Atmospheric Environment 186 (2018) 178–188

Contents lists available at ScienceDirect

Atmospheric Environment

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

Projection of near-future anthropogenic $PM_{2.5}$ over India using statistical approach

Abhishek Upadhyay[a,*], Sagnik Dey[a], Pramila Goyal[a,b], S.K. Dash[a]

[a] Centre for Atmospheric Sciences, IIT Delhi, Hauz Khas, New Delhi 110016, India
[b] The Northcap University, HUDA Sector 23 A, Gurgaon 122017, India

ARTICLE INFO

Keywords:
$PM_{2.5}$
Projection
Meteorology
Emission
MLR
ECLIPSE
India

ABSTRACT

Particulate matter smaller than 2.5 μm (referred to as $PM_{2.5}$) is the most important criteria pollutant impacting human health, environment and climate. India is already recognized as pollution hotspot where $PM_{2.5}$ has been increasing in the recent past. Here we project anthropogenic $PM_{2.5}$ for the near future (till 2040) over India using multiple linear regression (MLR) approach based on RegCM projected meteorology and ECLIPSE projected emission. MISR derived $PM_{2.5}$ concentration (μg/m^3) for the year 2010–2012 has been used to train the MLR model with reasonable accuracy ($R > 0.9$). The impact of the meteorological parameters under both RCP4.5 and 8.5 scenarios partially negates the impact of rising emission in future; more so in RCP8.5 than in RCP4.5 scenario. Air quality is projected to improve significantly with short lived climate pollutant (SLCP) 'mitigation' scenario in comparison with current legislation (CLE) 'baseline' emission scenario. Spatial analysis identifies a rapid increase in anthropogenic $PM_{2.5}$ in the eastern Indian states of Jharkhand, Chhattisgarh and Odisha, Peninsular India, and Delhi National Capital Region. Our results identify the near future pollution hotspots that would be useful in air quality management planning for the near future.

1. Introduction

Rapidly growing population, industrial activity, urbanization, vehicular emission and various other human activities like residential cooking, lighting and heating, construction activities etc. emit huge amount of pollutants to the atmosphere on a daily basis. As a result, air quality has degraded beyond its permissible limit in a large part of the world posing a serious threat to the biosphere. Particulate matter smaller than 2.5 μm in diameter (referred to as $PM_{2.5}$) is considered to be the best indicator of air quality and its health impacts (WHO, 2006). Each 10 μg/m^3 rise in $PM_{2.5}$ concentration has been found to be associated with a 4%, 6% and 8% increased risk of all cause, cardiopulmonary and lung cancer mortality respectively (Pope et al., 2002). Estimate of annual premature mortality burden from chronic ambient $PM_{2.5}$ exposure globally is quite large and alarming (Cohen et al., 2005, 2017 Anenberg et al., 2010; Lim et al., 2012). The severe impact on health from $PM_{2.5}$ exposure forced United States Environmental Protection Agency (USEPA) to adopt a new air quality standard. Strict regulations have resulted in a reduction in air pollution in the United States, which has experienced considerable health benefit (Correia et al., 2013).

In India, the annual and daily air quality guidelines for $PM_{2.5}$ are set at 40 and 60 μg/m^3 respectively (CPCB, 2009). However, the problem has been intensifying over the years and it becomes challenging to resolve it due to lack of adequate long-term in-situ $PM_{2.5}$ data, region-specific exposure-response function, poor understanding of the interplay of meteorology, emission and topography. Lack of robust in-situ data of $PM_{2.5}$ at the required spatial scale prompted the scientific community to use satellite data to infer $PM_{2.5}$ (van Donkelaar et al., 2010, 2016). Following this approach, Dey et al. (2012) converted satellite-retrieved columnar aerosol optical depth (AOD) to $PM_{2.5}$ using model derived conversion factor and created a $PM_{2.5}$ database for India. The data are further used to estimate district level premature mortality burden (Chowdhury and Dey, 2016). The spatio-temporal variability of $PM_{2.5}$ depends not only on emission of primary pollutants, but also on the meteorology and topography that modulate variability of gaseous precursors and the atmospheric processes leading to secondary $PM_{2.5}$ formation. $PM_{2.5}$ constitutes of many species and therefore the dependence of $PM_{2.5}$ on meteorological variable is more complex compared to other pollutants (Liao et al., 2006; Tai et al., 2012). Relation between meteorological parameters (e.g. wind speed, wind direction, rainfall, temperature and relative humidity) and PM concentration shows large spatial (Tai et al., 2010) and seasonal variations (Gupta et al., 2008). Air pollution and climate change are interlinked because the particulate

* Corresponding author.
E-mail addresses: abhiupadhyay777@gmail.com, asz138014@iitd.ac.in (A. Upadhyay).

https://doi.org/10.1016/j.atmosenv.2018.05.025
Received 17 November 2017; Received in revised form 31 March 2018; Accepted 12 May 2018
1352-2310/

matters have both climate and health impacts. Reduction of these short-lived climate pollutants would lead to co-benefits (Shindell et al., 2012). Therefore, it is important to understand how the distribution of $PM_{2.5}$ is modulated by the change in meteorology influenced by changing climate. Accuracy of the predicted $PM_{2.5}$ depending on the predicted changes in emission and meteorology in future relies on the representativeness of such correlation.

Analysis of past data can shed light into this. Silva et al. (2013) has shown that nearly 1500 and 2200 premature deaths per year can be attributed to climate change impacts on ozone and $PM_{2.5}$ in the past. They also noted large uncertainty in these estimates due to discrepancy in climate models. In their analysis, they revealed India as one region where the air quality has degraded rapidly. This is consistent with other studies showing a large increase in AOD over India in the last decade that can be attributed to rising anthropogenic sources (Dey and Di Girolamo, 2011; Krishna Moorthy et al., 2013) as opposed to a global decreasing trend (Mishchenko et al., 2007). These studies raise an important question. Will anthropogenic $PM_{2.5}$ concentration increase in the future under warming climate in India? Such information is of utmost importance to policymakers who can plan for adequate mitigation measures in advance. Using Mozart-4 chemical transport model (CTM), West et al. (2013) estimated that global air pollution related mortality is projected to increase in 2030 followed by a decrease till 2100 in reference as well as in RCP 4.5 scenario. But for south Asian region, it is estimated to increase up to 2050. Projections are showing more decrease in RCP 4.5 compared to reference scenario reflecting the beneficial impact of implementing climate policy. CMAQ coupled with downscaled GISS model output has shown a mean annual reduction of 23% (ranging from 9% to 32%) in $PM_{2.5}$ concentration over United States due to climate change (Tagaris et al., 2007). However, these projections have large spatial heterogeneity across the globe. Such estimate is lacking for India, the second most populous country in the world. Moreover, most climate models have large uncertainty in simulating the aerosol field in this region (Sanap et al., 2015).

The major objective of this work is to develop a statistical model to project anthropogenic $PM_{2.5}$ for the near future using the projected emission and meteorology for India. We consider projected meteorology under RCP (Representative Concentration Pathway) 4.5 and 8.5 scenarios, for $PM_{2.5}$ emission increasing as per business-as-usual (referred to here as baseline scenario) and decreasing due to technological intervention (mitigation scenario) (Klimont et al., 2017; Stohl et al., 2015). Statistical methods are computationally efficient than climate models (Mishra et al., 2015) and therefore, if the statistical model is able to represent the local conditions of the past and present reasonably well, then it can be used for future as well; provided that the dependency of $PM_{2.5}$ on meteorological variables and emission strengths remain unchanged in future. Our objective is to understand the variability of ambient $PM_{2.5}$ concentration in the near future under the 'mitigation' scenario relative to 'baseline' emission scenario, if the meteorological parameters change as per RCP4.5 and RCP8.5 scenarios. We show and discuss the projections for three representative years of the near future - 2020, 2030 and 2040 relative to the baseline 2010 in this study.

2. Methods

2.1. Multiple linear regression (MLR) analysis

MLR is a statistical process used to quantify the dependence of a variable on a number of independent variables. The dependent variables are fitted in a linear equation in order to predict independent variable (y). It can be represented as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 ... + \beta_n x_n + e \quad (1)$$

where, x_1, x_2 ......, x_n are the independent variables, β_1, β_2,......, β_n are regression coefficients, β_o is a constant and e is the estimated error term which is obtained from independent random sampling from the normal distribution with mean zero and constant variance. The task of regression modeling is to estimate the coefficients β_1, β_2,... β_n, which can be achieved through least square error technique. In this work, $PM_{2.5}$ concentration is the dependent variable that depends on meteorological parameters and anthropogenic $PM_{2.5}$ emission. The assumptions for a MLR include (1) selected variables are from reliable source and are well validated; (2) all variables are normally distributed; (3) there is a linear relationship between dependent and independent variables and (4) the variance of error is same across all level of independent variables (homoscedasticity). If a strong correlation exists between independent variables, it creates problem of multi-collinearity. This may induce errors in regression coefficients. Variance Inflation Factor (VIF) estimates severity of multi-collinearity in ordinary least square regression analysis and is used to address this issue in our work.

$$VIF = 1/(1 - R_i^2) \quad (2)$$

where R_i^2 denotes coefficient of determination when one predictor variable is regressed with other predictor variables for i = 1,2,3, …, n – 1 and n = number of predictor variable.

2.2. Meteorological data

Meteorology and emissions are the two most important factors that modulate the space-time variation of $PM_{2.5}$ concentration. For our work, we train the statistical model by meteorological outputs from a Regional Climate Model (RegCM) version 4.3, which include the meteorological parameters for present era and the future. The statistical model could have been trained by the meteorological data from India Meteorological Department (IMD) or available re-analysis data, but they are not available for future. Using dataset from RegCM for future in the statistical model trained with meteorological data from a different source will not allow us to understand and quantify the uncertainty in the estimates. Therefore, for the consistency, the model has been trained by meteorological data from RegCM.

RegCM was originally developed at National Centre for Atmospheric Research and now is being upgraded at International Centre for Theoretical Physics (ICTP), Italy. It is a hydrostatic and compressible model that runs on Arakawa B grid. The detail descriptions of this model for its current and earlier versions are given in Giorgi et al. (1993a, b, 2014) and Pal et al. (2007). The selected simulations are from a broad project - Coordinated Regional Climate Downscaling Experiment (CORDEX) RegCM4 Experiment Matrix abbreviated as CREMA (Giorgi, 2014). CORDEX is a modelling framework under auspices of world Climate Research Program with the aim of (i) access and improve regional climate downscaling techniques, (ii) produce 21st century regional climate projections over regions worldwide and (iii) bring up climate science and end-user communities and in particular the involvement of scientists from developing countries. Simulations were carried out from 1st January 1970 to 31st December 2100 over the South Asia CORDEX domain in two phases - phase I for the period 1970 to 2005 in the historical mode and phase II for the period 2006–2100 in projection mode. Initial and boundary conditions for simulations are generated by GFDL-ESM2M model, which is a general circulation model (GCM) developed at Geophysical Fluid Dynamics Laboratory (GFDL). Climate downscaling were performed for two IPCC emission scenarios - RCP 4.5 and 8.5. RegCM simulations were extensively validated in previous studies (e.g. Dash et al., 2014, Dash et al., 2015) against India Meteorological Department (IMD) and reanalysis data for its utility in projecting regional climate.

The output data from both the simulations have been used here to account variation in meteorological parameters within a probable range. Data over the Indian domain have been extracted from simulation outputs for the study period. It may be noted that only surface outputs have been chosen in this study. Surface output file includes 25