Available online at www.sciencedirect.com

ScienceDirect

Atmospheric Environment 41 (2007) 5831–5847

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

The use of conditional probability functions and potential source contribution functions to identify source regions and advection pathways of hydrocarbon emissions in Houston, Texas

Yulong Xie*, Carl M. Berkowitz

Natural Resource Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, Richland, WA 99352, USA

Received 18 December 2006; received in revised form 19 February 2007; accepted 22 March 2007

Abstract

In this study, we demonstrate the utility of conditional probability functions (CPFs), potential source contribution functions (PSCFs), and hierarchical clustering analysis (HAC) to identify the source region and transport pathways of hydrocarbons measured at five photochemical assessment monitoring stations (PAMS) near the Houston Ship Channel from June to October 2003. In contrast to scatter plots, which only show the pair-wise correlation of species, commonality in CPF figures shows both correlation and information on the source region of the species in question. In this study, we use over 50 hourly volatile organic compound (VOC) concentrations and surface wind observations to show that VOCs with similar CPF patterns likely have common transport pathways. This was established with the multivariate technique, which uses the hierarchical clustering analysis to define clusters of VOCs having similar CPF patterns. This method revealed that alkenes, and in particular those with geometric isomers such as *cis-/trans*-2-butene and *cis-/trans*-2-pentene, have similar CPF patterns and hence, a common area of origin. The alkane isomers often show CPF patterns among themselves, and similarly, aromatic compounds often show similar patterns. We also show how calculated trajectory information can be used in the PSCF analysis to produce a graphic picture that identifies specific geographic areas associated with a given VOC (or other pollutant). The use of these techniques in the chemically and meteorologically complex environment of Houston, Texas, suggests its further utility in other areas with relatively simpler conditions.

Keywords: Conditional probability function (CPF); Potential source contribution function (PSCF); Hierarchical clustering analysis (HAC); Houston; Volatile organic compound (VOC); Source receptor relationship

1. Introduction

Ozone is not a primary pollutant. It forms from a complex cycle involving emissions of nitric oxide and nitrogen dioxides ($NO+NO_2=NO_x$), volatile organic compounds (VOCs), sunlight and moisture. Much of the NO_x comes from motor vehicles and power generating stations. The sources for VOCs are much more diverse, and include many byproducts in the production and refinement of petroleum and other chemicals and biogenic emission from trees and plants. As a result, identifying the geographic area from which VOCs enter the atmosphere can be complex, yet this is the very

*Corresponding author. Tel.: +1 509 376 0062; fax: +1 509 376 5368.

E-mail address: yulong.xie@pnl.gov (Y. Xie).

1352-2310/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.atmosenv.2007.03.049

information needed to develop effective control strategies that will reduce ozone levels to concentrations considered safe for general public health (Sillman, 1999; TNRCC, 2001; Haagen-Smit and Fox, 1954).

Houston, Texas, and the surrounding area have unique air pollution problems as a direct result of the many emissions from petrochemical sources and meteorological conditions associated with its coastal location at low latitude that results in high temperatures, high humidity, and generally weak winds. As part of a multi-agency effort to identify specific processes and compounds associated with elevated levels of ozone, hourly measurements of more than 50 VOCs have been made at Photochemical Assessment Monitoring Stations (PAMS) throughout Houston, with particularly heavy monitoring being carried out in the immediate proximity of the Houston Ship Channel, home to one of the largest chemical manufacturing complexes in the world.

The present paper describes techniques to help identifying relatively well-defined geographic source areas associated with high concentrations of VOCs measured at the PAMS sites along the Ship Channel. The techniques to be described are rather general and have, we believe, the capability to help in similar analysis done at other areas with less complex emission patterns and meteorology. Conditional probability functions (CPFs) and potential source contribution functions (PSCFs) are the basis for this work. The CPF and PSCF techniques are widely used to identify transport paths and source areas by coupling observed chemical concentrations with meteorological information.

The CPF techniques have been successfully used in numerous source apportionment studies (Begum et al., 2004; Kim et al., 2003a, b, 2004; Xie and Berkowitz, 2006; Zhao et al., 2004; Zhou et al., 2004). However, most of the applications have been based on the factors from multivariate receptor modeling analysis, such as positive matrix factorization (PMF), instead on the individual species basis. In contrast, and what we believe is a novel application, we apply cluster analysis to CPF results, thus combining observed concentrations with trajectories, and objectively linking compounds having a common source area.

The motivation for developing these techniques stemmed from the fact that the large number of VOC species included in our analysis (55) made it difficult, if not impossible, to group species with a common direction of origin by visual inspections of the CPF plots alone. And this in turn led us to consider the use of hierarchical clustering analysis (HCA) to identify the groups of VOCs (Hartigan, 1975).

Cluster analysis objectively separates objects into groups whose identities are not known in advance. It represents a widely used multivariate technique in atmospheric science (Wilks, 2006). The uses of cluster analysis range from grouping weather observations into synoptic types for interpreting atmospheric chemical measurements (Brankov et al., 1998; Cape et al., 2000; Dorling and Davies, 1995; Harris and Kahl, 1990) to classifying a set of observations of either concentrations or size distributions into mutually exclusive groups (Dillner et al., 2005; Shaheen et al., 2005). For the work described here, the cluster analysis was based on the CPF patterns associated with the VOC species so that the grouped VOCs had common wind directions when all species within the resulting group had elevated concentrations of the species.

While local wind direction is used for the CPF analysis, the PSCF uses the entire wind trajectory and concentration information to identify possible transport pathways and source areas (Hopke, 2003). The PSCF was developed from residence time analysis (RTA) (Ashbaugh et al., 1985) and has been widely used to identify source locations and preferred transport pathways of atmospheric trace elements and particular species, e.g., sulfate, nitrate, ozone, black carbon, and mercury (Cheng et al., 1993a, b; Lin et al., 2001; Liu et al., 2003; Poirot and Wishinski, 1986; Polissar et al., 1999, 2001a; Stohl and Kromp-Kolb, 1994; Zeng and Hopke, 1989). The method has also been applied to locate the sources or source categories identified by multivariate receptor models, e.g., principal component analysis (PCA) and PMF (Poirot et al., 2001; Polissar et al., 2001b; Xie et al., 1999). While the PSCF was used primarily to identify remote sources, it has also been used to find local sources (Hsu et al., 2003). In this study, a PSCF was derived for each VOC species based on the observations from each the five sites and combining information from all five sites. The five PAMS sites that provided information on the 55 VOCs of this analysis were located at Channelview, Clinton, Haden Road, Lynchburg Ferry, and Wallisville, respectively (Fig. 1; Table 1). We used observations taken from June 1 to October 31, 2003. Additional