Atmospheric Environment 90 (2014) 59–70

Contents lists available at ScienceDirect

Atmospheric Environment

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

Proposal for estimating ground-level ozone concentrations at urban areas based on multivariate statistical methods

P. Pavón-Domínguez*, F.J. Jiménez-Hornero, E. Gutiérrez de Ravé

Dept. of Graphic Engineering and Geomatics, University of Córdoba, Gregor Mendel Building (3rd Floor), Campus Rabanales, 14071 Córdoba, Spain

HIGHLIGHTS

- A methodology for accurately estimating tropospheric ozone in urban areas is proposed.
- Urban areas have not been considered homogeneous for estimation purposes.
- Previous ozone concentration and solar radiation appear in all proposed models.
- Chemical variables are far more relevant in the city than in the outskirts.

ARTICLE INFO

Article history:
Received 12 December 2013
Received in revised form
12 March 2014
Accepted 17 March 2014
Available online 18 March 2014

Keywords:
Ozone
Air pollution
Urban areas
Estimation
Multivariate analysis

ABSTRACT

This study focuses on describing ozone patterns and estimating ozone concentrations in urban settings through the classification of an urban area into homogeneous typologies, according to hourly ozone concentrations, and the development of accurate estimation models for each typology. For these proposals, a hierarchical cluster analysis was conducted in order to define homogeneous subareas, and multiple linear regressions were subsequently applied with the aim of obtaining ozone predictions, employing chemical and meteorological variables as predictors. Seville metropolitan area (Spain) is a densely populated area of the Mediterranean Basin that exhibits environmental problems related to ozone pollution episodes. Ozone exceedances are a consequence of the combination of road traffic and industry emissions with hot temperatures and high solar radiation, mainly during anticyclonic events. Cluster analysis evince that this area can be divided into 3 categories according to hourly ozone concentration in summer. Cluster 1 is comprised of monitoring stations located in the outskirts of the city of Seville; Cluster 2 corresponds to monitoring stations located within the city of Seville; and Cluster 3 is comprised of a monitoring station specialized in traffic emissions. Multiple linear regression shows that the relative weight of meteorological variables decreases when moving from the urban periphery towards the urban center, whereas the weight of chemical variables increases. Coefficients of determination (R^2) values were 0.885, 0.890 and 0.830 and root mean squared error (RMSE) were 11.226, 11.874 and 11.260 μg m^{-3} for Cluster 1, 2 and 3, respectively.

1. Introduction and objectives

Nowadays ozone is considered as one of the most significant air pollutants owing to the fact that it severely affects plant tissues and human health. Ozone is formed as a result of photochemical reactions in presence of sunlight involving anthropogenic emissions, such as nitrogen oxides and volatile organic compounds (Crutzen, 1979; Derwent et al., 2003; Sillman, 1999). Ozone formation is favored by certain atmospheric conditions, such as atmospheric stability, high solar radiation and temperatures. Thresholds of ozone are often exceeded in the Mediterranean Basin (Fernández-Fernández et al., 2011; Güsten et al., 1994; Kalabokas et al., 2000; Ribas and Peñuelas, 2004). Mediterranean region is characterized not only by photochemical episodes in urban areas, but also high background ozone concentrations (Velchev et al., 2011). Stratosphere-to-troposphere transport is known to be a relevant factor which contributes to the observed background ozone levels in the troposphere. This vertical transport, which is important in the Mediterranean region, feeds stratospheric ozone into the upper troposphere, and these ozone-rich air masses can subsequently be introduced into the lower troposphere (Zanis et al., 2014). Indeed, during anticyclonic synoptic conditions, high ozone concentrations

* Corresponding author.
E-mail address: g22padop@uco.es (P. Pavón-Domínguez).

http://dx.doi.org/10.1016/j.atmosenv.2014.03.032
1352-2310/

are expected to be on the upper part of the boundary layer in the Mediterranean region (Kalabokas et al., 2013). Under these usual meteorological conditions in summer, surface ozone concentrations increase notably due to the downward movements of stagnant masses of ozone from the upper tropospheric air to the surface (Sánchez et al., 2008; Schürmann et al., 2009).

Surface wind regime also influences ozone concentration (Wang et al., 2001; Shan et al., 2010) due to the fact that both direct ozone transport and photochemical reactions take place during the transport of ozone chemical precursors from sources areas to surroundings. Therefore, ozone peak concentrations are related to high atmospheric pressure, slow air movement and high temperature and insolation (NCR, 1991). It is widely known that atmospheric pollution is governed by a complex variability of spatial and temporal scales (Dueñas et al., 2005). From a global point of view, ozone concentrations are influenced by chemical emissions, physical processes and atmospheric conditions. However, photochemical local conditions can also influence ozone concentrations, this being the reason why high ozone episodes are difficult to control and forecast. Multivariate analysis provides a broad range of methods for association, interpretation, modeling and forecasting from large datasets from environmental monitoring programs (Dominick et al., 2012). Cluster analysis is a useful procedure for simplifying and classifying the behavior of environmental pollutants in a specific region (Ludwig et al., 1995; Alonso et al., 2006; Jin et al., 2011; Monteiro et al., 2012). In this study a cluster analysis is carried out in a populated metropolitan area in the west of the Mediterranean Basin (Seville, Spain), where its ozone monitoring stations are grouped into homogeneous clusters. This technique has the advantage that it allows us to maintain ozone assessment in one area while some of the monitoring stations are subject to regular maintenance, and even to reduce the number of stations operating simultaneously.

Once clusters are identified, modeled ozone time series for each cluster are proposed for the summer season, when ozone concentrations are the highest. Models can be generated by applying several procedures, which are classified as deterministic, stochastic and empirical methods (Dueñas et al., 2005). Special attention has focused on ozone modeling in different geographical areas such as North America (Yi and Prybutok, 1996; Davis and Speckman, 1999; Prybutok et al., 2000), Europe (Dueñas et al., 2002; Lengyel et al., 2004; Schlink et al., 2006) and Asia (Abdul-Wahab et al., 2005). In this study, multiple linear regression was used as an estimation model, owing to the easiness of establishing relationships between ozone and variables involved in its behavior through an explicit equation. Physical and chemical terms are also considered it is worth due to the fact that the equation is based on chemical and meteorological variables (Barrero et al., 2006). Dominick et al. (2012) classified air quality patterns in Malaysia by jointly considering several chemical and meteorological parameters, and Pires et al. (2008) established a classification of the monitoring network of Oporto metropolitan area. While in the first one they were conducted in order to determine the main air pollutant which contributing to Air Pollution Index (API), in the second one no ozone estimations were developed after classification.

To the best of our knowledge, no ozone estimating works have been carried out considering the monitoring network as a whole and taking into account the heterogeneity of ozone behavior within the network, i.e., avoiding an aggregation bias. Thus, the main contribution of this study is to obtain estimation models of hourly ozone concentrations in previously established subareas by applying a cluster analysis. It is worth to mention that this work allows for the estimation of ozone values by zones, so it would not be affected during downtime periods in measuring stations. The goals of this research are (i) to classify the monitoring stations of the Seville metropolitan area in keeping with similar ozone patterns, (ii) to establish a seasonal comparison among clusters, and (iii) to develop a model that is capable of estimating hourly ozone concentrations quite accurately. This study aims to contribute to a deeper understanding of the temporal and spatial assessment of air quality pollution in densely populated areas, taking the advantage of the fact that the methodology can be easily extrapolated to other urban areas.

2. Methods

2.1. Site description and data collection

Seville is located at 37°23′N; 5°58′W and 20 masl, and its metropolitan area has a population of nearly 1 500 000 inhabitants, it being the largest in the southern region of Spain (see Fig. 1a and b). This area exhibits the typical Mediterranean climate with continental features, which combines cold temperatures in winter and rather hot temperatures in summer. Minimum and maximum average temperatures range from 6.3 °C in winter to 33.8 °C in summer, respectively. Annual rainfall is very irregular and is mainly concentrated in spring and autumn. In addition, Seville metropolitan area has nearly 2900 h of sun a year, most of them in the summer and spring seasons.

The Environmental Department of the Regional Government of Andalusia (southern region of Spain) manages an air quality monitoring network consisting of fifty ozone monitoring stations. Eight of these ozone monitoring stations are established in the Seville metropolitan area, and also register several pollutants such as CO, SO_2, PM_{10} and NO_2 (measured in μg m^{-3}). Whereas five stations are located in the city, the surrounding towns are monitored by three stations. Fig. 1c depicts the ozone monitoring network in this metropolitan area. Centro station (CE) is located in the Historic Quarter in Seville and surrounded by pedestrian areas and little traffic, thus air pollution levels corresponds to the typical background pollution of the city. Torneo station (TO) is located in the west inner ring road and mainly focuses on road traffic emissions. The monitoring station of Santa Clara (ST), which is mainly focused on ozone control, is located in the east of the city. Two additional monitoring stations are found in the city—on the one hand, San Jerónimo (JE) station, located in an industrial area in the north of the city; and on the other hand, Los Bermejales (BE), established in a residential area in the south. The stations in the outskirts are found in the municipalities of Mairena del Aljarafe (MA), Dos Hermanas (DH) and Alcalá de Guadaíra (AL).

10-min O_3 concentrations were collected in these 8 monitoring stations from January 1, 2006 to December 31, 2011. Table 1 and Table 2 summarize measurement sites and descriptive statistics, respectively. Chemical variables were also collected during this period in each monitoring station. It should be noted that $[SO_2]$ is not measured in ST and JE, $[PM_{10}]$ in CE, JE and DH, and [CO] in JE and MA. On the other hand, meteorology data were provided by the National Agency of Meteorology (AEMET). These data were comprised of measurements of the main meteorological variables involved in ozone behavior, such as temperature (°C), solar radiation (×10 kJ m^{-2}), wind speed (m s^{-2}), atmospheric pressure (hPa) and air humidity (%). Data temporal resolution was 1-h, except for atmospheric pressure, which is measured at 0, 7, 13 and 18 UTC. These measurements were collected at the nearest meteorological station to Seville, which is located at the airport at 37°25′N; 5°52′W (see Fig 1c). However, due to the fact that no Solar Radiation measurements are provided at this station, these data were collected from Jerez de la Frontera meteorological station (36°42′N; 6°7′W and 57 masl), a city about 80 km from Seville, this being the nearest AEMET station from Seville.