



# Serie temporal de ozono superficial de San Pablo de los Montes. Modelos estadísticos predictivos aplicados al ozono máximo diario

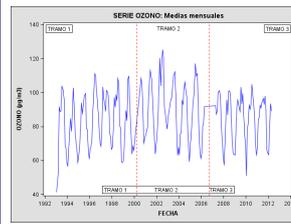
García-Marín, M. (milgarci@ucm.es)<sup>[1]</sup>; Martín, L.<sup>[2]</sup>; García, R.<sup>[2]</sup>; Valencia, J.L.<sup>[2]</sup>.  
(1) Escuela Universitaria de Estadística.  
(2) Agencia Estatal de Meteorología



La red EMEP/VAG/CAMP, gestionada por AEMET y vinculada a diferentes programas internacionales, está dedicada a mantener una vigilancia continua y eficaz de la composición química de la atmósfera, lejos de fuentes contaminantes. En ella se realizan mediciones de numerosos contaminantes atmosféricos para proporcionar datos científicos e información fiable sobre la contaminación atmosférica.

San Pablo de los Montes, situada en las estribaciones de los Montes de Toledo, es la estación más antigua de la red, formada en la actualidad por trece estaciones.

Su serie temporal de ozono se inicia en 1993 y presenta dos lagunas destacables: de enero a julio del 2000 y de mayo de 2006 a abril de 2007, que dividen a la serie en tres tramos diferenciados.



Test de Mann-Kendall					
TRAMO	Años	n	Media	Desv. Std.	Z Nivel signif.
1	1993-1999	64	81.75	16.39	1.32 0.09
2	2000-2006	68	80.30	17.36	-0.96 0.83
3	2007-2011	61	84.38	13.54	-0.35 0.64

## TENDENCIA

La tendencia de la serie de las medias mensuales se ha evaluado por tramos (test no paramétrico de Mann-Kendall), siendo el resultado que no existe tendencia estadísticamente significativa en ninguno de ellos, pues sus niveles de significación están por encima de  $\alpha=0,05$ :

- ✓ Tramo 1: tendencia creciente ( $Z > 0$ ) no significativa ( $p$ -valor=0,09)
- ✓ Tramos 2 y 3: tendencia decreciente ( $Z < 0$ ) no significativa ( $p$ -valores= 0,83 y 0,64)

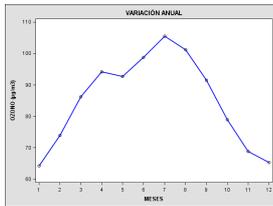
Los modelos se obtienen a partir de la serie temporal: "máximos diarios de ozono 2008-2011" de la estación y se validan con los valores observados del verano de 2012, datos que no se han utilizado en la elaboración de ninguno de ellos. Este estudio se ha realizado con el software estadístico SAS.

## MODELO ARIMA (p,d,q)

La metodología utilizada para ajustar un modelo ARIMA (p,d,q) a series temporales ha sido la de BOX-JENKINS, basada en:

- 1º Identificación del modelo: estacionariedad, ciclos y diferenciación
- 2º Estimación parámetros: ecuación
- 3º Validez del modelo: análisis de los residuos
- 4º Predicción

Al aplicar esta metodología a la serie "máximos diarios de ozono 2008-2011", no se obtienen modelos válidos, ya que los residuos resultantes no son ruido blanco, es decir, aun contienen información sobre la serie, que no puede desestimarse.



Este resultado induce a considerar la serie máximos diarios de ozono, verano 2008-2011" ( $x_t$ ), por ser el verano la época del año en la que el ozono alcanza los valores máximos en San Pablo de los Montes.

Tras aplicar la metodología BOX-JENKINS, se ve que la serie  $x_t$  no es estacionaria, presentando un ciclo principal de 92 días. La diferenciación de esta serie ( $x_t - x_{t-92}$ ) da como resultado la serie transformada  $z_t$  que sí cumple la condición de estacionariedad, siendo 92 el valor de "d".

En este caso, todos los modelos probados sobre  $z_t$  son válidos, pues sus residuos son ruido blanco. El óptimo, entre todos ellos, es un modelo ARIMA (3,92,0), que tiene el menor valor del criterio de Akaike (AIC) y es el más simple. Su ecuación es:

$$(1 - B)^{92} x_t = \frac{1}{(1 - 0,636B + 0,172B^2 - 0,186B^3)} a_t$$

$$\left( B^{92} x_t = x_{t-92} \right)$$

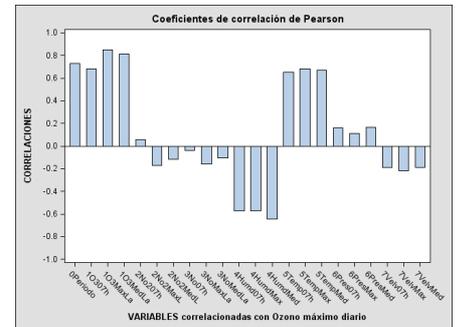
## MODELO REDES NEURONALES

Las redes neuronales permiten predecir el valor máximo de ozono en función de sus valores previos, así como de otras variables medidas en la misma estación y que pueden influir en su formación: monóxido y dióxido de nitrógeno (precursores), humedad, temperatura, presión, velocidad del viento.

Las características de formación del ozono llevan a estimar que su máximo diario dependería de las seis magnitudes citadas anteriormente, así como del propio ozono; siendo los valores propuestos de cada una de ellas:

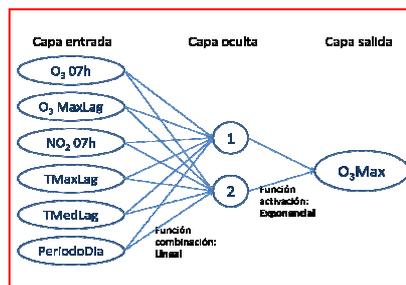
- ✓ Los registrados a las 7h de la mañana (*Var7h*)
  - ✓ Las medias del día anterior (*VarMedLag*)
  - ✓ Los máximos del día anterior (*VarMaxLag*)
- Además, para incluir la componente estacional del ozono se define la variable:
- ✓  $PeriodoDia = \cos((2 * \pi * DiaJuliano / 365) - 0,045)$

Todas ellas constituyen el conjunto inicial de variables.



Este gráfico representa los coeficientes de correlación de Pearson que cuantifican la relación entre el valor máximo diario de ozono y estas 22 variables.

Posteriormente, se crea un modelo lineal con los valores previos de  $O_3$ , con el  $NO_2$  como precursor más importante y con las variables meteorológicas (T y HR) de mayor correlación con el ozono máximo diario. La ecuación de este modelo lineal únicamente tiene como significativas 6 de estas variables que se utilizarán para la capa de entrada del modelo de red neuronal, representado a continuación.



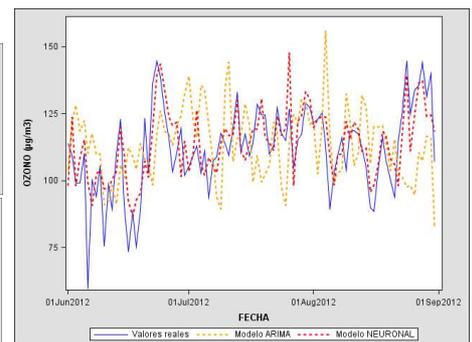
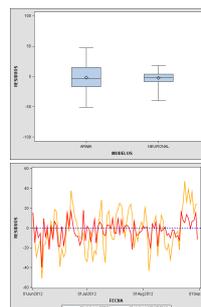
El procedimiento de redes neuronales se prueba variando múltiples parámetros: número de nodos ocultos, funciones de activación y algoritmos; utilizando la técnica de validación cruzada (VC). En cada paso se seleccionan los modelos que generan menos errores. Estos modelos óptimos se contrastan con los datos observados del verano de 2012; y se elige el que presenta un menor valor del AIC, "6DbldogExpB", obtenido con los siguientes parámetros:

- ✓ algoritmo *Dbldog*
- ✓ nodos ocultos 2
- ✓ función de activación: *Exponencial*
- ✓ con constante (*Bias*).

## COMPARACIÓN DE AMBOS MODELOS

### CONCLUSIONES:

- El modelo neuronal trabaja con la serie anual de los máximos diarios de ozono, mientras el modelo ARIMA sólo permite considerar la serie de los máximos diarios de ozono del verano, por lo que el primero aprovecha mejor la información disponible sobre las características estacionales del ozono
- El modelo neuronal utiliza valores de variables más próximos en el tiempo al valor buscado que el modelo ARIMA que necesita valores del verano anterior ( $d=92$ ), lo que facilita la aplicación del modelo
- El análisis de los residuos de ambos modelos indica que la distribución del modelo neuronal es menor, igual en media pero menos dispersa (diagrama de cajas). El ajuste de la salida del modelo neuronal sobre los valores observados es mucho mejor que en el caso del modelo ARIMA
- Por lo tanto, se considera más adecuado el modelo neuronal para predecir el máximo diario de ozono superficial.



### REFERENCIAS:

- Peña, Daniel. *Análisis de series temporales*. Alianza Editorial, 2010.
- Barrero M.A et al. *Prediction of daily ozone concentration maxima in the urban atmosphere*. Science Direct, 2005
- Portela, J. *Manual de programación SAS*, 2007

CONAMA. Congreso Nacional de Medio Ambiente Madrid, del 26 al 30 de noviembre de 2012