Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 705–716
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Interannual variation in meteorologically adjusted ozone levels in the eastern United States: A comparison of two approaches

Junyu Zheng[a,1], Jenise L. Swall[b,*,2], William M. Cox[c], Jerry M. Davis[a,3]

[a]*Atmospheric Modeling Division, National Exposure Research Laboratory, US Environmental Protection Agency, RTP, NC 27711, USA*
[b]*Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, RTP, NC 27711, USA*
[c]*Air Quality Assessment Division, Office of Air Quality Planning and Standards, US Environmental Protection Agency, RTP, NC 27711, USA*

Received 13 March 2006; received in revised form 22 August 2006; accepted 8 September 2006

Abstract

Assessing the influence of abatement efforts and other human activities on ozone levels is complicated by the atmosphere's changeable nature. Two statistical methods, the dynamic linear model (DLM) and the generalized additive model (GAM), are used to estimate ozone trends in the eastern United States and to adjust for meteorological effects. The techniques and resulting estimates are compared and contrasted for four monitoring locations chosen through principal components analysis to represent regional patterns of ozone concentrations. After adjustment for meteorological influence, overall downward trends are evident at all four locations from 1997 to 2004. The results indicate that the two methods' estimates of ozone changes agree well. When such estimates are needed quickly, or when many similar, but separate analyses are required, the ease of implementation and relative simplicity of the GAMs are attractive. The DLMs are much more flexible, readily addressing such issues as autocorrelation, the presence of missing values, and estimation of long-term trends or cyclical patterns. Implementation of DLMs, however, is typically more difficult, and especially in the absence of an experienced practitioner, they may be better reserved for in-depth analyses.

Keywords: Ozone trends; Dynamic linear model; Generalized additive model; Meteorological adjustment; Principal components analysis

1. Introduction

Ozone (O_3) is photo-chemically produced through a combination of chemical reactions involving a variety of volatile organic compounds (VOCs) and nitrogen oxides (NO_x), which are emitted by motor vehicles, by large stationary sources, and by natural sources. Because exceedingly high ozone levels may cause adverse health effects, some pollution control programs have been implemented in the United States (US) to reduce

*Corresponding author. Tel.: +1 919 541 7655; fax: +1 919 541 1379.
E-mail address: Jenise.Swall@noaa.gov (J.L. Swall).

[1]Present Address: School of Environmental Science and Engineering, South China University of Technology, Guangzhou, Guangdong Province 510641, PR China.

[2]In partnership with the US Environmental Protection Agency, National Exposure Research Laboratory.

[3]Visiting from the Department of Marine, Earth & Atmospheric Sciences, Box 8208, North Carolina State University, Raleigh, NC 27695, USA.

1352-2310/$ - see front matter
doi:10.1016/j.atmosenv.2006.09.010

emissions of the precursors in the years since 1990. These programs include the NO_x Reasonable Available Control Technology (RACT) rule, Clean Air Act Amendments (CAAA) Title IV, NO_x State Implementation Plan (SIP) Call, and other VOC reduction strategies (NRC, 2004). Due to the implementation of these emission control programs, significant reductions of regional NO_x and VOCs have occurred in the eastern United States. For example, during the period 1997–2004, there were 25% and 21% reductions in NO_x and VOC emissions, respectively, in the eastern portion of the country (EPA, 2005a, Chapter 2), with additional reductions anticipated in the future. These emission reductions are expected to reduce tropospheric ozone concentrations.

Recent EPA analyses of ozone trends show that ozone levels in this region have been gradually decreasing in recent years (EPA, 2005a). EPA's Report (2005a) also makes clear the strong association between ozone levels and meteorological conditions and notes that improved ozone air quality may result from favorable weather conditions, from continuing reductions in emissions of NO_x and VOCs, or from a combination of these. Therefore, it is necessary to investigate the effect of meteorological variation on ozone production in assessing ozone trends.

Various statistical methods for ozone meteorological adjustment have been used in the literature over the last decade. Thompson et al. (2001) broadly categorize these methods into three groups as: regression-based modeling, extreme value approaches, and space–time models. Since a thorough review of these methods has already been provided by Thompson et al. (2001), we focus our discussion on the techniques used in this work, which are the generalized additive model (GAM) and the dynamic linear model (DLM).

Both the GAM and the DLM fall into the category of regression-based methods. Regression-based approaches are probably the most commonly used in the literature for the meteorological adjustment of ozone, since this category incorporates a large variety of techniques which can estimate average ozone levels and account for the effects of other variables (Thompson et al., 2001). In addition to the GAM and the DLM, which we discuss in greater detail below, examples of such approaches used to investigate ozone trends include the multiple linear regression strategy of Fiore et al. (1998) and the non-linear regression approach of Bloomfield et al. (1996).

As pointed out by Thompson et al. (2001) and others, the linear relationships between ozone and associated meteorological variables assumed by many regression-based techniques may not capture the true nature of these relationships. The major advantage associated with GAMs is the ability to allow for more complicated, nonlinear relationships between the response variable, in this case ozone, and an explanatory variable, such as temperature or relative humidity, than would be the case with a simpler technique such as multiple linear regression. GAMs have been used in a wide variety of applications, but a primary reference point is the work of Hastie and Tibshirani (1990). Davis et al. (1998) analyzed 11 ozone monitoring sites for Houston, Texas, using singular value decomposition and clustering to define the meteorological regimes and GAMs to develop separately the relationship between ozone and meteorological variables for each. GAMs were also used by Davis and Speckman (1999) to make next-day predictions of ozone levels in the Houston area.

DLMs are commonly used for Bayesian time series analysis and forecasting and are described in great detail by West and Harrison (1997). These models are distinguished by the inclusion of parameters which are free to evolve through time, and so may be better able to represent the complex physical and chemical processes of ozone production and transport. The structure of these models is more flexible than many other regression-based methods, and components can be introduced to capture cyclical variation, autocorrelation, and many other elements. In addition, West and Harrison (1997, Chapter 1) argue that the Bayesian DLM framework better reflects the process by which we learn about a phenomenon as data is collected over time and provides a more coherent strategy for quantifying uncertainty. However, this flexibility is associated with additional burdens on the user, such as the need for customized computer code (DLMs are not available in standard statistical software packages), increased computational time, and greater complexity in interpretation. Perhaps for this reason, the environmental literature includes fewer applications of DLMs than many of the other regression-based approaches. Sansó and Guenni (1999) demonstrated how a stochastic model based on the DLM could be successfully used to model tropical rainfall. Shaddick and Wakefield (2002) developed a hierarchical model based on DLM methodology, which they use to model multivariate pollutant data at several sites simultaneously. A growing area of application is the use of DLMs for more complex spatiotemporal modeling. One relevant example can be found in the work of Huerta et al. (2004), who used DLMs in their analysis of ozone levels at 19 monitoring stations in Mexico City.

Studies have shown that ozone and its precursors can travel long distances across political boundaries and affect air quality far from the source of the pollution (EPA, 2005b). Because the production of ozone may involve a large geographical area, it is difficult to evaluate the roles of national or regional NO_x emission control programs in ozone level changes. In addition, ozone concentrations may be influenced by other factors, such as weather conditions or economic trends. Although this work does not directly address the issue of transport, statistical methods are used to group sites into regions possessing similar ozone patterns and to select sites for further examination which are representative of the various groups. Meteorological variables which have a strong statistical association with ozone concentrations are identified and used to adjust ozone concentrations, so that the levels from year to year can be assessed more rigorously. However, it is important to note that given the number the factors influencing ozone production, no statistical method can definitively diagnose a cause-and-effect relationship between the implementation of NO_x reduction programs and ozone concentrations.

The purposes of this study are: (1) to demonstrate how the GAM and DLM approaches can be applied to analyze ozone trends during ozone seasons 1997–2004; (2) to compare the results from DLM and GAM approaches to determine the extent to which the selection of statistical methods will affect the estimation of ozone trends; and (3) to assess changes in ozone concentrations at representative sites for the period during which controls were introduced. Specifically, the study investigates ozone trends for the states covered by the EPA's NO_x SIP Call. The EPA NO_x SIP Call, EPA's regional ozone transport rule, requires that 19 eastern states and the District of Columbia significantly reduce NO_x emissions by approximately 600,000 tons by the summer of 2004 and by nearly 1 million tons when fully implemented (EPA, 2005a).

2. Data

Hourly average concentrations of ozone and meteorological data were obtained from EPA's Clean Air Status and Trends Network (CASTNet) monitoring stations. Established by EPA in 1987, CASTNet (http://www.epa.gov/castnet) is an important source for data on dry acidic deposition and non-urban, ground level ozone. CASTNet consists of over 80 sites across the US. A primary advantage of using CASTNet data is that it provides the needed hourly meteorological data at the same monitoring sites, thus avoiding the extra variability arising from spatial separation between the ozone and meteorological monitoring sites.

In this study, data provided by CASTNet during the period 1 January 1997–30 November 2004 are utilized. Particular focus is placed on the data collected during ozone seasons 1997–2004, where the ozone season is defined as the period from 1 May to 30 September. For each day, the maximum 8-h ozone was calculated based on EPA's National Ambient Air Quality Standard guidelines (see http://www.epa.gov/ttn/amtic/files/ambient/criteria/reldocs/guidefin.pdf), using hourly average ozone values from all sites located in the eastern United States. The hourly meteorological data were also summarized on a daily basis for these same sites.

The number of missing observations is significant, so the data from these sites were first screened by inspecting the frequency of missing data. Those sites missing more than two-thirds of daily ozone observations over the course of at least one ozone season were removed, as were those sites which had more than one-third of the daily ozone observations missing for two (or more) ozone seasons. However, for those sites included in this study, most have less than 10% of data missing over the ozone seasons. Fig. 1 shows the locations of the 43 eastern CASTNet monitoring sites used in this study.

The data were then screened using regression methods to determine which of the meteorological parameters seemed to be most strongly associated with day-to-day fluctuations in daily maximum 8-h ozone. Table 1 contains a list of the main meteorological variables that were considered in this paper (temperature, dewpoint temperature, relative humidity, solar radiation, wind direction and speed). Each of these was summarized in several different ways, which yielded over 25 different potential explanatory variables for the analysis. For example, the maximum and minimum temperatures and the mean temperature for a specified time period were considered at two different levels above the surface. As another example, the U and V wind components were summarized for each day over

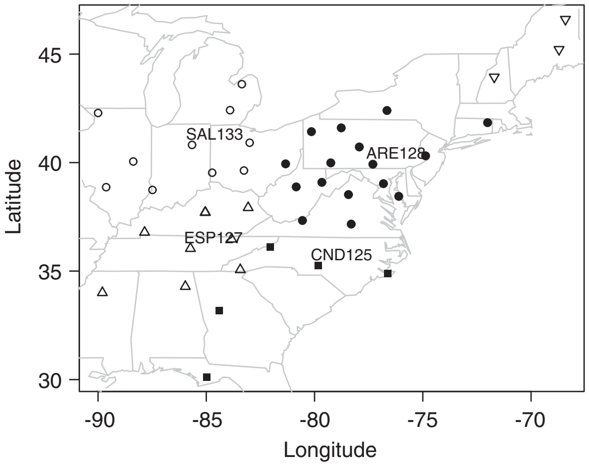

Fig. 1. Ozone influence regimes (denoted by differing symbols) and chosen representative sites (denoted by name).

Table 1
List of meteorological parameters considered

Temperature (°C)
Relative humidity (%)
Solar radiation ($W\,m^{-2}$)
Dew point (°C)
Wind speed ($m\,s^{-1}$)
U, V wind components ($m\,s^{-1}$)

time periods that reflected the diurnal variation in the structure of the planetary boundary layer. Other summaries for wind speed and direction were also examined. The screening process involved stepwise regression with inspection of the residuals versus each of the candidate meteorological variables listed in Table 1. Based on this process, it was generally found that of the available meteorological parameters, the two most strongly correlated with daily maximum 8-h ozone were daily maximum temperature and average daytime relative humidity. For many sites, the contribution from solar radiation (after controlling for temperature and humidity) was also important, though to a lesser extent. None of the wind-related variables in Table 1 displayed a strong association with daily maximum 8-h ozone.

It should be noted that many of the meteorological variables are highly correlated with one another, complicating the choice of variables. The inclusion of explanatory variables which are very closely related (e.g. multicollinearity) can sometimes lead to misleading error estimates or numerical problems with the regression procedure. Based on extensive exploratory data analysis, the results of the stepwise regression procedure, the combined strength of their statistical association with ozone levels, and the principle of parsimony, we selected daily maximum temperature, average mid-day relative humidity, and total solar radiation as explanatory variables for the statistical models used in this work. Though the retention of a relatively small number of meteorological variables may seem somewhat surprising, these choices are not out of keeping with other analyses in the literature. A summary of approaches to the meteorological adjustment of ozone levels provided by Thompson et al. (2001) shows that relevant meteorological variables differ greatly between applications, but that measures of temperature, humidity, and radiation are some of the more commonly utilized variables.

3. Approach

3.1. Principal components analysis

Principal components analysis (PCA) is a quantitative method which has been widely used to identify regions of homogeneity using observations of ozone (or other atmospheric quantities) taken over time (e.g., Eder et al., 1993, Davis et al., 1998, Lehman et al., 2004). It is specifically used in this work to help identify regions of the eastern United States in which ozone concentrations behave similarly during the 1997–2004 ozone seasons. Because this analysis so closely follows the work of Eder et al. (1993) and Lehman et al. (2004), the reader is urged to consult these references for details of the implementation not described here.

The technique was applied to the 43×43 correlation matrix derived from the set of 1224 daily maximum 8-h ozone concentrations (153 days in each ozone season $\times$ 8 ozone seasons) for each of the 43 sites depicted in Fig. 1. Missing data were imputed by incorporating site and daily effects. The goal of the procedure is to explain most of the variability in the data set with only a small subset of the total 43 orthogonal principal components. In this case, an argument could be made for the retention of either four or five components, which explain approximately 71% or 74% of the variability in ozone concentrations, respectively. As in Lehman et al. (2004), a decision was made to retain five principal components, based on the fact that after the fifth component, there is a substantial reduction in the gain added by the inclusion of each additional component.

As explained by Eder et al. (1993), the varimax rotation was then applied to these five components in order to better segregate and identify groups of sites with similar ozone patterns. To establish these groups, we used the loadings on the rotated principal components, which represent the correlations between the components and the sites. For each site, we noted the component on which it loaded most strongly; for purposes of discussion here, we denote this the "primary" component for the site. All sites with the same primary component were then assigned to the same group, yielding five groups. Fig. 1 displays a map of the CASTNet sites used in this study, in which the different symbols designate the various groups determined on the basis of this procedure.

The last step is to determine, for each group, the site which best represents the ozone season pattern in each region. First, we calculate the communality, which is defined as the proportion of the variance of a site explained by the retained components (Eder et al., 1993). The sites with the largest communalities are those whose variability is best explained by the retained principal components. In addition, for each site, we examined the loading on the primary principal component as a measure of the strength of the site's relationship with the underlying mode of variation common to all sites in its group. To adequately represent ozone behavior within its region, a site should ideally possess both a high communality and a high loading on its primary principal component. In some cases, there are several sites in a group which are similar in these measures; a choice may be made between them based on other factors, such as the prevalence of missing data.

3.2. Generalized additive models

Developed by Hastie and Tibshirani (1990), the GAM provides mechanisms for the modeling of more complex relationships than is found in the more commonly used method of linear regression. As mentioned above, a particular strength of the method is the ease with which the nature of a nonlinear relationship between the response variable and an explanatory variable can be investigated. This is relevant to this study, since previous work has indicated that such nonlinear relationships may exist between some meteorological variables and ozone levels (Davis et al., 1998; Davis and Speckman, 1999).

A GAM can be written in the following form:

$$g(\mu_i) = \beta_0 + f_1(x_{i1}) + \cdots + f_p(x_{ip}), \qquad (1)$$

where β_0 represents the overall mean and $f_j(x_{ij})$ is the value of the smoothing function associated with the ith value of explanatory variable j. Also, $g(\mu_i)$ represents the link function, which specifies the relationship between the linear formulation on the right side of Eq. (1) and the expected response. The concept of the "link" function was originally developed as a part of the generalized linear model methodology; for more information, see McCullagh and Nelder (1989). Depending on its formulation, a GAM is estimated using some combination of least squares, smoothing, back fitting, and other methods. These algorithms are included in many statistical software packages and can be employed efficiently, even with large datasets.

In this study, the GAM framework was used to model daily maximum 8-h ozone concentration as a function of the explanatory variables using a log link function (McCullagh and Nelder, 1989). A natural spline of the predictor variables was employed to allow for a nonlinear response between the meteorological parameters and daily ozone. In addition to three meteorological parameters (daily maximum temperature, mid-day relative humidity, and solar radiation), the explanatory variables included a term to account for seasonal effects (day of year) plus an additive factor associated with year to account for trends over time.

The predicted response for the year factor represents the expected seasonal average ozone after adjusting for the meteorological effects. Ninety-five percent confidence limits for this factor can also be readily obtained using various statistical software packages. Standard errors of the difference in adjusted ozone between 2 years (Tables 2 and 3) were estimated using the non-parametric bootstrap (Efron and Tibshirani, 1993).

3.3. Dynamic linear models

One particularly large and flexible class of statistical models for time series is that of DLMs. DLMs incorporate both the traditional elements of time series analysis, such as autoregressive and moving average components, as well as a dynamic regression strategy, into the framework of Bayesian analysis. This has several advantages, such as providing better estimates of uncertainty and allowing for missing data with greater ease.

Table 2
Comparison of meteorologically adjusted ozone differences for selected representative sites using both approaches

	Approach	Site name							
		Salamonie Reservoir, IN (SAL133)		Evins State Park, TN (ESP127)		Arendtsville, PA (ARE128)		Candor, NC (CND125)	
		DLM	GAM	DLM	GAM	DLM	GAM	DLM	GAM
Estimated difference (standard error)	02–04	−4.4 (1.3)	−6.8 (1.3)	−2.9 (1.5)	2.1 (1.4)	−6.9 (1.3)	−9.7 (1.6)	−5.3 (1.6)	−3.2 (1.6)
	97–02	−3.7 (1.2)	−2.7 (1.3)	−5.5 (1.5)	−3.9 (1.4)	1.4 (1.2)	2.7 (1.8)	−7.9 (1.7)	−5.3 (1.5)
	97–04	−8.1 (1.3)	−9.5 (1.1)	−8.5 (1.5)	−1.8 (1.6)	−5.5 (1.3)	−7.0 (1.5)	−13.1 (1.7)	−8.6 (1.7)

Table 3
Comparison of ozone differences without meteorological adjustment for selected representative sites using both approaches

	Approach	Site name							
		Salamonie Reservoir, IN (SAL133)		Evins State Park, TN (ESP127)		Arendtsville, PA (ARE128)		Candor, NC (CND125)	
		DLM	GAM	DLM	GAM	DLM	GAM	DLM	GAM
Estimated difference (standard error)	02–04	−10.6 (1.9)	−10.3 (2.2)	−9.1 (2.2)	−9.8 (2.1)	−11.6 (2.3)	−13.5 (2.5)	−10.8 (2.7)	−8.6 (2.9)
	97–02	3.4 (2.1)	5.1 (2.1)	−2.3 (2.1)	−1.6 (2.5)	4.4 (2.3)	7.8 (2.7)	−6.8 (2.6)	−7.1 (3.0)
	97–04	−7.3 (2.0)	−5.3 (2.0)	−11.4 (2.3)	−11.4 (2.0)	−7.2 (2.3)	−5.6 (2.2)	−17.7 (2.6)	−15.8 (2.5)

Extensive development and description of the entire class of DLMs can be found in the work of West and Harrison (1997).

DLMs are typically specified according to a hierarchical structure of two levels; using the notation of West and Harrison (1997), these can be written as follows:

Observation equation :

$$Y_t = F_t^{\mathrm{T}}\theta_t + v_t, \quad v_t \sim N[0, V_t], \tag{2}$$

System equation :

$$\theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t]. \tag{3}$$

For any day t, the observation equation in Eq. (2) describes the linear relationship between the 8-h ozone maximum (Y_t), explanatory variables (F_t), the time-evolving parameters (θ_t), and error (v_t). It should be noted that the DLM was applied to annual data, and not solely to daily data observed during ozone season. This makes more efficient use of the model's ability to allow parameters to evolve over time. In this study, the explanatory variables included in the DLMs were the meteorological variables temperature, the square of temperature, relative humidity, and solar radiation. Parameters express the relationships between the explanatory and response variables, but also include overall trend, cyclical components, and the correlation between the response values on sequential time days. The system equation, as shown in Eq. (3), describes the linear evolution of the parameters (θ_t) from day to day through an evolution matrix (G_t) and independent, normally distributed innovations (ω_t). Specifying a DLM, therefore, requires the specification of both levels of the system.

The exact formulation of the DLM will differ in every application, but the substantial features remain the same for each of the selected sites. The user first identifies the meteorological variables that are most closely associated with ozone levels; the selection of maximum daily temperature, daily average relative humidity, and daily total solar radiation was discussed previously in Section 2. Since the effect of meteorology is assumed to be the same across the years, the system equation specifies the corresponding parameters, say, θ_1, θ_2, and θ_3, to remain constant over time. For instance, this means that the increase in ozone associated with a rise in daily average relative humidity of one percentage point is the same for all years included in the analysis. It does not imply that the relative humidity is assumed to be the same across time.

In addition to accounting for the effect of meteorological conditions on ozone trends, preliminary analyses indicated that long-term trend, seasonal (or cyclical) patterns, and autoregressive components should also be included in the DLM parameterization. A first-order polynomial structure (West and Harrison, 1997, Chapter 2) is used to model the long-term trend component; this structure is based upon the concept of a slowly changing mean baseline value. The reduced Fourier form components representation described by West and Harrison (1997, Section 8.6.6) is used to describe the cyclical patterns. After a detailed exploration of the autocorrelated errors at each site, an AR(1) component is incorporated into the DLM using the formulation given by West and Harrison (1997, Section 9.4.7). The use of this component to account for the serial correlation in the errors negates the necessity of allowing for independent observational errors in Eq. (2), so we set v_t to zero.

After the DLM is formulated, Markov chain Monte Carlo techniques are used to sample from the joint distribution of the parameters, given the actual observations. West (1995) discusses this procedure in detail in the context of a DLM with cyclical components, similar to that used in this work. The simulations provide estimates for the effect of each meteorological component, as well as estimates of the states of the long-term, seasonal, and autoregressive processes on each day. The simulated values also provide an easy way to quantify the variability surrounding these estimates, both individually and as a group. Average ozone levels for each of the ozone seasons considered in the study can then be estimated using the simulated daily values over the entire summer period, with corresponding uncertainty estimates.

A complete explanation of the procedures used to estimate the variance associated with the error in the observation equation, as well as choices concerning the choice of the innovation variance W_t in the system equation, is beyond the scope of this paper. Further details about Bayesian modeling and simulation can be found in West (1995), in West and Harrison (1997), and in many other papers in the time series literature.

4. Results

4.1. Ozone groups and representative monitoring sites

The procedure discussed in Section 3.1 yielded five ozone groups, each making up a spatially coherent region within the eastern portion of the country. Fig. 1 shows each of the CASTNet sites used in this analysis; sites belonging to the various groups are denoted by the various symbols. The representative sites are further denoted with their CASTNet identifiers. The spatial groupings appear to be dependent on regional meteorological and/or emissions factors with a smooth regional pattern. This is probably because ozone is a secondary pollutant, so that its production, as well as subsequent transport and scavenging, depend greatly on meteorological conditions (Eder et al., 1994). Interestingly, the grouping resembles that noted by Eder et al. (1993), though there are some differences in the southeastern region that are probably attributable to the greater availability in that study of sites in the deep south to more closely delineate these regions.

The first group is comprised of 16 sites in the mid-Atlantic region, marked with solid circles in Fig. 1. The site ARE128, at Arendtsville, PA, is chosen as a representative of this group. Nine monitoring stations (including one co-located pair) fall into the second group and are denoted in the figure by triangles; these are located in the southwestern portion of the region. Site ESP127, in Evins State Park, TN, serves as the main representative of sites in this group. Solid squares mark the locations of a group of five sites in the southeastern portion of Fig. 1, and site CND125 (in Candor, NC) is chosen to represent this group. Marked with open circles in Fig. 1, group three contains ten sites in the midwestern region. For this group, we focus on site SAL133 at Salamonie Reservoir, IN. The three northeasternmost sites comprise the last group and are represented with inverted triangles. This group is not examined further in this work, due to the fact that none of these sites lie in states covered by the NO_x SIP Call and also due to the small size of the group. However, ozone levels in this region are primarily the result of atmospheric transport processes that occur in advance of an eastward moving frontal system. For an in-depth discussion of ozone characteristics in this and other regions in the eastern United States, see Eder et al. (1993, 1994).

4.2. Overall ozone trends for representative sites

Figs. 2 and 3 show raw 8-h maximum ozone time series data from 1997 to 2004 for sites SAL133 and ARE128, respectively. Strong seasonal patterns are evident in both, along with some decreases in ozone concentrations toward the end of this 8-year period. Similar features are found at most of the other locations.

Using the methods presented in Sections 3.2 and 3.3, we calculate estimated ozone seasonal average concentrations over the years 1997–2004, along with the associated uncertainties. Figs. 4 and 5 show the interannual variability in the average meteorologically adjusted summer ozone and the corresponding estimated confidence and credible intervals from both approaches for the four representative sites. In these figures, the circles represent estimated average concentrations during ozone season, and the dashed lines show the 95% confidence intervals (GAM) or 95% credible intervals (DLM) surrounding these estimates. As shown in Figs. 4 and 5, overall downward trends for these average levels are observed for all sites. Specifically, for each site, the results from the DLM and GAM approaches show similar trend behaviors, even for year-to-year fluctuations. For example, at site ESP127 in Tennessee, the results from both approaches show that there is a notable increase in average ozone levels from 1998 to 1999, and that the levels resume a downward trend the following year. The estimated average level in ozone season 2004 for ESP127 is also noteworthy, since a quick glance at Fig. 4 may seem to indicate that the models disagree about the estimated levels. This difference is not significant, since the confidence bounds on these sets of estimates overlap. However, it is clear that the estimated levels at ESP127 in the period from 2002 to 2004 are not dropping as fast as at the other sites, and may even have stabilized.

Both the GAM and DLM approaches produced similar estimates regarding changes in ozone trends as well as the associated uncertainty both before and

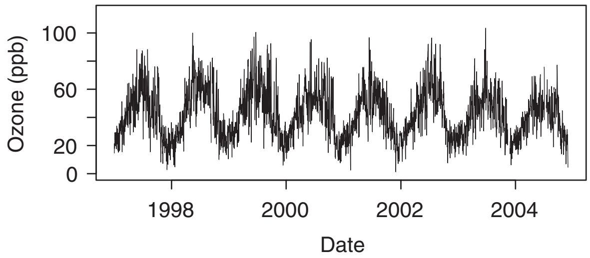

Fig. 2. Raw time series for site SAL133 (Salamonie Reservoir, IN).

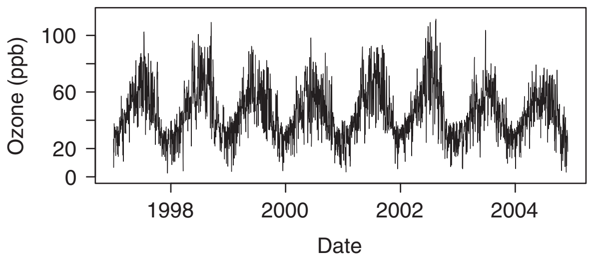

Fig. 3. Raw time series for site ARE128 (Arendtsville, PA).

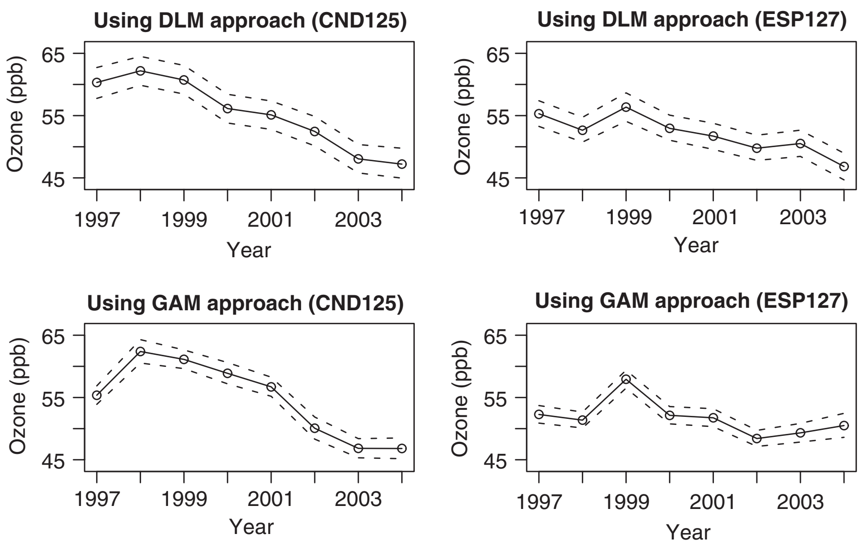

Fig. 4. Meteorologically adjusted average concentrations during ozone seasons using DLM and GAM approaches for sites CND125 (Candor, NC) and ESP127 (Evins State Park, TN).

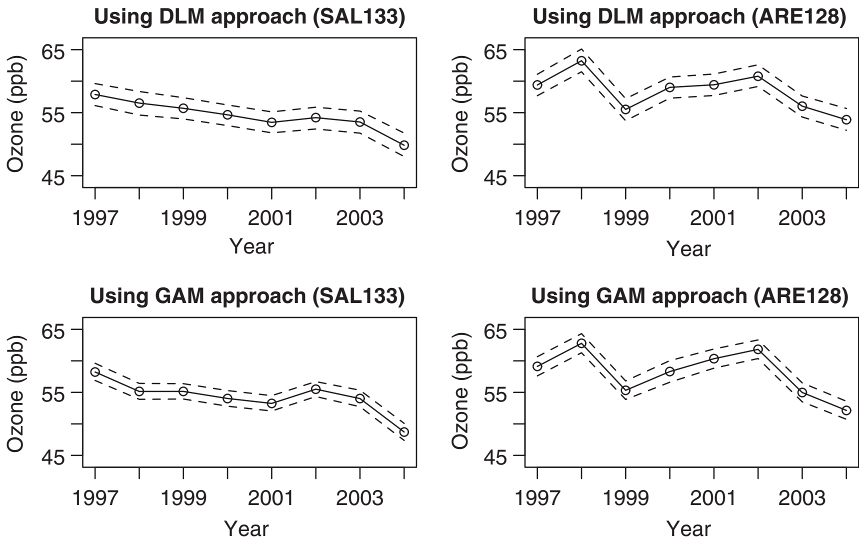

Fig. 5. Meteorologically adjusted average concentrations during ozone seasons using DLM and GAM approaches for sites Salamonie Reservoir, IN (SAL133) and Arendtsville, PA (ARE128).

after adjusting for meteorological conditions, as shown in Tables 2 and 3. For example, at site SAL133 in Indiana, the estimated meteorologically adjusted ozone decrease between 1997 and 2004 using the DLM and GAM approaches are 8.1 and 9.5 ppb, respectively, while the estimates are 7.3 and 5.3 ppb without meteorological adjustments using the two approaches for the same period. Though this discussion has focused on four sites chosen to represent the regions identified through PCA, similar changes in ozone levels from the two approaches are observed at other sites.

4.3. Roles of emission control programs in ozone level reductions

To assess whether or not emission control programs are associated with reductions in ozone levels, we selected 1997 and 2002 as two baseline years for analyzing ozone level changes during the 1997–2004 period. The time period from 1997 to 2002 saw the implementation of a number of programs to reduce emission of VOCs and NO_x; for example, these include the OTC NO_x Budget Program and the Acid Rain NO_x Reduction Program. After 2002, emissions from power generation began to be reduced to comply with the NO_x SIP Call (EPA, 2005a).

A comparison of Tables 2 and 3 shows that, as expected, meteorological conditions have important effects on estimated changes in ozone levels. For example, for site SAL133, estimates made without meteorological adjustments (Table 3) show an increase in average ozone on the order of 4 ppb (averaging the estimates yielded by the two methods) for the time period 1997–2002. However, decreases in ozone concentrations (approximately 3 ppb) after meteorological adjustment (Table 2) are observed for the same site during this period. Fig. 6 summarizes the ozone season (from May to September) daily maximum 1-h temperature for site SAL133. The boxplots for years 1997 and 2002 show that summer temperatures in 2002 were noticeably higher than those observed in 1997, both in terms of the median temperatures and in terms of the middle 50% of temperatures in those years. Similar patterns were observed at other sites in the eastern United States, which are not shown here. These results imply that the observed increases in ozone levels (Table 3) may be correlated with variations in meteorological conditions. In order to evaluate the effectiveness of emission control programs, ozone trend analysis results after meteorological adjustment should be used.

The decreases in ozone levels after meteorological adjustment imply that the implementations of emission control programs are associated with declining ozone levels. After the implementation of NO_x emissions controls as part of the OTC NO_x Budget Program (1999–2002) and the NO_x SIP Call (2002–2004), decreases in meteorologically adjusted ozone levels on the order of 2–10 ppb are produced

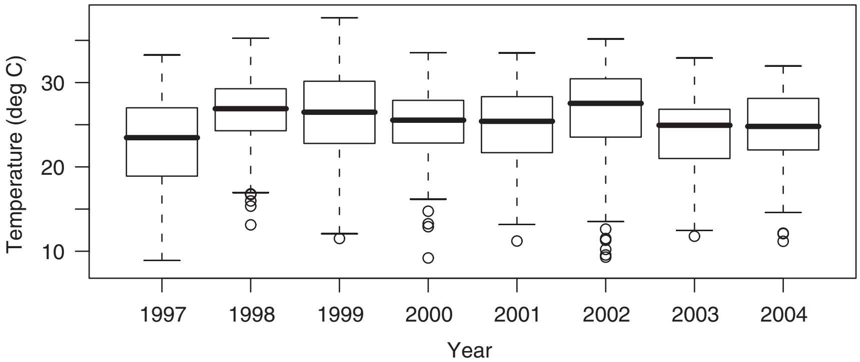

Fig. 6. Summary of daily maximum 1-h temperature during ozone seasons 1997–2004 for site SAL133 (Salamonie Reservoir, IN).

at many sites (not just those featured as representative sites in this paper) in the period from 1997 to 2004. This analysis indicates that the imposition of NO_x controls is likely an important factor in ozone reduction.

4.4. Cyclical properties of ozone trends

One advantage of the DLM is its ability to isolate various components of the variability, which together with random error make up the raw time series observed at a site. When the DLMs are utilized for the selected CASTNet sites, it is easy to detect the presence of cyclical patterns, on an annual and semi-annual basis, which together have amplitudes of about 10–20 ppb. As examples, Figs. 7 and 8 show these estimated cyclical patterns in black for sites SAL133 (in Indiana) and ARE128 (in Pennsylvania), respectively. The gray lines represent 95% credible intervals and are included to provide a sense of uncertainty levels surrounding these estimates.

Even after the effects of temperature, relative humidity, and solar radiation are statistically removed, there is still an annual peak in ozone around 1 May, though of course, in some years it occurs slightly earlier (end of April) or slightly later (early June). This is true not only for both sites pictured in Figs. 7 and 8, but also for all of the other sites that we inspected. Also, a smaller peak often appears in August or September, though this is not nearly as pronounced, nor is it observed consistently over the 8-year period. It is interesting to note that this secondary rise seems more dominant in the earlier years, and becomes less notable in the later years. A large dip occurs yearly in December or January; again, this is similar from site to site and year to year.

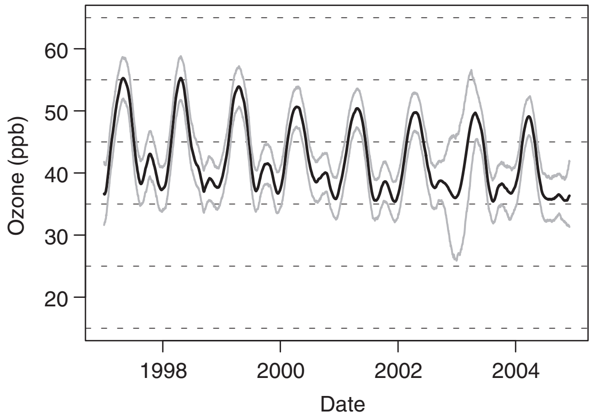

Fig. 7. Estimated cyclical pattern at site SAL133 (Salamonie Reservoir, IN).

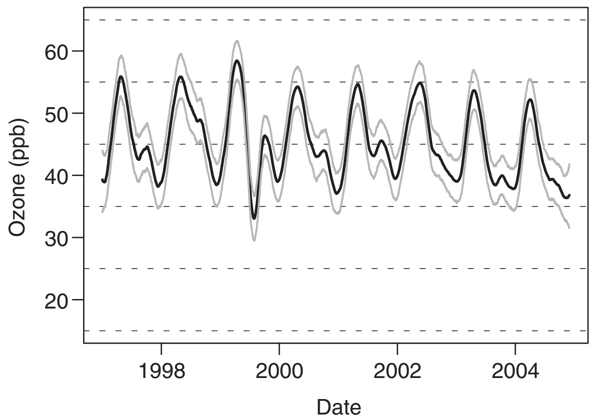

Fig. 8. Estimated cyclical pattern at site ARE128 (Arendtsville, PA).

Since these cyclical patterns are present even after the effects of temperature, relative humidity, and solar radiation are removed, it is unclear exactly what their causes are. It may be the case that the addition of other relevant meteorological variables to the DLM would be sufficient to explain most of the remaining seasonality in the time series. Unfortunately, the other meteorological information available to us at the CASTNet sites has not proved helpful in this regard, so this quantity may be one that is not currently being monitored (e.g., mixing depth) through this network. Other possible causes for these peaks include the springtime emission of biogenic VOCs and the stratospheric intrusion of ozone, and the late summer occurrence of stagnating weather conditions associated with the occurrence of quasi-stationary anticyclones. The weather conditions associated with these stagnating systems are very conducive to the formation of ozone (Eder et al., 1993). See Wayne (2000, Chapter 5) for a general discussion of the role that biogenic VOCs play in atmospheric gas-phase reactions. Although it is clear that these cyclical patterns occur at many sites, the causal agent is still under investigation.

5. Discussion

Homogeneous ozone groups are identified for 43 CASTNet sites in the eastern United States using PCA. The results of ozone trend analyses show that both GAM and DLM approaches produce very similar estimates of ozone changes. Obvious overall downward trends from 1997 to 2004 are observed for these sites after meteorological impacts are removed. The decreases in ozone levels appear to be associated with NO_x emission reductions due to implementation of NO_x control programs. These results may imply that NO_x control programs have played a role in reducing ground ozone levels.

Application of the GAM and DLM approaches to the trend analysis of ozone and their implementations are evaluated. The main advantage of the GAM approach is its relative ease of implementation. Given familiarity with related models, such as generalized linear models or multiple regression, an analyst can formulate and fit a GAM model using any number of statistical software packages, though producing and interpreting confidence bounds and uncertainty estimates still requires considerably more experience. Due to their simpler formulation, the algorithms required are much less time consuming and resource intensive.

The DLM approach is easily extended and therefore much more flexible than most other methods. Because the DLM approach can incorporate both traditional elements of time series analysis and a dynamic regression strategy into a Bayesian framework, it can better describe the complex process of ozone changes over years, provide better error estimates, and deal with missing data with greater ease. Furthermore, once the simulations are obtained based on the fitted DLM, the final results, season averages and credible intervals, can be calculated based on these simulated values. However, the DLM approach is substantially more difficult to implement compared to the GAM approach, partially because routines for estimating these models are not available in most statistical software packages. Beginners unfamiliar with Bayesian statistics or time series analysis can expect to pay the price of a steep learning curve.

Although both approaches can be applied to estimate the changes in average interannual ozone levels, both with and without meteorological adjustments, the selection of an approach depends on the purpose and depth of the planned analysis. In a situation in which estimates are needed quickly, but a sophisticated investigation into the components of the time series is not needed, GAM techniques are the clear choice. However, for more detailed analyses, in which an analyst needs to examine individually modeled components of the time series, the DLM is a superior tool. For instance, as described in Section 4.4, DLMs are able to easily pick up on cyclical patterns in the data which are much harder to uncover with the GAM approach used here. In addition, the DLM approach is able to take into account correlated errors and changing trend directions. Other traditional time series techniques can also be included into the DLM approach, including moving average and higher-order autoregressive components. This study also points to further research opportunities regarding ozone trend analysis and the effectiveness of emission control programs. For example, it may be possible to incorporate emission information (e.g., NO_x and VOCs) into models in order to explicitly investigate the relationship between decreases in ozone concentrations and emission reductions. Also, since the production and transportation of ozone exhibit obvious regional properties, a spatial-temporal model would likely capture these dependencies.

Acknowledgments

The authors would like to thank Brian Eder for many helpful scientific discussions and Steve Howard for his help in collating and managing the CASTNet data. We also gratefully acknowledge the comments and suggestions of several reviewers.

Disclaimer. The research presented here was performed under the Memorandum of Understanding between the US Environmental Protection Agency (EPA) and the US Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

References

Bloomfield, P.J., Royle, J.A., Steinberg, L.J., Yang, Q., 1996. Accounting for meteorological effects in measuring urban ozone levels and trends. Atmospheric Environment 30 (17), 3067–3077.

Davis, J.M., Speckman, P., 1999. A model for predicting maximum and 8 h average ozone in Houston. Atmospheric Environment 33, 2487–2500.

Davis, J.M., Eder, B.K., Nychka, D., Yang, Q., 1998. Modeling the effect of meteorology on ozone in Houston using cluster analysis and generalized additive models. Atmospheric Environment 32, 2055–2520.

Eder, B.K., Davis, J.M., Bloomfield, P.J., 1993. A characterization of the spatiotemporal variability of non-urban ozone concentrations over the eastern United States. Atmospheric Environment 27A, 2645–2668.

Eder, B.K., Davis, J.M., Bloomfield, P.J., 1994. An automated classification scheme designed to better elucidate the dependence of ozone meteorology. Journal of the Applied Meteorology 33, 1182–1199.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

EPA (US Environmental Protection Agency), 2005a. Evaluating ozone control programs in the eastern United States, EPA454-K-05-001, Washington, DC.

EPA (US Environmental Protection Agency), 2005b. Technical Support Document for the Final Clean Air Interstate Rule: Air Quality Modeling. Research Triangle Park, NC. 〈http://www.epa.gov/air/interstateairquality/pdfs/finaltech02.pdf〉

Fiore, A.M., Jacob, D.J., Logan, J.A., Yin, J.H., 1998. Long-term trends in ground level ozone over the contiguous United States, 1980–1995. Journal of Geophysical Research 103, 1471–1480.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, New York.

Huerta, G., Sansó, B., Stroud, J.R., 2004. A spatiotemporal model for Mexico City ozone levels. Applied Statistics 53 (2), 231–248.

Lehman, J., Swinton, K., Bortnick, S., Hamilton, C., Baldridge, E., Eder, B., Cox, B., 2004. Spatio-temporal characterization of tropospheric ozone across the eastern United States. Atmospheric Environment 38, 4357–4369.

McCullagh, P., Nelder, J.A., 1989. Generalized additive models. CRC, Boca Raton, FL.

NRC, 2004. Air Quality Management in the United States. National Academy Press, Washington, DC.

Sansó, B., Guenni, L., 1999. A stochastic model for tropical rainfall at a single location. Journal of Hydrology 214, 64–73.

Shaddick, G., Wakefield, J., 2002. Modeling daily multivariate pollutant data at multiple sites. Applied Statistics 51 (Part 3), 351–372.

Thompson, M.L., Reynolds, J., Cox, L.H., Guttorp, P., Sampson, P.D., 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. Atmospheric Environment 35, 617–630.

Wayne, R.P., 2000. Chemistry of Atmospheres. Oxford University Press, New York.

West, M., 1995. Bayesian interference in cyclical component dynamic linear models. Journal of the American Statistical Association 90, 1301–1312.

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, second ed. Springer, New York.