Science of the Total Environment 544 (2016) 1–9

Contents lists available at ScienceDirect

Science of the Total Environment

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

New classification scheme for ozone monitoring stations based on frequency distribution of hourly data

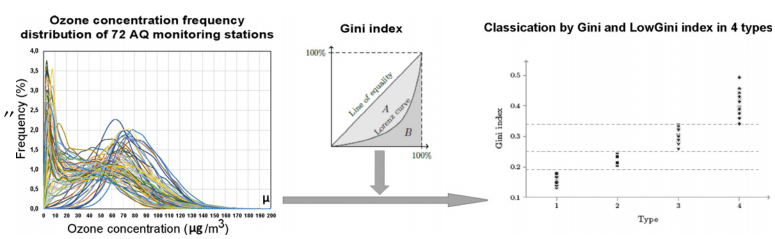
O. Tapia [a], M. Escudero [a,*], Á. Lozano [a], J. Anzano [b], E. Mantilla [c]

[a] Centro Universitario de la Defensa (CUD) de Zaragoza, Academia General Militar, Ctra. de Huesca s/n, E-50090 Zaragoza, Spain
[b] Laser Laboratory and Environment, Department of Analytical Chemistry, Faculty of Sciences, University of Zaragoza, Pedro Cerbuna 12, E-50009 Zaragoza, Spain
[c] Instituto Universitario Centro de Estudios Ambientales del Mediterráneo (CEAM-UMH), Parque Tecnológico C/ Charles R. Darwin 14, E-46980 Paterna, Spain

HIGHLIGHTS

- A new classification scheme for O_3 stations was developed.
- Based on frequency distribution curves four O_3 station types were identified.
- Proposed classification is suitable for assessing impact of reduction policies on O_3.
- The operational methodology to determine the station type is the Gini index.
- The robustness of the classification methodology was confirmed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 August 2015
Received in revised form 16 November 2015
Accepted 16 November 2015
Available online xxxx

Editor: D. Barcelo

Keywords:
O_3
Gini index
Classification
Monitoring stations
Frequency distribution

ABSTRACT

According to European Union (EU) legislation, ozone (O_3) monitoring sites can be classified regarding their location (rural background, rural, suburban, urban) or based on the presence of emission sources (background, traffic, industrial). There have been attempts to improve these classifications aiming to reduce their ambiguity and subjectivity, but although scientifically sound, they lack the simplicity needed for operational purposes. We present a simple methodology for classifying O_3 stations based on the characteristics of frequency distribution curves which are indicative of the actual impact of combustion sources emitting *NO* that consumes O_3 via titration. Four classes are identified using 1998–2012 hourly data from 72 stations widely distributed in mainland Spain and the Balearic Islands. Types 1 and 2 present unimodal bell-shaped distribution with very low amount of data near zero reflecting a limited influence of combustion sources while Type 4 has a primary mode close to zero, showing the impact of combustion sources, and a minor mode for higher concentrations. Type 3 stations present bimodal distributions with the main mode in the higher levels. We propose a quantitative metric based on the Gini index with the objective of reproducing this classification and finding empirical ranges potentially useful for future classifications. The analysis of the correspondence with the EUROAIRNET classes for the 72 stations reveals that the proposed scheme is only dependent on the impact of combustion sources and not on climatic or orographic aspects. It is demonstrated that this classification is robust since in 87% of the occasions the classification obtained for individual years coincide with the global classification obtained for the 1998–2012 period. Finally, case studies showing the applicability of the new classification scheme for assessing the impact on O3 of a station relocation and performing a critical evaluation of an air quality monitoring network are also presented.

* Corresponding author.
E-mail address: mescu@unizar.es (M. Escudero).

http://dx.doi.org/10.1016/j.scitotenv.2015.11.081
0048-9697/

1. Introduction

An adequate classification is a key instrument for interpretation and assessment of air quality data (especially for large data-sets covering large areas with a wide variety of types of locations); and for designing and evaluating mitigation measures and developing policy. Air quality managers need simple and effective tools for appropriate classification of monitoring sites. The standard European classification scheme for ozone (O_3) stations exclusively takes into account aspects related to the location and the presence of sources in the stations surroundings, and thus, do not use real data for the definition of the station classification. Other classification schemes that have been proposed are data-based although these generally demand complex analyses that may hinder their applicability for managing purposes. The classification scheme should respond to the actual O_3 behavior and, in particular, must be associated with the actual influence of the emissions of species which act as precursors or consumers of O_3 In this work a classification scheme with these characteristics is presented and tested.

Tropospheric ozone is formed in the atmosphere from reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the presence of sunlight (Atkinson, 2000). Another source is stratospheric intrusions (Holton et al., 1995; Cristofanelli et al., 2006). Ozone acts as an important oxidant agent generating negative effects on health such as increased morbidity and mortality, alterations in the respiratory tract and other chronic effects (WHO, 2008). Harmful effects on vegetation and ecosystems have also been demonstrated in the presence of excessive O_3 concentrations such as visible leaf symptoms, defoliation, senescence, and reduction in crop productivity (Paoletti, 2006; Sicard et al., 2011). Ozone can also cause corrosion on natural and synthetic rubber, and damage on plastic materials, surface coatings, and building materials (de Leeuw, 2000; Screpanti and Marco, 2009). Finally, O_3 alters the Earths energy balance acting as greenhouse gas (IPCC, 2013). The study of tropospheric O_3 concentrations in Europe is motivated by the elevated concentrations registered in wide areas of the continent, especially in the Mediterranean basin which exhibits the highest concentrations during spring and summer (Cristofanelli and Bonasoni, 2009; Sicard et al., 2013). These high concentrations result in exceedances of the concentrations designated as safe for health and vegetation protection established in European Union (EU) legislation (EEA, 2013). The designated European Commission Air Quality Standard for O_3 is 120 $\mu g/m^3$ maximum daily 8 h mean permitted exeedences being 25 days averaged over a 3 year.

Monitoring and evaluation of air quality standards are based on measurements performed at fixed automated stations that, according to EU Directive on ambient air quality and cleaner air for Europe 2008/50/EC, are classified according to location and type of sources existing in the vicinity of the station (EUROAIRNET classification). O_3 stations are designated as Urban, Suburban, Rural, and Rural Background depending on their location, for the remaining pollutants (particulate matter, nitrogen oxides, sulfur dioxide, benzene, metals and benzo[a]pyrene) only three categories are considered (Urban, Suburban and Rural). In parallel, stations are also classified as Background, Industrial, or Traffic depending on the sources that most likely affect the station. However, in practice, the classification of a monitoring site is generally made subjectively without prior knowledge of the pollutants behavior in that area. In addition, the classification of sites resulting from the combination of meta-data only reflect aspects related to location and the presence of sources, but does not include information on relevant factors such dispersion, transport, and physico-chemical interactions. The importance of these processes for understanding atmospheric O_3 combined with its complex reactivity, makes it even more difficult to identify comparable classes of monitoring sites for this pollutant with the current classifications.

One of the main motivations for establishing an improved classification scheme based on O_3 and other pollutant behavior is the need for evaluating the considerable number of Air Quality plans that have been implemented across Europe in the last decade. If the classification of a monitoring station is not directly associated with the phenomenology of the pollutants, and, more specifically with the actual impact of the sources, it can be difficult to evaluate the real effectiveness of mitigation measures. It is also of great importance to compare stations of the same type between different countries across Europe to evaluate if a similar action can have a different impact on pollutants levels depending on factors such as climate, urban structure, and orography. In the case of O_3, it is particularly relevant, since previous work has exposed the risk of increments in urban concentrations in the context of reductions in NO_x emissions (Sicard et al., 2013; Escudero et al., 2014; Querol et al., 2014). In summary, it is important to develop classifications for O_3 stations that respond to atmospheric dynamics of photochemical processes in a way that sites with the same typology are subject to similar mechanisms and, therefore, would require actions of the same nature from the viewpoint of air quality management.

Different classification schemes for O_3 monitoring stations have been proposed in previous work. Esser (1993) presented a classification based on a priori information for the pollution and local meteorological conditions at the observational site, and hence, has some degree of subjectivity. Larssen et al. (1999) developed a classification for air quality monitoring stations based on a distance to emission sources, in a similar way to the directive one. Flemming et al. (2005) used hierarchical clustering where the regimes were characterized by a pronounced weekly cycle. Definition of the clusters was done according to their previous O_3 classification of Enke et al. (1997). Tarasova et al. (2007) developed a classification of O_3 rural background stations employing hierarchical agglomeration clustering on seasonal-diurnal variations. Similarly to the last study, Kovac-Andric et al. (2010) proposed a classification for remote monitoring sites (EMEP) based on two indexes which take into account the diurnal behavior of ozone and the amount of high values in a particular site, however the authors only tested their methodology for 12 stations. A last relevant contribution was the work of Joly and Peuch (2012) in which a methodology to classify air quality stations over Europe was proposed based on statistical filtering of time series in order to extract some essential features such as the diurnal cycle, weekend effect, high frequency variability (periods lower than 3 days), and taking also into account the summer/winter differences. All these works offered valuable contributions towards a more reliable classification of monitoring stations measuring O_3 and other pollutants. Common aspects can be identified in most of these references such as the premise of classifying on the basis of the data recorded in the stations and not using a priori considerations. Although any classification scheme is a subjective choice adopted for an intended use of the classification, a data-based approach is more flexible than the standard approach because numerical classification rules can be used. Moreover, the classifications are pollutant-specific recognizing that a common type for different parameters may be ambiguous. The most relevant limitation of these studies is the use of complex procedures such as clustering techniques and other elaborated statistical analyses which can be effective from a scientific perspective, but are in general too complicated for operational implementation for management purposes. Moreover, a requirement for a classification scheme to be operative is to have a sufficiently reduced (manageable) number of groups, well differentiated, and clear and simple criteria to assign each station to one of the categories.

Here we present simple and powerful method based on the distribution of the density function of ozone concentrations that is closely linked to O_3 dynamics. After using O_3 data from 1998 to 2012 collected at 72 monitoring stations in Spain to define 4 different types of sites based on the frequency distribution of the concentrations, a quantitative metric was derived using the Gini index that is a statistical parameter indicating the inequality of the distribution of a variable in its range. Correspondence between the proposed classification and the standard one was analyzed and, finally, the robustness of the classification method was tested.