

Spatial-Temporal Autoregressive Dynamic Model[†]

Naresh Kumar¹ & Jacob Oleson²

¹ *Department of Geography, University of Iowa, Iowa City, IA & PSTC, Brown University, Providence, RI, USA*

² *Department of Biostatistics, University of Iowa, Iowa City, IA, USA*

Abstract: Although a myriad of methods have been advanced to tackle spatial and temporal structures in data separately, it becomes difficult to analyze these data using classical linear regression models when spatial-temporal structures coexist, especially when the data size is relatively large. In this article, we demonstrate a simple to implement method to handle spatial-temporal structures simultaneously. Building on the classical linear regression model, we propose to account for spatial-temporal autocorrelation in the residuals from the linear regression model iteratively until the main assumptions of the model are satisfied. Neighbors and their weights are defined using the strength of correlation at different spatial-temporal lags.

Using the proposed methodology aerosol optical depth (AOD) from satellite data will be corrected for meteorological conditions and spatial-temporal structure to predict air quality for Delhi and its neighboring areas. This methodology will have greater applications for estimating air quality at unprecedented spatial-temporal resolutions for air quality surveillance and management, epidemiological and environmental justice research, because the existing network of air pollution monitoring stations has limited spatial-temporal coverage.

Keywords: Spatial-temporal autoregressive models, air quality, aerosol optical depth.

[†] This is the preliminary draft. Please do not quote and/or cite from this manuscript.

1. INTRODUCTION

The space-time domain is a complicated framework to deal with in a manner that is both scientifically justifiable and computationally feasible^[1]. Although the spatial and temporal structures in cross-sectional and longitudinal data, respectively, have long been addressed, their simultaneous implementation is complicated by the fact that they require three dimensional search to determine covariance structure by time and space. The space-time framework has seen some advancement in the Bayesian arena^[3]. In the spatial domain, we have public health data which consists of counts in designated geographic regions or irregular lattices. We present an approach to incorporate spatial-temporal correlations together. Space-time models have been presented in many contexts such as economics, meteorology, ecology, and public health^[2, 4, 5]. The need for this methodology was felt to estimate air quality which was free from the biases of meteorological conditions and spatial-temporal structures.

There have been significant advances in atmospheric remote sensing in terms of repetitive coverage and spatial resolution, for example Terra and Aqua satellites have a daily repetitive global coverage at 250m resolution. These advances have been exploited to derive indirect estimates of air quality, especially for developing countries, which lack air pollution monitoring at high spatial-temporal resolutions^[6-8]. In principle, electromagnetic radiation l_s from the earth surface changes to l_0 while interacting with the suspended and liquid particles present in the atmosphere prior to reaching the sensor mounted onto a satellite. Radiative transfer models are employed to compute aerosol optical depth (AOD) using the change in l_s to l_0 ^[6, 9-12]. Metrological conditions that can last for several days and weeks can greatly impact aerosol loading, therefore AOD is highly autocorrelated in time.

AOD has three components – aerosols generated by human activities (AOD_h), aerosols generated by natural factors (AOD_n), such as change in meteorological conditions, and aerosols produced through $AOD_h \cap AOD_n$. For air quality studies, however, we are interested in AOD_h . While it is difficult to extract AOD_h from AOD with the aid of radiative transfer models, we propose spatial-temporal autoregressive dynamic modeling approach to estimate AOD_h , which is free from the influence of meteorological conditions and spatial-temporal structures. This, in turn, could be used to study the time space dynamic of air pollution and for air quality surveillance and management worldwide^[13].

This paper is organized as follows. We develop the theoretical model in Section 2. An alternative approach to determining a neighborhood structure is discussed in Section 3. The data and results are described in Section 4 and a discussion follows in Section 5.

2. THEORETICAL MODEL

2.1 *Spatial structure*

In the model that we develop here, we study the autoregressive models both temporally and spatially. Our model builds on the classic autoregressive models. We begin with a conditional autoregressive (CAR) spatial framework. A CAR model is commonly used for irregular lattice models. Suppose that we observe counts in n spatial regions such that

Y_i is the observed value at location i for $i = 1, \dots, n$. Consider \mathbf{Y}_{-i} to be a vector that contains all observations except for the i^{th} case. If we assume each conditional distribution is normal, then the conditional autoregressive model can be written with conditional mean

$$E[Y_i|\mathbf{Y}_{-i}] = \beta X'_i + \gamma \sum_{j=1}^n (c_{ij} Y_j) \quad (1)$$

where γ is a measure of the correlation between neighboring spatial locations. The c_{ij} values equal one if region i is a neighbor of region j and zero otherwise which compile the symmetric matrix \mathbf{C} . The conditional variance can be written as

$$Var[Y_i|\mathbf{Y}_{-i}] = \sigma^2 \quad (2)$$

Equation (1) through (6) show the componentwise structure. Written in matrix notation with $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$, we see

$$Y_i|\mathbf{Y}_{-i} \sim MVN_n(\alpha + \beta X, \sigma^2(\mathbf{I} - \gamma \mathbf{C})^{-1}) \quad (3)$$

2.2 Temporal Structure

We will use a similar approach to design the temporal framework. We let Y_t be the observed value at time t for $t = 1, \dots, T$, Y_{t-1} be the observed value at time $t - 1$ for $t = 2, \dots, T$ and Y_{t+1} be the observed value at time $t + 1$ for $t = 1, \dots, T - 1$. Again we assume that \mathbf{X} is a design matrix of covariates and β is the vector of coefficients. We write the conventional autoregressive model of order 1

$$Y_t|Y_{-t} = \alpha + \beta X'_t + \rho Y_{t-1} + \rho Y_{t+1} + \varepsilon_t \quad (4)$$

where Y_{-t} represents all values of Y other than Y_t and ρ is the correlation between Y_t and values one time point away after controlling for the covariates in \mathbf{X} . The residuals ε_t are assumed independent and normally distributed with mean zero and variance τ^2 . Following (1) we can write

$$E[Y_t|Y_{-t}] = \beta X'_t + \rho \sum_{l=1}^k (d_{il} Y_{t-l}) \quad (5)$$

In this case Y_{t-l} is value of Y at l lag(s) from time t and the neighborhood matrix $D = \{d_{ij}\}$ where the matrix D is $T \times T$ with zeroes down the main diagonal, ones down the diagonal immediately above and immediately below the main diagonal and zeroes elsewhere. The diagonal of ones would represent the neighbors in a temporal setting.

Again following the CAR framework we set the conditional variance as

$$Var[Y_i|\mathbf{Y}_{-i}] = \tau^2 \quad (6)$$

It then follows that $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_T)$ and we have

$$Y_i|Y_{-i} \sim MVN_T(\alpha + \beta X + \tau^2(\mathbf{I} - \rho \mathbf{D})^{-1}) \quad (7)$$

2.3 Spatial-temporal structure

A separable spatial-temporal structure could combine (1) and (5) by including both γ and ρ in the same model as

$$Y_{it} = \alpha + \beta X'_{it} + \gamma \sum_{j=1}^n (c_{ij} Y_{jt}) + \rho \sum_{l=1}^k (d_{ij} Y_{it \pm l}) + \varepsilon_{it} \quad (8)$$

Although computationally it is straight forward, there are two major problems with this model. First, it does not allow for interaction between time and space. Second, incorporating γ and ρ separately is likely to result in overestimation of spatial and temporal correlation, for example the location (i) of event in time t and $t \pm l$ will be same and if there is serial correlation it will obviously result in spatial correlation because the location of the event remain the same. Therefore, we propose a computationally feasible model that uses the linear structure as in (3) and (7) and we will combine the spatial and temporal correlations (θ) into one term. We write this model as

$$Y_{it} = \beta X'_{it} + \theta \sum_{j=1}^n \sum_{l=1, l \neq j, t \neq t \pm l}^k (c_{(ij)} d_{t(t \pm l)} Y_{j(t \pm l)}) + \varepsilon_{it} \quad (9)$$

Here $cd_{(ij)(tl)}$ is a three dimensional matrix a product of C and D as in equation (3) and (7), in which neighbors within n distance and k lag apart are treated as one and zero otherwise. Following the CAR framework we could assume the conditional variance as

$$Var[Y_{it} | \mathbf{Y}_{-i-t}] = \varphi^2 \quad (10)$$

It then follows that $\mathbf{Y}' = (Y_{11}, Y_{12}, \dots, Y_{21}, Y_{nr})$ and we have

$$Y_{it} | \mathbf{Y}_{-i-t} \sim MVN_{nT}(\alpha + \beta X' + \varphi^2 (\mathbf{I} - \theta(\mathbf{C} \times \mathbf{D}))^{-1}) \quad (11)$$

Given the data size (of over 100k observations) computationally it is difficult to implement equation (11), we propose an alternate which will correct the residuals from OLS for spatial-temporal autocorrelation iteratively unless the main assumptions of OLS are satisfied. We write this alternate approach as

$$Y_{it} = \beta X'_{it} + \tilde{\theta} Z_{it} + \tilde{\varepsilon}_{it} \quad (12)$$

where $\tilde{\theta}$ is a proxy of spatial-temporal autocorrelation estimated using linear regression, $\tilde{\varepsilon}_{it}$ is the error term *iid* and Z_{it} is the distance and time weighted average of neighboring

residuals ($\varepsilon_{j(t\pm l)}$) from OLS within n distance range and k lag interval from the i^{th} location and t^{th} time and can be written as

$$Z_{it} = \frac{1}{\sum_{j=1}^n \sum_{l=1, l \neq j, t \neq t \pm l}^k c_{ij} d_{t \pm l}} \sum_{j=1}^n \sum_{l=1}^k \varepsilon_{j(t \pm l)} c_{ij} d_{t(t \pm l)} \quad (13)$$

Where c_{ij} is one when the distance between the j^{th} neighbor and i^{th} location $\leq n$, 0 otherwise. Likewise, $d_{t(t \pm l)} = 1$ if the time lag is $\pm k$ time interval. Treating all neighbors equally within n distance and k lags does not make much sense, because the intensity of correlation with neighbors tends to declines with increasing distance and time. We explain the methodology for defining neighbors and their weights in the next section.

3. NEIGHBORHOOD STRUCTURE

In Section 2 we presented a general model of autocorrelation over both space and time. In this model, there are two matrices that determine where the correlations should exist. In other words, the C and D matrices determine where the neighbors are in both space and in time. Often these neighbors are determined *apriori*. That is, we know which geographic regions are thought to be similar to one another. This is particularly true in irregular lattice structures. For example, we can specify which counties are neighbors of each other in a particular state. In the neighborhood matrix, we specify a one where two counties border each other and a zero otherwise.

Specifying neighbors *apriori* may not always be the best approach, because determining neighbors by lattice boundaries oversimplifies the true underlying spatial phenomenon that one is trying to capture. For example, air pollutants move across time and space and there are no fixed jurisdictional bounds to define spatial-temporal autocorrelation in air pollutants. Therefore, a data driven approach could be more appropriate. Often empirical semi-variograms are used to define the neighbors and their weight for spatial-temporal autocorrelation, separately. In equation (9) $c_{ij} \rightarrow n$ and $d_{t(t \pm l)} \rightarrow k$. Following a geostatistical approach empirical semivariogram can be constructed to estimate n , which is the distance at which fitted semivariogram begins to flatten, which means the intensity of spatial-autocorrelation at this distance becomes statistically insignificant. For example, figure 1 and figure 2 show the semivariogram of the residuals from OLS by different time lags and distance lags, respectively. Looking at these figures it is evident that the temporal and spatial autocorrelations extends up to 75 days and up to 1.2 degree, and gradually decline with the increasing time and distance, respectively. Following this approach the weights of neighbors can be specified by modifying equation (13) as

$$Z_{it} = \frac{1}{\sum_{j=1}^n \sum_{l=1, l \neq j, t \neq t \pm l}^k c_{ij}^{-a} d_{t \pm l}^{-b}} \sum_{j=1}^n \sum_{l=1}^k \varepsilon_{j(t \pm l)} c_{ij}^{-a} d_{t(t \pm l)}^{-b} \quad (14)$$

where c_{ij} is the geographic distance between i^{th} location and j^{th} neighbor, and $d_{t(t \pm l)}$ is the number of days before and after the t^{th} time, $c_{ij} \rightarrow n$ and $d_{t(t \pm l)} \rightarrow k$, and $-a$ and $-b$ are inverse

weights of distance and time, respectively. The shape of semivariogram is generally used to determine the weighting factor and in the absence of *a priori* one could begin with -1. In the analysis we used -1.5 for both time and distance, because the semivariance declines gradually with increase in distance and time intervals.

Although the semivariogram in figure 1 and 2 could provide some insight into the extents of spatial and temporal correlation separately, temporal autocorrelation can influence spatial autocorrelation and vice-versa and could result in over-estimation if specified separately as in equation (8), because

$$\gamma + \rho \neq \theta \quad (15)$$

Therefore the conventional way of defining distance range (n) and the extent of time lags (k) using a two dimensional empirical semivariogram will be of little use. Defining neighborhood structure across time and space will require a 3-D semi-variogram, in which semi-variance is plotted against different time and distance lags/intervals. It is rather difficult to implement a 3-D variogram graphically; we propose the use of a correlation matrix in which the relationship of a value at time t and location i is mapped with the neighboring values at different spatial-temporal lags together (Table 1a & 1b). We experimented with different spatial-temporal ranges and distance and time intervals, and finally learned that neighbors are correlated within 0.15 degree distance and 3 months time lags (for our dataset), and obviously neighbors, which are closer in time and space, have stronger association and we write the final weighting scheme as

$$\tilde{Z}_{it} = \frac{1}{\sum_{m=1}^M r_{it(m)}} \sum_{m=1}^M Z_{it(m)} r_{it(m)} \quad (16)$$

where r is the coefficient of correlation between residuals from OLS (ε_{it}) and Z_{it} at m distance and time lags.

3. DATA AND METHODS

For demonstrating the application of the proposed model, we aim to correct AOD for the bias of meteorological conditions and spatial-temporal structure, needed for predicting air quality at high spatial-temporal resolutions. The data for this analysis come from two different sources – (1) AOD from NASA, and (2) meteorological data from the National Climatic Data Center (NCDC).

3.1 Satellite Data: We used data from MODIS onboard Terra satellite (10:30A local equatorial crossing time)^[14]. National Aeronautics and Space Administration (NASA) aerosol land team has developed an algorithm to extract aerosol over land and ocean^[15]. Using this algorithm daily AOD are extracted globally at various spatial resolution and made available through Goddard Earth Sciences Distribution Active Archive Center (DAAC) and from Level 1 and Atmosphere Archive and Distribution System (LAADS)

in recent years^[16]. These 10km (at nadir) aerosol over land and ocean product from 2000 to 2006 for MODIS Terra were acquired from DAAC. The algorithm used to estimate AOD is discussed in Remer et al.^[15] The AOD products were arranged in granule and stored in HDF4 format^[17]. Not only does each granule covers very large geographic area (more than 2300² km longitudinal extent in each granule) but also a large number of variables. A computer application was written in .net C++ to extract AOD within the geographic extent from 75.45E to 78.55E and 27.45N and 29.55N, which covers the entire National Capital Region (NCR) including Delhi metropolitan and 11 districts of the three states surrounding Delhi.

3.2 Meteorological Data: Hourly meteorological data, such as wind direction, atmospheric (sea level) pressure, visibility, temperature and dew points, from 2000 to 2006 were downloaded from the Global Surface Hourly database maintained by the National Climatic Data Center^[18] for all meteorological stations within NCR and these data were extracted within ± 90 minutes of the time stamp of AOD data (i.e. 9:00-12:00h), an optimal time window^[7] and integrated with the closest AOD pixel.

4. RESULTS

Daily AOD for NCR from 2000 to 2006 were aggregated to monthly averages at 5km grid cell. The entire study area was covered by a total of over 2118 grids. The results are presented in Table 2. We begin describing these results with the second column in Table 2 that utilizes the classic OLS model. Although most meteorological conditions were associated significantly with AOD, we used only dew point, because not only does this variable has the most significant association with AOD, but it was also highly correlated with other meteorological conditions, which is makes this variable as the most representative of meteorological conditions. We used a dummy variable, i.e. *monsoon* category, to control for seasonality, because aerosol loading and their type can differ from *monsoon* to non-*monsoon* season^[19].

The next seven columns in table 2 show the influence of spatial-temporal structure using three different schemes of defining neighborhood structure. In the first, spatial-temporal structure is defined in three different bands of temporal-spatial lags. The second includes the cumulative effects of temporal-spatial structure up to three lags, and the final scheme accounts for the influence of neighbors (across diagonal spatial-temporal lags) weighted by the correlation with neighbors at each lag as shown in table 1b. As evident from this table, diagonal first lag show the highest correlation and it gradually becomes insignificant by the third diagonal lag. In cumulative lags, which include all neighbors from the previous lags, the relationship, however, is statistically significant up to the third lag (Table 1a), because all neighbors included in the previous lags are also included in the succeeding lags and correlation could extend beyond the actual spatial-temporal lags. Therefore, to define neighbors' weight we used correlation at different spatial-temporal lags (as in Table 1b) instead of incremental lags.

Our results clearly indicate that dew point shows a statistically significant positive association with AOD and account for one fifth of the total variability in the OLS model. As evident from Figure 3 there is a stronger component of spatial-temporal

autocorrelation in the AOD, and we tried to correct for this using the proposed model in equation (12). From the comparison of seven different spatial-temporal lags across three different schemes, the final scheme that assigns weight to neighbors by their correlation with the residuals at i^{th} location at time t^{th} emerge as the best correction for spatial-temporal autocorrelation (Table 2).

Monthly AOD and averages of AOD at 5km grid cell at original scale are contrasted with the AOD corrected for dew point and spatial-temporal structure, and the results are presented in table 3 and 4 and figure 3 and 4. Although the spatial temporal autocorrelation still remains in AOD, the intensity of spatial-temporal correlation certainty declines and the 95% confidence interval are more robust in the corrected AOD (Table 4), which mean the corrected AOD are more reliable measure of air quality. The corrected estimates will be used to study the time-space dynamics of air pollution in Delhi and its surrounding areas to study in response to a series of air quality regulations implemented in Delhi from 2000 to 2002.

5. DISCUSSION

The proposed methodology is computationally feasible to control for spatial-temporal autocorrelation and provide more robust estimates of air quality than the AOD at the original scale. Using this methodology we were able to correct AOD for dew point and spatial-temporal autocorrelation to a greater extent, the spatial-temporal structure still persists in the corrected AOD. There could be two potential reasons behind it. First, in the analysis we used the (constant) global specification of temporal and spatial lags. The extent of spatial-temporal correlation, however, may vary by seasons and spatial extent. Therefore, the future research will be geared towards accounting for local spatial-temporal autocorrelation as against the global specification used in this analysis. Second, other research suggests that there is an inherent spatial-temporal structure in the air pollution distribution^[8], because the location of different sources of air pollutants are constant and these pollutants disperse with time.

An interesting thing that emerges from this research is the specification of neighborhood structure. Temporal autocorrelation could influence spatial autocorrelation and vice-versa, because plotting autocorrelation across different time lags suggests that autocorrelation persists up to 72 days and within 1.2 degree distance, but when spatial-temporal autocorrelation was examined simultaneously it becomes statistically insignificant after 0.1 degree and within 2 month lag. Our analysis builds on trial and error and we tried all sort of combinations at different spatial-temporal lags to determine the neighborhood structure and its weight. Using a 3-D semivariogram, however, could be more efficient for determining the global extents of spatial-temporal ranges for defining neighbors and their weights. Although these ranges will be inappropriate for the local specification of spatial-temporal autocorrelation, these could serve as seeds for searching for local spatial-temporal autocorrelation.

Reference

1. Giacomini, R. and C.W.J. Granger, Aggregation of space-time processes. *Journal of Econometrics*, 2004. **118**(1-2): p. 7-26.
2. Chen, X.H. and T.G. Conley, A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 2001. **105**(1): p. 59-83.
3. Wikle, C.K., R.F. Milliff, D. Nychka, and L.M. Berliner, Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, 2001. **96**(454): p. 382-97.
4. Neas, L.M., D.W. Dockery, P. Koutrakis, D.J. Tollerud, and F.E. Speizer, The Association of Ambient Air-Pollution with Twice-Daily Peak Expiratory Flow-Rate Measurements in Children. *American Journal of Epidemiology*, 1995. **141**(2): p. 111-22.
5. Schwartz, J., D. Wypij, D. Dockery, J. Ware, S. Zeger, J. Spengler, and B. Ferris, Daily Diaries of Respiratory Symptoms and Air-Pollution - Methodological Issues and Results. *Environmental Health Perspectives*, 1991. **90**: p. 181-87.
6. Chu, D.A., Y.J. Kaufman, G. Zibordi, J.-D. Chern, J.-M. Mao, L. C., and H.B. Holben, Global Monitoring of Air Pollution over Land from EOS-Terra MODIS. *J. Geophys. Res.*, 2003. **108**(D21): p. 4661.
7. Kumar, N., A. Chu, and A. Foster, An Empirical Relationship between PM_{2.5} and Aerosol Optical Depth in Delhi Metropolitan. *Atmospheric Environment*, 2007. **41**(21): p. 4492-503.
8. Kumar, N., A. Chu, and A. Foster, Remote Sensing of Ambient Particles in Delhi and Environs: Estimation and Validation. *International Journal of Remote Sensing (forthcoming)*, 2007.
9. Kaufman, Y.J., N. Gobron, B. Pinty, J.-L. Widlowski, and M.M. Verstraete, Relationship between surface reflectance in the visible and mid-IR used in MODIS aerosol algorithm - theory. *Geophysical Research Letters (0094-8276)*, 2002. **29**(23): p. 31-1 to 31-4.
10. Christopher, S.A., J. Chou, J. Zhang, X. Li, and R.M. Welch, Shortwave Direct Radiative Forcing of Biomass Burning Aerosols Estimated From VIRS and CERES. *Geophys Res. Lett.*, 2000. **27**: p. 2197-200.
11. Dubovik, O. and M.D. King, A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. *Journal of Geophysical Research-Atmospheres*, 2000. **105**(D16): p. 20673-96.
12. Remer, L.A., D. Tanré, and Y.J. Kaufman, *Algorithm for remote sensing of tropospheric Aerosol from MODIS*, NASA, Editor. 2006, Greenbelt, MD: Goddard Space Flight Center, NASA.
13. Kumar, N. and D.A. Foster, Air Quality Interventions and Spatial Dynamics of Air Pollution in Delhi and its Neighboring Areas. *International Journal of Environment and Waste Management*, 2007. **(forthcoming)**.
14. NASA. *MODIS Web*. 2007 [cited; Available from: <http://modis.gsfc.nasa.gov/>].
15. Remer, L.A., Y.J. Kaufman, D. Tanre, S. Mattoo, D.A. Chu, J.V. Martins, R.R. Li, C. Ichoku, R.C. Levy, R.G. Kleidman, T.F. Eck, E. Vermote, and B.N. Holben, The MODIS aerosol algorithm, products, and validation. *Journal Of The Atmospheric Sciences*, 2005. **62**(4): p. 947-73.

16. NASA. *the Level 1 and Atmosphere Archive and Distribution System*. 2007 [cited; Available from: <http://ladsweb.nascom.nasa.gov/>].
17. THG. *Hierarchical Data Format 4 (HDF4)*. 2007 [cited; Available from: <http://www.hdfgroup.org/>].
18. NCDC. *National Climatic Data Center*. 2007 [cited; Available from: <http://www.ncdc.noaa.gov/oa/ncdc.html>].
19. Kumar, N., A.D. Foster, S.N. Tripathi, and A. Chu, Aerosol Optical Depth and PM₁₀ Association in Kanpur. (*manuscript*), 2007.

Acknowledgement: The material is upon work supported by the National Science Foundation under Grant No. _____ and National Institute of Health under grant No. NICHD/NIH (5 R21 HD046571-02).

Table 1a: Product moment coefficient of correlation of residuals (Z_{it}) from OLS at different cumulative lags by time and space

	Residuals of AOD from OLS at cumulative distance and time lags					
	T=0 D=0	T=0-1 D=0-0.05	T=0-2 D=0-0.1	T=0-3 D=0-0.15	T=0-4 D=0-0.2	T=0-5 D=0-0.25
T=0 D=0	1					
T=0-1 D=0-0.05	0.7572	1				
T=0-2 D=0-0.1	0.5986	0.811	1			
T=0-3 D=0-0.15	0.3542	0.5007	0.7888	1		
T=0-4 D=0-0.2	0.0198	0.1298	0.4263	0.8088	1	
T=0-5 D=0-0.25	-0.3186	-0.2905	0.0071	0.4336	0.7967	1

T= Time lag in months
D=Distance lag in degree

Table 1a: Product moment coefficient of correlation of residuals from OLS at different lags (bands) by time and space

Time and Distance Lags	Residuals of AOD from OLS at different distance and time lags					
	T=0 D=0	T=1 D=>0-0.05	T=2 D=0.05-0.1	T=3 D=0.1-0.15	T=4 D=0.15-0.2	T=5 D=0.2-0.25
T=0 D=0	1					
T=1 D=>0-0.05	0.7127	1				
T=2 D=0.05-0.1	0.401	0.5539	1			
T=3 D=0.1-0.15	-0.0065	0.0346	0.3995	1		
T=4 D=0.15-0.2	-0.3664	-0.328	-0.0187	0.535	1	
T=5 D=0.2-0.25	-0.5757	-0.6411	-0.2547	0.2574	0.5582	1

T= Time lag in months

D=Distance lag in degree

Table 2: OLS and spatial-temporal autoregressive model using three different specification of neighborhood structure

	OLS Model	Spatial-temporal lags			Cumulative spatial-temporal lags			Weighted lags at 0.1 degree and 2 months
		T=1 D=1	T=2 D=2	T=3 D=3	T=1 D=1	T=1-2 D=1-2	T=1-3 D=1-3	T=1-2 ^Φ D=1-2
Dew Point(F)	0.023	0.017	0.021	0.023	0.016	0.019	0.021	0.017
	(9.29)**	(5.93)**	(8.12)**	(9.09)**	(5.23)**	(7.05)**	(8.15)**	(6.19)**
Monsoon Category	-0.141	-0.243	-0.161	-0.137	-0.236	-0.188	-0.151	-0.235
	-1.32	(2.41)*	-1.52	-1.27	(2.33)*	-1.79	-1.44	(2.34)*
Spatial-temporal component	NA	0.735	0.541	0.059	0.781	0.801	0.564	0.922
		(3.81)**	(2.86)**	-0.36	(3.77)**	(3.92)**	(2.60)*	(4.60)**
Constant	-1.86	-1.083	-1.405	-1.809	-0.992	-1.123	-1.423	-0.957
	(14.29)* *	(4.41)**	(7.01)**	(10.71)**	(3.70)**	(5.04)**	(6.62)**	(4.08)**
Observations	111678	111557	111663	110605	111678	111678	111678	111542
R-squared	0.21	0.23	0.23	0.21	0.23	0.23	0.22	0.24
Robust t statistics in parentheses * significant at 5%; ** significant at 1% $\Phi = ((y_{i-t-1} \times 0.71) + (y_{i-2t-2} \times 0.4)) / (0.71 + 0.4)$								

Table 3: Mean and 95% (CI) of AOD at original scale and AOD corrected for relative humidity and spatial-temporal autocorrelation using the neighborhood specification in the last column of table 2, 2000-2006

Location	Original		Corrected for dew point & spatial-temporal autocorrelation	
	AOD	ln(AOD)	AOD	ln(AOD)
Delhi (n=2998)	0.732±1.371	-0.417±1.586	0.518±0.790	-0.734±1.358
Outside Delhi (n=108680)	0.602±0.198	-0.625±0.281	0.416±0.112	-0.963±0.244
Total (n=111678)	0.605±0.197	-0.620±0.277	0.419±0.112	-0.957±0.241

Table 4: Monthly original AOD and AOD corrected for dew point, seasonality and spatial-temporal autocorrelation in Delhi and Outside Delhi, 2000-2006

Month	Delhi		Outside Delhi		Total	
	AOD Original	AOD Corrected	AOD Original	AOD Corrected	AOD Original	AOD Corrected
Jan	0.488	0.461	0.451	0.408	0.452	0.409
Feb	0.449	0.415	0.376	0.336	0.378	0.337
March	0.524	0.443	0.418	0.348	0.420	0.350
April	0.751	0.591	0.581	0.449	0.587	0.454
May	1.020	0.681	0.920	0.588	0.924	0.591
June	1.262	0.805	1.043	0.626	1.051	0.632
July	1.516	0.802	1.322	0.674	1.328	0.678
Aug	0.909	0.466	0.734	0.372	0.739	0.374
Sept	0.616	0.377	0.514	0.304	0.516	0.306
Oct	0.807	0.506	0.721	0.436	0.723	0.437
Nov	0.629	0.463	0.583	0.412	0.584	0.414
Dec	0.531	0.444	0.494	0.399	0.495	0.400
Total	0.732	0.518	0.602	0.416	0.605	0.419

Table 3: Monthly AOD.

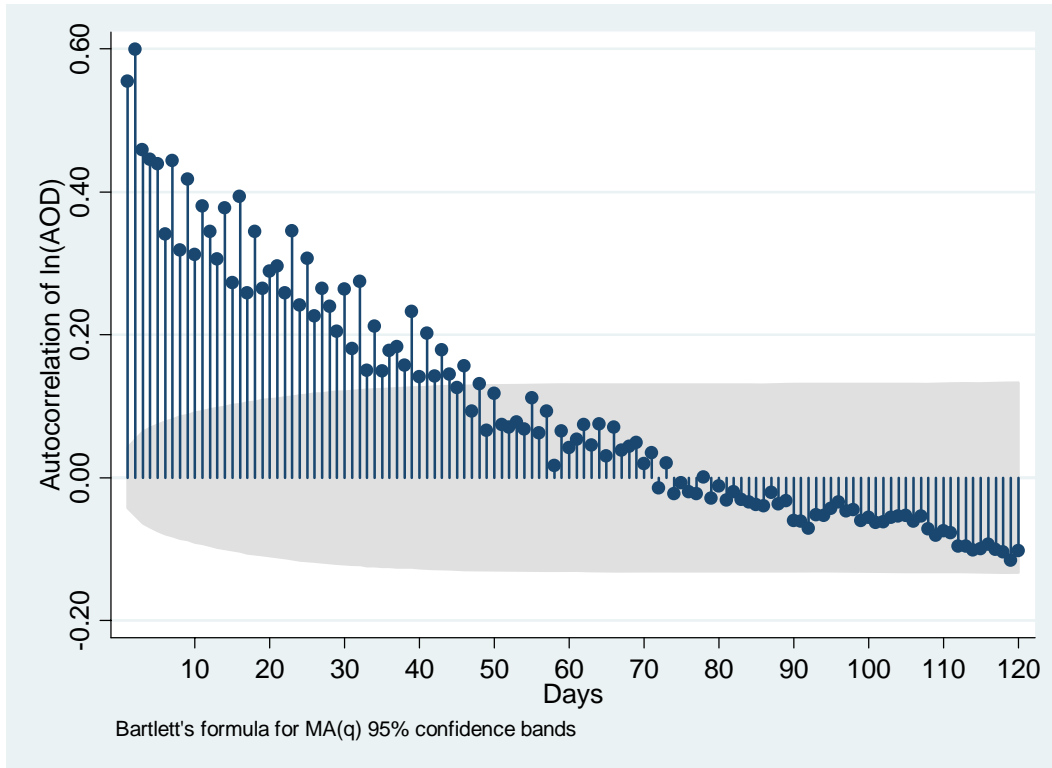


Figure 1: Temporal autocorrelation in AOD in the National Capital Regions, 2000-2006, showing the autocorrelation extends up to 75 days.

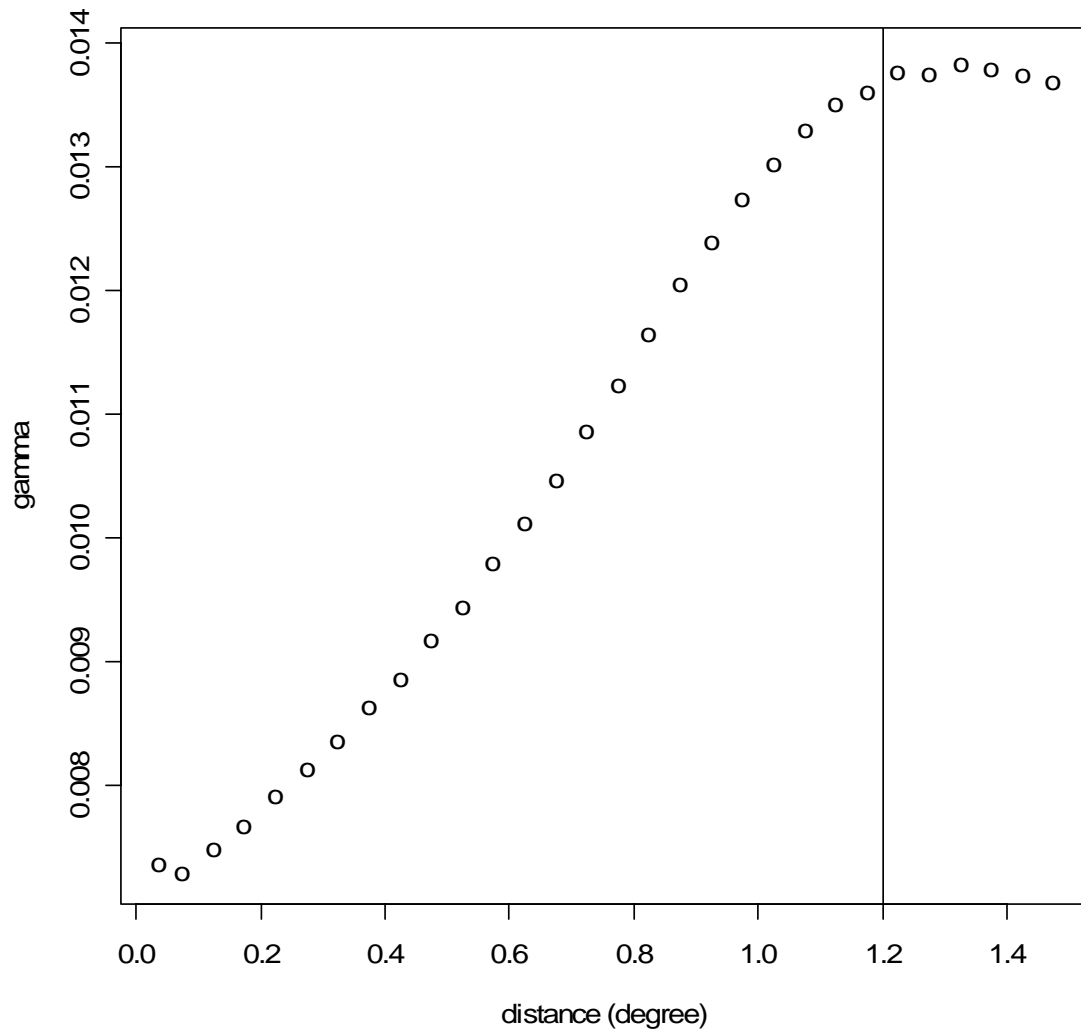


Figure 2: semivariogram of AOD in NCR, 2000-2006 showing the spatial autocorrelation extends up to 1.2 degree.

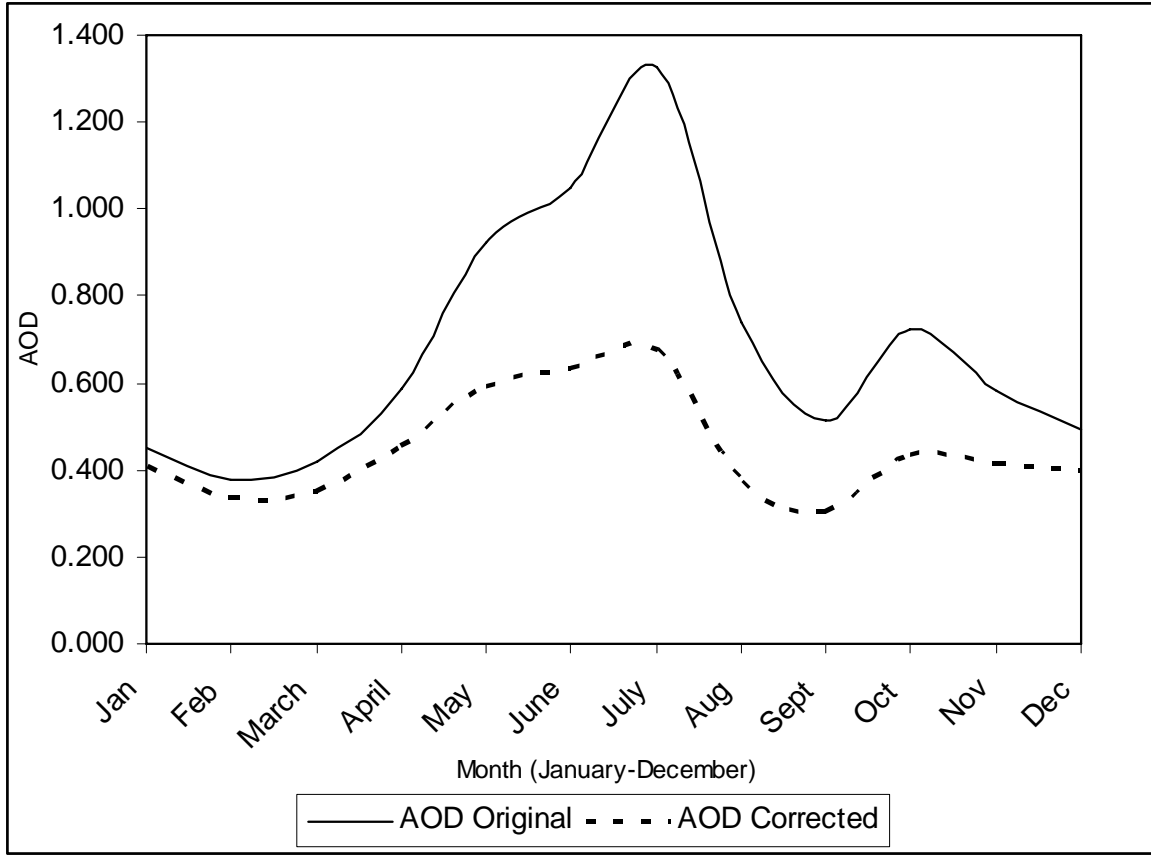


Figure 3: Monthly AOD at original scale and AOD corrected for dew point and spatial-temporal autocorrelation in NCR, 2000-2006

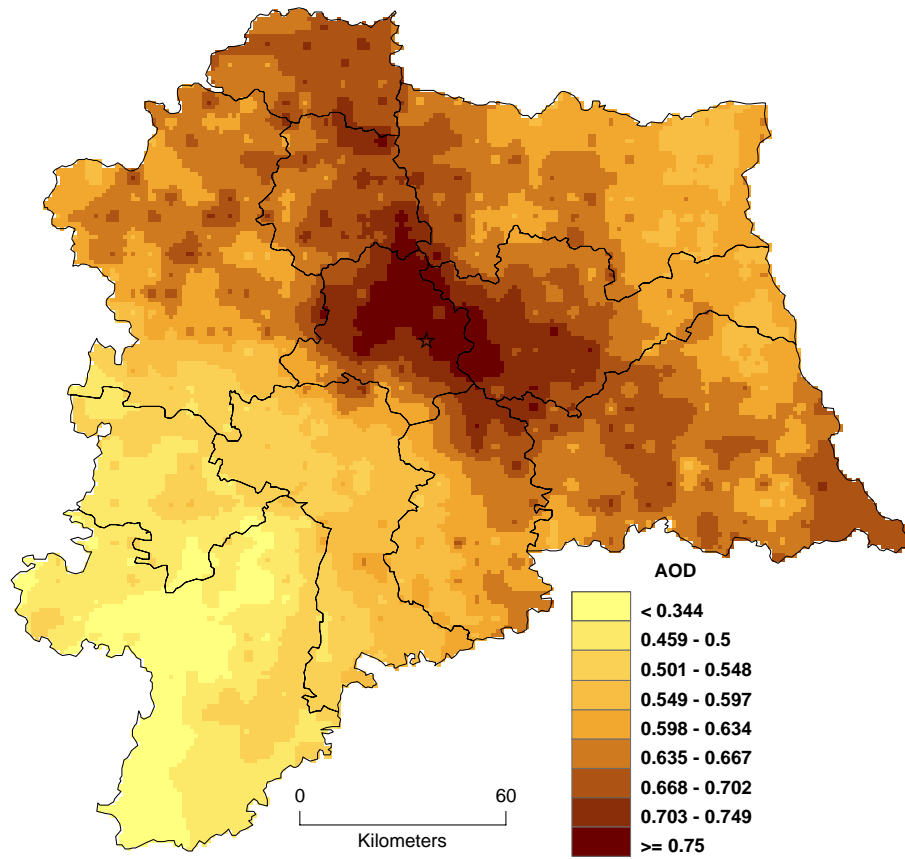


Figure 4a: AOD distribution in NCR, 2000-2006. This surface was interpolated using the AOD (at original scale) values aggregated to 5km grid cell from 2000-2006.

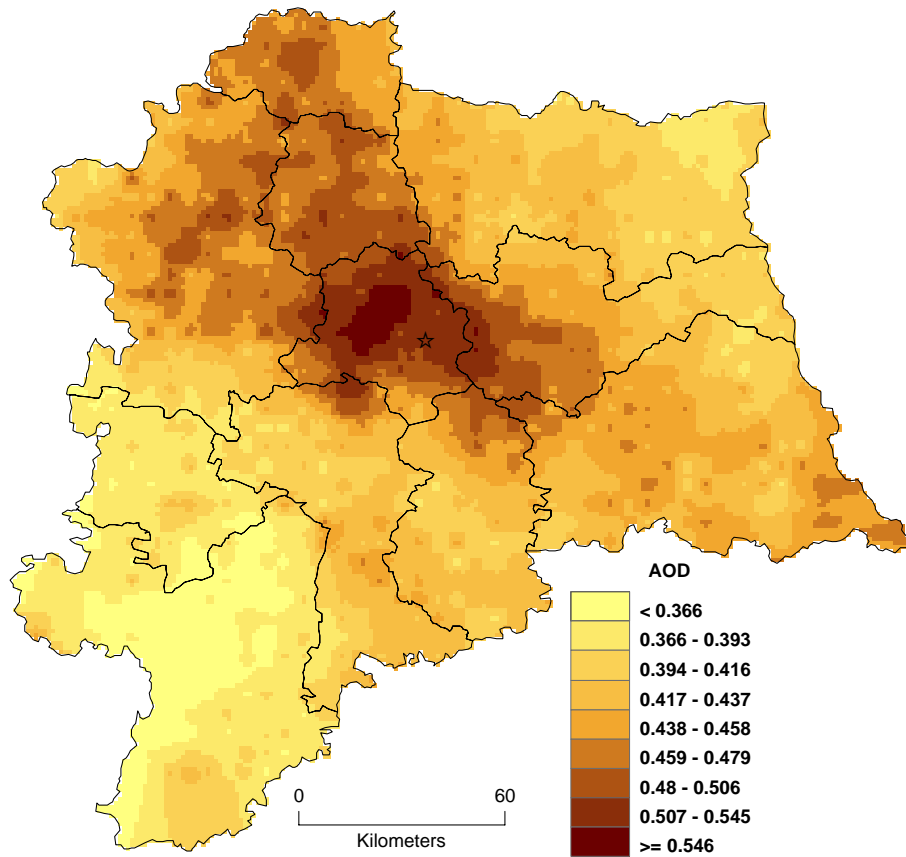


Figure 4b: Distribution of AOD corrected for dew point and spatial-temporal autocorrelation in NCR, 2000-2006. Corrected AOD were aggregated to 5km grid from 2000-2006 to interpolate this surface map.