Statistical analysis of the Los Angeles catalyst study data

ARTICLE · SEPTEMBER 1979

Johannes Ledolter
University of Iowa

Statistical Models for Ambient Air Pollutants, with Special Reference to the Los Angeles Catalyst Study (LACS) Data

Johannes Ledolter*

Department of Statistics, University of Iowa, Iowa City, Iowa 52242

George C. Tiao

Department of Statistics, University of Wisconsin, Madison, Wis. 53706

■ In a recent article, Tiao and Hillmer presented a statistical analysis of pollutant data gathered under the Los Angeles Catalyst Study (LACS) program. This paper extends their analysis using the most recent data on pollutant concentrations, meteorological variables, and traffic. A model is developed that jointly predicts CO concentrations on and on each side of the freeway. Also, their Pb model is further refined, and it is shown that the observed Pb increase in 1977 is due to an additional traffic lane that resulted in higher traffic speed.

An extensive data collection was undertaken to study the environmental impact of the catalytic converter in a real world setting. The converter has been adopted by American car manufacturers since the 1975 model year to reduce the emissions of carbon monoxides and hydrocarbons. Furthermore, since catalyst-equipped cars require unleaded gasoline, its use should lead to a reduction in lead concentrations.

Four air monitoring stations, A, B, C, and D, two on each side of the San Diego Freeway in Los Angeles, were established by the Environmental Protection Agency (Figure 1). At each location, data on a number of pollutants including CO and Pb, and on meteorological variables such as wind speed and wind direction, have been collected since June 1974. A fifth measurement site, F, for CO concentrations at the median strip was added in January 1977. In addition, hourly traffic counts and average traffic speeds for each traffic lane have been recorded since September 1976.

Detailed statistical analyses of the data are given elsewhere (*1–4*). In this paper we first propose a joint model that relates hourly CO concentrations at locations A, B, F, and C to traffic, wind speed, wind direction, and distance of the measurement site from the freeway median. Such a model can be used to predict CO concentrations for locations at any given distance from the median and, in contrast to the EPA HIWAY model (*5*), is capable of predicting CO concentrations for downwind and upwind sites. Furthermore, we present a modified version of the model for Pb given in ref *1* that takes into account the fact that Pb emissions increase with the traffic speed. Using this new model, it is shown that the observed Pb increase in 1977 at site C can be explained by the opening of an additional northbound traffic lane, which resulted in higher traffic speed.

Analysis of CO Data

Preliminary Analysis of Traffic, Wind, and CO. It is argued by Tiao and Hillmer (*1*) that two major factors influencing the ambient CO concentrations are traffic and the wind vector. We first present some preliminary graphical analyses of the traffic, wind, and CO data.

Traffic. Since September 1976, traffic counts and average traffic speeds have been recorded for each lane of the San Diego Freeway. Until January 1977, the freeway consisted of four lanes in each direction, separated by a concrete median strip. In February 1977, a fifth northbound lane was added. Figures 2a–d show diurnal plots of hourly traffic count and average traffic speed for weekdays (Monday–Thursday) and weekends (Saturday, Sunday) in 1976 and 1977. In all diurnal plots we have adopted the convention that the value for time t corresponds to the average for the hour t to $t + 1$ ($t = 0, 1, \ldots, 23$). Weekday traffic counts are characterized by two peaks, occurring during the morning and afternoon rush hours. Furthermore, during rush hours the traffic speed is considerably reduced. On weekends, the morning peak of the traffic count is absent and the average traffic speed is rarely less than 50 mph.

A comparison of 1976 and 1977 traffic figures shows that the additional northbound lane led to an increase in the average weekday traffic speed, especially during the afternoon rush hour period. The weekend traffic speed stayed largely unaffected. Furthermore, a slight increase in traffic volume can be noticed.

Tiao and Hillmer (*1*) observed that during the afternoon rush hours, because of traffic congestion, the traffic counts do not reflect the magnitude of the CO concentrations. They proposed to consider the traffic density, TD, which is the ratio

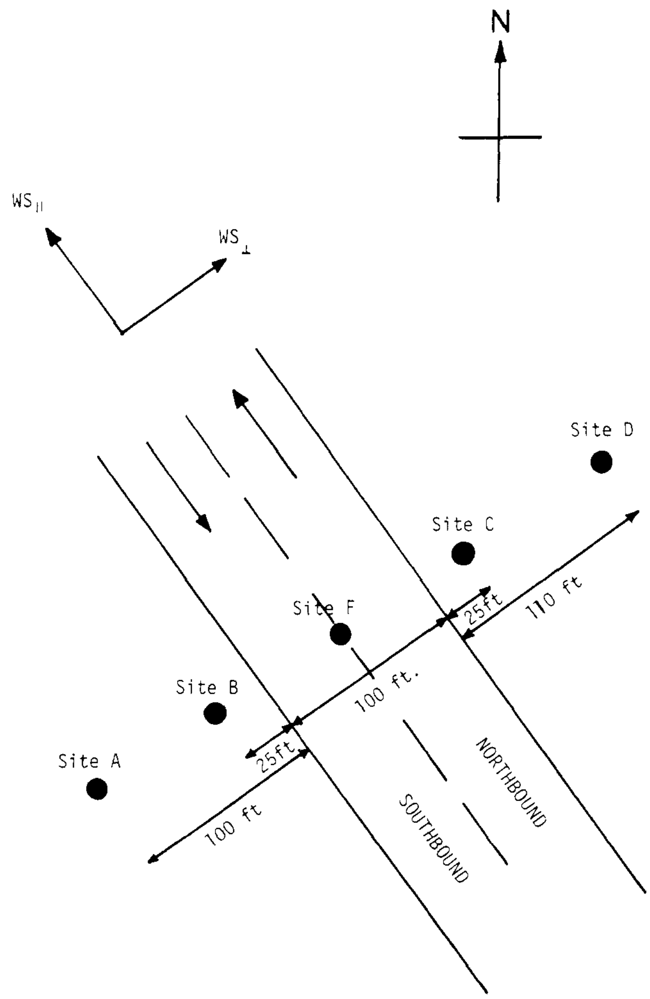

Figure 1. Map of Los Angeles air monitoring stations at San Diego Freeway

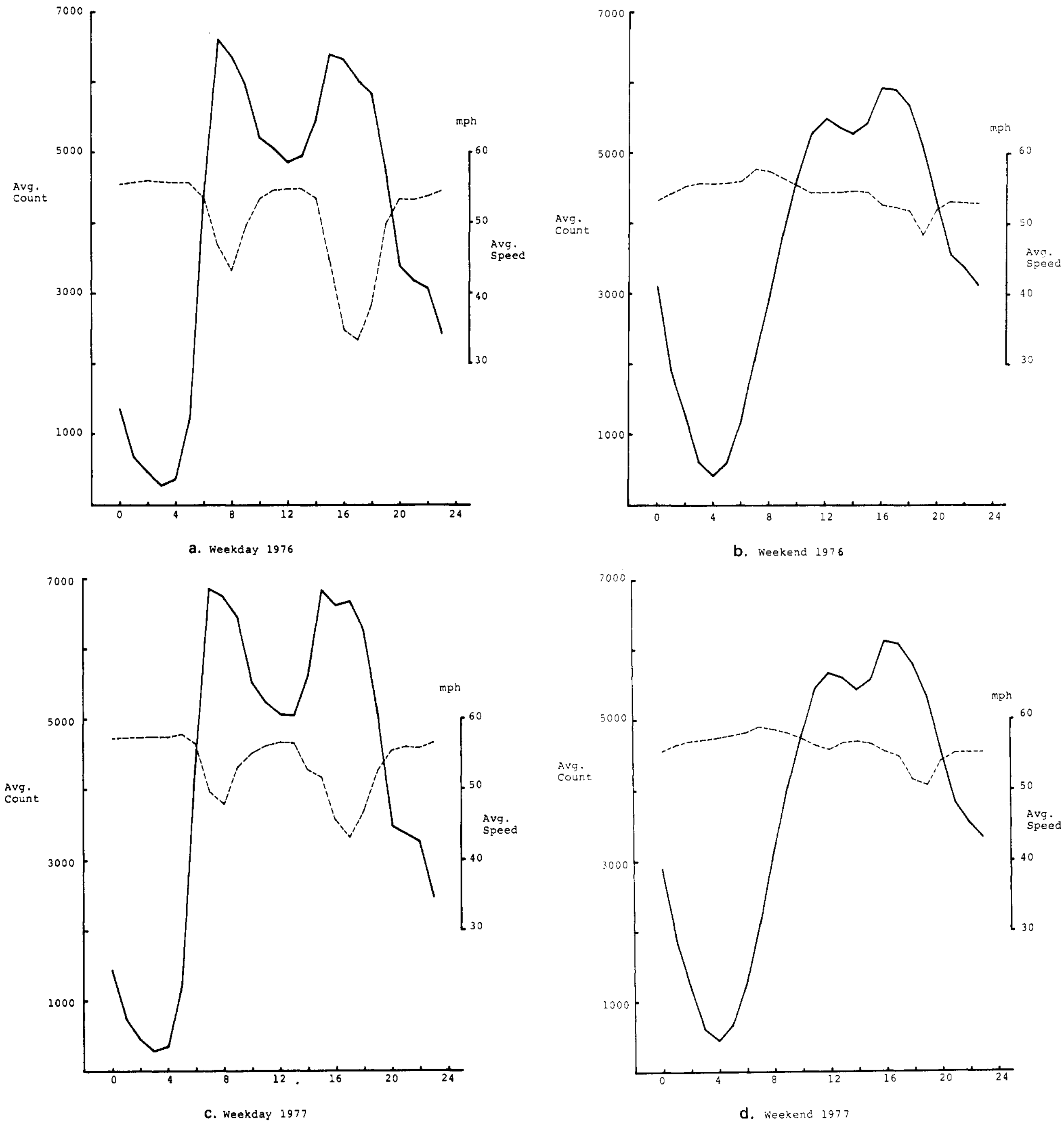

Figure 2. (a–d) Diurnal plots of traffic count and traffic speed; weekday–weekend, 1976–1977: (—) count; (- - -) speed

of traffic count to traffic speed. Figures 3a and 3b present the 1976 and 1977 density figures for weekdays and weekends. Whereas the weekend traffic densities are largely unaffected by the additional northbound lane, one can notice a rather substantial decrease in weekday afternoon traffic densities from 1976 to 1977.

Wind Speed and Wind Direction. Wind direction (WD) is an important factor, since it controls the direction of transport of the pollutants. Since the freeway is at 145° north, winds from 145 to 325° will transport the freeway contribution toward C, while locations A and B act as background stations. If the winds are from 325 to 145°, the freeway contribution will be recorded at locations A and B. Wind speed (WS) is important, since it controls transport as well as diffusion. At higher wind speeds more CO will be diffused. Following ref *1* and *6*, we decompose the hourly wind vector into two components, one perpendicular ($WS_{\perp}$) and the other parallel ($WS_{\parallel}$) to the freeway:

$$WS_{\perp} = WS \cos (WD - 235°)$$

$$WS_{\parallel} = WS \sin (WD - 235°)$$

The positive directions of these two components are given in Figure 1. In Figure 4 we plot the diurnal pattern of $WS_{\perp}$ for the summer months (May–October 1977). The $WS_{\perp}$ is nearly zero from midnight to 9 a.m., and reaches its peak around midafternoon when WS is high and WD is consistently perpendicular to the freeway.

CO. Hourly CO data have been collected at locations A and C since June 1974 and at locations B and F since February 1977. Diurnal plots of CO averages for weekdays and weekends, May–October 1977, are shown in Figures 5a,b. The

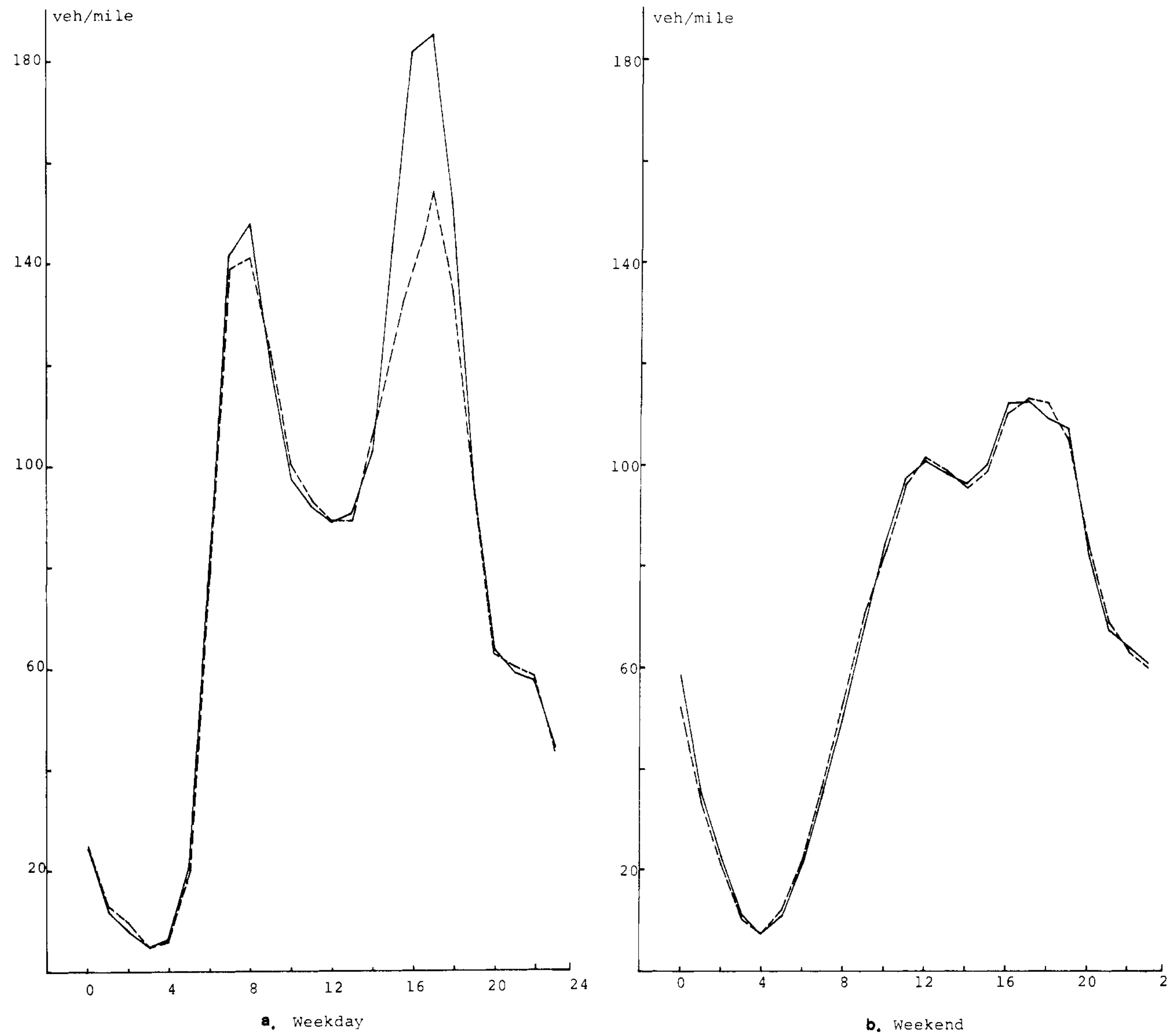

Figure 3. (a,b) Diurnal plots of traffic density TD; weekday-weekend comparison: (—) 1976; (- - -) 1977

weekday diurnal patterns of CO at the median strip F and at the downwind site C exhibit two peaks corresponding to the morning and afternoon rush hour traffic. At the upwind sites A and B, lower CO concentrations are recorded during most of the daytime hours, since the wind is consistently blowing across the freeway toward C. At locations A and B, the CO concentrations are highest during the morning rush hours when $WS_\perp$ is low. The weekend diurnal curves of CO differ from the weekday figures due to changed traffic patterns.

A Model for CO at Location C. In ref *1*, the following model was proposed to relate the diurnal behavior of CO at location C to traffic density and the perpendicular wind component:

$$[\mathrm{CO}]_t = \alpha + k\mathrm{TD}_t e^{-b(\mathrm{WS}_{\perp t}-\omega)^2} + a_t \tag{1}$$

$[\mathrm{CO}]_t$, TD_t, and $\mathrm{WS}_{\perp t}$ are the observed CO concentration, traffic density, and the perpendicular wind component at hour t; a_t is an error term; α is a parameter measuring background; k is a parameter proportional to emissions; $e^{-b(\mathrm{WS}_{\perp t}-\omega)^2}$ is a dispersion factor involving the perpendicular wind component and parameters b and ω. This simple model was extensively verified on summer, winter, weekday, and weekend averages (*1*).

A Model for CO at the Median. Ledolter et al. (*2*) proposed a model for the CO concentrations at the median strip. The model is of the form:

$$[\mathrm{CO}]_t = (\alpha + k\mathrm{TD}_t)e^{-\gamma\sqrt{|\mathrm{WS}_{\perp t}|}} + a_t \tag{2}$$

The summer 1977 CO averages at location F and the corresponding averages of $\mathrm{WS}_\perp$ and TD were used to estimate the parameters and verify the model.

A Joint Model for CO at Sites A, B, C, and F. We now consider a model that relates CO at sites A, B, C, and F jointly to traffic and the wind. As shown in Figure 1, A is 150 ft and B and C are both 75 ft from the median site F. Any joint model covering these sites must include the distance from the center of the freeway as an explanatory variable. Furthermore, since the wind vector controls the transport of the freeway contribution, the model must reflect that sites A and B, and site C, are on opposite sides of the freeway. Thus, we have chosen d to be an indicator variable for the distance such that $d_\mathrm{A} = -2$, $d_\mathrm{B} = -1$, $d_\mathrm{F} = 0$, and $d_\mathrm{C} = 1$, where one unit corresponds to 75 ft. The proposed model takes the form:

$$[\mathrm{CO}]_{jt} = g(\mathrm{WS}_{\perp t},\mathrm{TD}_t,d_j;\alpha,k,\beta,\eta,\omega)e^{-\gamma\sqrt{|\mathrm{WS}_{\perp t}|}} + a_{jt} \tag{3}$$

where

$$g(\mathrm{WS}_{\perp t},\mathrm{TD}_t,d_j;\alpha,k,\beta,\eta,\omega) = \alpha + k\mathrm{TD}_t \frac{e^{[-\eta|d_j|/(1+\beta|d_j|)](\mathrm{WS}_{\perp t} - \mathrm{sgn}(d_j)\sqrt{|d_j|}\omega)^2}}{\sqrt{1+\beta|d_j|}}$$

where $[\mathrm{CO}]_{jt}$ is the CO concentration at site j (A, B, F, or C)

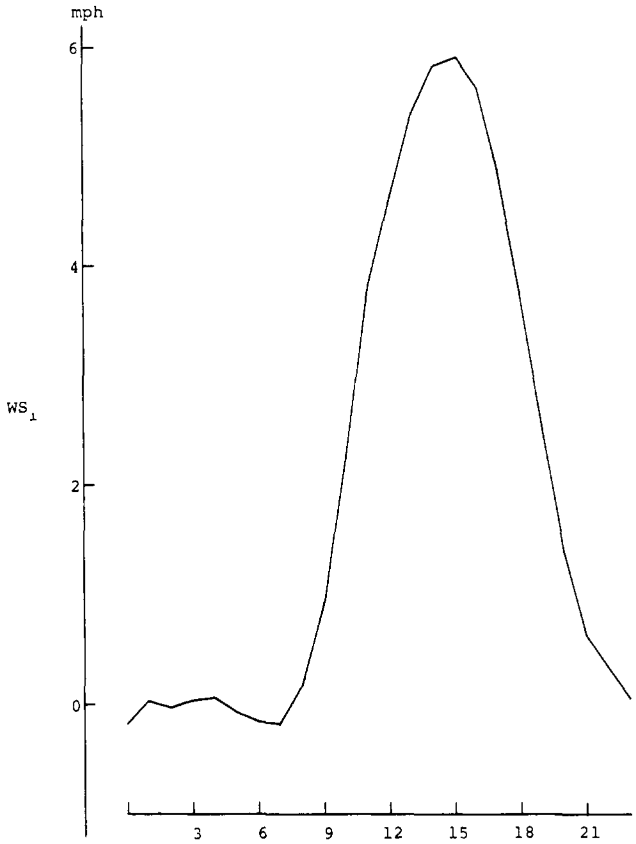

Figure 4. Diurnal plot of $WS_{\perp}$; summer 1977

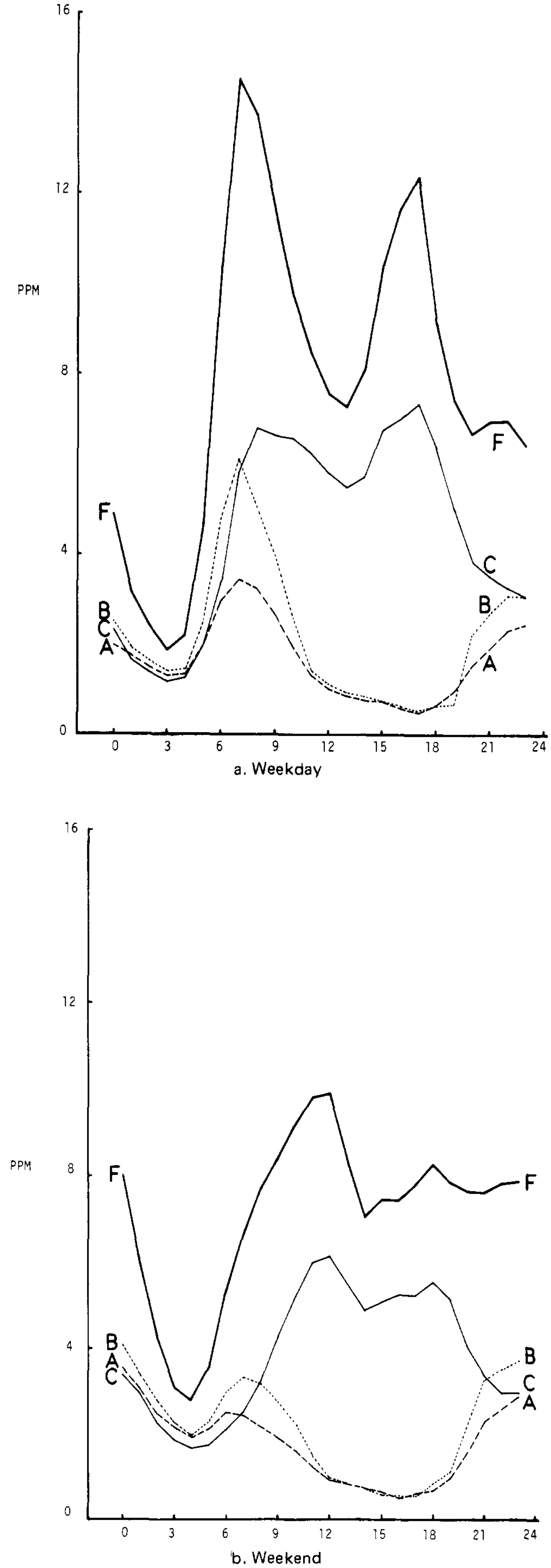

Figure 5. (a,b) Diurnal plots of CO summer (May–October 1977) averages at locations A, B, C, and F; weekday–weekend

and hour t, TD_t is the traffic density at hour t, d_j is the distance of site j from the median strip as defined above, $WS_{\perp t}$ is the perpendicular wind component at hour t, sgn is defined as $\mathrm{sgn}(x) = 1$ if $x > 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$, a_{jt} is an error term, and α, k, β, η, ω, and γ are parameters to be estimated from the data.

The model in Equation 3 is fitted by multivariate nonlinear least squares using 1977 summer weekday averages, assuming that the errors a_{jt} are independent over time, but possibly correlated across sites. The fitting results are:

parameter	estimate	std error
α	1.58	0.11
k	0.051	0.0015
η	0.088	0.014
β	1.29	0.19
ω	4.12	0.32
γ	0.17	0.016

Predicted and observed CO averages are plotted in Figure 6. The model in Equation 3 is quite parsimonious (apart from the covariances, only six parameters have to be estimated) and produces a good overall fit for all four locations.

Interpretation of the Joint CO Model. (i) The joint model in Equation 3 can be used to predict CO concentrations at any point near the freeway, irrespective of whether it is in the downwind or in the upwind direction. This is an advantage over the EPA HIWAY model (*5*), which can be used only for downwind locations.

(ii) If $d_j = 0$ (median strip), the model in Equation 3 reduces to Equation 2, the univariate model for CO concentrations at site F.

(iii) For any fixed $d_j \neq 0$, the factor $g(\cdot)$ in Equation 3 is of the same form as the model in Equation 1. In addition, the background concentration α and the freeway contribution involving the traffic density and the distance from the median are diffused by the factor $e^{-\gamma\sqrt{|WS_{\perp}|}}$.

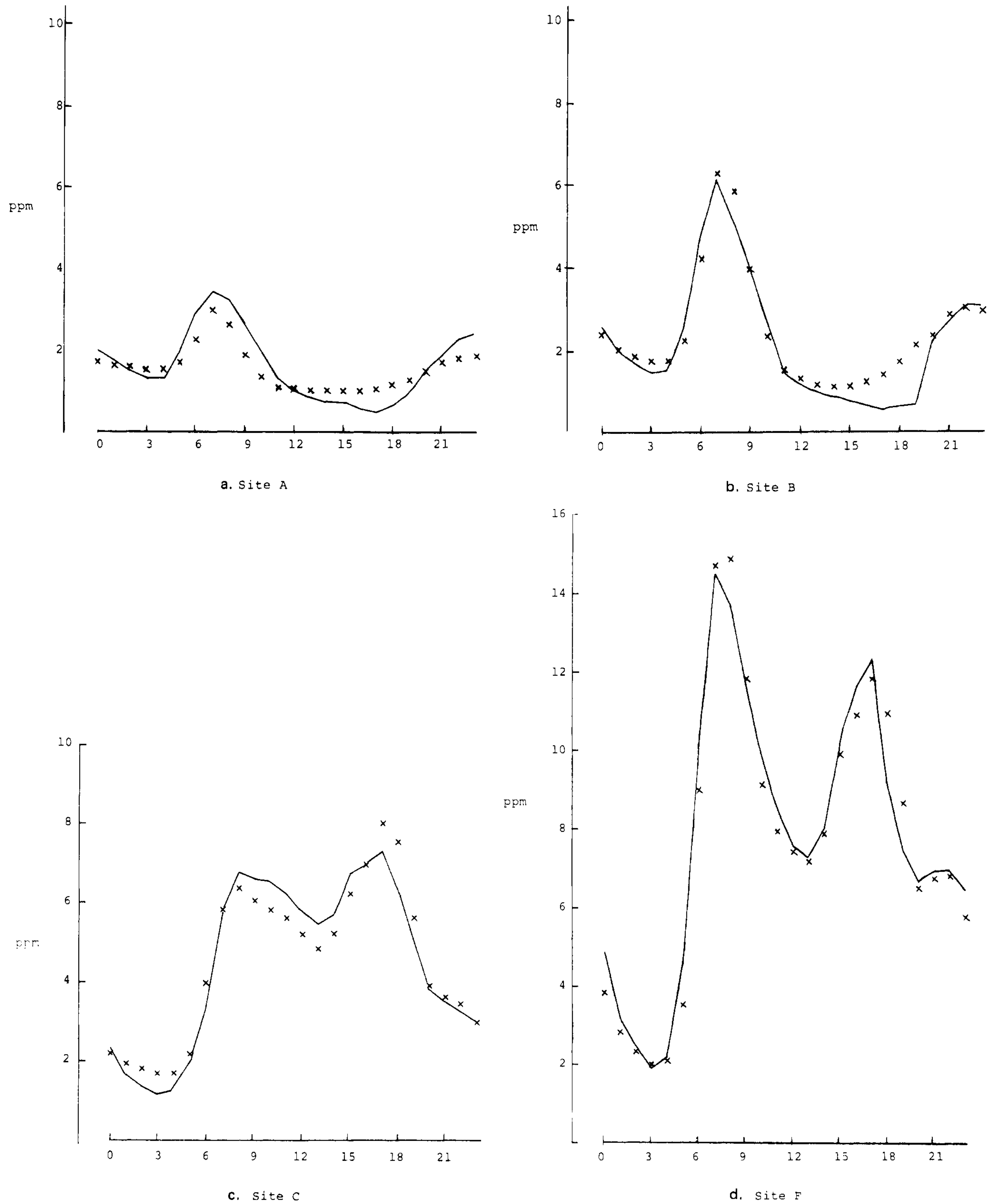

Figure 6. (a–d) Joint CO model summer 1977 weekday fit: (X) predicted; (—) observed

(*iv*) For fixed $WS_{\perp}$, the predicted CO concentrations generally decrease as the distance $|d|$ increases. This decrease is less at downwind sites.

(v) The form of the joint model in Equation 3 is relatively simple and can be applied to data from other areas. The parameters will vary, however, and should be reestimated for new data sets.

Analysis of Pb Data

A Preliminary Graphic Analysis. Since catalyst equipped cars require unleaded gasoline, their use should lead to a reduction in lead concentrations. The Pb data consist of 4-h afternoon high-volume concentrations at sites A and C from June 1974 through November 1977. At sites B and D, data are available only through May 1977. The 4-h afternoon Pb con-

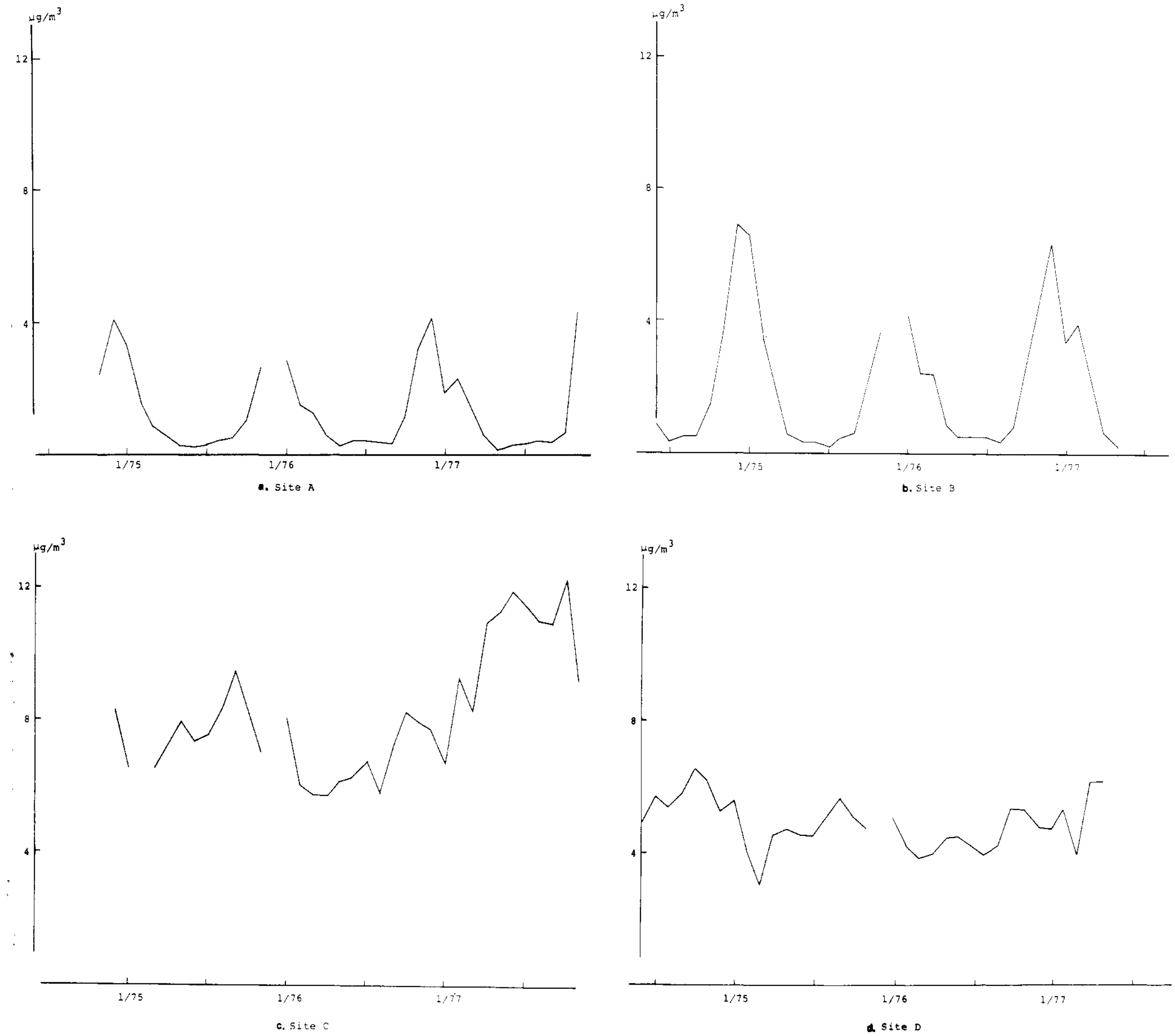

Figure 7. (a–d) Monthly means of 4-h afternoon (3–7 p.m.) Pb concentrations

centrations at locations A, B, C, and D are given in Figures 7a–d. It can be seen that: (i) Pb levels at the downwind sites C and D are substantially higher than those at the upwind sites A and B. (ii) During the summer, when afternoon winds are predominantly blowing across the freeway toward C, the Pb levels at A and B are very low. This indicates that there is very little background Pb and that most of the recorded Pb comes from automobile emissions on the freeway. (iii) The 4-h afternoon Pb concentrations decrease gradually until the end of 1976. From summer 1975 to summer 1976, the reduction is 17% at site C and 11% at site D. Such a decrease is consistent with the increase in catalyst equipped cars. (iv) Starting in February 1977, however, Pb concentrations increase sharply, especially at location C, which is closest to the freeway on the downwind site. At first sight this increase is surprising, since on the contrary one would expect a further decrease of Pb concentrations. However, this sudden increase can be explained by the fact that in February 1977 an additional fifth northbound lane was added to the freeway. This extra lane reduced traffic congestion and led to an increase in the afternoon traffic speed on the northbound lanes. Laboratory studies by Hirschler et al. (*7*) have shown that Pb emissions increase with traffic speed.

Weekday and Weekend Pb Comparison. To further support the above explanation, we compare weekday and weekend afternoon Pb readings in Figure 8. It was reported before (*1, 6*) that on weekends the afternoon Pb concentrations are substantially higher than on weekdays, even though on weekends fewer cars pass through the measurement station. This weekday–weekend Pb difference was attributed to the fact that on weekends the afternoon traffic speed is higher than on weekdays, resulting in increased weekend Pb emissions.

A comparison of traffic data for 1976 and 1977 shows that the average weekday afternoon traffic speed increased from 37 mph in 1976 to 47 mph in 1977. The northbound traffic speed increased even more, from 27 to 47 mph. At the same time there is very little change in the weekend traffic speed between 1976 and 1977. Therefore, the speed difference between weekdays and weekends became smaller in 1977, and thus the difference between weekday and weekend afternoon Pb concentrations should be reduced. This is confirmed by Figures 8a,b for locations C and D.

A Model for 4-h Afternoon Pb at Location C. We now concentrate our analysis on the downwind site C. We shall use daily afternoon averages of traffic count, traffic speed, $WS_\perp$, and Pb from the two periods, (i) August 27 through December 27, 1976 and (ii) September 14 through November 30, 1977,

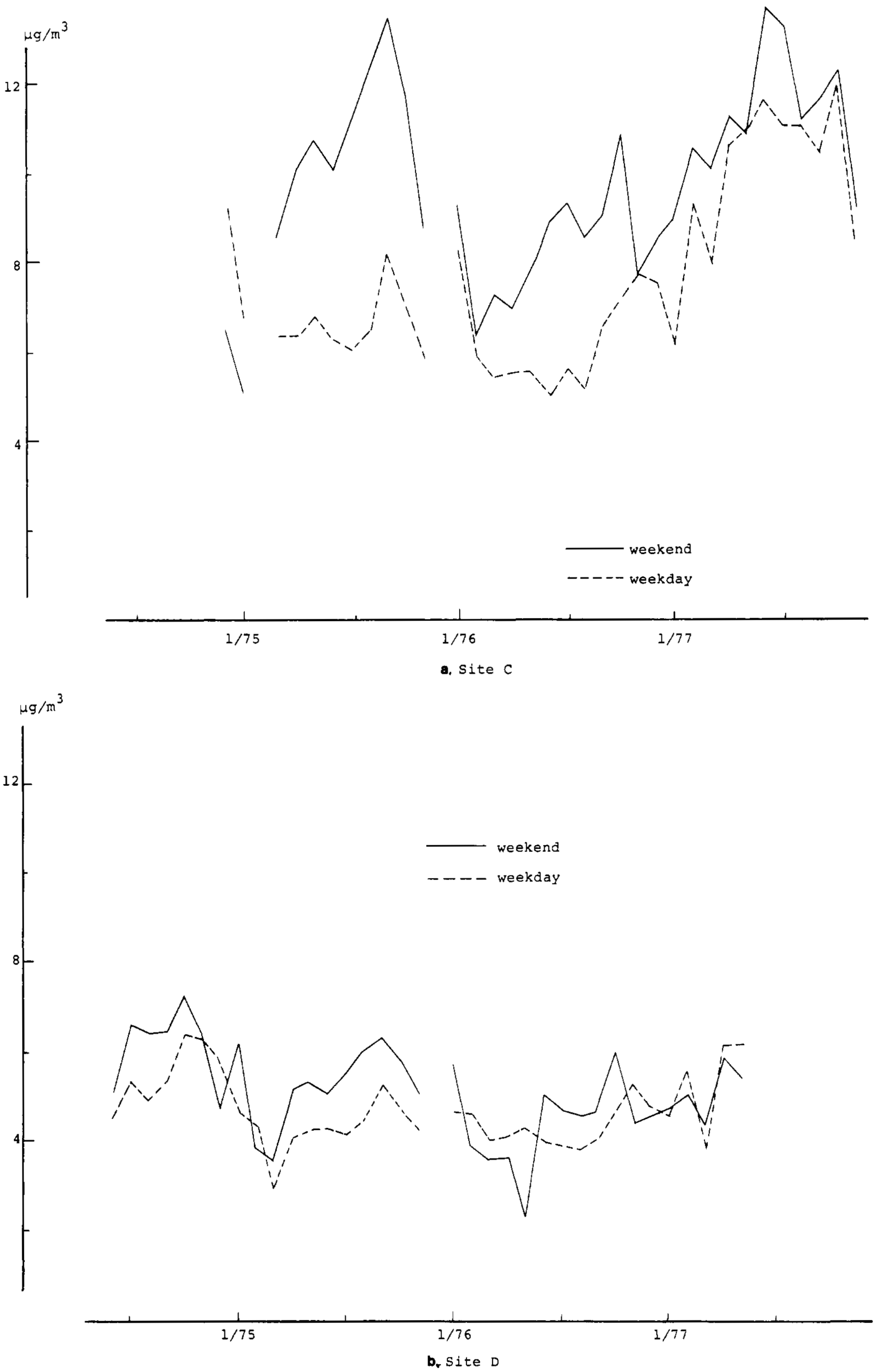

Figure 8. (a,b) Monthly means of 4-h afternoon (3–7 p.m.) Pb concentrations; weekday–weekend comparison: (—) weekend; (- - -) weekday

to quantify the dependence of Pb concentrations on traffic and meteorological variables. The above periods were chosen since reliable data on traffic count and traffic speed were available.

The average of the 4-h afternoon Pb readings for the 1976 period is 7.91 $\mu g/m^3$. For the 1977 period, after the additional northbound lane was opened, it is 10.87 $\mu g/m^3$. The proposed model will help to determine whether this increase can be explained by changed meteorological and/or traffic conditions.

It was mentioned earlier that laboratory studies showed that Pb emissions increase with traffic speed. Such a relationship is also noticed in the observed data. Figure 9 shows a graph of the ratio of afternoon Pb to total afternoon traffic count, [Pb]/TC, vs. the average afternoon traffic speed, TS. The ratio increases roughly in proportion to traffic speed, implying an approximate relationship of the form: $[Pb]/TC = k TS$ or equivalently $[Pb] = k TC \times TS$.

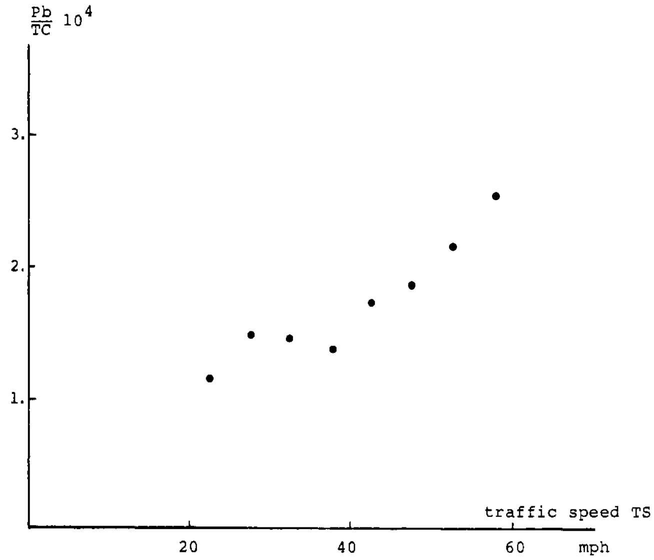

Figure 9. Plot of [Pb]/TC against TS

The wind serves two functions: transport and diffusion. Its influence on Pb concentrations is illustrated in Figure 10, where averages of [Pb]/(TC × TS) are plotted against successive $WS_\perp$ intervals. A parsimonious representation of such a relationship is to take [Pb]/(TC × TS) proportional to the diffusion factor $e^{-b(WS_\perp-\omega)^2}$, where b and ω are appropriate constants. Thus, a model for Pb concentrations can be written as:

$$[\mathrm{Pb}] = k\mathrm{TC} \times \mathrm{TS}e^{-b(\mathrm{WS}_\perp-\omega)^2} \quad (4)$$

In addition, we introduce parameters $k = k_1$ for the 1976 period and $k = k_2$ for the 1977 period. This will allow us to test whether the parameter k, which is proportional to the Pb emissions after adjustment of speed, has changed over time. Furthermore, we separate the effects of northbound and southbound traffic, since, due to the proximity to location C, a larger Pb contribution can be expected from the northbound lane.

Thus we are led to the following model:

$$[\mathrm{Pb}]_t = k_i\{\rho \mathrm{TC}_t^{\mathrm{N}}\mathrm{TS}_t^{\mathrm{N}} + (1-\rho)\mathrm{TC}_t^{\mathrm{S}}\mathrm{TS}_t^{\mathrm{S}}\}e^{-b(\mathrm{WS}_{\perp t}-\omega)^2} + a_t \quad (5)$$

where $[\mathrm{Pb}]_t$ is the observed 4-h afternoon Pb concentration at location C and day t, $\mathrm{TC}_t^{\mathrm{N}}$ ($\mathrm{TC}_t^{\mathrm{S}}$) is the total afternoon northbound (southbound) traffic count, $\mathrm{TS}_t^{\mathrm{N}}$ ($\mathrm{TS}_t^{\mathrm{S}}$) is the average afternoon northbound (southbound) traffic speed, $\mathrm{WS}_{\perp t}$ is the average afternoon perpendicular component of the wind vector, a_t is an error term, k_1 and k_2 are parameters proportional to Pb emissions ($k = k_1$ for the 1976 period; $k = k_2$ for the 1977 period), ρ is a parameter measuring the relative contribution of the northbound traffic lane, and b and ω are parameters in the diffusion factor involving the perpendicular wind component.

Using the available data on Pb, traffic, and $WS_\perp$, the five parameters k_1, k_2, ρ, b, and ω in Equation 5 are estimated by nonlinear least squares. The fitting results are:

parameter	estimate	std error
k_1 (1976)	8.77×10^{-6}	0.39×10^{-6}
k_2 (1977)	8.75×10^{-6}	0.52×10^{-6}
ρ	0.75	0.10
b	0.011	0.0035
ω	2.19	0.56

Diagnostic checks including plots of residuals against fitted values and explanatory variables do not reveal any apparent shortcomings of the above Pb model.

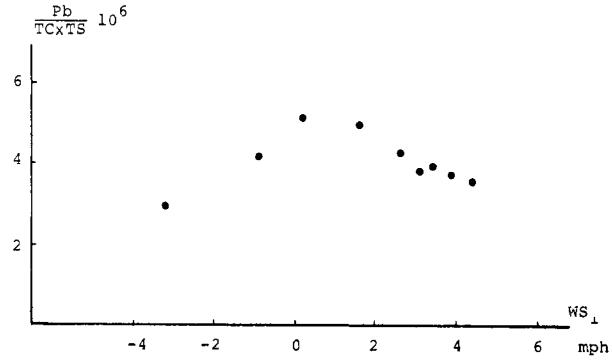

Figure 10. Plot of [Pb]/(TC × TS) against $WS_\perp$

Interpretation of the Pb Model and Conclusions. (i) The parameter estimate of ρ is 0.75. Thus, the northbound lane, which is closer to the receptor at C, exhibits a larger contribution to the Pb concentration than the southbound lane. This is explained by the fact that Pb particles settle fairly quickly.

(ii) The estimates of k_1 and k_2 are not significantly different. This indicates that Pb emissions (adjusted for speed) for 1976 and 1977 are roughly the same, and that the observed 37% increase in the average Pb concentration is due to the change in traffic speed. This example shows clearly that correct trend assessments can only be made when changes in explanatory variables such as traffic and meteorological conditions are properly taken into account.

(iii) The model in Equation 4 differs from the earlier Pb model proposed in ref *1*. Here we consider Pb concentrations proportional to the product of traffic count and traffic speed. Such a relationship is verified both by laboratory experiments (*7*) and by the LACS data. On the other hand, the earlier model treated Pb as proportional to the traffic density, which implies an inverse proportionality of Pb emissions and traffic speed. The trend analysis of Pb given in ref *1* still remains valid, however, because during the period 1974 to 1976 considered there, no appreciable changes in speed occurred. The earlier model would be inappropriate when applied to the 1976 and 1977 trend comparisons.

Acknowledgment

The authors thank G. Hudak, J. T. Hsieh, and S. Graves for computing assistance.

Literature Cited

(1) Tiao, G. C., Hillmer, S. C., *Environ. Sci. Technol.*, **12**, 820 (1978).
(2) Ledolter, J., Tiao, G. C., Hudak, G. B., Hsieh, J. T., Graves, S. B., "Statistical Analysis of the Los Angeles Catalyst Study Data", Technical Report 539, Department of Statistics, University of Wisconsin, 1978.
(3) Evans, G. F., Rodes, C. E., "Summary of LACS Continuous Data", in USEPA-600/4-77-034, June 1977.
(4) Rodes, C. E., Evans, G. F., "Summary of LACS Integrated Pollutant Data", in USEPA-600/4-77-034, June 1977.
(5) Zimmerman, J. R., Thompson, R. S., "Users Guide for HIWAY, A Highway Air Pollution Model", USEPA, Feb 1975.
(6) Phadke, M. S., Tiao, G. C., Hillmer, S. C., "Statistical Evaluation of the Environmental Impact of the Catalytic Converter", Pollution Control Association, Toronto, June 1977.
(7) Hirschler, D. A., Gilbert, L. F., Lamb, F. W., Niebylski, L. M., *Ind. Eng. Chem.*, **49**, 1131 (1957).

Received for review November 27, 1978. Accepted June 14, 1979. This research was supported by the Environmental Protection Agency under Contract No. 68-02-2261.