

UN SISTEMA DE PREDICCIÓN DE LOS NIVELES DE OZONO MÁXIMOS DIARIOS

E., Agirre,⁽¹⁾ A. Anta⁽²⁾, L.J.R., Barrón⁽³⁾.

⁽¹⁾ *Universidad del País Vasco, Escuela Universitaria de Ingeniería Técnica Industrial, La Casilla, 3 – 48010, Bilbao, elena.agirre@ehu.es*

⁽²⁾ *Analista programador, Vitoria-Gasteiz.*

⁽³⁾ *Universidad del País Vasco, Facultad de Farmacia, Paseo de la Universidad, 7 – 01006, Vitoria-Gasteiz.*

Resumen

Este artículo presenta la eficacia de las redes neuronales artificiales como herramienta en el diseño de un modelo de predicción de los niveles de ozono (troposférico) máximos diarios. Se diseñaron diversos modelos para predecir los niveles de ozono máximos diarios, todos ellos basados en la utilización del perceptrón multicapa. Para construir los modelos se tomaron las mediciones horarias de las variables de meteorología y de las variables de contaminantes atmosféricos más relevantes en la formación del ozono registradas en la Red de Vigilancia y Control de la Calidad del Aire de la Comunidad Autónoma del País Vasco (C.A.P.V.) en el periodo 2001-2003. La bondad de ajuste del modelo fue cuantificada estadísticamente sobre los valores de ozono medidos en el año 2004. El ajuste global y el estudio de la bondad de los modelos en el periodo en el que se pueden superar los umbrales de protección de la salud humana (abril-septiembre) concluyeron con la elección del modelo más eficaz, destacándose la relevancia de la utilización de variables meteorológicas como variables de entrada.

Palabras clave: ozono troposférico, modelo de predicción, redes neuronales artificiales, perceptron multicapa.

1 Introducción

El ozono troposférico (O₃) es uno de los contaminantes atmosféricos más preocupantes en la actualidad. Este contaminante se produce como consecuencia de las reacciones entre los óxidos de nitrógeno (NO_x) con los componentes orgánicos volátiles (COVs) en presencia de la radiación solar. La exposición a ciertos niveles de ozono puede causar daños en la vegetación y en diversos materiales, llegando a ser nocivo para la salud humana. Así, las exposiciones prolongadas a niveles elevados de ozono pueden aumentar estos efectos negativos tanto para los humanos como para el medio ambiente. Consecuentemente, la aplicación de modelos que predigan dichos niveles de ozono de forma eficaz permitiría avisar a la población ante situaciones en las que se fueran a superar los umbrales para proteger la salud humana (tabla 1).

Umbrales	Valor ($\mu\text{g}/\text{m}^3$)	Periodo de referencia
Alerta a la población	240	Promedio horario
Información a la población	180	Promedio horario
Protección de la salud humana	120	Media móvil octohoraria

Tabla 1. Umbrales de ozono de protección de la salud humana (R.D. 1796/2003).

Por otra parte, la capacidad de pronóstico probada mediante la utilización de redes neuronales artificiales en el diseño de modelos de predicción de contaminantes atmosféricos (Gardner y col., 1999; Pérez y col., 2000; Elkamel y col., 2001) y la experiencia de los autores de este artículo (Agirre y col., 2006; Agirre y col., 2005; Agirre y col., 2003) fue determinante en la selección del perceptrón multicapa como herramienta base en la construcción del modelo diseñado para predecir a un día vista los niveles de ozono máximos diarios. El modelo fue elaborado y validado en diversas estaciones de la C.A.P.V., especialmente de la zona de Kostaldea. En este artículo se van a mostrar los resultados obtenidos en la estación de Mundaka.

2 Evolución del ozono en el periodo 2001-2004

Antes de iniciar la construcción del modelo de predicción en cuestión se realizó un análisis estadístico descriptivo de las concentraciones de ozono y de diversas variables meteorológicas registradas en la Red de Control y Vigilancia de la Calidad del Aire de la C.A.P.V. en el periodo 2001-2004 en las estaciones de estudio. En general, en las estaciones de Kostaldea entre el año 2001 y el año 2004 la variación anual del ozono mostró una tendencia creciente, despuntándose los elevados valores de ozono registrados durante el caluroso verano del año 2003. Se observó el carácter cíclico del ozono con concentraciones horarias más elevadas en el periodo comprendido entre los meses de abril y septiembre, coincidente con la época más cálida del año. Asimismo, el estudio horario de los niveles de ozono mostró en general una tendencia creciente que se inicia a las 09:00 h y vuelve a decrecer aproximadamente a partir de las 19:00 h, coincidente con el tiempo en el que se produce más intensamente tanto la actividad antropogénica como la actividad fotoquímica. En general, los picos de ozono diarios se registraron principalmente entre las 14:00 h y las 16:00 h. Finalmente, respecto de las superaciones del umbral de ozono referentes a la salud humana, en ningún caso se superó el umbral de alerta a la población en las estaciones del estudio en el periodo 2001-2004; el umbral de ozono de información a la población se superó entre 5 y 12 ocasiones durante los días 29 de mayo, 21 de junio, 4 de agosto y 8 de agosto del año 2003. Tal y como se puede observar, este umbral sólo se superó en la época de verano del año 2003. Finalmente, las superaciones del umbral de protección de la salud humana se produjeron principalmente entre los meses de abril y septiembre de 2001-2004 y el mayor número de superaciones de este umbral se produjo principalmente durante el mes de agosto de 2003.

3 Elaboración y validación del modelo de predicción

3.1 Determinación del modelo matemático-computacional

Los modelos de predicción basados en el uso de redes neuronales artificiales son modelos matemático-computacionales que poseen una elevada capacidad de generalización y de tratamiento de problemas tanto lineales como no lineales, y no requieren del conocimiento exhaustivo de la distribución de la variable de estudio. Los diferentes tipos de conexión entre las neuronas y los algoritmos de aprendizaje dan lugar a diversos tipos de redes neuronales artificiales. El perceptrón multicapa (MLP, *Multilayer Perceptron*) es la red neuronal artificial con mayor número de aplicaciones prácticas (Rumelhart y col., 1986).

El perceptrón multicapa consta de al menos tres capas: la capa de entrada, la(s) capa(s) intermedia(s) o capa(s) oculta(s) y la capa de salida. Se pueden diseñar una o varias capas intermedias. La figura 1 muestra un perceptrón multicapa con 9 neuronas en la capa de entrada, 4 neuronas en una única capa intermedia y 1 neurona en la capa de salida.

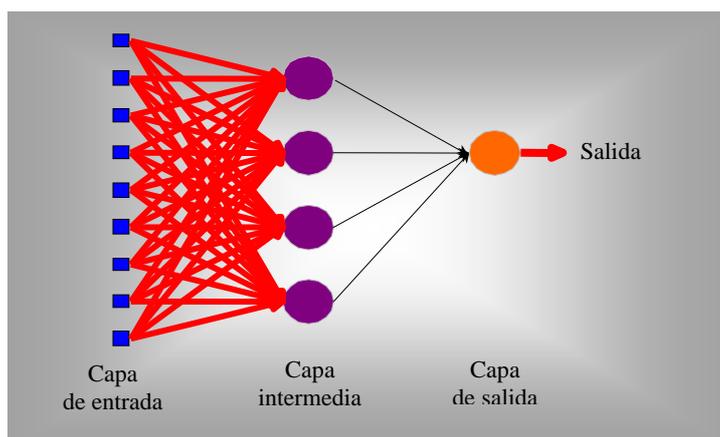


Figura 1. Perceptrón multicapa 9-4-1.

La capa de entrada recibe la información que entra del exterior a la red. Esta información se propaga hacia delante de cada neurona de la capa de entrada a cada neurona de la capa intermedia, y de cada neurona de la capa intermedia hasta cada neurona de la capa siguiente, hasta llegar a cada neurona de la capa de salida. Las entradas se multiplican por los pesos sinápticos y la suma total de estos productos llega a las neuronas de la capa intermedia. A la suma resultante en cada neurona de la capa intermedia se le aplica una función de transferencia, y el resultado de esta operación es considerado como entrada de la siguiente capa. Finalmente, la capa de salida produce el resultado del perceptrón multicapa.

Se activa un proceso de aprendizaje entre los valores objetivo y las salidas del perceptrón multicapa, que es equivalente al proceso de minimización del error E (ecuación 1):

$$E = \frac{1}{L} \sum_{k=1}^L (t_k - y_k)^2 \quad (\text{ecuación 1})$$

donde $(t_1, t_2, \dots, t_L)^t$ es el objetivo, $(y_1, y_2, \dots, y_L)^t$ la salida y L es el número de neuronas de la capa de salida. La salida del perceptrón multicapa se compara con el objetivo, y si el error no es el mínimo, se retrocede hacia atrás en la red neuronal, produciéndose un ajuste en los pesos de la red para que la diferencia entre la salida de la red y la salida objetivo sea menor en la siguiente iteración. Este método de aprendizaje es conocido como método de aprendizaje con retropropagación (*backpropagation*). Tras un número finito de iteraciones, el aprendizaje finaliza una vez alcanzado el mínimo de la función error E .

3.2 Elaboración del modelo

En la elaboración del modelo de predicción de los niveles de ozono máximos diarios primeramente se recogieron los patrones disponibles para elaborar los perceptrones multicapa en cada estación. En siguiente lugar se diseñaron perceptrones multicapa N-S-1, con N neuronas en la capa de entrada, S neuronas en la única capa intermedia y 1 neurona en la capa de salida. N variables de entrada (variables predictoras) a determinar formaron la capa de entrada; una neurona (la predicción del valor máximo de ozono diario en cada estación) formó la capa de salida y para determinar el número de neuronas S de la capa intermedia se aplicó un procedimiento de error-prueba (Amari y col., 1997). En la conexión de la capa de entrada con la capa intermedia se utilizó la función tangente sigmoideal y la función lineal estableció la conexión correspondiente entre la capa intermedia y la capa de salida. El algoritmo de aprendizaje utilizado en este trabajo es el algoritmo de descenso de gradiente escalado, que es un algoritmo derivado del algoritmo con retropropagación o *backpropagation*.

La base de datos utilizada en la elaboración del modelo de predicción de los niveles máximos de ozono diarios a un día vista ($maxO_3(d+1)$) la forman los valores horarios de las variables meteorológicas de temperatura (TE), humedad relativa (HU), presión (PR), radiación solar (RA), dirección del viento (DV) y velocidad del viento (VV) y de las variables de ozono (O_3) y dióxido de nitrógeno (NO_2) registrados en el periodo 2001-2004 en las estaciones de Kostaldea. El modelo de predicción en cuestión se construyó sobre el conjunto de datos del periodo 2001-2003 y su validación fue realizada sobre los datos del año 2004.

3.2.1. Utilización de variables meteorológicas como variables de entrada

Las variables meteorológicas desempeñan un papel relevante en la formación del ozono (Spitchinger y col., 1996), por lo que son consideradas como variables de entrada en la elaboración de modelos de predicción de dicho contaminante atmosférico. Así, en este trabajo se construyeron diferentes perceptrones multicapa con una estructura común, pero con distintas variables de entrada, con el fin de obtener el valor máximo diario de ozono del día siguiente ($maxO_3(d+1)$). Se probó con distintas combinaciones de las siguientes variables de entrada: el valor máximo diario de ozono del día d en el que se

realiza la predicción ($maxO3(d)$), el valor máximo diario de ozono de los tres días anteriores $maxO3(d-1)$, $maxO3(d-2)$ y $maxO3(d-3)$ y los valores de óxido de nitrógeno, temperatura, humedad relativa, presión, radiación solar, dirección del viento y velocidad del viento $NO2(tmax)$, $TE(tmax)$, $HU(tmax)$, $PR(tmax)$, $RA(tmax)$, $DV(tmax)$ y $VV(tmax)$ registrados en el momento $tmax$ en el que se registra el nivel máximo de ozono del día en el que se realiza la predicción. También se utilizaron las variables periódicas $sen(2\pi d/365)$ y $cos(2\pi d/365)$, siendo $d = 1, 2, \dots, 365$ el día del año. De este modo se construyeron los modelos denominados MLP2, MLP3, MLP3*, MLP4 y MLP9, cuyas variables de entrada se recogen en la tabla 2:

Modelo	Variables de entrada
MLP2	$maxO3(d), maxO3(d-1)$
MLP3	$maxO3(d), maxO3(d-1), maxO3(d-2)$
MLP3*	$maxO3(d), maxO3(d-1), cos(2\pi d/365)$
MLP4	$maxO3(d), maxO3(d-1), maxO3(d-2), maxO3(d-3)$
MLP9	$maxO3(d), maxO3(d-1), NO2(tmax), TE(tmax), HU(tmax), PR(tmax), RA(tmax), DV(tmax), VV(tmax)$

Tabla 2. Modelos y variables de entrada

3.3 Bondad del ajuste del modelo

El cálculo de los valores de los estadísticos del Kit de Validación del Modelo (Hanna y col., 1991) sobre el conjunto de test (año 2004) permitió evaluar cuantitativamente la bondad de ajuste de cada modelo, y como consecuencia se seleccionó el modelo más adecuado. El Kit de Validación del Modelo lo forman los cinco estadísticos siguientes:

(i) el coeficiente de correlación, $R = \frac{(C_o - \bar{C}_o)(C_p - \bar{C}_p)}{(SC_p)(SC_o)}$; (ii) el error cuadrático medio

normalizado (*Normalized Mean Square Error*) $NMSE = \frac{(C_o - C_p)^2}{(C_o)(C_p)}$; (iii) el factor de dos

(*factor of two*) $FA2$ que muestra en qué proporción están los valores de la proporción valor predicho/valor real en el intervalo $[0,5, 2]$; (iv) el sesgo fraccional (*Fractional*

Bias) $FB = 2 \frac{\bar{C}_o - \bar{C}_p}{\bar{C}_o + \bar{C}_p}$ y (v) la varianza fraccional (*Fractional Variance*)

$FV = 2 \frac{SC_o - SC_p}{SC_o + SC_p}$, donde C_p es la predicción, C_o el valor real u observación, \bar{C}_p y \bar{C}_o

son las medias aritméticas de C_p y C_o y SC_p y SC_o son las desviaciones típicas de estos valores respectivamente.

4 Resultados

En este trabajo se presentan los resultados obtenidos en la estación de Mundaka. La tabla 3 presenta los valores de los estadísticos del Kit de Validación del Modelo

calculados sobre el conjunto de test (2004) para cada modelo en la estación de Mundaka.

Mundaka	<i>R</i>	<i>NMSE</i>	<i>FA2</i>	<i>FB</i>	<i>FV</i>
Observación	1,0000	0,0000	1	0,0000	0,0000
MLP2	0,9900	0,0072	1	0,0127	0,3807
MLP3	0,9816	0,0055	1	0,0061	0,2878
MLP3*	0,9659	0,0076	1	0,0054	0,3180
MLP4	0,9237	0,0106	1	0,0149	0,2564
MLP9	0,9470	0,0085	1	0,0143	0,2637

Tabla 3. Kit de validación del modelo en Mundaka (2004)

En los cuatro modelos diseñados los valores *R* y *FA2* próximos de la unidad y los valores de *NMSE*, *FB* y *FV* cercanos a 0 prueban la eficacia de predicción de éstos. Según indican globalmente los cinco valores del Kit de Validación del Modelo, el modelo MLP2 que utiliza como variables de entrada el nivel máximo de ozono del día en que se efectúa la predicción ($maxO3(d)$) y el nivel máximo de ozono del día anterior ($maxO3(d-1)$) muestra el ajuste más adecuado para obtener una predicción del máximo de ozono para el día siguiente ($maxO3(d+1)$). En este sentido, la figura 2 permite observar la eficacia de las predicciones de la variable $maxO3(d+1)$ obtenidas mediante los modelos MLP2, MLP3, MLP3*, MLP4 y MLP9 respecto de la observación registrada en la estación de Mundaka durante los meses de junio, julio y agosto del periodo de test (año 2004), meses en los que los niveles máximos diarios de ozono fueron más elevados.

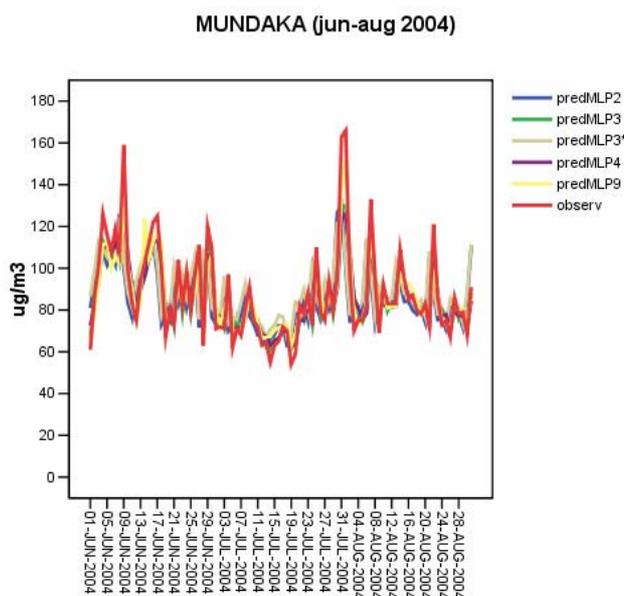


Figura 2. Predicción vs. observación de $maxO3(d+1)$ en Mundaka (verano de 2004).

En este periodo el modelo MLP9 presentó el ajuste más adecuado en las concentraciones más elevadas de la variable $maxO3(d+1)$.

La figura 3 y la figura 4 muestran comparativamente el ajuste de la predicción (output) de la variable $maxO3(d+1)$ respecto de la observación (target) en la primera quincena de agosto de 2004. La figura 3 y la figura 4 muestran una tendencia muy similar, pero tal y como se deduce de estas figuras la predicción del valor máximo de ozono para el día 01 de agosto de 2004 (cuando se registró el valor máximo de ozono anual en la estación de Mundaka) obtenida mediante el modelo MLP9 se ajusta mejor que la obtenida mediante el modelo MLP2.

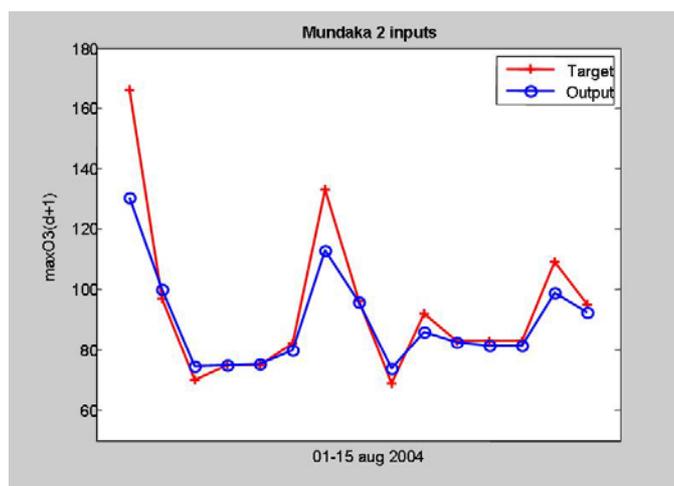


Figura 3. Evolución de la predicción de $maxO3(d+1)$ mediante MLP2 en Mundaka (01/08/2004-15/08/2004).

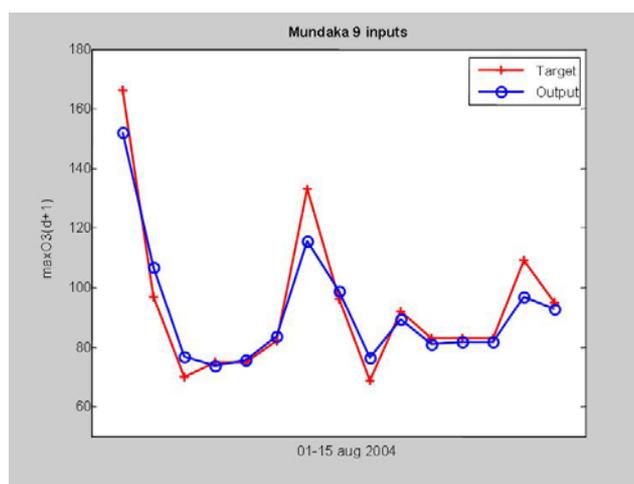


Figura 4. Evolución de la predicción de $maxO3(d+1)$ mediante MLP9 en Mundaka (01/08/2004-15/08/2004).

Luego, es esperable que el modelo MLP9 que utiliza variables meteorológicas como variables de entrada sea más eficaz a la hora de predecir posibles superaciones de los

umbrales de información a la población o de alerta a la población. Por ello, aunque el modelo MLP2 mostrara globalmente el mejor ajuste sobre el conjunto de test total (año 2004), la eficacia del modelo MLP9 en el cálculo de la predicción de los valores más elevados de la variable $maxO3(d+1)$ fue determinante en la selección del modelo MLP9 como modelo de predicción de los niveles de ozono máximos diarios a un día vista con vistas a predecir la superación de los umbrales de protección de la salud humana.

Finalmente, se calculó la diferencia entre la predicción de la variable $maxO3(d+1)$ mediante el modelo MLP9 y de la observación, de modo que en la estación de Mundaka dicha diferencia en valor absoluto no supera los 15 ug/m^3 sobre el conjunto de test total (año 2004), y en particular, en verano de 2004 es a lo más igual a 10 ug/m^3 .

5 Conclusiones

Los resultados obtenidos en este trabajo prueban la eficacia del uso del perceptrón multicapa en el diseño de modelos para predecir los niveles máximos de ozono diarios. Se construyeron diversos modelos en los que la única diferencia residió en la selección de las variables de entrada empleadas. Mientras que los modelos MLP2, MLP3, MLP3* y MLP4 utilizan como variables de entrada los valores máximos de ozono diarios registrados el día en el que se efectúa la predicción y los días anteriores (hasta tres días atrás), el modelo MLP9 introduce también variables meteorológicas como variables de entrada. El cálculo de los valores de los estadísticos del Kit de Validación del Modelo (R , $NMSE$, $FA2$, FB y FV) sobre el conjunto de test (año 2004) y el estudio de las predicciones obtenidas mediante los modelos en el periodo en el que se pueden superar los umbrales de protección de la salud humana (abril-septiembre de 2004) concluyó con la elección del modelo MLP9 como modelo de predicción más eficaz de los valores máximos de ozono diarios. El modelo fue construido sobre la base de datos formada por las concentraciones horarias de los niveles de ozono (O_3), dióxido de nitrógeno (NO_2), temperatura (TE), humedad relativa (HU), presión (PR), radiación solar (RA), dirección del viento (DV) y velocidad del viento (VV) registradas en las estaciones de la zona de estudio en el periodo 2001-2003 y utilizó las nueve variables de entrada siguientes: el valor máximo diario de ozono del día en el que se calcula la predicción ($maxO3(d)$), el valor máximo diario de ozono del día anterior $maxO3(d-1)$ y las variables $NO_2(tmax)$, $TE(tmax)$, $HU(tmax)$, $PR(tmax)$, $RA(tmax)$, $DV(tmax)$ y $VV(tmax)$ registradas en el momento $tmax$ en el que se midió la concentración máxima de ozono del día d .

Se seleccionó el modelo MLP9 por ser potencialmente el modelo más eficaz a la hora de predecir tanto los niveles elevados de las concentraciones de ozono máximas diarias como las superaciones de los umbrales de ozono de información a la población y de alerta a la población. Se podría afirmar que la utilización de variables meteorológicas como variables de entrada del modelo contribuyó a la reducción del fenómeno de subestimación de los niveles de ozono máximos diarios.

Finalmente, el estudio del valor absoluto de la diferencia entre la predicción obtenida mediante el modelo MLP9 y el valor observado del nivel máximo de ozono diario a un día vista probó la eficacia del modelo en la estación de Mundaka, tanto en los meses de verano como durante el resto del año 2004.

Se obtuvieron resultados muy similares en el resto de las estaciones de Kostaldea, que garantizan la utilidad del modelo MLP9 en distintas estaciones, independientemente de su ubicación dentro de la C.A.P.V. y del tipo de estación.

Agradecimientos

Los autores de este trabajo agradecen especialmente al departamento de Medio Ambiente y Ordenación del Territorio del Gobierno Vasco la facilitación de los datos de la Red Automática de Control y Vigilancia de la Calidad del Aire, así como la financiación de los trabajos de investigación que han permitido obtener los resultados mostrados en este artículo.

Referencias

Amari, S.I., Murata, N., Müller, K.R., Finke, M., Yang, H., 1997. *Asymptotic statistical theory of overtraining and cross-validation*. IEEE Transactions on Neural Networks 8, 985-996.

Agirre, E., 2003. *Bilbo hiriko O₃ eta NO₂-aren mailak denbora errealean epe laburrean aurreateko eredu estatistikoaren eraketa*. Tesis doctoral Universidad del País Vasco, Bilbao.

Agirre, E., Anta, A., L.J.R. Barrón, 2005. *Elaboración y evaluación de un modelo de predicción, en tiempo real, de los niveles de ozono en Kostaldea y Rioja Alavesa basado en el uso del perceptrón multicapa* (Gobierno Vasco, nº expediente 175P/2005).

Agirre, E., Ibarra, G., Madariaga, I., 2006. *Regresión and multilayer perceptron based models to forecast hourly O₃ and NO₂ levels in the Bilbao area*. Environmental Modelling and Software 21, 430-446.

Elkamel, A., Abdul-Wahab, S., Bouhamra, W., Alper, E., 2001. *Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach*. Advances in Environmental Research 5, 47-59.

Gardner, M.W., Dorling, S.R., 1999. *Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London*. Atmospheric Environment 33, 2627-2636.

Hanna, S.R., Strimaitis, D.G., Chang, J.C., 1991. *Hazard response modeling uncertainty (a quantitative method). User's guide for software for evaluating hazardous gas dispersion models*. American Petroleum Institute, Washington.

Perez, P., Trier, A., Reyes, J., 2000. *Prediction of PM_{2,5} concentrations several hours in advance using neural networks in Santiago, Chile*. Atmospheric Environment 34, 1189-1196.

- Rumelhart D.E., Hinton G.E., Williams R.J., 1986 “*Learning representations by backpropagation errors*” *Nature* 323, 533-536
- Spitchinger, N., Winterhalter, M., Fabian, P., 1996. *Ozone and Grosswetterlagen. Analysis for the Munich Metropolitan area.* *Environmental Science and Pollution Research* 3, 145-152.