1, 1]$$



# Técnicas de regresión



## Coeficiente de correlación

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \in [-1, 1]$$

**r=+1** Dependencia lineal total en sentido positivo (cuanto mayor es X, mayor es Y).

**r=-1** Dependencia lineal total en sentido negativo (cuanto mayor es X, menor es Y).



# Técnicas de regresión



## Coeficiente de correlación

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \in [-1, 1]$$

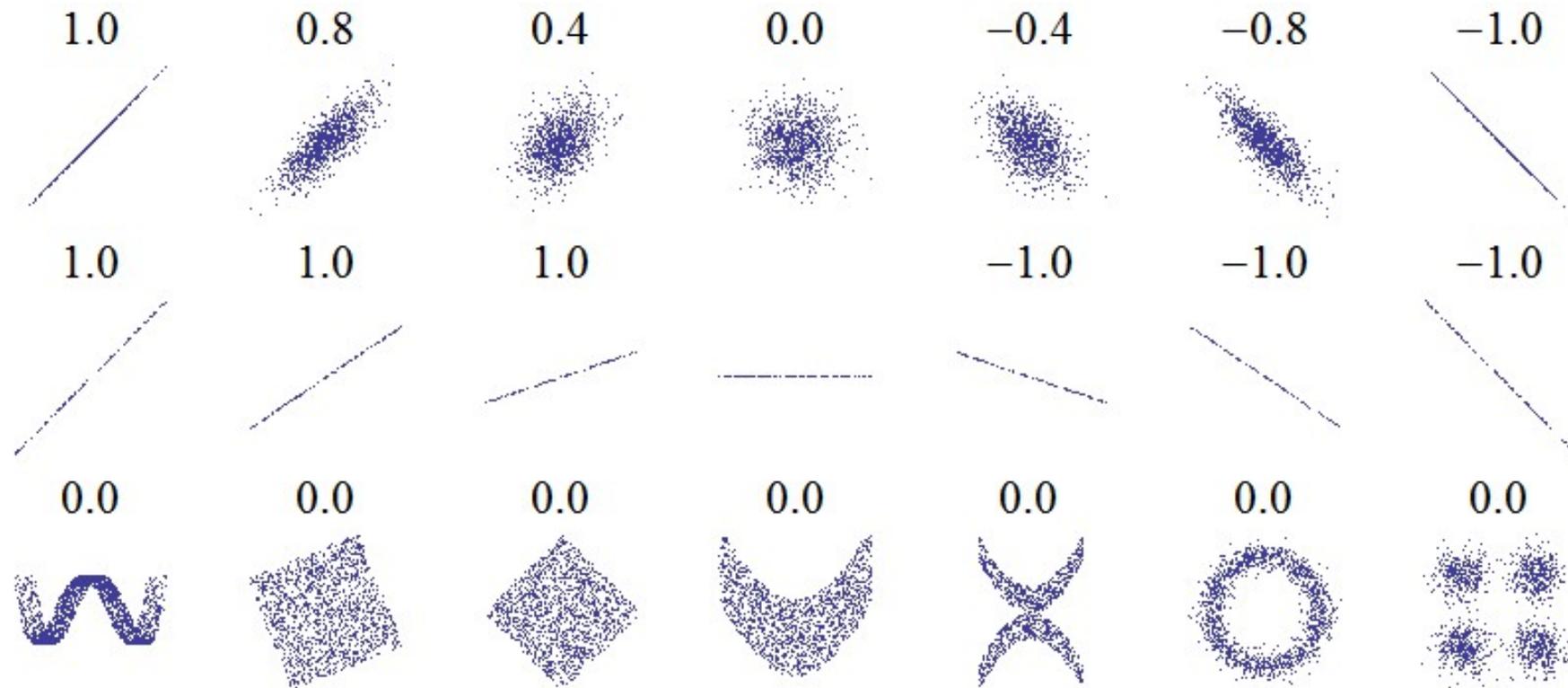
- $r > 0$**  Existe una dependencia positiva.  
Cuanto más se acerque a 1, mayor es ésta.
- $r < 0$**  Existe una dependencia negativa.  
Cuanto más se acerque a -1, mayor será.
- $r = 0$**  No podemos afirmar nada.



# Técnicas de regresión



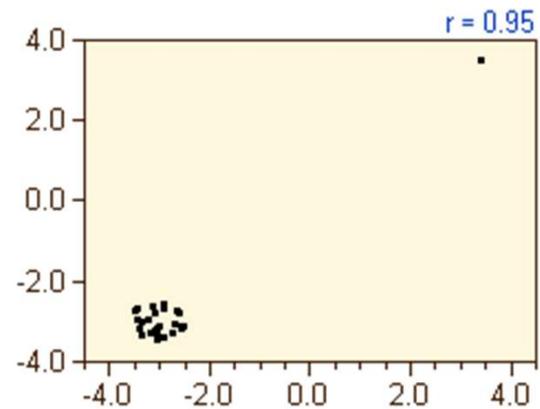
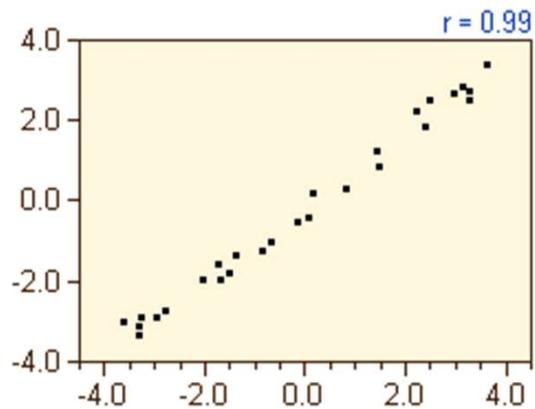
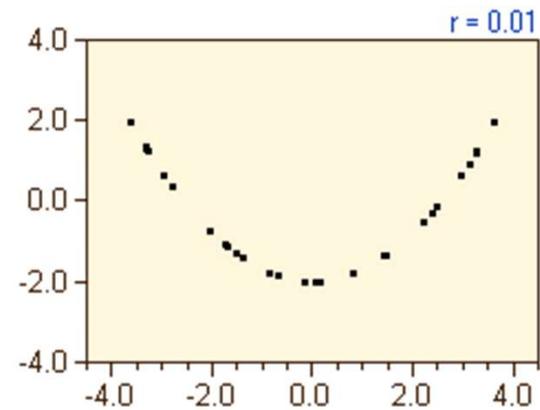
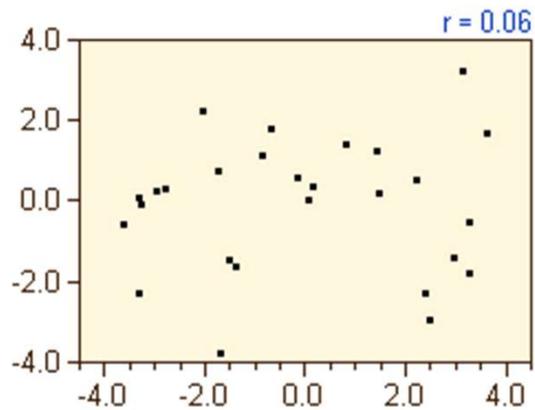
## Coeficiente de correlación



# Técnicas de regresión



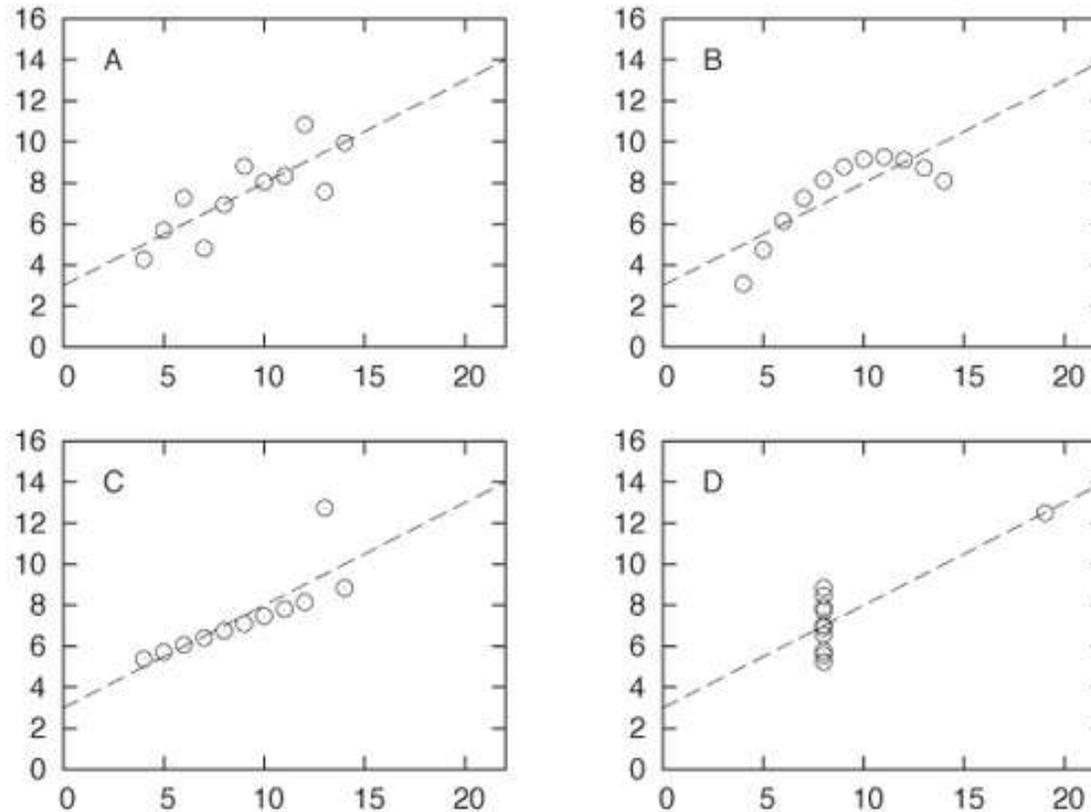
## Coeficiente de correlación



# Técnicas de regresión



## Coeficiente de correlación



El cuarteto de Anscombe

(4 conjuntos de datos con el mismo coeficiente de correlación)



# Técnicas de regresión



## Coeficiente de correlación

### Ventaja de $r$

- No depende de las unidades usadas en la medición.

### Limitaciones de $r$

- Sólo mide dependencia lineal entre las variables.

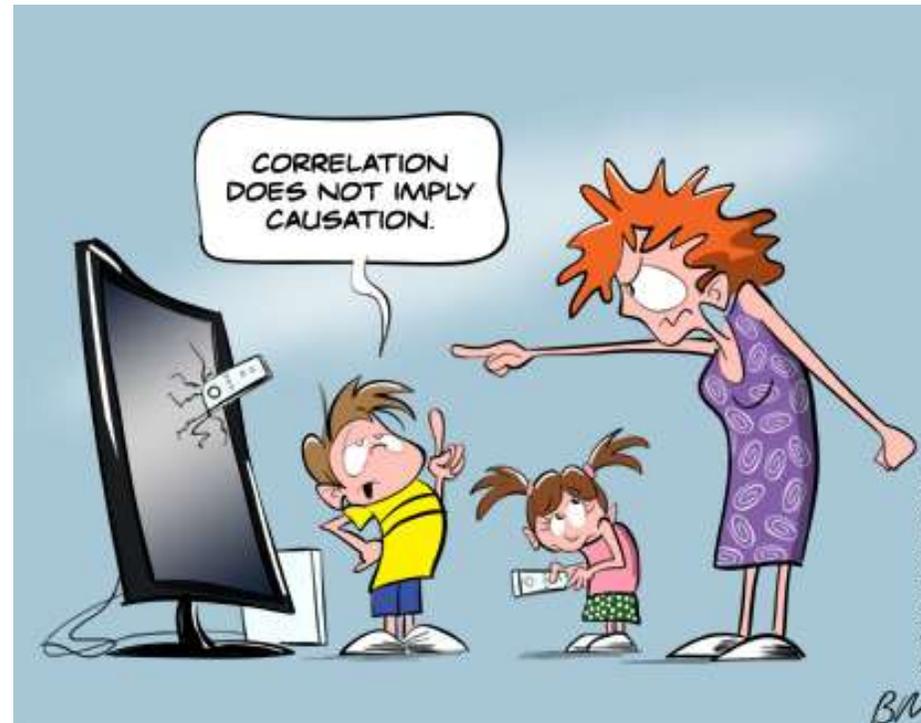
¡OJO! La correlación no implica causalidad...



# Técnicas de regresión



## Coeficiente de correlación



"Correlation is not causation but it sure is a hint."

-- Edward Tufte



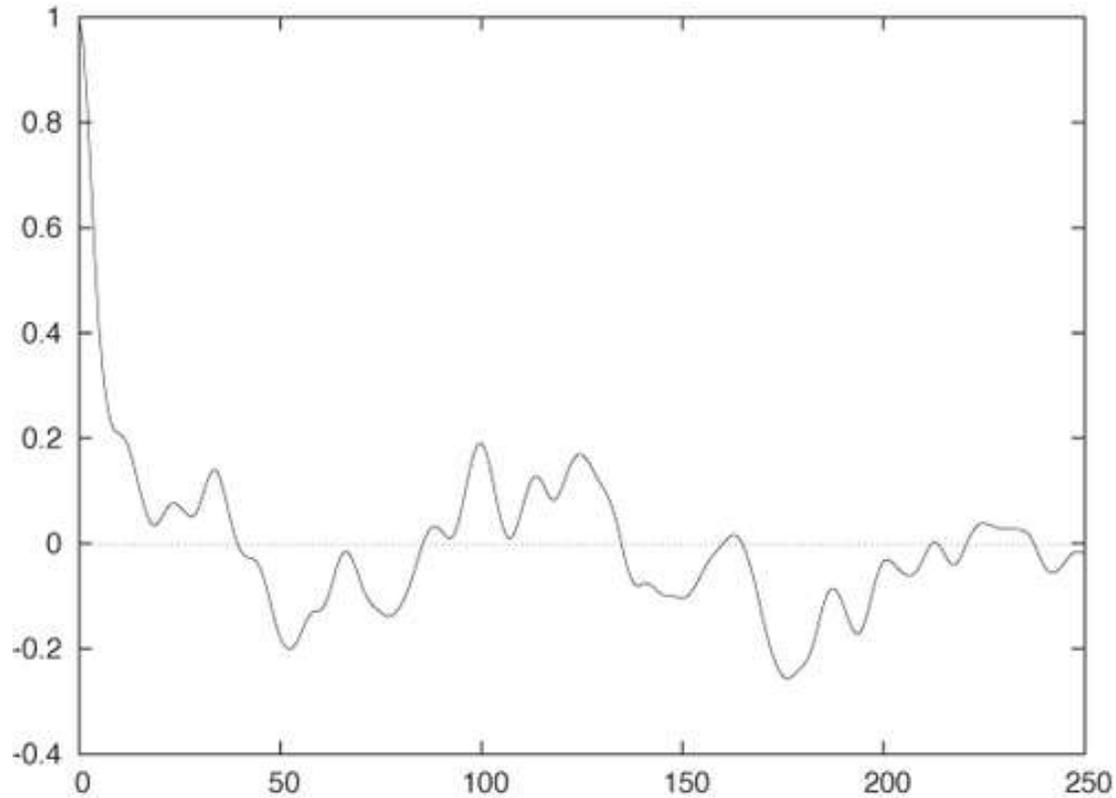
# Función de autocorrelación



$$c(k) = \frac{\sum_{i=1}^{N-k} (x_i - \mu)(x_{i+k} - \mu)}{\sum_{i=1}^N (x_i - \mu)^2} \quad \text{con} \quad \mu = \frac{1}{N} \sum_{i=1}^N x_i$$



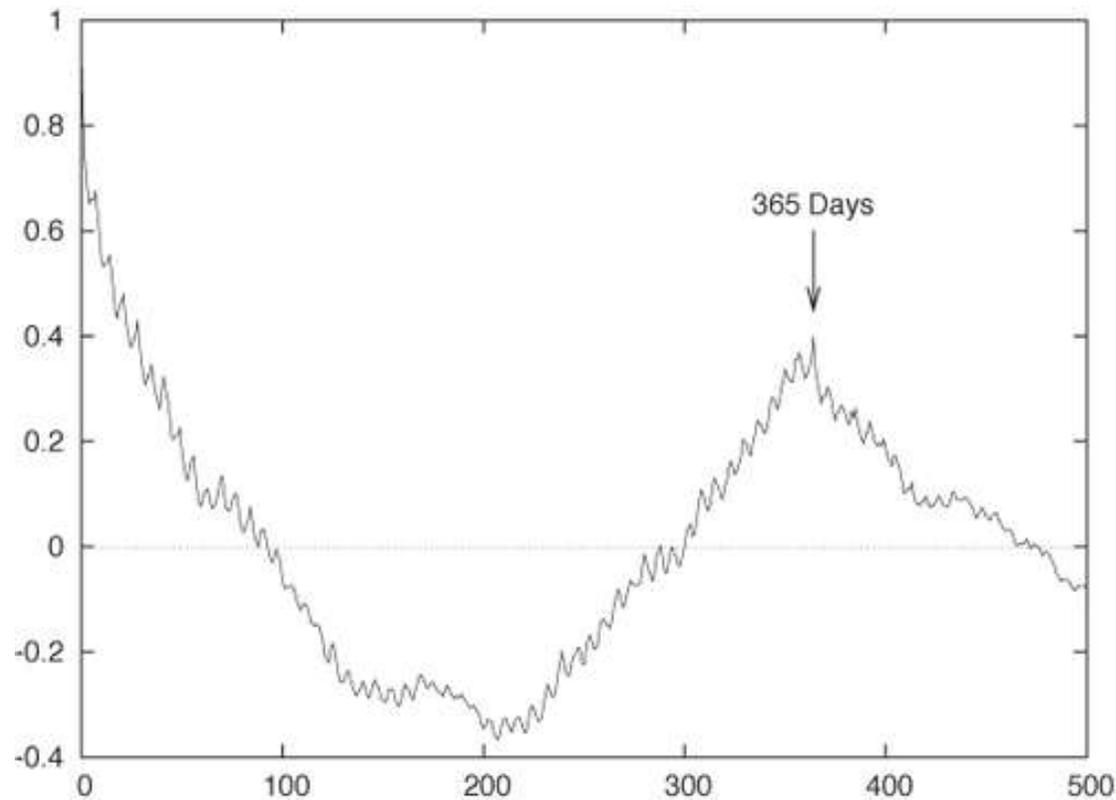
# Función de autocorrelación



Autocorrelación para la salida de gas de una caldera



# Función de autocorrelación



Autocorrelación en las llamadas a un call-center

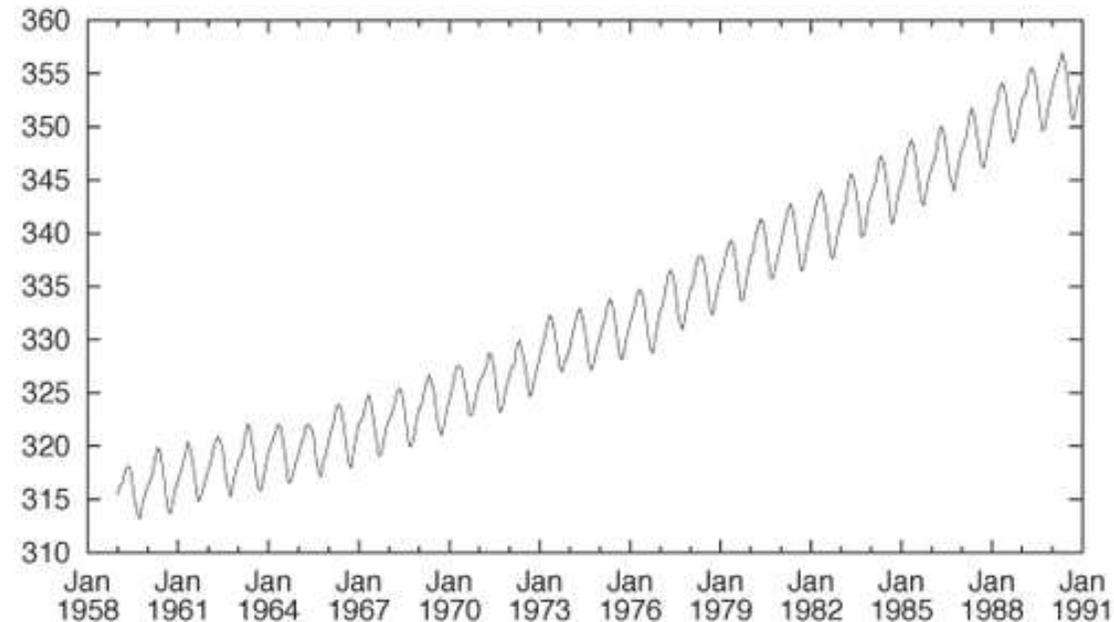


# Caso práctico



## Conjunto de datos

Mediciones de CO<sub>2</sub> en Mauna Loa (Hawaii)



Adaptado de Philipp K. Jannert:

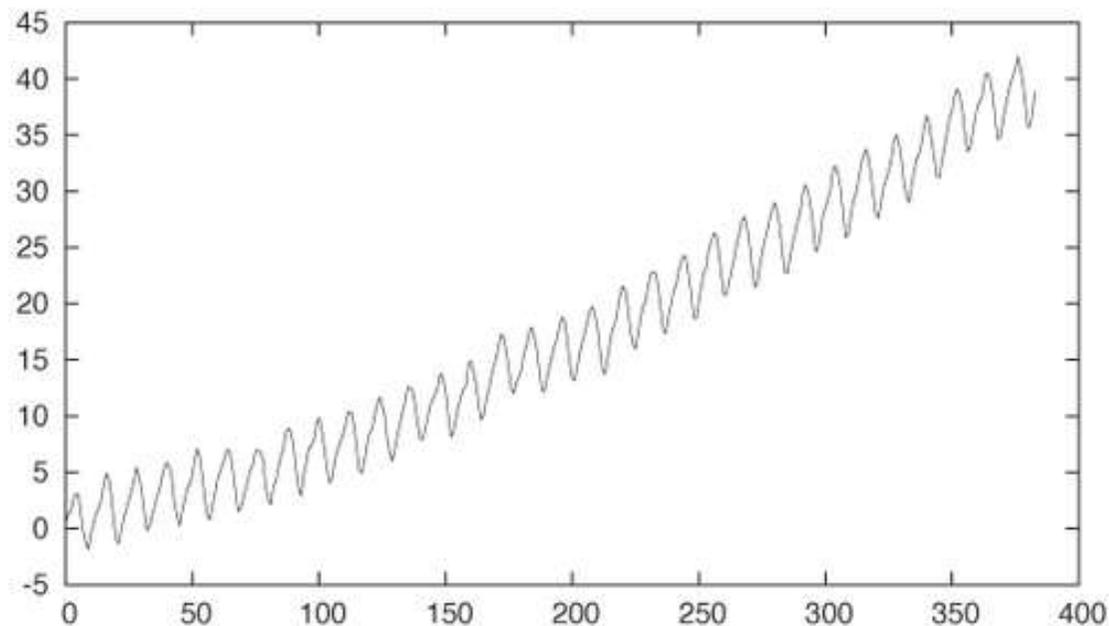
“Intermezzo: A Data Analysis Session” [capítulo 6]



# Caso práctico



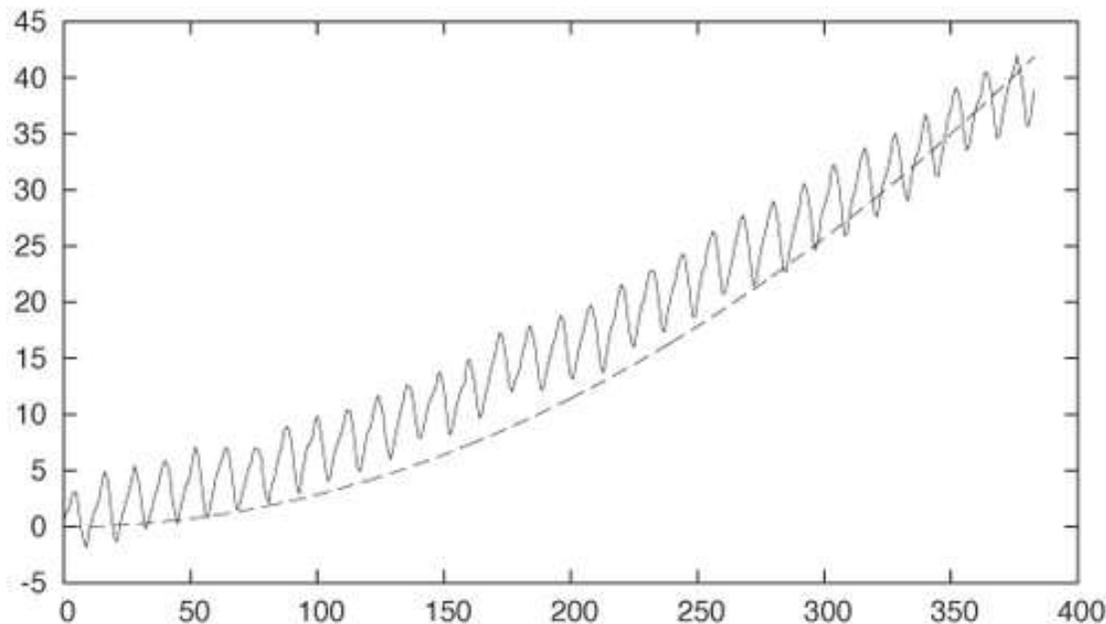
A partir de las mediciones mensuales (1959-1991), eliminamos las fechas del eje X y hacemos que la serie empiece de cero:



# Caso práctico



**Tendencia:** Apreciamos una tendencia no lineal:  
Intentamos ajustarla con una función de la forma  $x^k$   
Nota: Todas las curvas de ese tipo pasan por (0,0) y (1,1)



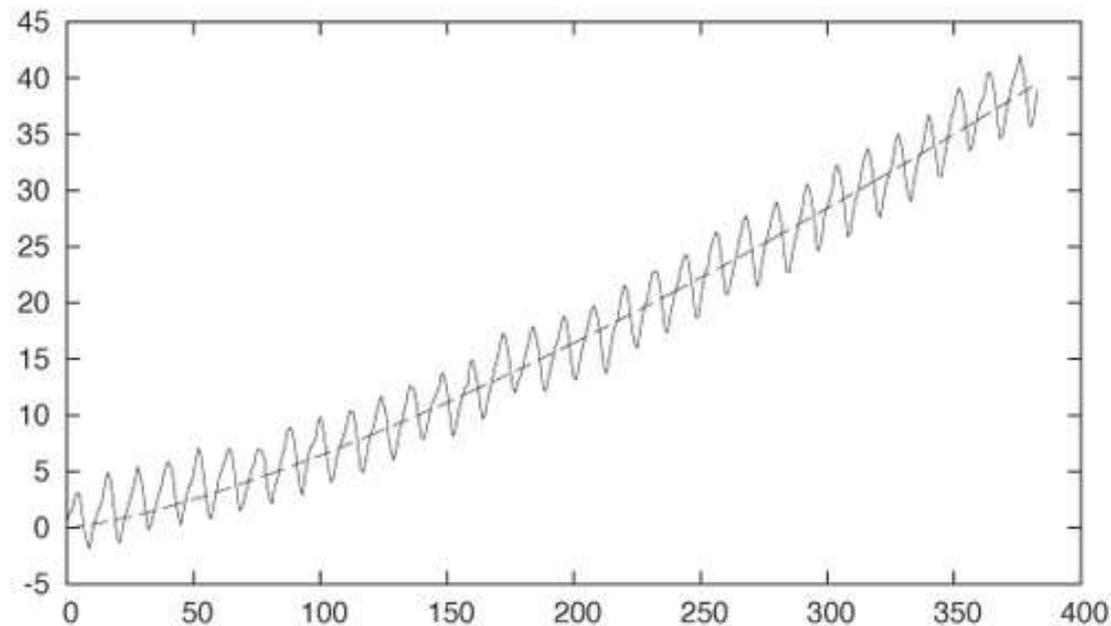
Con  $k=2$ , tenemos  $35 \cdot (x/350)^2$ , pero parece que nos hemos pasado...



# Caso práctico



Afinamos un poco más y usamos un valor menor:



OK!

$$k=1.35$$

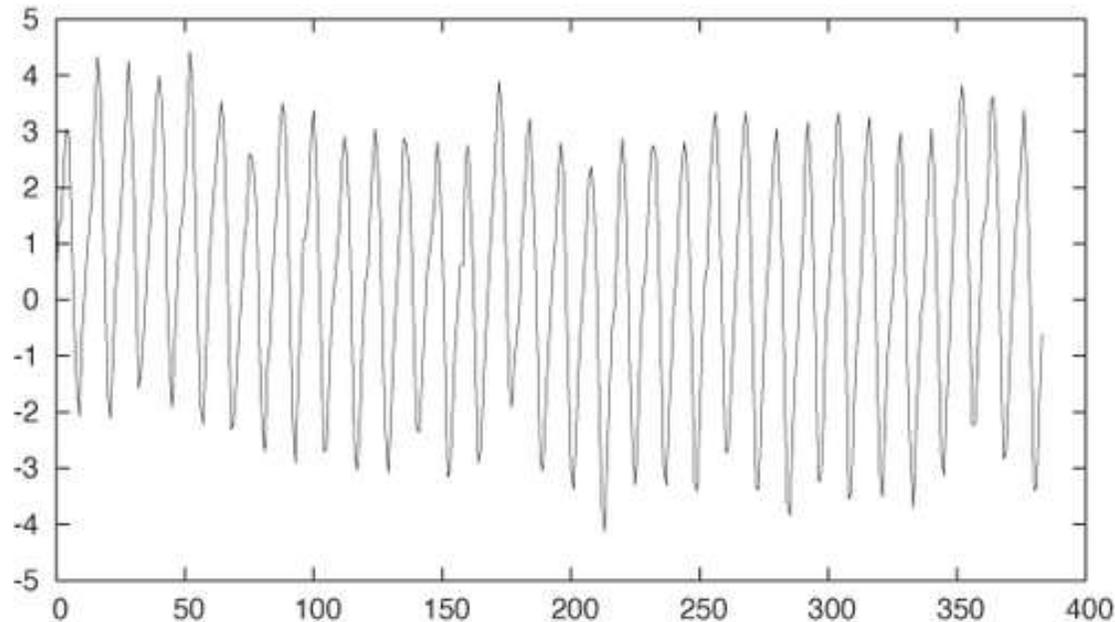
Ajuste de la función  $f(x) = 35 \cdot (x/350)^{1.35}$



# Caso práctico



Para comprobar que no vamos mal,  
calculamos los residuos (valor original – aproximación):



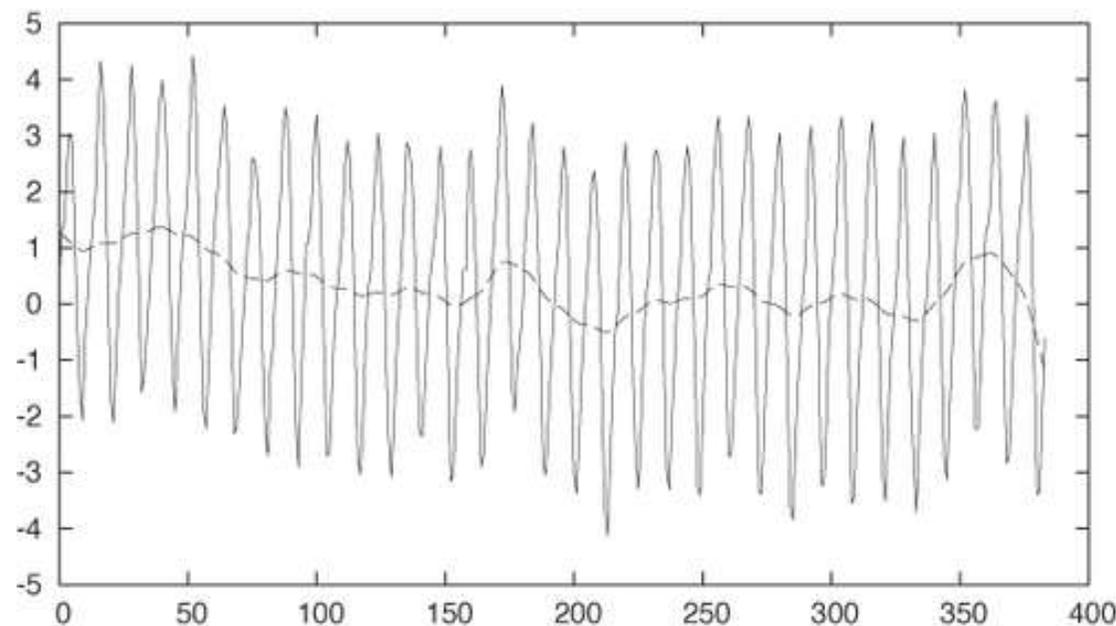
Residuos del ajuste  $f(x) = 35 \cdot (x/350)^{1.35}$



# Caso práctico



Si nuestro ajuste de la tendencia es correcto, los residuos no deben exhibir tendencia alguna (deberían aparecer balanceados en torno a  $y=0$ ):



Suavizamos los residuos para comprobar si aún existe algún tipo de tendencia en los residuos...

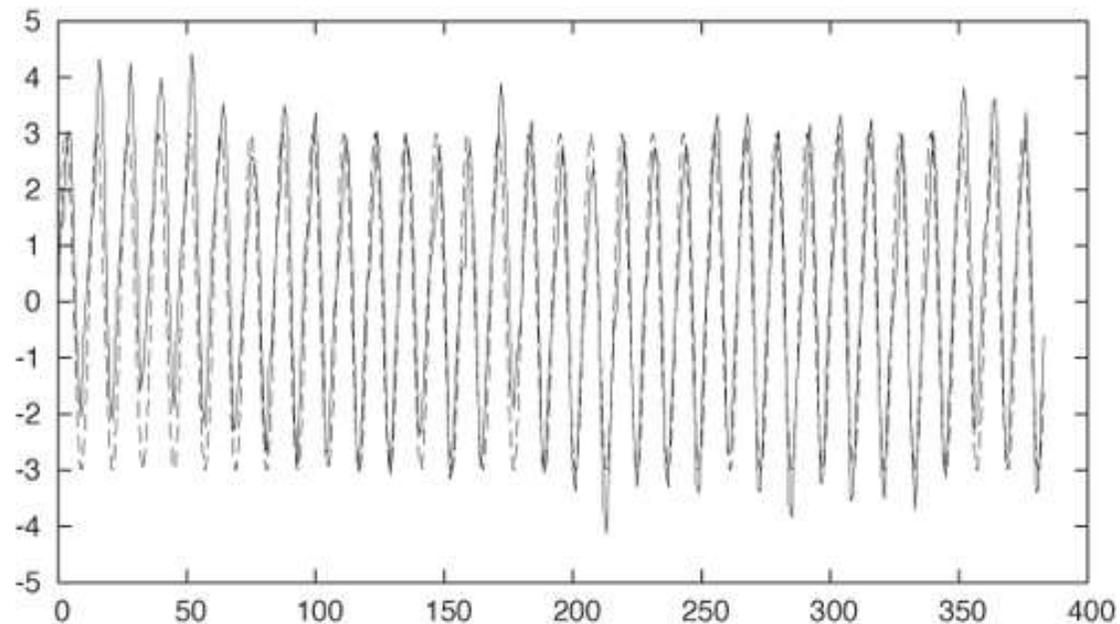


# Caso práctico



## Estacionalidad:

Apreciamos una periodicidad anual (cada 12 valores)



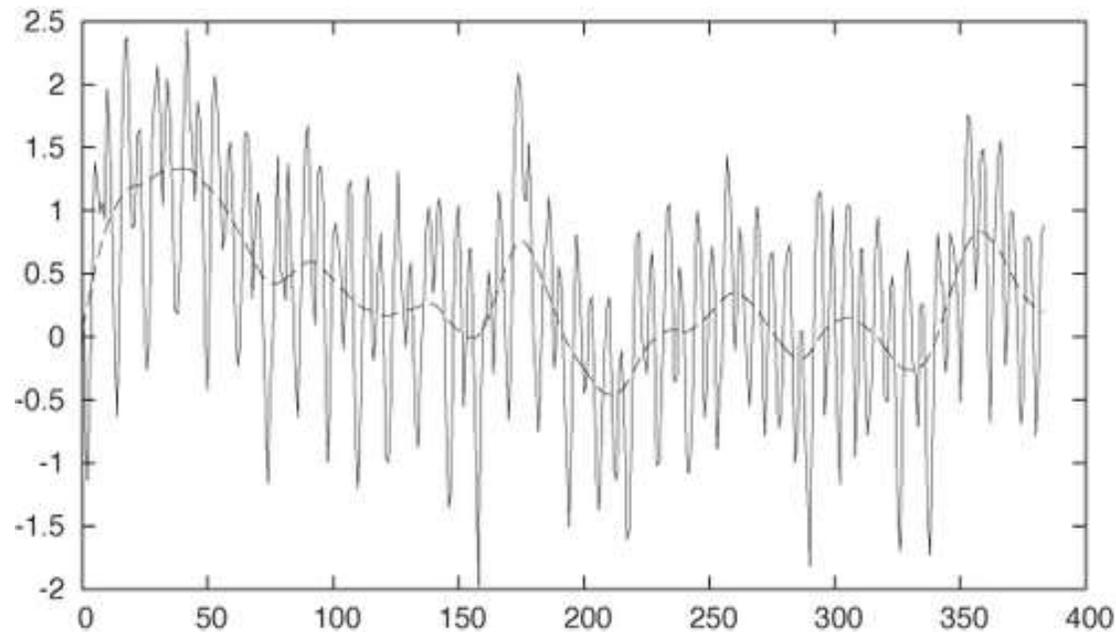
Ajustamos con una función senoidal  $3 \cdot \sin(2 \cdot \pi \cdot x / 12)$



# Caso práctico



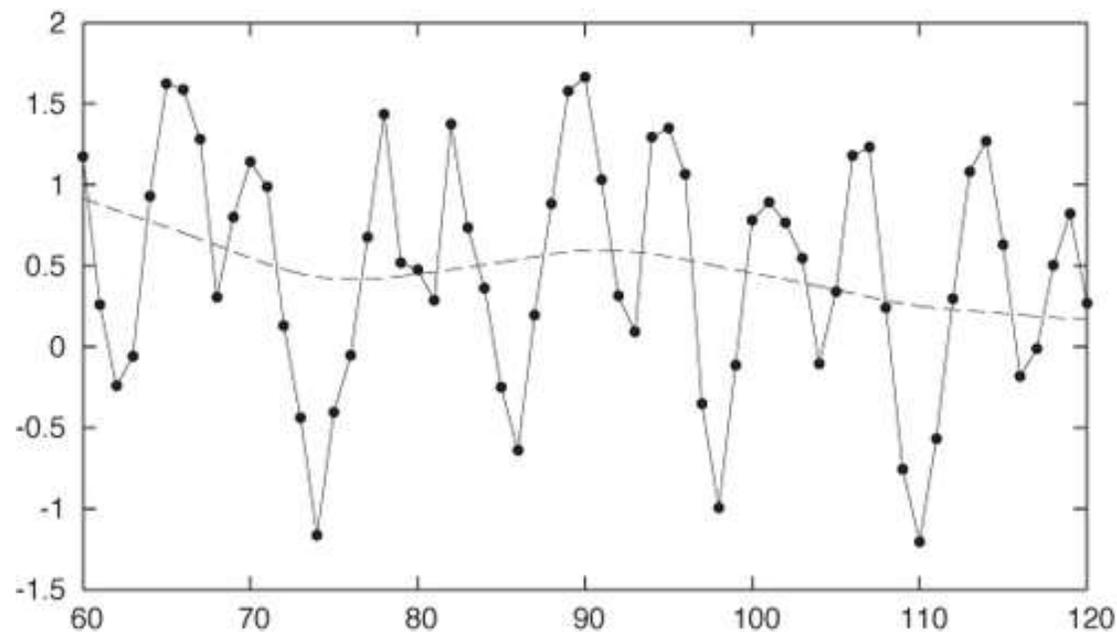
Calculamos los residuos tras nuestras aproximaciones  
(valor original – tendencia – estacionalidad)



# Caso práctico



En la figura anterior no se ve mucho... hacemos zoom:



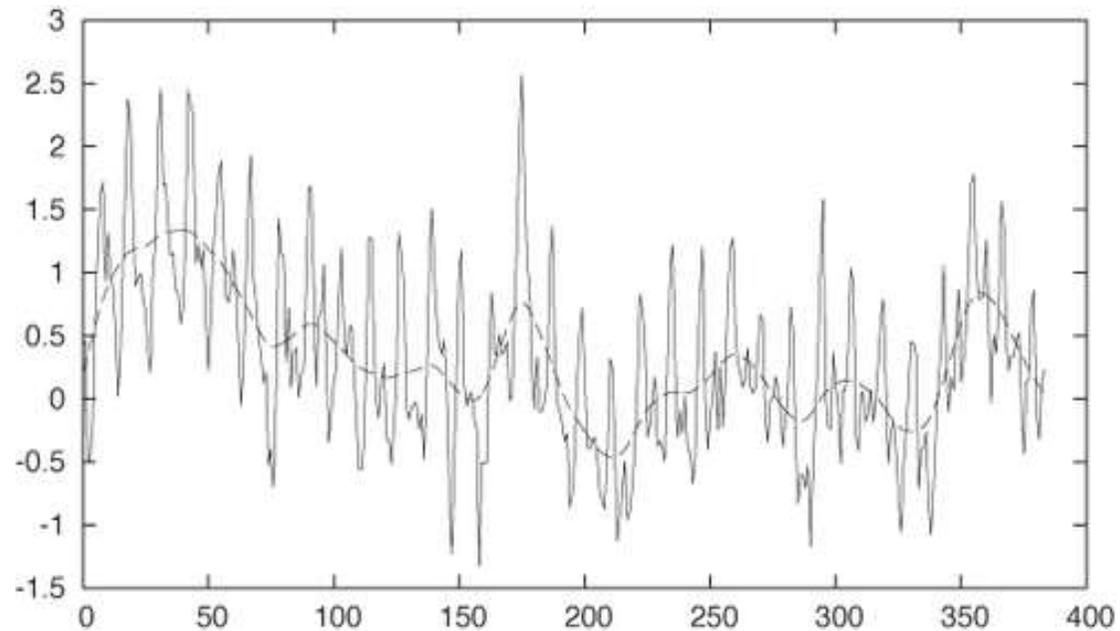
Se sigue apreciando cierta periodicidad, por lo que usamos un segundo armónico  $-0.75 \cdot \sin(2 \cdot \pi \cdot x / 6)$



# Caso práctico



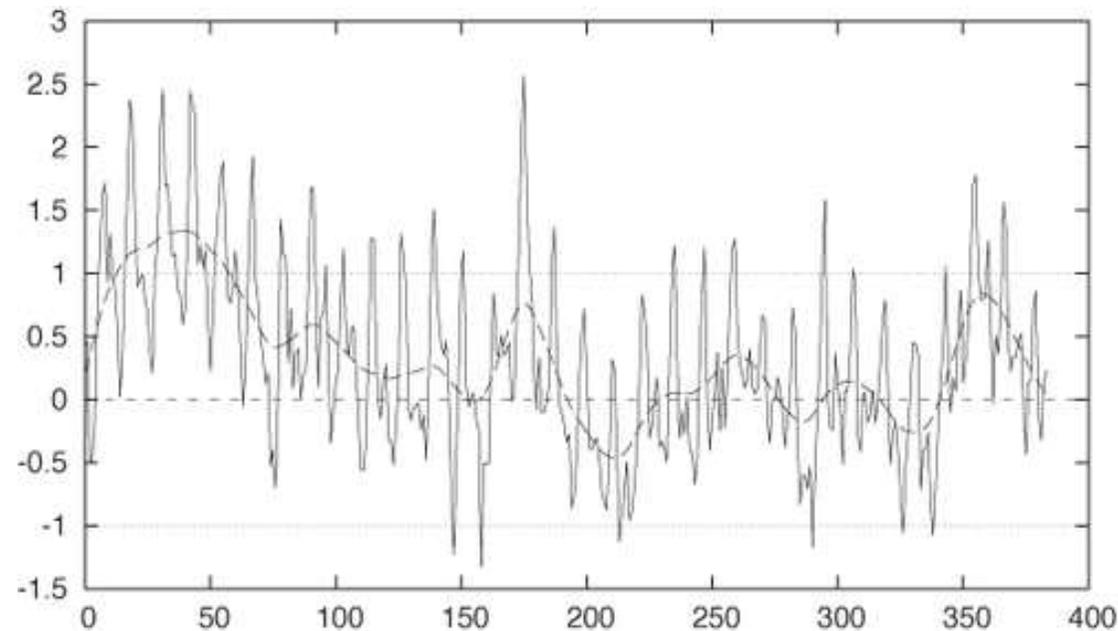
Residuos tras eliminar la tendencia y los dos primeros armónicos correspondientes a la estacionalidad:



# Caso práctico



Añadimos líneas que nos ayuden a ver si los residuos están sesgados:



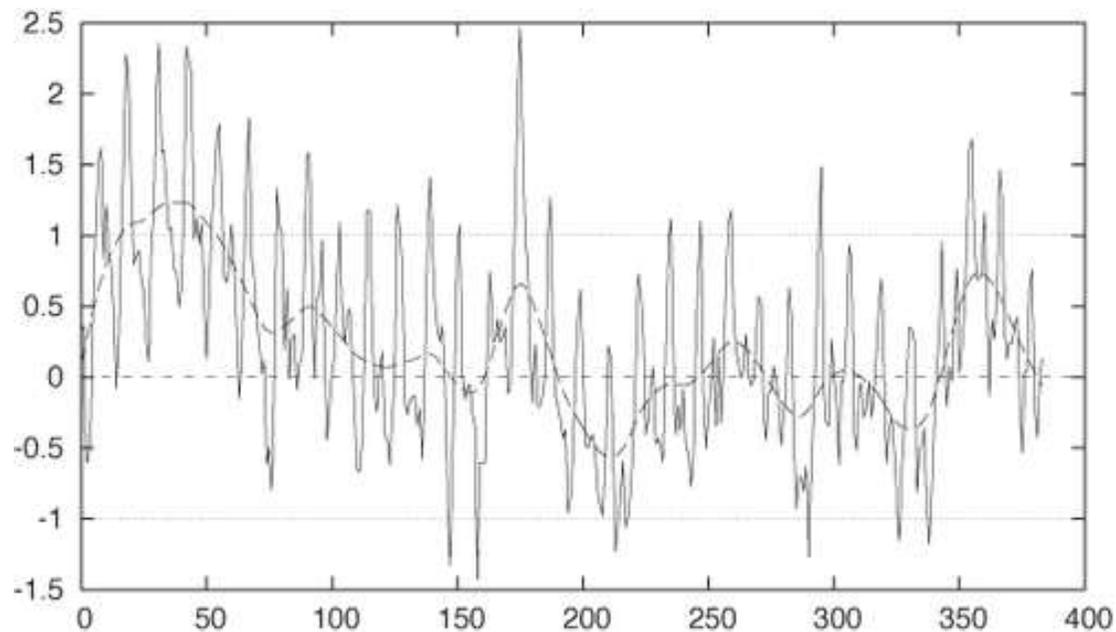
Parece sesgado hacia arriba,  
por lo que añadimos un desplazamiento de  $+0.1$



# Caso práctico



Los residuos de nuestra aproximación final:



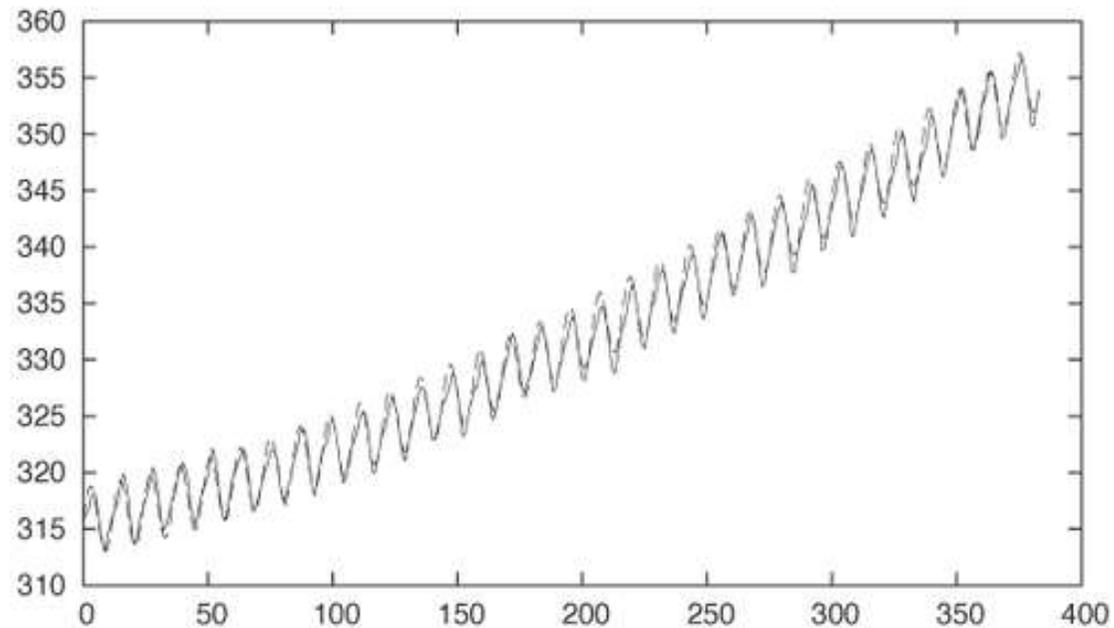
$$\begin{aligned} f(x) = & 315 + 35*(x/350)**1.35 \\ & + 3*\sin(2*pi*x/12) - 0.75*\sin(2*pi*x/6) \\ & + 0.1 \end{aligned}$$



# Caso práctico



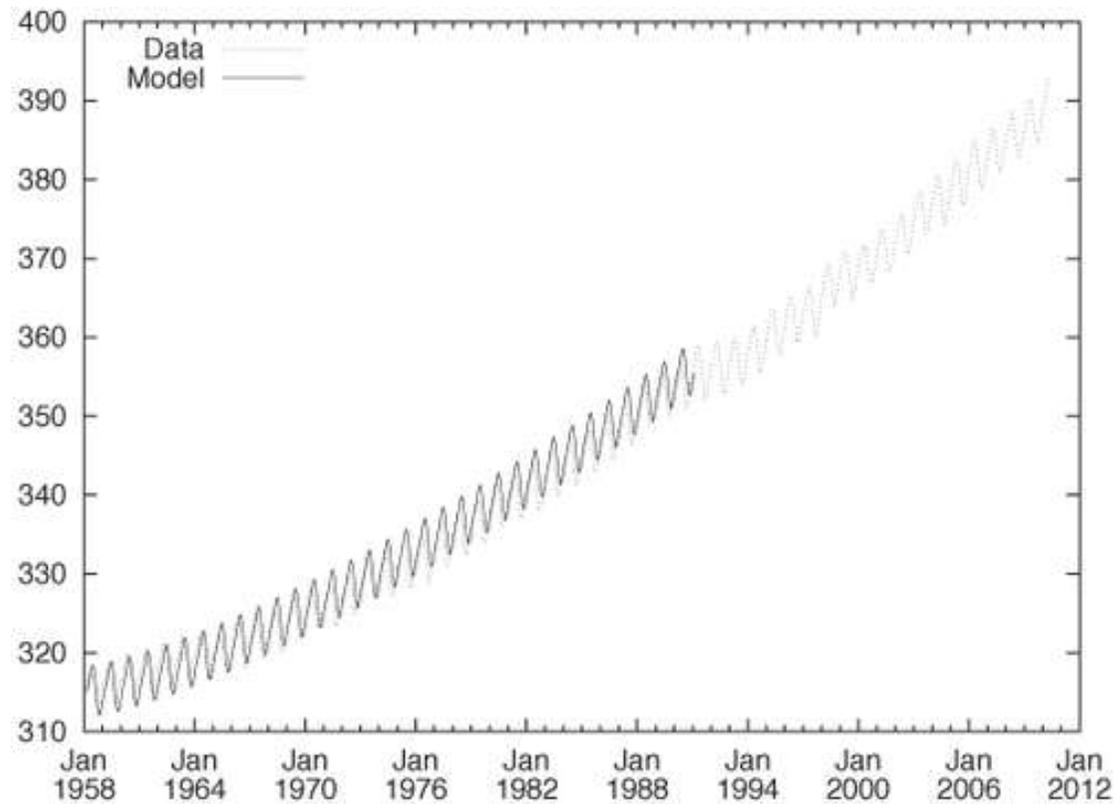
El ajuste que hemos realizado (1959-1990):



# Caso práctico



Nuestra predicción del futuro (1991-2010)



# Predicción de series temporales

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----



**skforecast**



# Predicción de series temporales

Time series

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

Exogenous variable

a	b	c	d	e	f	g	h	i	j
---	---	---	---	---	---	---	---	---	---

X

y

1	2	3	4	5	f	6
2	3	4	5	6	g	7
3	4	5	6	7	h	8
4	5	6	7	8	i	9
5	6	7	8	9	j	10

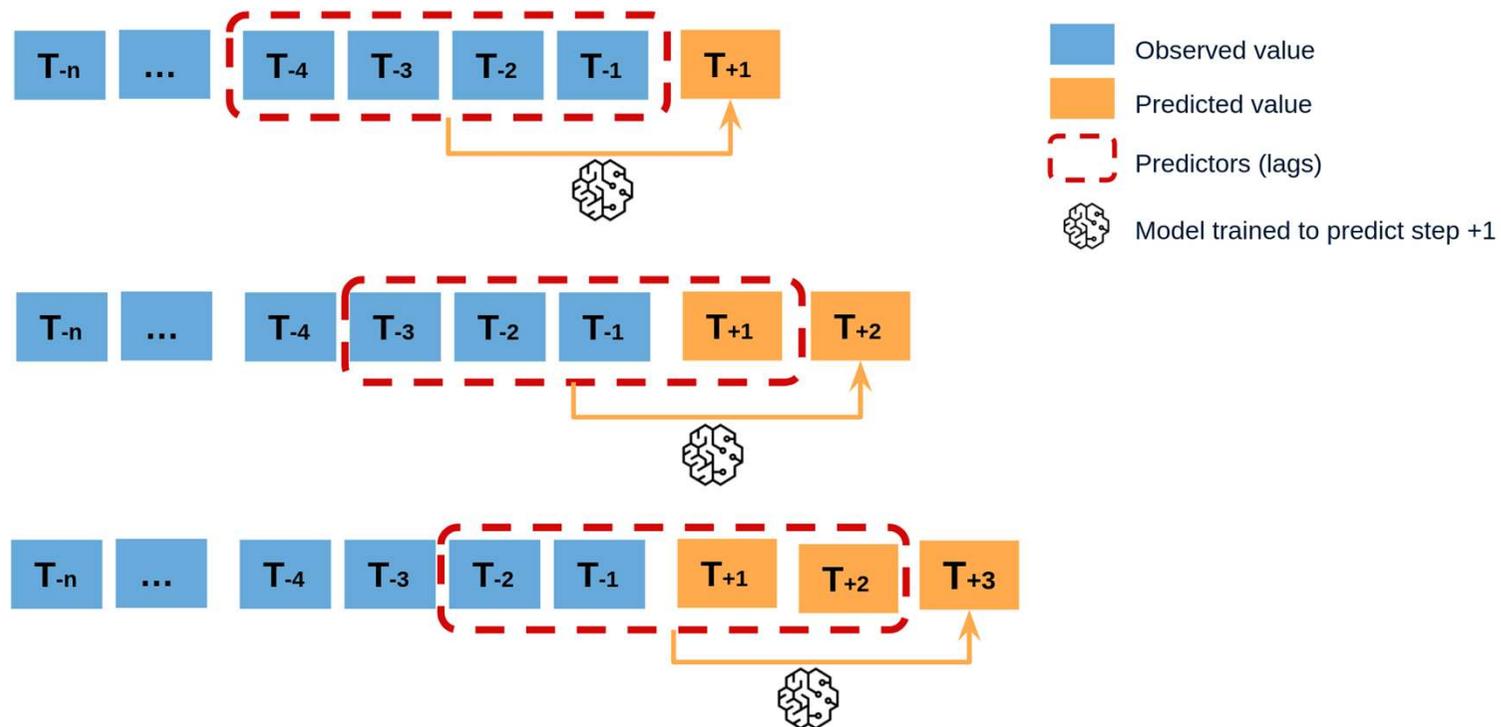


skforecast



# Predicción de series temporales

## Recursive multi-step forecasting

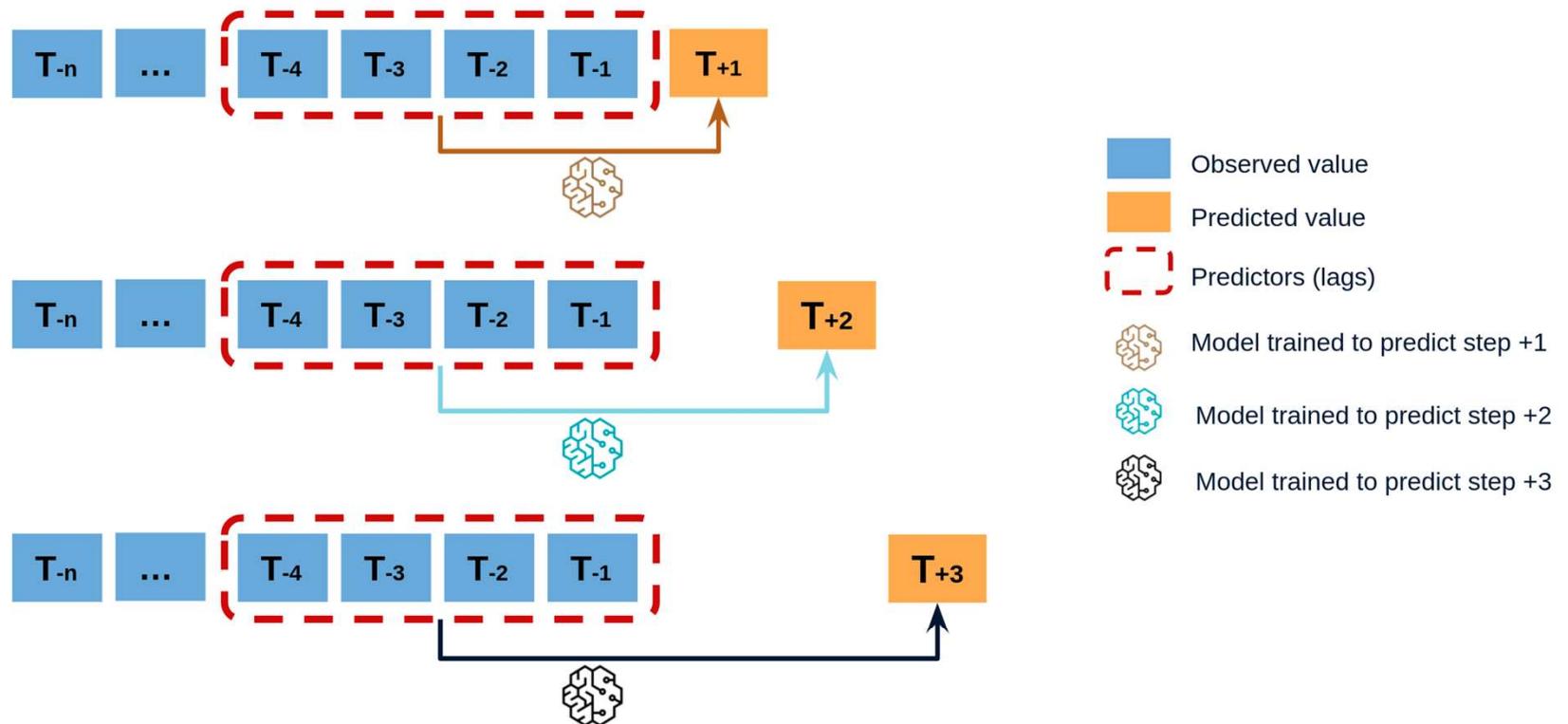


skforecast



# Predicción de series temporales

## Direct multi-step forecasting



skforecast



# Predicción de series temporales

y	1	2	3	4	5	6	7	8	9
exog 1	A	B	C	D	E	F	G	H	I
exog 2	a	b	c	d	e	f	g	h	i

lags=3, steps=2, create\_train\_X\_y()

y_train		X_train						
step 1	step 2	X_lags			X_exog_1		X_exog_2	
		lag 1	lag 2	lag 3	step 1	step 2	step 1	step 2
4	5	3	2	1	D	E	d	e
5	6	4	3	2	E	F	e	f
6	7	5	4	3	F	G	f	g
7	8	6	5	4	G	H	g	h
8	9	7	6	5	H	I	h	i

filter\_train\_X\_y\_for\_step(step=1)

filter\_train\_X\_y\_for\_step(step=2)

y_train	X_train				
step 1	X_lags			X_exog_1	X_exog_2
	lag 1	lag 2	lag 3	step 1	step 1
4	3	2	1	D	d
5	4	3	2	E	e
6	5	4	3	F	f
7	6	5	4	G	g
8	7	6	5	H	h

Training data for model of step 1

y_train	X_train				
step 2	X_lags			X_exog_1	X_exog_2
	lag 1	lag 2	lag 3	step 2	step 2
5	3	2	1	E	e
6	4	3	2	F	f
7	5	4	3	G	g
8	6	5	4	H	h
9	7	6	5	I	i

Training data for model of step 2

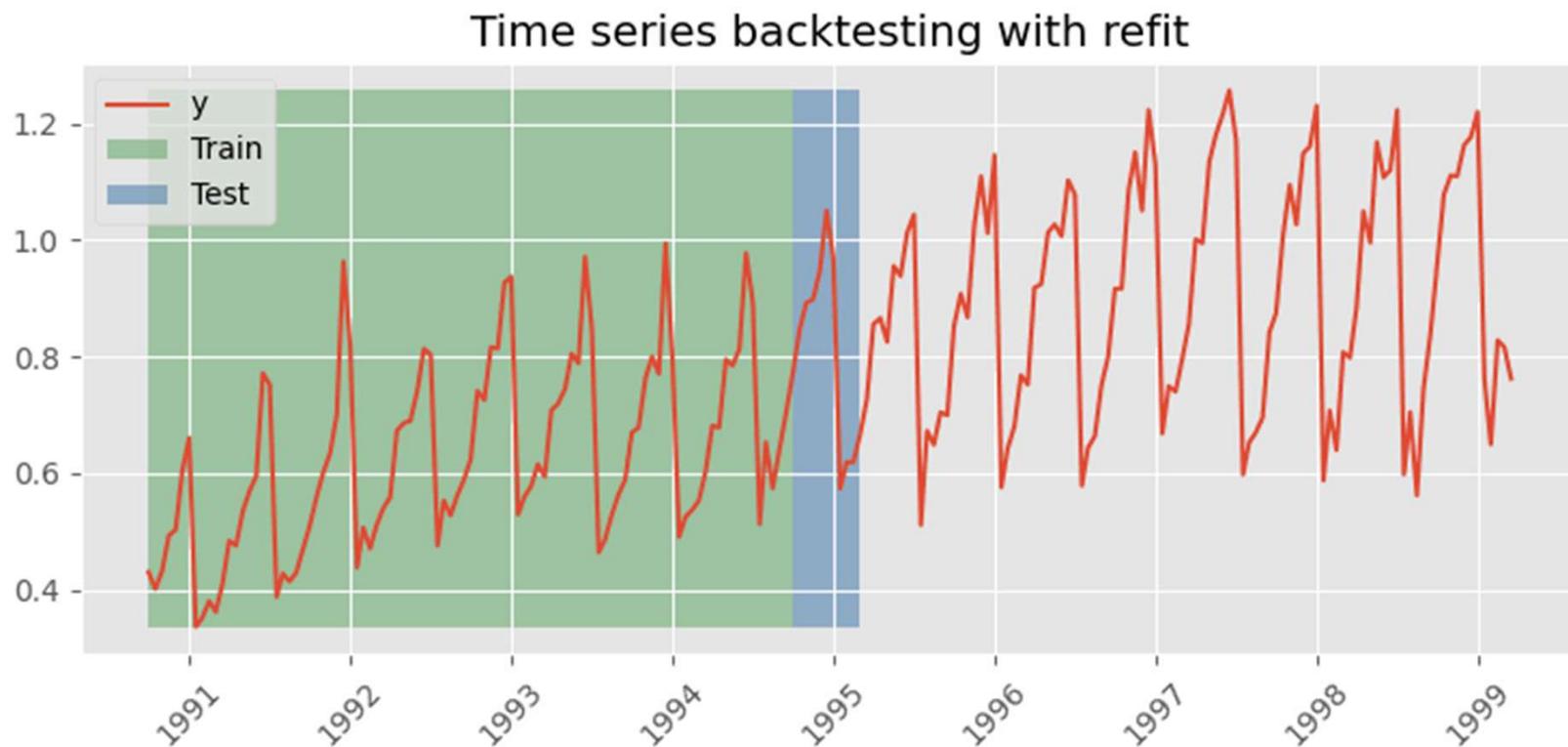


skforecast



# Predicción de series temporales

**Backtesting** (evaluación del rendimiento del modelo):



skforecast



# Predicción de series temporales

**Backtesting** (evaluación del rendimiento del modelo):

Métricas de evaluación

- MAE [Mean Absolute Error]

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|$$

- MSE [Mean Square Error]

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2$$



# Predicción de series temporales

**Backtesting** (evaluación del rendimiento del modelo):

Métricas de evaluación

- RMSE [Root Mean Square Error]

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2}$$

- MAPE [Mean Absolute Percentage Error]

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{|y_t|} \times 100\%$$



# Predicción de series temporales

**Backtesting** (evaluación del rendimiento del modelo):

Métricas de evaluación

- sMAPE [Symmetric MAPE]  $\in [0\%, 200\%]$

$$sMAPE = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{\frac{|y_t| + |\hat{y}_t|}{2}} \times 100\%$$

- MASE [Mean Absolute Scaled Error]

$$MASE = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{Q} \quad Q = \frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|$$



# Predicción de series temporales



# Predicción de series temporales



## Forecasting

<http://en.wikipedia.org/wiki/Forecasting>



# Modelos estadísticos



Modelos autorregresivos

## **AR(p)**

Modelo autorregresivo de orden p:

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

- AR(0): Sin parámetros, sólo error
- AR(1): Un único parámetro  $\varphi_1$  (positivo)
- AR(2): Dos parámetros  $\varphi_1$  (positivo) y  $\varphi_2$



# Modelos estadísticos



Medias móviles

**MA(q)**

Modelo de medias móviles de orden q:

$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$



# Modelos estadísticos



Autorregresión (AR) + Medias móviles (MA)

## **ARMA(p,q)**

Con  $p$  términos autorregresivos y  $q$  medias móviles:

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Peter Whittle: "Hypothesis testing in time series analysis,"  
Ph.D. thesis, Uppsala, Sweden, 1951



# Modelos estadísticos



Cuando la media no es estacionaria...

## **ARIMA(p,d,q)**

**autoregressive integrated moving average**

Se añaden  $d$  pasos de diferenciación ("integración") para eliminar la tendencia de la serie (media no estacionaria):

- $d=0$ :  $X_t^* = X_t$
- $d=1$ :  $X_t^* = X_t - X_{t-1}$
- $d=2$ :  $X_t^* = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$

Y se aplica un modelo ARMA sobre  $X_t^*$ ...



# Modelos estadísticos



Cuando la media no es estacionaria...

**ARIMA(p,d,q)**

**autoregressive integrated moving average**

$$X_t = -(\Delta^d X_t - X_t) + \varphi_0 + \sum_{i=1}^p \varphi_i \Delta^d X_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

donde

$$\Delta X_t = X_t - X_{t-1}$$

George E.P. Box & Gwilym Jenkins: *Time Series Analysis: Forecasting and Control*, 2<sup>nd</sup> ed., Holden-Day, 1976



# Modelos estadísticos



Cuando la media no es estacionaria...

## **ARIMA(p,d,q)**

**autoregressive integrated moving average**

- $ARIMA(1,0,0) = AR(1)$
- $ARIMA(0,1,0) = I(1) = \text{Random walk (paseo aleatorio)}$
- $ARIMA(0,0,1) = MA(1)$
- $ARIMA(0,1,2) = \text{Damped Holt's model}$
- $ARIMA(0,1,1) = \text{Suavizado exponencial básico}$
- $ARIMA(0,2,2) = \text{Suavizado exponencial doble, a.k.a. Método lineal de Holt con errores aditivos.}$



# Modelos estadísticos



ARIMA estacional

**SARIMA(p,d,q)(P,D,Q)m**

**Seasonal** autoregressive integrated moving average

- (p,d,q) Parte no estacional
- (P,D,Q) Parte estacional  
P términos AR estacionales  
D diferencias estacionales  
Q términos MA estacionales
- m Observaciones por año

p.ej.

SARIMA(1,0,1)(2,1,0)12       $y_t - y_{t-12} = W_t$

$$W_t = \mu + \phi_1 W_{t-1} + \phi_2 W_{t-12} + \phi_3 W_{t-24} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$



# Modelos estadísticos



ARIMA estacional

**SARIMA(p,d,q)(P,D,Q)m**

**Seasonal** autoregressive integrated moving average

**SARIMA(1,0,m+1)(0,1,0)m**

equivale a

**método de Holt-Winters aditivo**

(suavizado exponencial triple)



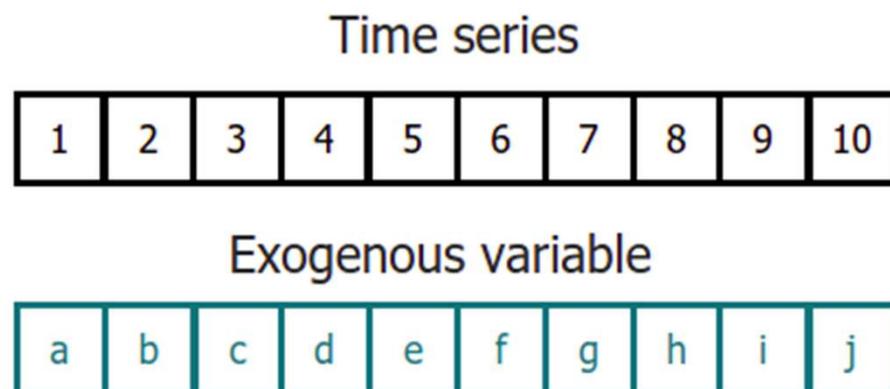
# Modelos estadísticos



Con variables exógenas

## **SARIMAX**

**Seasonal Autoregressive Integrated Moving-Average  
with eXogenous regressors**



X						y
1	2	3	4	5	f	6
2	3	4	5	6	g	7
3	4	5	6	7	h	8
4	5	6	7	8	i	9
5	6	7	8	9	j	10

# Modelos estadísticos



Extensiones vectoriales,  
para predecir series temporales multivariable  
(i.e. múltiples series temporales relacionadas):

- VAR [Vector AutoRegressive]
- VMA [Vector Moving Average]
- VARMA [Vector ARMA]
- VARIMA [Vector ARIMA]
- ...

p.ej. Econometría



# Modelos estadísticos



Cuando la varianza del error no es estacionaria pero se asume que sigue un modelo autorregresivo (AR):

**ARCH(q)**

**AutoRegressive Conditional Heteroskedasticity**

$$\varepsilon_t = \sigma_t Z_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

Robert F. Engle: "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation," *Econometrica*, 50(4):987–1007, 1982



# Modelos estadísticos



Cuando la varianza del error no es estacionaria pero se asume que sigue un modelo ARMA:

**GARCH(p,q)**  
**Generalized ARCH**

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

Tim Bollerslev: "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31(3): 307–327, 1986



# Modelos estadísticos



Los modelos ARCH/GARCH proporcionan una buena descripción de dos características de las series temporales en finanzas:

- Volatilidad cambiante en el tiempo
- Distribuciones leptocúrticas (colas más gruesas que en una distribución normal)

Sin embargo, la descripción de la varianza en términos de cuadrados no tiene en cuenta el signo de los cambios, cuando se ha observado que existen asimetrías en la correlación entre la volatilidad actual y los valores pasados de una serie (shocks positivos/negativos), p.ej. en los retornos de activos financieros.



# Modelos estadísticos



Variantes que rompen la simetría de GARCH:

- **EGARCH** [Exponential GARCH]

Daniel B. Nelson: "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59(2):347–370, 1991

- **QGARCH** [Quadratic GARCH]

Enrique Sentana: "Quadratic ARCH Models," *Review of Economic Studies*, 62(4):639-661, 1995

- **TGARCH** [Threshold GARCH]

Jean-Michel Zakoian: "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18(5):931-955, 1994

- ...



# Sistemas de predicción



## Prophet (Facebook)

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t)$$

Modelo aditivo:

- $g(t)$  Tendencia  
(crecimiento con saturación o lineal a trozos)
- $s(t)$  Estacionalidad [seasonality]  
(series de Fourier)
- $h(t)$  Vacaciones [holidays]  
(indicador 0/1 en función del día)
- $\varepsilon(t)$  Error

<https://facebook.github.io/prophet/>

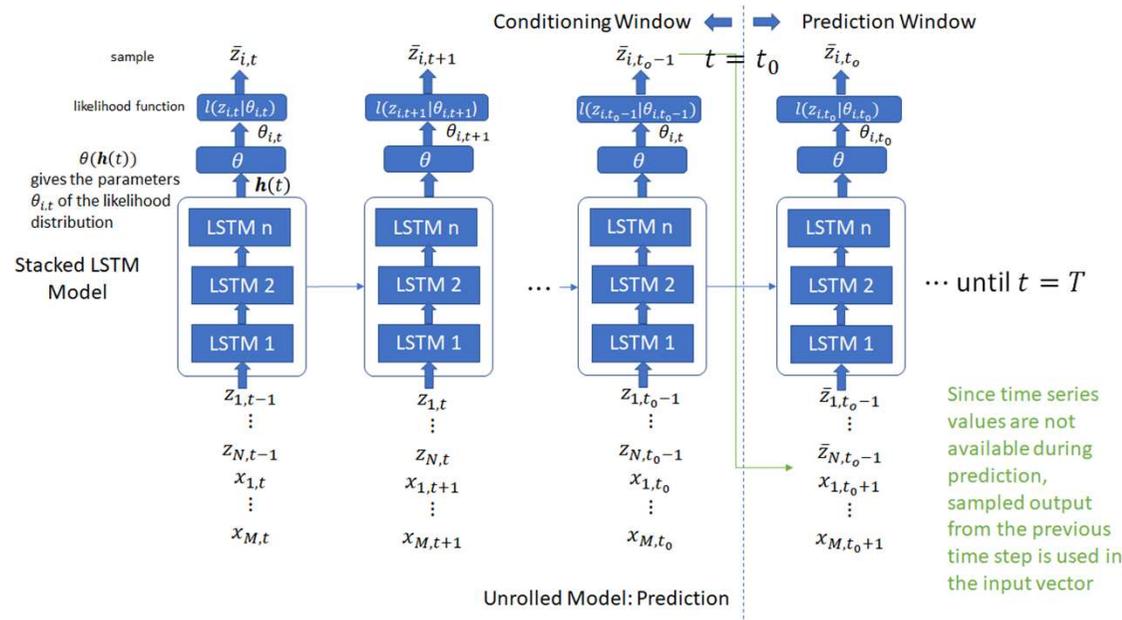
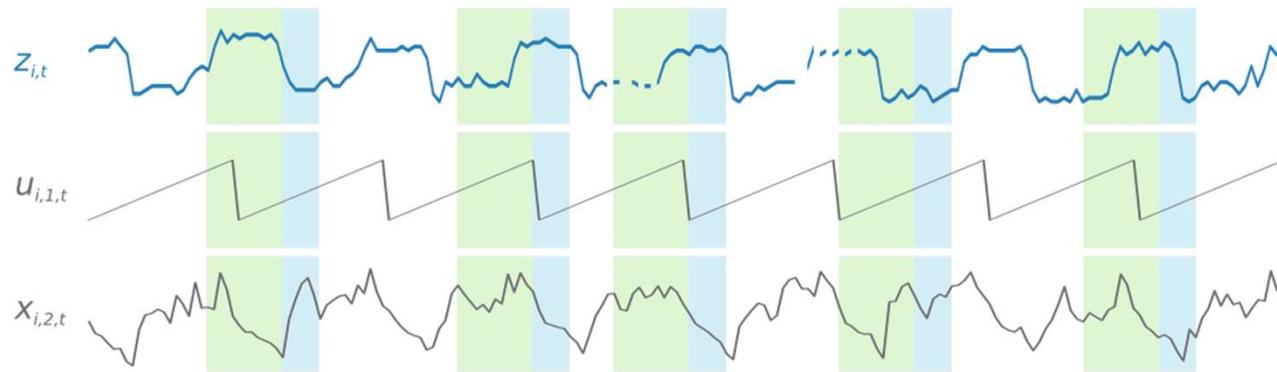


# Sistemas de predicción



## DeepAR (Amazon)

### Red neuronal recurrente

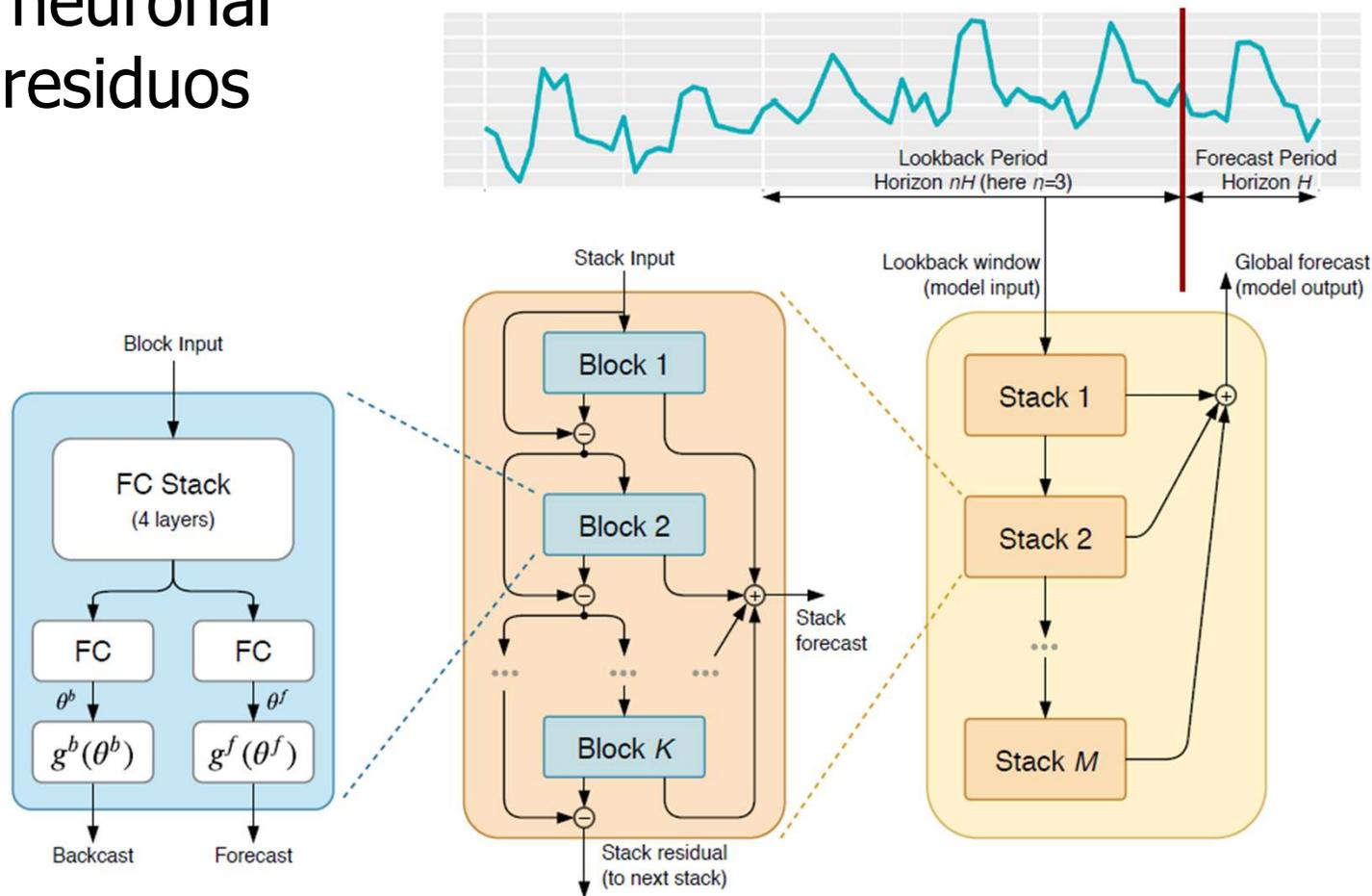


# Sistemas de predicción



## N-BEATS (ICLR'2020)

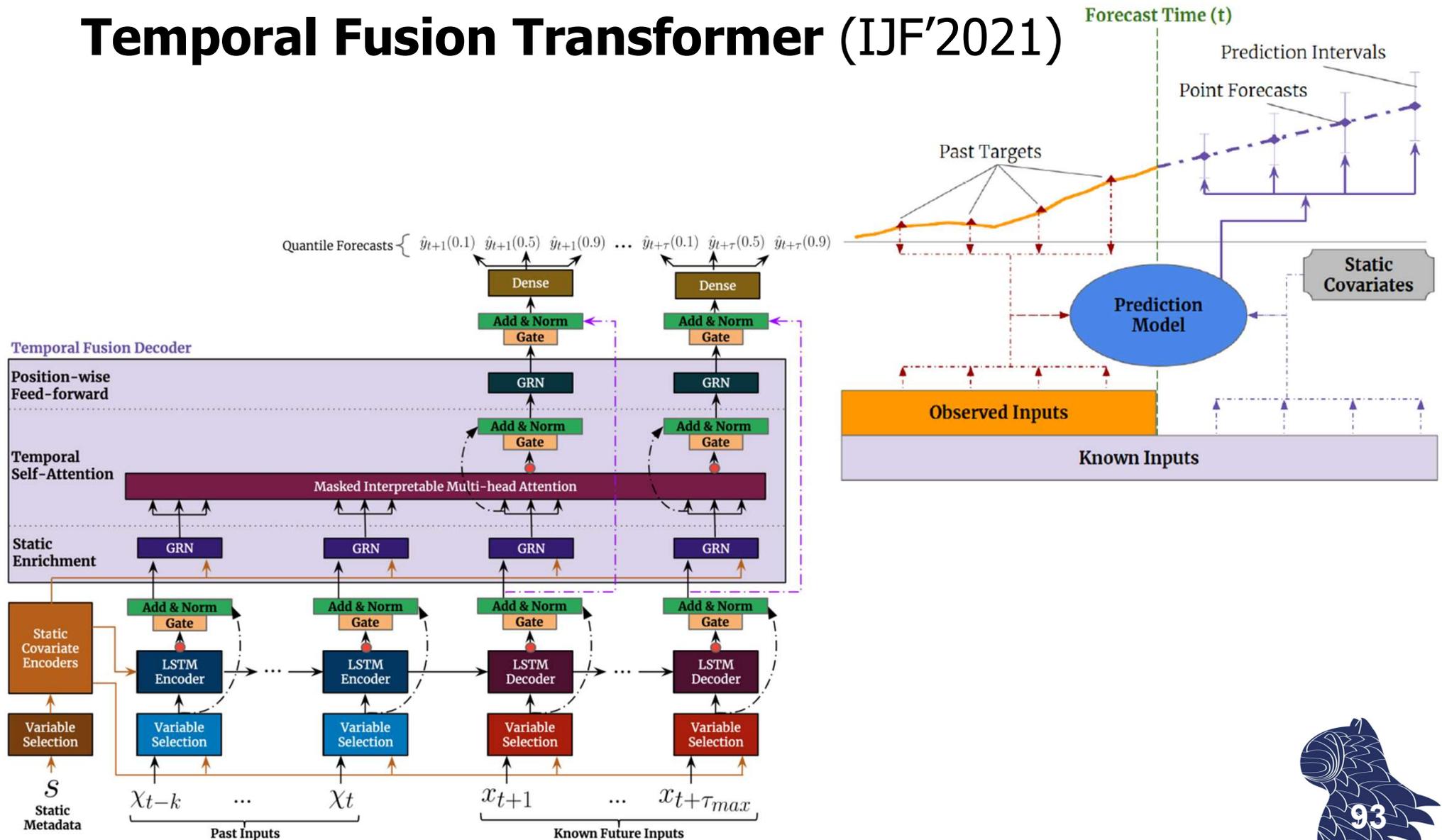
Red neuronal  
con residuos



# Sistemas de predicción



## Temporal Fusion Transformer (IJF'2021)

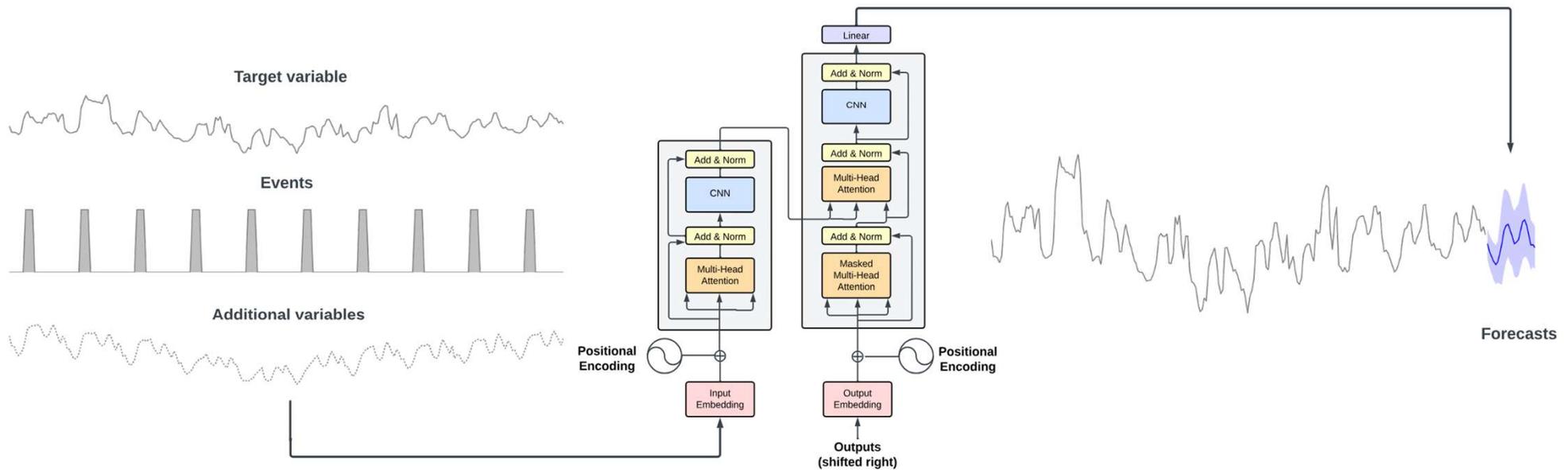


# Sistemas de predicción



## TimeGPT (arXiv'2023)

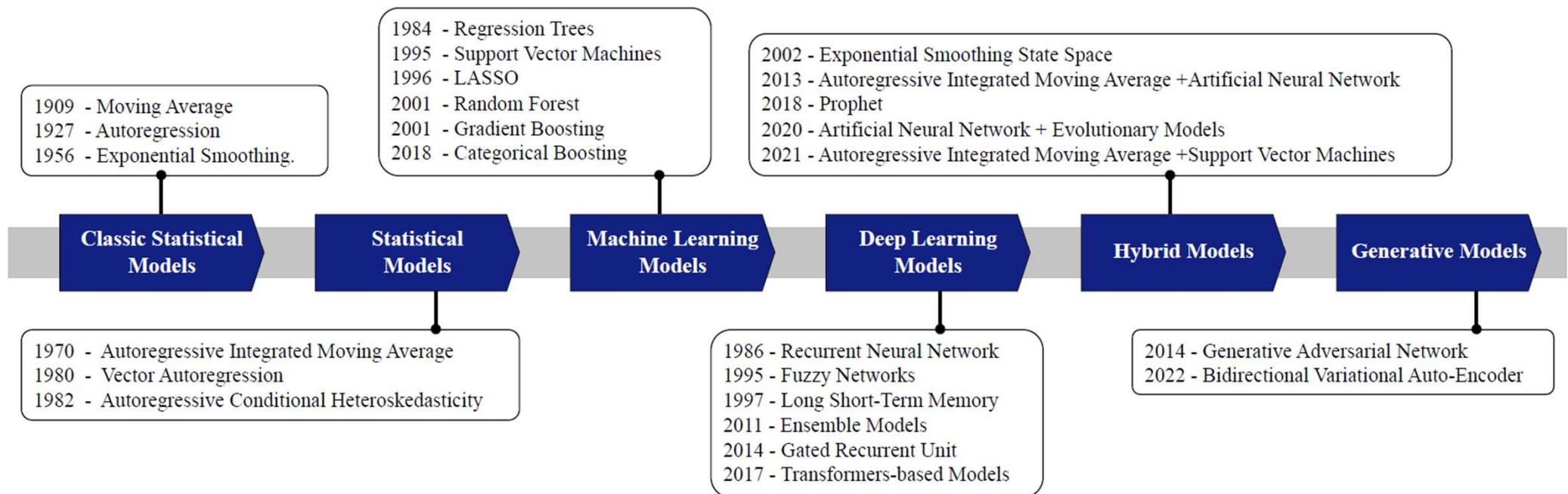
<https://github.com/Nixtla/nixtla>



# Sistemas de predicción



## Desarrollo cronológico de modelos de predicción de series temporales



“A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems”, Machines’2024  
<https://doi.org/10.3390/machines12060380>



# Sistemas de predicción



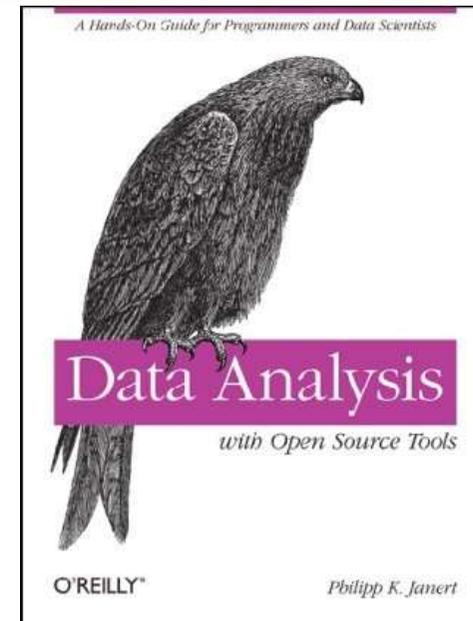
Model	Advantages	Disadvantages
ARIMA	<ul style="list-style-type: none"><li>- Handles linear relationships</li><li>- Can model stationary time-series data</li></ul>	<ul style="list-style-type: none"><li>- Requires data to be stationary</li><li>- Struggles with non-linear patterns</li></ul>
ANN	<ul style="list-style-type: none"><li>- Can model complex non-linear relationships</li><li>- Flexible architecture</li></ul>	<ul style="list-style-type: none"><li>- Requires large datasets</li><li>- Black box nature</li></ul>
RNN	<ul style="list-style-type: none"><li>- Can capture sequences and patterns over time</li><li>- Suitable for time-series data</li></ul>	<ul style="list-style-type: none"><li>- Can struggle with long-term dependencies</li></ul>
LSTM	<ul style="list-style-type: none"><li>- Addresses vanishing gradient problem of RNNs</li><li>- Can model long-term dependencies</li></ul>	<ul style="list-style-type: none"><li>- Computationally intensive</li><li>- Requires careful tuning</li></ul>
KNN	<ul style="list-style-type: none"><li>- Non-parametric</li><li>- Simple to understand</li></ul>	<ul style="list-style-type: none"><li>- Computationally expensive</li><li>- Sensitive to irrelevant features</li></ul>
Fuzzy Networks	<ul style="list-style-type: none"><li>- Can handle uncertainty and vagueness</li><li>- Flexible modeling</li></ul>	<ul style="list-style-type: none"><li>- Requires expert knowledge</li><li>- Can be computationally intensive</li></ul>
Regression Models	<ul style="list-style-type: none"><li>- Interpretable</li><li>- Can handle multiple predictors</li></ul>	<ul style="list-style-type: none"><li>- Assumes linear relationship</li><li>- Sensitive to outliers</li></ul>
Prophet	<ul style="list-style-type: none"><li>- Handles seasonality and holidays</li><li>- Robust to missing data</li></ul>	<ul style="list-style-type: none"><li>- Assumes additive model</li><li>- Might require domain-specific adjustments</li></ul>
MeanAbsolute	<ul style="list-style-type: none"><li>- Simple to compute</li><li>- Provides a baseline measure</li></ul>	<ul style="list-style-type: none"><li>- Does not capture complex patterns</li></ul>

“A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems”, Machines’2024  
<https://doi.org/10.3390/machines12060380>

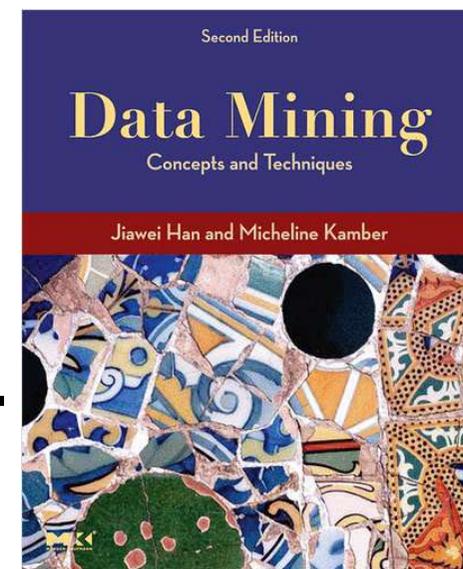


# Bibliografía

- Philipp K. Janert:  
**Data Analysis  
with Open Source Tools**  
O'Reilly, 2010.  
ISBN 0596802358  
Part I - Graphics: Looking at Data



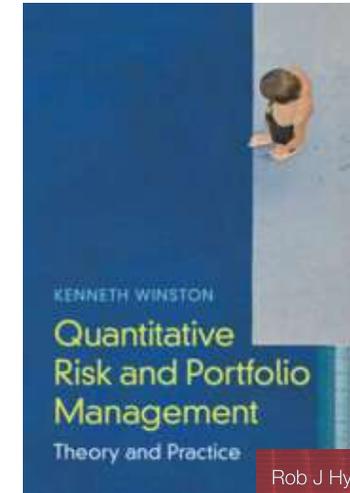
- Jiawei Han  
& Micheline Kamber:  
**Data Mining:  
Concepts and Techniques**  
2<sup>nd</sup> edition, Morgan Kaufmann, 2006.  
ISBN 1558609016  
8.2 Mining Time-Series Data



# Bibliografía

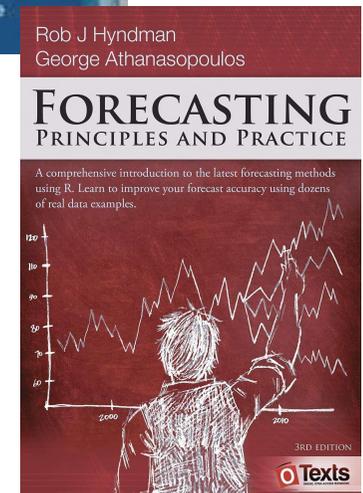


- Kenneth J. Winston:  
**Quantitative Risk and Portfolio Management: Theory and Practice**  
Cambridge University Press, 2023  
ISBN 1009209043



<https://doi.org/10.1017/9781009209090>

- Rob J. Hyndman & George Athanasopoulos:  
**Forecasting: Principles and Practice**  
OTexts, 3<sup>rd</sup> edition, 2021  
ISBN 0987507133



<https://otexts.com/fpp3/>

