SANTA FE RESEARCH CORPORATION

DEVELOPMENT AND APPLICATION OF METHODS FOR ESTIMATING INHALABLE AND FINE PARTICLE CONCENTRATIONS FROM ROUTINE HI-VOL DATA

EXECUTIVE SUMMARY

by

John Trijonis
and
Marilyn Davis

December 1981

Submitted to: California Air Resources Board
P.O. Box 2815
Sacramento, California 95812

Dr. Douglas Lawson, Project Officer

Contract Number AO-076-32

The statements and conclusions in this report are those of the Contractor and not necessarily those of the State Air Resources Board. The mention of commercial products, their source or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

INTRODUCTION

The existing California and Federal Ambient Air Quality Standards for atmospheric particles pertain to the mass concentration of total suspended particulate matter (TSP). As measured by the Hi-Vol sampler, TSP consists of the mass of particles less than approximately 50 microns in diameter. There is growing recognition both nationally and in California that particulate standards based on TSP alone are inadequate. An increasing body of scientific evidence has established that the health and welfare effects of particles depend significantly on the size distribution and chemical composition of the particles. As an important step in refining and improving air quality standards for particulate matter, both the EPA and California ARB are considering a revision of the standards that will take into account the most significant aspects of particle size distributions.

Physiological experiments have established the basic relationship between particle size distribution and the penetration/deposition properties of particles within the human respiratory system. Typically, only those particles smaller than 10 or 15 microns in size reach the lower respiratory tract, and only those particles less than 2 or 3 microns in size penetrate to the deepest part of the lungs, the alveoli. Because of this relationship between particle size and penetration in the respiratory system, a formal distinction has been made of inhalable particles (IP), those less than 15 microns in diameter, and fine particles (FP), those less than 2.5 microns in diameter (Miller et al. 1979).

Similarly, many of the important welfare effects of particles depend on their size. For example, particle light scattering, which is usually the dominant component of regional hazes, basically arises from those particles in the size range of 0.1 to 1.0 micron. Fortunately for the purpose of simplicity, the mass of particles in the 0.1 to 1.0 micron range is nearly the same as FP mass, because nearly all the particle mass less than 2.5 microns resides in a mode (called the accumulation mode) between 0.1 and 1.0 micron.

Within the past 2 or 3 years, EPA, the State of California, and other agencies have started to collect ambient data on IP and FP using dichotomous particulate samplers with particle size cut-offs at 15 μm and 2.5 μm. Because the dichotomous sampling networks are so new, the geographical coverage and historical time coverage of the dichotomous data are small compared to the spatial and temporal coverages of Hi-Vol data. In order to ease the expansion from TSP standards and monitoring to TSP/IP/FP standards and monitoring, there is a pressing need for simple empirical formulae that can be used to compare the new dichotomous data with the Hi-Vol data. In this study, we develop, evaluate, and apply simple equations for estimating IP and FP from Hi-Vol data for TSP, sulfates ($SO_4^=$), and lead (Pb).

Several researchers have investigated the statistical relationships between TSP and IP or FP (Trijonis et al. 1980; Spengler et al. 1980; Wendt and Torre 1981; Feldman et al. 1981; Evans et al. 1981). The present study extends these previous investigations in several major respects. First, each of the previous studies has examined only a limited number of sites in a restricted geographical area: ten sites in St. Louis (Trijonis et al. 1980), 11 sites in California (Wendt and Torre 1980), and six sites in the Eastern U.S. (Spengler et al. 1980; Feldman et al. 1981; Evans et al. 1981). In this study, we use simultaneous dichotomous sampler data and Hi-Vol sampler data at 75 locations nationwide, including 11 sites in California. Second, because of the limited number of sites examined, the previous studies could not address geographical or site-type variations in the relationship between dichotomous data and Hi-Vol data. This study does quantify the geographical, site-type, and seasonal variations in the relationships. Third, prior studies have been restricted to just univariate analyses, e.g. IP versus TSP, or FP versus TSP. In this study, we perform multivariate analyses relating IP or FP to Hi-Vol data for TSP, $SO_4^=$, Pb, and NO_3^- (although the NO_3^- variable is later excluded from our recommended equations). The addition of $SO_4^=$ and/or Pb to the equations is important because these parameters provide information concerning the particle mass in the fine aerosol mode. Fourth, the prior studies have focused on purely statistical relationships. Our hybrid approach -- physico-chemical and empirical -- makes the equations more credible,

adaptable, and interpretable, while losing essentially no accuracy compared to the best-fit statistical approach. Finally, previous studies have not proceeded to the application phase. This study includes a major application using five years of Hi-Vol data at 226 California sites; also, the ARB staff has begun applying our formulae to historical health effects studies.

DATA BASE AND METHODOLOGY

After reviewing various monitoring networks that provide dichotomous data for IP and FP as well as Hi-Vol data for TSP, $SO_4^=$, Pb, and NO_3^-, we select two of them -- the EPA IP Network and the ARB Network -- as being most appropriate for this study. The EPA Network provides, by far, the greatest amount of data nationwide; it is also most pertinent to our planned applications because it contains several sites where historical health effects studies have been performed and several sites in California. The ARB Network is chosen to enlarge the data base for California as much as possible.

A data quality screening procedure for the dichotomous and Hi-Vol data is formulated based on both physical and statistical considerations. Application of the data quality screen eliminates about 5% of the EPA Network data but none of the ARB Network data.

The final, quality-screened data base contains 930 simultaneous 24-hour measurements of IP, FP, TSP, $SO_4^=$, Pb, and NO_3^-. These data are from 75 sites nationwide, including 11 in California. For the purposes of our analysis the sites are organized into eight geographical regions: three regions in California (San Francisco Bay Area, Central Valley, and Los Angeles Area) and five other regions nationwide (Pacific Northwest, Arid Southwest, North Central, Northeast, and Southeast). The study locations are also classified according to three site types: metropolitan, suburban, and nonurban.

A special "annual" data base is constructed using sites that have a full year (or nearly a full year) of data. The annual data base is used to verify that seasonal biases are absent from our statistical results and to evaluate the errors in our predictive equations as applied to annual means. A special data base is also constructed for comparing SSIP (IP data taken using Hi-Vols with size-selective inlets) to the routine Hi-Vol data.

The linear equations relating IP or FP to the Hi-Vol parameters are derived using a hybrid approach. The $SO_4^=$ and Pb coefficients are based on physico-chemical principles. The coefficient for TSP (or NO_3^-) is determined statistically by a least-squares regression analysis.

The hybrid approach offers important advantages. Credibility is enhanced because the number of free coefficients is reduced and because physical meaning is attached to the coefficients in the equations. Prespecifying the $SO_4^=$ and Pb coefficients not only assigns these coefficients well-defined physical meanings but also makes the TSP coefficient more easily interpretable on physical grounds. The hybrid equations are also much more adaptable to new situations -- other locations and other years. These advantages are accrued at very little loss in predictive accuracy; the hybrid equations produce negligible increases in error compared to best-fit statistical equations.

The predictive equations for IP and FP are developed in a stepwise manner, adding Hi-Vol parameters (the independent variables) in the order: TSP, $SO_4^=$, Pb, and NO_3^-. The stepwise analysis permits greater flexibility in applying the results (depending on data availability) and yields an assessment of how accuracy improves as each variable is added.

The statistical measures used to evaluate the equations are the degree of correlation and the standard error in predicting IP or FP. Because the standard error appears to increase nearly in proportion with the magnitude of IP or FP, we specify the standard error as percentage errors rather than absolute errors.

RELATIONSHIP OF IP AND FP TO HI-VOL DATA

On a national aggregate basis, the equation relating IP to TSP is IP = 0.61 TSP (see Table 1). This equation yields a 31% standard error in predicting annual mean values of IP (see Table 2). The error in predicting IP can be reduced slightly by adding the variables $SO_4^=$, Pb, and NO_3^- (i.e. proceeding down the columns of Table 2). If $SO_4^=$ and Pb data are available, it is reasonable to use equations containing these variables. However, we recommend against using equations with the NO_3^- variable because the coefficient for NO_3^- turns out to be unstable and statistically insignificant.

TABLE 1 NATIONAL AND (CALIFORNIA) REGIONAL COEFFICIENTS FOR IP (≤ 15 μm) and FP (≤ 2.5 μm) PREDICTIVE EQUATIONS

EQUATION	VALUES OF THE COEFFICIENT "b"			
	NATIONAL VALUE	CALIFORNIA AREAS: ANNUAL VALUE (SUMMER, WINTER)		
		San Francisco Area	Central Valley	Los Angeles Area
IP PREDICTIONS				
$IP = b \cdot TSP$	0.61	0.53* (0.43,0.60)**	0.51* (0.43,0.59)**	0.66* (0.70,0.62)**
$IP = 1.2\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	0.56	0.50* (0.38,0.58)**	0.49* (0.39,0.57)**	0.64* (0.68,0.60)**
$IP = 1.2\ SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	0.50	0.41* (0.33,0.48)**	0.41* (0.32,0.50)**	0.59* (0.66,0.53)**
FP PREDICTIONS				
$FP = b \cdot TSP$	0.30	0.26 (0.19,0.31)**	0.23 (0.16,0.29)**	0.28 (0.29,0.27)
$FP = 1.1\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	0.21	0.20 (0.11,0.27)**	0.18 (0.11,0.25)**	0.23 (0.24,0.23)
$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	0.14	0.11 (0.05,0.16)**	0.10 (0.04,0.16)**	0.16 (0.20,0.13)**

*Regional coefficient differs from national value at 95% confidence level.

**Summer-winter difference is significant at 95% confidence level.

Note: Coefficients for site-types and other regions, as well as the standard errors for all coefficients, are given in the tables of Chapter 3.

TABLE 2 PERFORMANCE OF VARIOUS MODELS IN PREDICTING IP ($\leq$ 15 μm).

Table 2a Percentage Errors in Predicting Daily IP.

EQUATION	1-EQUATION NATIONAL MODEL	3-EQUATION SITE-TYPE MODEL	8-EQUATION REGIONAL MODEL	16-EQUATION REGIONAL/ SEASONAL MODEL
$IP = b \cdot TSP$	31.3%	31.0%	29.6%	28.5%
$IP = 1.2\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	29.5%	29.2%	28.0%	26.8%
$IP = 1.2\ SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	27.8%	27.6%	26.5%	25.7%
$IP = 1.2\ SO_4^{=} + 15\ Pb + b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2 NO_3^{-}$	27.3%	26.9%	25.5%	24.8%

Table 2b Percentage Errors in Predicting Annual Mean IP.

EQUATION	1-EQUATION NATIONAL MODEL	3-EQUATION SITE-TYPE MODEL	8-EQUATION REGIONAL MODEL	16-EQUATION REGIONAL/ SEASONAL MODEL
$IP = b \cdot TSP$	15.8%	16.2%	14.3%	Not Applicable
$IP = 1.2\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	15.5%	15.9%	13.1%	Not Applicable
$IP = 1.2\ SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	15.8%	16.0%	13.0%	Not Applicable
$IP = 1.2\ SO_4^{=} + 15\ Pb + b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2 NO_3^{-}$	15.0%	15.0%	13.0%	Not Applicable

The "free" coefficient in the predictive equations for IP exhibits statistically significant variations according to site-type, region, and region/season (see Table 1). Some of these variations make sense in terms of known site-type, regional, and regional/seasonal patterns in aerosol composition. Although the variations in the coefficient are statistically significant, many are not of great practical significance in the sense that they are small in absolute magnitude. Disaggregating the IP predictive scheme by site-type produces essentially no reduction in overall error (compare first and second columns of Table 2). Disaggregating the IP predictive equations by region and region/season produces a slight reduction in error. The most complex of the recommended schemes -- regional/seasonal equations using TSP, $SO_4^=$, and Pb -- yields a 26% error in predicting daily IP and a 13% error in predicting annual mean IP.

The national aggregate equation relating FP to TSP is FP = 0.30 TSP. This equation is rather imprecise, yielding a 56% standard error in predicting daily FP and a 30% standard error in predicting annual mean FP (see Table 3). The prediction of FP can be made significantly more accurate by adding the $SO_4^=$ and Pb variables. The national aggregate equation using three Hi-Vol variables -- FP = 1.1 $SO_4^=$ + 11 Pb + 0.14 (TSP - 1.4 $SO_4^=$ - 15 Pb) -- has an error of 40% for daily FP and 17% for annual mean FP. For the same reasons noted previously, we recommend against using equations with the NO_3^- variable.

Disaggregating the FP predictive scheme by site-type produces essentially no increase in accuracy (compare first and second columns of Table 3). Disaggregating the FP equations by region and/or season increases accuracy very slightly. The most complex of the recommended schemes -- regional/seasonal equations using TSP, $SO_4^=$, and Pb -- has an error of 38% for daily FP and 16% for annual mean FP.

Measurements of SSIP (IP data taken with Hi-Vols that have size-selective inlets) can be predicted accurately from TSP data alone. The aggregate national equation, SSIP = .74 TSP, represents a correlation level of 0.97. The error in this equation is 18% for individual daily values of SSIP and 11% for annual mean values of SSIP.

TABLE 3 PERFORMANCE OF VARIOUS MODELS IN PREDICTING FP ($\leq$ 2.5 μm).

Table 3a Percentage Errors in Predicting Daily FP.

EQUATION	1-EQUATION NATIONAL MODEL	3-EQUATION SITE-TYPE MODEL	8-EQUATION REGIONAL MODEL	16-EQUATION REGIONAL/ SEASONAL MODEL
$FP = b \cdot TSP$	56.3%	55.9%	53.1%	51.4%
$FP = 1.1 + b(TSP - 1.4\ SO_4^{=})$	46.1%	45.9%	44.9%	42.8%
$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	39.7%	39.6%	39.2%	38.2%
$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2 NO_3^{-}$	37.9%	37.5%	36.2%	34.3%

Table 3b Percentage Errors in Predicting Annual Mean FP.

EQUATION	1-EQUATION NATIONAL MODEL	3-EQUATION SITE-TYPE MODEL	8-EQUATION REGIONAL MODEL	16-EQUATION REGIONAL/ SEASONAL MODEL
$FP = b \cdot TSP$	29.9%	29.7%	24.3%	Not Applicable
$FP = 1.1 + b(TSP - 1.4\ SO_4^{=})$	19.7%	20.1%	18.2%	Not Applicable
$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	16.6%	16.8%	16.4%	Not Applicable
$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2 NO_3^{-}$	15.8%	16.5%	18.1%	Not Applicable

APPLICATION TO CALIFORNIA HI-VOL DATA

The ARB computerized files include 287 monitoring sites that reported some Hi-Vol data during the years 1976 to 1980. The records for 226 of these sites contain adequate quantities of data to be included in our application study. In applying our formulae to predict IP and FP at these sites, we choose to use the complex, disaggregated, regional/seasonal models because regional and seasonal variations are especially significant in California. Also, for as many sites as possible, we include $SO_4^=$ and/or Pb data in addition to the TSP data.

At each site, we convert the daily Hi-Vol data into estimates of IP and FP. Then, using all available data for 1976-1980 at each site, we calculate the annual mean concentration, seasonal average concentrations, and yearly maximum concentration for TSP, IP, and FP.

Figures 1, 2, and 3 present approximate isopleth maps indicating the general geographical patterns of annual mean values for TSP, IP, and FP, respectively. The most notable features of these maps are the high particulate concentrations in the South Coast Air Basin (Los Angeles area) and the San Joaquin Valley Air Basin. Considerable portions of the South Coast Air Basin experience annual mean values of TSP > 125 $\mu g/m^3$, IP > 85 $\mu g/m^3$, and FP > 40 $\mu g/m^3$. The southern part of the San Joaquin Valley, from just south of Fresno down to Bakersfield, experiences annual means of TSP > 150 $\mu g/m^3$, IP > 70 $\mu g/m^3$, and FP > 30 $\mu g/m^3$.

The very limited data available for southeast California suggest that the Imperial Valley may also be a significant hot-spot for particulate concentrations. In the future, it would be worthwhile to add dichotomous samplers and expand the Hi-Vol network in the Imperial Valley.

The lowest particulate concentrations in California occur in the eastern edge of the state along the Nevada border, where annual means of TSP, IP, and FP are generally less than 50 $\mu g/m^3$, 25 $\mu g/m^3$, and 10 $\mu g/m^3$, respectively. A band of low particulate concentrations also apparently exists in the northwest part of the state, from Trinity County down to Lake County.

The geographical patterns of particulate concentrations in California

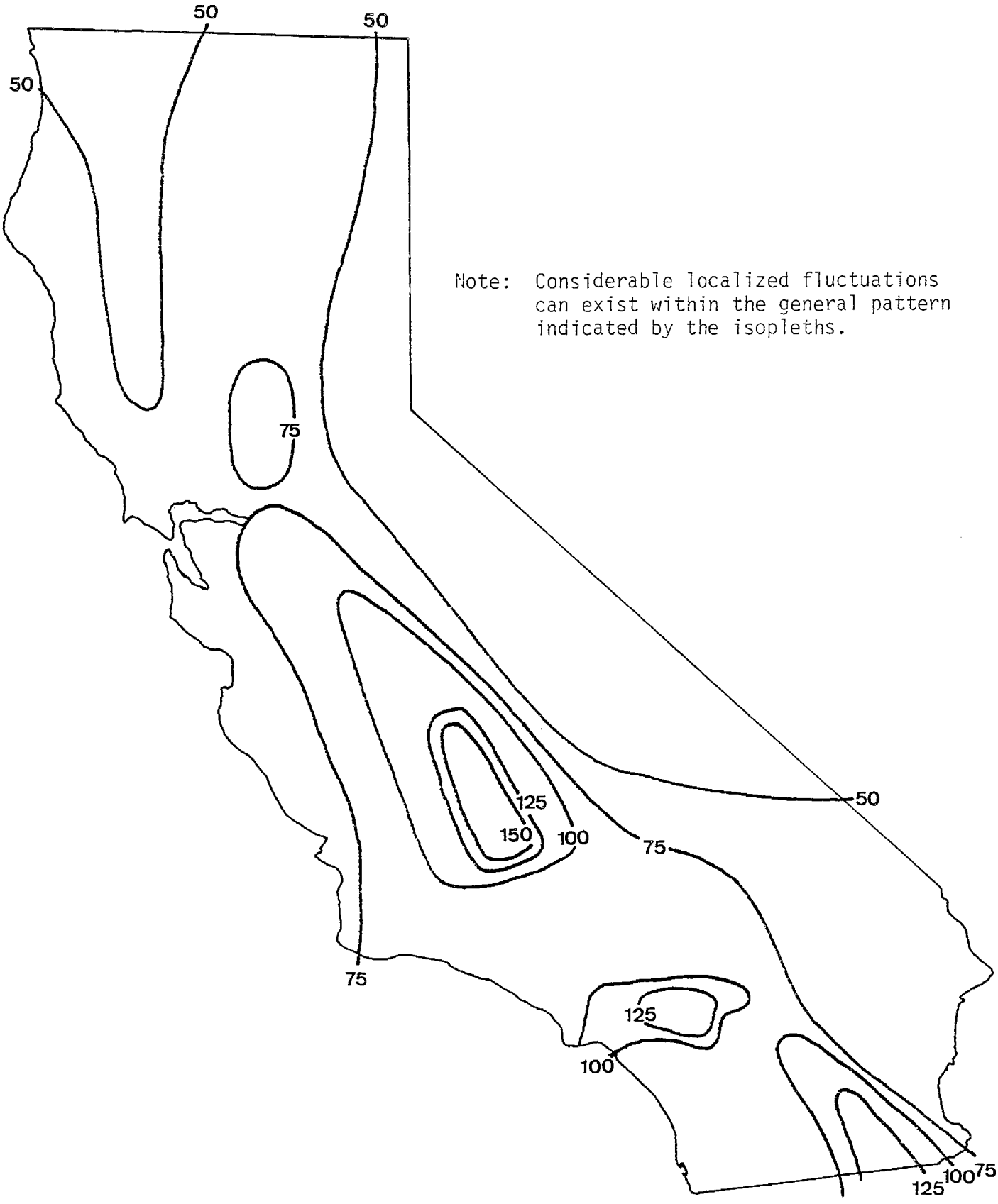

Figure 1 Isopleths illustrating the general spatial pattern of annual TSP concentrations ($\mu g/m^3$) in California.

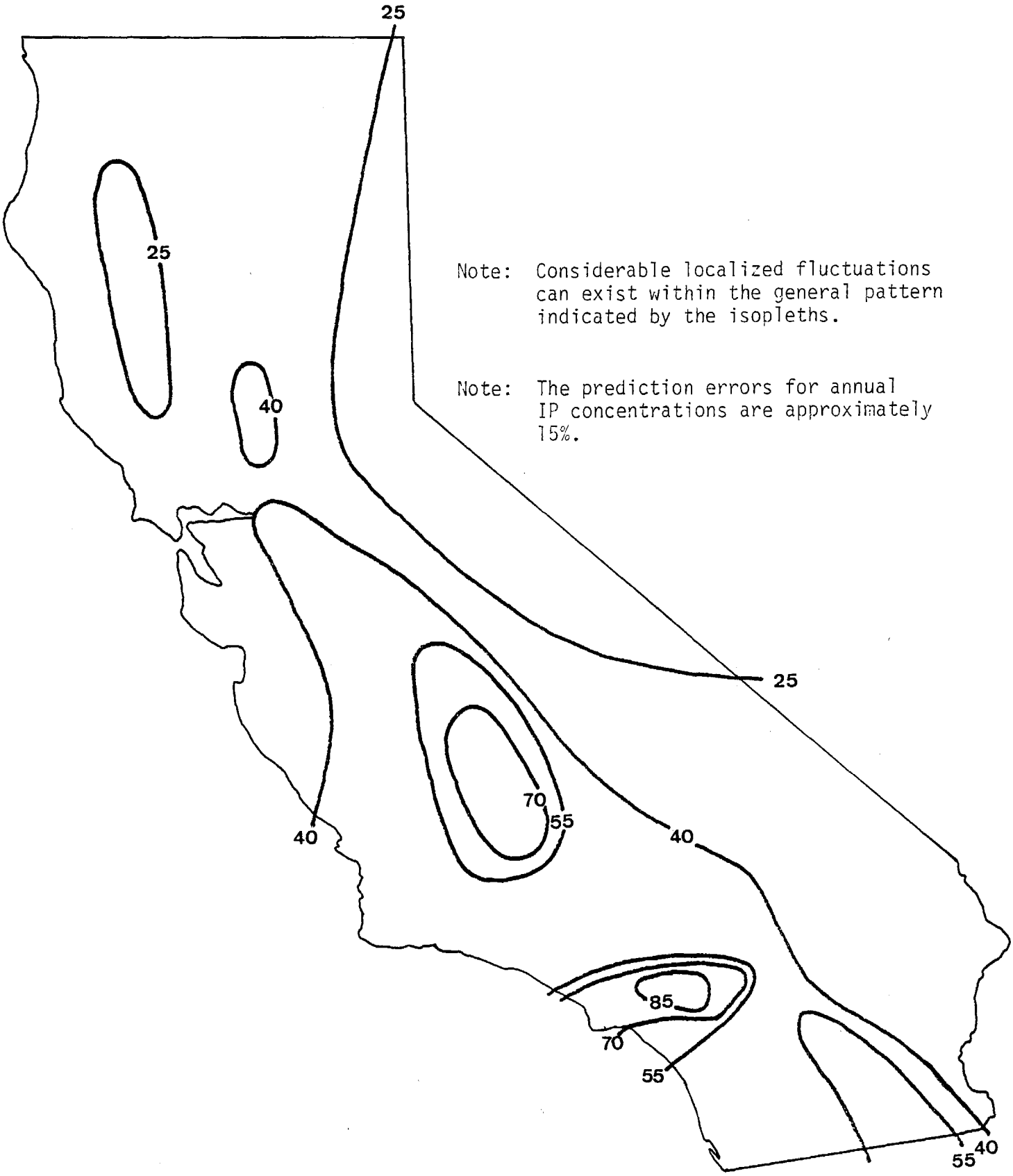

Figure 2 Isopleths illustrating the general spatial pattern of predicted annual IP (≤ 15 μm) concentrations (μg/m^3) in California.

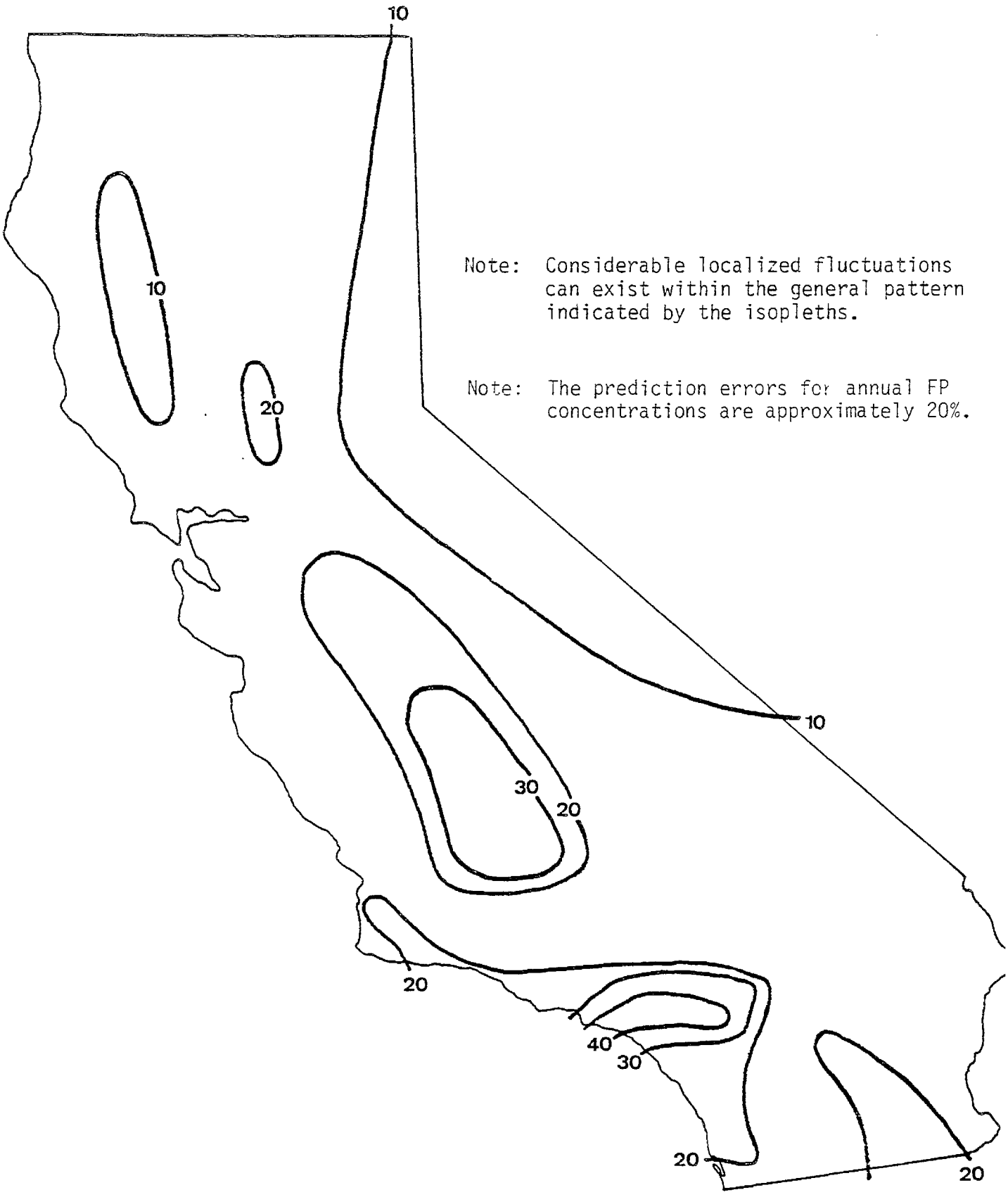

Figure 3 Isopleths illustrating the general spatial pattern of predicted annual FP (≤ 2.5 μm) concentrations (μg/m^3) in California.

generally make sense in terms of the spatial distribution of emissions for primary particles and for gaseous precursors of secondary aerosols. In particular, the Los Angeles area and southern San Joaquin Valley stand out as hot-spots for particulate and SO_x emissions, while the Los Angeles area stands out for NO_x and hydrocarbons.

The geographical distribution of fine particle concentrations corresponds fairly well with the geographical distribution of visibility in California. The results of the present study add further support to the conclusion by Trijonis (1980) that the very low visibilities in the South Coast and San Joaquin Valley Air Basins are essentially caused by excessive levels of anthropogenic fine aerosols.

The seasonal patterns of FP often diverge significantly from the seasonal patterns of TSP. As one would expect, the seasonal variations of IP are intermediate to those of TSP and FP. Also as expected, the seasonal pattern of visibility corresponds better (in an inverse sense) to that of FP than to that of TSP or IP. The seasonal variation of visibility most closely tracks the seasonal variation of FP in those air basins where man-made visibility impacts are most severe (e.g. the San Joaquin, Sacramento, San Francisco Bay Area, and South Coast Air Basins), but the seasonal variation in visibility more closely follows seasonal variations in meteorology in some of the cleaner areas of California.

The Northeast Plateau, Sacramento Valley, San Francisco Bay Area, and San Joaquin Valley Air Basins undergo their highest FP levels and lowest visibility during the fourth (fall) quarter, with winter being the second worst season. Most notably, the southern part of the San Joaquin Valley experiences average FP concentrations of 45 to 65 $\mu g/m^3$ and average visibilities of 6 to 7 miles during the fall quarter.

The South Central Coast, South Coast, and Southeast Desert Air Basins experience their highest FP levels and lowest visibility levels during the third (summer) quarter, with spring being the second worst season. The valleys and eastern inland areas of the South Coast Air Basin undergo average FP levels of 40 to 60 $\mu g/m^3$ and average visibilities of 5 to 6 miles during the summer.

REFERENCES

Evans, J.S., J.D. Spengler, and D.W. Cooper, "A Comparison of Three Particulate Metrics in Four U.S. Cities," Harvard School of Public Health, Submitted to Environmental Science and Technology, 1981.

Feldman, H.J., W.A. Turner, J.S. Evans, and J.D. Spengler, "Investigation of the Relationship Between IP (da≤ 15 μm) and TSP Mass Concentrations in Six U.S. Cities," Presented at the 74th Annual Meeting of the Air Pollution Control Association, Philadelphia, June 1981.

Miller, F.J., D.E. Gardner, J.A. Graham, R.E. Lee, W.E. Wilson, and J.D. Bachman, "Size Considerations for Establishing a Standard for Inhalable Particles," J. of the Air Pollution Control Association, 29, pp. 610-615, 1979.

Spengler, J.D., W.A. Turner, and D.W. Dockery, "Comparison of Hi-Vol, Dichotomous, and Cyclone Samplers in Four U.S. Cities," Presented at the 73rd Annual Meeting of the Air Pollution Control Association, Montreal, June 1980.

Trijonis, J., "Visibility in California," Prepared under contract A7-181-30 for the California Air Resources Board, also accepted for publication in the J. of the Air Pollution Control Association, 1980.

Trijonis, J., J. Eldon, J. Gins, and G. Berglund, "Analysis of the St. Louis RAMS Ambient Particulate Data, Volume I: Final Report," EPA-450/4-80-006a, 1980.

Wendt, J.G. and K.J. Torre, "Field Test of Four Size-Segregated Particulate Samplers," Presented at the 74th Annual Meeting of the Air Pollution Control Association, Philadelphia, June 1981.