Published in final edited form as:

Environ Ecol Stat. 2013 June 1; 20(2): 297–314. doi:10.1007/s10651-012-0221-4.

Spatiotemporal modeling of irregularly spaced Aerosol Optical Depth data

Jacob J. Oleson¹, Naresh Kumar², and Brian J. Smith¹

¹Department of Biostatistics, The University of Iowa, Iowa City, IA, USA

²Department of Geography, The University of Iowa, Iowa City, IA, USA

Abstract

Many advancements have been introduced to tackle spatial and temporal structures in data. When the spatial and/or temporal domains are relatively large, assumptions must be made to account for the sheer size of the data. The large data size, coupled with realities that come with observational data, make it difficult for all of these assumptions to be met. In particular, air quality data are very sparse across geographic space and time, due to a limited air pollution monitoring network. These “missing” values make it difficult to incorporate most dimension reduction techniques developed for high-dimensional spatiotemporal data. This article examines aerosol optical depth (AOD), an indirect measure of radiative forcing, and air quality. The spatiotemporal distribution of AOD can be influenced by both natural (e.g., meteorological conditions) and anthropogenic factors (e.g., emission from industries and transport). After accounting for natural factors influencing AOD, we examine the spatiotemporal relationship in the remaining human influenced portion of AOD. The presented data cover a portion of India surrounding New Delhi from 2000 – 2006. The proposed method is demonstrated showing how it can handle the large spatiotemporal structure containing so much missing data for both meteorologic conditions and AOD over time and space.

Keywords

air quality; AOD; autoregressive; Bayesian; spatial correlation; temporal correlation

1 Introduction

Effects of exposure to poor air quality include increased respiratory symptoms, decreased lung function, aggravated asthma, and chronic bronchitis (Environmental Protection Agency, 2010). In addition, particles in the air can make lakes and streams acidic, change the nutrient balance along coasts and river basins, damage forests and crops at high elevations, and affect diverse ecosystems. Thus, it is important to understand as much as we can about sources of air pollution, how it tends to vary over space and time, and what can be done to improve polluted air. Steps to improve polluted air will rely upon knowing what the causes of poor air quality are. Some causes occur from natural phenomena while others will be human influenced, and it is important to be able to differentiate between the two if we are to take corrective measures to improve air quality.

Given the limited spatiotemporal coverage of *in situ* monitoring of air pollution, researchers are increasingly seeking alternate measures of air quality. Among these measures, satellite remote sensing (SRS) offers unprecedented opportunity to develop time-space resolved estimates of air quality in a cost-effective manner (Gupta and Christopher, 2009, Kumar, 2010, Kumar et al., 2007, 2011). Various sensors, aboard many satellites orbiting the earth, have daily global coverage, and capture data at various spatial and spectral resolutions. Among many applications of these data, aerosol optical depth (AOD), computed using a

radiative transfer model, is being utilized extensively for climate change and weather prediction models. The radiative transfer model, in the context of this paper and atmospheric remote sensing, is a method of conceptualizing the change in electromagnetic radiation traveling from the earth's surface to sensors orbiting the earth. See Remer et al. (2006) for additional details. It seems that researchers have in recent years become more interested in developing air quality models using AOD. AOD is the depletion of reflected solar irradiation, caused by the presence of solid and liquid aerosols in the atmosphere, as measured by sensors aboard satellites. A fraction of AOD also captures airborne particulate matter (PM) of different sizes near the surface of the earth (Kumar et al., 2011). PM ($2.5\mu\text{m}$ and $10\mu\text{m}$ in aerodynamic diameter [$\text{PM}_{2.5}$ and PM_{10} , respectively]) is a widely accepted measure of air quality worldwide (World Health Organization, 2006). If sources of the remaining fraction of AOD are identified, AOD can be utilized to develop PM concentration at a given location and time.

AOD represents columnar presence of solid and liquid aerosols, and has three distinct components: aerosols generated by human activities (AOD_h), aerosols generated by natural processes (AOD_n), and aerosols generated through the interaction of AOD_h with AOD_n . Note that the average life time of aerosols is approximately one week and since the locations of emission sources do not change frequently, AOD_h is likely to exhibit a strong spatial pattern. To develop air quality estimates using AOD, either we need to extract AOD_h from AOD or identify covariates of the AOD_n component. While it is difficult to extract AOD_h from AOD using radiative transfer models, this paper presents a Bayesian spatiotemporal model to derive a method to measure AOD_h . For the purposes of this paper, we do not separate out the interaction of AOD_h with AOD_n but assume that this interaction quantity is, at least partly, measured within AOD_n . Although we cannot completely separate AOD_h and AOD_n , we can begin to quantify AOD_h by accounting for many of the factors that drive AOD_n amounts. Such a model can be useful to develop indirect measures of air quality worldwide – needed for air quality surveillance and management – especially for areas that lack adequate *in situ* monitoring of air pollution, such as cities of developing countries.

AOD data recorded by MODerate Resolution Imaging Spectroradiometer (MODIS) aboard the Terra satellite can be utilized to develop daily air quality estimates. Details regarding the Terra satellite are given in section 2.1. However, statistical analysis of the AOD data is complicated for three important reasons: large data dimensions, spatiotemporal misalignment and missing values due to cloud cover and other factors. Paciorek et al. (2009) present a spatiotemporal model for monthly averaged sparse $\text{PM}_{2.5}$. Paciorek and Liu (2009) discuss the importance of and limitations in comparing AOD with $\text{PM}_{2.5}$. Since $\text{PM}_{2.5}$ measurements are often very sparse, work is being done to use AOD as a proxy in improving $\text{PM}_{2.5}$ predictions (Kumar et al., 2011, Liu et al., 2007, McMillan et al., 2009).

This paper aims to address these issues and develops daily estimates of air quality for the National Capital Region (NCR) of Delhi, India. We capture the AOD_n through the mean model. Although we cannot measure every AOD_n process in the mean model, we present an approach that begins to separate out the various components of AOD using factors known to influence AOD_n . We believe that these factors drive most of the spatiotemporal structure in AOD that is due to natural factors. We aim to examine the spatiotemporal structure of AOD_h through variation unaccounted for in the mean model. Of course, the remaining spatiotemporal structure will include not only human induced AOD, but also will include unaccounted for natural variability, non-linearities, or variability that may be due to different environmental variables. However, we believe that a large portion of the spatial structure will be due to human induced factors because of the static nature of the emissions sources. We employ a Bayesian model with interaction between space and time. Spatiotemporal interaction methods typically require the inversion of large matrices or involves dimension

reduction, both of which are not suitable for MODIS data sets because they are either too large or too incomplete. The method that we propose allows for the irregular and missing observations of AOD data points. We apply the model to correct for spatiotemporal variability in AOD resulting from meteorological conditions, and spatiotemporal structure inherent in the AOD measurements.

The remainder of this paper is organized as follows. The data sources are presented in Section 2. We develop the proposed model in Section 3. The modeling results are described in Section 4, and the paper concludes with a discussion of the modeling approach in Section 5.

2 Data

The data for this analysis come from two different sources: AOD from NASA (NASA, 2010) described in Section 2.1, and meteorological data from the National Climatic Data Center (NCDC) (National Climatic Data Center, 2007) described in Section 2.2.

2.1 Satellite Data

We used data from MODIS onboard the Terra satellite (10:30A local equatorial crossing time). The Terra satellite was launched on December 19, 1999 and MODIS became operational on February 24, 2000. The NASA aerosol land team has developed an algorithm to extract aerosol over both land and ocean (Chu et al., 2003, Kaufman et al., 2002, Levy et al., 2007, Remer et al., 2006). Using this algorithm, daily AOD values are extracted globally and made available through Level 1 Atmosphere Archive and Distribution System (LAADS) (NASA, 2010). MODIS data (and data from other sensors as well) are gridded based on the geographic coverage of the viewing angle of the sensor and geographic extent along the satellite path, referred to as granule. The geographic extent of MODIS data is approximately 2300km. Given the large size of the satellite data on 36 different channels, these data are stored in hierarchical data format (HDF) to optimize storage, access, and retrieval (TheHDGGroup, 2012).

We acquired Level 2 AOD products from NASA in HDF format. The spatial resolution of the Level 2 AOD is roughly 10km at nadir (i.e., the satellite path). We extracted AOD (using an application written in C++) within the geographic extent from 75.45E to 78.55E and 27.45N to 29.55N, which covered most of the National Capital Region (NCR) including the Delhi metropolitan area and 11 districts within the three states surrounding Delhi. The details on the study area are available in Kumar et al. (2008), Kumar and Foster (2009).

Because the satellite path is not the same every day, the centroid of each AOD pixel does not remain the same every day. The same satellite path repeats every 16th day. In addition, there are missing AOD values due to cloud cover and cloud contamination. Given the ad-hoc spatial arrangement of the daily AOD grid, a systematic grid of 0.05 degrees (~5.56km) was overlaid onto the study area (Figure 1) and AOD values (if their centroid was within the $0.05^\circ \times 0.05^\circ$ [5.56km $\times$ 5.56km] cell) were averaged to develop a systematic grid of daily AOD at 0.05° spatial resolution. Thus, we have regular observations across space and regular observations across time except for missing values on cloudy days. This resulted in 1,932 cells in a grid that was 46 cells wide and 42 cells high. AOD measurements are highly right skewed and a natural log transformation was taken to improve normality, $\ln(\text{AOD})$. All computations in this paper are taken on the $\ln(\text{AOD})$ scale.

The dates collected for this particular study were from February 26, 2000 through December 31, 2006—a total of 2,501 days. An important note is that none of the grid cells had observations on all 2,501 days. The cell with the fewest observations had recordings on only

68 total days while the cell with the most observations had recordings on 2,079 total days. The values of $\ln(\text{AOD})$ averaged across the entire region are shown in a time series in the top figure in Figure 2. The large number of missing values can be seen in most years, particularly 2000 and 2003. Note that the missing values were for the entire study region and not just individual grid locations.

Observed $\ln(\text{AOD})$ values are presented in the first plot in Figure 3, which are averaged over the year 2000, as no locations had measurements on every day throughout that time period. In viewing Figure 3, we see the largest average concentrations of $\ln(\text{AOD})$ are around the capital city, Delhi, and its surrounding region. This area has the largest number of people and is home to many manufacturing industries and services. Because of regional differences in human demographics and emission sources such as these, it is important to separate AOD values into human, AOD_h , and natural, AOD_n , sources.

2.2 Meteorological Data

Global surface hourly data on meteorological conditions including relative humidity (%), surface temperature ($^{\circ}\text{F}$), ceiling height (hundreds of feet), wind speed (miles per hour), and dew point ($^{\circ}\text{F}$) from 2000 to 2006 were acquired from NCDC for all meteorological stations within NCR. These data were extracted within 90 minutes of the time stamp of AOD data (i.e. 9:00-12:00h), an optimal time window (Kumar et al., 2008), and integrated with the closest AOD pixel. The year 2000 averaged values are plotted for the aforementioned meteorological conditions in Figure 3. Meteorological data are important for spatiotemporal analysis of AOD because meteorological conditions can influence AOD greatly through chemical and physical atmospheric processes and transportation of aerosol across geographic space and time. For example, the value of AOD increases with a rise in relative humidity, not only because it increases concentration of water vapors, but it also inflates particle size (Ramachandran, 2007). Other factors, such as wind speed, can influence aerosols mixing within the boundary layer height and transport aerosols from other areas.

Although only 2% of the meteorologic data are missing, imputation was used to ensure a complete set of values for use as predictors in the planned linear model. Bloomfield et al. (1996) impute missing meteorologic values as predictors when ozone was the outcome. Assuming the data are missing at random (MAR), missing values were imputed using a regression approach described as single imputation in Little and Rubin (2002). In the imputation approach, each meteorologic variable was regressed on the other variables plus continuous time using STATA 10.1 (StataCorp, 2007). In this case, given the relatively small number of missing values, summaries from the original (incomplete) dataset and the imputed (complete) dataset are very similar.

To assess whether all variables contributed unique information for modeling AOD, the variables were entered into a multiple regression model. The variance inflation factor (VIF), which is a method to assess multicollinearity, was calculated and all VIFs were less than 9. The traditional rule of thumb is any VIF greater than 10 is a sign of multicollinearity. Thus, we continue with the five meteorologic variables in the analysis. Note, in a regression analysis that ignored spatiotemporal correlation, the five meteorologic variables explained 17% of the variability in $\ln(\text{AOD})$.

3 Methodology

3.1 Data Model

The space-time domain is a complicated framework to derive a model that is both scientifically justifiable and computationally feasible (Giacomini and Granger, 2004).

Although the interaction between space and time for the irregularly spaced AOD data must be accounted for in some way, setting up correlation matrices between all 1,932 grid cells and 2,501 days is inefficient and impractical. In addition, air quality data, such as these, are regularly not fully observable.

We consider modeling the spatial and temporal relationships using the autoregressive format. The irregularities in the missing data makes it impracticable to fit a dynamic space $\times$ time interaction (Wikle et al., 1998). Consequently, to practically account for the interaction while accommodating data irregularities, we develop an extended version of the type of models proposed by Waller et al. (1997), Bernardinelli et al. (1995), and Sun et al. (2000), in which heterogeneity effects and spatial effects are nested within time. Our approach differs in that the time scale for the nesting is at a different scale (larger) than we are interested in. Therefore, an additional parameter is included to account for fine scale time changes.

The modeling was performed on the $\ln(\text{AOD})$ scale as done in Kang et al. (2010). Our model can be expressed as

$$\ln(\text{AOD}_{ij(k)}) = \mathbf{X}_{ij(k)}\beta + \mathbf{Z}_{i(k)} + \mathbf{T}_{j'} + \mathbf{e}_{ij(k)} \quad (1)$$

where $k = 2000, \dots, 2006$ denotes year and $i = 1, \dots, N$ represents the $N = 1,932$ five kilometer square grids. Note that indices for daily measurements are denoted by either j or j' . Let $j = 1, \dots, J_k$ represent the days of year k in the study period such that $J_k = M$ and $j' = 1, \dots, M$ represents the effects of day j' with a total of $M = 2,501$ days.

The mean model, $\mathbf{X}_{ij(k)}\beta$, is assumed to capture the aerosols generated by natural factors, AOD_n . Not only does it capture the mean, but the factors of the meteorological conditions are spatiotemporal and will capture some of the space-time structure within AOD_n . Thus, the parameters $\mathbf{Z}_{i(k)}$ and $\mathbf{T}_{j'}$ capture part of the space and time relationships, respectively, of AOD resulting from human activities, AOD_h , along with unaccounted for AOD_n measures, as previously mentioned. The individual space and time differences are modeled as $\mathbf{Z}_{i(k)} + \mathbf{T}_{j'}$. The p covariates represented in the columns of $1 \times p$ dimensional $\mathbf{X}_{ij(k)}$ are included in the model as fixed effects with $\beta = (\beta_1, \dots, \beta_p)'$ representing the impact of the p covariates on $\ln(\text{AOD}_{ij(k)})$. The columns of $\mathbf{X}$ were standardized to a mean of zero and variance of one so that the relative sizes of β could be compared. Even though the model would allow for yearly or seasonal coefficients on the fixed effects, the coefficients in our model are assumed constant over both space and time. It should be noted that seasonality is still captured by the seasonal fluctuation in the daily recordings of the fixed effects values; we just don't allow for different multiplying effects per season. The $\mathbf{Z}_{i(k)}$ are the effects of the i^{th} grid cell nested within year k , which attempt to capture spatially structured clustering among cells per year. The $\mathbf{e}_{ij(k)}$ are extra variation effects for the i^{th} cell of year k which are intended to reflect unstructured heterogeneity, i.e., account for the unexplained factors of the model and are assumed $N(0, \delta_{e(k)})$.

A simpler model using a single spatial term that did not vary by year, $\mathbf{Z}_i$ rather than $\mathbf{Z}_{i(k)}$, was investigated. Using the single spatial model, we computed the posterior predictive p-value (Gelman et al., 1996) for each year based on the goodness-of-fit discrepancy measure

$$X^2(y_{ij(k)}; \theta) = \sum_{ij} \frac{(y_{ij(k)} - E(y_{ij(k)}|\theta))^2}{\text{Var}(y_{ij(k)}|\theta)}$$

where θ represents the model parameters. Posterior predictive p-values that deviate from a value of 0.50 indicate differences between $\ln(\text{AOD})$ values that were observed and those that would be predicted by the model. Since significant p-values were observed for some years, this indicated that the spatial patterns should be modeled separately across years.

In model (1), there are no fixed effects terms specified for human factors as we are concerned with modeling the natural factors explicitly so that remaining variability can be assumed to be from human activities. Known contributors to AOD_h could be part of the design matrix, X_{ij}^k , if that were the goal. Here, we only wish to partition AOD_n and AOD_h .

Multiple missing observations in grid cells across time is clearly a problem that cannot be ignored. Imputing values and using a dimension reduction process is a possible solution. However, imputations would require detailed knowledge of the AOD spatiotemporal structure from human sources, natural sources, and the interaction of the two. An advantage of the proposed model is that it does not require observations in every cell of the grid to define the correlation matrix. Indeed, a true interaction model would be extremely difficult given the large number of missing observations. Details are given in the following two sections.

3.2 Distribution of Spatial Effects

The spatial effects per year, $Z_{(k)} = (Z_{1(k)}, \dots, Z_{N(k)})'$ are designed to capture any remaining spatial relationship not accounted for by the natural factors. Again, the $Z_{i(k)}$ are the effects of the i^{th} grid cell nested within year k . We use the conditionally autoregressive (CAR) models originally introduced by Whittle (1954) and Besag (1974) to induce spatial correlation. In these models, spatial association is defined through a neighborhood structure where one region is correlated with other regions that are defined to be neighbors. We define regions to be neighbors if their five kilometer square grids are adjacent (i.e., share a border).

Specifically, define an $N \times N$ adjacency matrix as $C = (C_{uv})$ where $C_{uv} = 1$ if two regions u and v share a common boundary and 0 otherwise, with $C_{uu} = 0$. In this framework, the conditional mean and variance are

$$E\left(Z_{i(k)} | Z_{-(i)(k)}\right) = \rho_{z(k)} \sum_{j=1}^N C_{ij} Z_{j(k)} \quad \text{and} \quad \text{Var}\left(Z_{i(k)} | Z_{-(i)(k)}\right) = \delta_{z(k)},$$

where the parameters $\delta_{z(k)}$ represents the degree of spatial precision in year k and $Z_{-(i)(k)}$ are all the elements in $Z_{(k)}$ other than location i for year k . This particular specification of the spatial distribution (the proper autoregressive model) is proper if $\lambda_1^{-1} \leq \rho_z(k) \leq \lambda_N^{-1}$ where $\lambda_1 \dots \lambda_N$ represent the ordered eigenvalues of C (Sun et al., 2001). For the spatial grid of interest, $\lambda_1^{-1} = -0.2512$ and $\lambda_N^{-1} = 0.1255$. It is expected that $\rho_{z(k)}$ will be near the upper boundary to produce spatial similarity. See Banerjee et al. (2004) for additional details on CAR models.

Within any given year, the replicate observations at every grid cell are averaged together since this structure assumes the same spatial structure for every day within that year. Furthermore, if a grid cell being modeled has no recorded data in it, then the CAR model will draw strength from its neighboring cells concentrating on the local average. If the neighboring cells have no information then the value moves toward the global average.

3.3 Distribution of Temporal Effects

A difficulty in specifying a temporal model is the vast amount of missing data. Of the 2,501 days in the study period, the entire region is missing AOD values for 309 (12%) of them. With irregular time points, using a spatial correlation matrix for time is a crafty option because it allows for the observed value to be related to another observed value, no matter how many time points away it would be. Unfortunately, the period of time in this study is too long to be computationally feasible for the matrix inversion involved. The temporal CAR creates a random walk and is another option if there isn't too much missing data. Aggregating days and modeling weekly or monthly would be another approach, but we wish to model daily, because the goal of the proposed method is to develop daily estimates of air quality.

There is some seasonality seen in time series of observed $\ln(\text{AOD})$ values in the top figure of Figure 2. Each point represents spatially averaged $\ln(\text{AOD})$ values with lower values in winter months and higher values in summer months. In fact, preliminary analyses from $\ln(\text{AOD})$ measurements suggested an AR(1) prior for the temporal trend to be appropriate, so we assume a basic AR(1) temporal process. The seasonal component is modelled as an annual harmonic with amplitudes and phases. The $T = (T_1, \dots, T_M)'$ parameter represents the temporal effects of each day in the model.

Assume that the $\ln(\text{AOD})$ value is an annual harmonic correlated with the previous day so that the AR(1) is written as

$$T_{j'} = \gamma_1 \cos(\omega j') + \gamma_2 \sin(\omega j') + \rho_t T_{j'-1} + \epsilon_{j'} \quad (2)$$

$$\epsilon_{j'} \sim N(0, \delta_t)$$

where $\omega = 2\pi/365$ for the annual harmonic and days $j' = 1, 2, \dots, 2501$. Let T_0 have a vague normal prior with mean zero and large variance. Note that γ_1 and γ_2 are the 'cosine' and 'sine' coefficients of the series, respectively, while ρ_t captures the dependence between successive days. This model does not need to be stationary as ρ_t does not need to be restricted to be less than 1.

When there are days with no observed values for the entire study region, then the effect draws strength from the previous days; and if multiple days in a row are missing, then the effect moves toward the global average. That is, the value of $T_{j'}$ will converge towards the global (historical) average. This phenomenon is most noticeable when there are multiple days in a row with no observations. Modeling in this way allows the time series to fully utilize all observed data without needing to aggregate days.

3.4 Hyperparameters

The model parameters are assigned prior distributions in the Bayesian hierarchical framework. Assume conjugate priors for β_q for $q = 1, \dots, p$ such that $\beta_q \sim N(0, 1000)$. For the temporal correlation parameters, ρ_t has a $N(0, 1000)$ prior so that stationarity is not assumed. The seasonal trend parameters are independent normal random variables $\gamma_1 \sim N(0, 1000)$ and $\gamma_2 \sim N(0, 1000)$. The $\rho_{z(k)}$ parameters are uniform $(-0.2512, 0.1255)$. The variance components $\delta_{z(k)}$, δ_t , and δ_e are $IG(0.01, 0.01)$. Results are insensitive to a shift in the prior means.

3.5 Estimation Via MCMC

The design matrix, X , has five columns representing five variables of meteorologic conditions. In order, the five columns are dew point, relative humidity, wind speed, ceiling height, and surface temperature with corresponding beta coefficients $\beta_1, \dots, \beta_5$.

Bayesian computation via numerical integration is not feasible for this hierarchical model. Instead, we use Gibbs sampling for the computation. Full conditional distributions were derived and the model was programmed in Fortran 95. We used a sample of 50,000, and discarded a burn-in sample of 11,000. Geweke convergence diagnostics (Geweke, 1992) have p-values greater than 0.05 for all parameters. In addition, visual assessment of the chains yielded convergence for all parameters. The goodness of fit was measured by posterior predictive assessment (Gelman et al., 1996) indicating no evidence of lack of fit (p-value=0.4948). In addition, posterior p-values were computed for each year and all p-values were approximately 0.5.

4 Results

A summary of the MCMC samples from the posterior distribution are displayed in Table 1 for most of the parameters included in the model. None of the meteorological effects have 95% intervals covering zero. Thus, even though the meteorologic variables are clearly related to one another, they contribute unique information to $\ln(\text{AOD})$. The coefficients are standardized so that a larger absolute posterior estimate signifies a more important variable conditional on all other variables. Therefore, ordered from largest to smallest, the most important variables are dew point, surface temperature, relative humidity, wind speed, and ceiling height.

The plots in Figure 3 show $\ln(\text{AOD})$ along with each of the covariates averaged across the year 2000. Spatially, we see definite similarities between $\ln(\text{AOD})$ and the variables dew point and relative humidity. The lowest values appear in the southwest corner with values increasing to the northeast. Interestingly, even though the highest $\ln(\text{AOD})$ values appear in and around Delhi, none of the meteorologic variables display this particular characteristic.

Figure 3 gives a snapshot in time for each of the covariates. To display the meteorological changes over time, we combine the variables into AOD_n by computing the posterior mean estimates of $X\beta$ as shown in Figure 4. These figures are yearly averages per grid cell for each year (2000 – 2006). Much of the pattern would appear to come from dew point and relative humidity which show the same southwest to northeast gradient. An interesting result to take from Figure 4 is the spatiotemporal interaction picked up by the covariates. Although most of the years do display a similar spatial pattern, there are clear yearly differences which are particularly evident in 2003 and 2005. It should also be pointed out that the large values around Delhi are not captured by mean AOD_n .

We now turn to the amount of spatial and temporal structure remaining after adjusting for the meteorologic variables. The estimated temporal pattern of T gives an assessment of the AOD_h temporal structure and can be viewed as the bottom figure in Figure 2. The x-axis displays the 2,501 days from the start of study period on February 26, 2000 until the ending day December 31, 2006. There appears to be variation that was not captured by the AOD_n given the large level of temporal variance (0.1392). The peaks in Figure 2 appear approximately during the summer months and the troughs during the winter months. Specifically, the three highest values in Figure 2 are June 28, 2000, July 27, 2002, and July 14, 2006. The four lowest values appear on December 20, 2000, February 24, 2004, March 2, 2004, and March 23, 2006. The posterior mean of the daily correlation in the temporal process was estimated at 0.8566. That amount of correlation is a significant amount of

unexplained variation that would be explained by human influenced AOD_h not explained by the meteorologic variables. It is also interesting to examine the annual harmonic amplitude ($amp = (\gamma_1^2 + \gamma_2^2)^{0.5}$) estimated at 0.039 and phase ($phs = \tan^{-1}(\gamma_2/\gamma_1)$) estimated at -0.874 . As expected the phase peak occurs in July (i.e. $amp * \sin(2\pi j/365 + phs)$, $j = 1, \dots, 365$ is a maximum at $j = 142$, corresponding to a maximum on July 17).

The estimated posterior means of the spatial effects, $Z_{(k)}$, are displayed in Figure 5. The estimates capture the remaining spatial structure of aerosols not explained by the meteorologic conditions. That is, the estimates represent the spatial structure in human generated aerosols, AOD_h . Values higher than zero indicate locations that have a higher than average AOD_h value. As expected, since none of the meteorologic variables picked up the large readings around Delhi, these maximum cells are left to be explained as AOD_h . The spatial correlation term from the CAR model is at the upper boundary of 0.1255 in each year indicating strong spatial relationships. The high correlation and spatial variance values indicate that, even though some of the spatial pattern in $\ln(AOD)$ is captured by meteorologic conditions, a substantial amount of spatial variability remains unaccounted for. Clearly, the spatial process parameters capture the maximum around Delhi.

These results also validate the use of separate spatial structures per year. Although the correlations are all very similar, the estimated spatial variance components are substantially different with the most variability appearing in 2006. The disparity in variance estimates indicates the need for some form of spatiotemporal interaction when modeling $\ln(AOD)$. The final terms that require discussion are the model errors. Separate model errors were used for each year. Interestingly, the yearly spatial terms don't capture all of the remaining year-to-year variation which requires the additional error terms.

5 Discussion

It is clear that the meteorological variables do not capture the highest levels of $\ln(AOD)$ centered around Delhi, India. Thus, we can begin to separate measurements of AOD into the components that are due to nature, AOD_n , and those that are due to human intervention, AOD_h . Outside of the metropolitan area of Delhi, the spatial pattern of AOD tends to relate highly to the meteorologic conditions, particularly dew point and relative humidity. Additional variables can easily be incorporated into the Bayesian hierarchical model. We could consider other known human influences of AOD or perhaps natural influences. Of course, the present version of the model excludes additional effects such as natural variation, non-linearities or data manipulations, and different environmental variables as part of the AOD_h component. These relationships need to be further explored in future research in order to take the AOD values and better dissect the data into human and natural components.

We show here, a method for modeling AOD that are measured daily and at a fine spatial grid. AOD measurements at that fine of a scale inherently include missing values, which can be handled by our spatiotemporal modeling approach. Furthermore, our approach allows for spatiotemporal interaction through the use of different spatial structures per year. A finer scale time structure is additionally used to model daily differences. Given that the data are too sparse to inform fine scale spatio-temporal interaction, our approach provides an informative method of incorporating necessary spatial changes over time but still modeling at a fine temporal scale.

Even with the modeling shown here, further work is needed to deal with outstanding issues such as missing covariate data, and separating out the interaction of AOD_h and AOD_n . Future work will also develop relationships between AOD_h , AOD_n , and PM_{10} values. Part of this work will be to involve more specific effects for the Delhi region. The work in this

paper focused solely on meteorologic conditions and we intend to further develop the sources due to human intervention in future work. Incorporating relationships with PM_{10} values also requires additional work because it cannot be measured at such a fine scale. If PM_{10} can be shown to have the same properties as AOD_h , then we can use the fine scale of AOD_h to relate to health outcomes or, at least, develop indirect estimates of air quality.

Since AOD and PM_{10} are independent measures of air quality, we would like to understand their relationship better. One of the advantages of AOD is that hourly measurements are available worldwide since measurements come from satellite readings. Our modeling approach further increases the utility of these measurements by breaking AOD into two important components, AOD_h and AOD_n .

Acknowledgments

This work was partially supported by NIH grant R21 ES014004-01A2. We also thank two reviewers for their constructive comments on an earlier version of the manuscript.

References

- Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall; 2004.
- Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*. 1995; 14:2433–2443. [PubMed: 8711279]
- Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological*. 1974; 36:192–236.
- Bloomfield P, Royle JA, Steinberg LJ, Yang Q. Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment*. 1996; 30:3067–3077.
- Chu DA, Kaufman YJ, Zibordi G, Chern JD, Mao J, Li C, Holben BN. Global monitoring of air pollution over land from the earth observing system terra moderate resolution imaging spectroradiometer (MODIS). *Journal of Geophysical Research*. 2003; 108:4661.
- Environmental Protection Agency. Particulate Matter. 2010. Available from: <http://www.epa.gov/air/particlepollution/>
- Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*. 1996; 6:733–807.
- Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Disc: P189-193). In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*; Clarendon Press [Oxford University Press]; 1992. p. 169-188.
- Giacomini R, Granger CWJ. Aggregation of space-time processes. *Journal of Econometrics*. 2004; 118:7–26.
- Gupta P, Christopher SA. Particulate matter air quality assessment using integrated surface, satellite, and meteorologic products: 2. a neural network approach. *Journal of Geophysical Research - Atmospheres*. 2009; 114:D20205. doi:10.1029/2008JD011497.
- Kang EL, Cressie N, Shi T. Using temporal variability to improve spatial mapping with application to satellite data. *The Canadian Journal of Statistics*. 2010; 38:271–289.
- Kaufman YJ, Gobron N, Pinty B, Widlowski J-L, Verstraete MM. Relationship between surface reflectance in the visible and mid-IR used in MODIS aerosol algorithm - theory. *Geophysical Research Letters*. 2002; 29:2116.
- Kumar N. A hybrid approach for predicting $PM_{2.5}$ and aerosol optical depth in Delhi metropolitan. *Environmental Health Perspectives*. 2010; 118:a425–a425. doi:10.1289/ehp.1002706. [PubMed: 20884398]
- Kumar N, Chu A, Foster A. An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi metropolitan. *Atmospheric Environment*. 2007; 41:4492–4503. [PubMed: 22180723]

- Kumar N, Chu A, Foster A. Remote sensing of ambient particles in delhi and its environs: estimation and validation. *International Journal of Remote Sensing*. 2008; 29:3383–3405. [PubMed: 22162895]
- Kumar N, Chu AD, Foster AD, Peters T, Willis R. Satellite remote sensing for developing time-space resolved estimates of ambient particulate in Cleveland, OH. *Aerosol Science and Technology*. 2011; 45:1090–1108. [PubMed: 22238503]
- Kumar N, Foster A. Air quality interventions and spatial dynamics of air pollution in Delhi and its surroundings. *International Journal of Remote Sensing*. 2009; 29:3383–3405. [PubMed: 22162895]
- Levy RC, Remer LA, Dubovik O. Global aerosol optical properties and application to moderate resolution imaging spectroradiometer aerosol retrieval over land. *Journal of Geophysical Research-Atmospheres*. 2007; 112:D13210. doi:10.1029/2006JD007815.
- Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. Wiley; 2002.
- Liu Y, Franklin M, Kahn R, Koutrakis P. Using aerosol optical thickness to predict ground-level pm_{2.5} concentrations in the St. Louis area: a comparison between MISR and MODIS. *Remote Sensing of Environment*. 2007; 107:33–44.
- McMillan NJ, Holland DM, Morara I M, Feng J. Combining numerical model output and particulate data using Bayesian spacetime modeling. *Environmetrics*. 2009; 21:48–65.
- NASA. The Level 1 and Atmosphere Archive and Distribution System. 2010. Available from: <http://ladsweb.nascom.nasa.gov/>
- National Climatic Data Center. NOAA Satellite and Information Service. 2007. Available from: <http://www.ncdc.noaa.gov/oa/ncdc.html>
- Paciorek CJ, Liu Y. Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environmental Health Perspectives*. 2009; 117:904–909. [PubMed: 19590681]
- Paciorek CJ, Yanosky JD, Puett RC, Laden F, Suh HH. Practical large-scale spatio-temporal modeling of particulate matter concentrations. *The Annals of Applied Statistics*. 2009; 3:370–397.
- Ramachandran S. Aerosol optical depth and fine mode fraction variations deduced from Moderate Resolution Imaging Spectroradiometer (MODIS) over four urban areas in india. *Journal of Geophysical Research - Atmospheres*. 2007; 112:D16207. doi:10.1029/2007JD008500.
- Remer, LA.; Tanré, D.; Kaufman, YJ. *Algorithm for remote sensing of tropospheric Aerosol from MODIS*. Goddard Space Flight Center; NASA, Greenbelt, MD: 2006.
- StataCorp. *Stata Statistical Software Release 10*. 2007.
- Sun D, Tsutakawa RK, He Z. Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica*. 2001; 11:77–95.
- Sun D, Tsutakawa RK, Kim H, He Z. Spatio-temporal interaction with disease mapping. *Statistics in Medicine*. 2000; 19:2015–2035. [PubMed: 10900449]
- TheHDGGroup. *Hierarchical data format: Information, support and software*. 2012. <http://www.hdfgroup.org/>
- Waller LA, Carlin BP, Xia H, Gelfand AE. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*. 1997; 92:607–617.
- Whittle P. On stationary process in the plane. *Biometrika*. 1954; 41:434–449.
- Wikle CA, Berliner LM, Cressie N. Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*. 1998; 5:117–154.
- World Health Organization. *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide – Global update 2005 – summary of risk assessment*. Geneva: 2006.

Figure 1.
Dots represent the corners of the 1,932 $0.05^\circ \times 0.05^\circ$ grid locations. Delhi is denoted by the star and the boundaries are the 11 districts surrounding Delhi.

Figure 2.

Time series of $\ln(\text{AOD})$ values. The top figure contains daily geographically averaged $\ln(\text{AOD})$ values for 2000 – 2006. The bottom figure contains daily posterior estimates of T_j .

Figure 3. Observed values of $\ln(\text{AOD})$ and each of the covariates. Values for each spatial location are averaged across all available days in 2000.

Figure 4.

Average value of AOD_n as it varies across years. Each plot is an average for the given year from 2000 - 2006.

Figure 5.
Posterior estimates of spatial parameter $Z_{(k)}$.

Table 1

Hyperparameters and posterior results

Parameter	Prior Distribution	Posterior Mean	95% Credible Interval
β_1	$N(0, 1000)$	0.0453	(0.0419, 0.0499)
β_2	$N(0, 1000)$	-0.0155	(-0.0185, -0.0108)
β_3	$N(0, 1000)$	0.0049	(0.0037, 0.0065)
β_4	$N(0, 1000)$	-0.0047	(-0.0061, -0.0033)
β_5	$N(0, 1000)$	-0.0229	(-0.0277, -0.0181)
T_0	$N(0, 1000)$	-1.4966	(-2.3511, -0.6376)
ρ_t	$N(0, 1000)$	0.8566	(0.8317, 0.8806)
γ_1	$N(0, 1000)$	-0.0234	(-0.0452, -0.0016)
γ_2	$N(0, 1000)$	0.0316	(0.0102, 0.0531)
ρ_{z1}	$\text{Uniform}(-0.2512, 0.1255)$	0.1249	(0.1240, 0.1254)
ρ_{z2}	$\text{Uniform}(-0.2512, 0.1255)$	0.1249	(0.1239, 0.1254)
ρ_{z3}	$\text{Uniform}(-0.2512, 0.1255)$	0.1243	(0.1225, 0.1253)
ρ_{z4}	$\text{Uniform}(-0.2512, 0.1255)$	0.1242	(0.1223, 0.1253)
ρ_{z5}	$\text{Uniform}(-0.2512, 0.1255)$	0.1249	(0.1240, 0.1254)
ρ_{z6}	$\text{Uniform}(-0.2512, 0.1255)$	0.1243	(0.1225, 0.1253)
ρ_{z7}	$\text{Uniform}(-0.2512, 0.1255)$	0.1237	(0.1215, 0.1252)
δ_{e1}	$\text{IG}(0.01, 0.01)$	0.0481	(0.0475, 0.0488)
δ_{e2}	$\text{IG}(0.01, 0.01)$	0.0471	(0.0465, 0.0477)
δ_{e3}	$\text{IG}(0.01, 0.01)$	0.0587	(0.0580, 0.0594)
δ_{e4}	$\text{IG}(0.01, 0.01)$	0.0584	(0.0576, 0.0592)
δ_{e5}	$\text{IG}(0.01, 0.01)$	0.0709	(0.0700, 0.0719)
δ_{e6}	$\text{IG}(0.01, 0.01)$	0.0682	(0.0674, 0.0690)
δ_{e7}	$\text{IG}(0.01, 0.01)$	0.0876	(0.0867, 0.0886)
δ_t	$\text{IG}(0.01, 0.01)$	0.1392	(0.1308, 0.1482)
δ_{z1}	$\text{IG}(0.01, 0.01)$	0.0262	(0.0231, 0.0297)
δ_{z2}	$\text{IG}(0.01, 0.01)$	0.0134	(0.0121, 0.0159)
δ_{z3}	$\text{IG}(0.01, 0.01)$	0.0441	(0.0400, 0.0486)
δ_{z4}	$\text{IG}(0.01, 0.01)$	0.0243	(0.0221, 0.0268)
δ_{z5}	$\text{IG}(0.01, 0.01)$	0.0244	(0.0221, 0.0270)
δ_{z6}	$\text{IG}(0.01, 0.01)$	0.0186	(0.0169, 0.0205)
δ_{z7}	$\text{IG}(0.01, 0.01)$	0.1024	(0.0941, 0.1116)