A multilayer perceptron-based model to forecast hourly ozone levels in Kostaldea

E. Agirre*¹, A. Anta², L. J. R. Barrón³, M. Albizu⁴

¹University of the Basque Country, Department of Applied Mathematics, Bilbao, Spain; ²Egailan, Vitoria-Gasteiz, Spain; ³University of the Basque Country, Department of Food Technology, Vitoria-Gasteiz, Spain; ⁴Basque Government, Environmental Department, Bilbao, Spain

2000 Mathematics Subject Classification. 93A30

Nowadays tropospheric ozone (\(O_3\)) is one of the most relevant air pollutants in the Basque Country (North Central Spain). High ozone levels cause human health problems and damage plants and certain materials and long exposures to ozone increase the negative effects. Several techniques as multiple linear regression, time series analysis and more recently artificial neural networks have been applied in order to predict ozone levels [1] Previous work proved that MLP based models are more efficient in the forecast of ozone and nitrogen dioxide eight hours ahead in urban areas [2]. In the same way, the aim of the present study was to build a prognostic model to forecast \(O_3\) levels in real time up to eight hours ahead at five stations located on the coast of the Basque Country, in the area known as Kostaldea. The data used were hourly concentrations of ozone, nitrogen dioxide, temperature, pressure, relative humidity, radiation, wind speed and wind direction in Kostaldea during the period 2001–2004. Those data were measured from the air pollution network managed by the Environmental Department of the Basque Government. Additionally the variables \(\sin(2\pi h/24), \cos(2\pi h/24), \sin(2\pi d/7)\) and \(\cos(2\pi d/7)\) were calculated, being \(d = 1, 2, \ldots, 7\) the day of the week and \(h = 1, 2, \ldots, 24\) the hour of the day. Data from the period 2001–2003 were used to build the models. The model was validated on the test set, formed by data from 2004. The statistics of the Model Validation Kit [3] determined the goodness of the fit in a quantitative manner, by calculating the correlation coefficient between the observed values and the predictions, the normalized mean square error, the factor of two, the fractional bias and the fractional variance. The results shown in this study were developed in a project sponsored by The Environmental Department of The Basque Government.