Atmospheric Research 178–179 (2016) 415–425

Contents lists available at ScienceDirect

Atmospheric Research

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

Potential emission flux to aerosol pollutants over Bengal Gangetic plain through combined trajectory clustering and aerosol source fields analysis

D. Bharath Kumar, S. Verma *

Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 14 March 2016
Accepted 13 April 2016
Available online 22 April 2016

Keywords:
Fuzzy c-mean clustering
Trajectory cluster analysis
Concentrated weighted trajectory
Aerosol fields analysis
Aerosol optical depth

ABSTRACT

A hybrid source-receptor analysis was carried out to evaluate the potential emission flux to winter monsoon (WinMon) aerosols over Bengal Gangetic plain urban (Kolkata, Kol) and semi-urban atmospheres (Kharagpur, Kgp). This was done through application of fuzzy c-mean clustering to back-trajectory data combined with emission flux and residence time weighted aerosols analysis. WinMon mean aerosol optical depth (AOD) and angstrom exponent (AE) at Kol (AOD: 0.77; AE: 1.17) were respectively slightly higher than and nearly equal to that at Kgp (AOD: 0.71; AE: 1.18). Out of six source region clusters over Indian subcontinent and two over Indian oceanic region, the cluster mean AOD was the highest when associated with the mean path of air mass originating from the Bay of Bengal and the Arabian sea clusters at Kol and that from the Indo-Gangetic plain (IGP) cluster at Kgp. Spatial distribution of weighted AOD fields showed the highest potential source of aerosols over the IGP, primarily over upper IGP (e.g. Punjab, Haryana), lower IGP (e.g. Uttarpradesh) and eastern region (e.g. west Bengal, Bihar, northeast India) clusters. The emission flux contribution potential (EFCP) of fossil fuel (FF) emissions at surface (SL) of Kol/Kgp, elevated layer (EL) of Kol, and of biomass burning (BB) emissions at SL of Kol were primarily from upper, lower, upper/lower IGP clusters respectively. The EFCP of FF/BB emissions at Kgp-EL/SL, and that of BB at EL of Kol/Kgp were mainly from eastern region and Africa (AFR) clusters respectively. Though the AFR cluster was constituted of significantly high emission flux source potential of dust emissions, the EFCP of dust from northwest India (NWI) was comparable to that from AFR at Kol SL/EL.

1. Introduction

Studies on evaluation of source origin of aerosols are important to examine the role of pollutant emissions on climate and human health. These studies are helpful towards developing effective air quality management strategies. In order to understand the effect of pollutant sources on pollutant concentration at a receptor location, knowledge of pathways of air parcel between source and receptor location is required. In addition, it is also required to know the residence time of air parcel in a given source region during its movement from source to the receptor. Trajectory modelling has widely been used to estimate the transport pathway of an air parcel on regional- to continental- and global-scales (Kahl et al., 1989; Methven et al., 2001). Classification of large number of trajectories according to transport sector has been applied in previous studies to identify source areas (Simmonds et al., 1997; Solberg et al., 1996). However, it was difficult to identify individual source areas in these studies (Stohl, 1996). Cluster analysis algorithm (such as average-link clustering, K-means clustering), a multivariate statistical tool is being used to overcome this difficulty (Bratchell, 1989; Dorling et al., 1992; Brankov et al., 1998; Moy et al., 1994; Stohl, 1998; Cape et al., 2000; Methven et al., 2001; Chan et al., 2002). Cluster analysis has also been carried out recently using new artificial intelligence techniques such as artificial neural networks and self organization map (Owega et al., 2006; Kassomenos et al., 2010; Mingoti and Lima, 2006). Cluster analysis classifies the trajectory dataset into number of groups which differ from each by a specified distance, such as the Euclidean distance between trajectories (Stohl, 1996; Markou and Kassomenos, 2010; Kong et al., 2013). Though cluster analysis is helpful to segregate various source areas influencing receptor, it is solely not sufficient to identify the potential source areas delineated by distinct emission flux of aerosol constituents contributing to pollutants concentration at receptor.

In order to identify the probable location of emission sources contributing to aerosol pollutants at the receptor, statistical methods combining the back trajectory information with measurement data at receptor location have widely been used in various studies (Salvador et al., 2008). This method is applied to obtain the potential source contribution function (PSCF) linking residence time of air parcel in upwind

* Corresponding author.
E-mail address: shubha@iitkgp.ac.in (S. Verma).

http://dx.doi.org/10.1016/j.atmosres.2016.04.012
0169-8095/

area with high concentrations through a conditional probability field (Ashbaugh et al., 1985; Zeng and Hopke, 1989; Wang et al., 2006; Jeong et al., 2011). Further, the concentration field method was developed which is based on carrying out concentration-weighted trajectory (CWT) or residence time weighted concentration (RTWC) calculations in each grid cell (Seibert et al., 1994; Stohl, 1996; Hsu et al., 2003).

In order to evaluate the relative influence of aerosols emitted from source categories and geographical regions in south Asia, aerosol transport simulations in a general circulation model (GCM) with emissions in the model tagged by region and sector have been carried out (Verma et al., 2007; Verma et al., 2008). However, carrying out such simulations on a fine grid resolution are computationally more complex and costly and hence lead to difficulty in locating fine grid resolved regional "hot spots" or evaluating potential emission sources and their gradient at a finer resolution on a regional scale.

In the present study, we carry out a combined source-receptor analysis to examine the potential emission flux of aerosol constituents to winter monsoon (WinMon) aerosols over Bengal Gangetic plain. This is done using a novel fuzzy-c mean cluster analysis of back-trajectory dataset combined with emission flux information and conditional probability field analysis using CWT. A comparative study of aerosol optical properties and potential emission flux to aerosols is carried out between two of the receptor locations over Bengal Gangetic plain in an urban and semi-urban atmospheres.

The specific objectives of the present study are to (i) compare semi-urban and urban atmosphere aerosol optical properties which is being used in estimation of aerosol source fields over Bengal Gangetic plain.

(ii) identify source region clusters corresponding to receptor locations in urban and semi-urban atmospheres through clustering of back-trajectory data and estimating the emission flux source potential (EFSP) of identified clusters, (iii) evaluate the conditional probability field of moderate to strong transport sectors through CWT calculations and estimate the emission flux contribution potential (EFCP) from source region clusters to aerosols at receptor locations. To facilitate reading, we have also provided a list of acronyms as an appendix to this manuscript.

2. Method of study

2.1. AOD measurements

Measurements of AOD were carried out at rooftop of building (about 10 to 12 m above the ground level) during the period November 2009 to February 2010; these measurements in a semi-urban atmosphere (Kharagpur (Kgp), 22.19 °N, 87.19 °E) were done in the main campus of Indian Institute of Technology (IIT) Kharagpur and that in an urban atmosphere (Kolkata (Kol), 22.57 °N, 88.42 °E) at an extension centre of IIT Kharagpur campus located on the outskirts of Kolkata city. Location of the study sites (Kolkata and Kharagpur) over Bengal Gangetic plain is shown in Fig. 1a.

Aerosol optical properties measured in a semi-urban atmosphere in the present study are compared with that in an urban atmosphere (Kol) in eastern India over Bengal Gangetic plain. Information of aerosol optical properties from measurements at Kol is obtained from a previous study (Verma et al., 2014).

Hourly measurements of AOD were carried out using hand-held Microtops-II sun-photometer (Solar Light Company, USA) at five wavelength channels (340, 440, 500, 870, and 1020 nm) from 0800 to 1600 local time, weather permitting, during the period from November to February corresponding to winter monsoon (WinMon) season.

Cloud screened data (Verma et al., 2014) including 182 samples of AOD at Kharagpur and 133 samples at Kolkata were used in analyses.

The instrument is factory-calibrated, maintaining the interval between calibration seasons to be of order of one year (www.solarlight.com/products/sunphoto.html).

AOD values obtained at 0.44 and 0.87 μm wavelength channels are used to compute Angstrom Exponent (AE) using the following formula (Schuster et al., 2006):

$$AE = -\frac{log\left[\frac{\tau(\lambda1)}{\tau(\lambda2)}\right]}{log\left[\frac{\lambda1}{\lambda2}\right]} \quad (1)$$

where, $\tau(\lambda1)$ and $\tau(\lambda2)$ are AOD value at wavelengths of λ1 and λ2, respectively. AE is an indicative of spectral dependence of AOD and is useful to characterize the columnar aerosol size distribution, e.g. an increase in its value is an indicative of increase in the number of small-size particles relative to larger ones, and vice versa (Eck et al., 1999). Relative humidity (RH) and temperature were recorded using sensors operating at an accuracy of ±2% for RH and ±0.5 °C for temperature. Data of wind speed and direction at Kharagpur were obtained at an interval of 30 min from sensor (propeller type) installed at a meteorological tower located inside the IIT campus and operating at an accuracy of ±0.3 m s^{-1} and ±3 ° respectively. This data at Kolkata were obtained from the archives at weather underground (http://www.wunderground.com) corresponding to the period of measurement; the reported uncertainty in this data archives for wind speed and wind direction being ±0.5 m s^{-1} and ±5 °, respectively.

2.2. Source region clusters

Cluster of source regions is defined through grouping of seven-day air mass back trajectory origin or end points using fuzzy c-mean clustering (Falasconi et al., 2010; Mingoti and Lima, 2006) method.

Seven-day back trajectory calculations were done taking into account the residence time of aerosols of the order of a week in the lower atmosphere (Jaenicke, 1981). These calculations were carried out using NOAA HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) (Version 4) model (Draxler and Hess, 1998) corresponding to days of measurement during November 2009 to February 2010.

Back trajectories were calculated at 1200 h Indian Standard Time (+0530 h GMT) at heights of 10 m, 100 m, 500 m, 1000 m, and 5000 m above ground level. These calculations were done taking velocity and temperature fields from NCEP data archives (ftp://www.arl.noaa.gov/pub/archives/fnl). Mean path of air mass originating from identified clusters (discussed later) in the present study is shown in Fig. 1b to e.

Fuzzy c-mean clustering optimization method allows grouping on the basis of similarity or distance between the centroids and its neighbouring elements (endpoints of back trajectory in the present study). Alternating optimization scheme implemented in fuzzy c-mean clustering algorithm (Falasconi et al., 2010) is used to conduct the simulation of clustering of the back trajectory origin or endpoint data set. This scheme consists of updating the membership matrix as well as cluster centroids. Cluster centres or centroids of each cluster are obtained and probable membership (between 0 and 1) is given to every end point based on its Euclidean distance from the centroids of cluster (Dorling et al., 1992; Owega et al., 2006; Cape et al., 2000; Kassomenos et al., 2010; Karaca and Camci, 2010). Endpoint is then allocated to a cluster for which its membership is higher. A brief description on fuzzy c-mean algorithm implemented in the present study is given in a supplementary material provided with this manuscript.

The fuzzy c-mean clustering of back trajectory end points data is carried out separately at the surface layer (at heights of 10 m, 100 m, 500 m above ground level) and elevated layer (1000 m, 5000 m above ground level). This is because most of the origin points at surface were found over Indian region while that at elevated were over far-off locations of world regions. Initially (phase 1) the number of cluster centroids were suitably chosen as six which lead to obtain distribution of centroids broadly over regions of interest including Indian subcontinent,