Analysis of Ambient Carbon Monoxide Data

PAUL E. BENSON

ABSTRACT

Several current methods for estimating worst-case ambient carbon monoxide levels are critically reviewed. The distributions by month, day, and hour of seasonal maximum ambient levels measured at 12 California stations are presented. These distributions are used to develop an observed maximum method for estimating the second annual maximum concentration from limited field measurements. The method is based on the combined use of the binomial distribution and combinatorial analysis. The binomial distribution is used to generate the expected number of occurrences of ambient concentrations within the top six ranks of the seasonal statistics given scheduling and duration of sampling. Combinatorial analysis is used to predict the distribution of seasonal maximums among these top six ranks. The resulting models are verified both separately and together by using the California data. A table is produced that can be used to design sampling plans that will yield observed maximum concentrations equal to or close to the second annual maximum.

The determination of second annual maximum 1-hr and 8-hr ambient carbon monoxide (CO) concentrations from limited field-monitoring data is an important component in transportation air quality impact studies. Because the significance of an air quality impact is judged on the basis of comparison with an absolute standard rather than between alternatives, accurate estimation of the ambient or background concentration is critical. It can often mean the difference between the finding of an acceptable or unacceptable impact. This is particularly true when project-related impacts are small relative to background concentrations. Many highway improvement projects in urban areas fall into this category.

In this paper the problems underlying the current method used by the California Department of Transportation (Caltrans) for extrapolating second annual maximum concentrations from field measurements are examined. A simpler, more accurate scheme is developed in which scheduling and duration of sampling are used to yield a high probability of sampling a value equal to or close to the second annual maximum. Data analysis is reduced by using the maximum value sampled as a direct estimate of the second annual maximum. The new method eliminates overly conservative assumptions and time-consuming analytical procedures.

LITERATURE REVIEW

The method currently used by Caltrans to estimate second annual maximum 1-hr and 8-hr CO concentrations was first introduced by R. I. Larsen in 1971 (1). It was developed empirically from aerometric data collected at eight urban sites from 1962 through 1968. A two-parameter lognormal distribution was used by Larsen to extrapolate expected maximum values from random field measurements. A computerized version of Larsen's model was developed by Caltrans in 1976 (2).

Since its introduction, Larsen's two-parameter lognormal model has been studied and in some ways improved on. The weaknesses of the original model primarily involve three areas:

1. The suitability of the two-parameter lognormal distribution.
2. The implicit assumption that sequential aerometric measurements are independent and evolve from a stationary process, and
3. The requirement that a random sampling scheme be followed.

Several authors, including Larsen, recognized that the two-parameter lognormal distribution was not appropriate for all cases. In 1977 Larsen proposed a three-parameter lognormal distribution for use on data collected at urban and source-affected sites (3). Mage and Ott recommended use of a censored three-parameter lognormal distribution in 1978 (4). In 1975 Curran and Frank proposed the use of a one- or two-parameter exponential distribution fit exclusively to the highest observed concentrations (5).

In 1973, Patel objected to the implicit assumption of independence between sequential aerometric measurements contained in Larsen's model (6). Neustadter and Sidik later showed that the assumption of independence was reasonable for successive measurements made 3 to 6 days apart (7). Horowitz and Barakat concluded that serial correlation between sequential measurements would not seriously limit the usefulness of Larsen's model but that deviations from the implicit assumption of stationarity could (8).

A survey conducted by Meisel and Dushane (9) showed that continuous aerometric sampling over a period of 3 weeks to 3 months was the normal field practice. Random sampling by day or by hour was characterized by respondents as inconsistent with efficient field operations. Respondents found it more efficient to site a sampler for a fixed block of time and sample on a 3- to 5-day weekly schedule. Meisel and Dushane developed an analytical methodology consistent with this type of quasi-continuous sampling plan. Their project, funded by NCHRP, was published as NCHRP Report 200 in 1979 (9).

NCHRP 200 METHOD

The NCHRP methodology is designed to amplify limited project-specific CO data by the use of an auxiliary data set collected concurrently at a nearby, year-round monitoring station. The method assumes a significant temporal correlation between the two sets of data. The auxiliary station data set is used to estimate the number of adverse days sampled at the project site target station. An adverse day is defined as a day containing an 8-hr daily maximum ranking in the upper 20 percent for the year.

The target station data may be analyzed in one of three ways: the distribution, observed maximum, or combination methods. The last is simply a weighted average of results from the first two methods. In the distribution method, an exponential distribution is fitted by least squares to the 8-hr daily maximums measured at the target station during adverse days. The second annual maximum is extrapolated from this distribution. In the observed maximum method, the number of adverse days occurring during the sampling period is used as a prequalification. If there are at least 6 adverse days in a 1-month period or 10 in a 2-month period, the highest 8-hr daily maximum observed during the period is used as an estimate of the second annual maximum. In cases where no auxiliary data are available, NCHRP 200 recommends that sampling periods be prequalified on the basis of nationwide or statewide monthly distributions of adverse days.

During implementation of NCHRP 200 by Caltrans, it became clear that acquiring and processing auxiliary data required an inordinate amount of time and effort. It was decided that the method for determining adverse days from statewide monthly distributions should be followed. However, this approach also presented a problem. The distribution of adverse days by months given in NCHRP 200 contained data from only two regions in California. In addition, possible differences between the monthly distributions of adverse days and second annual maximums had not been investigated. The lack of temporal resolution in the final result also had to be considered. For example, an observed 8-hr maximum that occurred late in the evening might be used as representative of a morning commute-hour ambient. Without organizing the analysis by time of day, similar examples of data mismatching might occur.

These problems and others were addressed by examining data from a number of California monitoring stations and developing several modifications to the original NCHRP 200 method.

ANALYSIS OF AMBIENT CO DATA FROM SELECTED CALIFORNIA STATIONS

In order to help develop and verify the intended modifications to NCHRP 200, a large representative data set was required. Fortunately, historical data from a comprehensive network of air quality monitoring stations throughout California were readily available from the California Air Resources Board in an edited, machine-readable form. Twelve stations with relatively complete records over a period of years were chosen from this data base. They represented a variety of geographic and demographic settings typical of California.

The selected data set was composed of daily records of 1-hr averaged CO concentrations. In cases where missing data were encountered, the NCHRP 200 interpolation method was used. If gaps within a 31-hr period (midnight to 7:00 a.m. of the following day) exceeded the size and frequency criteria set down in NCHRP 200, the entire day was dropped from the data set. After being edited for missing values, the data set was stratified by CO season, starting July 1 and ending June 30 of the following year. The seasonal stratification was made in lieu of a calendar year division so that the monthly distribution of maximums would accurately represent the distribution encountered when sampling was done within a season. Seasons with more than 10 days missing in any single month from October through February or more than 25 days missing over this entire 5-month wintertime period were deleted from the data set. This left a total of 112 station-seasons in the data set, each composed of an average of 349 days' worth of twenty-four 1-hr CO concentrations.

Peak 8-hr averages were determined for each day. Calculations were made by crossing midnight with the start hour of the 8-hr period to determine the date of the maximum. Overlapping 8-hr daily maximums were not permitted. The daily maximums within each station-season were then ranked. Dates and start hours of the top six ranks were retained in the final version of the data set. Tied values were assigned the same rank, making multiple occurrences of seasonal maximums possible. A similar treatment was given to 1-hr maximums, with the exception that multiple annual maximums within the same day were allowed. A description of the final data set is given in Table 1.

The distributions of the seasonal high 8-hr daily maximums by month, day of week, and hour of day were key elements in development of the modified sampling procedure. Instead of using the top 20 percent of the data, as in NCHRP 200, the modified method focused on the probabilities of encountering maximums within the top six ranks of the seasonal statistics. Study was limited to these seasonal maximums rather than adverse days because their temporal distributions were expected to follow the distribution of second annual maximums more closely.

The modified method assumes that the monthly distribution of seasonal maximums is independent of averaging time and rank. A categorical analysis of variance performed on the 12-station data set showed no significant differences between monthly distributions of 1-hr and 8-hr maximum and first- through sixth-ranked seasonal high 8-hr maximums. There was a significant difference in the distribution of maximums by station. However, this was slight enough to justify the aggregation of results over the 12 stations as representative of a composite California location. The final sampling plan was specified by using the aggregated monthly distribution of the proportion of days containing 8-hr seasonal maximums shown in Figure 1.

The distribution of seasonal maximums by day of week, shown in Figure 2, was also important in the development of the modified method. For weekdays the fraction of 1-hr seasonal maximums is somewhat greater than 8-hr maximums. For weekends this difference is reversed. The relatively short duration of weekday traffic peaks and the broader temporal distribution of traffic volumes on weekends is consistent with this pattern. The day-to-day trends in Figure 2 are roughly similar for both 1-hr and 8-hr averaging times. There are gradually more seasonal maximums occurring through the week until a peak is reached on Friday. The number of seasonal maximums then drops significantly for Saturday and reaches a minimum on Sunday. Cross-stratification of the data by time of day and day of week revealed that the additional Friday occurrences, as well as many of the Saturday occurrences, take place in the late evening hours. The few occurrences of Sunday maximums also take place in the evening about 1 hr later than weekday commute peaks.

These temporal patterns exhibited by the seasonal maximums closely follow expected traffic distributions reported by Shirley (10). A composite version of the 1-hr and 8-hr day-of-week distributions was used to determine the probabilities of encountering seasonal maximums associated with different day-of-week sampling plans.

The distributions of 1-hr and 8-hr seasonal maximums by start hour are shown in Figure 3. The distributions are quite dissimilar, particularly regarding the occurrence of morning maximums. The most concentrated number of 1-hr maximums occurs between

TABLE 1 California Ambient CO Data Set

1980 Metropolitan Population (000s)	Station	Area/Site Code	Years Studied	Total Seasons	Second Maximum, 1981-1982 (ppm)	
					1-Hr	8-Hr
<100	Pittsburg	700/430	1969–1982	12	8	4.9
	Lancaster	7000/82	1971–1982	10	9	4.9
	Escondido	8000/115	1975–1982	7	12	8.0
	Santa Barbara	4200/355	1974–1982	7	15	8.1
	Salinas	2700/544	1976–1982	5	4	2.9
100 to 500	Bakersfield	1500/203	1972,1973,1976–1979,1981–1982	5	14	10.1
	Stockton	3900/252	1965–1967,1979–1982	5	14	7.5
	Redwood City	4100/541	1968–1982	13	10	5.5
	Sacramento	3400/282	1972–1980	7	11[a]	7.4[a]
	Pomona	7000/75	1966–1982	14	12	9.6
>500	San Diego	8000/120	1973–1982	8	12	8.6
	Burbank (L.A.)	7000/69	1963–1982	19	25	20.1

[a] 1979-1980 season.

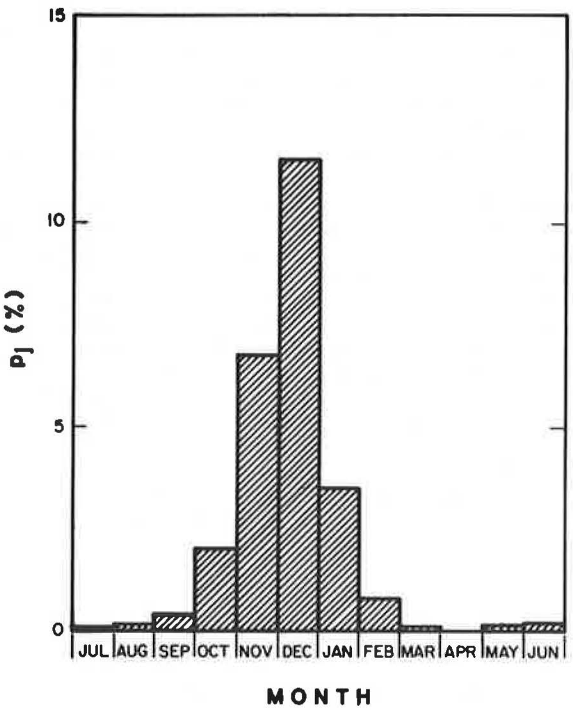

FIGURE 1 Monthly distribution of proportion of days (p_j) containing 8-hr daily maximums within top six seasonal ranks.

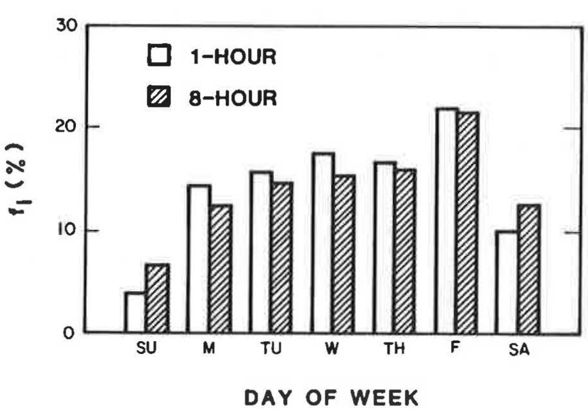

FIGURE 2 Fraction of 1-hr and 8-hr seasonal maximums (f_i) distributed by day of week.

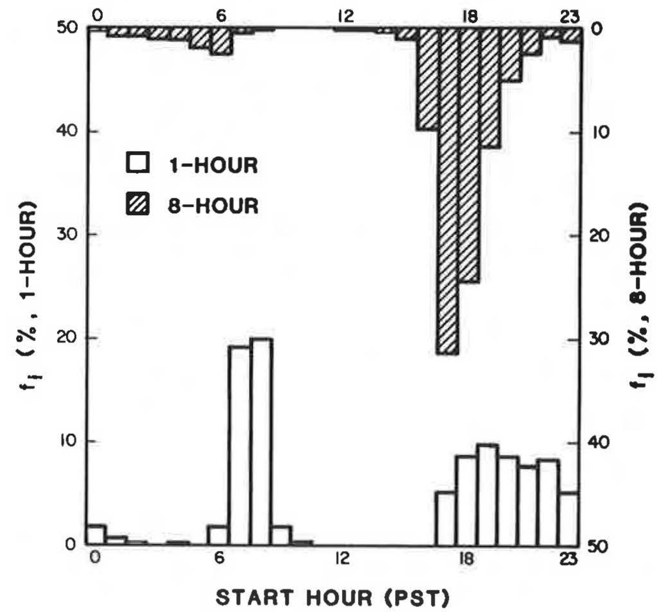

FIGURE 3 Fraction of 1-hr and 8-hr seasonal maximums (f_i) distributed by start hour.

the morning commute hours of 7:00 and 9:00 a.m. Approximately 40 percent of the 1-hr maximums and 10 percent of the 8-hr maximums occur during this period. The explanation for this difference lies in the combined temporal distributions of traffic and meteorology. During the evening hours, these two factors combine over sufficiently long periods to yield the bulk of the 8-hr daily maximums. In the morning hours, a pronounced morning commute and stable meteorological conditions lead to a substantial number of 1-hr seasonal maximums. However, the short duration of the morning commute peak and the rapid shift to unstable meteorological conditions typically following this peak limit the number of morning 8-hr seasonal maximums to a small percentage of the total.

The time-of-day distributions were used to determine an appropriate division of the modified observed maximum analysis into four time periods: morning, midday, evening, and nocturnal (Table 2).

DEVELOPMENT OF A MODIFIED SAMPLING PROGRAM

The fundamental difference between the NCHRP 200 method and the modified method developed in this

TABLE 2 Time Periods for Analysis of Ambient CO Concentrations

Time Period	Start Hour: 1-Hr Maximum	Start Hour: 8-Hr Maximum	Occurrence of Seasonal Maximum (%): 1-Hr	Occurrence of Seasonal Maximum (%): 8-Hr
Morning	6:00	1:00	43	10
Midday	10:00	8:00	1	>1
Evening	5:00	11:00	32	80
Nocturnal	9:00	8:00	24	10

paper involves the procedure for determining the duration of sampling. Instead of sampling for a fixed 30 days with a minimum of 6 adverse days required, the modified procedure calls for a sampling program the duration of which varies with the month or months sampled. The duration of sampling is chosen so as to yield an extremely high probability of attaining as an observed maximum an unbiased estimate of the expected second annual maximum. This is achieved through the combined application of the binomial distribution and combinatorial analysis. The specific sampling intervals recommended in this paper are based on the California data set. However, the principles can be extended to any comparable data set.

A distribution of randomly chosen, independent events characterized by two mutually exclusive outcomes can be described by the binomial expansion $(q + p)^n$, where q and p represent the probabilities of occurrence attached to each outcome. The rth term of the expansion equals the probability that the outcome, the underlying probability of which is denoted by p, will occur r times in n samples. This can be stated as follows:

$$P(r|p,n) = [n!/r!(n - r)!]\ p^r q^{n-r} \quad (1)$$

The binomial expansion was used to generate expected monthly probabilities of encountering r seasonal maximums (defined as daily 8-hr maximums within the top six ranks for the season) in an n-day sampling period based on the underlying probabilities shown in Figure 1. Thus, for a full-month sample taken during the jth month,

$$P(r|j) = P(r|p_j,n_j) \quad (2)$$

where p_j and n_j equal, respectively for the jth month, the probability of encountering seasonal maximums and the number of days in the month. Equation 2 was used to predict the distribution of occurrences of seasonal maximums for the full-month sampling periods of October through February. These are compared in Table 3 with the observed distributions taken from the California data set. Probabilities are rounded to the nearest whole percent, so totals may not exactly equal 100 percent.

Use of the binomial distribution assumes that the 8-hr daily maximums are randomly chosen, independent events. In fact, they are a set of sequentially sampled, autocorrelated events. The assumption of randomness is not seriously violated provided sampling is of sufficient duration to incorporate a majority of winter meteorological conditions. The assumption of independence between daily maximums presents a more serious problem, however. Examination of the California data set showed that clusters of consecutive seasonal maximums occur with greater frequency than would be expected from a series of independent events. This was most evident for small clusters of two to three seasonal maximums; 26 percent of the paired values and 8 percent of the groups of 3 occurred on successive days.

Clustering of seasonal maximums is caused by short periods of calm, stable meteorological conditions between winter storms. The effect of clustering on the overall distribution of seasonal maximums is apparent in Table 3. There is a consistently higher percentage of months with no occurrences of seasonal maximums than would be expected from a truly independent distribution. This higher percentage is caused by the clustering of maximums in other months. By the same token, the overestimation of months with only one occurrence can be attributed to the likelihood that seasonal maximums will occur in clusters rather than as isolated events.

The binomial distribution does reasonably well at predicting the observed pattern for months having two or more occurrences as well as for the entire distribution for November. Therefore, the assumptions of independence and randomness, although not entirely valid, were considered satisfactory for purposes of approximating the number of seasonal maximums (r) within a given sampling interval.

The probability that an observed maximum equals the mth-ranked seasonal maximum, given a sample containing r seasonal maximums, can be stated as follows:

$$P(m|r) = [{}_{(\ell + 1)}C_r - {}_{\ell}C_r]/{}_6C_r \quad (3)$$

where ℓ is the number of ranks less than the mth rank (6 - m) for the r seasonal maximums and

$${}_nC_r = \begin{cases} n!/r!(n - r)! & n \geq r \\ 0 & n < r \end{cases}$$

This formulation assumes that there are no multiple occurrences of seasonal maximums. For the California data set there were tied ranks, however, yielding an average of eight maximums within the top six ranks

TABLE 3 Observed and Predicted Probabilities of Encountering r Seasonal Maximums by Month

No. of Seasonal Maximums (r)	Oct. Observed	Oct. Predicted	Nov. Observed	Nov. Predicted	Dec. Observed	Dec. Predicted	Jan. Observed	Jan. Predicted	Feb. Observed	Feb. Predicted
0	64	52	14	12	9	2	45	33	82	79
1	19	34	26	27	6	9	24	37	14	19
2	10	11	26	28	13	18	15	20	2	2
3	4	2	21	19	22	22	11	7	1	0
4	3	0	7	9	17	20	5	2	1	0
5	1	0	4	3	15	14	1	0		
6			2	1	10	8				
7			1	0	5	4				
8					2	1				
9					1	0				

per season. The following modification to Equation 3 was derived to account for this:

$$P(m|r) = (1/{}_8C_r)\ \{(q^2/36)[{}_{(\ell+1)}C_r - {}_{\ell}C_r] + (q/18)[{}_{(\ell+2)}C_r - {}_{\ell}C_r] + (q\ell/18)[{}_{(\ell+2)}C_r - {}_{(\ell+1)}C_r] + (1/36)[{}_{(\ell+3)}C_r - {}_{\ell}C_r] + (\ell/18)[{}_{(\ell+3)}C_r - {}_{(\ell+1)}C_r] + (\ell^2/36)[{}_{(\ell+3)}C_r - {}_{(\ell+2)}C_r]\} \quad (4)$$

where g is the number of ranks greater than the mth rank (m − 1).

Equation 4 is based on the assumption that the two extra seasonal maximums are randomly distributed among the top six ranks. Overall, there is a slight increase with descending rank of the number of ties in the California data set. However, for months containing three or more seasonal maximums, the distribution of tied ranks is approximately uniform among the top six ranks.

Equation 4 was used to model the distribution of observed maximums among the top six ranks as a function of the number of occurrences of seasonal maximums. These values can be compared in Table 4 to the rank distribution of monthly observed maximums categorized by number of occurrences (r). The discrepancies between observed and predicted probabilities are primarily due to the nonuniform distribution of tied ranks and the tendency for first and second seasonal maximums to be associated with clustered results.

To test the validity of the combined use of the binomial distribution and Equation 4, the modeled distributions of observed maximums by rank for the months October through February were generated by the following:

$$P(m|j) = \sum_{r=1}^{7} [P(r|j) \cdot P(m|r)] + [P(r \geq 8|j) \cdot P(m|r=8)] \quad (5)$$

where $P(m|j)$ equals the probability of the observed maximum during the jth month coming from the mth rank. The second term in Equation 5 relates the diminishing probabilities generated by the bionomial distribution for $r \geq 8$ to the fixed probability for $r = 8$ derived from the combinatorial analysis. The predictions generated by Equation 5 compare favorably with the observed distributions obtained from the California data set (Table 5).

In general terms, the combined model can be expressed as follows:

$$P(m|p_j,n) = \sum_{r=1}^{n} [P(r|p_j,n) \cdot P(m|r)] \quad (6)$$

where n is the number of days sampled in the jth month and $P(m|r) = P(m|r = 8)$ for all $r \geq 8$. Given known values for p_j and the average differences between concentrations by rank from the California data set, Equation 6 can be used to approximate an unbiased sampling program. The values for p_j are given in Figure 1, whereas the distribution of differences between the second seasonal maximum (m = 2) and the first through sixth seasonal maximums for the California data set are summarized in Table 6. The maximum and overall average differences are given as follows:

Seasonal Maximum (r)	$\|d_{m-2}\|_{max}$ (ppm)	$\|\bar{d}_{m-2}\|$ (ppm)
1	8.6	1.27
3	7.2	0.69
4	7.7	1.15
5	7.9	1.52
6	8.4	1.85

TABLE 4 Observed and Predicted Probabilities That Observed Maximum Will Equal mth Seasonal Maximum Given r Occurrences

Rank (m)	Probability of Occurrence (%) by No. of Seasonal Maximums (r): 1		2		3		4		5		6		7		8+	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
1	14	17	23	32	44	46	56	59	74	71	77	82	100	91	100	100
2	14	17	20	26	25	29	25	28	17	23	15	17	0	9		
3	11	17	27	20	18	16	17	10	9	5	0	2				
4	23	17	13	14	9	7	3	2	0	0	8	0				
5	19	17	13	7	4	2	0	0								
6	19	17	4	1	0	0										

Because there was no significant difference between the monthly distributions of 1-hr and 8-hr seasonal maximums, a sampling program based on 8-hr values was assumed equally valid for 1-hr estimates.

The sampling program was designed so that the duration of sampling would be sufficient to guarantee a fixed probability (P_c) of obtaining one or

TABLE 5 Observed and Predicted Probabilities by Month That Observed Maximum Will Equal mth Seasonal Maximum

Rank (m)	Probability of Occurrence (%) by Month									
	Oct.		Nov.		Dec.		Jan.		Feb.	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
1	9	10	27	31	51	50	16	17	3	4
2	4	9	25	21	15	23	7	14	3	4
3	5	8	13	14	13	12	11	12	3	4
4	8	7	9	10	8	6	9	10	3	4
5	6	7	6	7	4	3	8	8	3	3
6	4	6	6	5	1	2	5	6	5	3

TABLE 6 Distribution of Differences Between Second and mth Seasonal Maximums

$\|d_{m-2}\|$ (ppm)	Probability of Occurrence (%) by No. of Seasonal Maximums (r)				
	1	3	4	5	6
≤0.5	38	62	30	14	6
0.6-1.0	22	21	29	29	25
1.1-1.5	17	10	22	20	21
1.6-2.0	6	2	7	17	17
2.1-2.5	6	1	4	5	11
2.6-3.0	2	2	3	4	6
>3.0	9	3	5	11	14

more seasonal maximums. To select a proper value for P_c, the probabilities described by Equation 6 were used in combination with the average differences ($\bar{d}_{m-2}$) given earlier. Because the binomial distribution turned out to be relatively insensitive to values of p_j ranging from 0.01 to 0.10 given a fixed value of P_c, an averaged value ($\bar{p}$) of 0.05 was used in the following final design equation:

$$B = \sum_{r=1}^{7} \left\{ P(r|\bar{p},P_c) \cdot \sum_{m=1}^{6} [P(m|r) \cdot \bar{d}_{m-2}] \right\} + [P(r \geq 8|\bar{p},P_c) \cdot \bar{d}_{1-2}] + [(1 - P_c) \cdot \bar{d}_o] \qquad (7)$$

In this equation, B equals the expected bias in parts per million given P_c, and $\bar{d}_o$ represents an estimate of the average difference between the second seasonal maximum and observed maximums occurring outside the top six-rank interval. In cases where a seasonal maximum is not encountered during the sampling period, there is still a high probability (P_c') that a daily maximum within the top 12 ranks will be found. It can be shown that

$$P_c' = 1 - \exp [\ln(1 - P_c)\ln(1 - 2\bar{p})/\ln(1 - \bar{p})] \qquad (8)$$

assuming that the average underlying probability of encountering a daily maximum within the top 12 ranks is twice the probability of encountering a maximum in the top 6 ranks. Thus, the selection of the design probability (P_c) was based on B approaching zero in Equation 7 and P_c approaching 1 in Equation 8. The value for $\bar{d}_o$ was determined by extrapolating values of $\bar{d}_{m-2}$ for m = 7 to 12 by using the average differences given earlier and compositing these values as follows:

$$\bar{d}_o = \sum_{r=1}^{7} \left\{ [P(r|\bar{p},P_c)/P_c] \cdot \sum_{m=7}^{12} [P(m-6|r) \cdot \bar{d}_{m-2}] \right\} + \left\{ [P(r \geq 8|\bar{p},P_c)/P_c] \cdot \bar{d}_{7-2} \right\} \qquad (9)$$

Equation 9 deals essentially with the small probability ($P_c' - P_c$) that an observed maximum will fall outside the top 6 ranks but within the top 12 ranks for the season. The same combined probabilities used to model the distribution of the top 6 ranks are used to weight the values of $\bar{d}_{m-2}$ for m = 7 to 12 when the composite result ($\bar{d}_o$) is developed.

By trial and error, a design value for P_c of 0.93 was derived. This yields a value of 0.995 for P_c'. Simply stated, this means that given a sampling period of sufficient duration to assure a 93 percent chance of encountering at least one seasonal maximum, one can be 99.5 percent confident that the observed maximum is an unbiased estimate of the expected second annual maximum.

The distribution of differences given in Table 6 represents a wide range of exposures. The average differences were used in this paper strictly for the purpose of approximating an unbiased sampling program. It was assumed that

$$\bar{d}_{1-2}/\bar{d}_{m-2} = \text{constant} \qquad (10)$$

for each value of m regardless of location, season, or averaging time. The absolute random error that one can expect in terms of parts per million for any given location will depend on the magnitude of the seasonal maximums at that location, not the average differences derived from the California data set.

RECOMMENDED SAMPLING PLAN

Scheduling and duration of sampling are the key elements in the modified observed maximum method. They are used to minimize bias and to assure a reasonable probability of encountering a maximum value equal to or near the second seasonal maximum. Sample scheduling determines the probability (p_{jk}) of encountering a seasonal maximum given the jth month and the kth day-of-week sampling plan. Sampling duration determines the probability of encountering one or more seasonal maximums given p_{jk}. If the probability $[P(r \geq 1|p_{jk})]$ equals the design probability (P_c), the observed maximum represents an unbiased estimate of the expected second seasonal maximum.

To facilitate selection and design of sampling plans, a table listing values of $P(r=0|p_{jk}) = 1 - P(r \geq 1|p_{jk})$ for 1-week periods as a function of month and days sampled during the week was constructed. The following simplified form of the binomial distribution for r = 0 was used to compute these probabilities:

$$P(r=0|p_{jk},n_k) = (1 - p_{jk})^{n_k} \qquad (11)$$

where n_k is the number of days sampled per week.

Values of p_j taken from Figure 1 were modified according to the following equation:

$$p_{jk} = p_j \cdot \left[7 \sum_{i=1}^{7} (D_{ik} \cdot \bar{f}_i) / \sum_{i=1}^{7} D_{ik} \right] \qquad (12)$$

where $\bar{f}_i$ equals the average probability from Figure 2 of encountering a seasonal maximum on the ith day, and D_{ik} = 1 if the ith day is included in the kth day-of-week sampling plan or D_{ik} = 0 if it is not. Equation 12 is simply a means of accounting for the significant difference in the distribution of seasonal maximums by day of week shown in Figure 2. Eight day-of-week sampling plans were considered. The results are given in Table 7.

TABLE 7 Probability of Encountering Zero Seasonal Maximums in a 1-Week Sampling Period by Month and Day-of-Week Sampling Plan

Days Sampled	$P(r=0 \mid p_{jk})$				
	Oct.	Nov.	Dec.	Jan.	Feb.
M–W	0.94	0.80	0.68	0.89	0.97
Tu–Th	0.93	0.79	0.66	0.89	0.97
W–F	0.92	0.76	0.62	0.87	0.97
M–Th	0.91	0.74	0.59	0.86	0.96
Tu–F	0.90	0.71	0.55	0.84	0.96
M–F	0.88	0.66	0.49	0.81	0.95
M–Sa	0.87	0.63	0.44	0.79	0.94
M–Su	0.86	0.61	0.43	0.78	0.94

To use Table 7, one simply selects the entry or entries for the month or months and day-of-week sampling plan or plans being considered. Treating each probability as independent, the combined probability of encountering zero seasonal maximums over a given sampling period will be the product of the individual probabilities taken from Table 7. For instance, if a proposed sampling plan calls for three weeks of M-F sampling in December followed by two weeks of Tu-F sampling and one week of M-Sa sampling in January, the combined probability of encountering zero seasonal maximums would be given by

$$P(r=0) = (0.49)^3 \cdot (0.84)^2 \cdot (0.79) = 0.07.$$

The criterion for accepting a proposed sampling plan is

$$\prod_{j,k} P(r=0 \mid p_{jk})^{w_{jk}} \simeq 1 - P_c \qquad (13)$$

where w_{jk} equals the number of weeks the kth day-of-week sampling plan will be repeated in the jth month. Because P_c = 0.93, the sample plan cited earlier meets this criterion.

By using Table 7 and Equation 13, a field supervisor can choose a sampling plan that will yield as an observed maximum a relatively unbiased estimate of the expected second seasonal maximum. If the need arises, a prearranged plan can even be changed midstream and still meet the criterion stated in Equation 13. After sampling is concluded and the data checked for outliers, the 1-hr and 8-hr observed maximums by time period can be considered accurate estimates of their respective second annual maximums.

CONCLUSION

Development of the sampling criterion specified by Table 7 and Equation 13 was presented in general form so that it might be applied to a variety of situations. For locations where the monthly distribution of seasonal maximums differs significantly from the 12-station California data set, a more appropriate version of Table 7 could be constructed from local aerometric data by using the same design equations. In cases where the duration of sampling falls short of the recommended period because of time, funding, or staff limitations, the probability of encountering the second or higher seasonal maximum can still be determined by using Equation 6. Proposed revisions to the National Ambient Air Quality Standards for CO by increasing the allowable number of exceedances to five per year (11) can also be accommodated by shifting the reference rank in Equation 7 from the second to the sixth seasonal maximum and modifying Equation 4 for use with either the second or third observed maximum. If the proposed revisions are adopted, sampling-duration requirements would probably be reduced.

The observed-maximum method recommended in NCHRP 200 and the modified sampling procedure developed in this paper will soon be implemented by Caltrans. It is anticipated that the new procedures will save considerable time in the collection and analysis of aerometric data for project-level transportation air quality studies. In addition, more accurate estimates of ambient maximums are expected.

ACKNOWLEDGMENT

This report was produced as part of a research project funded by FHWA. The author would like to express his appreciation to Gary Honcoop of the California Air Resources Board for providing the 12-station California data base, Jim Quittmeyer for performing numerous analyses of the data on the computer, and Earl Shirley for his review and interest in the final document.

REFERENCES

1. R. I. Larsen. A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards. Report AP-89. U.S. Environmental Protection Agency, Nov. 1971.
2. P. D. Allen, A. J. Ranzieri, and E. C. Shirley. Application of Larsen's Mathematical Model for Air Quality Studies. Report TL-7082-12-76-25. Transportation Laboratory, California Department of Transportation, Sacramento, 1976.
3. R. I. Larsen. An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions, Part 4: A Three-Parameter Averaging-Time Model. Journal of the Air Pollution Control Association, Vol. 27, May 1977, pp. 454-459.
4. D. T. Mage and W. R. Ott. Refinements of the Lognormal Model for Analysis of Aerometric Data. Journal of the Air Pollution Control Association, Vol. 28, Aug. 1978, pp. 796-798.
5. T. C. Curran and N. H. Frank. Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations. Presented at the 68th Annual Meeting of the Air Pollution Control Association, Boston, Mass., 1975.
6. N. R. Patel. Comment on a New Mathematical Model of Air Pollution Concentration. Journal of the Air Pollution Control Association, Vol. 23, April 1973, pp. 291-292.
7. H. E. Neustadter and S. M. Sidik. On Evaluating Compliance with Air Pollution Levels Not to be Exceeded More Than Once a Year. Journal of the Air Pollution Control Association, Vol. 24, June 1974, pp. 559-563.
8. J. Horowitz and S. Barakat. Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Non-Stationarity. Atmospheric Environment, Vol. 13, 1979, pp. 811-818.
9. W. S. Meisel and T. E. Dushane. Monitoring

Carbon Monoxide Concentrations in Urban Areas. NCHRP Report 200. TRB, National Research Council, Washington, D.C., April 1979.
10. E. C. Shirley. Traffic Information Requirements for Estimates of Highway Impact on Air Quality. Report MR657082S(3)-72-09. California Division of Highways, Sacramento, 1972.
11. A. Samberg and J. Elston. The Proposed Revision to the NAAQS for Carbon Monoxide and Its Effect on State Designation. Journal of the Air Pollution Control Association, Vol. 32, Dec. 1982, pp. 1228-1230.

Publication of this paper sponsored by Committee on Transportation and Air Quality.

Cost-Effectiveness Model for the Analysis of Trade-Offs Between Stationary and Transportation Emission Controls in Baltimore

ARMANDO M. LAGO, SALVATORE BELLOMO, KEVIN HOLLENBECK, SAUD SIDDIQUE, and JOE MEHRA

ABSTRACT

The application of a cost-effectiveness model for the attainment of ozone standards in Baltimore is described. Cost-effectiveness programs for Baltimore are designed taking into account direct implementation costs and user costs. The mix of controls in the cost-effective solution varies when either direct implementation costs or social costs are considered. The economic and social impacts of the cost-effective solutions are discussed. Finally, the results of the Baltimore application are contrasted with the results of an earlier study in Philadelphia.

Under the provisions of the Clean Air Act (40 CFR 50, revised July 1, 1980) each state must prepare a state implementation plan (SIP) for meeting air quality goals. The SIP, which is usually prepared by a designated metropolitan planning organization (MPO), contains programs for control of mobile sources of air pollution [including transportation control measures (TCMs)] and stationary sources to meet the air quality emission goals. However, the plan may also consider other important socioeconomic, mobility, and environmental factors in the design and choice of pollution abatement and control strategies. A review of SIPs conducted by BKI Associates, Inc. (1) found that the SIP planning methodologies were applied separately to control of transportation sources and to stationary sources, a procedure that limited the opportunities for coordination and trade-off of mobile-source and stationary-source controls with the concomitant loss of information and opportunities for optimization of the strategies contained in the SIPs. The results of the development and application of a cost-effectiveness model for the analysis of trade-offs between controls of stationary sources and transportation sources for hydrocarbons in the Baltimore standard metropolitan statistical area (SMSA) are summarized.

BASIC COST-EFFECTIVENESS CONCEPTS

Cost-effectiveness analysis provides an efficient method for coordinating and trading off stationary-source and mobile-source control options. The method consists of defining a measure of effectiveness (MOE), in this case the reduction of hydrocarbon (HC) emissions, and then estimating costs of abatement control per unit of HC removed. Abatement strategies in specific pollutant-emitting industries are next ranked in terms of cost-effectiveness ratios (i.e., dollar costs divided by units of HC removed). Then, given an objective of total HC reductions derived from air quality standards for the region, the least-cost package of abatement strategies for meeting the standard is selected by picking those strategies with lowest cost per unit of HC removed and avoiding the higher-cost strategies and alternatives. Designing the least-cost package, sometimes called the cost-effective package, enables environmental planners to consider and trade off abatement strategies, such as those for stationary point sources versus those for mobile sources, an important feature sometimes lacking in the methodology used in developing the SIPs.

At the outset it should be noted that cost-effectiveness analysis per se is neutral with respect to the definition of the target level of emission reductions. In addition, cost-effectiveness analysis assumes that the effectiveness target is valuable in