Science of the Total Environment 488–489 (2014) 197–207

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Can artificial neural networks be used to predict the origin of ozone episodes?

T. Fontes [a,b,*], L.M. Silva [c,d], M.P. Silva [a], N. Barros [a], A.C. Carvalho [e]

[a] University Fernando Pessoa, Global Change, Energy, Environment and Bioengineering Center (CIAGEB), Praça 9 de Abril, 349, 4249-004 Porto, Portugal
[b] University of Aveiro, Department of Mechanical Engineering/Centre for Mechanical Technology and Automation, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[c] University of Aveiro, Department of Mathematics, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[d] INEB — Instituto de Engenharia Biomédica, Rua do Campo Alegre, 823, 4150-180 Porto, Portugal
[e] New University of Lisbon, Faculty of Sciences and Technology/Center for Environmental and Sustainability Research (CENSE), Quinta da Torre, 2829-516 Caparica, Portugal

HIGHLIGHTS

- ANN can classify the origin of an O_3 episode with a mean error around 2-7%.
- The best classification is obtained when a simpler input combination is used.
- ANN can help authorities to foster O_3 action plans to control exceedances.

ARTICLE INFO

Article history:
Received 5 February 2014
Received in revised form 7 April 2014
Accepted 20 April 2014
Available online xxxx

Editor: P. Kassomenos

Keywords:
Human health
Ozone
Stratosphere
Troposphere
Classification
Artificial neural network

ABSTRACT

Tropospheric ozone is a secondary pollutant having a negative impact on health and environment. To control and minimize such impact the European Community established regulations to promote a clean air all over Europe. However, when an episode is related with natural mechanisms as Stratosphere–Troposphere Exchanges (STE), the benefits of an action plan to minimize precursor emissions are inefficient. Therefore, this work aims to develop a tool to identify the sources of ozone episodes in order to minimize misclassification and thus avoid the implementation of inappropriate air quality plans. For this purpose, an artificial neural network model – the Multilayer Perceptron – is used as a binary classifier of the source of an ozone episode. Long data series, between 2001 and 2010, considering the ozone precursors, ^{7}Be activity and meteorological conditions were used. With this model, 2–7% of a mean error was achieved, which is considered as a good generalization. Accuracy measures for imbalanced data are also discussed. The *MCC* values show a good performance of the model (0.65–0.92). *Precision* and *F_1-measure* indicate that the model specifies a little better the rare class. Thus, the results demonstrate that such a tool can be used to help authorities in the management of ozone, namely when its thresholds are exceeded due natural causes, as the above mentioned STE. Therefore, the resources used to implement an action plan to minimize ozone precursors could be better managed avoiding the implementation of inappropriate measures.

1. Introduction

Due to its oxidative characteristics, the adverse effects of tropospheric ozone (O_3) on human health and on the environment are fully recognized (Agrawal et al., 2003; Dimitriou et al., 2011; Heal et al., 2013). The major health issues found to be related with high ozone concentrations are associated with the respiratory system, provoking or increasing lung irritations and asthma (Halonen et al., 2010; Hamade et al., 2008; Ebi and McGregor, 2008; WHO, 2006; Srebot et al., 2009). This pollutant is classified as a greenhouse gas and is responsible for inducing a reduction of the photosynthetic process thus affecting vegetation growth and reproduction, and hence crop productivity (EEA, 2011). Thereby, the prediction of ozone concentrations and, as possible the identification of its sources are fundamental to improve the effectiveness of public awareness and policies for human health protection, as well as vegetation, and increase the knowledge on the interactions between the ozone concentrations, weather and climate.

To promote a cleaner air in Europe, guidelines, programmes and standards from the World Health Organization (WHO) were included in the Directive 2008/50/EC, of 21 May 2008. This Directive defines the main rules concerning the ambient air quality as well as strategies

* Corresponding author at: University of Aveiro, Department of Mechanical Engineering/Centre for Mechanical Technology and Automation, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.
E-mail address: trfontes@ua.pt (T. Fontes).

http://dx.doi.org/10.1016/j.scitotenv.2014.04.077
0048-9697/

to reduce, minimize and inform citizens about the adverse effects of air pollution. As a result, information thresholds (180 $\mu g \cdot m^{-3}$) and alert thresholds (240 $\mu g \cdot m^{-3}$) were defined for ozone. Moreover, for places where one of these limits is exceeded, Member States shall ensure the implementation of a programme, and/or an action plan, to reduce the ozone concentration and minimize its effects. Nevertheless, the recommended measures may give results when anthropogenic sources are the main ozone precursors and can be useless if the origin of the episode is natural.

The main origin of tropospheric ozone is a series of complex photochemical reactions regulated by natural and anthropogenic precursor emissions, such as volatile organic compounds (VOCs) and nitrogen oxides (NO_x), in the presence of solar radiation (wavelength < 424 nm) (Crutzen et al., 1999; Fishman and Crutzen, 1978). However, the ozone concentrations can also increase due to Stratosphere–Troposphere Exchanges (STE). Fishman and Crutzen (1978) reasoned that the enhancement of tropospheric ozone concentration over the Northern Hemisphere is based on photochemical reactions but its flux is not sufficient to explain the measured concentrations. Stevenson et al. (2006), and Hess and Zbinden (2013) show the importance of the STE in the tropospheric ozone burden. Studies conducted by Hess and Lamarque (2007), and Solomon et al. (2007), conclude that STE are responsible for nearly 20% of the global tropospheric ozone burden and Hegglin and Shepherd (2009) suggest that the stratospheric flux of ozone has been increasing at a nearly constant rate in the North Hemisphere of approximately 2% per decade since 1970. Therefore, this last process cannot be neglected as a dynamical mechanism that also contributes to the tropospheric ozone budget.

The exchange of air masses between the stratosphere and the troposphere in the extratropics results from irreversible diabatic erosion of cut-off lows and diffusive mixing connected with tropopause folds (Elbern et al., 1997), associated with upper level frontogenesis and rapid surface cyclogenesis (Davies and Schuepbach, 1994). Other mechanisms include mesoscale convective systems, thunderstorms and breaking gravity waves (Stohl et al., 2000). The shape of the STE in the extratropics is highly variable in space and time, which induces a high variability in the surface measured ozone concentrations influenced by these dynamical processes.

Stratospheric intrusions are responsible for high or very high levels of ozone in the troposphere with an unknown impact at a regional to local scale (Arsićet et al., 2011; Barros et al., 2004; Carvalho et al., 2005; Fontes et al., 2013; San José et al., 2005). Generally, the signature of this phenomenon follows fine three-dimensional atmospheric movements that only by chance are measured. Very few studies report these high ozone levels. In Madrid (Spain) between 2 h00 and 6 h00 in the 29th of April, 2000 several monitoring stations reported ozone concentration levels up to 1190 $\mu g \cdot m^{-3}$ (San José et al., 2005). In Bor (Servia) ozone concentration reached values of 3000 $\mu g \cdot m^{-3}$ in two episodes, which lasted 3–5 days in November 2010 (Arsić et al., 2011). Other studies report similar results (e.g. Fadnavis et al., 2010; Ganguly, 2012).

In spite of scarcity of extremely high surface ozone measurements, the actual contribution to ozone episodes near the legal values presented above is up to now, difficult to justify and forecast. Unknowing the real reason for an ozone episode may be costly when action plans are enforced by law (EEA, 1999). Thus, to improve the effectiveness of these plans more research on this field is needed in order to identify the origin of these episodes.

In the last decades, the research has been focused in quantifying the ozone levels from different sources and understanding the main processes related with their formation. In fact, few studies have been developed in order to present tools which can help authorities to optimize the costs of an action plan and thus minimize the negative effects of ozone. Some exceptions are the studies presented by Grewe (2006) and Emmons et al. (2012). Grewe (2006) applied a chemistry–climate model E39/C to determine the origin of ozone (using a classification model), and Emmons et al. (2012) present a technique of tagging ozone from various source regions by using artificial tracers of NO and its oxidation products. These are interesting approaches to identify the origin of ozone, but complex and time consuming to implement on the air quality management networks and services.

Instead of using deterministic models, as the ones used in the abovementioned works, other groups of researchers have been focused on the prediction of ozone concentrations using statistical models. These linear and non-linear methods can be used for constructing non-deterministic models that can be used not only to forecast the ozone concentration but also to predict the concentrations of other pollutants as PM_{10} and NO_x (using regression model). Although linear models are acceptable, non-linear models capture the non-linearity of ozone response and thus have been one of the preferred statistical tools for ozone prediction in the last years (Chattopadhyay and Bandyopadhyay, 2007; Hájek and Olej, 2012; Sellitto et al., 2011; Taormina et al., 2011; Tsai et al., 2009; Wang et al., 2003).

In these studies, chemical and/or meteorological input variables were used. Most of the studies use chemical variables related with ozone and its precursors' concentrations (NO, NO_2) (e.g. Chattopadhyay and Chattopadhyay, 2012; Moustris et al., 2012). In addition, in some of the studies other pollutants, such as carbon oxide — CO, particles — PM (PM_{10} and/or $PM_{2.5}$), and sulfur dioxide — SO_2, are also included (e.g. Hájek and Olej, 2012; Özbay et al. 2011; Zhang et al., 2012). Concerning meteorological variables, the most commonly used variables are temperature, wind speed, wind direction, solar radiation and relative humidity. Additionally, some authors include other meteorological variables such as the Solar Zenith Angle (SZA), the reflectance data and the Total Ozone Column (TOC) (Chattopadhyay et al., 2012; Sellitto et al., 2011), rain (Chattopadhyay and Chattopadhyay, 2012; Özbay et al., 2011) and the cloud cover (Chattopadhyay and Chattopadhyay, 2012; Tsai et al., 2009).

In general, the literature review shows a lack of knowledge about the prediction of the origin of ozone episodes, with the underlying assumption that the near surface ozone concentrations are mainly resulting from photochemistry reaction products. In addition, the literature review on ozone regression models shows that the large majority of studies were developed to predict the surface ozone concentrations using chemical and meteorological data. Although these regression models (e.g. Hájek and Olej, 2012; Sellitto et al., 2011) are simpler than those classification models (e.g. Emmons et al., 2012; Grewe, 2006) they give information only on the ozone concentration level but their origin is not addressed.

Therefore, the main purpose of the work presented here is to fill this gap and include the ozone origin into an artificial neural network model for ozone forecast purpose. To achieve this goal the MLP architecture was used to build a binary classifier of the origin of ozone episodes, anthropogenic vs natural. The objective of the development here presented is to understand if an intelligent computational tool can be used to help authorities to improve the efficiency of ozone concentration management mainly when ozone concentrations exceed threshold values due to natural causes, like STE. Although STE are classified as rare events, in this case the implementation of control programmes is clearly ineffective. For this reason, knowing the origin of ozone may be of paramount importance in order to improve the actuation of the air quality management services improving the use of the available resources.

2. Materials and methods

To predict the origin of ozone episodes between anthropogenic (e.g. photochemical due to the presence of anthropogenic precursors) and natural (in the present study mainly due to STE), four main steps were defined: (i) firstly, air quality and meteorological data were collected from different sources (see Subsection 2.1); (ii) secondly, the classification of ozone episodes was then performed by experts in the field and also based on knowledge available in the literature (see Subsection 2.2); (iii) in the third step, different input scenarios were evaluated, in