MASTER DE INGENIERÍA MEDIOAMBIENTAL .

Seminario: Series temporales (Metodología clásica...y algo más!!)

All models are wrong, but some are useful.” George E. P. Box (1979, p. 202)

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Introducción.

Una serie temporal es un conjunto de observaciones realizadas secuencialmente en el tiempo (una maldita cosa después de otra R.A. Fisher).

En la mayoría de aplicaciones el tiempo entre observaciones es constante; de hecho toda la teoría de procesado de series temporales se basa en esta suposición (además de que existe algún tipo de relación entre las muestras).

Esta información temporal es la que se puede utilizar para extraer más información sobre la señal. Se tiene un parámetro (el tiempo) que nos indica la posible periodicidad de la serie, la velocidad de crecimiento, disminuición, la presencia o no de una tendencia lineal ,etc. En este breve seminario veremos cómo podemos extraer esta información de la serie temporal.

En cuanto a las aplicaciones se tienen en todos los campos del conocimiento; economía (valores en bolsa, stocks, ventas); meteorología (precipitaciones, temperaturas, humedad); Salud (bioseñales, farmacocinética/farmacodinámica, control de señales en una UCI, etc).

Objetivos en el análisis de una serie temporal.

Desarrollo de modelos.	Construcción de un modelo para representar los datos
Estimación.	Obtención de los parámetros que se tienen en el modelo que se quiere construir para analizar la serie temporal.
Comprobación.	Validación de un determinado modelo mediante el ajuste de la serie temporal.
Comprensión.	Uso del modelo para entender el mecanismo que genera la serie temporal.
Filtrado.	Separación del ruido de la señal.
Testeo de hipótesis	Comprobar la validez de hipótesis mediante la serie temporal (¿se tiene un cambio climático actualmente?).
Predicción.	Obtención de valores futuros de la serie temporal.
Simulación.	Uso del modelo para generar valores que no se tiene posibilidad (o que son muy difíciles) de obtener de la serie temporal.

Elementos en una serie temporal.

DESCOMPOSICIÓN CLÁSICA

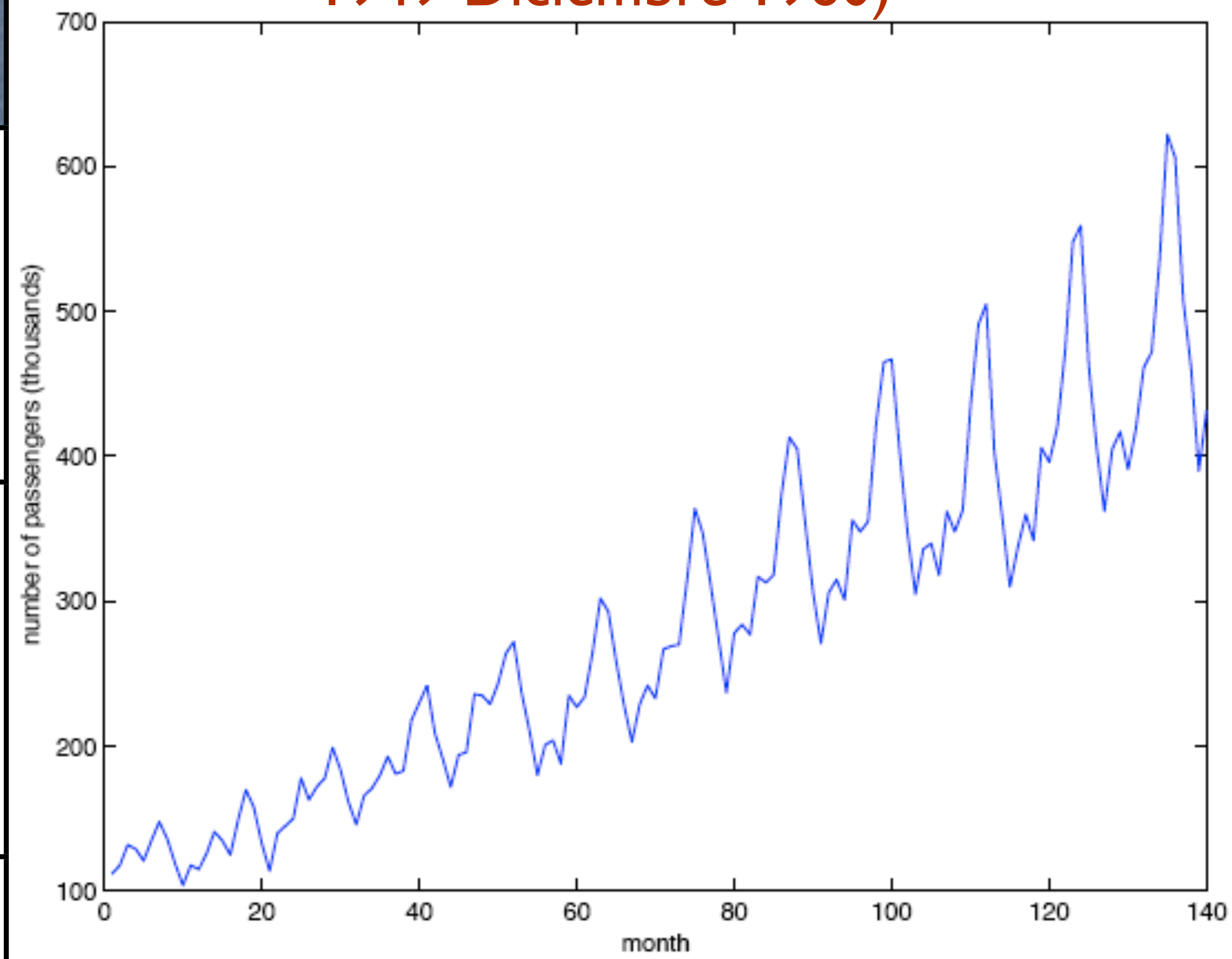
Tendencia: variación de la serie temporal relacionada con una función polinómica de un determinado orden.

Variación estacional: fluctuación de la serie temporal que se repite cada cierto tiempo.

Ciclo: variaciones como las anteriores pero sin seguir un patrón estacional.

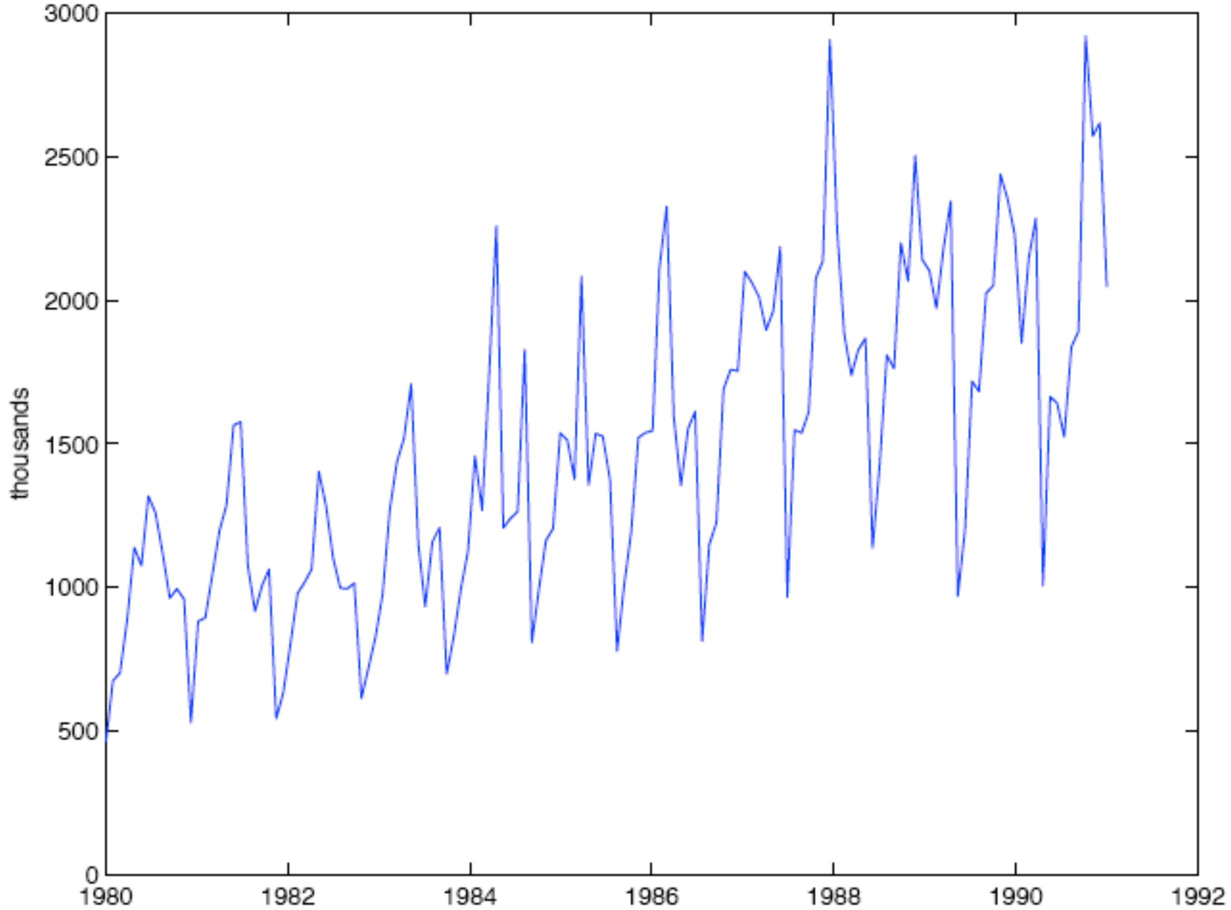
Residuos: variaciones aleatorias y todos los fenómenos no explicados en los puntos anteriores.

Número de pasajeros de avión (Enero 1949-Diciembre 1960)

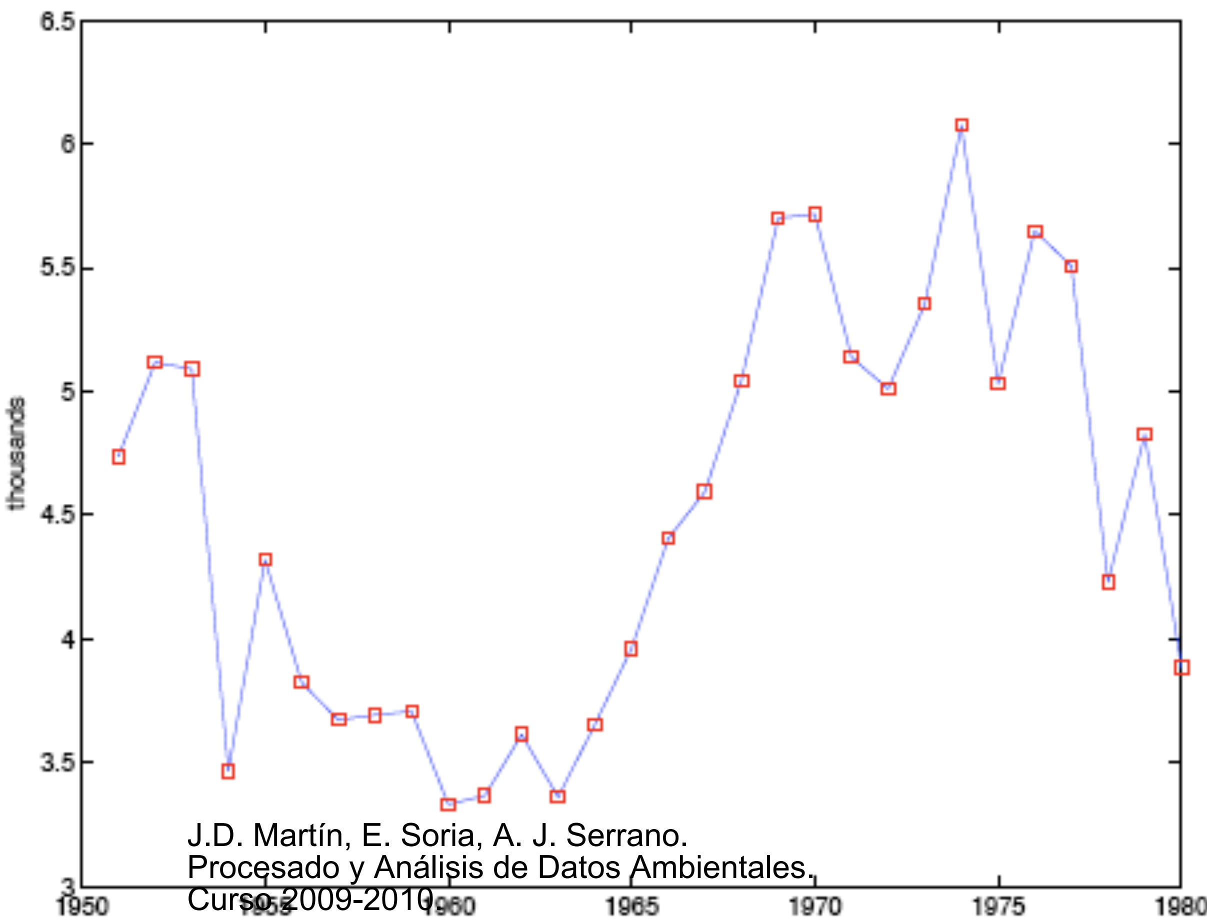


La observación temporal es muy importante para observar las posibles tendencias y los ciclos estacionales que hay que eliminar para hacer la serie estacionaria.

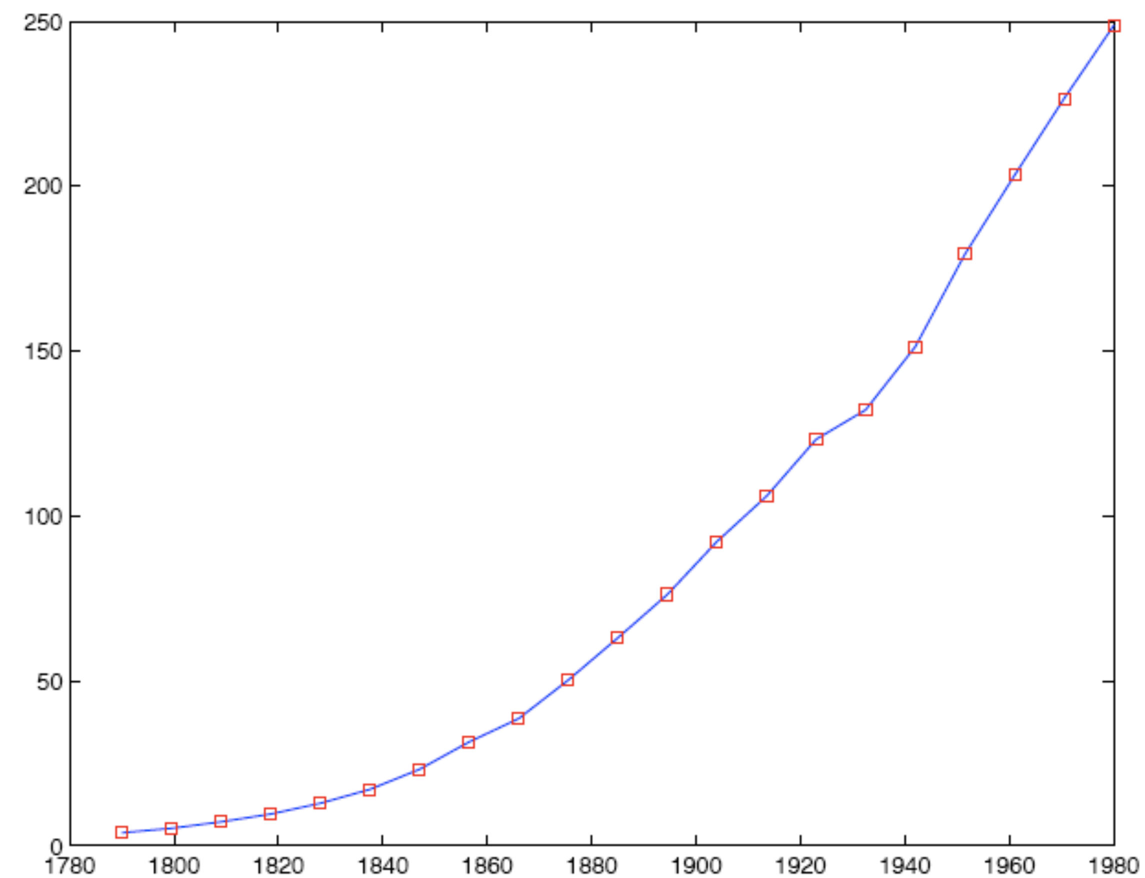
Ejemplos



Vino australiano
(Enero 1980 a
Diciembre 1991)



Huelgas en
EEUU



Población en EEUU (cada diez años)

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Pasos en el análisis de una serie temporal

El primer paso es representar la serie temporal para determinar:

Posibles tendencias (hay que eliminarlas antes de procesar la señal).

Posibles periodos en la señal (tendencias cíclicas)

¿Las variaciones que se observan se pueden modelizar de forma temporal?

¿Aparecen cambios bruscos en el comportamiento de la serie temporal?

¿Tenemos outliers que pueden conducirnos a grandes errores en la modelización o bien a conclusiones erróneas?

Y lo siguientes serían.....

Transforma la serie si es necesario (por ejemplo aplicando una función logaritmo para analizar un crecimiento exponencial).

Elimina la posible tendencia así como los posibles ciclos (la operación más usual es diferenciar la serie temporal).

Escoge un modelo para ajustar la serie final obtenida (tema central en el análisis de series temporales)

Implementa el modelo y .....COMPRUÉBALO!

Parámetros y estacionariedad.

Valor medio; es el valor que mejor modeliza a la serie temporal como una constante

$$\mu_{X(t)} = E(X_t)$$

Covarianza; este parámetro determina una posible relación lineal entre los instantes t y s de la serie temporal.

$$\gamma_X(s,t) = E[(X_s - \mu_X(s)) \cdot (X_t - \mu_X(t))]$$

Varianza; es el parámetro que determina la dispersión de la serie temporal

$$\sigma^2_{X(t)} = E[(X_t - \mu_X(t))^2]$$

Una serie temporal se dice que es estacionaria si sus propiedades estadísticas no cambian al considerar una versión desplazada de dicha serie temporal

En el análisis de serie temporales se puede reducir esta condición a la que se conoce como estacionariedad en sentido amplio, o estacionariedad de segundo orden que consiste en: a) tener un valor medio independiente de t; b) el valor de la covarianza sólo depende de la diferencia entre los tiempos considerados en el cálculo de este parámetro y c) varianza constante para cada t.

Sea X_t una serie temporal estacionaria se define su autocovarianza como

$$\gamma_{X(h)} = E[(X_{t+h} - \mu_X) \cdot (X_t - \mu_X)]$$

Sea X_t una serie temporal estacionaria se define su autocorrelación (ACF) como

$$\rho_{X(h)} = \frac{\gamma_{X(h)}}{\gamma_{X(0)}} = corr(X_{t+h}, X_t)$$

ESTAS CANTIDADES DESCRIBEN LA DEPENDENCIA ENTRE MUESTRAS DE LA SERIE TEMPORAL.

Ruido blanco.

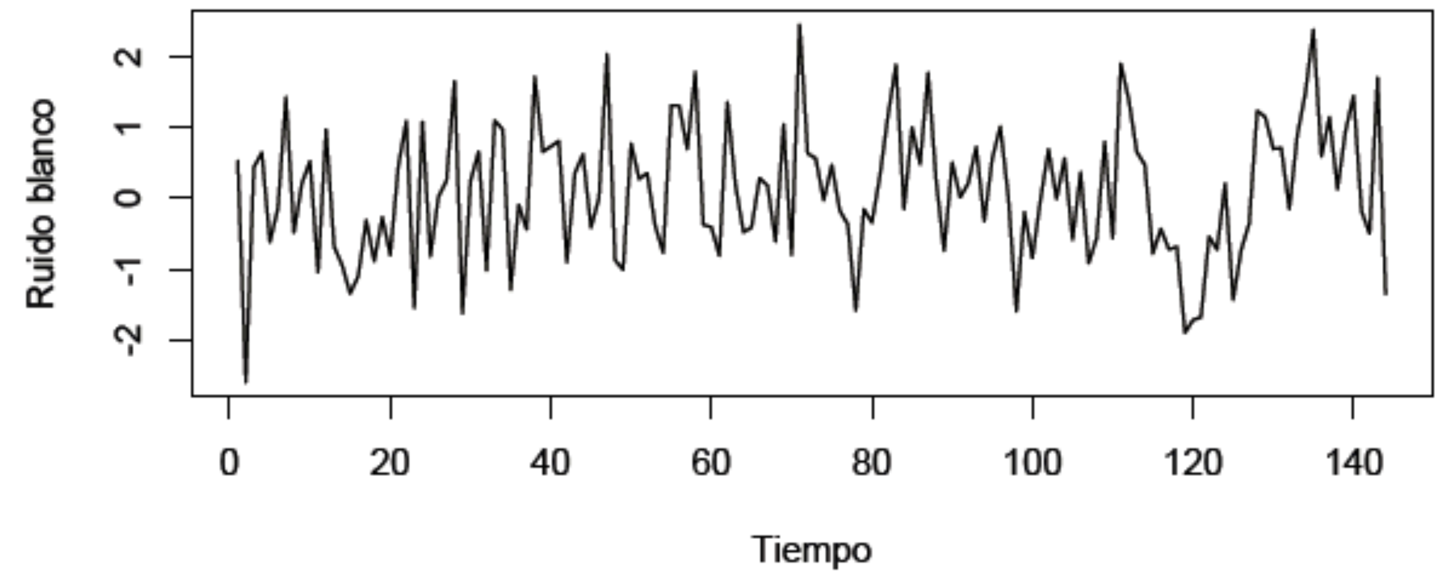
Esta secuencia se considera en todos los modelos de series temporales que se plantean y tiene las siguientes características:

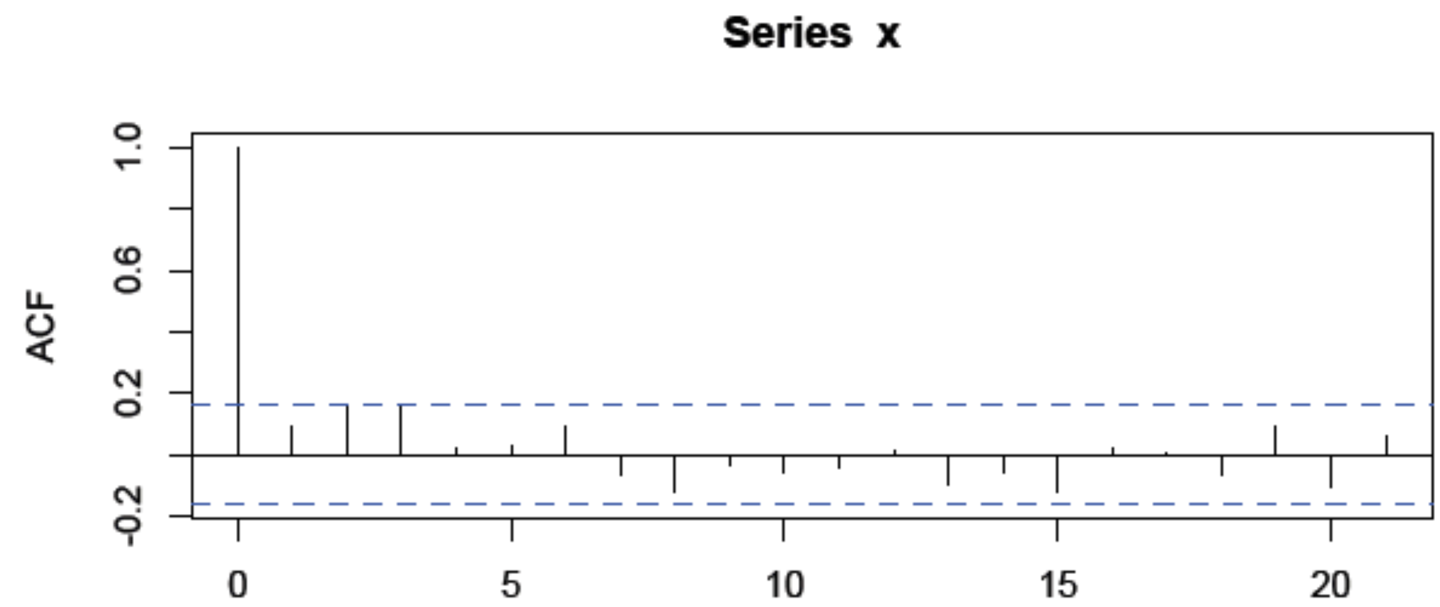
$$E(X_t) = 0$$

$$E(X_t^2) = \sigma^2 < \infty$$

$$\gamma_{X(h)} = 0; h \neq 0$$

Es una secuencia con valor medio cero, varianza finita y en la que no se tiene una relación lineal (o existe una relación no lineal) entre muestras; esta característica queda definida por el valor de la autocorrelación.





Este modelo sirve como un bloque básico para el resto de modelos de series temporales que se analizarán a continuación.

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Modelo Promediado Móvil, *MA*

Una serie temporal X_t es un proceso de tipo MA(1) si queda definido de la siguiente forma:

$$X_t = Z_t + \theta \cdot Z_{t-1}$$

Donde Z_t es un proceso del tipo ruido blanco.

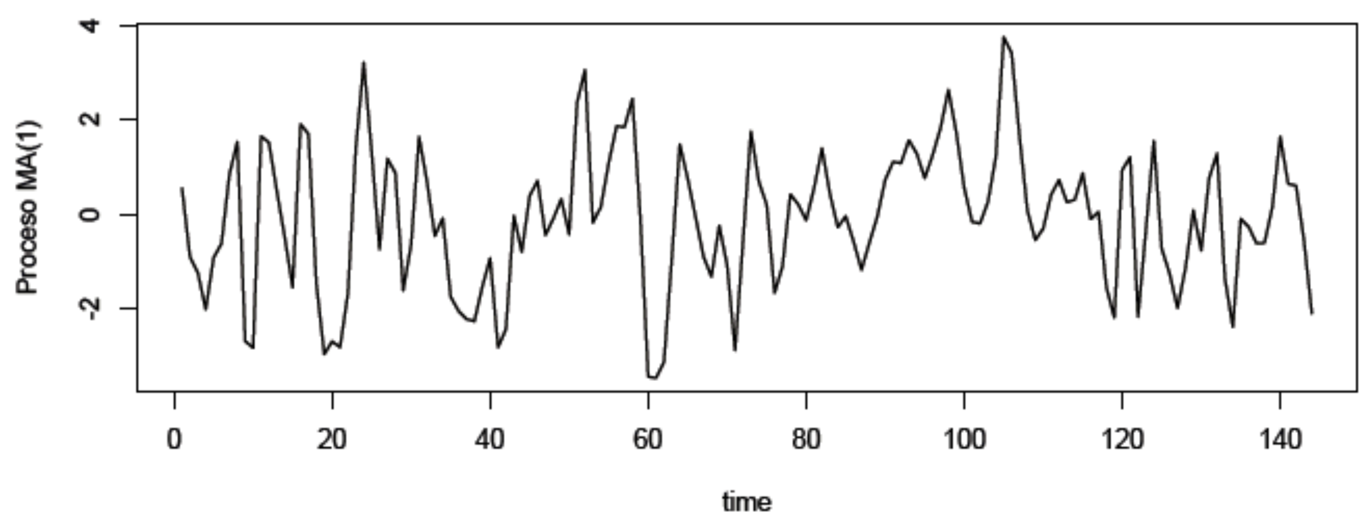
La autocovarianza de esta serie temporal viene dada por la siguiente expresión:

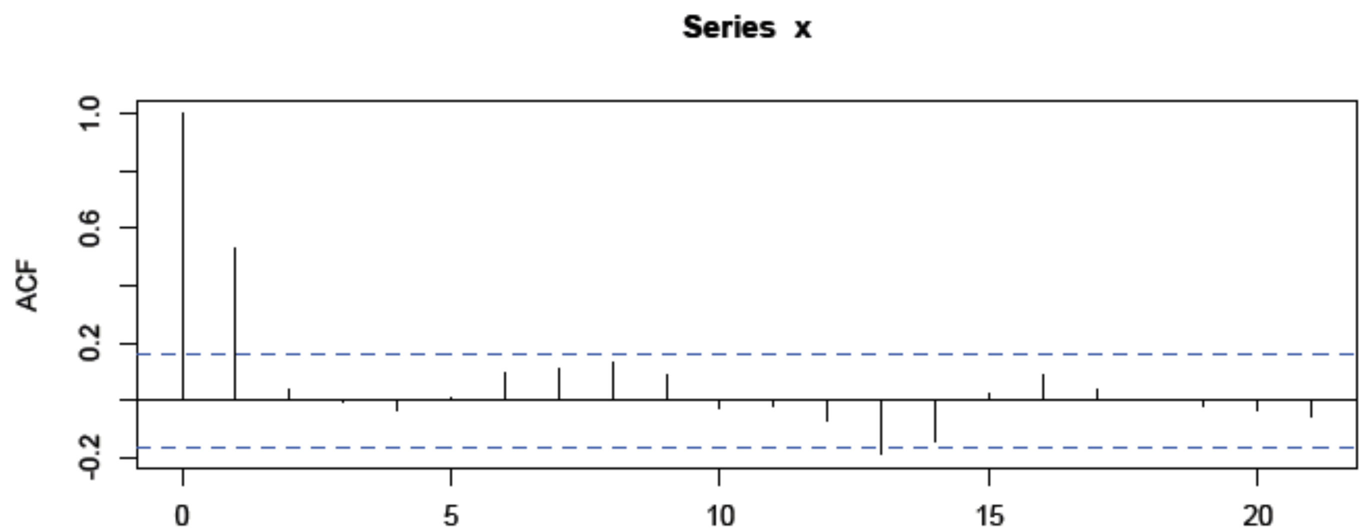
$$\gamma_X(h) = \begin{cases} \sigma^2 \cdot (1+\theta^2), & \text{if } h = 0; \\ \sigma^2 \cdot \theta, & \text{if } h = \pm 1; \\ 0 & \text{if } |h| > 1; \end{cases}$$

Y la autocorrelación (ACF) por la expresión

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0; \\ \frac{\theta}{1+\theta^2}, & \text{if } h = \pm 1; \\ 0 & \text{if } |h| > 1; \end{cases}$$

Proceso MA(1); parámetro=0.9





Es una serie temporal donde se tiene una memoria finita del proceso (existe una relación entre la muestra n y la n-1).

Modelo Promediado Móvil, *MA*

Una serie temporal X_t es un proceso de tipo MA(q) si queda definido de la siguiente forma:

$$X_t = Z_t + \theta_1 \cdot Z_{t-1} + \ldots \theta_q \cdot Z_{t-q} \qquad X_t = \sum_{i=0}^{q} \theta_i \cdot Z_{t-i}$$

Donde Z_t es un proceso del tipo ruido blanco.

La autocorrelación del proceso (ACF) queda definida por la siguiente expresión:

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0; \\ \frac{-\theta_h + \theta_1 \cdot \theta_{h+1} + \theta_2 \cdot \theta_{h+2} + \cdots + \theta_{q-h} \cdot \theta_q}{1 + \sum_{i=1}^{q} \theta_i^2}, & \text{if } h = 1, 2, \ldots q; \\ 0, & \text{if } h > q; \end{cases}$$

Si X_t es una serie temporal con una correlación no nula hasta el índice s entonces puede ser representada por un proceso de tipo MA(s)

Modelo Autorregresivo, *AR*

Una serie temporal X_t es un proceso de tipo AR(1) si queda definido de la siguiente forma:

$$X_t = \alpha \cdot X_{t-1} + Z_t$$

Donde Z_t es un proceso de tipo ruido blanco.

La autocovarianza de este proceso viene dado por la siguiente expresión:

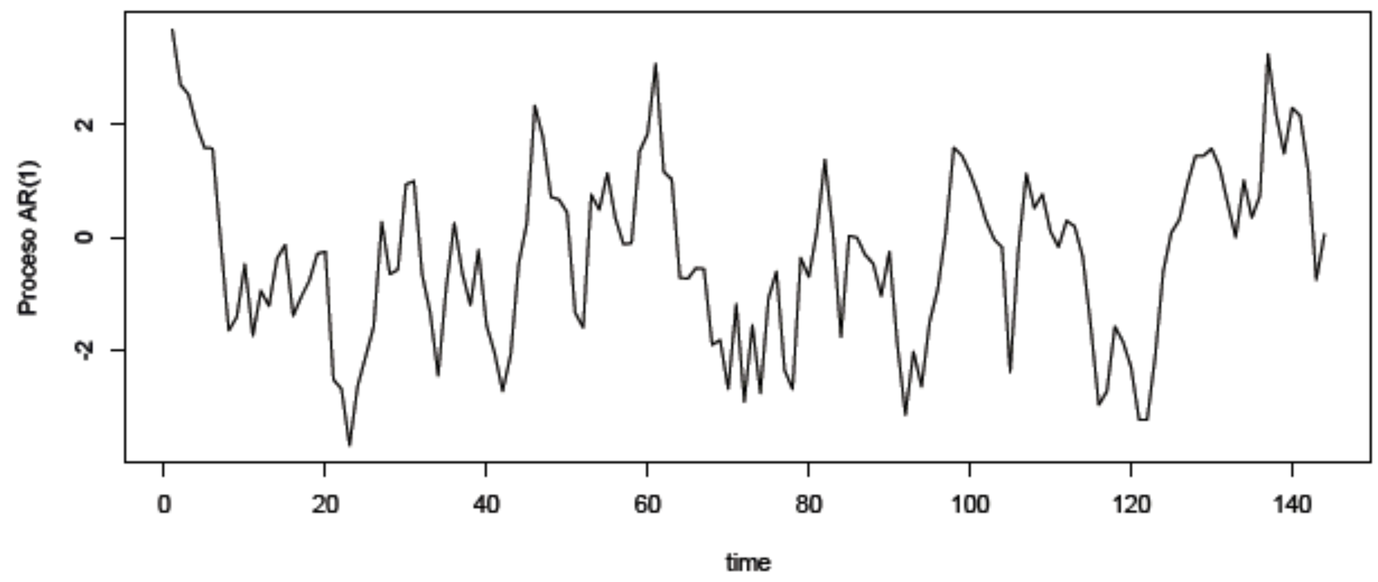
$$\gamma_X(h) = \sigma^2 \cdot \frac{\alpha^{|h|}}{1-\alpha^2}$$

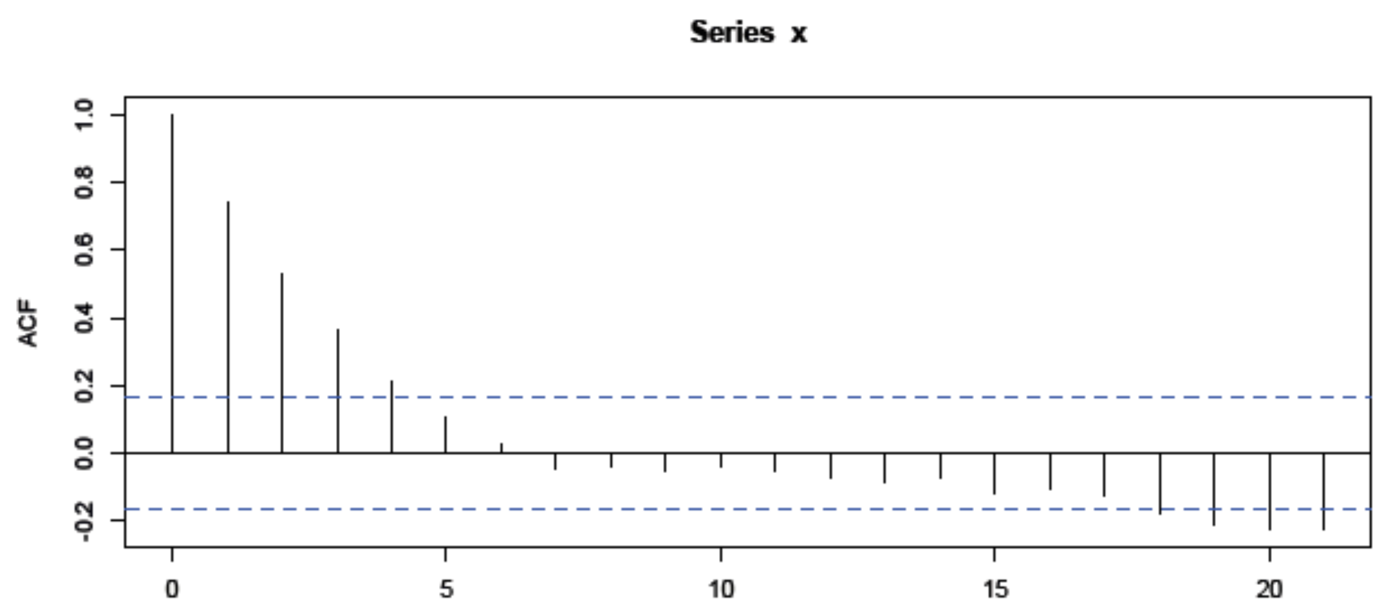
La autocorrelación (ACF) de este proceso viene dada por la siguiente expresión:

$$\rho_X(h) = \alpha^{|h|}$$

En este proceso se debe cumplir que $|\alpha|<1$

Proceso AR(1); parámetro=0.9





EN ESTE TIPO DE PROCESOS SE ASUME QUE EXISTE UNA MEMORIA A LARGO PLAZO

Modelo Autorregresivo, *AR*

Una serie temporal X_t es un proceso de tipo AR(p) si queda definido de la siguiente forma:

$$X_t = \alpha_1 \cdot X_{t-1} + \alpha_2 \cdot X_{t-2} + \cdots + \alpha_p \cdot X_{t-p} + Z_t$$

Donde Z_t es un proceso de tipo ruido blanco.

Se tienen que cumplir las siguientes condiciones (son necesarias pero no suficientes) para que sea una serie estacionaria

$$\begin{cases} \sum_{i=1}^{p} \alpha_i < 1 \\ |\alpha_k| < 1 \end{cases}$$

La autocorrelación de este tipo de procesos puede dar lugar a diferentes expresiones y viene dada por la solución del siguiente sistema de ecuaciones (ecuaciones de *Yule-Walker*).

$$\begin{cases} \rho_1 = \alpha_1 + \alpha_2 \cdot \rho_1 + \alpha_3 \cdot \rho_2 + \cdots + \alpha_p \cdot \rho_{p-1} \\ \rho_2 = \alpha_1 \cdot \rho_1 + \alpha_2 + \alpha_3 \cdot \rho_1 + \cdots + \alpha_p \cdot \rho_{p-2} \\ \cdots \\ \rho_p = \alpha_1 \cdot \rho_{p-1} + \alpha_2 \cdot \rho_{p-2} + \alpha_3 \cdot \rho_{p-3} + \cdots + \rho_p \end{cases}$$

Dependiendo de los coeficientes se obtienen una serie de soluciones para la secuencia de autocorrelación.

Modelo ARMA

Una serie temporal X_t es un proceso de tipo ARMA(1,1) si queda definido de la siguiente forma:

$$X_t = \phi \cdot X_{t-1} + Z_t + \theta \cdot Z_{t-1}$$

Donde Z_t es un proceso del tipo ruido blanco.

La autocorrelación de este proceso viene dada por

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0; \\ \frac{(1+\theta\cdot\phi)\cdot(\theta+\phi)}{1+2\cdot\theta\cdot\phi+\theta^2} \cdot \phi^{h-1}, & \text{if } h \geq 1; \end{cases}$$

Una serie temporal X_t es un proceso de tipo ARMA(p,q) si queda definido de la siguiente forma:

$$X_t = \phi_1 \cdot X_{t-1} + \phi_2 \cdot X_{t-2} + \ldots \phi_p \cdot X_{t-p} + Z_t + \theta_1 \cdot Z_{t-1} + \theta_2 \cdot Z_{t-2} + \ldots \theta_q \cdot X_{t-q}$$

Donde Z_t es un proceso del tipo ruido blanco.

La obtención de la secuencia de autocorrelación en este tipo de modelos es más complicada que en los anteriores (debido a la existencia de los dos términos AR y MA).

Modelo ARIMA

Una manera de hacer una señal estacionaria es diferenciándola, los procesos ARIMA consideran esta situación.

Una serie temporal Y_t es un proceso de tipo ARIMA(p,d,q) si la d-diferencia de Y_t sigue un proceso ARMA(p,q). A nivel práctico, normalmente, es suficiente con una primera o segunda derivada como mucho.

Por ejemplo, un proceso ARIMA(1,1,1) vendría dado por:

$$W_t = \phi \cdot W_{t-1} + Z_t + \theta \cdot Z_{t-1}$$

que, desarrollando

$$Y_t - Y_{t-1} = \phi \cdot (Y_{t-1} - Y_{t-2}) + Z_t + \theta \cdot Z_{t-1}$$

finalmente, se obtiene

$$Y_t = (1+\phi) \cdot Y_{t-1} - \phi \cdot Y_{t-2} + Z_t + \theta \cdot Z_{t-1}$$

Si el proceso no tiene parte autorregresiva se conoce como IMA(d,q) y, si tiene solo parte autoregresiva, se conoce como ARI(*p*,*d*).

Metodología de Box-Jenkins

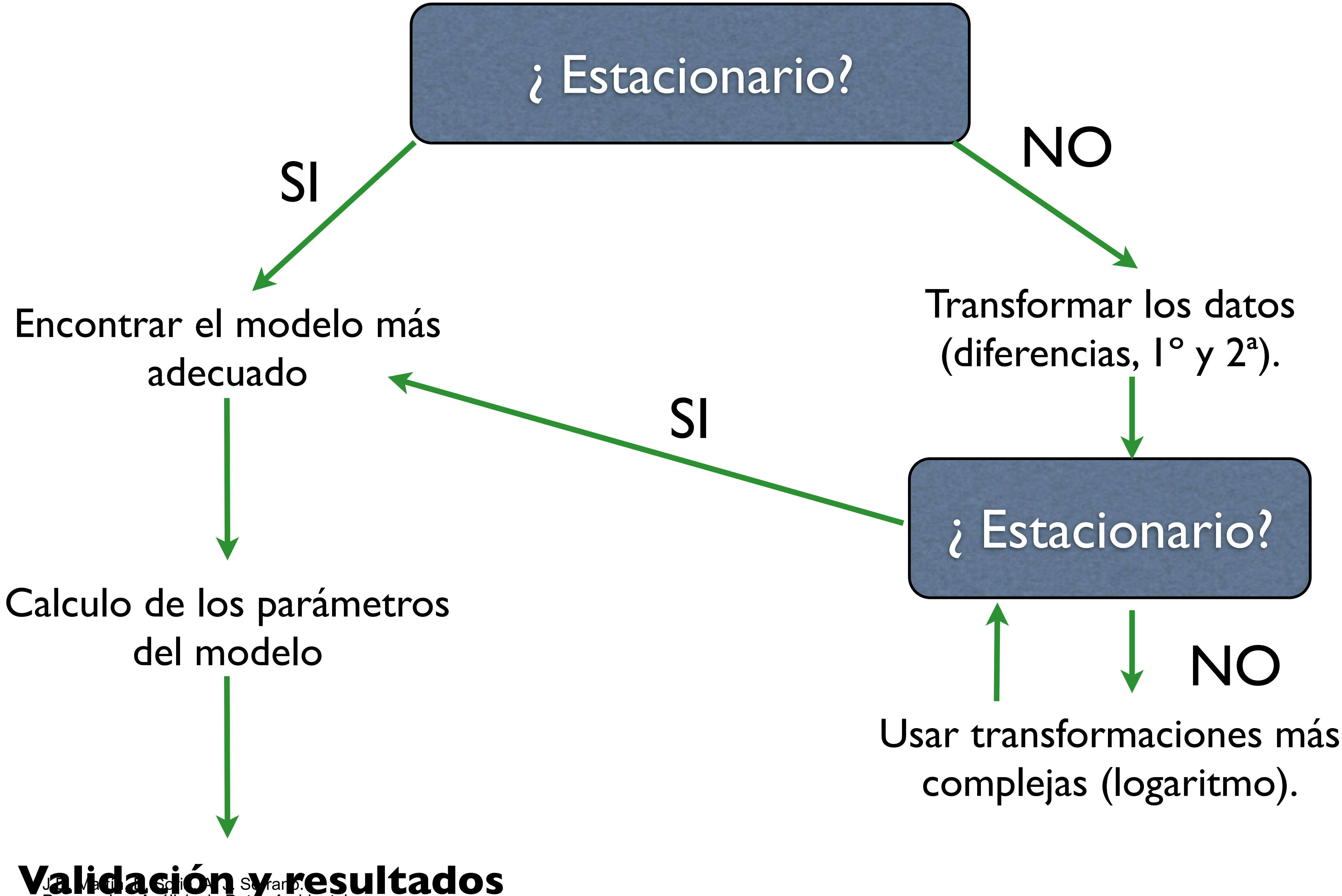
Identificación del modelo (tentativa en base a la información que puede proporcionar la autocorrelación).

Estimación de los parámetros (coeficientes) del modelo.

Evaluación del modelo en base a los ajustes obtenidos.

Generación de resultados.

Todo el proceso......



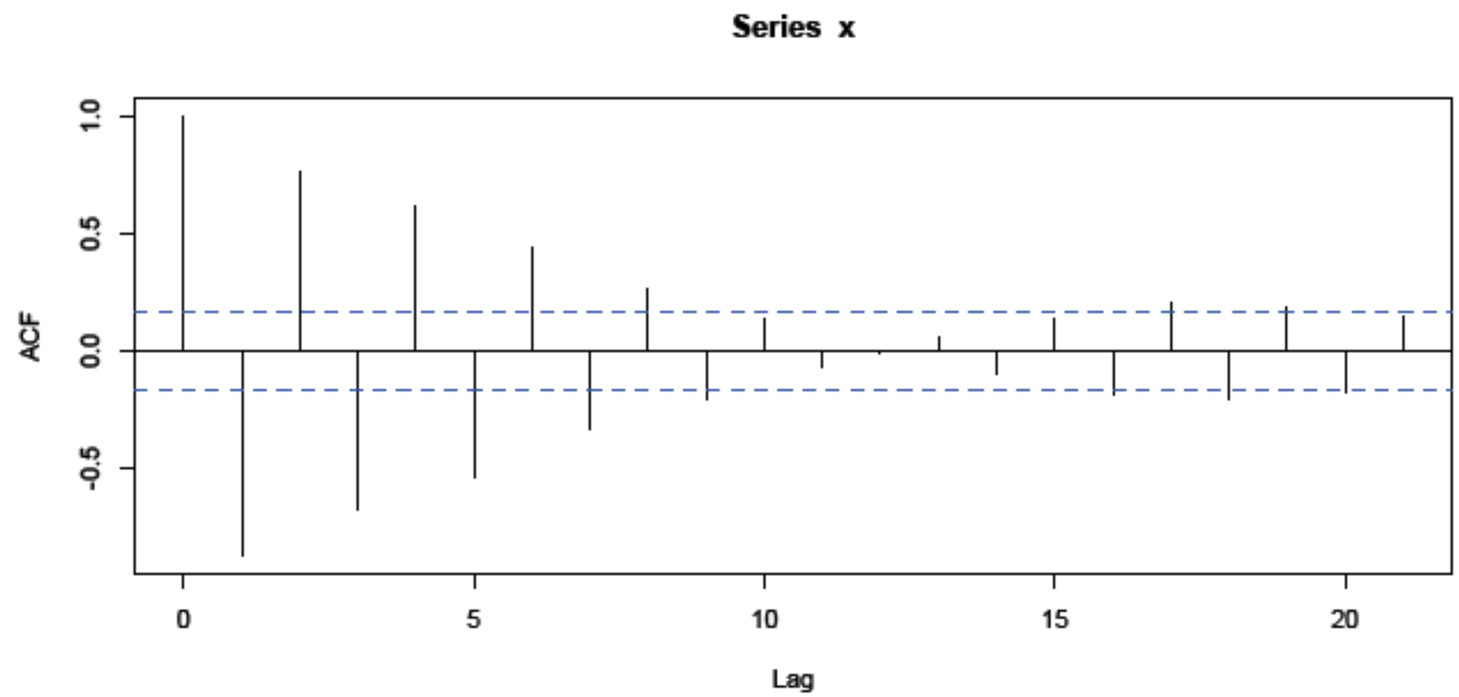
Autocorrelación parcial (PACF)

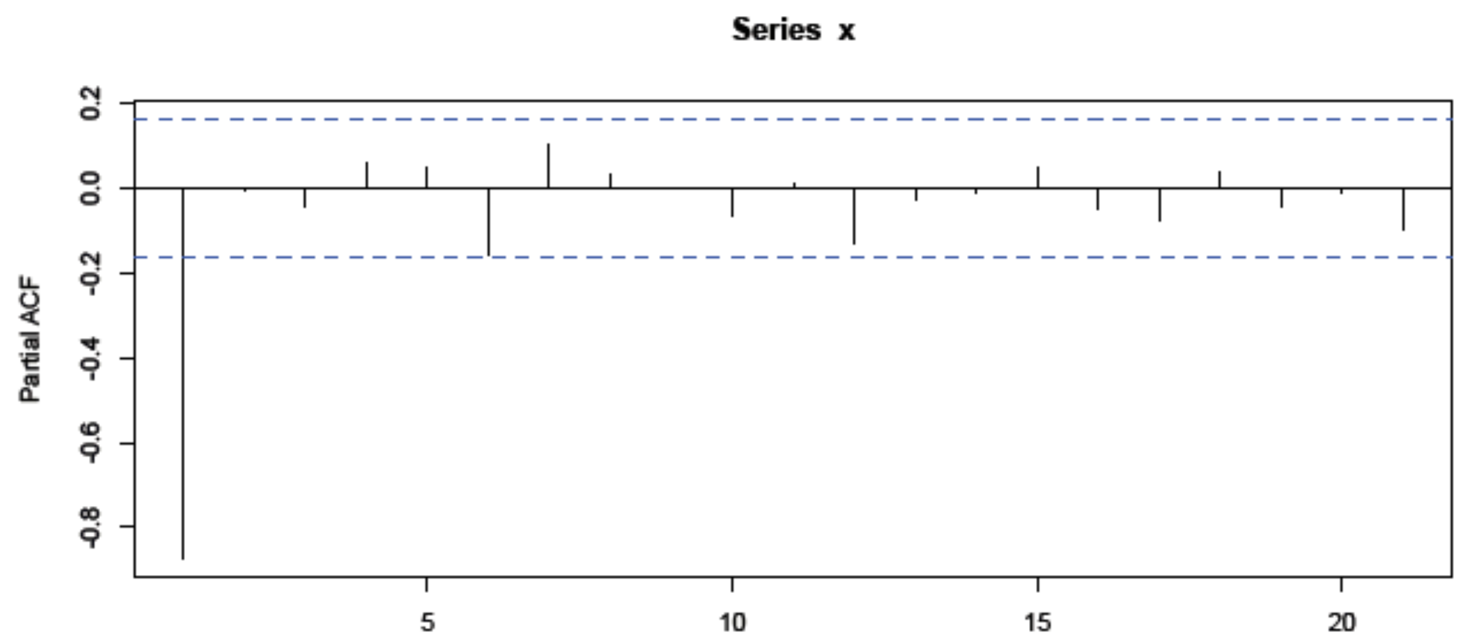
Dado un proceso AR(1) se tiene la siguiente relación:

$$X_t = \theta \cdot X_{t-1} + Z_t$$

Si se itera se llega a:

$$X_t = \theta^2 \cdot X_{t-2} + \theta \cdot Z_{t-1} + Z_t$$





Se tiene una relación entre el término t-2 t t pero no de una forma directa sino a través de la relación entre el término t y el anterior y el t-1 y su anterior.

En el coeficiente de autocorrelación parcial de orden *k*, se calcula la correlación entre parejas de valores separados esa distancia pero eliminando el efecto debido a la correlación producida por retardos anteriores a *k*.

Este coeficiente es muy útil en la identificación de posibles términos autoregresivos

Identificación del modelo usando ACF y PACF

Un proceso de tipo MA(q) tiene una autocorrelación que se corta en el valor q. La autocorrelación parcial de un MA(q) presenta un decaimiento; este decaimiento puede oscilar (tipo seno/coseno amortiguado).

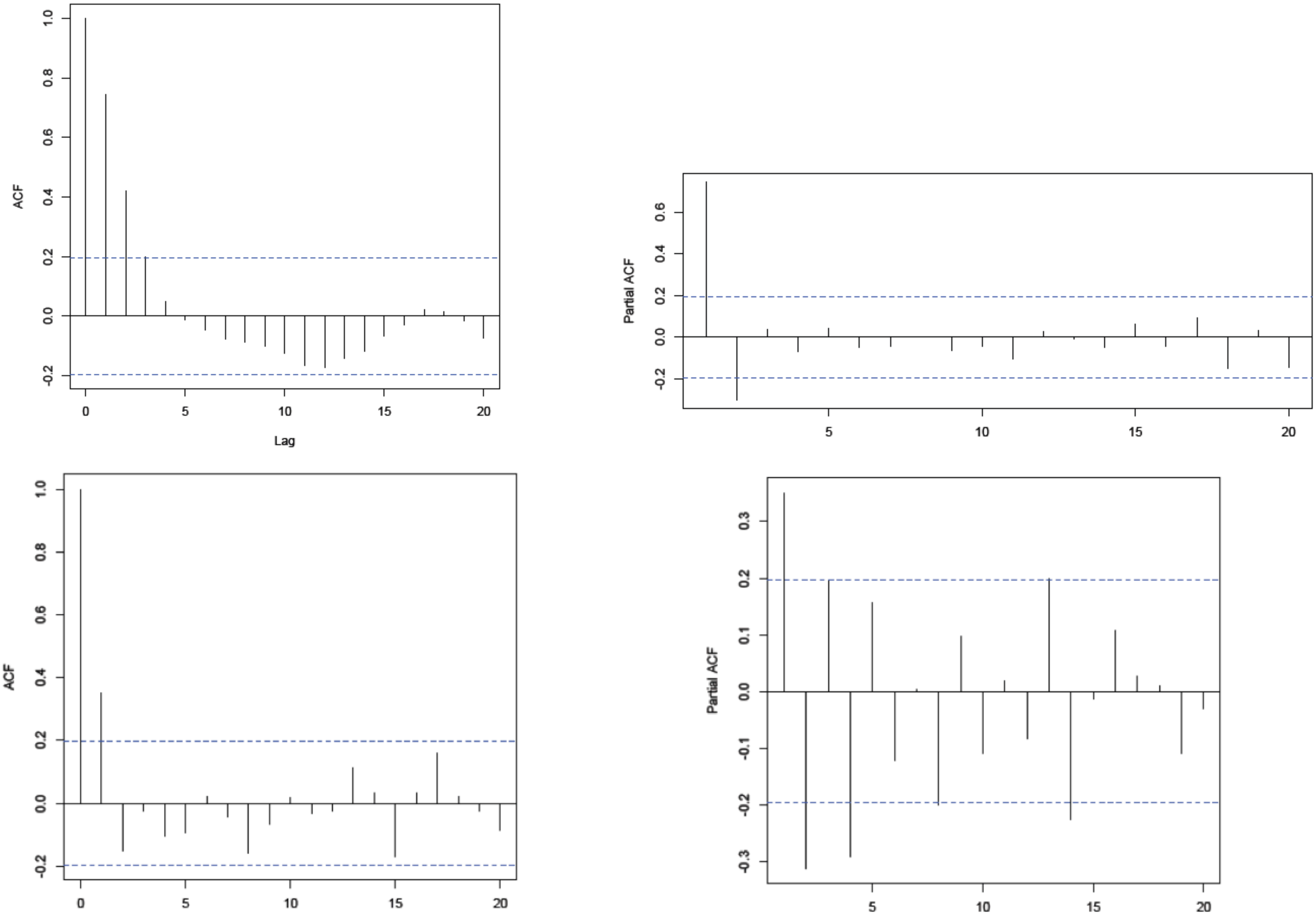
Un proceso AR(p) tiene una autocorrelación que presenta un decaimiento que puede oscilar. La autocorrelación parcial en un AR(p) presenta un punto de corte en p.

Se piensa en un proceso de tipo ARMA(p,q) si se presenta un decaimiento en ambas. El patrón general es que, normalmente, las autocorrelaciones en un ARMA(p,q) siguen un patrón de un proceso AR(p) mientras que las autocorrelaciones parciales siguen un comportamiento de un MA(q).

Si la función de autocorrelación no parece decaer, puede ser que: a) el proceso no es estacionario en covarianza o b) que es un proceso de memoria larga.

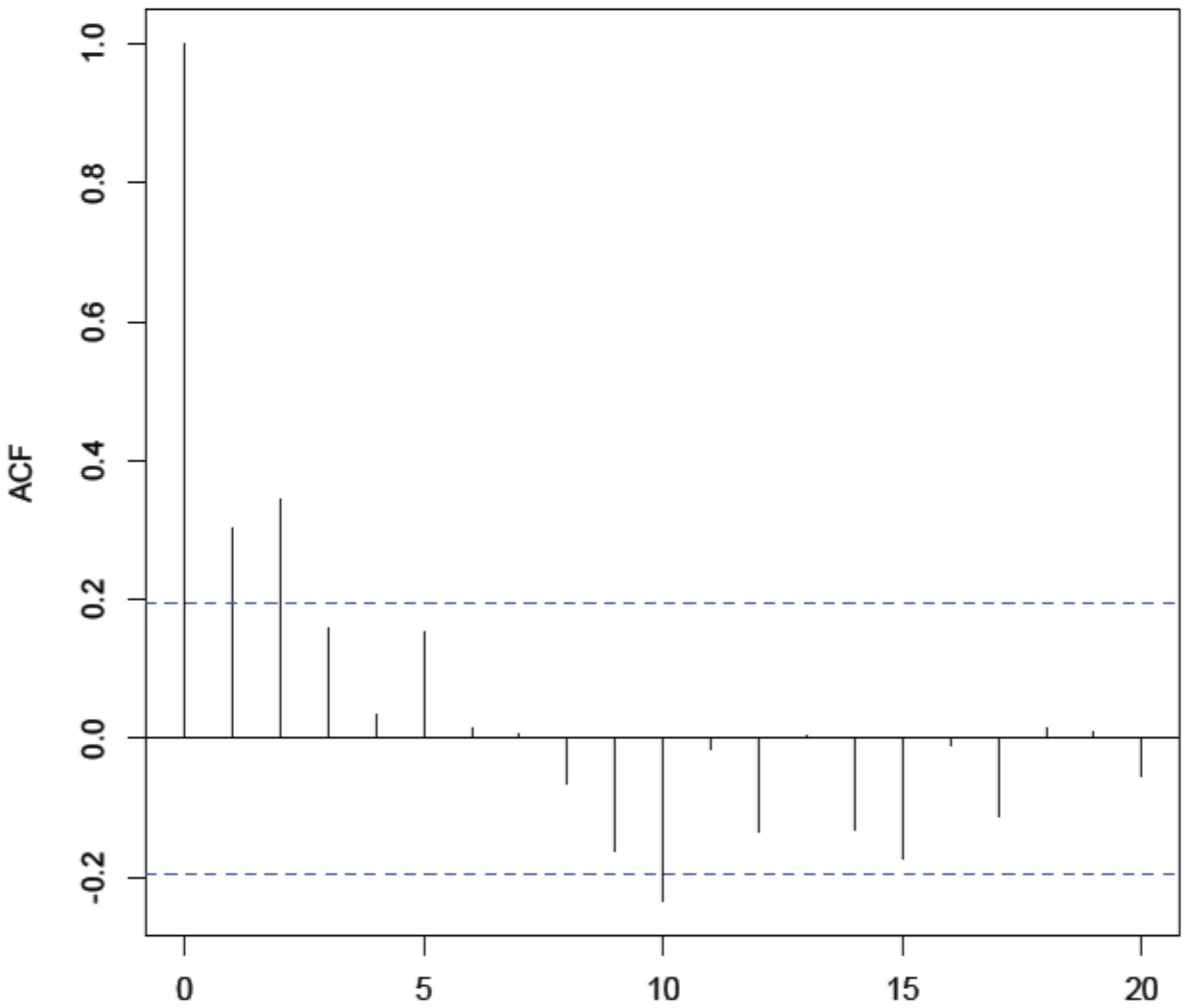
Existen otros criterios basados en teoría de la información o que consideran la longitud del modelo como factor que penaliza dicho modelo. En general, ante la posibilidad de dos modelos se escogerá el más sencillo (navaja de *Occam*).

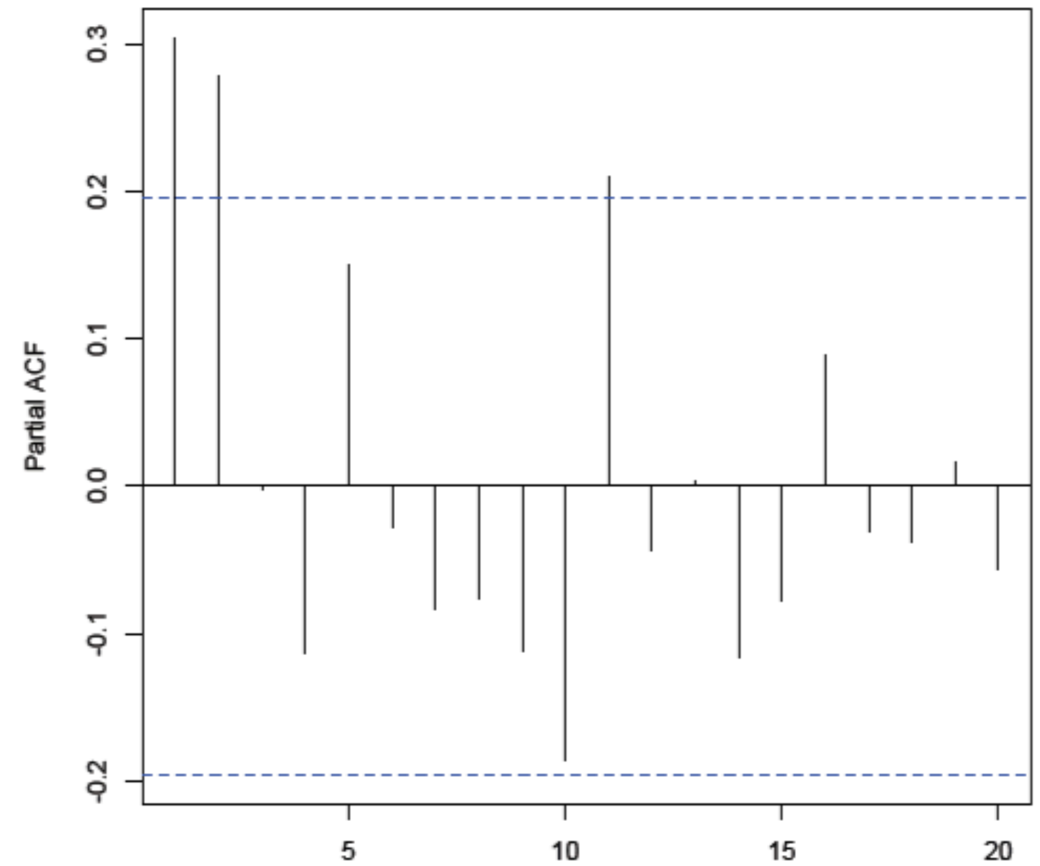
Identificación del modelo usando ACF y PACF



J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Identificación del modelo usando ACF y PACF





J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Problemas de esta metodología

Estacionariedad. Para la aplicación de estos modelos es necesario que la serie temporal sea estacionaria. Muchos procesos naturales son no estacionarios (por ejemplo un EEG presenta características no estacionarias). Una posible solución podría ser considerar (si con los métodos clásicos no se consigue esa propiedad) es analizar la serie por tramos.

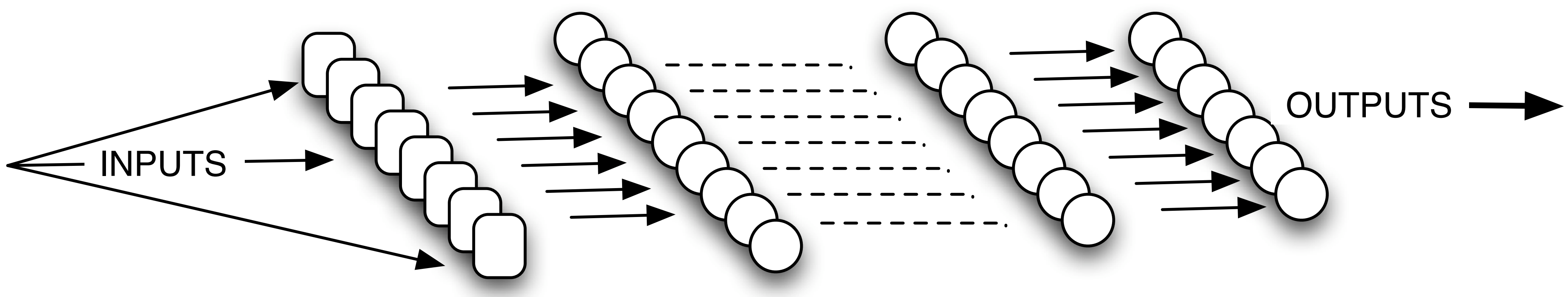
Elección del orden. Como se ha visto en la transparencia anterior la elección del modelo y del orden no deja de ser un poco "artesanal". La influencia del ruido en la serie puede modificar el tipo y orden del modelo escogido y, por tanto, de los resultados obtenidos.

Muestreo no uniforme. En esta metodología se asume que las muestras se consideran a intervalos regulares de tiempo; es muchos problemas clínicos esto no se va a tener (por ejemplo toma de sangre en un hospital).

Dependencias no lineales. En todos los modelos desarrollados las dependencias entre muestras es siempre mediante un coeficiente constante; no se plantean otras dependencias por lo que modelizaciones de sistemas más complejos no es posible con esta metodología.

¿Tenemos alternativas?

La alternativa a los problemas anteriormente comentados es....¡nuestro perceptrón multicapa!



Capa de entrada

Se toman como entradas x(n-1),x(n-2)...así como el resto de factores que influyen en la salida

Se consideran también los tiempos (muestreo no uniforme) que intervienen en la modelización.

Capas ocultas

Se pueden plantear realimentaciones entre neuronas o entre capas (se complica el algoritmo de aprendizaje). Además se pueden usar filtros digitales en lugar de los pesos sinápticos.

Capa de salida

Se pueden considerar diferentes salidas (podemos plantearnos diferentes objetivos a resolver).

Podemos establecer realimentaciones salida-entrada para establecer predicciones a diferentes periodos temporales

Ejemplo: predicción de la concentración de ozono troposférico

Se planteó un modelo para obtener la predicción de la concentración de ozono a partir de datos climáticos y de concentración de otros gases (además del propio ozono en el instante anterior). Se probaron modelos lineales neuronales pero funcionaron mejor los modelos neuronales por:

Es robusto a los datos con ruido.	Es capaz de modelizar relaciones no lineales.	No necesita de ninguna suposición a priori sobre el problema	Son *aproximadores universales*

Además de encontrar un modelizador de la evolución del ozono se buscaba encontrar aquellas variables que tuvieran más importancia en esa evolución mediante un análisis de sensibilidad de los resultados de la red.

Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration
Atmospheric Environment, Volume 40, Issue 32, October 2006, Pages 6173-6180 ; *Juan Gómez-Sanchis, José D. Martín-Guerrero, Emilio Soria-Olivas, Joan Vila-Francés, José L. Carrasco, Secundino del Valle-Tascón*

Primer paso: los datos y su procesado.

Los datos de este estudio corresponden a los meses de Abril de 1997, 1999 y 2000 (el año 1998 no se consideró por falta de datos debido a fallos en los aparatos de medida). Se escogió este mes porque es uno de los meses donde se tiene una mayor concentración de ozono.

El conjunto de datos se dividió en dos subconjuntos: entrenamiento (primeros 20 días del mes) y validación (últimos 10 días del mes); con el primero se entrenaba la red y, con el segundo, se comprobaba su funcionamiento con datos que no había visto. Este conjunto además permitía controlar el grado de generalización de la red neuronal.

Como entradas a la red neuronal se siguieron los consejos de los expertos; se usaron la concentraciones de aquello componentes que entraban en la evolución del ozono (el propio O_3 actual, NO y NO_2) además de variables climatológicas (temperatura, velocidad del viento, humedad relativa, radiación solar y presión atmosférica). Como salida se tomó el valor de la concentración de ozono troposférico para la siguiente hora

Los contaminantes se midieron cada 50 minutos y la variables climatológicas cada 15 minutos por lo que se interpolaron/promediaron para tener valores horarios

Para evitar sesgos de la red los datos se estandarizaron; a cada variable se le quitó su valor medio y se dividió por su varianza (las nuevas variables tenían valor medio cero y varianza unidad).

Segundo paso: la red y su entrenamiento.

Una vez que se tienen los datos preprocesado toca entrenar la red para ello (y por nuestra experiencia) y para ello hay que considerar:

Hacer un barrido en el número de capas ocultas (máximo 2) variando el número de neuronas. No exceder el límite de 2 pesos sinápticos por dato que se tiene	**Para cada arquitectura hacer un número de pruebas con diferentes inicializaciones de los coeficientes. La razón es para no caer en mínimos locales.**	**Probar algoritmos de tipo batch y on-line. Los primeros tienen una menor carga computacional pero no suelen funcionar tan bien como los segundos en problemas de predicción de series**

Esto en Matlab se traduce a tres instrucciones!! (si se tiene la librería Neural Networks Toolbox; si no es así hay que programársela).

newff	Crea una nueva red; podemos fijar el número de capas, número de neuronas por capa y el tipo de función de activación por capa.
train	Entrena la red creada anteriormente, se puede escoger el algoritmo de aprendizaje así como los parámetros de éste.
sim	Sirve para determinar la salida de la red neuronal una vez que se ha entrenado.

Y después de entrenar la red SIEMPRE HAY QUE ANALIZAR LO RESULTADOS; SI ESTOS SON BUENOS TODAVÍA MÁS!!!!

Tercer paso: el análisis de la red (I)

Este análisis va a ser diferente si se tiene un problema de clasificación (se asigna el patrón de entrada a una serie predeterminada de clases) o un problema de modelización/predicción. Para un problema de clasificación de dos clases se tiene

		Disease: Present	Disease: Absent	
Test	+	A	B	A + B
	–	C	D	C + D
		A + C	B + D	

Sensitivity	$= A/(A+C)$	$= \dfrac{\text{true positives}}{\text{true positives} + \text{false negatives}}$
Specificity	$= D/(B+D)$	$= \dfrac{\text{true negatives}}{\text{false positives} + \text{true negatives}}$
False-positive rate	$= B/(B+D)$	$= \dfrac{\text{false positives}}{\text{false positives} + \text{true negatives}}$
False-negative rate	$= C/(A+C)$	$= \dfrac{\text{false negatives}}{\text{true positives} + \text{false negatives}}$
Positive predictive value	$= A/(A+B)$	$= \dfrac{\text{true positives}}{\text{true positives} + \text{false positives}}$
Negative predictive value	$= D/(D+C)$	$= \dfrac{\text{true negatives}}{\text{true negatives} + \text{false negatives}}$
Accuracy	$= \dfrac{A+D}{(A+B+C+D)}$	$\dfrac{\text{true positives} + \text{true negatives}}{\text{all positives} + \text{all negatives}}$

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Medidas para problema de modelización/predicción de series temporales (aquí d_k es la señal deseada de la red y o_k la salida de la red neuronal)

Medidas de precisión (dependen del rango de la señal deseada).

$$MAE = \frac{1}{N} \cdot \sum_{i=0}^{N} |d_i - o_i|$$

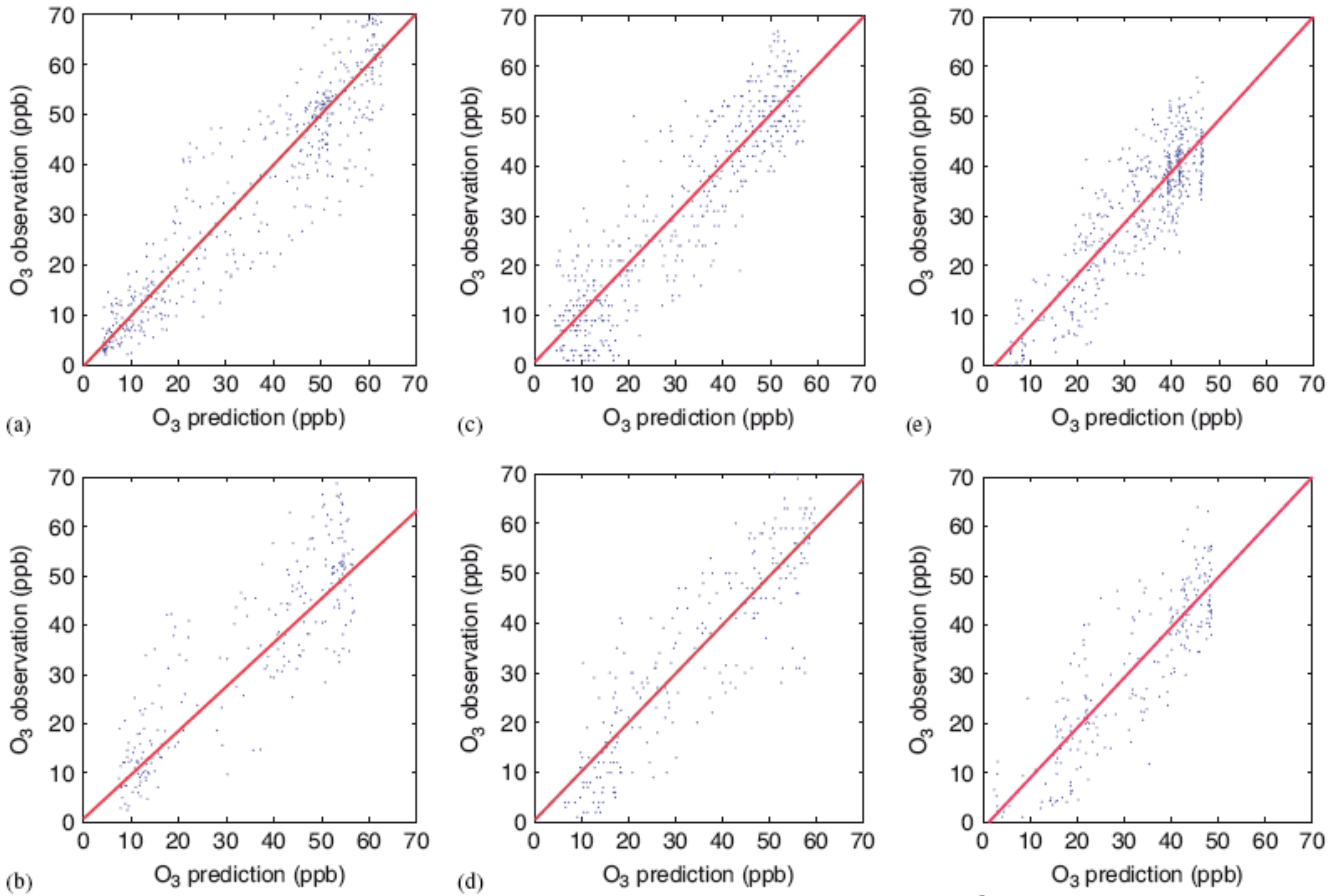
$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=0}^{N} |d_i - o_i|^2}$$

Medida de sesgo

$$ME = \frac{1}{N} \cdot \sum_{i=0}^{N} (d_i - o_i)$$

Tercer paso: el análisis de la red (II)

Existen otras medidas que se obtienen de la representación de la señal deseada frente a la señal obtenida por la red neuronal

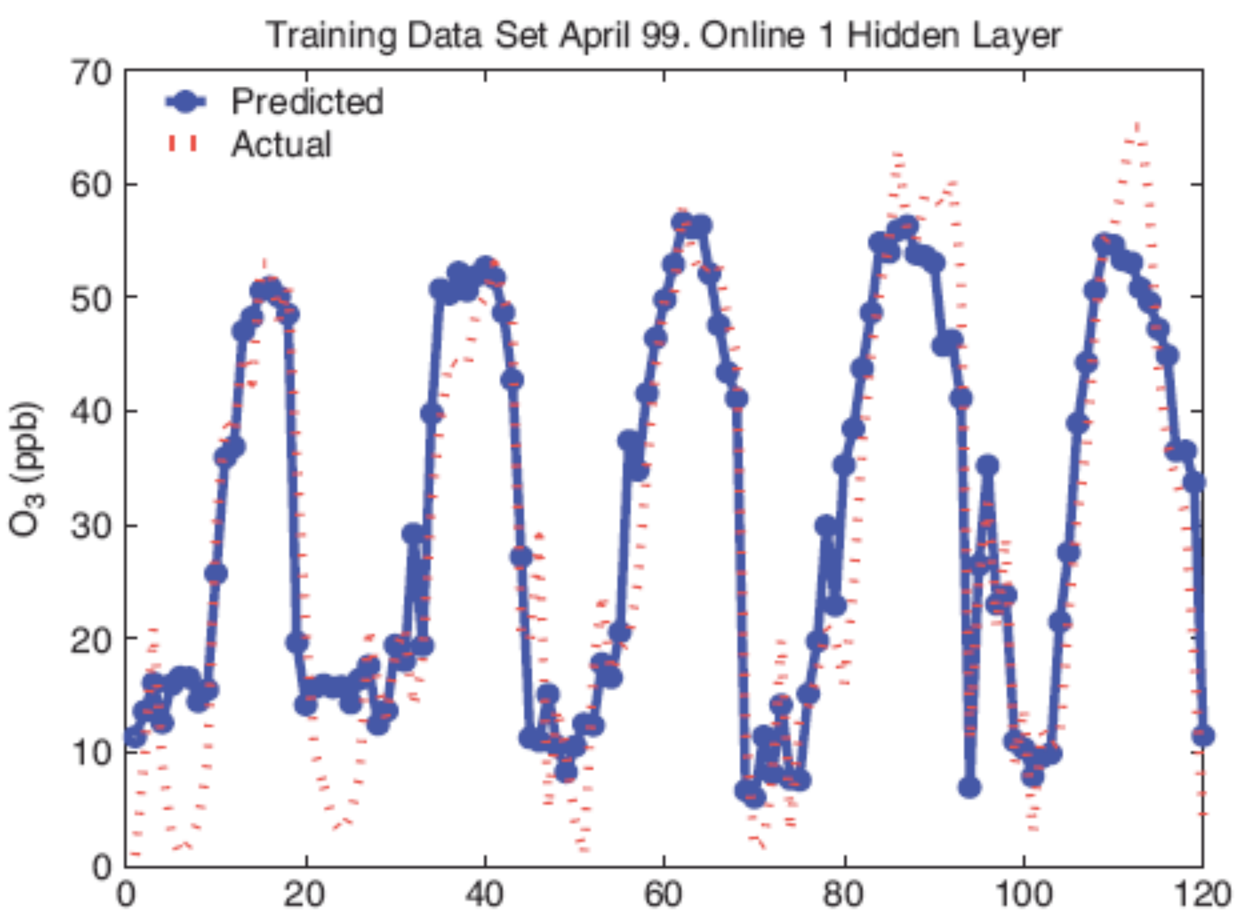


Esas medidas son los índices de ajuste de la recta que pasa por los puntos que se tienen. La pendiente, en una predicción/ modelización ideal debe dar un valor cercano a uno y la ordenada en el origen debe tener un valor cercano a cero.

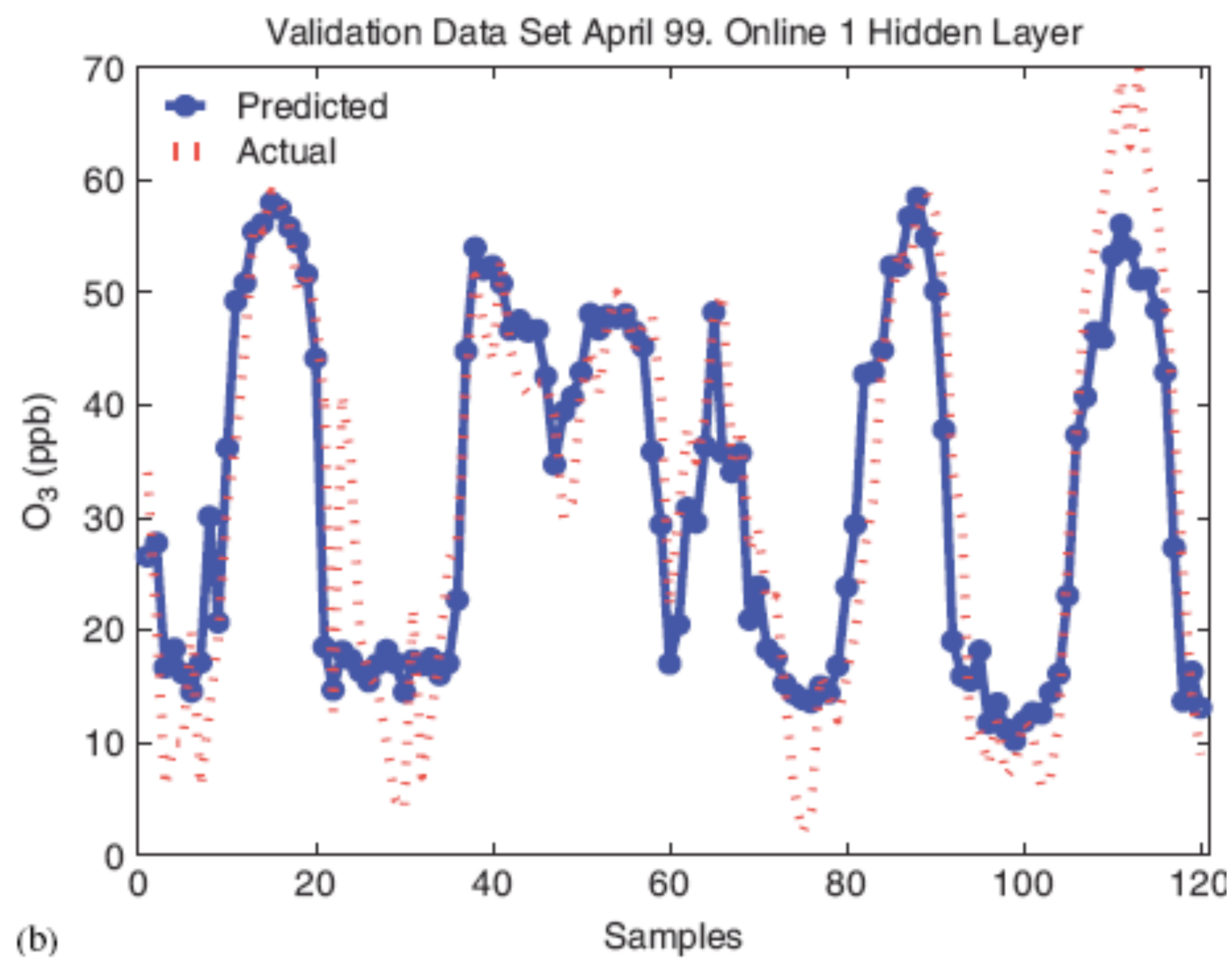
J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Tercer paso: el análisis de la red (III)

	1997		1999		2000	
	Training	Validation	Training	Validation	Training	Validation
MAE	8.94	7.21	6.36	7.05	6.68	5.87
MSE	127.46	81.82	64.15	81.05	66.27	53.44
RMSE	11.29	9.04	8.01	9.03	8.14	7.31
A	0.67	0.74	0.82	0.78	0.60	0.73
B	15.09	10.26	5.07	7.58	12.77	8.58
$RMSE_s$	7.83	4.92	3.31	4.04	5.12	4.21
$RMSE_u$	8.12	7.59	7.29	8.07	6.32	5.98
PSE	0.69	0.54	0.41	0.45	0.63	0.57
d	0.98	0.98	0.94	0.93	0.86	0.93



(a)



(b)

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.

Cuarto paso: análisis de sensibilidad (I).

Este análisis consiste en obtener los N- mejores modelos neuronales y obtener posteriormente la importancia de las variables de entrada. Para ello se determina la diferencia entre la salida con esa variable y anulando esa variable. Si no tiene mucha importancia esa diferencia será cercana a cero; dará igual considerarla o o no. Posteriormente se suman los valores absolutos de esas diferencias para todos los patrones y se ordenan las variables por ese valor; cuanto mayor sea ese valor mayor importancia tendrá la variable .

Sensitivity analysis of input variables

Dataset	Network	Ranking of input variables						
		First	Second	Third	Fourth	Fifth	Sixth	Seventh
1997	□	WS	SI	T	NO_2	RH	P	NO
1997	△	WS	SI	RH	NO	NO_2	T	P
1997	◇	T	SI	WS	P	RH	NO_2	NO
1999	□	T	NO_2	RH	WS	SI	P	NO
1999	△	T	NO_2	WS	SI	RH	P	NO
1999	◇	T	NO_2	RH	SI	WS	P	NO
2000	□	NO_2	RH	WS	T	NO	P	SI
2000	△	RH	NO_2	T	SI	P	WS	NO
2000	◇	RH	NO_2	T	SI	WS	P	NO

□, first best model; △, second best model; ◇, third best model.

Temperatura (T), Velocidad del viento (WS), Humedad relativa (RH), Radiación solar (SI), presión atmosférica (P) y contaminantes.

Cuarto paso: análisis de sensibilidad (II).

Ese análisis lo podemos hacer fijando diferentes horas...

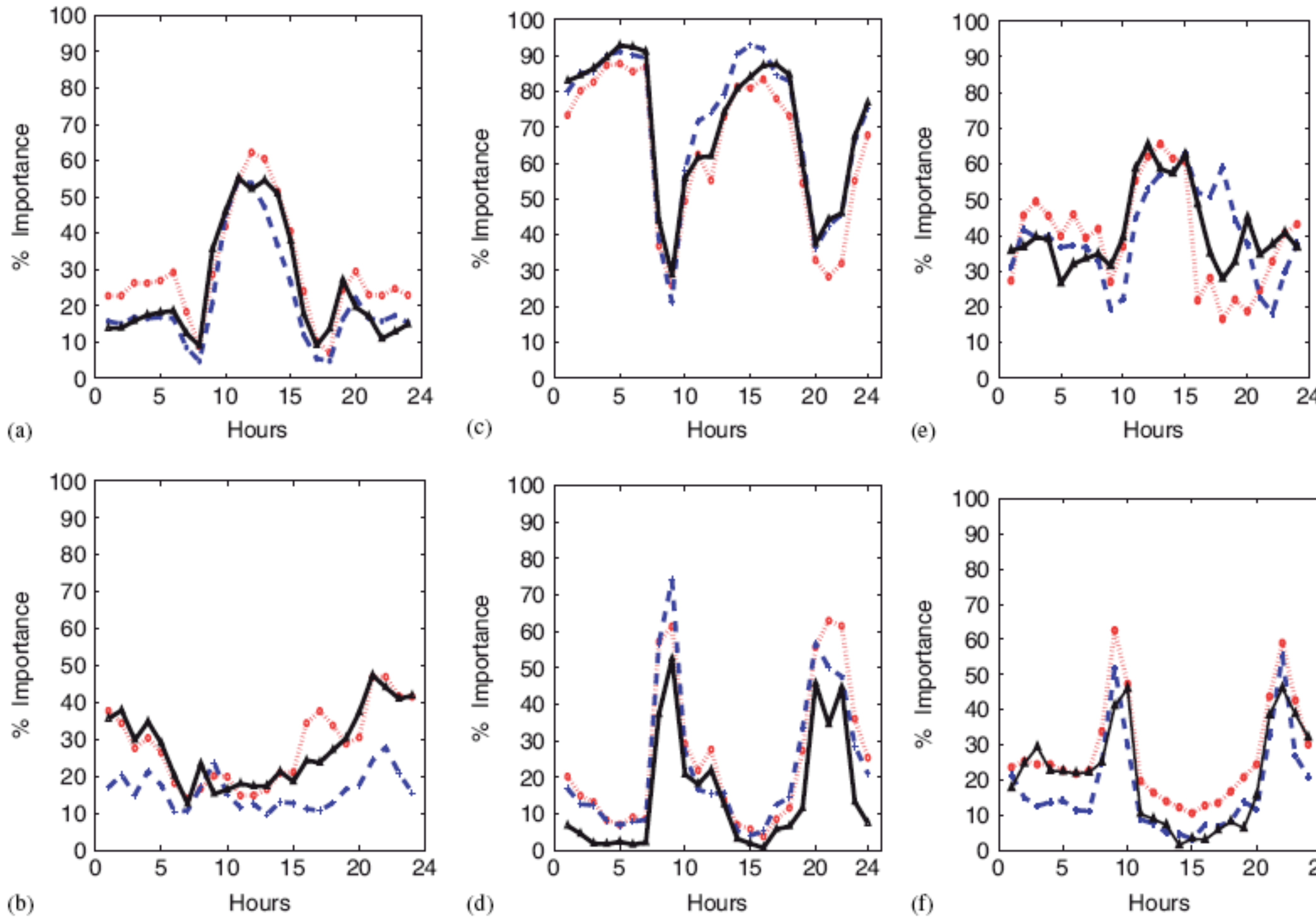


Fig. 4. Relevance of different variables: (a) solar irradiation (1997); (b) wind speed (1997); (c) temperature (1999); (d) NO_2 (1999); (e) relative humidity (2000) and (f) NO_2 (2000). The three lines on each plot represent the three best models.

VNIVERSITAT DE VALÈNCIA

MASTER DE INGENIERÍA BIOMÉDICA.

Métodos de ayuda al diagnóstico clínico.

Seminario: Series temporales (modelos lineales)

All models are wrong, but some are useful." George E. P. Box (1979, p. 202)

J.D. Martín, E. Soria, A. J. Serrano.
Procesado y Análisis de Datos Ambientales.
Curso 2009-2010.