Forecasting Ozone Levels using Artificial Neural Networks

Elena Agirre
Department of Applied Mathematics
University of the Basque Country, Spain

Albaro Anta
Computer Analyst
Spain

Luis Javier R. Barron
Department of Food Technology
University of the Basque Country, Spain

1 Introduction

Ground-level ozone (O$_3$) is one of the photochemical oxidants causing air quality problems. It is formed from chemical reactions among primary air pollutants, such as nitrogen oxides and volatile organic compounds in the presence of sunlight [Finlayson-Pitts and Pitts, 1999]. It can cause adverse impacts on human health and on the environment [WHO, 2003]. O$_3$ could irritate the respiratory system, reduce lung functions, and aggravate asthma problems and chronic lung diseases [EPA, 2008]. Active children, active adults, and people with respiratory diseases are people groups particularly sensitive to ozone. Recently the European Commission has introduced new legislative measures establishing ozone thresholds to inform (180 μg/m$^3$) and warn the population (240 μg/m$^3$) [European Commission, 2008]. Hence, it would be of interest to develop an efficient forecasting model that could be used to inform or alert the population even before these thresholds are exceeded.

The accuracy of artificial neural networks (ANNs) has been proven in forecasting air pollutant levels. The multilayer perceptron (MLP), the first ANN application, was developed in Slovenia and was aimed at predicting SO$_2$ levels [Boznar et al., 1993]. More recently, SO$_2$ predictions from the use of MLP have been contrasted with corresponding predictions based on applied linear regression analysis techniques; MLP predictions were proven to be more accurate [Chelani et al., 2002]. In the same way, ANN-based models have been designed to predict ozone levels and to obtain better predictions than models designed for the same purpose (i.e., ARIMA models or linear regression models) [Yi & Prybutok, 1996, Kao & Huang, 2000]. An ANN-based model was developed in Sweden to forecast NO$_2$ levels [Kolehmainen et al., 2001]. Meanwhile, in London, regression models and ANN-based models designed to predict NO$_x$ and NO$_2$ levels have been contrasted; the ANN-based models yielded better results [Gardner and Dorling, 1999].

AireKal, a research team formed by Agirre, Anta, and Barron, has elaborated on and evaluated
prognostic models based on mathematical-computational methods, such as ANNs. The purpose is to forecast, in real time, hourly ozone levels up to eight hours in advance and daily maximum ozone levels at several stations in the Air Quality Monitoring Network of the Basque Country [Agirre et al., 2006a, Agirre et al., 2007]. After having determined the necessary mathematical equations, a mathematical-computational prognostic model is now being studied in an ongoing research using a computational software developed by the team [Agirre et al., 2009]. Currently, the computational software named airEsan is being exclusively used by the Environmental Department of the Basque Country [Agirre et al., 2008].

This article introduces an ANN-based methodology to elaborate prognostic models with the goal of forecasting ozone levels.

2 Artificial neural networks: The multilayer perceptron

Artificial neural networks (ANNs) are mathematical-computational structures that attempt to simulate the human nervous system, with the neuron being the fundamental element. Artificial neural networks possess the aptitude to learn from the patterns introduced to them and from the errors measured in the learning process. In the end, they are capable of identifying patterns that were never seen previously. ANNs have become famous especially because of their ability to establish non-linear relationships [Chen & Billings, 1990] and due to their broad range of applications in different areas, such as air quality [Gardner & Dorling, 1998], medicine [Blazek et al., 1991], economy [Schöneburg, 1990], and meteorology [Marzban & Stumpf, 1996].

There are different types of ANNs, depending on the structure of the connections, the characteristics of the neurons, and the learning algorithm of the ANN [Hagan et al., 1996]. The multilayer perceptron is one of the most known ANNs.

2.1 The multilayer perceptron

The multilayer perceptron (MLP) is an ANN where neurons are organized in layers. Every neuron at each layer connects to each of the neurons in the following layer. Figure 1 shows an MLP with three layers.

![Figure 1: A three-layer MLP](image)

The first layer consists of the input variables and receives external input information. This information
propagates forwards so that each input is multiplied by the synaptic weights. The total sum of the products is connected to each neuron of the hidden layer. A transfer function is applied to this sum and the result becomes the input of the following layer. The MLP could be formed by one or more hidden layers. Finally, the output layer produces the output of the MLP.

The following equation defines in a matrix form the output of the MLP:

\[
y = f^o(W^o(f^h(W^h x + b^h)) + b^o)
\]

where \( W^o \) is the weight matrix that connects the hidden layer to the output layer, \( W^h \) is the weight matrix that connects the input layer to the hidden layer, \( f^o \) is the transfer function that activates the hidden layer to the output layer, \( f^h \) is the transfer function that activates the input layer to the hidden layer, \( b^o \) is the bias for the output layer, \( b^h \) is the bias for the hidden layer, \( x \) represents the input vector of the MLP and \( y \) defines the output vector of the MLP.

The synaptic weights are determined using a training algorithm. This process requires training patterns \((x_1, t_1), (x_2, t_2), \ldots, (x_p, t_p)\) consisting of \( p \) ordered pairs of \( N \)- and \( M \)-dimensional vectors, where \( N \) is the number of inputs and \( M \) is the number of outputs. It is said that an adequately trained ANN has high generalization capability. Learning starts by changing the values of the synaptic weights of the MLP. The learning procedure is mathematically equivalent to the minimization of the error function,

\[
E = \frac{1}{2} \sum_{i=1}^{p} \|y_i - t_i\|^2
\]

where \( y \) is the output vector and \( t \) is the specified target. The output is compared to the target, and the error is propagated backwards through the network to produce an adjustment in the synaptic weights and biases of the network. Hence, the global difference between the output of the network and the target is minimized. Once the minimum of the difference is reached, the learning ends. This method is known as backpropagation (BP) [Rumelhart et al., 1986].

Afterwards, in the design of the MLP, some factors need to be determined: the number of layers, transference functions between different layers, normalization of the data according to the chosen transference function, input and output variables, the chosen training algorithm, and the rule used to determine the number of neurons in the hidden layer.

### 2.1.1 Number of layers

A single hidden layer MLP is a universal function approximator [Hornik et al., 1989]. Hence, in several studies the three-layer MLP has been chosen as a basic technique in elaborating the prediction model. The input variables (i.e., most often, the real past concentrations of the target variable) are entered into the input layer. The output of the input layer is the input to the hidden layer, and the output of the hidden layer is the input to the output layer. The output layer obtains the output of the MLP.

### 2.1.2 Transfer functions

The transfer function activates the relationships between the neurons of the different layers. Table 1 presents some of the most usual transfer functions:
<table>
<thead>
<tr>
<th>Function</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard limit</td>
<td>( f(x) = \begin{cases} 0, &amp; x &lt; 0 \ 1, &amp; x \geq 0 \end{cases} )</td>
</tr>
<tr>
<td>Symmetric hard limit</td>
<td>( f(x) = \begin{cases} -1, &amp; x &lt; 0 \ +1, &amp; x \geq 0 \end{cases} )</td>
</tr>
<tr>
<td>Linear</td>
<td>( f(x) = x )</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>( f(x) = \frac{1}{1 + e^{-x}} )</td>
</tr>
<tr>
<td>Hyperbolic tangent</td>
<td>( f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} )</td>
</tr>
<tr>
<td>Hard limit</td>
<td>( f(x) = \begin{cases} 0, &amp; x &lt; 0 \ 1, &amp; x \geq 0 \end{cases} )</td>
</tr>
</tbody>
</table>

Table 1: Transfer functions

The designer of the ANN has to select the transfer function for specific training algorithms. Due to their derivability and their non-linearity, the most common transfer functions are the hyperbolic tangent and the sigmoid function.

### 2.1.3 Data normalization

In numerous occasions where sigmoid functions are used mainly as transference functions between the different layers of the MLP, data needs to be normalized. Hence, the use of the hyperbolic tangent function requires normalization of the source data, so that they all have values within the range \([-1, 1]\). Once the output of the MLP is produced, vectors must be de-normalized.

### 2.1.4 Input variables and output variables

The most relevant variables used as input variables to obtain the desired output variable (target) need to be selected.

### 2.1.5 Training algorithm

BP algorithms and its derivations are most commonly used in the learning process of the MLP. Initially, an input is entered; consequently, the MLP produces an output. This output is contrasted with the target, and it is propagated backwards, correcting the values of the weights and thresholds. Thereafter, it propagates forward again and obtains another output that is contrasted with the target. This process is repeated iteratively until it finds the minimum of the function dependent on the difference between the output produced in the MLP and the desired target. The algorithm used in newly designed models needs to be determined from various existing learning algorithms.
2.1.6 Avoiding overtraining

The main objective of the MLP is to prove its generalization capability, that is, while preventing the neural network from learning the singularities of the training set. In situations where the number of training patterns is high, it is convenient to apply techniques to avoid overtraining of the ANN. In this case, the early stopping technique can be applied [Sarle, 1995]. This technique is based on the subdivision of the database into three sets: training set, validation set, and test set. During the training period, the network learns with the data from the training set. It also calculates the values for the synaptic weights and adjusts to them accordingly. The number of iterations in the learning is determined by the data in the validation set so that when an error in the validation set increases, the training is stopped. Finally, the data from the test set, which the MLP does not know until then, is used to measure the goodness of the fit of the results.

2.1.7 Number of neurons in the hidden layer(s)

In the case of a three-layer MLP, the architecture of the generic MLP of the ozone level prediction model is made up of \( N \) neurons in the input layer, \( S \) neurons in the hidden layer, and \( M \) neurons in the output layer. Once input variables are selected, a rule needs to be chosen in order to establish the number of neurons in the hidden layer(s).

3 Methodology in the design of ANN-based models to forecast ozone levels

This section presents a summary of the steps that the authors of this article have followed in the elaboration of models based on the MLP. The aim is to forecast daily maximum ozone levels (in some studies) and hourly concentrations of ozone up to eight hours in advance (in other studies) at several stations of the Air Quality Monitoring Network of the Basque Country.

3.1 Database

The Air Quality Monitoring Network of the Basque Country measures several hourly meteorological parameters and air pollution variables at each station. Meteorological variables are relevant in ozone formation and in the elaboration of prognostic models to forecast the levels of air pollutants [Spitchinger et al., 1996, Bloomfield et al., 1996]. Hence, the database considered in the aforementioned studies was formed based on the hourly measures of ozone (O\(_3\)), nitrogen dioxide (NO\(_2\)), temperature, relative humidity, pressure, solar radiation, wind speed and wind direction registered at several stations of the Air Quality Monitoring Network of the Basque Country (Figure 2).

The selected period should be wide enough to separate the database into two subsets: the first one to be used in building the model, and the second to be used for testing. For this purpose, database containing records measured over three years was used in building the model, and database with data from a single year was used for testing.
3.2 Building the prognostic model to forecast ozone levels

The effectiveness of the MLP for prediction purposes, its ability to generalize, and its applicability in non-linear problems have been the primary reasons behind its selection as a base tool for models designed to predict ozone levels at several stations of the Air Quality Monitoring Network of the Basque Country. The features of the general architecture of the prediction models designed by the authors of this article during their research given in the references at the end of this article, are hereinafter summarized.

3.2.1 A three-layer MLP

Indeed, because the single hidden layer MLP is a universal function approximator, the prediction model to be elaborated was based on the use of the MLP with an input layer, a single hidden layer, and an output layer.

3.2.2 Selection of the transfer functions

The hyperbolic tangent function was selected as the transfer function between the input layer and the hidden layer, and the linear function connected the hidden layer to the output layer.

3.2.3 Normalization of the database

Because the transfer function (hyperbolic tangent) that connects the input layer to the hidden layer takes values from –1 to 1, the variables entered into the MLP must be normalized so that they are also within the range [-1, 1]. For this reason, all of the data have been normalized according to the following equation

\[ v_{\text{norm}} = 1 - 2(v - v)/(v_{\text{max}} - v_{\text{min}}) \]  (3)

where \( v \) is the original variable, \( v_{\text{norm}} \) is the normalized variable, \( v_{\text{max}} \) represents the maximum of the \( v \) va-
riable and \( v_{\text{min}} \) represents the minimum of the \( v \) variable. This way, diverse variables obtain the same treatment. Once the output of the MLP is obtained, the variables are de-normalized according to the following equation:

\[
v = v_{\text{max}} + 0.5(v_{\text{norm}} - 1)(v_{\text{max}} - v_{\text{min}})
\]

### 3.2.4 Inputs and outputs of the model

In the research carried out by the authors of this chapter, the output layer of each MLP was made up of a single neuron: the output of the model. In some studies, this was the daily maximum ozone one day in advance; in other studies, this was the hourly concentration of 1, 2, \ldots, 8 days in advance. To obtain such targets, several MLPs were designed, by which historical values of the most relevant meteorological variables, such as temperature, relative humidity, pressure, solar radiation, wind speed, wind direction, and the past values of \( O_3 \) and \( NO_2 \), were entered as inputs. Furthermore, the precursors of ozone, as well as the activities of the transmission sources, showed temporal dependency that consequently revealed the ozone concentrations that were present [Derwent and Davies, 1994]. This temporal dependency led to the introduction of periodical components in the ozone prediction models [Gardner & Dorling, 2000, Agirre et al., 2006b]. Therefore, temporally dependant variables of the functions of \( \sin \) and \( \cos \) could be entered as input variables. Hence, the following variables could be used: \( \sin(2\pi h / 24) \), \( \cos(2\pi h / 24) \), \( \sin(2\pi d / 7) \), and \( \cos(2\pi d / 7) \), where \( h = 1, 2, \ldots, 24 \) is the hour of the day and \( d = 1, 2, \ldots, 7 \) is the day of the week.

### 3.2.5 The Scaled Conjugate Gradient learning algorithm

In the abovementioned ozone prediction models, the \textit{Scaled Conjugate Gradient (SCG)} algorithm was used [Moller, 1991]. This algorithm converges faster than other conjugate gradient algorithms and has been proven more efficient than the standard BP algorithm [Moller, 1993]. Furthermore, this algorithm is independent on the weights set at the beginning of the training process.

### 3.2.6 Early stopping

In all prognostic models questioned, to avoid overtraining problems, the \textit{early stopping} technique was applied so that the total database in each study was divided into three subsets and used accordingly: 50% of the initial data for the training, 25% of the following data for validation, and the remaining 25% of the data for testing.

### 3.2.7 Establishment of the number of neurons in the hidden layer

To determine the number of neurons in the unique hidden layer of the MLP, a rule using trial and error procedures was applied: The number of patterns of the MLP has to be at least 30 times the number of its parameters [Amari et al., 1997].

### 3.3 Goodness of the fit of the prognostic model

After building the prognostic model to forecast the ozone levels, the calculation of several statistics on the test set allows the study to determine quantitatively the accuracy of the forecasts with respect to the ozone observations. In the five statistics of the Model Validation Kit [Hanna et al., 1991] listed below, \( C_o \) represents the observed value, \( C_p \) is the forecasted value, \( \text{Mean} \) is the mean value, and \( S \) is the standard
deviation:

1. The correlation coefficient, $R$:

$$R = \frac{\text{Mean}[(C_o - \text{Mean}(C_o))(C_p - \text{Mean}(C_p))]}{\sqrt{\text{Mean}(SC_o)(SC_p)}}$$

(5)

2. Normalized Mean Square Error, $\text{NMSE}$:

$$\text{NMSE} = \frac{\text{Mean}(C_o - C_p)^2}{\text{Mean}(C_o)\text{Mean}(C_p)}$$

(6)

3. The factor of two, $FA2$, which gives the percentage of forecasted cases in which the values of the ratio $C_o/C_p$ are in the range $[0.5, 2]$:

$$0.5 \leq C_o/C_p \leq 2$$

(7)

4. Fractional Bias, $FB$:

$$FB = 2\frac{\text{Mean}(C_o) - \text{Mean}(C_p)}{\text{Mean}(C_o) + \text{Mean}(C_p)}$$

(8)

5. Fractional Variance, $FV$:

$$FV = 2\frac{SC_o - SC_p}{SC_o + SC_p}$$

(9)

The values $R = FA2 = 1$ and $\text{NMSE} = FB = FV = 0$ indicate a very efficient output or forecast.

4 Results of the designed prognostic models

A joint study of the values obtained from the statistics of the Model Validation Kit could show the adequacy of the MLP-based models in forecasting ozone maximum levels one day in advance and to forecast short-term ozone concentrations (up to 8 hours in advance) at several locations in the Basque Country (see the values of the corresponding statistics in Agirre et al., 2006b, and Agirre et al., 2007).

Additionally, a graphical representation of the output of the model (forecasts of ozone) versus the target (observations of ozone) could show the goodness of fit. As an example, Figure 3 presents the forecast of the daily maximum ozone one day in advance (in blue) and the observations of the daily maximum ozone levels (in red) at the Mundaka Station from 2001 to 2004. Data from 2001 to 2003 were used to build the prediction model while data from 2004 composed the test set.

It is possible to assert that the model fits accurately, but overfitting and underfitting cases have been noticed. Overfitting occurs in cases where the daily maximum ozone levels are low. Hence, it is not advisable to improve the model’s fit in these occasions. However, it is advisable to improve the fit of the model in cases where the daily maximum ozone concentrations are higher (overfitting). Specifically, the applicability of the prediction model can be improved as a tool to inform and alert the population on certain situations, such as when the thresholds of $180 \mu g/m^3$ and $240 \mu g/m^3$ are likely to exceed.
It is well known that in the Basque Country, higher ozone concentrations are recorded during the summer period. Therefore, a study on the performance of the model at this period would be particularly interesting. Figure 4 shows the performance of the MLP-based model in forecasting the daily maximum ozone levels one day in advance at the Izki Station from July 2004 to August 2004, the warmest period when the test set was conducted.

Furthermore, the study showed that the difference between the prediction of the daily maximum ozone level one day in advance, and the observed value was limited to the range [-15, 15] (sometimes to lower ranges, as in Figure 5) in at least 85% of the cases during aforementioned test period in summer 2004.
Figure 5. Difference between the forecast and the observed value of the daily maximum ozone levels at Mundaka (June to August 2004).

Figure 5 shows the behavior of the difference between the forecast and the observed value of the daily maximum ozone levels at the Mundaka Station in June to August 2004. In this case, the difference was limited to the range \([-10, 10]\). In addition, this estimation of the forecast of the daily maximum ozone, with a confidence level of at least 94%, spread through the whole year of 2004.

Using the models that the authors of this article have designed the ozone-related air quality indexes [Basque Government, 2009] listed in Table 3 could also be forecasted.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Air quality description</th>
<th>Ozone (µg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>Good</td>
<td>(0, 90]</td>
</tr>
<tr>
<td>Green</td>
<td>Acceptable</td>
<td>(90 – 160]</td>
</tr>
<tr>
<td>Yellow</td>
<td>Moderate</td>
<td>(160 - 180]</td>
</tr>
<tr>
<td>Red</td>
<td>Bad</td>
<td>(180 - 270]</td>
</tr>
<tr>
<td>Brown</td>
<td>Very bad</td>
<td>(270 - 360]</td>
</tr>
<tr>
<td>Purple</td>
<td>Dangerous</td>
<td>&gt;360</td>
</tr>
</tbody>
</table>

Table 3. Air quality index related to ozone in the Basque Country (2009)

Figure 6 shows the daily maximum ozone forecasts (in blue) and observations (in white) at Mundaka from January 2008 to May 2009, where all the ozone concentrations were either “acceptable” or “good.”

Types of results, such as these, can be obtained through airEsan, which was developed by the research team composed of the authors of this article, and which is currently being exclusively used by the Environmental Department of the Basque Country.
5 Conclusions

This article shows the steps that need to be followed in the elaboration and validation of a model to predict ozone levels using a generic model based on the MLP. In addition, it showed several of the results obtained from research as it has been carried out by the authors of the article. The experience of the authors in this research area, as well as the interest shown by the Environmental Department of the Basque Country in the prediction models, led the authors to develop a computational software named airEsan. This software currently obtains the daily maximum ozone concentration one day in advance at several stations of the Air Quality Monitoring Network of the Basque Country. Outputs generated by airEsan are based on the use of a mathematical model. Mathematical-computational prognostic models based on the design of artificial neural networks for forecasting purposes can be very useful in decision-making, planning, and evaluation of air quality, especially if they are capable of forecasting possible exceedances in ozone thresholds for the protection of human health.

Acknowledgments

The authors wish to thank the Environmental Department of the Basque Country for providing them with data from their Air Quality Monitoring Network and for financial support.
References


