Available online at www.sciencedirect.com

SCIENCE DIRECT®

Atmospheric Environment 40 (2006) 3070–3091

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

The use of positive matrix factorization with conditional probability functions in air quality studies: An application to hydrocarbon emissions in Houston, Texas

Yulong Xie, Carl M. Berkowitz*

Pacific Northwest National Laboratory, Richland, WA 99352, USA

Received 30 September 2005; received in revised form 13 December 2005; accepted 21 December 2005

Abstract

In this paper, we describe two advanced statistical techniques suited to address the following questions: which source categories of emissions affect given areas and where do these source categories come from? A source category is defined as a combination of volatile organic compounds (VOCs) associated with a specific industrial process. A discussion of the positive matrix factorization (PMF) multivariate receptor model is presented, and this PMF technique applied to hourly average concentrations of VOCs measured at five Photochemical Assessment Monitoring Stations (PAMS) located near the emission-rich Houston Ship Channel region in Texas. The observations were made between June and October 2003, and the PMF analysis was limited to nighttime measurements (21:00–06:00 CDT) to remove the complexity of photochemical processing and associated changes in the concentrations of primary and secondary VOCs. Six to eight VOCs source categories were identified for the five Ship Channel sites. Specific geographic areas associated with each source category were identified through the use of conditional probability functions that identify source regions when superimposed on maps of VOC emissions.

Keywords: Receptor modeling; VOC; Positive matrix factorization (PMF); Conditional probability function (CPF); Houston; Air quality

1. Introduction

Practical questions to be addressed when implementing any emission control strategy are which emissions are affecting which areas and where do these emissions come from? The answers to such questions require knowledge of the ambient chemistry and local meteorology, with the latter knowledge needed to describe the geographic path followed by air parcels carrying the emissions of concern. In this paper, we describe techniques suited for obtaining such knowledge and present results using an extensive set of hydrocarbon measurements and surface meteorological observations from Houston, Texas.

The compounds of interest for our study are the 55 volatile organic compounds (VOCs) measured along and downwind of the Houston Ship Channel, which is a major waterway of Galveston Bay and home to one of the largest petrochemical processing facilities in the world. These VOCs, in conjunction with NO_x emissions, have long been known to mix in the presence of sunlight to form ozone. Ozone

---

*Corresponding author. Tel.: +1 509 372 6183; fax: +1 509 372 6168.

E-mail address: carl.berkowitz@pnl.gov (C.M. Berkowitz).

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2005.12.065

concentrations in this area are among the highest in North America not only because of the ample ozone-precursors but also because of the abundance of moisture and a coastal circulation pattern that allows an air mass to make multiple passes over key VOC source regions.

The basis for our analysis is a sophisticated analytical technique called positive matrix factorization (PMF) (Paatero, 1997; Hopke, 2000, 2003). This technique was used to derive a set of source composition profiles, each identifying a mix of compounds associated with a particular category of emissions (e.g., on-road emissions, solvent manufacturing, etc.). It also gave a measure of the relative importance of each profile at each of the Photochemical Assessment Monitoring Stations (PAMS) sites shown in Fig. 1. The major VOC point sources from a recent Texas Commission on Environmental Quality (TCEQ) emissions inventory also are shown in Fig. 1.

We begin this paper by reviewing the theory of PMF and defining the Conditional Probability Functions (CPF) used to identify the wind directions with which each source category, or 'factor,' was associated. Next, a short statistical overview is given of all the observations. We then present results of the PMF analysis, limiting the analysis to measurements made at night so as to remove the uncertainty in the interpretation associated with daytime photochemistry. Five major factors are identified, and their roles in the chemical mix at each of the measurement sites are discussed. We concluded that if these techniques work in an area as complex as Houston, they are probably suitable for use in other areas.

2. A brief review and summary of PMF and CPF

While the basic ideas of PMF are straightforward, their implementation can be quite detailed. For this reason, we present a short review of how the PMF calculations are carried out and how the results can be interpreted.

2.1. Positive matrix factorization

The PMF method was comprehensively described by Paatero and Tapper (1994) and Paatero (1997) and has been used in many source identification studies involving particulate matter (Hopke, 2003; Polissar et al., 1998; Xie et al., 1999; Willis, 2000) and VOCs (Miller et al., 2002; Roberts et al., 2004; Zhao et al., 2004). The implementation of PMF in the PMF2 program (Paatero, 2000) was used in this study. Additional guidance on the use of PMF for receptor modeling also was published by Hopke (2000). Here, we present a brief overview of the technique, noting modifications needed for its application to the PAMS data set.

The two-way formulation of PMF solves the general receptor model, which is formulized as a chemical mass balance problem in terms of contributions from p independent sources to all chemical species measured in a given sample (Miller et al., 1972; Hopke, 1985, 1991):

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}, \tag{1}$$

where x_{ij} is the jth species concentration measured in the ith sample, g_{ik} is the emitted mass concentration

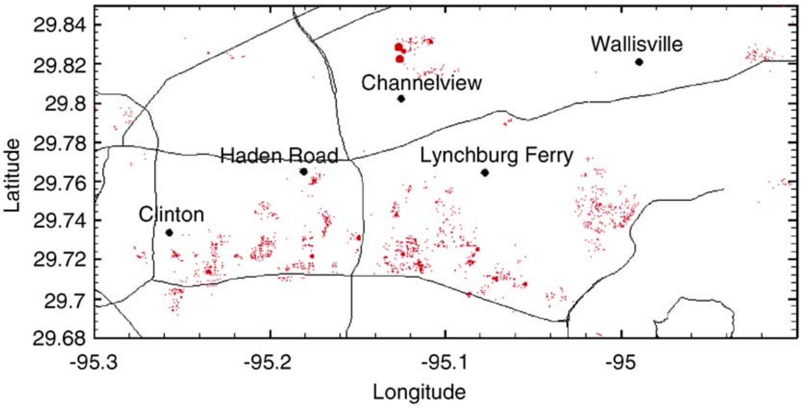

Fig. 1. Map of the Houston area showing the locations of the PAMS auto-gas chromatographs that provided the observations used in this study. Also shown are major point sources of VOC emissions.