Impact of Urban Growth on Air Quality in Indian Cities Using Hierarchical Bayesian Approach

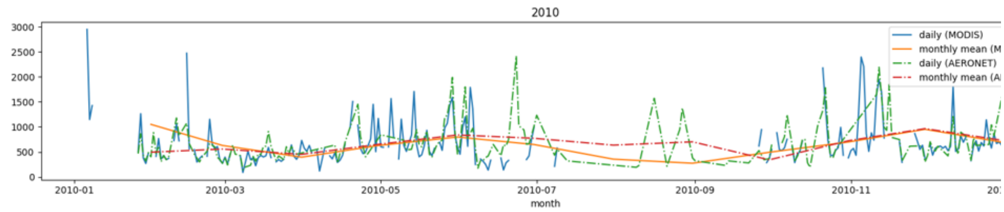
Prakhar Misra (1), Ryoichi Imasu (2), Wataru Takeuchi (1)
[1] Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[2] Atmosphere and Ocean Research Institute, The University of Tokyo, Tokyo, Japan
Email: mprakhar@iis.u-tokyo.ac.jp, imasu@aori.u-tokyo.ac.jp, wataru@iis.u-tokyo.ac.jp

KEYWORDS: - MODIS AOD, $PM_{2.5}$, LULC, urban pollution

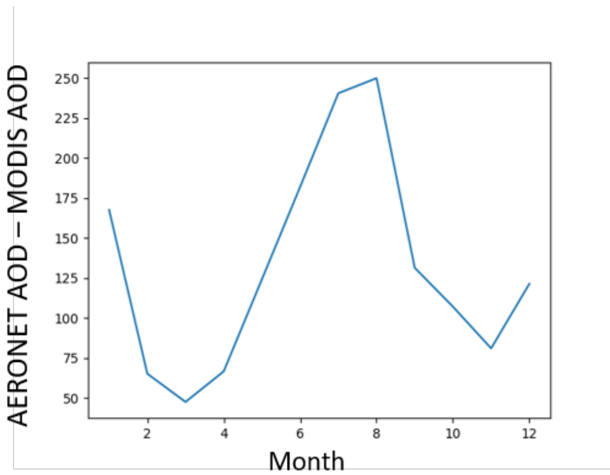
1 Supporting figures and tables

1.1 Missing AOD observation due to cloud

During Monsoon months of June, July and August, there are several missing daily MODIS AOD retrievals due to cloud coverage. A comparison of MODIS AOD with AOD from a ground based station, such as AERONET can help in estimating the expected AOD when the retrievals are missing. Figure S1 (a), shows the missing AOD retrievals. If the number of continuously missing AOD retrievals is small, rolling mean based interpolation can approximate mean monthly mean AOD. However if AOD retrievals are continuously missing for more than 15 days, the monthly mean value cannot be relied upon. Fortunately, high number of continuously missing AOD retrieval happens during the monsoon rain, during which wet scavenging leads to low AERONET AOD. This can be seen in Figure S1 (b). However as pointed by other researchers that final values can vary by as much as 30% and neither technique has been judged to be superior.



(a) Daily and mean AOD for 2010



(b) Difference in mean monthly AOD

Figure S1: MODIS retrievals are missing during months of June, July and August due to which monthly mean of MODIS AOD is much lower than AERONET AOD over the Kanpur city, India (a). In (b) the difference mean monthly AOD from AERONET and MODIS show difference is maximum during the Monsoon months.

Table S1: Assumed distribution of seasonal emission activity parameter (SEA_{mon}). For any source SEA_{mon} represents ratio of emissions in any month *mon* compared to its maximum monthly emission.

Source	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Residential (A_R)	1.0	0.9	0.9	0.85	0.85	0.85	0.75	0.75	0.75	0.8	1.0	1.0
Commercial (A_C)	1.0	0.9	0.9	0.85	0.85	0.85	0.75	0.75	0.75	0.8	1.0	1.0
Industrial (A_I)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Crop fire (A_{agro})	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Brick-kiln (A_{BK})	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0
Vehicle (A_V)	1.0	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	1.0

Table S2: Pearson correlation between coefficient obtained R_{est} and R_{obs} when the *SEA* was set according to Table S1.

City	correlation	p-value
Agra	0.58	0.0000
Ahmedabad	-0.02	0.0000
Allahabad	0.54	0.0000
Chennai	0.63	0.0000
Kanpur	0.58	0.0000
Lucknow	0.50	0.0000
Ludhiana	0.30	0.0000
Patna	0.76	0.0000
Raipur	0.42	0.0000
Hyderabad	0.44	0.0000
Jaipur	-0.27	0.0003
Bangalore	0.58	0.0000
Kolkata	0.64	0.0000
NewDelhi	0.61	0.0000
Mumbai	0.45	0.0000

1.2 Model sensitivity to the choice of *SEA* parameters

Owing to lack of seasonal emission activity data for each source, relative monthly emissions deduced from the REASv2 SO_2 inventory were also experimented as SEA_{mon} for residential, commercial, industries and vehicles. The SEA_{mon} values used for *LU* and *V* are shown in Table S1. It can be SEA_{mon} in REASv2 varies by smaller amount compared the SEA_{mon} used by us in the main text. The Pearson's correlation coefficient obtained between R_{est} and $_{obs}$ shows that REASv2 derived *SEA* are not better than the *SEA* obtained by us for the polluted cities in Indo-Gangetic plain. Although REASv2 based SEA_{mon} gives higher correlations for cities in the southern which have tropical climate all round year. The correlations are shown in Table S2. We conclude that s and r are similar to 1 for southern Indian cities where annual temperature range is smaller than northern Indian cities.

1.3 Model sensitivity to the missing *R* values in rainy months

To further assess whether the 15-day rolling day interpolation on missing values is better than no interpolation, we trained the model without any observation from the month of June, July and August. The correlation between R_{est} and R_{obs} in this case is shown in Table S3. The overall correlation decreased in 12 out 15 cities when rainy month observations were discarded, although their magnitude varied. This shows that some form of interpolation is better than no interpolation case. More advanced techniques like data assimilation, which consider both the modeling and observation have been used in literature before. Performing comparison of different techniques is desirable, specially to figure out the under what conditions simple statistical approach would suffice. However such comparison is out of scope in this study.

1.4 Model sensitivity to the choice of training period

We would like to recall that originally we prepared the urban land-use dataset over 2001 and 2011 due to data availability of the original ASTER and AW3D30 DSM in only those years. This implicitly assumes linear urban expansion between 2001 to 2011 and beyond. However the non-uniform annual urban land-use expansion could take place in the intervening years under influence of population growth and per capita GDP growth. This may bias *EC* estimates of base year (2001) and diminish its applicability for future predictions. To test this hypothesis, *EC* was also estimated on the basis of interpolated land-use area. Three additional models were prepared by training over pair of annual datasets, Model1: 2001 and 2005, Model2: 2005 and 2010, and Model3: 2010 and 2015. The comparison

Table S3: Pearson correlation between coefficient obtained R_{est} and R_{obs} to compare predictions from the model trained on interpolated R values for rainy months, and the model trained on non-interpolated and rainy months discarded R values.

City	rain month discarded	interpolated rain-months (same as main text)
Chennai	0.63	0.62
Mumbai	0.30	0.46
NewDelhi	0.63	0.61
Bangalore	0.36	0.43
Hyderabad	0.30	0.35
Kolkata	0.61	0.63
Agra	0.62	0.46
Ahmedabad	-0.17	0.22
Allahabad	0.58	0.66
Kanpur	0.52	0.61
Lucknow	0.50	0.52
Ludhiana	0.21	0.28
Patna	0.70	0.78
Raipur	0.45	0.46
Jaipur	-0.33	-0.11

Table S4: EA_O and EC_{LU} parameters depending on year used for training - Model1: 2001, 2005; Model2: 2005, 2010; Model3: 2010, 2015.

Tier	Model	EA_{O_m}	EC_{R_m}	EC_{C_m}	EC_{I_m}	EC_{V_m}
1	Model1 (2001,2005)	19.45	0.0005	0.0003	0.0005	6×10^{-9}
1	Model2 (2005,2010)	20.45	0.0004	0.0002	0.0006	2×10^{-8}
1	Model3 (2010,2015)	19.82	0.0005	0.0002	0.0002	6×10^{-9}
2	Model1 (2001,2005)	27.44	0.0009	0.0006	0.0004	7×10^{-8}
2	Model2 (2005,2010)	29.35	0.0007	0.0004	0.0006	1×10^{-7}
2	Model3 (2010,2015)	30.55	0.0009	0.0006	0.0004	1×10^{-7}

of Tier level EC_m estimates from the three 5-year interval models with the current model is shown in Table S4. EC of Tier-1 and Tier-2 in Model1 are same as current model since they both have 2001 itself as starting year. For Tier-1, EC_{I_m} decreased from Model1 to Model3. Sensitivity of EC to the training dataset could be because of the interpolated data itself that was used as input to A_{LU}. This sensitivity could also be due to decreasing unit industrial emission coefficient in and after year 2010. So although total area under industrial units may be rising, but their mean emission is decreasing. For example, decrease in emissions due to economic recession or a structural shift to less emitting industries. This could imply that mean emission per unit area of industries is lowered after 2010. For EC_{R_m} and EC_{C_m}, there is no apparent trend, signifying no change in their unit emission coefficients. The EC_V in both Tier-1 and Tier-2 cities has also increased. This is contrary to emission policies set for vehicles which are stricter in 2010 compared to 2001. Similarly contribution EA_{O_m} also increases from Model1 to Model3 for Tier-2 cities. This analysis needs to be performed more systematically with the availability of land-use datasets of the intervening years.