Space and Time Scales in Ambient Ozone Data

S. T. Rao,* I. G. Zurbenko,+ R. Neagu,+ P. S. Porter,# J. Y. Ku,@ and R. F. Henry@

ABSTRACT

This paper describes the characteristic space and time scales in time series of ambient ozone data. The authors discuss the need and a methodology for cleanly separating the various scales of motion embedded in ozone time series data, namely, short-term (weather related) variations, seasonal (solar induced) variations, and long-term (climate–policy related) trends, in order to provide a better understanding of the underlying physical processes that affect ambient ozone levels. Spatial and temporal information in ozone time series data, obscure prior to separation, is clearly displayed by simple laws afterward. In addition, process changes due to policy or climate changes may be very small and invisible unless they are separated from weather and seasonality. Successful analysis of the ozone problem, therefore, requires a careful separation of seasonal and synoptic components.

The authors show that baseline ozone retains global information on the scale of more than 2 months in time and about 300 km in space. The short-term ozone component, attributable to short-term weather and precursor emission fluctuations, is highly correlated in space, retaining 50% of the short-term information at distances ranging from 350 to 400 km; in time, short-term ozone resembles a Markov process with 1-day lag correlations ranging from 0.2 to 0.5. The correlation structure of short-term ozone permits highly accurate predictions of ozone concentrations up to distances of about 600 km from a given monitor. These results clearly demonstrate that ozone is a regional-scale problem.

1. Introduction

To discern whether a physical process is dynamically important in any particular situation, meteorologists introduce *scales of motion*. The presence of various scales of motion in time series of meteorological and air quality variables can complicate analysis and interpretation of data. Separation of time series of ozone and meteorological data into synoptic, seasonal, and long-term components is necessary since the processes occurring at different frequencies are caused by different physical phenomena: the synoptic-scale component is attributable to weather and short-term fluctuations in precursor emissions, seasonal scale to variation in the solar angle, and long-term scale to changes in climate, policy, and/or economics (Rao and Zurbenko 1994; Rao et al. 1995; Porter et al. 1996). Spatial and temporal information in ozone and meteorological data, obscure prior to separation, may become clear afterward.

Episodes of exceedances of ozone standards may be viewed as particular realizations of the ozone process. Whereas each episode is a unique event with little predictability, the ozone process in total has well-defined spatial and temporal scales that permit accurate prediction of average conditions and probablistic statements about ozone concentrations at particular times and places. Separation of the temporal scales in ozone is also important when assessing changes due to pollution control actions or climate changes. Such changes may be quite small and invisible unless they are separated from weather and seasonality. An addi-

*Department of Earth and Atmospheric Sciences, University at Albany, Albany, New York.
+Department of Biometry and Statistics, University at Albany, Albany, New York.
#Department of Civil Engineering, University of Idaho, Idaho Falls, Idaho.
@New York State Department of Environmental Conservation, Albany, New York.
Corresponding author address: Dr. S. T. Rao, Office of Science and Technology, NYS Dept. of Environmental Conservation, 50 Wolf Rd., Albany, NY 12233-3529.
E-mail: strao@dec.state.ny.us
In final form 23 May 1997.

tional reason for separately analyzing temporal components of ozone is that short-term phenomena contain information related to the transport of ozone and conditions conducive to ozone accumulation. Seasonal components, on the other hand, lend little insight to pollutant transport issues.

The purpose of this paper is to illustrate the temporal and spatial information available in spectrally decomposed ozone and temperature data. The Kolmogorov–Zurbenko (KZ) filter (Zurbenko 1986) is used to separate data into short-term, seasonal, and long-term processes. Characteristics of the KZ filter, including parameter choices given a cutoff frequency, and its transfer function are described in detail. This method is used because it provides effective separation of frequencies (Eskridge et al. 1997) and because it does not require special treatment for missing data. The anomaly (perturbation) method was not used because noticeable amounts of energy from all parts of the spectrum are present in every component (Eskridge et al. 1997). The wavelet transform method described by Lau and Weng (1995) has very low leakage between temporal components (Eskridge et al. 1997) but requires special considerations for missing data.

Analysis of 437 sites in the United States for the period 1983–94 indicates that only about 2% of the total ozone variance is in the long-term component, with the remainder divided roughly equally between short-term and seasonal components. The variance of short-term and seasonal components varies considerably across the United States.

Temporal and spatial scales for ozone components are described in terms of the decay of the process correlation in time (serial correlation) or space (correlations between different stations as a function of distance) to a value of $1/e$. Process correlations for temporal and spatial scales follow a Markov process and exponential decay, respectively. The $1/e$ index of scale has also been used by the National Climatic Data Center to design monitoring networks (Wallis 1996). Whereas raw ozone data exhibit very slow decay in serial correlations, serial correlation in the short-term component alone decays in 1 to 3 days to that expected for white noise. The resemblance between short-term ozone and white noise suggests that ozone events can be simulated by superimposing "noise" (computer-generated random numbers) over the baseline (defined as the sum of long-term and seasonal components) (Rao et al. 1996). Similarly, Eskridge et al. (1997) suggest that time series of temperature data can also be represented as the sum of "noise" and baseline components. Reproduction of the synoptic (stochastic) component around the baseline will also re-create exceedances (Rao et al. 1996). Thus, ozone air quality management efforts can be addressed in baseline terms.

The correlations among baseline components of meteorological variables are much stronger than those among their short-term components. Baseline ozone retains global information on the scale of more than 2 months in time and about 300 km in space. Short-term ozone is highly correlated in space, retaining 50% of short-term information at distances ranging from 350 to 400 km. The correlation structure of short-term (weather related) ozone permits accurate predictions of ozone concentrations up to distances of 600 km from a monitor. In contrast to baseline and short-term ozone, the information available in exceedance events is only about 15 km.

2. Spectral decomposition of time series data

a. Database

Hourly concentrations of ozone (in ppb) at all monitoring stations in the United States were extracted from the Environmental Protection Agency's (EPA) Aerometric Information Retrieval System (AIRS) for the 1983–94 period. Also, temperature data (in °F) for the same time period were obtained from the nearest and most representative National Weather Service stations. From this dataset, time series of daily maxima of hourly ozone and temperature were constructed and analyzed here.

b. Conceptual model

Concentrations of atmospheric ozone depend on atmospheric variables and precursor concentrations that have strong seasonal and synoptic components (Rao and Zurbenko 1994). Successful analysis of the ozone problem requires a careful separation of seasonal and synoptic components. Therefore, time series of ozone and temperature data will be represented by

$$O(t) = e(t) + S(t) + W(t); \; T(t) = e'(t) + S'(t) + W'(t), \quad (1)$$

where $O(t)$ is the natural logarithm of the original ozone time series and $T(t)$ is the temperature time series, $e(t)$ is the long-term (trend) component, $S(t)$ is seasonal change, $W(t)$ is short-term variation, and t is time. Whereas, $O(t)$ is log-transformed with a view that

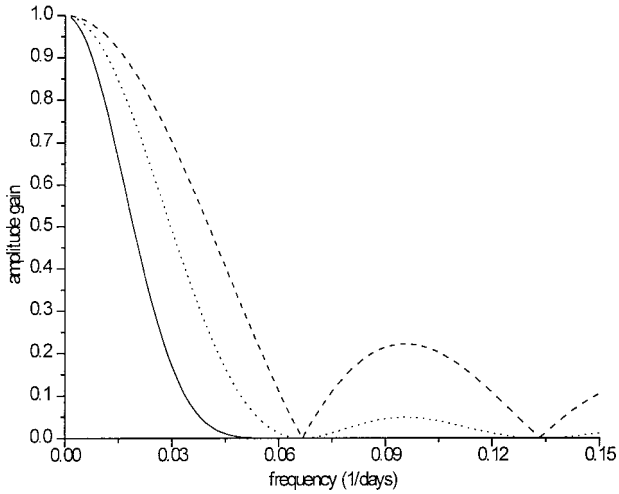

FIG. 1. Absolute value of the transfer function of the KZ filter.

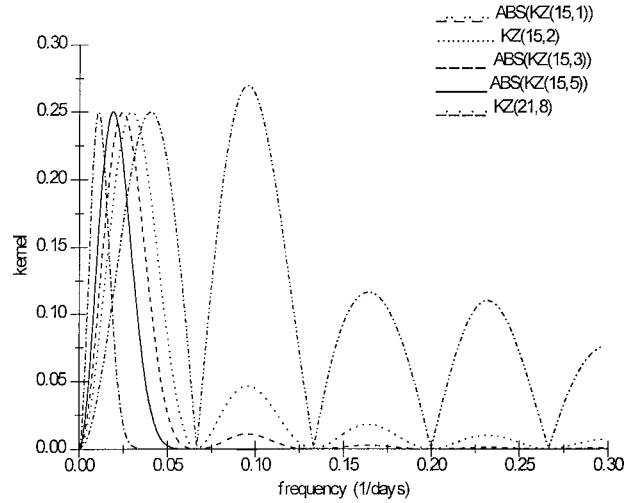

FIG. 2. Kernels (absolute value if k is odd) for several sets of filter parameters.

statistical analysis can be performed successfully, $T(t)$ is not log-transformed since it already follows an additive model; the components of this additive model, $W'(t)$, $S'(t)$, and $e'(t)$ have the same meanings as the ones in the first equation. Long-term, seasonal, and synoptic components are described by completely different physical and chemical processes.

c. Separation techniques

The choice of separation techniques is crucial. Anomaly techniques based on monthly, seasonal, and annual averages are far from adequate because 5% of the energy and 22% of the amplitude of each component is mistakenly attributed to the others (Eskridge et al. 1997). Such poor separation completely destroys the possibility of drawing accurate inferences through anomaly-based models. In this paper, components in the data due to different scales of motion are separated using the KZ filter (Zurbenko 1986, 1991). Among the several high-resolution filters available, the KZ filter is distinguished by its simple algorithm and the preservation of true information when applied in a nonequally spaced and/or missing data environment (Zurbenko et al. 1996; Eskridge et al. 1997).

The KZ(m,k) filter is defined as k applications of a simple moving average of m points. The moving average can be expressed as

$$Y_t = \frac{1}{m} \sum_{s=-(m-1)/2}^{(m-1)/2} X(t+s), \qquad (2)$$

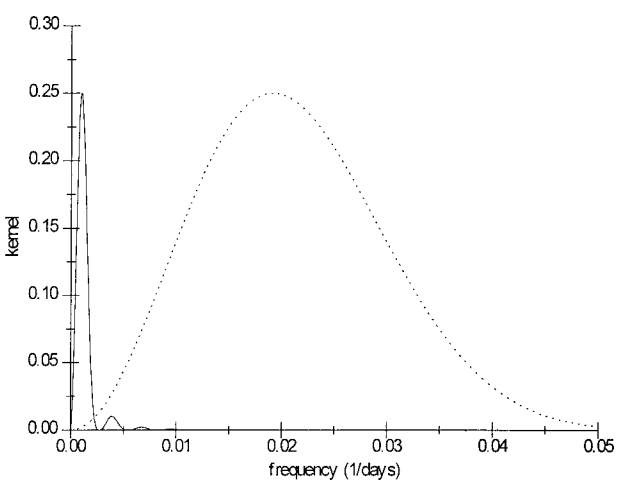

FIG. 3. Same as Fig. 2 for KZ(15,5) and KZ(365,3).

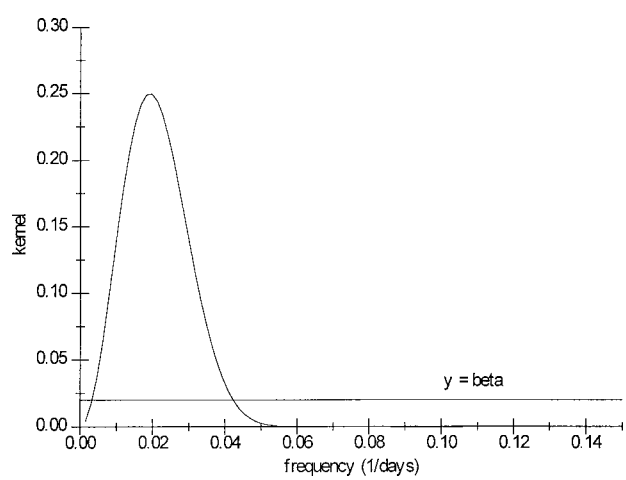

FIG. 4. Same as Fig. 2 for KZ(15,5) and a β-level interval.

TABLE 1. Cutoff frequency for given m, k, and α.

α	m	k	Cutoff frequency
1/2	5	4	0.046405
1/2	11	5	0.018605
1/2	13	3	0.020205
1/2	15	5	0.013619
1/2	11	8	0.014748

TABLE 2. Parameters m and k that give cutoff frequency 0.01492 with precision 0.001.

α	Precision	m	k	Cutoff frequency
1/2	0.001	11	8	0.01474
1/2	0.001	13	5	0.01573
1/2	0.001	15	4	0.01520
1/2	0.001	17	3	0.01543
1/2	0.001	11	7	0.01576

where X is the original time series and t is time (in days). The series Y_t becomes the input for the second pass, and so on. The time series produced by k iterations of the filter described by (2) is denoted $Y_t^{(k)}$.

The square transfer function of the KZ(m,k) [see Eskridge et al. (1997) for details on the transfer function] is given by

$$\left|\phi_{m,k}(\omega)\right|^2 = \left[\frac{1}{m}\frac{\sin(\pi m\omega)}{\sin(\pi\omega)}\right]^{2k}, \qquad (3)$$

where ω has units of cycles per day (frequency). Equation (3) shows, among other things, that the KZ is a low-pass filter (Fig. 1).

The parameter k controls the level of noise suppression. For example, if a value for k is chosen such that the height of the additional peaks in the squared transfer function are to be less than 10^{-5}, the resulting value for k will be ≥ 4 (Fig. 1). Once k is fixed, m is chosen such that

$$\omega_0 \approx \frac{\sqrt{6}}{\pi}\sqrt{\frac{1-(1/2)^{1/2k}}{m^2-(1/2)^{1/2k}}}, \qquad (4)$$

where ω_0 is the desired separating frequency. The right side of (4) is the approximate solution to the equation $|\phi_{m,k}(\omega)|^2 = 1/2$.

Thus far, these values of m and k are only approximations; their accuracy should be checked by solving the equation $|\phi_{m,k}(\omega)|^2 = 1/2$ for ω [which we call the *cutoff frequency of the KZ(m,k) filter*] and assess its proximity to ω_0, the desired cutoff frequency. A computer program that solves the above equations for any values of m and k is available from the authors. The results of the program have been tabulated (Table 1) so that one can easily find the cutoff frequency given (m,k).

Usually the problem appears in reverse: analysis of a periodogram indicates the separating frequency; knowing

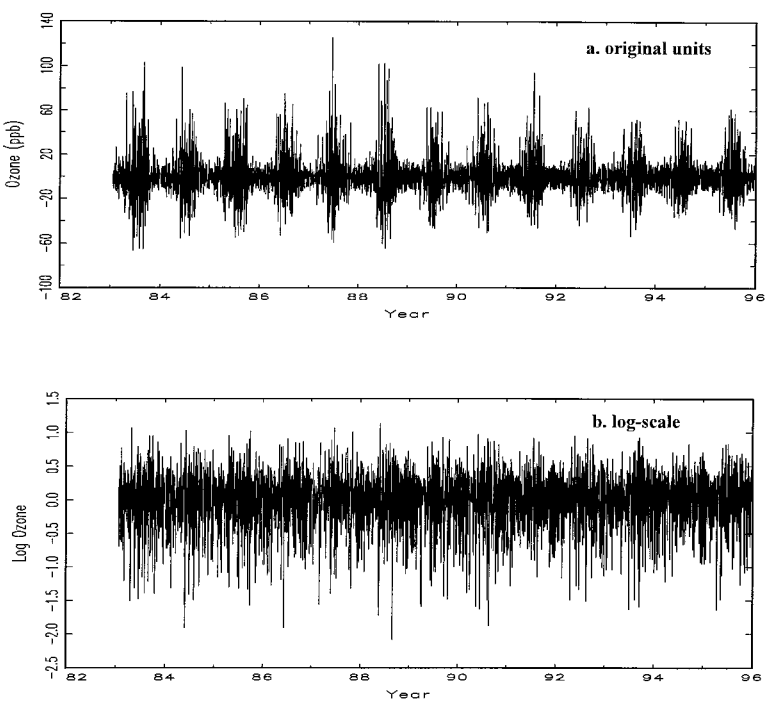

FIG. 5. Comparison of short-term components for original and log ozone at Cliffside Park.

TABLE 3. β-level interval for $\beta = 0.3$ (30%).

m	*k*	Interval	Kernel center
5	4	(0.023174, 0.10683)	0.065002
11	5	(0.009274, 0.04297)	0.026126
13	3	(0.010107, 0.04636)	0.028233
15	5	(0.006786, 0.03146)	0.019124
11	8	(0.007336, 0.03419)	0.020766

this, we have to find the pair (m,k) that will produce a cutoff frequency closest to the separating frequency. A program that takes as input a number (corresponding to the separating frequency) and returns pairs (m,k) that will produce KZ(m,k) filters having cutoff frequencies within a given interval around the input frequency is also available (see Table 2 for an example).

The above analysis was also done for an arbitrary level of reduction (substitute one-half with arbitrary α between 0 and 1) and tables with results for other values of α are available from the authors.

d. Criteria for the effectiveness of separation techniques

The energy of separated processes should be concentrated at different frequencies (spectral domain) and the information in the natural physical processes that cause these energies should be independent of each other. The degree to which this is accomplished is evident in the filter's squared transfer function (gain), which shows the transfer of energy to each component affected by a separation technique (see Eskridge et al. 1997). A good separation technique is characterized by a gain function that concentrates energy at the timescale of interest and does not mix energies from different timescales.

The squared transfer function (gain) gives us information about the transfer of energy to resulting components. It does not tell us whether these energies are mixed. More precisely, we want to have a measure of the correlation between resulting components of the KZ filter. This is important when making inferences about the general process, having reasoned only on a specific component.

Calculation of the covariance between the result of KZ filtration and the residual, [1 – KZ], gives

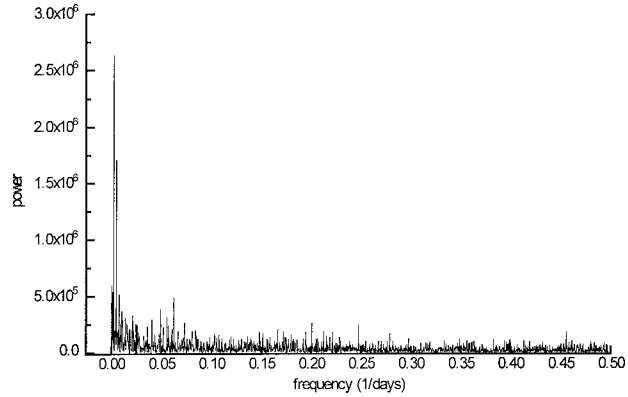

FIG. 6. Raw periodogram of log ozone at Cliffside Park, New Jersey.

$$R = \int f(\omega)\,\phi_{m,k}(\omega)\left[1-\phi_{m,k}(\omega)\right]d\omega. \qquad (5)$$

Therefore, the kernel for the covariance between KZ and [1 – KZ] is given by

$$k(\omega) = \phi_{m,k}(\omega)\left[1-\phi_{m,k}(\omega)\right]. \qquad (6)$$

There is high correlation between filtered data and residuals for small values of k (Fig. 2); also, the width of the kernel gets smaller as m and k increase. The asymptotical convergence of the correlation to 0 is given by

$$O(R) = \left\{\frac{24\pi}{\left[(2m+1)^2-1\right]k}\right\}^{1/2} + o\left(\frac{1}{m\,k^{3/2}}\right). \qquad (7)$$

The frequencies ω_c, at which $k(\omega)$ is concentrated, have been tabulated. The interval around ω_c, inside which the kernel takes values bigger than $\beta/4$, where β is

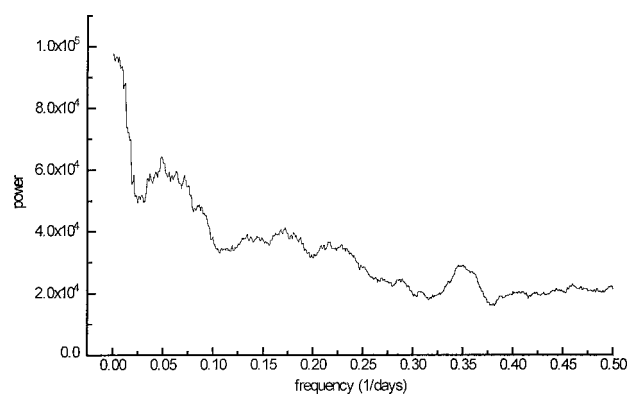

FIG. 7. Same as Fig. 6 except smoothed with Daniell smoother.

an arbitrary number between 0 and 1, have also been tabulated (Table 3). We call this the β-level interval. Approximations for the end points of the β-level interval can easily be calculated using the estimate of the cutoff frequency given above. Figure 3 shows the kernels used to separate high-frequency (short term) and low-frequency (long term) components in ozone (m,k = 15,5 and 365,3, respectively).

What is the meaning of this β-level interval? We choose β (depending on the physics of the process) such that we can assume $k(\omega) = 0$ outside the β-level interval. Now, if the frequency interval we are interested in is outside the β-level interval, the components are not correlated and we can perform further analyses. Otherwise, a retuning of the parameters m and k may be needed using, for example, Table 3. Figure 4 shows an example of a β-level interval for the kernel of the KZ(15,5) filter.

e. Log scale for ozone

We will denote the natural logarithm of ozone O(t,x,y), where t is time and x,y is the monitor location. Log scales are as essential to the clear separation of the components of ozone as the choice of separation technique. The effect of working in the log scale can be illustrated using both raw and log-transformed ozone concentrations at Cliffside Park, New Jersey. Filtration of raw ozone concentrations leaves a short-term component, $W(t)$, that is clearly seasonal (Fig. 5a). Filtration of log-transformed ozone, on the other hand, leaves a short-term stationary variable that is nearly independent of seasonal influences (Fig. 5b). The statistical analyses of short-term components presented later in this paper would be extremely difficult, if not impossible, were the short-term components seasonal. In addition, without a log transformation, all higher-order nonlinear terms and effects in ozone data are not separated and quantile–quantile (QQ) plots of the short-term component are nonlinear.

Separation of O(t,x,y) in time by KZ filtration (with appropriate parameter choices) provides baseline (O^B) and short-term or synoptic (O^S) components defined by

$$O(t,x,y) = O^B(t,x,y) + O^S(t,x,y). \quad (8)$$

Equation (8) is practically realizable only when O^B and O^S components are cleanly separated. Poor separation leaves together in each component completely different physical phenomena. *Even when working only with summer season ozone observations, which have smaller variation in the baseline relative to a complete year, separation is needed to provide the correct short-term component.*

f. Partition of variability

The ozone time series data for 437 stations located across the United States, extracted from the EPA's AIRS for the 1983–94 time period (daily maximum 1-h ozone concentrations in ppb and daily maximum temperature in °F), were separated into high-frequency (weather related) and low-frequency components (seasonal and long term) using the KZ filter. Baseline ozone is defined as the sum of the seasonal and long-term components:

$$\text{baseline}(t) = e(t) + S(t). \quad (9)$$

Rao et al. (1996) simulated ozone time series data as

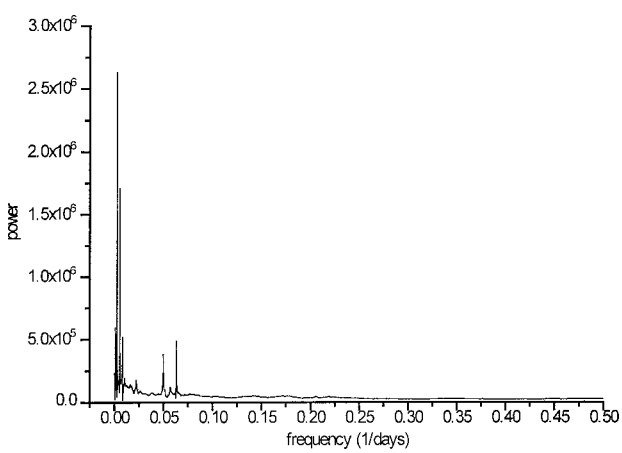

FIG. 8. Same as Fig. 6 except adaptively smoothed; Cs = 0.008%.

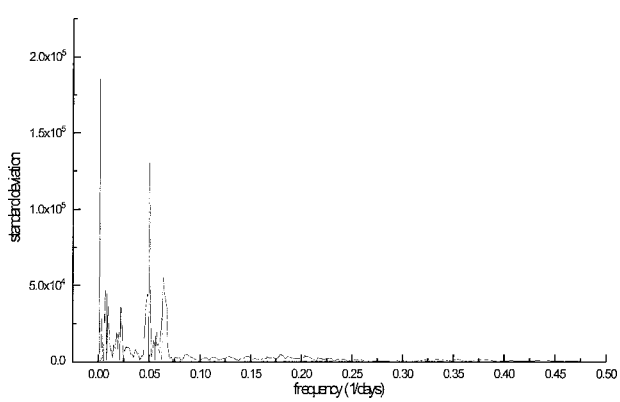

FIG. 9. Standard deviation of adaptively smoothed periodogram; Cs = 0.008%.

$$O(t) \cong \text{baseline}(t) + N\left(0, \sigma_o^2\right), \tag{10}$$

where N is the normally distributed random variable with a zero mean and variance of σ_o^2. Similarly, Eskridge et al. (1997) were able to recreate the temperature time series data as

$$T(t) \cong \text{baseline}(t) + N\left(0, \sigma_T^2\right), \tag{11}$$

where σ_0 and σ_T are the standard deviations of the high-frequency processes (the short-term component) in the ozone and temperature time series data, respectively.

An estimate of the baseline was provided by the filtered ozone or temperature:

$$\text{estimate of baseline}(t) = \text{KZ}_{15,5}, \tag{12}$$

where $\text{KZ}_{15,5}$ refers to five passes of a simple moving average of width 15 days. The effective filter width is approximately $15(5^{1/2})$ days, or approximately 33.5 days, which results in an approximate separation frequency of $0.5/[15(5^{1/2})] = 0.0149$, or separation time of approximately 67 days. From Table 1, we see that the exact time of separation for KZ(15,5) is 1/0.0136 = 73.5 days. The baseline contains phenomena that have a period longer than 73.5 days, and the residuals of the filter, $[O(t) - \text{KZ}_{15,5}]$ or $[T(t) - \text{KZ}_{15,5}]$, contain high-frequency processes.

One might ask why $\text{KZ}_{15,5}$ rather than some other choice of parameters. For any set of parameters, one might also consider how well the synoptic variation is separated from the baseline; that is, are the resulting components independent or is there some correlation left between them (and, therefore, mixed energies)? To answer these two questions, we will follow the procedures outlined in sections 2c and 2d above.

Typically, one would find the separating frequency (cutoff frequency) from the power spectrum. In the case of ozone data, however, it is difficult to find the frequencies at which the energy is concentrated (Figs. 6 and 7). The classical solution to this inconvenience is to smooth the periodogram, which we have done using the algorithm (DZ algorithm) constructed by DiRienzo and Zurbenko (1997) and DiRienzo et al. (1997). When there is information about $\{X_t\}$ at ω, it is reflected by a sharp change in the spectra at ω, motivating one to construct a spectral estimate with variable window width selected as follows. At each point of estimation ω_k, extend the width of the spectral window $\Phi_N(\bullet)$ until the local squared variation of the periodogram $I_N(\bullet)$ within reaches a prespecified constant value C_s. This amounts to considering spectral estimates $\hat{f}_N(\omega_k)$:

$$\hat{f}_N(\omega_k) = \frac{1}{2r_k + 1} \sum_{j=-r_k}^{r_k} I_N\left(\omega_{k+j}\right), \quad k = -N/2 + 1, \ldots, N/2, \tag{13}$$

where $2r_k + 1$ is the length of the window of the adaptive smoothing procedure, chosen such that r_k is the largest integer satisfying:

$$\sum_{j=-r_k}^{r_k - 1} \left[I_N\left(\omega_{k+j+1}\right) - I_N\left(\omega_{k+j}\right) \right]^2 \preceq C_S. \tag{14}$$

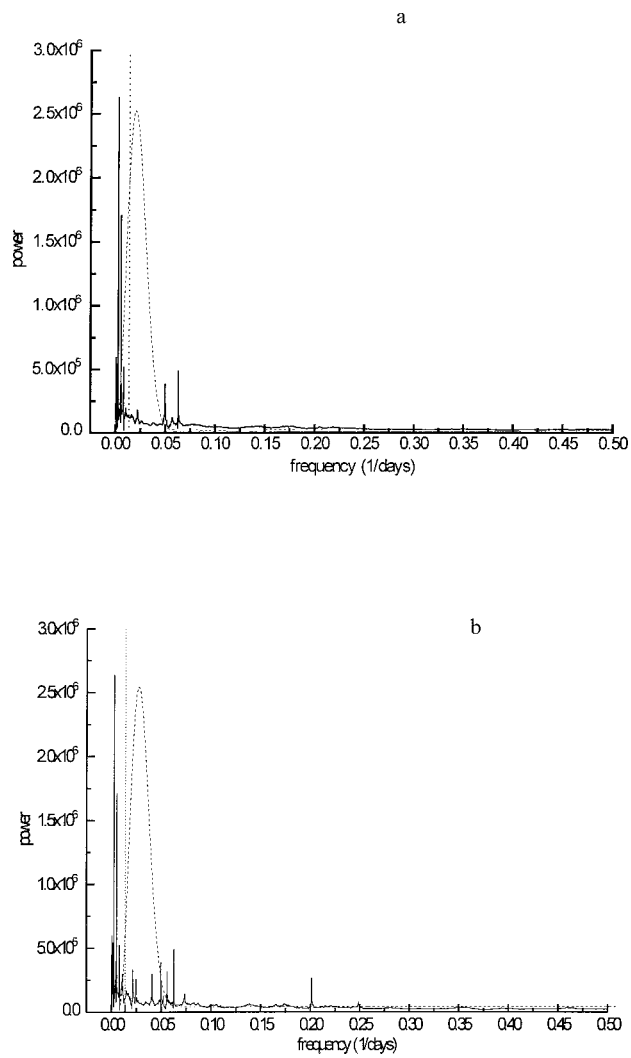

FIG. 10. Kernel for KZ(15,5) for (a) cut = 0.0136, Cs = 0.008%; (b) cut = 0.0136, Cs = 0.004%.

The *adaptively smoothed periodogram* uses a spectral window that varies with location (different from the S-plus smoothing procedure of Fig. 7, which uses a modified Daniell smoother that has a fixed value throughout the periodogram); the spectral window width is proportional to the local variation of the periodogram. Variation in the periodogram is an indication of information present in the data at that frequency; in order not to destroy this information by oversmoothing, we need to control the width of the spectral window. We compute the total variation (TV) of the periodogram, and we smooth on intervals where its variation does not exceed a given constant C_s. This constant is better understood as a percentage of the total variation, that is, C_s/TV%. Therefore, at a given frequency, we extend the spectral window until the local variation reaches the value C_s; this results in averaging few points around frequencies with high energies (information present) and averaging over large sets of points where there is low energy (noise present).

Following the DZ algorithm (DiRienzo et al. 1997; DiRienzo and Zurbenko 1997), we produced an *adaptively smoothed periodogram* that clearly shows the high energy frequencies (Fig. 8). The baseline and short-term components are clearly identified in the *adaptively smoothed periodogram*, making selection of the cutoff frequency an easy task. It is obvious from Fig. 8 that the energy concentrated at the first peak (mostly seasonal) is significant, but we are also concerned with whether the second group of peaks at about 20 days contains any information or should be thought of as noise. Therefore, we computed the standard deviation for the adaptively smoothed periodogram (Fig. 9) and compared the height of the 20-day peak (5×10^5) with the standard deviation at that period (1.25×10^5). The proportion of the two (the height of the peak is approximately 4 times larger than its standard deviation) is very strong evidence that the level of energy at that period is significant, indicating the presence of some information in the short-term component. Although several of the peaks in the short-term component appear to be significant, from Fig. 8 we observe that the energy is not concentrated at a single location. Further, the magnitudes of energies decrease exponentially (see Fig. 10b), an indication of a Markov process. While this information may be significant, it represents only a small portion (about 2%) of the total variation of the short-term component at this location.

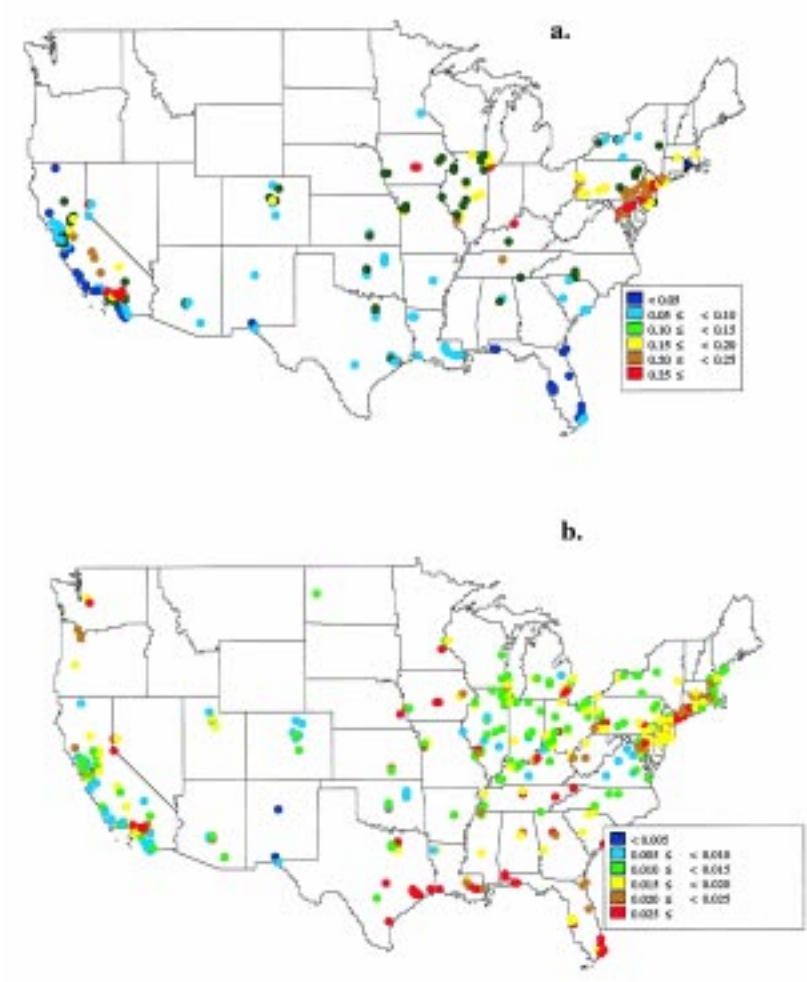

FIG. 11. Variance of the ozone seasonal component for (a) year-round monitoring and (b) 120-day summer season (15 May to 15 September).

Figures 10a and 10b show the cutoff frequency for the $KZ_{15,5}$ and also the kernel function as a measure of the correlation between resulting components of the filter $KZ_{15,5}$. Because the two frequency regions responsible for most of the variation in the data are clearly outside a β-level interval (with small β), we can conclude that there will be very low correlation between resulting components of $KZ_{15,5}$.

One would not go too far wrong if they used filter parameters (15,5) for all ozone monitoring sites. However, the procedure outlined above for choosing filter parameters will lead to slight differences among ozone (and temperature)

time series and, hence, should be repeated to assure optimal results.

Sample estimates of individual components were obtained as follows:

$$\text{estimate of } W(t) = O(t) - \mathrm{KZ}_{15,5}, \quad (15)$$

$$\text{estimate of } e(t) = \mathrm{KZ}_{365,3}, \quad (16)$$

$$\text{estimate of } S(t) = \mathrm{KZ}_{15,5} - \mathrm{KZ}_{365,3}. \quad (17)$$

The total variance of $O(t)$ can be written as a sum of the variances and covariances of the ozone components separated by the filter:

$$\sigma^2(O) = \sigma^2(e) + \sigma^2(S) + \sigma^2(W) + 2\,\mathrm{cov}(e,S) + 2\,\mathrm{cov}(e,W) + 2\,\mathrm{cov}(S,W). \quad (18)$$

The sum of covariance terms was typically less than 2% of the total variance, indicating good separation of components, as described above.

For data at 437 U.S. stations, long-term fluctuations are a small fraction of the total. More than 90% of year-round and 95% of summer season monitoring sites have long-term components that are less than 10% of the total variation. Median values (fraction of total variance in the long term component) for year-round and summer season fractions are 2.4% and 3.6%, respectively. Summer season data have a greater relative contribution of long-term and synoptic fluctuations because the seasonal component is reduced by about half.

Seasonal fluctuations account for up to 73% and 60% of the total variance in year-round and summer season data, respectively, with median values of 51% and 12%. Seasonal variance patterns are quite different for year-round and summer season data. For year-round monitoring, seasonal variance is lowest near coastal areas and highest in a band from the Midwest to the East Coast (Fig. 11a). For summer season–only monitoring, there is no clear pattern to seasonal variance. Values tend to be highest on the West Coast and in the Midwest, and low in the Southeast. However, sites with low seasonal variance can be found throughout the United States.

Median synoptic (short term) fluctuations are 45% and 77% for year-round and summer season data, respectively. In relative terms, synoptic fluctuations tend to be low in the Midwest and high in coastal areas. For sites monitoring year-round or summer season–only, areas of high short-term variance include the Northeast and parts of the Gulf Coast (Fig. 12).

Taken together, Figs. 11 and 12 illustrate the amount of energy associated with seasonal and synoptic processes across the United States. Long-term variances are small in both relative and absolute senses. Seasonal forces increase relative to synoptic forces in a northeasterly direction. In an absolute sense, the Northeast is a high energy region, having large seasonal and synoptic variances.

g. Analysis of short-term (synoptic) component

As previously noted, ozone data are log-transformed prior to filtering. Variance stabilization does not lead automatically to normalization, however, and the synoptic components are negatively skewed for

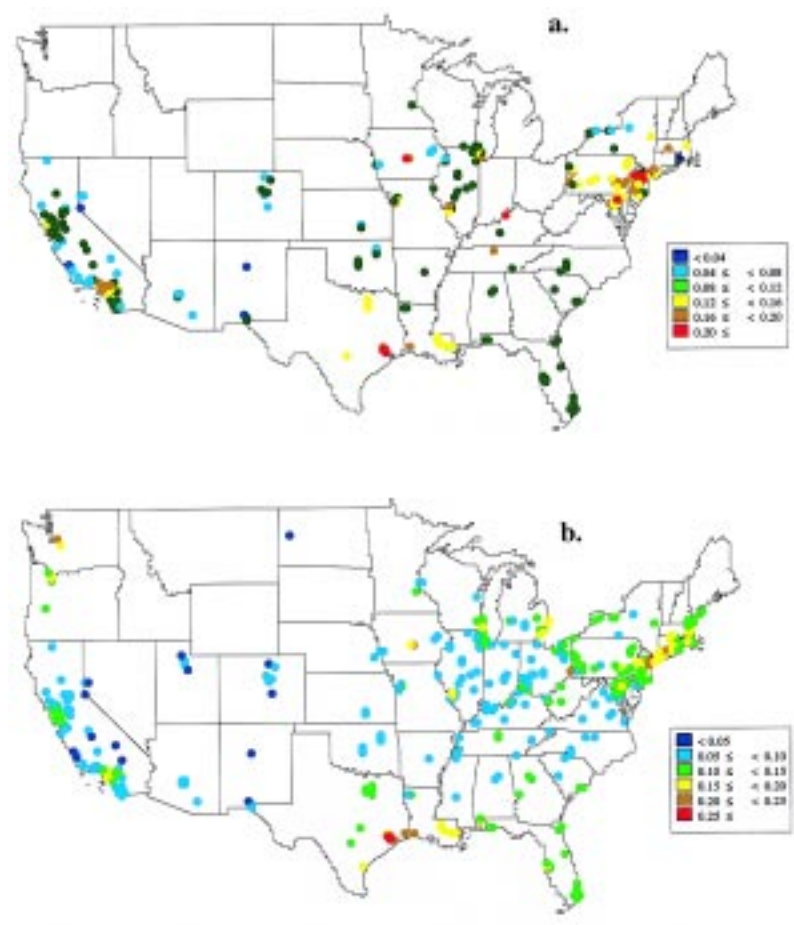

FIG. 12. Variance of the ozone short-term component for (a) year-round monitoring and (b) 120-day summer season (15 May to 15 September).

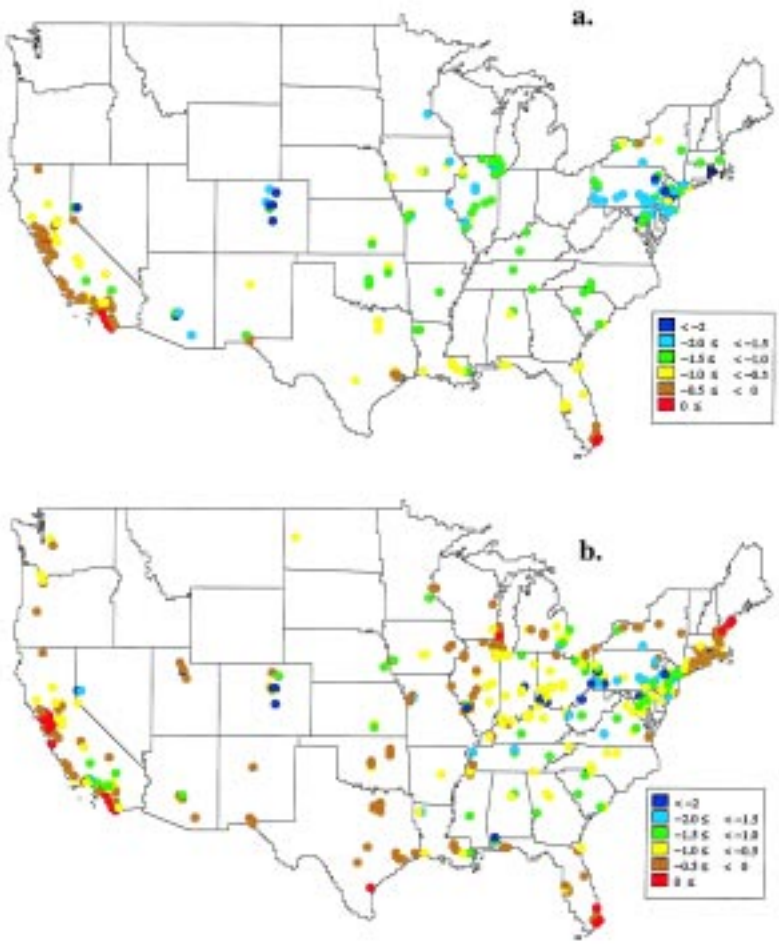

FIG. 13. Coefficient of skewness of the ozone short-term component for (a) year-round monitoring and (b) 120-day summer season (15 May to 15 September).

nearly every monitoring station (Fig. 13), probably attributable to the variability in the synoptic-scale processes responsible for the ozone accumulation and removal. Other transformations, such as Box–Cox, square root, etc., may produce $W(t)$ that closely approach normality at every location. However, each monitoring station would require a slightly different transformation.

The independence of the $W(t)$ were assessed by analyzing their correlograms, which are consistent with an ARMA(2,1) or ARMA(1,1) model:

$$[O(t) - \mathrm{KZ}(15,5)] = \Theta e(t-1) + \Phi_1 \mathrm{W}(t-1) + \Phi_2 \mathrm{W}(t-2) + e(t) \qquad (19)$$
$$e \sim IID(0,\sigma^2),$$

where *IID* refers to independent and identically distributed random variables. The moving average component is due to the moving average filtration [since the estimate of $W(t)$ is a difference between $O(t)$ and a moving average], while the autoregressive component is attributable to correlation in day-to-day weather phenomena. One-day lag correlations ranged from 0.20 to 0.5, with a median value of 0.35.

The filter effect is illustrated by Fig. 14a, which shows the expected correlogram of independent random numbers subject to the $\mathrm{KZ}_{15,5}$. Significant negative serial correlation extends to about half the effective filter length. A purely autoregressive process produces a correlogram resembling Fig. 14b. The correlograms of the estimated $W(t)$ for nearly all stations resemble Fig. 14c.

A few stations have very strong weekly cycles, which produce significant autocorrelations every 7 days. One might also notice the resemblance between the spectra of short-term ozone at a typical site (Fig. 15a) and that for a Markov process subject to filtration (Fig. 15b). The decay in the height of the peaks follows a similar pattern in both figures.

Some improvement in the simulation of $W(t)$ would result from using (19) but at a cost of increasing complexity. In particular, (19) requires four parameters, whereas (10) and (11) need only 1.

The *e*-folding distance (the distance at which the correlation drops to the $1/e$ value) for short-term (synoptic forcing) ozone is on the order of 600 km. One can address the timescale for the short-term component either by applying a mean wind speed to the spatial scale or by examining the temporal (serial) correlation in the short-term ozone time series data. Based on the former, mean wind speeds in the range of 8–20 km h^{-1} lead to timescales of about 1–3 days for ozone. For the latter approach, the short-term component was assumed to follow an AR(1) process. The *e*-folding times in the autocorrelation coefficient of an AR(1) process were then calculated at 437 ozone and 270 temperature monitoring stations in the United States (Fig. 16). After accounting for the effects of filtration, the *e*-folding times for short-term ozone were found to range from about 0.5 to 2.5 days (Fig. 16b). Ozone timescales are shorter in the Northeast than in the Southeast. Examination of the timescales for short-term (weather induced) surface temperature in

Fig. 16a suggests that ozone is associated with slower-moving synoptic patterns in the Southeast and faster-moving weather systems in the Northeast. High ozone levels in the eastern United States are often associated with a slow-moving high pressure system, a Bermuda high, which results in near stagnant and disorganized flow conditions conducive to the ozone accumulation in the Southeast (Vukovich 1995). When the flow becomes organized as the high pressure system advances, the Northeast might then receive ozone, and its precursors from transport around the high pressure system and the flow could be fairly rapid in the Northeast. These results are consistent with those derived from time-lagged intersite correlation analyses by Brankov et al. (1997).

3. Space scales in baseline, short term, and exceedances

a. Spatial information in baseline and short-term components

Spatial correlations among the baseline components in ozone remain relatively unchanged over a distance of 300 km, indicating similar seasonal influences over this scale (Fig. 17a). Synoptic ozone $[O^S(t,x,y)]$ contains a very precise spatial law related to the directional decay of their spatial correlations (Rao et al. 1995); that is there is a strong exponential decay relationship in the correlation between two monitors m_0 and m' and their distance of separation, d, in direction ϕ:

$$\text{correlation}[O^S(t,m_0), O^S(t,m')] \approx \exp[-a(m_0, \phi)d + b], \quad (20)$$

where "a" depends on reference monitor location (x_0,y_0) and spatial direction ϕ (Fig. 17b). Stated in other words, there is a Markov relationship between short-term (synoptic) components in space.

Obvious in a physical sense is that a station m'' downwind of station m' does not have information about m_0: ozone transported to m' or produced there cannot be distinguished farther downwind at m''. The Markov property for a Gaussian process yields

$$O^S(t,m') = \exp\{-a\,d + b\}O^S(t,m_0) + O^{\text{loc}}(t,m'), \quad (21)$$

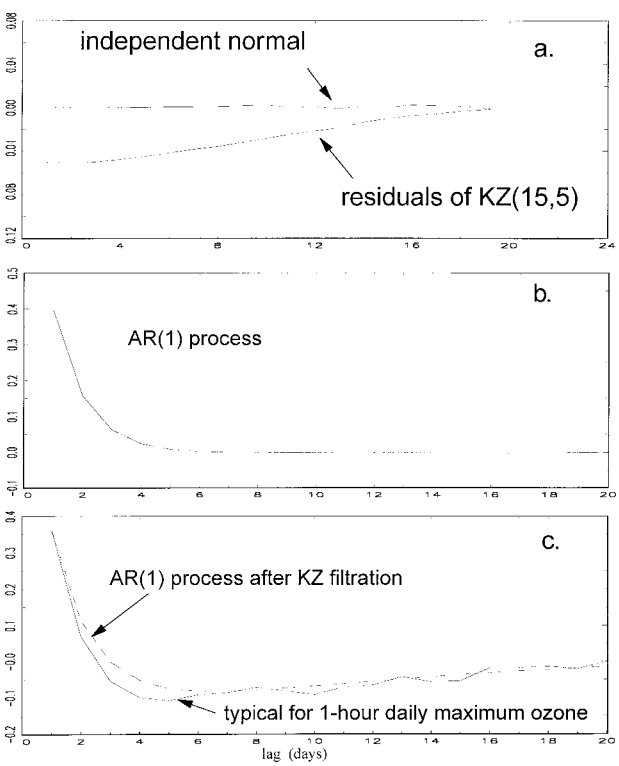

FIG. 14. Correlograms for (a) Gaussian white noise residuals of KZ[N(0,1), 15,5] and (b) AR(1) process; (c) AR(1) process after KZ(15,5) and typical of 1-h daily maximum ozone.

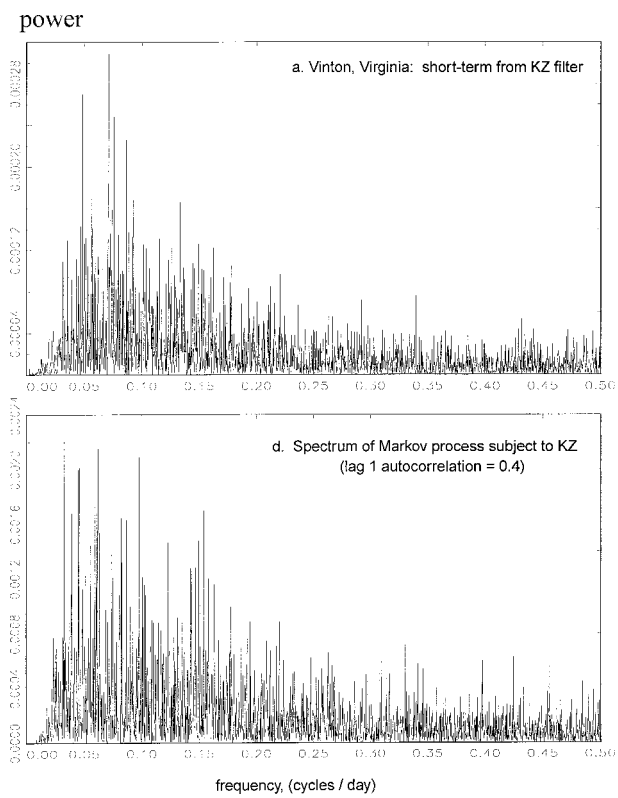

FIG. 15. Spectrums of (a) short-term ozone and (b) Markov process.

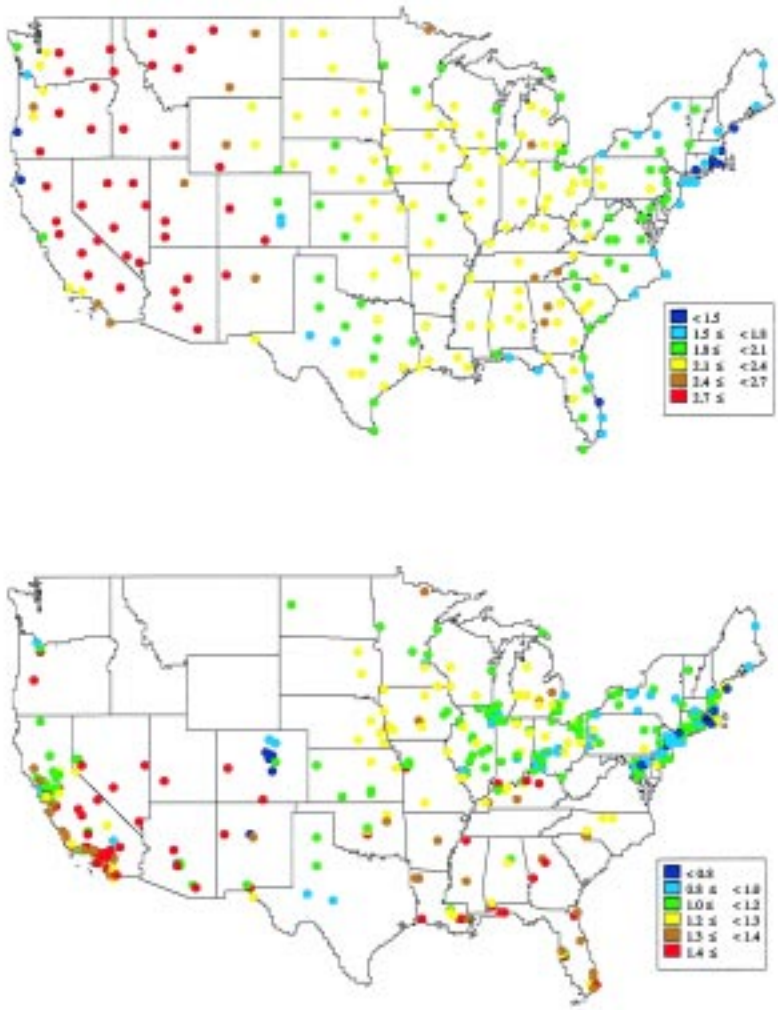

FIG. 16. *E*-folding timescales (in days) for (a) surface temperature and (b) ozone.

where $O^{loc}(t,m')$ is a local contribution portion at the location m'. Therefore, (21) separates the transport portion of $O^{s}(t,m')$ from m_0 to m' and local production. At distances of about 300 km, correlations in baselines exceed 90% and contain about 60% of total process energy. Figure 17b displays a 50% decline at distances of about 400 km produced by (21) in the direction of the prevailing wind (maximum transport direction).

Spatial correlations among short-term components initially have a much greater rate of decay than is indicated by the exponential relationship in Fig. 17b (distances less than 15 km.). Fifteen kilometers represents the scale of "direct" transport of ozone, as well as the scale at which the exceedances (concentrations exceeding a given threshold) in combination with deterministic models can yield accurate predictions. Unfortunately, a monitoring network with stations every few kilometers is practically impossible. At short distances (< 80 km), the decline in a correlation plot determined by "*b*" from (21) is related to the very short survival time of ozone (Hales 1996; Loibl et al. 1994) and can be examined separately if spatially dense ozone data are available.

The exponential term in (21) is related to the synoptic transport of the pollution, weather conditions, and the emissions that created ozone. The coefficient a (ϕ,x,y) smoothly depends on (ϕ,x,y) and can be plotted in space using a spatial filter [see for example Rao et al. (1995) and Zurbenko et al. (1995)]. Correlations in the short-term component for the same distance are about 70% and contain about 40% of total process energy. Since the baseline and short-term components are nearly orthogonal, the spatial prediction error in total ozone over a distance of about 150 km will be less than 20%. The decay of correlations among short-term (synoptic) components in temperature data is evident in Fig. 17d. The exponential decay of the correlations with distance along the direction of the prevailing wind at Charlotte, North Carolina, and Cincinnati, Ohio, presented in Fig. 18, reveal *e*-folding distances (scales) for ozone that are similar to those extracted from the data in the Northeast.

b. Spatial relationship among ozone exceedances

Spatial correlations for exceedances (ozone concentrations exceeding the 0.12-ppm level) were calculated from

$$\text{correlation}_t\{E(t,x_0,y_0), E(t,x',y')\} = C_E(d,\phi), \quad (22)$$

where E(t,x,y) is an indicator of exceedances on day t at location x,y; $E(t,x,y) = 1$ when an exceedance occurs and 0 otherwise; d is the distance between m_0 and m'; and ϕ is direction. Correlations for exceedances in space are calculated in the same way as for the synoptic component. Spatial correlations for exceedances extend information only about 15 km and are indistinguishable from zero for distances of more than 50 km (Fig. 17c). There are many examples where the 3-yr total exceedances are completely different within

the range of 15 km. A plausible explanation for the rapid decay of correlations for ozone exceedances may be the spatial inhomogeneity in the NOx levels in an urban area and the importance of chemical lifetime of ozone relative to the advective time. Such high spatial variability necessarily requires a monitoring density on a scale of about 5 km to capture all the exceedances. *Thus, the space and time scales associated with the ozone exceedance metric render its use as a control variable impractical.*

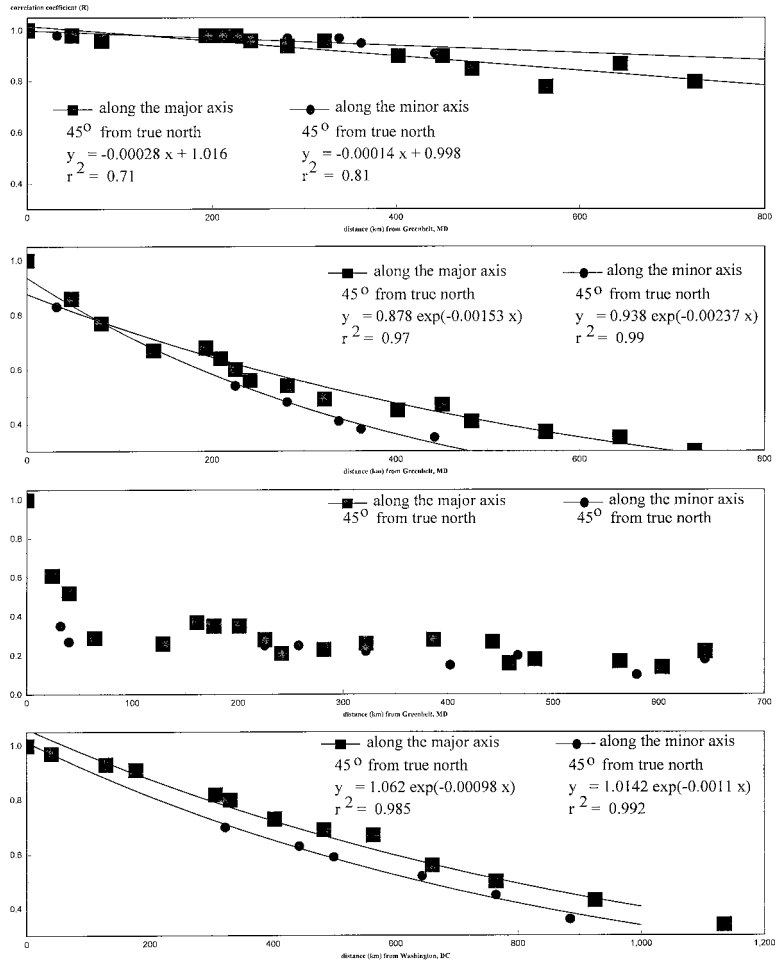

FIG. 17. Correlation coefficient as a function of distance (km) from Washington, DC, in the direction of the prevailing wind between (a) baselines of ozone, (b) short-term components of ozone, (c) ozone exceedances, and (d) short-term components of temperature.

4. Summary

In this paper, we have demonstrated the need for separating the various spectral components of ozone time series. Separation leads to a clearer understanding of ozone and its relationships to meteorological and precursor variables. Any spectral-decomposition technique (e.g., a wavelet transform) capable of creating statistically independent short-term, seasonal, and long-term components will achieve the goal of providing a better understanding of the underlying physical processes that affect ambient ozone concentrations.

Among the useful results of this analysis are descriptions of spatial–temporal information in ozone data, needed for ozone management efforts. This information, unavailable in a meaningful form from raw ozone data, can be clearly displayed in the separated components. In addition, baseline ozone retains global information on the scale of more than 2 months in time and about 300 km in space. The short-term ozone component, attributable to weather fluctuations, is highly correlated in space, retaining 50% of the short-term information at distances ranging from 350 to 400 km; in time, short-term ozone resembles a Markov process with 1-day lag correlations ranging from 0.2 to 0.5. Furthermore, the correlation structure of short-term ozone permits highly accurate predictions of ozone concentrations up to distances of about 600 km from a monitor. The ozone timescales in the United States range from approximately 1 to 2.5 day depending upon the monitoring location.

The geographical variation in the timescales for ozone and temperature reveal that ozone in the Southeast is associated with slow-moving synoptic conditions and fast-moving weather systems in the Northeast. The timescale of approximately 1–2.5 day and the space scale of about 600 km in ambient ozone data imply that ozone in the eastern United States is a regional-scale problem.

Acknowledgments. This research is supported by the Environmental Protection Agency under Grant R825260-01 and by the Electric Power Research Institute under Contract WO4447-01.

References

Brankov, E., S. T. Rao, and P. S. Porter, 1997: A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants. *Atmos. Environ.*, in press.

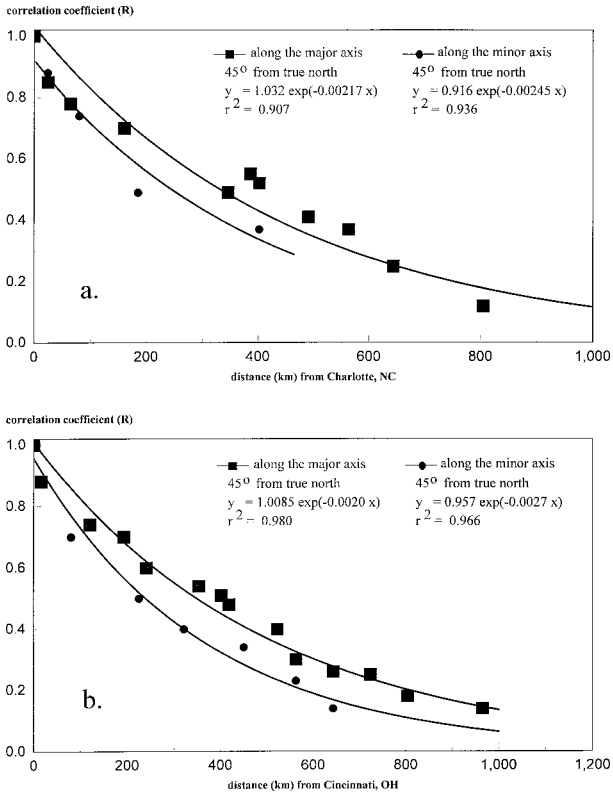

FIG. 18. Correlation coefficient between short-term components of ozone as a function of distance (km) in the direction of the prevailing wind from (a) Charlotte and (b) Cincinnati.

DiRienzo, A. G., and I. G. Zurbenko, 1997: Adaptive nonparametric spectral estimation. *J. Comput. Graphical Stat.*, in press.

——, ——, and D. O. Carpenter, 1997: Time series analyses of *Aplysia* total motion activity. *Biometrics*, in press.

Eskridge, R. E., J. Y. Ku, S. T. Rao, P. S. Porter, and I. G. Zurbenko, 1997: Separating different scales of motion in time series of meteorological variables. *Bull. Amer. Meteor. Soc.*, **78,** 1473–1483.

Hales, J., 1996: Scientific background for AMS policy statement on atmospheric ozone. *Bull. Amer. Meteor. Soc.*, **77,** 1249.

Lau, K. M., and W. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.*, **76,** 2391–2402.

Loibl, W., W. Winiwarter, A. Kopsca, and J. Zueger, 1994: Estimating the spatial distribution of ozone concentrations in complex terrain. *Atmos. Environ.*, **28,** 2557–2566.

Porter, P. S., S. T. Rao, I. G. Zurbenko, E. Zalewsky, R. F. Henry, and J. Y. Ku, 1996: Statistical characteristics of spectrally-decomposed ambient ozone time series data. Washington Univ. [Available on-line from http:\\capita.wustl.edu\otag\reports\StatChar\otagrep.htm.]

Rao, S. T., and I. G. Zurbenko, 1994: Detecting and tracking changes in ozone air quality. *J. Air Waste Manage. Assoc.*, **44,** 1089–1092.

——, E. Zalewsky, and I. G. Zurbenko, 1995: Determining spatial and temporal variations in ozone air quality. *J. Air Waste Manage. Assoc.*, **45,** 57–61.

——, I. G. Zurbenko, P. S. Porter, J. Y. Ku, and R. F. Henry, 1996: Dealing with the ozone non-attainment problem in the Eastern United States. *Environ. Manager*, January, 17–31.

Vukovich, F. M., 1995: Regional-scale boundary layer ozone variations in the eastern United States and the association with meteorological variations. *Atmos. Environ.*, **29,** 2250–2273.

Wallis, T. W. R., 1996: Report on the search for core stations and suggested products from the Comprehensive Aerological Reference Data Set (CARDS). [Available from National Climatic Data Center, 151 Patton Ave., Asheville, NC 28806.]

Zurbenko, I. G., 1986: *The Spectral Analysis of Time Series.* North Holland, 241 pp.

——, 1991: Spectral analysis of nonstationary time series. *Int. Stat. Rev.*, **59,** 163–173.

——, S. T. Rao, and R. F. Henry, 1995: Mapping ozone in the eastern United States. *Environ. Manager*, February, 24–30.

——, P. S. Porter, S. T. Rao, J. Y. Ku, R. Gui, and R. E. Eskridge, 1996: Detecting discontinuities in time series of upper air data: Development and demonstration of an adaptive filter technique. *J. Climate*, **9,** 3548–3560.