SANTA FE RESEARCH CORPORATION

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

FINAL REPORT

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 A0-076-32

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.

ABSTRACT

Formulae for translating Hi-Vol data into estimates of inhalable particles (IP) and fine particles (FP) are developed, evaluated, and applied. The equations are developed using simultaneous data from dichotomous samplers and Hi-Vol samplers at 75 locations nationwide, including 11 locations in California. The formulae are multivariate in the sense that they include the Hi-Vol parameters, TSP, $SO_4^=$, and Pb; the formulae are hybrid in the sense that the coefficients are partly physico-chemical and partly statistical.

Several sets of equations are presented with varying degrees of complexity. The Hi-Vol parameters are added in a stepwise fashion -- TSP, then $SO_4^=$, then Pb. Also, there are national aggregate equations (e.g. IP = 0.61 TSP or FP = 0.30 TSP) as well as equations disaggregated by site-type, region, and region/season. Depending on the level of complexity, the predictive errors are as follows: 26 to 31% for individual daily values of IP, 13 to 16% for annual mean values of IP, 39 to 56% for individual daily values of FP, and 16 to 30% for annual mean values of FP.

A major application study using 5 years of Hi-Vol data at 226 California sites allows us to investigate the statistical, geographical, and seasonal patterns of TSP, IP, and FP throughout California. The most salient features of the application study involve the extremely high particulate concentrations in the Los Angeles area and the San Joaquin Valley. The predictive formulae for IP and FP can also be usefully applied to historical health effects studies based on Hi-Vol data.

TABLE OF CONTENTS

FIGURES

TABLES

1. INTRODUCTION AND SUMMARY

Because the health and welfare effects of airborne particles depend significantly on the size of the particles, the California ARB is considering the possibility of expanding or replacing the present air quality standards for total suspended particulate matter (TSP) with size-specific standards for inhalable particles (IP) and/or fine particles (FP). In considering new size-specific particulate standards, one major difficulty is the paucity of health studies and ambient data for IP and FP as compared to the abundance of studies and data for TSP. To support various technical analyses concerning potential standards for IP and/or FP, there is a need for methods of estimating IP and FP concentrations from routine measurements of TSP and other Hi-Vol parameters.

In this report, we develop, evaluate, and apply linear equations for estimating IP and FP from Hi-Vol data for TSP, sulfates ($SO_4^=$), and lead (Pb). The equations are based on a hybrid approach; some of the coefficients are determined from physico-chemical principles, but one coefficient in each equation is determined statistically. The equations are developed using all simultaneous recordings by dichotomous samplers and Hi-Vol samplers available nationwide from the EPA IP Network and in California from the ARB Network. The equations are evaluated with respect to errors in predicting both annual means and individual 24-hour values. The application uses five years of California Hi-Vol data and yields a comprehensive description of the spatial/seasonal patterns of IP and FP throughout California.

1.1 BACKGROUND

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.

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 5 years of Hi-Vol data at 226 California sites; also, the ARB staff has begun applying our formulae to historical health effects studies.

1.2 REPORT AND PROJECT ORGANIZATION

The present study is organized into two phases, a development phase and an application phase. Chapters 2 and 3 of this report deal with the development phase. The first part of Chapter 2 summarizes the data base (930 simultaneous readings by dichotomous samplers and Hi-Vol samplers at 75 locations) and discusses our data quality screening procedure. The remainder of Chapter 2 discusses the methodology -- a hybrid approach with stepwise addition of Hi-Vol parameters. The hybrid approach is part physico-chemical, part statistical. The Hi-Vol parameters are added stepwise (in the order TSP, $SO_4^=$, Pb, and NO_3^-) in order to investigate the trade-off between simplicity and accuracy and in order to provide a method that is flexible depending on the number of Hi-Vol parameters that are available in various applications.

Chapter 3 develops and evaluates predictive formulae for IP and FP. At each stage in the stepwise addition of independent variables (Hi-Vol parameters), the predictive equations are developed and evaluated on an aggregate national basis, a site-type basis, a regional basis, and a regional/seasonal basis. The errors in the various sets of equations are assessed both for predictions of annual means and for predictions of individual daily values.

Chapter 4 describes the application of the methodology to five years of Hi-Vol data at 226 California sites. At each location, we estimate annual means, seasonal averages, and expected yearly maxima for both IP and FP. A comparison is then made of the geographical and seasonal patterns for TSP, IP, FP, and visibility throughout California.

1.3 SUMMARY

The following subsections summarize our findings and conclusions. For convenient referral, the summary is organized according to the order of the chapters.

Data Base and Methodology (Chapter 2)

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 use in 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 8 geographical regions: 3 regions in California (San Francisco Bay Area, Central Valley, and Los Angeles Area) and 5 other regions nationwide (Pacific Northwest, Arid Southwest, North Central, Northeast, and Southeast). The study locations are also classified according to 3 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 $SO_4^=$ terms represent ammonium sulfate, with adjustments made to account for the different particle size cut-offs and different artifact sulfate properties of dichotomous and Hi-Vol samplers. Lead is used to represent contributions due to primary particulate matter from all road vehicles (gasoline vehicle exhaust, diesel exhaust, tire wear, and brake wear); again adjustments are made for the different particle size cut-offs for FP, IP, and TSP. 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 equations for the four steps are as follows:

$$IP = b \cdot TSP$$
$$IP = 1.2\ SO_4^= + b(TSP - 1.4\ SO_4^=)$$
$$IP = 1.2\ SO_4^= + 15\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$$
$$IP = 1.2\ SO_4^= + 15\ Pb + b_1(TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2 NO_3^-$$

and

$$FP = b \cdot TSP$$
$$FP = 1.1\ SO_4^= + b(TSP - 1.4\ SO_4^=)$$
$$FP = 1.1\ SO_4^= + 11\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$$
$$FP = 1.1\ SO_4^= + 11\ Pb + b_1(TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2 NO_3^-$$

The coefficients "b" are determined through zero intercept regressions. It makes sense to supress the intercept in the regressions because the intercept is usually insignificant statistically, because the equations can be more readily interpreted without an intercept, and because the intercept parameter can be eliminated with negligible loss in predictive accuracy.

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 (Chapter 3)

On a national aggregate basis, the equation relating IP to TSP is IP = 0.61 TSP (see Table 1.1). This equation yields a 31% standard error in predicting individual daily values of IP and a 16% standard error in predicting annual mean values of IP (see Table 1.2). 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 1.2). If $SO_4^=$ and Pb data

TABLE 1.1 NATIONAL AND (CALIFORNIA) REGIONAL COEFFICIENTS FOR IP AND FP 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 1.2 PERFORMANCE OF VARIOUS MODELS IN PREDICTING IP (≤ 15 μm).

Table 1.2a 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%
$IP = 1.2\ SO_4^{=} + 15\ Pb + b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2NO_3^{-}$	27.3%	26.9%	25.5%	24.8%

Table 1.1b 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
$IP = 1.2\ SO_4^{=} + 15\ Pb + b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2NO_3^{-}$	15.0%	15.0%	13.0%	Not Applicable

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.

The "free" coefficient in the predictive equations for IP exhibits statistically significant variations according to site-type, region, and region/season (see Table 1.1). 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 1.2). 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 1.3). 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 1.3). 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

TABLE 1.3 PERFORMANCE OF VARIOUS MODELS IN PREDICTING FP ($\leq 2.5\ \mu m$).

Table 1.3a 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 1.2b 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

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. These errors can be reduced very slightly (by about 1%) by adding the $SO_4^=$ variable or by disaggregating the predictive scheme by region.

Application to California Hi-Vol Data (Chapter 4)

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. The yearly maximum concentrations are computed for an every sixth day sampling schedule (61 samples per year) by interpolating the actual frequency distributions of the data. We find that the yearly max/mean ratios for TSP, IP, and FP are generally in the range of 2 to 4, although a few sites exhibit max/mean ratios significantly greater than 4.

Figures 1.1, 1.2, and 1.3 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 $\mu g/m^3$.

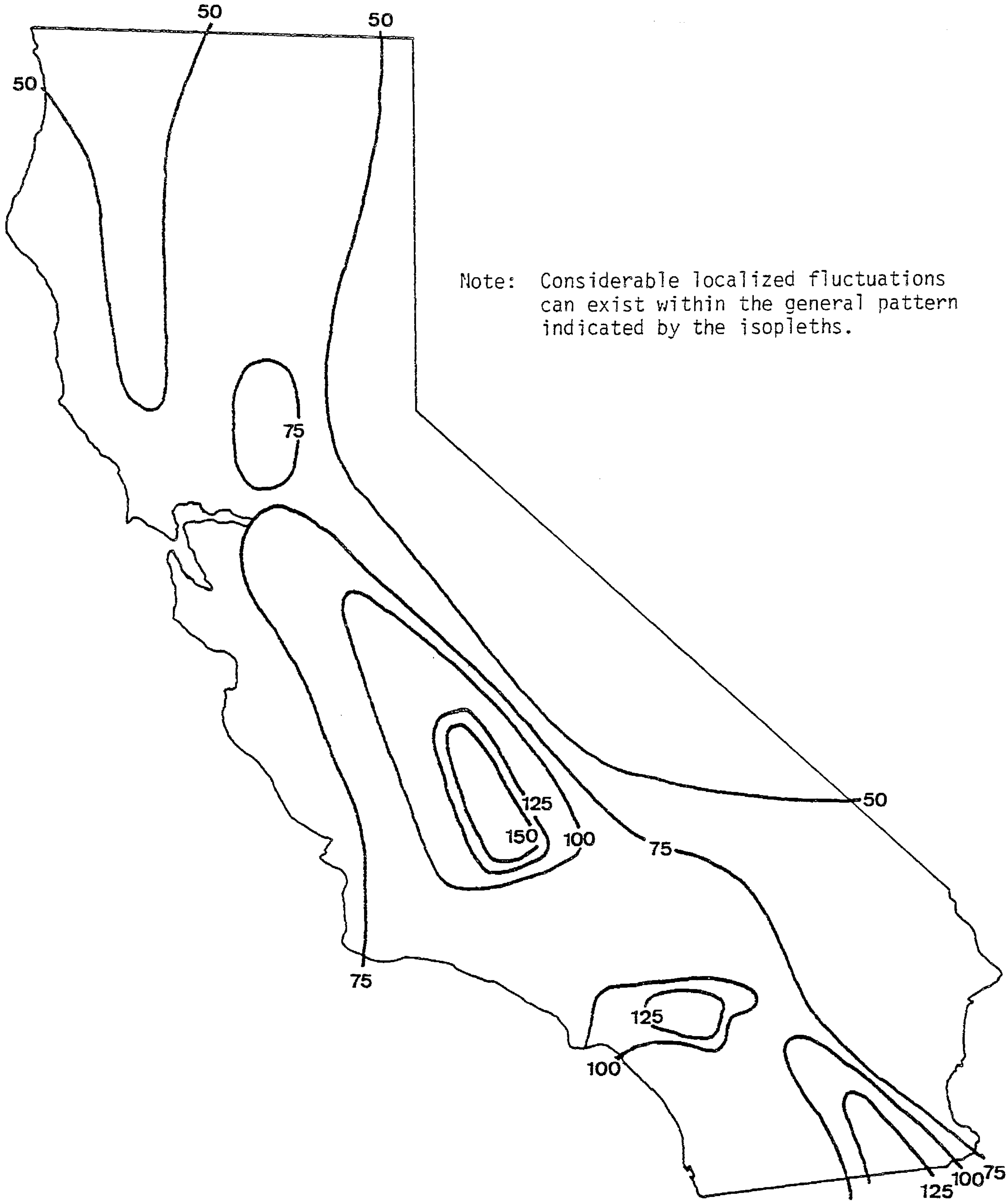
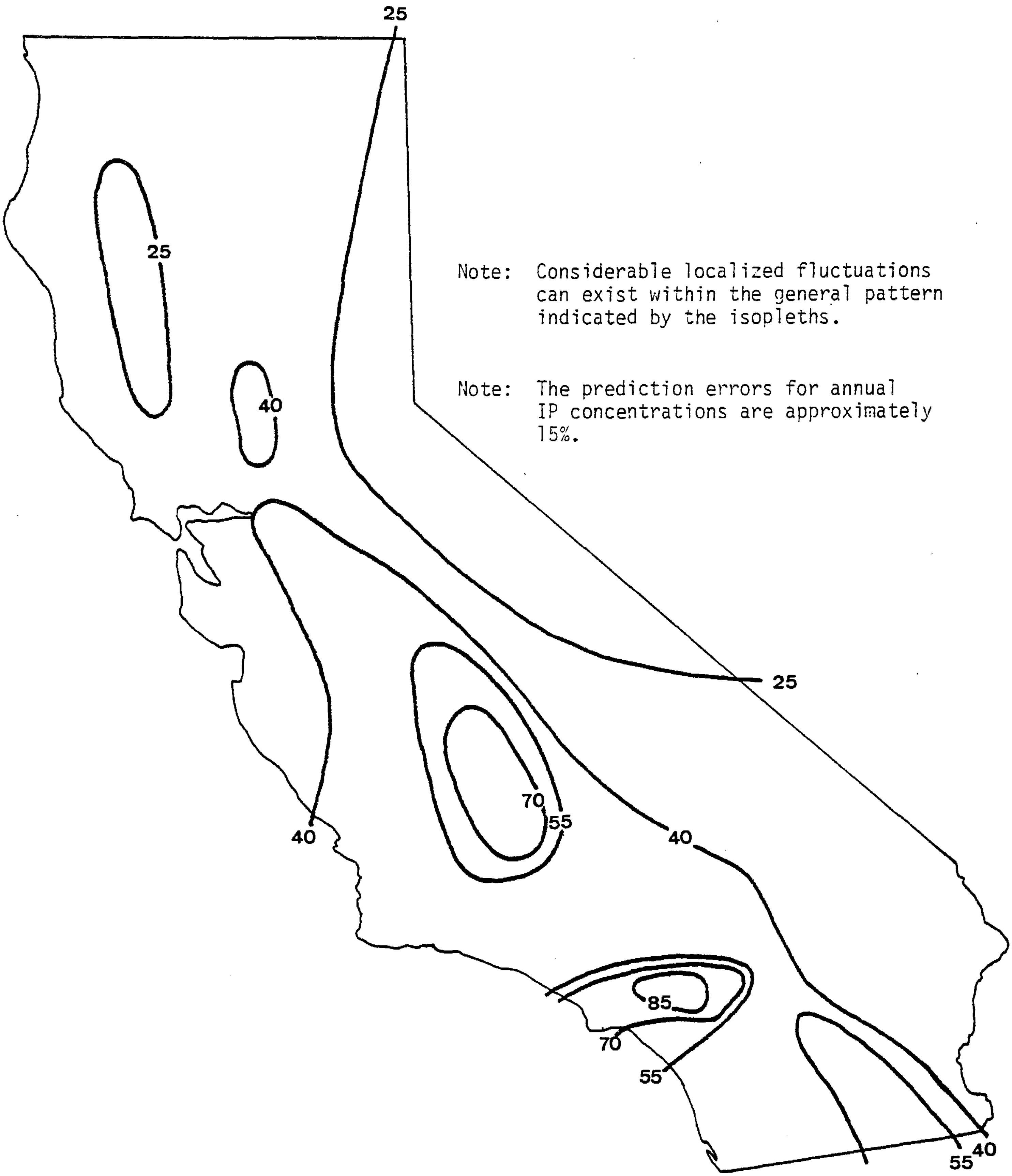


Figure 1.1 Isopleths illustrating the general spatial pattern of annual TSP concentrations ($\mu g/m^3$) in California.



Note: Considerable localized fluctuations can exist within the general pattern indicated by the isopleths.

Note: The prediction errors for annual IP concentrations are approximately 15%.

Figure 1.2 Isopleths illustrating the general spatial pattern of predicted annual IP ($\leq$ 15 μm) concentrations ($\mu g/m^3$) in California.

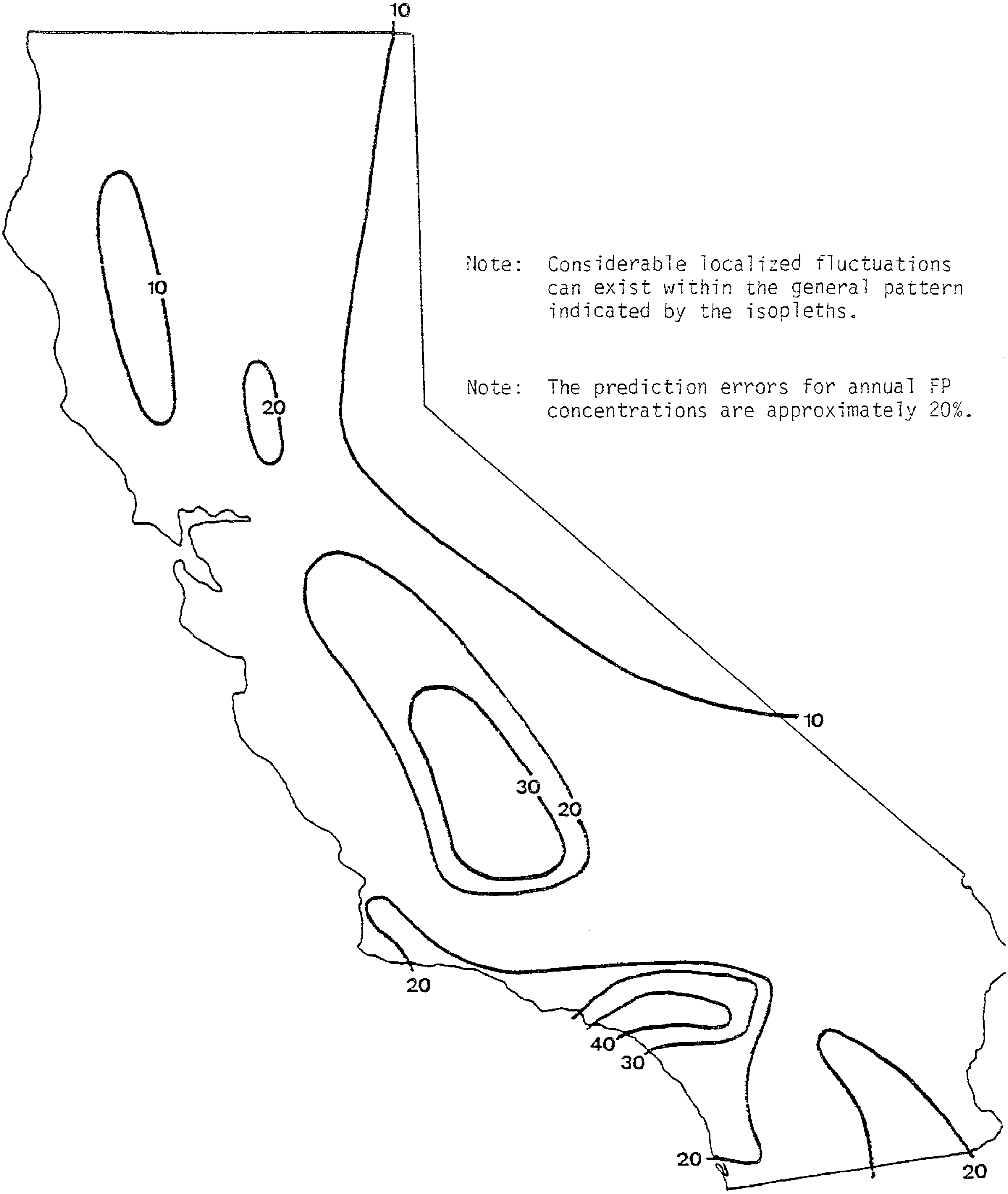


Figure 1.3 Isopleths illustrating the general spatial pattern of predicted annual FP ($\leq$ 2.5 μm) concentrations ($\mu g/m^3$) in California.

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 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.

2. DATA BASE AND METHODOLOGY

This chapter describes the data base and methodology used to develop relationships between dichotomous parameters and Hi-Vol parameters. The data base, discussed in Section 2.1, consists of simultaneous measurements made with dichotomous samplers and Hi-Vol samplers at 75 locations nationwide, including 11 sites in California. The methodology, discussed in Section 2.2, is a hybrid approach involving predictive equations based on both physico-chemical principles and statistical results.

It should be noted that other data sets are used in the application phase of this project. These other data sets are described in Chapter 4 which deals with the California application study. The present chapter concerns only the data and methodology for the development phase of the project.

2.1 DATA BASE FOR DEVELOPING RELATIONSHIPS

The data base required to develop the required relationships consists of simultaneous measurements made by dichotomous samplers and Hi-Vol samplers. The relevant dichotomous parameters are IP and FP, while the relevant Hi-Vol parameters are TSP, $SO_4^=$, Pb, and NO_3^-. We are also interested in comparing SSIP (IP data taken using Hi-Vols with size-selective inlets) to the routine Hi-Vol parameters. This section discusses the data sets selected for the study, describes our data quality analyses, and summarizes the resultant data base.

2.1.1 Available Data Sets

Table 2.1 lists five major monitoring networks that provide dichotomous data for IP and FP as well as Hi-Vol data for TSP, $SO_4^=$, Pb, and NO_3^-. Although there are some other monitoring programs that include both dichotomous sampling and Hi-Vol sampling (Lioy et al. 1980), we focused only on these five networks because they contain the greatest number of sites, cover the longest time periods, and measure all (or nearly all) of the required parameters.

After a thorough review of the available data sets, we selected only two

TABLE 2.1 MAJOR MONITORING PROGRAMS PROVIDING SIMULTANEOUS HI-VOL AND DICHOTOMOUS SAMPLER MEASUREMENTS.

MONITORING NETWORK	SITES	TIME PERIOD OF AVAILABLE DATA
EPA IP NETWORK	Presently there are 73 usable sites, including 9 in California. Eventually, there will be nearly 200 sites.	1 - 15 months, depending on the site. Overall the data start in mid-1979 and are now available through late 1980.
CALIFORNIA ARB NETWORK	Bakersfield and Riverside	11 months at Bakersfield in 1979-1980. 3 months at Riverside in 1979.
NEW YORK STATE NETWORK	6 locations in Buffalo	9 months in 1978-1979.
EPA RAMS NETWORK	10 sites in and near St. Louis	20 months from middle 1975 to early 1977.
HARVARD NETWORK	6 sites in the Northeast and Midwest	18 months during late 1970s.

of them -- the EPA IP Network and the ARB Network -- for this study. The EPA Network is an obvious and necessary choice; it provides, by far, the greatest amount of data. Also, the EPA Network is most pertinent to our later applications because it contains several sites where health effects studies have been performed and several sites in California. The ARB Network is also a necessary choice because we want as many sites as possible in California. Both the EPA and ARB Networks contain all of the relevant parameters -- dichotomous IP and FP; Hi-Vol TSP, $SO_4^=$, Pb, and NO_3^-; and Hi-Vol SSIP.

We also acquired data from the New York State Network in Buffalo. Buffalo is of special interest because of health effects studies that have been conducted there. However, after acquiring all available data for Buffalo (Kolak et al. 1979; Delaware 1981), we found that none of the sites with chemical composition data for $SO_4^=$, Pb, and NO_3^- corresponded to the sites with co-located dichotomous and Hi-Vol samplers. This problem would have been correctable if the $SO_4^=$, Pb, and NO_3^- levels agreed very closely from site to site; then, we could have used the chemical composition data at sites which had such data as a surrogate for the missing data at the sites of interest. Unfortunately, a statistical analysis for the sites with available data revealed that the inter-site agreement of daily levels of $SO_4^=$, Pb, and NO_3^- was not nearly good enough to justify this surrogate method. Thus, we excluded the New York State data for Buffalo in our final data base. This exclusion is not a significant loss because we have two EPA Network sites in the Buffalo area.

The Harvard data and St. Louis RAMS data were not acquired for two reasons. First, each of these data sets had a feature that made it imperfect for the purposes of our study. The Harvard Network did not include Hi-Vol measurements of Pb and NO_3^- (Spengler 1980); the St. Louis RAMS data for IP corresponded to an upper size cut-off of 20 to 25 μm (Dzubay 1980; Lioy et al. 1980) rather than the 15 μm cut-off for all other data sets. Second, both of these networks are in the Eastern United States, where the EPA Network provides a plethora of data. It did not seem that the Harvard and St. Louis RAMS data sets would improve our study enough to justify the time and effort spent in acquiring and organizing those data sets.

In an effort to expand our data base for California as much as possible, we constructed new data sets using co-located ARB Hi-Vol and EPA dichotomous samplers. Chemical composition measurements on the EPA Hi-Vols were taken only every twenty-fourth day; by using the ARB Hi-Vol data we hoped to increase the number of data points with $SO_4^=$, Pb, and NO_3^-. Unfortunately, at some locations, the ARB Hi-Vol and EPA dichotomous sampling schedules were one or two days out of phase so that no simultaneous data occurred. At Azusa, Pasadena, Richmond, and Rubidoux, however, we managed to expand our data base considerably by matching ARB Hi-Vol measurements with EPA dichotomous sampler measurements.*

2.1.2 Data Quality Analysis

In reviewing the EPA data base, we noticed that there were a few problems evident in the data. For example, recordings existed for certain sites and days with IP values two to five times greater than TSP values. Because TSP essentially represents the mass of particles less than 50 μm in diameter while IP only represents the mass of particles less than 15 μm in diameter, and because TSP also tends to be inflated over IP due to "artifact" collection of gases on Hi-Vol filters, such recordings are obviously unreasonable. We decided to eliminate these and other highly unreasonable data points by formulating a data quality screening procedure.

It should be noted that EPA does perform a statistical screening test on the IP Network data; this test flags a significant number of data points (Rodes 1981). However, EPA only eliminates a data point when there is an obvious cause for the error. As a result, most of the questionable points are left in the data base.

Our data quality screens are based on ratios among three basic parameters: IP (for the dichotomous sampler), SSIP (for the size-selective Hi-Vol sampler), and TSP (for the routine Hi-Vol sampler). We selected cutoffs for these ratios using both physical and statistical considerations. The physical considerations involved what would be reasonable given the particle size ranges measured by the samplers and the possibility of artifacts

*At these sites there were some days when we had EPA dichotomous data matched with both EPA Hi-Vol data and ARB Hi-Vol data. For those days, we averaged the EPA and ARB Hi-Vol data to arrive at a single set of values.

with the Hi-Vol samplers. The statistical considerations were based on identification of outliers in simple histogram plots of the ratios (see, for example, Figure 2.1 for a histogram of the IP/TSP ratio). We chose the cut-offs so that only the most egregious outliers, at most a few percent of the data, would be eliminated. The data quality screen that we finally selected is as follows: a data point is eliminated if

(1) SSI/TSP > 1.4
(2) TSP/SSI > 5
(3) IP/TSP > 1.5
(4) TSP/IP > 4 and FP/IP > 0.6
(5) IP/SSI > 2
(6) SSI/IP > 2.5

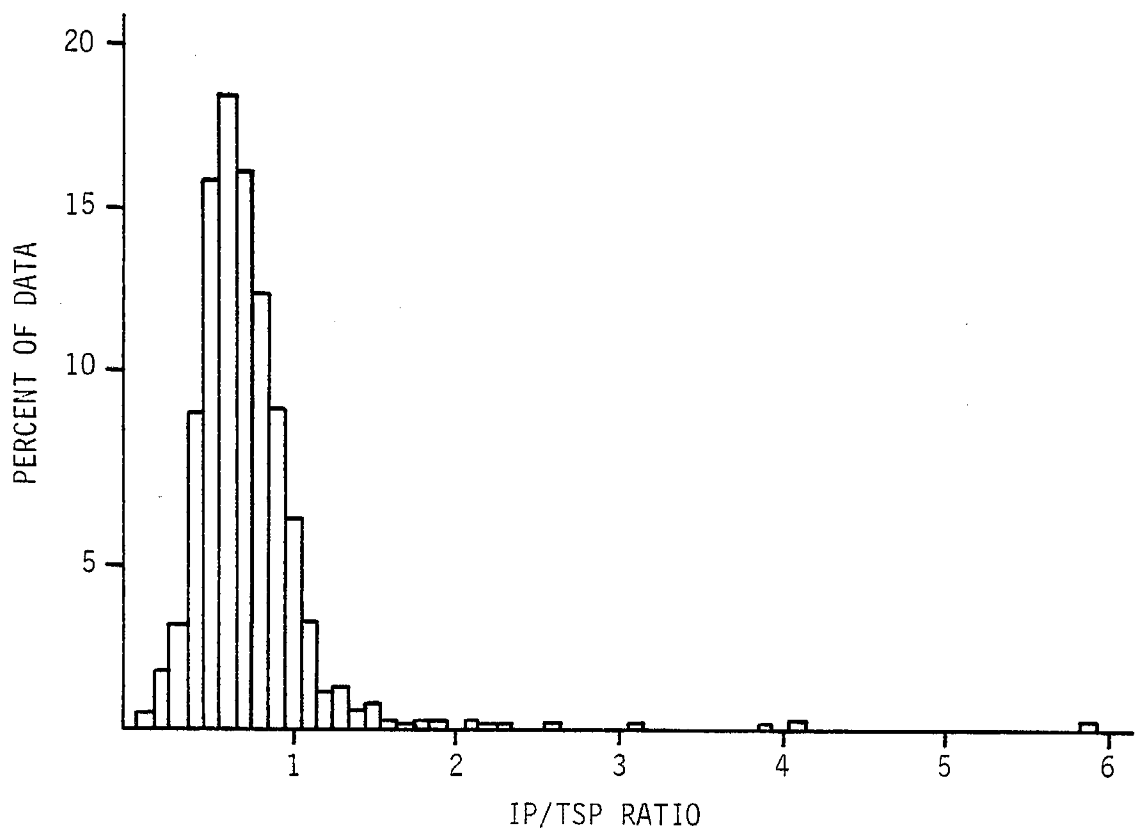


Figure 2.1 Histogram of the IP/TSP ratio for data from the EPA Network.

Table 2.2 summarizes the results of applying the data quality screen to the EPA Network data. Out of 928 data points, 42 are eliminated, leaving 886. Nearly half of the eliminated points are screened out by the single criteria IP/TSP > 1.5.

TABLE 2.2 RESULTS OF THE DATA QUALITY SCREEN APPLIED TO EPA NETWORK DATA.

TOTAL NUMBER OF DATA POINTS:		928
POINTS ELIMINATED BY VARIOUS CRITERIA:		
	SSI/TSP > 1.4	2
	TSP/SSI > 5	1
	IP/TSP > 1.5	20
	TSP/IP > 4 and FP/IP > 0.6	12
	IP/SSI > 2	3
	SSI/IP > 2.5	8
TOTAL POINTS ELIMINATED: (Note that this total does not equal the sum of the previous column because of 4 duplicated eliminations.)		42
REMAINING NUMBER OF POINTS AFTER THE DATA QUALITY SCREEN:		886

It is worthwhile to investigate the degree to which our data quality screen improves the statistical fit between the dichotomous data and Hi-Vol data. For the 928 EPA Network data points before the screen, a simple multiple regression against TSP, $SO_4^=$, Pb, and NO_3^- yields a correlation coefficient of 0.80 with IP as the dependent variable and 0.79 with FP as the dependent variable. For the 886 data points after the screen, the correlation is 0.90 for IP and 0.84 for FP. Thus, the statistical fit, which is rather good even for the unscreened data base, is improved moderately but significantly by the data quality test. We think that it is important to eliminate the data flagged by our data quality screen, not only because it improves the statistical fit somewhat, but even more because our final results will be partly based on regression coefficients which can be quite sensitive to outliers.

We also applied the same data quality test to the ARB Network and to the data base that we assembled using ARB Hi-Vol recordings and EPA dichotomous samplers. It is noteworthy that none of the data points involving ARB measurements were eliminated by the data quality screen.

2.1.3 Summary of the Data Base

The 75 locations included in this study are illustrated in Figure 2.2 and listed in Table 2.3. Table 2.3 also lists the number of data points at each site (after applying the data quality screen) and indicates the period of record covered by the data (note that there are significant gaps in the data record at certain sites). Of the 75 study locations, all but two -- Bakersfield and Riverside -- are EPA Network sites. As noted previously, the data base at 4 EPA Network sites in California (Azusa, Pasadena, Richmond, and Rubidoux) was expanded a great deal (30 additional data points) by combining ARB Hi-Vol measurements with EPA dichotomous sampler measurements.

Both Table 2.3 and Figure 2.2 distinguish the study locations as to geographical region, site-type, and an "annual" designation. There are 8 geographical regions: California - San Francisco Bay Area, California - Central Valley, California - Los Angeles Area, Pacific Northwest, Arid Southwest, North Central, Northeast, and Southeast. The major regions were chosen based on climatology (NOAA 1977); the geographical distribution of emission sources (EPA 1979); spatial patterns in the IP/TSP, FP/IP, and FP/TSP ratios; and the natural geographical groupings of the sites. California was afforded the special treatment of being split into three relatively small regions because spatial air quality gradients are most severe in California (Trijonis 1980), because the IP/TSP ratio shows obvious spatial variations in California (see Figure 2.3), and because California is of special concern in this study.

We found that it is very difficult to arrive at a satisfactory "site-type" classification. In the end we settled on three classes: "metropolitan",

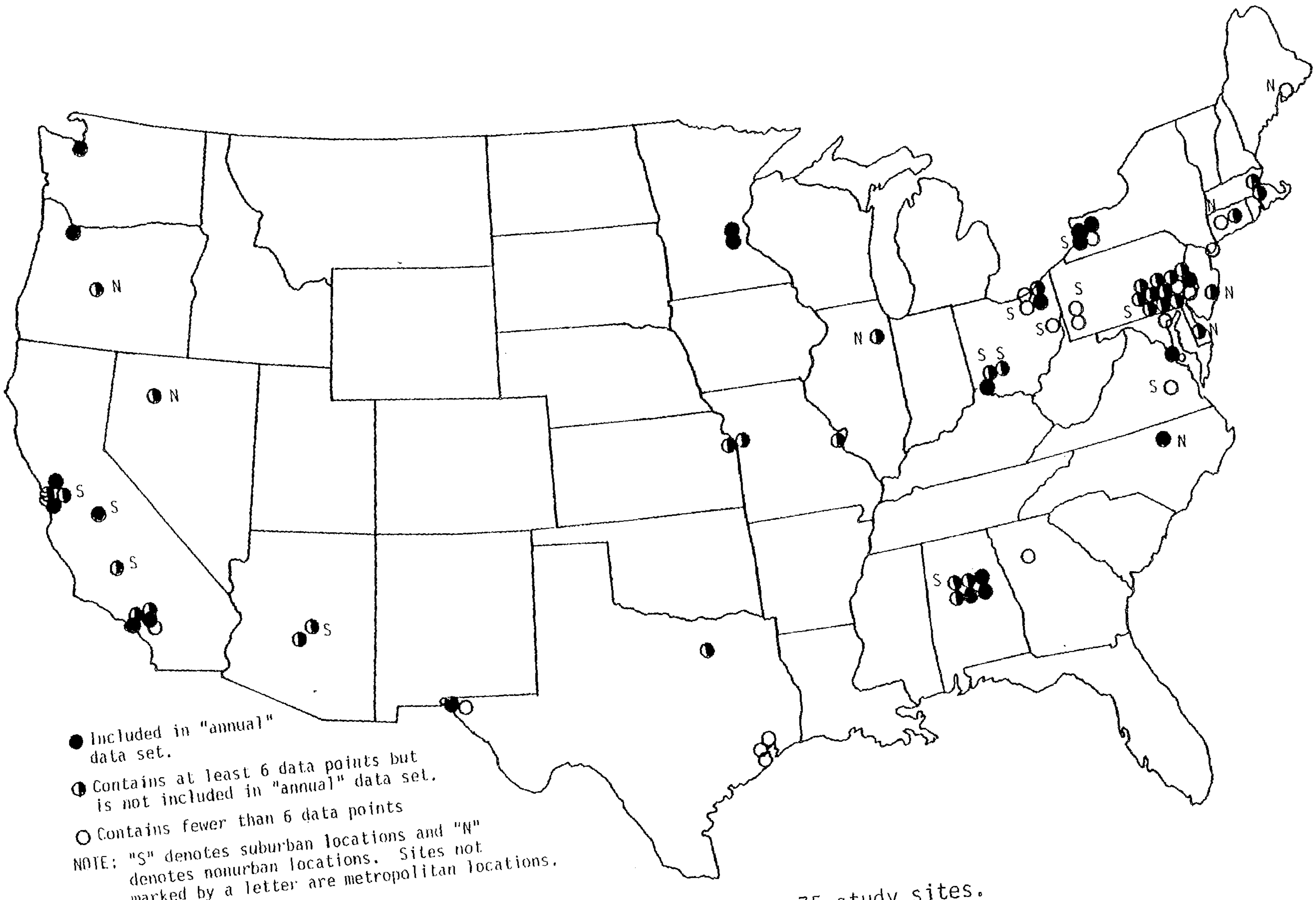


Figure 2.2 Locations of the 75 study sites.

TABLE 2.3 LIST OF THE 75 STUDY SITES.

SITE	SITE TYPE	ANNUAL DATA SET		1979 M	J	J	A	S	O	N	D	1980 J	F	M	A	M	J	J	A	S	O	N	D
CAL-SAN FRANCISCO BAY AREA																							
Livermore	S		8					—	—				—	—	—	—			—				
Richmond†	M	●	19					—												—			
San Francisco	M		6									—	—			—	—	—					
San Jose	M	●	20					—	—	—	—	—	—	—	—	—	—	—	—	—			
CAL-CENTRAL VALLEY																							
Bakersfield*	S		22	—	—	—	—	—	—	—	—												
Fresno Cnty.	S	●	13						—	—	—	—	—	—	—	—	—	—	—				
CAL-LOS ANGELES AREA																							
Azusa†	M		12					—	—	—	—	—	—										
Los Angeles	M	●	16			—	—	—	—	—	—	—	—	—	—	—	—	—	—				
Pasadena†	M		13						—	—	—	—											
Riverside*	M		3					—	—	—	—												
Rubidoux†	M	●	31				—	—	—	—	—	—	—	—	—	—	—	—	—				
PACIFIC NORTHWEST																							
Deschutes Cnty., OR	N		10						—	—	—	—	—		—	—	—	—					
Portland, OR	M	●	16							—	—	—	—	—	—	—	—	—	—				
Seattle, WA	M	●	16							—	—	—	—	—	—	—	—	—	—	—			
ARID SOUTHWEST																							
Carefree, AZ	S		6					—	—	—	—	—											
Phoenix, AZ	M		11					—	—	—	—	—	—										
El Paso (0002), TX	M		5								—					—	—	—	—				
El Paso (0004), TX	M	●	16							—	—	—	—	—	—	—	—	—	—				

Period of data record

Number of days with complete data for all variables

TABLE 2.3 LIST OF THE 75 STUDY SITES (Continued).

SITE	SITE TYPE	ANNUAL DATA SET	M J J A S O N D J F M A M J J A S O N D
Winnemucca, NV	N		11
NORTH CENTRAL			
Will Cnty., IL	N		9
Kansas City, KS	M		15
Minneapolis (0049), MN	M	●	22
Minneapolis (0051), MN	M	●	17
Afton, MO	M		15
Kansas City, MO	M		12
NORTHEAST			
Hartford, CT	M		11
Litchfield Cnty., CT	N		2
Dover, DE	M		8
Washington, DC	M	●	10
Acadia Nat'l Park, ME	N		2
Baltimore, MD	M		3
Boston (0012), MA	M		8
Boston (0013), MA	M		11
Ocean Cnty., NJ	N		21
Buffalo (0003), NY	M	●	18
Buffalo (0010), NY	M	●	11
Erie Cnty., NY	S	●	17
Lackawanna, NY	M		1
New York City, NY	M		1
Akron, OH	M	●	21

TABLE 2.3 LIST OF THE 75 STUDY SITES (Continued).

SITE	SITE TYPE	ANNUAL DATA SET	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Cincinnati, OH	M	●		18		—	—	—	—	—	—	—	—	—	—	—	—					
Cleveland (0013), OH	M				6		—	—	—	—	—	—					—					
Cleveland (0021), OH	M									3		—	—				—					
Medina, OH	S													5	—	—	—					
Middletown (0006), OH	S										32		—	—	—	—	—					
Middletown (0007), OH	S										30		—	—	—	—	—					
Steubenville, OH	S						1		—													
Allegheny Cnty., PA	S						3		—	—	—											
Downingtown, PA	S			24		—	—															
Philadelphia (0003), PA	M		28 —				—	—	—	—	—	—	—	—	—	—						
Philadelphia (0019), PA	M					16		—	—	—	—	—										
Philadelphia (0020), PA	M		3	—	—																	
Philadelphia (0024), PA	M	●	24	—	—	—			—	—	—	—	—	—	—	—	—	—	—			
Philadelphia (0036), PA	M		15	—	—	—			—	—			—			—	—	—	—			
Philadelphia (0037), PA	M		2	—	—																	
Philadelphia (0038), PA	M					13		—	—	—	—											
Philadelphia (0040), PA	M					16		—	—	—	—	—										
Philadelphia (0041), PA	M					17		—	—	—	—	—										
Philadelphia (0042), PA	M					24		—	—	—	—	—										
Philadelphia (0043), PA	M					30		—	—	—	—	—										
Philadelphia (0044), PA	M					27		—	—	—	—	—										
Pittsburgh, PA	M							2		—	—											
Hopewell, VA	S																1		—			

TABLE 2.3 LIST OF THE 75 STUDY SITES (Continued).

SITE	SITE TYPE	ANNUAL DATA SET	M J J A S O N D J F M A M J J A S O N D
SOUTHEAST			
Birmingham (0003), AL	M	•	10
Birmingham (0023), AL	M		6
Birmingham (0026), AL	M	•	12
Center Point, AL	S		11
Mountain Brook, AL	M		11
Tarrant, AL	M	•	14
Atlanta, GA	M		1
Durham, NC	N	•	13
Dallas, TX	M		12
Harris Cnty., TX	M		3
Houston, TX	M		1
Seabrook, TX	M		4

† Includes some data with ARB Hi-Vol matched to EPA dichotomous sampler.

* ARB monitoring program.

NOTE: Site types are M - metropolitan, S - suburban, and N - nonurban.

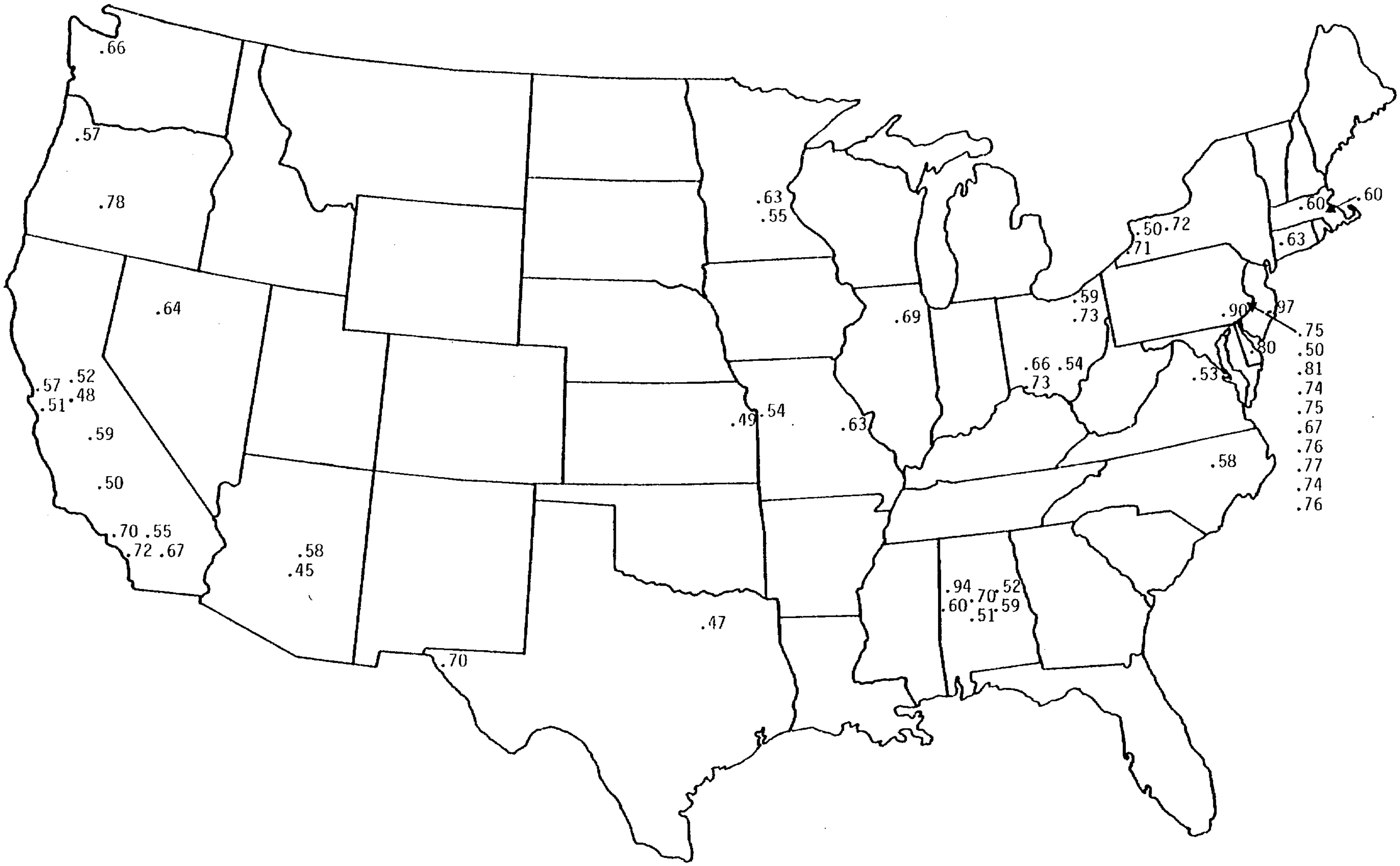


Figure 2.3 Geographical distribution of the ratio of average IP to average TSP (note that only sites with at least six data points are included).

"suburban", and "nonurban". Metropolitan sites are those locations within the densely populated parts of large urban centers (population exceeding 500,000). Nonurban sites are in towns with less than 30,000 population that are at least 15 miles from the outskirts of large urban centers. Suburban sites are the remainder -- either sites in cities of 30,000 to 500,000 population or sites in small towns that are within 15 miles of large urban centers. It should be noted that our classification differs considerably from the EPA site-type categorization which focuses on the local environment. For example, we call Rubidoux CA "metropolitan", while EPA categorizes it as "rural", and we call Winnemucca NV "nonurban", while EPA categorizes it as "center-city".

A special "annual" data base was constructed with the measurements from 21 sites that had nearly a full year of data.* There are two purposes for this "annual" data base, First, we want to evaluate the errors in our equations (relating IP or FP to Hi-Vol parameters) for both annual mean predictions and daily predictions. The yearly averages for the sites in the "annual" data base permit a direct calculation of the errors in our predictions of annual means. Second, we want to use the "annual" data base to check that seasonal biases are not introduced by the spotty records that exist at most of the sites in the full data base. With respect to this second purpose, we have been able to show that the predictive equations derived from the full data base are not seasonally biased. This is demonstrated by the fact that we consistently obtain very similar results in various regression models whether we use the full data base (930 points) or the "annual" data base (348 points). To give the reader a hint of the similarity between the full and "annual" data bases, Table 2.4 presents the means for the most important parameters. Table 2.4 shows that the two data bases are close in

*Sites qualified for the "annual" data base if they had a twelve month period of record with at most one 3-month gap. The gap could occur at one end, as well as in the middle, so that nine full months of data would qualify a site. If there were more than twelve months of data at a site, we selected the twelve month period with the greatest number of data points.

an average sense; the similar regression results obtained for the two data bases show that they are also close with respect to the basic correlations among parameters.

TABLE 2.4 COMPARISON OF AVERAGE VALUES FOR THE FULL DATA BASE AND ANNUAL DATA BASE.

		FULL DATA BASE	ANNUAL DATA BASE
NUMBER OF SITES		75	21
NUMBER OF DATA POINTS		930	348
AVERAGES FOR PARAMETERS			
	$\overline{IP}$	47.5 $\mu g/m^3$	46.3 $\mu g/m^3$
	$\overline{FP}$	24.6 $\mu g/m^3$	23.2 $\mu g/m^3$
	$\overline{TSP}$	74.5 $\mu g/m$	73.9 $\mu g/m^3$
	$\overline{SO_4^=}$	9.5 $\mu g/m^3$	8.8 $\mu g/m^3$
	$\overline{Pb}$	0.53 $\mu g/m^3$	0.47 $\mu g/m^3$
	$\overline{NO_3^-}$	3.63 $\mu g/m^3$	3.98 $\mu g/m^3$
	$\overline{FP/IP}$	0.52	0.50
	$\overline{IP/TSP}$	0.64	0.63

The reader might ask why we did not choose to analyze <u>only</u> the annual data base. The answers are breadth of coverage and statistical robustness. By using the full data base we have a much wider spatial sample, 75 locations rather than 21 locations. Also, by using the full data base, there are 930 rather than 348 data points to develop and evaluate the predictive equations.

Table 2.5 indicates the number of points in the full data base by geographical region, site type, and season.* One outstanding feature in Table 2.5 is the large number of data points for the Northeast. Much of this is due to the large data set (239 points) for the sites in-and-near Philadelphia. Also, Table 2.5 shows that nearly three-fourths of the data

*We have defined summer as April to September and winter as October to March.

nationwide are from metropolitan sites rather than suburban or nonurban sites.

TABLE 2.5 SUMMARY OF DATA POINTS IN THE FULL DATA SET.

	NUMBER OF DATA POINTS		
	All Data	Summer	Winter
FULL DATA SET	930	440	490
BY REGION:			
Cal.-San Francisco	53	27	26
Cal.-Central Valley	35	16	19
Cal.-Los Angeles	78	26	52
Pacific Northwest	42	19	23
Arid Southwest	49	22	27
North Central	90	49	41
Northeast	485	220	265
Southeast	98	61	37
BY SITE TYPE:			
Metropolitan	681	---	---
Suburban	173	---	---
Nonurban	76	---	---

2.2 METHODOLOGY FOR DEVELOPING RELATIONSHIPS

The purpose of the development phase of this study is to derive simple linear equations that predict IP (or FP) in terms of Hi-Vol data for TSP, $SO_4^=$, Pb, and NO_3^-. In deriving these equations, we have adopted a hybrid approach -- part physico-chemical, part statistical. Specifically, for the $SO_4^=$ and Pb terms, we can (and do) use physico-chemical concepts to specify what the coefficients should be. For the TSP and NO_3^-, we cannot unambiguously determine the correct coefficients from physical principles, so we estimate them statistically.

This hybrid approach has obvious advantages over a purely statistical methodology. First, credibility increases when physical meaning is attached to the coefficients in the equations. Pre-specifying 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 (see later discussion). A second, related advantage involves the fact that our equations will be applied to new situations -- other locations and other years. In the case of a purely statistical model, how can we be sure that the coefficients are still applicable in these new situations? Or, how can we even estimate the degree to which they might be in error? In the case of our hybrid model, we can state why and how the $SO_4^=$ and Pb coefficients should change. Furthermore, because the TSP coefficient also has a physical interpretation, we can do a reasonable sensitivity analysis on how much it might change. These factors make the hybrid model much more adaptable to new and varied situations. Finally, these advantages of credibility and adaptability are accrued at very little loss in predictive accuracy. Table 2.6 shows that the correlations are only slightly less and the predictive errors only slightly greater for a hybrid model than they are for a best-fit statistical model.

TABLE 2.6 PERFORMANCE OF HYBRID FORMULA COMPARED TO BEST-FIT STATISTICAL FORMULA.*

	CORRELATION COEFFICIENT	PREDICTION ERROR Absolute	Percent
STATISTICAL IP EQUATION	.901	12.6 $\mu g/m^3$	26.5%
HYBRID IP EQUATION	.899	12.8 $\mu g/m^3$	26.9%
STATISTICAL FP EQUATION	.840	8.9 $\mu g/m^3$	36.2%
HYBRID FP EQUATION	.828	9.2 $\mu g/m^3$	37.4%

*Both the statistical and hybrid formula are single nationwide equations evaluated against the entire data base (930 data points).

2.2.1 General Approach

Our objective is to develop predictive formulae for IP and FP that are simple linear equations involving the Hi-Vol parameters TSP, $SO_4^=$, Pb, and NO_3^-. As discussed above, the hybrid approach uses physico-chemical coefficients for $SO_4^=$ and Pb with statistical coefficients for TSP and NO_3^-. This section describes the methodology for deriving those coefficients.

For sulfates, we must answer the question: To what extent does a Hi-Vol measurement of $SO_4^=$ represent a direct contribution to IP (or FP)? In answering this question, we first account for the mass of the cation associated with $SO_4^=$. Assuming that the chemical form is ammonium sulfate, the total sulfate mass is 1.38 x $SO_4^=$. Next, we note that Hi-Vol measurements include more sulfate than IP or FP due to artifact sulfate formation on the Hi-Vol glass fiber filters. To account for this artifact, we reduce our 1.38 factor by about 15% (Coutant 1977; Stevens et al. 1978; Tanner et al. 1978) to 1.2. Finally, based on sulfate size distribution data reported by Lundgren (1970), Stevens et al. (1978, 1980), Whitby and Sverdrup (1978), Tanner (1979), and Trijonis et al. (1980), we assume that all of the sulfate is inhalable (less than 15 μm in size) and 90% is fine (less than 2.5 μm in size). Taking all of the above into account, we arrive at the end result:

$$\text{IP sulfate} \simeq 1.2 \text{ Hi-Vol } SO_4^= \quad (1\text{-a})$$

$$\text{FP sulfate} \simeq 1.1 \text{ Hi-Vol } SO_4^= \quad (1\text{-b})$$

Lead will be used to represent the contribution to IP (or FP) due to primary particulate matter from all road vehicles (gasoline vehicle exhaust, diesel exhaust, tire wear, and brake wear). The coefficient of interest is the ratio of inhalable (or fine) vehicular particulate emissions to suspendable vehicular Pb emissions. For the time period of our data sets, 1979-1980, calculations and data presented in Appendix A indicate that the appropriate coefficients are as follows:

$$\text{IP vehicular} \simeq 15 \times \text{Hi-Vol Pb} \quad (2\text{-a})$$

$$\text{FP vehicular} \simeq 11 \times \text{Hi-Vol Pb} \quad (2\text{-b})$$

It should be noted that the appropriate Pb coefficients change over time, depending on the percentage of catalyst-equipped vehicles and on the amount of lead in leaded gasoline. Appendix A explicitly shows how these coefficients have varied historically; for example the IP coefficient was 7 and the FP coefficient was 5 in the early 1970s. As will be discussed later in the application phase of this project, the lead coefficients must be changed when applying our equations to data sets from years other than 1979-1980.

The TSP and NO_3^- coefficients in our equations will be estimated statistically rather than calculated from physical principles. Actually, the TSP coefficient in our equations does have a physical interpretation; it basically represents the fraction of non-sulfate, non-vehicular TSP that is inhalable (or fine). Despite this simple physical meaning, we cannot calculate the TSP coefficient from first principles due to inadequate information. There is insufficient knowledge regarding emission factors for fugitive dust sources, particle size-distributions for various sources (both conventional and fugitive), the relative dispersion characteristics and relative ambient contributions for fugitive and conventional sources, and the relative contributions of primary aerosols versus non-sulfate secondary aerosols.

For nitrates, one might initially think that a calculation can be performed similar to the one performed above for sulfates. However, unlike sulfates, which have a relatively minor artifact measurement problem, nitrates involve major measurement errors due to both artifacts and interferences (Spicer and Schumacher 1979; Appel et al. 1979; Harker et al. 1977). Because of these severe measurement difficulties, we will treat NO_3^- as a purely statistical variable and will not assign the NO_3^- coefficient a physical meaning. In the next section, the predictive equations for IP and FP will be organized in a hierarchy based on the number of variables included. Because the NO_3^- coefficient is the most difficult to interpret and understand, the NO_3^- variable will be introduced only in the last step.

The methodology for estimating the TSP and NO_3^- coefficients can be best explained by example. For instance, one set of formulae for IP and FP will be based on data for TSP, $SO_4^=$, and Pb. The TSP coefficient (b) is determined through a (zero intercept) univariate regression equation:

$$\left.\begin{array}{l} IP - 1.2\ SO_4^= - 15\ Pb \\ \text{or} \\ FP - 1.1\ SO_4^= - 11\ Pb \end{array}\right\} = b(TSP - 1.4\ SO_4^= - 15\ Pb) \qquad (3)$$

That is, "b" is chosen to provide a least squares fit to the independent variable IP - 1.2 $SO_4^=$ - 15 Pb or FP - 1.1 $SO_4^=$ - 11 Pb.

In order to give the coefficient "b" a more meaningful physical interpretation, 1.4 $SO_4^=$ (the Hi-Vol ammonium sulfate) and 15 Pb (suspendable vehicular particles measured by the Hi-Vol) are subtracted from TSP so that the dependent variable represents non-sulfate, non-vehicular TSP. The final predictive formula for IP or FP is obtained by simple algebraic manipulation, i.e.

$$IP = 1.2\ SO_4^= + 15\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb). \tag{4}$$

Another set of formulae for IP and FP will be based on all four Hi-Vol variables. In this case, the TSP and NO_3^- coefficients, b_1 and b_2, are determined through a (zero intercept) <u>multiple</u> regression equation:*

$$\left.\begin{array}{l} IP - 1.2\ SO_4^= - 15\ Pb \\ \text{or} \\ FP - 1.1\ SO_4^= - 11\ Pb \end{array}\right. = b_1\ (TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2\ NO_3^-. \tag{5}$$

There are very strong reasons for our decision to exclude the intercept parameter (an arbitrary additive constant) in the regressions. In most of our regression analyses, we found that the intercept was statistically insignificant. Accordingly, adding an intercept produced negligible reduction in the errors of the predictive equations. Also, excluding an intercept made our coefficients more interpretable in many cases. For example, the coefficient "b" discussed above is just the fraction of non-sulfate, non-vehicular TSP that is inhalable (or fine). Furthermore, a zero intercept allowed us, in some instances, to examine site to site variations simply by taking ratios. Finally, greater credibility is added to our results in the sense that we have obtained good statistical fits with only one free parameter when NO_3^- is not included, or only two free parameters when NO_3^- is included.

2.2.2 Stepwise Addition of Variables

In developing predictive formulae for IP and FP, we will add independent variables in a stepwise fashion, starting with TSP and ending with all four Hi-Vol variables. This stepwise progression will allow us to investigate the trade-off between simplicity (fewer independent variables) and

*Note that 1.3 NO_3^- (equivalent to ammonium nitrate) is subtracted from the TSP term to make the two regression variables physically "independent."

accuracy. Even more importantly, we must develop equations involving subsets of the variables (e.g. just TSP, or just TSP and $SO_4^=$) because data for some of the variables will not be available in certain applications (e.g. some routine Hi-Vol monitoring sites provide only TSP data or only TSP and $SO_4^=$ data).

The variables will be added in the order: TSP, $SO_4^=$, Pb, and NO_3^-. This is approximately the order in which the variables are most frequently available from routine Hi-Vol monitoring programs. That is -- TSP is always available from Hi-Vol monitoring, sulfate is the most commonly measured chemical species, and Pb and NO_3^- are the next most commonly measured chemical species. Fortunately, as shown in Table 2.7, the chosen sequence is also the approximate order in which the variables best correlate to IP and FP.

TABLE 2.7 CORRELATION OF IP AND FP TO INDIVIDUAL HI-VOL PARAMETERS.*

	CORRELATION COEFFICIENTS			
	TSP	$SO_4^=$	Pb	NO_3^-
IP	0.87	0.48	0.52	0.57
FP	0.61	0.64	0.55	0.47

*Based on all 930 nationwide data points.

The first step in our analysis uses data only for TSP. Zero-intercept, one-parameter regressions are run using the equation:

$$IP \text{ (or FP)} = b \, TSP \tag{6}$$

In this case, "b" can be interpreted simply as the fraction of TSP that is inhalable (or fine). The second step assumes Hi-Vol data for TSP and $SO_4^=$. The regression form is then,

$$\left.\begin{array}{c} IP - 1.2\ SO_4^= \\ \text{or} \\ FP - 1.1\ SO_4^= \end{array}\right\} = b(TSP - 1.4\ SO_4^=) \tag{7}$$

with "b" now representing the fraction of non-sulfate TSP that is inhalable (or fine). The third step -- using data for TSP, $SO_4^=$, and Pb -- is based on the regression equation,

$$\left.\begin{array}{l} IP - 1.2\ SO_4^{=} - 15\ Pb \\ \text{or} \\ FP - 1.1\ SO_4^{=} - 11\ Pb \end{array}\right\} = b(TSP - 1.4\ SO_4^{=} - 15\ Pb) \qquad (8)$$

also given previously as Equation (3). As discussed in the previous section, the coefficient "b" in this equation represents the fraction of non-sulfate, non-vehicular TSP that is inhalable (or fine). The final step, based on all four Hi-Vol parameters, uses a multiple regression equation.

$$\left.\begin{array}{l} IP - 1.2\ SO_4^{=} - 15\ Pb \\ \text{or} \\ FP - 1.1\ SO_4^{=} - 11\ Pb \end{array}\right\} = b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2\ NO_3^{-}$$

At each step, the regression equations are first determined for the entire national data base, 930 daily data points. Next, the regressions are run individually for the three site-types, for the eight geographical regions, and for the eight regions separated into two seasons. Using a statistical significance level of 95% (two standard deviations), an evaluation is then made as to whether the site-type/regional/seasonal results differ from the national results.

The single national equation, the three site-type equations, the eight regional equations, and the sixteen regional/seasonal equations are each evaluated as "predictive models". The errors in the national equation and the disaggregated site-type/regional/seasonal models are evaluated on a daily basis against the entire 930 data points, on an annual basis against the 21 yearly means in the special "annual" data set, and on a site-type/regional/seasonal basis against the data for the specific site-types, regions, and seasons. The evaluations against the entire 930 data points and the 21 "annual" data points allow us to see how much the site-type/regional/seasonal equations improve over the national equation in an overall sense. The individualized evaluations on a site-type/regional/seasonal basis allow us to see how much the site-type/regional/seasonal results improve over the national equation for various site-type/regional/seasonal subsets of the data.

The error measure used in the evaluations is the routine root-mean-square error (i.e. standard deviation). This error can be expressed either in an absolute sense (e.g. in $\mu g/m^3$) or in a relative sense (e.g. as a percent of average IP or FP). For example, our national equation based on all four Hi-Vol parameters [e.g. Equation (9)] yields an error of 12.9 $\mu g/m^3$ in predicting daily IP, which is 27% of average IP, and an error of 9.3 $\mu g/m^3$ in predicting daily FP which is 38% of average FP.

In applying our formulae to new data sets, it is important for the user to know if he should use our absolute error or our percent error. For example, in a specific application, if a user arrives at a predicted IP of 60 $\mu g/m^3$, should he assume the absolute error of 12.9 $\mu g/m^3$, or should he assume a 27% error, which now represents 16.2 $\mu g/m^3$? To examine whether a constant absolute error applies over the entire range of IP and FP or whether the error tends to grow in proportion to IP and FP, we have evaluated the error as a function of the magnitude of IP and FP for our four-parameter national equation. This has been done simply by sorting the data into various magnitude ranges of IP and FP and by evaluating the error for each of these ranges. The results are shown in Figures 2.4 and 2.5. These figures show that the error, as a function of the magnitude of IP and FP, tends to fall between the two lines represented by a constant absolute error and a proportional percentage error. The errors, however, are much closer to the proportional line, indicating that a percentage error is more accurate than an absolute error. Accordingly, all of our results will be specified in terms of percentage error.

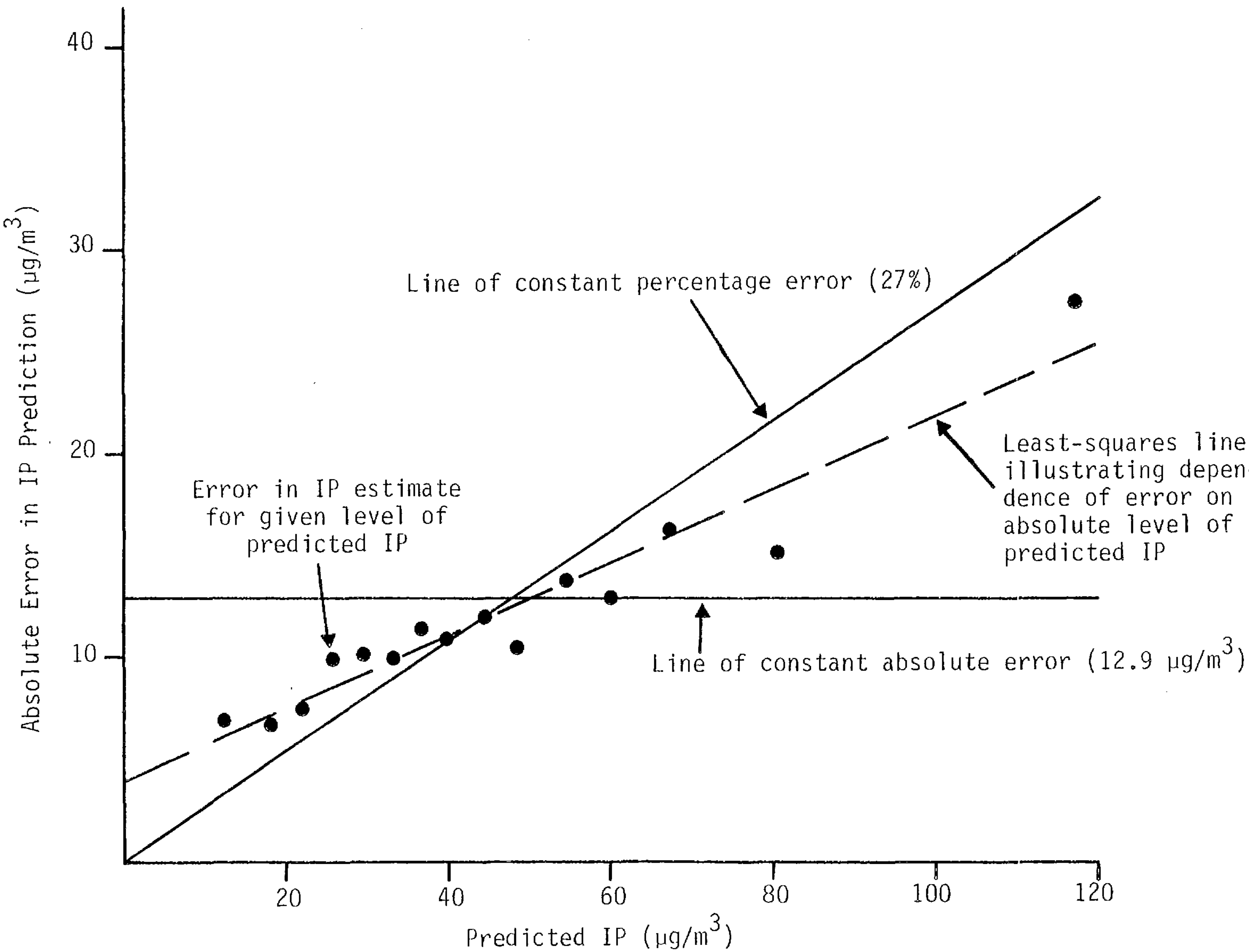


Figure 2.4 Error in IP estimates as a function of the magnitude of predicted IP (based on the nationwide equation using all four Hi-Vol variables).

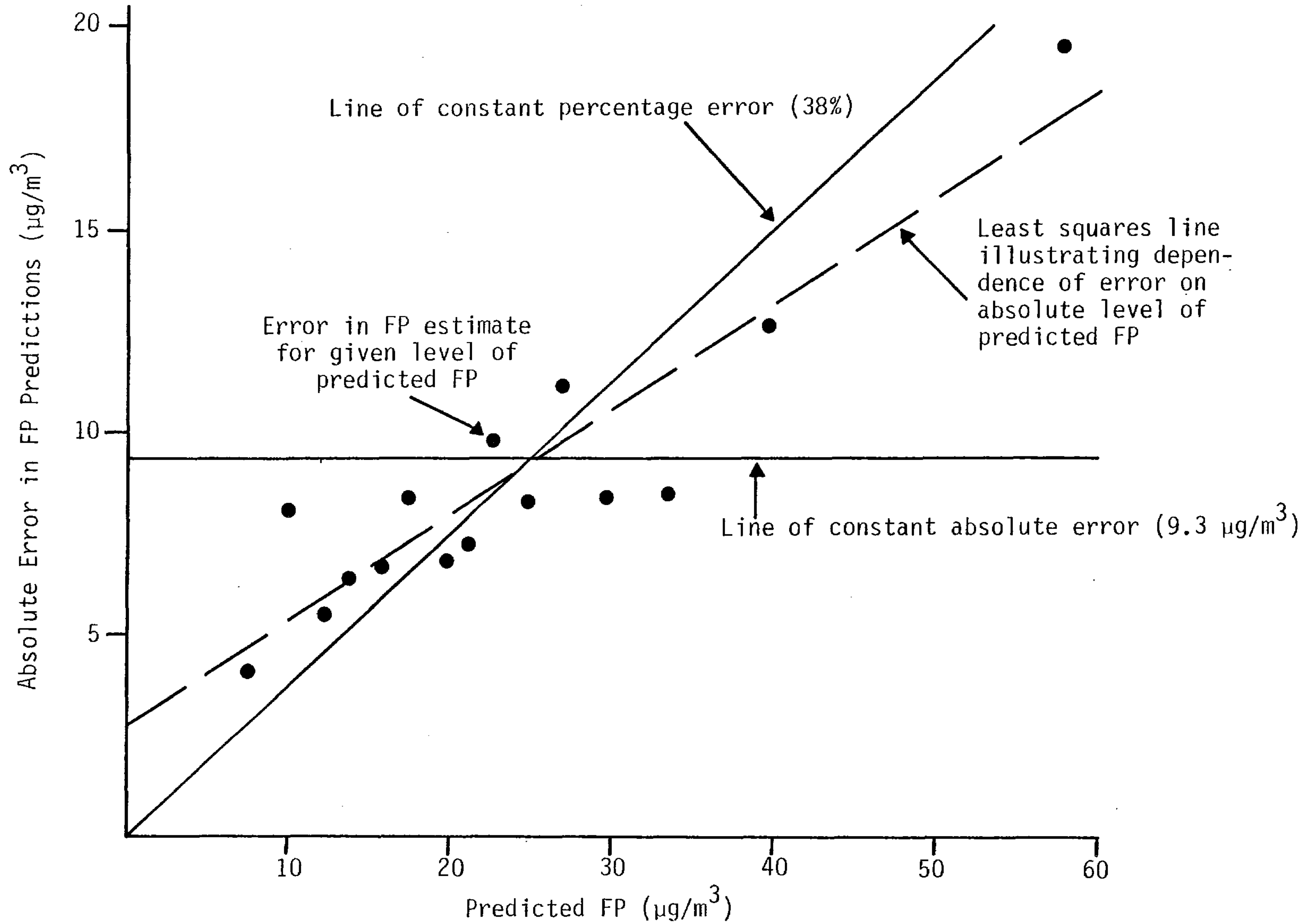


Figure 2.5 Error in FP estimates as a function of the magnitude of predicted FP (based on the nationwide equation using all four Hi-Vol variables).

3. RELATIONSHIP OF IP AND FP TO HI-VOL DATA

Using the data base and methodology described in the previous chapter, this chapter develops formulae for estimating dichotomous parameters (IP and FP) from Hi-Vol parameters (TSP, $SO_4^=$, Pb, and NO_3^-). Formulae are also developed for estimating size-selective-inlet Hi-Vol data (SSIP) from routine Hi-Vol data. The chapter is organized into three major sections dealing with IP, FP, and SSIP, respectively. Within each section, the independent variables are added in a stepwise progression starting with TSP and ending with all four Hi-Vol parameters. At each step, predictive equations are developed and evaluated on an aggregate nationwide basis, a site-type basis, a regional basis, and a regional/seasonal basis. Although the reader is free to use any of the equations presented herein, we do conclude each section with recommendations concerning the equations that we think are most appropriate.

3.1 ESTIMATION OF IP

This section develops and evaluates equations for estimating IP from Hi-Vol data. The first four subsections deal with the sequential addition of independent variables (Hi-Vol parameters), starting with TSP and ending with TSP, $SO_4^=$, Pb, and NO_3^-. The last subsection summarizes our findings and recommendations.

3.1.1 IP Versus TSP

In the simplest case, we assume that only TSP data are available for estimating IP. As indicated in the methodology discussion (Section 2.2), the equation for predicting IP is then

$$IP = b \cdot TSP \tag{10}$$

The coefficient "b" -- determined from a zero intercept, one-parameter regression -- simply measures that fraction of TSP that represents IP.

Table 3.1 lists the values for "b" determined from the national, site-type, regional, and regional/seasonal data bases. Table 3.1 also includes the standard error for "b". Those site-type or regional coefficients that

TABLE 3.1 COEFFICIENTS FOR THE EQUATION IP = b·TSP.

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)	REGIONAL/SEASONAL: Summer		REGIONAL/SEASONAL: Winter
NATIONAL:	0.61 ± .006			
SITE-TYPE:				
Metropolitan	0.62 ± .006			
Suburban	0.57* ± .014			
Nonurban	0.68* ± .029			
REGIONAL:				
Cal.-San Francisco Area	0.53* ± .024	0.43 ± .021	**	0.60 ± .036
Cal.-Central Valley	0.51* ± .029	0.43 ± .036	**	0.59 ± .036
Cal.-Los Angeles Area	0.66* ± .016	0.70 ± .026	**	0.62 ± .019
Pacific Northwest	0.55* ± .024	0.49 ± .020	**	0.70 ± .036
Arid Southwest	0.58 ± .021	0.60 ± .030		0.56 ± .031
North Central.	0.54* ± .015	0.53 ± .018		0.58 ± .027
Northeast	0.66* ± .008	0.63 ± .011	**	0.67 ± .011
Southeast	0.57* ± .016	0.57 ± .019		0.57 ± .033

*Differs from national aggregate value at 95% confidence level.

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

differ from the national value (0.61) at a 95% confidence level are marked with an asterisk. The summer/winter differences that are statistically significant at a 95% confidence level are marked by a double asterisk.

Table 3.1 reveals that there are statistically significant differences among the regions in the fraction of TSP that is IP (i.e. in the coefficient "b"). The fraction is lowest in the San Francisco area and in the Central Valley of California (about 0.52) and highest in the Los Angeles area and in the Northeast United States (about 0.66). The fraction is slightly lower than the national value (0.61) in the Pacific Northwest, Arid Southwest, North Central area, and Southeast. The high values in Los Angeles and the Northeast indicate that a greater fraction of the aerosol is inhalable (e.g. of smaller size) in those regions. This may be directly related to the fact that Los Angeles and the Northeast are the national hot spots for high sulfate concentrations and low visibility (e.g. high fine particle concentrations) (EPA 1975; Trijonis and Shapland 1979; Trijonis 1980).

Table 3.1 also reveals statistically significant differences in the coefficients among site-types. Specifically, suburban sites have a relatively low IP/TSP ratio, and nonurban sites have a relatively high IP/TSP ratio. The reason for these differences is not obvious, but the high values at nonurban sites may reflect relatively greater secondary aerosol levels (compared to total TSP levels). The high values for nonurban sites could also represent lesser artifact formation on the Hi-Vol filters at nonurban locations (due to lower SO_2 and nitric acid levels); less artifact on the Hi-Vol would lead to lower TSP levels and higher IP/TSP ratios.

Most of the seasonal differences in Table 3.1 are also statistically significant. The San Francisco Bay area and Central Valley have a much higher IP/TSP ratio during the winter than during the summer; this likely reflects the fact that winter is the season of high fine particle concentrations (high sulfate, high nitrate, and low visibility) in San Francisco and the Central Valley (Trijonis 1980). The low IP/TSP ratio in San Francisco and the Central Valley during the summer, however, could also reflect greater dust concentrations during the dry summer season (leading to an upward shift in the particle size distribution). Opposite to the pattern in San Francisco and the Central Valley, the Los Angeles area has a higher IP/TSP ratio during the summer, when Los Angeles experiences extremely low values of visibility and extremely high values of sulfates and other photochemical aerosols (Trijonis 1980). It is noteworthy that the seasonal patterns differ significantly from region to region. This is the reason why we chose not to present seasonal differences on an aggregated national scale.

We have also performed the regressions with Equation (10) using the 348 data points in the "annual" data set and using the 21 annual means derived from the "annual" data set. In both cases we again obtain the coefficient 0.61. As discussed earlier in Section 2.1.3, this reflects the fact that our aggregate data base of 930 points does not contain a significant seasonal bias.

Although most of the site-type, regional, and regional/seasonal variations in "b" are significant from a statistical standpoint, it is not obvious that they are significant from a practical standpoint. To investigate this issue, we have evaluated the predictive errors in four types of models: (1) the simple national equation, (2) a set of site-type equations, (3) a set of regional equations, and (4) a set of regional/seasonal equations. The aggregate national equation is simply, IP = 0.61 TSP, applied to all the data points. The site-type model involves three equations, corresponding to the three site-type coefficients listed in Table 3.1, and applied individually to the data from the three specific site types. Similarly, the regional model involves eight equations, and the regional/seasonal model involves sixteen equations; these are applied to data for the specific region or region/season.

The (root-mean-square) errors for the four models have been evaluated on both an individualized (site-type, regional, or regional/seasonal) basis and a generalized national basis. Table 3.2, organized by site-type, region, and region/season, compares the errors in predicting daily values for the various specific equations to the corresponding errors obtained by just using the national equation. Table 3.3 summarizes the overall errors on a national basis for predicting both daily values and annual means.

TABLE 3.2 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATION OF THE FORM IP = b·TSP.

PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET
(PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)

		REGIONAL/SEASONAL	
		Summer	Winter
SITE-TYPE:			
Metropolitan	29.3 (29.3)		
Suburban	35.7 (36.8)		
Nonurban	38.7 (40.1)		
REGIONAL:			
Cal.-San Francisco Area	37.6 (41.5)	26.4 (51.4)	35.2 (35.2)
Cal.-Central Valley	38.0 (43.7)	33.8 (55.7)	33.3 (33.5)
Cal.-Los Angeles Area	24.1 (25.5)	22.0 (26.9)	24.1 (24.2)
Pacific Northwest	30.7 (32.7)	21.1 (36.1)	26.6 (30.1)
Arid Southwest	29.3 (29.8)	26.0 (26.1)	32.9 (34.5)
North Central	27.1 (30.3)	24.6 (29.7)	30.4 (31.0)
Northeast	28.9 (29.8)	28.1 (28.2)	28.8 (30.7)
Southeast	30.3 (31.4)	28.1 (29.3)	35.8 (36.5)

TABLE 3.3 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM IP = b·TSP.

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.860	31.3%	0.913	15.8%
3 SITE-TYPE EQUATIONS	0.863	31.0%	0.909	16.2%
8 REGIONAL EQUATIONS	0.877	29.6%	0.930	14.3%
16 REGIONAL/ SEASONAL EQUATIONS	0.886	28.5%	Not Applicable	

It is obvious from Tables 3.2 and 3.3 that the three disaggregated models are not that much more accurate than the national equation. As shown in Table 3.2, even on an individualized site-type and regional basis, the national equation performs almost as well as the specific equations for various site-types and regions. It is only when we consider certain individual regions and seasons (e.g. San Francisco area in the summer, Central Valley in the summer, and Pacific Northwest in the summer) that the national equation sometimes performs substantially worse than the specific equations. As a caution, however, we note that these particular regions/seasons are among those with the fewest data points (see previous Table 2.5, page 32).

On an overall basis, the national equation predicts daily values with a 31.3% error and annual means with a 15.8% error. As shown in Table 3.3, the overall, daily prediction errors are reduced only a slight amount by the disaggregated models -- to 31.0% for the site-type model, to 29.6% for the regional model, and to 28.5% for the regional/seasonal model. The regional model reduces the error in annual mean predictions from 15.8% to 14.3%, but the site-type model actually increases the error when it is evaluated against the annual mean data base. Summarizing, we find little if any improvement over the national equation with the site-type model, slight improvement with the regional model, and further slight improvement with the regional/seasonal model.

3.1.2 IP Versus TSP and $SO_4^=$

The second step in deriving prediction equations for IP assumes that Hi-Vol data are available for TSP and $SO_4^=$. As discussed in Section 2.2, physical considerations lead us to fix the $SO_4^=$ coefficient and to conduct a zero-intercept regression of the form:

$$IP - 1.2\ SO_4^= = b(TSP - 1.4\ SO_4^=). \tag{11}$$

Here, the coefficient "b" represents the fraction of non-sulfate TSP that is inhalable. The final predictive equation is then

$$IP = 1.2\ SO_4^= + b(TSP - 1.4\ SO_4^=). \tag{12}$$

Table 3.4 lists values for "b" determined from the national, site-type, regional, and regional/seasonal data bases. As was the case previously in Table 3.1, we find many statistically significant variations in the coefficient among the site-types, regions, and seasons. In fact, the site-type, regional, and seasonal variations in Table 3.4 are very similar to those in Table 3.1. In the previous section, we offered the explanation that these variations might be related to patterns in sulfate and other fine particle concentrations (e.g. sulfate and other fine particle concentrations may be relatively high compared to TSP in Los Angeles and the Northeast, at nonurban locations, during the winter season in San Francisco and the Central Valley, etc.). That very similar variations remain once we have explicitly discounted for sulfates in our equation suggests that non-sulfate fine particles may be at least as important as sulfates in this explanation.

As was done in the previous section, we again have evaluated four types of predictive models: (1) the simple national equation, (2) a set of three site-type equations, (3) a set of eight regional equations, and (4) a set of sixteen regional/seasonal equations. Table 3.5 compares the percentage errors in applying the national equation to percentage errors in applying the equations specific to each site-type, region, and region/season. As

TABLE 3.4 COEFFICIENTS FOR THE EQUATION IP = 1.2 $SO_4^=$ + b(TSP - 1.4 $SO_4^=$).

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)	REGIONAL/SEASONAL: Summer		REGIONAL/SEASONAL: Winter
NATIONAL:	0.56 ± .006			
SITE-TYPE:				
Metropolitan	0.58* ± .007			
Suburban	0.51* ± .014			
Nonurban	0.59 ± .036			
REGIONAL:				
Cal.-San Francisco Area	0.50* ± .027	0.38 ± .026	**	0.58 ± .038
Cal.-Central Valley	0.49* ± .031	0.39 ± .039	**	0.57 ± .037
Cal.-Los Angeles Area	0.64* ± .016	0.68 ± .028	**	0.60 ± .019
Pacific Northwest	0.52 ± .024	0.46 ± .021	**	0.67 ± .039
Arid Southwest	0.56 ± .023	0.58 ± .032		0.54 ± .032
North Central	0.48* ± .015	0.47 ± .018		0.50 ± .030
Northeast	0.60* ± .009	0.54 ± .013	**	0.63 ± .012
Southeast	0.50* ± .017	0.51 ± .021		0.50 ± .033

*Differs from national aggregate value at 95% confidence level.

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

TABLE 3.5 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATION OF THE FORM $IP = 1.2\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$.

PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET
(PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)

		REGIONAL/SEASONAL	
		Summer	Winter
SITE-TYPE:			
Metropolitan	27.8 (27.9)		
Suburban	32.7 (33.8)		
Nonurban	36.2 (36.4)		
REGIONAL:			
Cal.-San Francisco Area	38.8 (40.7)	29.1 (49.6)	34.9 (35.1)
Cal.-Central Valley	37.7 (40.7)	34.0 (50.5)	32.6 (32.7)
Cal.-Los Angeles Area	23.3 (26.4)	21.8 (28.7)	23.0 (24.1)
Pacific Northwest	28.4 (29.4)	20.5 (30.4)	25.2 (29.2)
Arid Southwest	28.9 (28.9)	26.0 (26.3)	32.0 (32.3)
North Central	23.7 (27.6)	21.2 (26.7)	27.7 (29.0)
Northeast	27.1 (27.5)	25.2 (25.3)	27.2 (28.7)
Southeast	27.6 (29.0)	26.1 (27.5)	30.6 (32.0)

before, we find that the national equation yields almost as low an error as the equations specific to each site-type and region. Also, as before, we find that the national equation yields substantially greater error than the specific equations for certain regional/seasonal data bases (e.g. San Francisco in the summer, Central Valley in the summer, Los Angeles in the summer, and Pacific Northwest in the summer).

Table 3.6 summarizes the overall performance of the national equation and the disaggregated models, evaluated on both a daily and annual mean basis. As was the case previously, we see little if any improvement with the site-type equations, a slight improvement with the regional equations, and a further slight improvement with the regional/seasonal equations.

TABLE 3.6 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM IP = 1.2 $SO_4^{=}$ + b(TSP - 1.4 $SO_4^{=}$).

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.877	29.5%	0.916	15.5%
3 SITE-TYPE EQUATIONS	0.880	29.2%	0.912	15.9%
8 REGIONAL EQUATIONS	0.890	28.0%	0.941	13.1%
16 REGIONAL/ SEASONAL EQUATIONS	0.900	26.8%	Not Applicable	

3.1.3 IP Versus TSP, $SO_4^{=}$, and Pb

In the third step, we assume that data are available for TSP, $SO_4^{=}$, and Pb. With the $SO_4^{=}$ and Pb coefficients constrained by the physico-chemical considerations discussed in Section 2.2, the regression equation takes the form:

$$IP - 1.2\ SO_4^{=} - 15\ Pb = b(TSP - 1.4\ SO_4^{=} - 15\ Pb) \quad (13)$$

The final predictive equation is then,

$$IP = 1.2\ SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb) \quad (14)$$

The coefficient "b" represents the fraction of non-sulfate, non-vehicular TSP that is inhalable.

Table 3.7 presents the national, site-type, regional, and regional/seasonal values for the coefficient "b" in Equation (14). The site-type, regional, and regional/seasonal patterns in Table 3.7 are very similar to those in Table 3.1 (the TSP equation) and Table 3.4 (the TSP and $SO_4^{=}$ equation). As discussed previously in Section 3.1.1, these site-type, regional, and regional/seasonal patterns appear to be related to similar variations in the concentrations of sulfates and other fine particles. The fact that the same patterns remain after we have explicitly discounted for sulfates and primary vehicular particles indicates that non-sulfate, non-vehicular fine particles are important contributors to the observed patterns.

Table 3.8 lists the percentage errors that result when the specific equations and the national equation are applied to each site-type, region, and region/season. As before, the national equation yields only a slightly greater error than the specific equations for the various site-types and regions, but a substantially greater error than the specific equations for summertime in the San Francisco, Central Valley, and Los Angeles areas.

Table 3.9 summarizes the overall performance of the national equation, the 3-equation site-type model, the 8-equation regional model, and the 16-equation regional/seasonal model. As before, the site-type model produces little if any improvement over the national equation; the regional model yields a slight improvement, and the regional/seasonal model yields a further slight improvement.

3.1.4 IP Versus TSP, $SO_4^{=}$, Pb, and NO_3^{-}

The fourth and final step of the analysis uses data for all four Hi-Vol parameters. A multiple, zero-intercept regression is run of the form:

$$IP - 1.2\ SO_4^{=} - 15\ Pb = b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2 NO_3^{-} \quad (15)$$

TABLE 3.7 COEFFICIENTS FOR THE EQUATION IP = 1.2 $SO_4^=$ + 15 Pb + b(TSP - 1.4 $SO_4^=$ - 15 Pb).

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)	REGIONAL/SEASONAL: Summer		REGIONAL/SEASONAL: Winter
NATIONAL:	0.50 ± .006			
SITE-TYPE:				
Metropolitan	0.52 ± .007			
Suburban	0.46* ± .016			
Nonurban	0.56 ± .036			
REGIONAL:				
Cal.-San Francisco Area	0.41* ± .026	0.33 ± .025	**	0.48 ± .039
Cal.-Central Valley	0.41* ± .035	0.32 ± .045	**	0.50 ± .045
Cal.-Los Angeles Area	0.59* ± .019	0.66 ± .030	**	0.53 ± .021
Pacific Northwest	0.48 ± .023	0.44 ± .021	**	0.60 ± .047
Arid Southwest	0.53 ± .027	0.56 ± .036		0.50 ± .040
North Central	0.45* ± .015	0.45 ± .018		0.46 ± .028
Northeast	0.53* ± .010	0.50 ± .013	**	0.54 ± .013
Southeast	0.47 ± .018	0.47 ± .022		0.42 ± .040

*Differs from national aggregate value at 95% confidence level.

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

TABLE 3.8 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATION OF THE FORM $IP = 1.2\ SO_4^= + 15\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$.

	PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET (PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)	REGIONAL/SEASONAL	
		Summer	Winter
SITE-TYPE:			
Metropolitan	25.8 (25.9)		
Suburban	32.7 (33.4)		
Nonurban	34.3 (35.0)		
REGIONAL:			
Cal.-San Francisco Area	31.6 (35.2)	25.8 (43.9)	29.6 (29.7)
Cal.-Central Valley	37.0 (40.5)	34.9 (49.6)	33.3 (33.3)
Cal.-Los Angeles Area	23.3 (26.5)	21.5 (31.2)	21.2 (21.6)
Pacific Northwest	24.9 (25.1)	19.7 (23.1)	24.9 (27.4)
Arid Southwest	31.8 (32.3)	27.5 (29.4)	36.1 (36.1)
North Central	22.5 (23.8)	21.3 (23.1)	24.4 (25.0)
Northeast	24.8 (25.0)	23.6 (23.6)	25.2 (25.6)
Southeast	27.3 (27.8)	25.4 (25.8)	31.4 (33.0)

TABLE 3.9 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM IP = 1.2 $SO_4^=$ + 15 Pb + b(TSP - 1.4 $SO_4^=$ - 15 Pb).

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.892	27.8%	0.913	15.8%
3 SITE-TYPE EQUATIONS	0.894	27.6%	0.911	16.0%
8 REGIONAL EQUATIONS	0.902	26.5%	0.942	13.0%
16 REGIONAL/ SEASONAL EQUATIONS	0.909	25.7%	Not Applicable	

The final predictive equation is then

$$IP = 1.2\ SO_4^= + 15\ Pb + b_1(TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2NO_3^- \qquad (16)$$

We will not present tables of the coefficients "b_1" and "b_2" for Equation (16) because this model does not perform well. The basic problem is that the NO_3^- coefficient, "b_2", is <u>unstable</u> and has <u>high standard errors</u>. Among the regions and seasons, the coefficient "b_2" varies erratically from -6.8 to +6.9. Part of this instability may be due to colinearity between NO_3^- and "TSP - 1.4 $SO_4^=$ - 15 Pb" (the correlation between these two "independent" variables is typically around 0.60). The high standard errors imply that "b_2" is statistically insignificant for most of the regions and seasons. Related to the fact that "b_2" is often statistically insignificant, we find that Equation (16) does not reduce the overall prediction error for IP very much compared to Equation (14) (the model that included only TSP, $SO_4^=$, and Pb). The marginal improvement of Equation (16) over Equation (14) will be presented quantitatively in the next section. One further reason for discarding this model is that, unlike our previous models, Equation (16) lacks a well-defined physical interpretation.

While we are on the subject of the NO_3^- variable, as an aside, we would like to point out a very interesting regional peculiarity. Table 3.10 indicates the relative importance of the four Hi-Vol variables in predicting IP or FP, as measured by the order in which the variables are selected by multiple, step-wise regressions. For example, Table 3.10 indicates that TSP is always the first variable selected in predictions of IP. The regional peculiarity involves the step-wise regressions for FP. For all five non-California regions, the NO_3^- variable is entered last. In contrast, NO_3^- is entered first in two of the California regions and third in the other California region. Visibility/aerosol regressions yield a similar finding -- that NO_3^- tends to be much more important in California than in other regions (Trijonis 1979, 1980; Trijonis and Yuan 1978; White and Roberts 1977). These results suggest that California probably has a very real nitrate aerosol problem (i.e. that the excessive Hi-Vol NO_3^- levels measured in California are not all or nearly all due to artifact formation).

TABLE 3.10 RELATIVE IMPORTANCE OF THE INDEPENDENT VARIABLES AS MEASURED BY ORDER OF ENTRANCE INTO A MULTIPLE STEPWISE REGRESSION.

	ORDER OF VARIABLE ENTRANCE IN STEPWISE REGRESSIONS							
	IP PREDICTIONS				FP PREDICTIONS			
	1	2	3	4	1	2	3	4
NATIONAL	TSP	$SO_4^=$	Pb	NO_3^-	$SO_4^=$	Pb	NO_3^-	TSP
REGIONAL								
Cal.-San Francisco Area	TSP	Pb	NO_3^-	$SO_4^=$	Pb	$SO_4^=$	NO_3^-	TSP
Cal. Central Valley	TSP	NO_3^-	$SO_4^=$	Pb	NO_3^-	$SO_4^=$	TSP	Pb
Cal.-Los Angeles Area	TSP	$SO_4^=$	NO_3^-	Pb	NO_3^-	Pb	$SO_4^=$	TSP
Pacific Northwest	TSP	$SO_4^=$	Pb	NO_3^-	$SO_4^=$	Pb	TSP	NO_3^-
Arid Southwest	TSP	Pb	$SO_4^=$	NO_3^-	TSP	Pb	$SO_4^=$	NO_3^-
North Central	TSP	$SO_4^=$	Pb	NO_3^-	$SO_4^=$	TSP	Pb	NO_3^-
Northeast	TSP	$SO_4^=$	Pb	NO_3^-	$SO_4^=$	Pb	TSP	NO_3^-
Southeast	TSP	$SO_4^=$	NO_3^-	Pb	$SO_4^=$	Pb	TSP	NO_3^-

3.1.5 Summary of Predictive Relationships for IP

Table 3.11 presents matrices summarizing the overall performance of the various IP predictive models described in the previous sections. The simplest model (the single national equation, IP = 0.61 TSP) has an error of 31.3% in predicting daily values of IP and 15.8% in predicting annual mean IP. The errors decrease as more parameters are added to the equations (downward in the matrices) and as site-type, regional, or regional/seasonal complexity is added (to the right in the matrices). These decreases in error are (and must be) monotonic for Table 3.11a which involves tests against the 930 daily data points used to derive the equations but are not monotonic for Table 3.11b which involves tests against 21 "independent" annual means. The most complex model (16-regional/seasonal equations using all four Hi-Vol parameters) achieves an overall error of 24.8% in predicting daily values and 13.0% in predicting annual means.

Noting that the errors in the various complex models are not that much smaller than the error in the simplest model (the single national equation, IP = 0.61 TSP), the reader might ask: "Why not just use the simplest model?" We think this simple approach is not unreasonable, and we would not argue strongly against it. We also note, however, that adding certain of the complexities is also very reasonable. For example, we think that the equations involving $SO_4^=$ (or $SO_4^=$ and Pb) should be used when data are available for $SO_4^=$ and/or Pb. The equations involving $SO_4^=$ and/or Pb reduce the predictive errors somewhat, and they disaggregate the predictions in a well-defined physical way. Furthermore, in some areas, especially in California, it makes sense to use the regional or regional/seasonal equations. The only two complexities that we definitely recommend not to use are the "site-type" equations (which produce little error reduction) and the equations involving NO_3^- (which produce little error reduction and lack credibility because of unstable coefficients).

It should be noted that, with a single national equation, one can either use the single overall national error estimate or use the errors in the national equation as determined individually for specific regions and seasons. Tables 3.2, 3.5, and 3.8 summarized the errors in the single

TABLE 3.11 PERFORMANCE OF VARIOUS MODELS IN PREDICTING IP.

Table 3.11a 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 3.11b 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

national equation (for 1, 2, or 3 variables, respectively) specific to various regions and seasons. With respect to this issue, we do not have a strong recommendation.

3.2 ESTIMATION OF FP

This section develops and evaluates equations for predicting FP from Hi-Vol data. The organization of the discussion is entirely analagous to that of the previous section. The first four subsections sequentially add the independent variables (TSP, $SO_4^=$, Pb, and NO_3^-); the last subsection summarizes our findings and recommendations.

3.2.1 FP Versus TSP

In the simplest case, we assume that Hi-Vol data are available only for TSP. The equation for predicting FP is derived from a zero-intercept, one-parameter regression:

$$FP = b \cdot TSP \tag{17}$$

The coefficient "b" represents the fraction of TSP that is FP (i.e. the FP/TSP ratio).

Table 3.12 lists the values and standard errors for "b" determined for the national, site-type, regional, and regional/seasonal data bases. The national coefficient is 0.30, but there are many significant variations in the coefficient by site-type, region, and region/season.

The site-type variations in "b" are similar to the site-type variations for the IP predictive equation (see previous Table 3.1). As discussed previously for IP in Section 3.1.1, the high FP/TSP ratio for nonurban locations could represent relatively greater secondary aerosol levels compared to total TSP levels at nonurban sites or lesser artifact formation on Hi-Vol filters at nonurban sites.

Some of the regional and regional/seasonal patterns in Table 3.12 are similar to the patterns we observed for the IP/TSP ratio (see Table 3.1). This supports our previous hypothesis that many of the regional and regional/seasonal patterns in the IP/TSP ratio are related to patterns in the concentrations of fine aerosols. The specific similarities between the FP/TSP ratio and the IP/TSP ratio are as follows: both ratios are especially high

TABLE 3.12 COEFFICIENTS FOR THE EQUATION FP = b·TSP.

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)	REGIONAL/SEASONAL: Summer		REGIONAL/SEASONAL: Winter
NATIONAL:	0.30 ± .005			
SITE-TYPE:				
Metropolitan	0.30 ± .006			
Suburban	0.27* ± .012			
Nonurban	0.39* ± .022			
REGIONAL:				
Cal.-San Francisco Area	0.26 ± .026	0.19 ± .018	**	0.31 ± .044
Cal.-Central Valley	0.23 ± .026	0.16 ± .027	**	0.29 ± .037
Cal.-Los Angeles Area	0.28 ± .015	0.29 ± .022		0.27 ± .021
Pacific Northwest	0.21* ± .027	0.14 ± .025	**	0.38 ± .038
Arid Southwest	0.21* ± .021	0.21 ± .029		0.22 ± .031
North Central	0.25* ± .014	0.22 ± .013	**	0.33 ± .027
Northeast	0.35* ± .007	0.34 ± .010		0.36 ± .010
Southeast	0.28 ± .012	0.27 ± .014	**	0.35 ± .026

*Differs from national aggregate value at 95% confidence level.

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

in the Northeast; both tend to be low in the Central Valley, Pacific Northwest, and North Central area; and both exhibit a strong seasonal pattern -- peaking in the winter -- in the San Francisco area, Central Valley, and Pacific Northwest. There are however, also some differences in the regional and regional/seasonal patterns of the FP/TSP and IP/TSP ratios. The FP/TSP ratio does not stand out as being as especially high in the Los Angeles area as the IP/TSP ratio, or as especially low in the San Francisco area. The FP/TSP ratio, unlike the IP/TSP ratio, is remarkably low in the Arid Southwest. Furthermore, the FP/TSP ratio exhibits less of a seasonal pattern than the IP/TSP ratio in Los Angeles and more of a seasonal pattern in the Southeast. The causes of these differences would be interesting to speculate about but very difficult to verify. Because of the complex and sometimes puzzling regional and regional/seasonal patterns in the FP/TSP and IP/TSP ratios, it is fair to say that no simple explanation readily accounts for all the patterns. Even the general explanation we offered previously -- that high FP/TSP and IP/TSP ratios tend to be observed in areas and times of high concentrations of sulfates and other fine particles -- falls short in certain instances.

As was the case with IP, we have formulated four types of predictive models: (1) the simple national equation, (2) a set of three site-type equations, (3) a set of eight regional equations, and (4) a set of sixteen regional/seasonal equations. Table 3.13 compares the percentage errors that result when the specific equations and the national equation are applied to each site-type, region, and season. As was generally the case with IP, the national equation performs almost as well as the specific equations for the three site-types. For certain regions and regions/seasons, however, the specific equations produce substantially less error than the national equation. This is particularly true in the San Francisco area during the summer, in the Central Valley (especially during the summer), in the Pacific Northwest (especially in the summer), in the Arid Southwest, and in the North Central area during the summer.

Table 3.14 summarizes the overall performance of the national equation and the disaggregated models, evaluated on both a daily and annual mean

TABLE 3.13 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATIONS OF THE FORM FP = b·TSP.

PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET
(PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)

		REGIONAL/SEASONAL	
		Summer	Winter
SITE-TYPE:			
Metropolitan	54.2 (54.2)		
Suburban	61.9 (63.2)		
Nonurban	52.3 (57.7)		
REGIONAL:			
Cal.-San Francisco Area	84.1 (86.4)	54.3 (86.5)	82.9 (83.0)
Cal.-Central Valley	73.9 (82.1)	63.3 (107.8)	67.7 (67.8)
Cal.-Los Angeles Area	54.8 (55.3)	44.9 (45.0)	61.2 (62.2)
Pacific Northwest	72.1 (80.9)	70.8 (127.6)	47.7 (52.5)
Arid Southwest	74.8 (87.7)	73.2 (89.6)	77.7 (87.7)
North Central	53.0 (57.5)	45.6 (62.1)	52.4 (53.0)
Northeast	47.3 (49.9)	46.4 (48.1)	47.6 (50.8)
Southeast	43.2 (43.6)	41.1 (42.4)	43.2 (45.8)

basis. The site-type model offers little if any improvement over the national equation. The regional and regional/seasonal models both offer slight improvements over the national equation.

TABLE 3.14 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM FP = b·TSP.

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.539	56.3%	0.328	29.9%
3 SITE-TYPE EQUATIONS	0.548	55.9%	0.347	29.7%
8 REGIONAL EQUATIONS	0.607	53.1%	0.640	24.3%
16 REGIONAL/ SEASONAL EQUATIONS	0.638	51.4%	Not Applicable	

In comparing Tables 3.13 and 3.14 to the corresponding results for IP (see Tables 3.2 and 3.3), we find that the errors are nearly twice as great in predicting FP from TSP than in predicting IP from TSP. This makes physical sense because IP (particle mass in the size range less than or equal to 15 μm) more closely represents TSP (particle mass in the size range less than or equal to 50 μm) than does FP (particle mass in the size range less than or equal to 2.5 μm). Most importantly, IP and TSP each contain some contribution from both the fine (.1 - 1 μm) and coarse (3 - 50 μm) particle mass modes, while FP only contains the fine mass mode.

3.2.2 FP Versus TSP and $SO_4^=$

The second step in deriving predictive equations for FP uses Hi-Vol data for both TSP and $SO_4^=$. As discussed in Section 2.2, physical considerations lead us to conduct a zero-intercept regression of the form:

$$FP - 1.1\ SO_4^{=} = b(TSP - 1.4\ SO_4^{=}) \qquad (18)$$

The coefficient "b" now represents the fraction of non-sulfate TSP that is fine. The final predictive equation for FP is then

$$FP = 1.1\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=}) \qquad (19)$$

Table 3.15 lists the values of "b" for Equation (19) on a national, site-type, regional, and regional/seasonal basis. The site-type and regional patterns in Table 3.15 are less pronounced than the corresponding patterns in the FP/TSP ratio (see previous Table 3.12). The fact that the patterns become less pronounced when we explicitly discount for sulfates suggests that sulfate concentrations are a significant factor contributing to the original site-type and regional patterns for the FP/TSP ratio. The seasonal variations in Table 3.15, however, are just as strong as the seasonal variations in Table 3.12. This indicates that sulfate concentrations are not a major factor accounting for seasonal variations in the FP/TSP ratio.

Table 3.16 compares the percentage errors in applying the national equation to the percentage errors in applying the equations specific to each site-type, region, and region/season. Table 3.16 shows that the national equation yields almost as low an error as the equations specific to each site-type and region. However, for certain regions/seasons (e.g. San Francisco in the summer, Central Valley in the summer, Pacific Northwest in the summer, and North Central in the summer), the national equation produces a substantially greater error than the equations specific to those regions/seasons.

Table 3.17 summarizes the overall performance of the national equation and disaggregated models on both a daily basis and an annual mean basis. As has generally been the case, the site-type model yields little if any improvement over the national equation; the regional model offers a slight improvement; and the regional/seasonal model offers further slight improvements.

TABLE 3.15 COEFFICIENTS FOR THE EQUATION $FP = 1.1\ SO_4^{=} + b\ (TSP - 1.4\ SO_4^{=})$.

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)			
NATIONAL:	0.21 ± .005			
SITE-TYPE:				
Metropolitan	0.21 ± .006			
Suburban	0.18* ± .011			
Nonurban	0.24 ± .023			
REGIONAL:		REGIONAL/SEASONAL:		
		Summer		Winter
Cal.-San Francisco Area	0.20 ± .030	0.11 ± .024	**	0.27 ± .047
Cal.-Central Valley	0.18 ± .027	0.11 ± .029	**	0.25 ± .038
Cal.-Los Angeles Area	0.23 ± .016	0.24 ± .022		0.23 ± .021
Pacific Northwest	0.15* ± .025	0.09 ± .023	**	0.32 ± .038
Arid Southwest	0.17 ± .022	0.16 ± .032		0.17 ± .030
North Central	0.14* ± .012	0.12 ± .012	**	0.20 ± .027
Northeast	0.23 ± .007	0.18 ± .009	**	0.26 ± .010
Southeast	0.18* ± .010	0.17 ± .011	**	0.26 ± .022

*Differs from national aggregate value at 95% confidence level.

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

TABLE 3.16 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATION OF THE FORM $FP = 1.1\ SO_4^= + b(TSP - 1.4\ SO_4^=)$.

PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET
(PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)

		REGIONAL/SEASONAL	
		Summer	Winter
SITE-TYPE:			
Metropolitan	45.1 (45.1)		
Suburban	48.9 (50.0)		
Nonurban	39.7 (40.0)		
REGIONAL:			
Cal.-San Francisco Area	87.5 (87.6)	64.4 (84.5)	82.3 (85.2)
Cal.-Central Valley	72.2 (73.1)	63.4 (85.7)	64.7 (66.9)
Cal.-Los Angeles Area	50.6 (51.4)	40.8 (41.9)	57.1 (57.7)
Pacific Northwest	62.0 (66.1)	60.9 (97.2)	42.0 (48.7)
Arid Southwest	71.4 (74.4)	74.7 (78.5)	69.8 (72.2)
North Central	40.8 (47.7)	35.4 (52.7)	42.3 (42.4)
Northeast	37.2 (37.6)	31.3 (31.9)	38.7 (40.7)
Southeast	29.0 (30.3)	25.7 (29.1)	30.7 (33.0)

TABLE 3.17 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM FP = 1.1 $SO_4^=$ + b(TSP - 1.4 $SO_4^=$).

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.724	46.1%	0.781	19.7%
3 SITE-TYPE EQUATIONS	0.727	45.9%	0.772	20.1%
8 REGIONAL EQUATIONS	0.740	44.9%	0.818	18.2%
16 REGIONAL/ SEASONAL EQUATIONS	0.768	42.8%	Not Applicable	

It is worthwhile to note that the addition of the $SO_4^=$ variable in this step has produced a substantial reduction in predictive errors for FP. For example, comparing Table 3.17 and 3.14, we see that adding the sulfate variable reduces the FP predictive error for the national equation from 56% to 46% on a daily basis, and from 30% to 20% on an annual mean basis. This reduction in FP predictive error is much greater than the error reduction we observed in adding the $SO_4^=$ variable to the IP prediction scheme. This makes sense because we earlier observed that TSP alone is a good predictor of IP but a poor predictor of FP. Adding the sulfate variable means that we are adding explicit information about an important FP component (i.e. fine sulfate aerosols); it is not surprising that the prediction errors for FP are thereby reduced considerably.

3.2.3 FP Versus TSP, $SO_4^=$, and Pb

The third step uses data for TSP, $SO_4^=$, and Pb. With the $SO_4^=$ and Pb coefficients constrained by the physical/chemical considerations discussed in Section 2.2, the regression is

$$FP - 1.1\ SO_4^= - 11\ Pb = b(TSP - 1.4\ SO_4^= - 15\ Pb) \tag{20}$$

The final prediction equation is

$$FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb) \quad (21)$$

with the coefficient "b" representing the fraction of non-sulfate, non-vehicular TSP that is fine.

Table 3.18 presents the national, site-type, regional, and regional/seasonal values for the coefficient "b" in Equation (21). Almost none of the site-type and regional variations in the coefficient are statistically significant. The fact that the site-type and regional patterns are suppressed when we explicitly include sulfates and vehicular aerosols indicates that sulfates and vehicular aerosols are major causes of the original site-type and regional patterns that we observed in the FP/TSP ratio (i.e. in previous Table 3.12). The seasonal patterns, however, remain pronounced, indicating that sulfates and vehicular aerosols do not account for the seasonal variations in the FP/TSP ratio.

Table 3.19 lists the percentage errors that result when the specific equations and the national equation are applied to each site-type, region, and region/season. The national equation generally performs as well as the specific equations for the various site-types and regions. For a few of the regions/seasons (San Francisco in the summer, Central Valley in the summer, and Pacific Northwest in the summer), the national equation performs moderately worse than the specific equations.

Table 3.20 summarizes the overall performance of the national equation, the 3-equation site-type model, the 8-equation regional model, and the 16-equation regional/seasonal model. As before, the site-type model yields little if any improvement over the national equation. Also, in this case, the regional and regional seasonal models offer only a very slight improvement over the national model.

As was the case with the addition of the $SO_4^{=}$ variable, the addition of the Pb variable produces a substantial decrease in predictive errors for FP, much more of a decrease than was produced for IP. Again, this reflects the fact that TSP alone is a very poor predictor of FP; much better predictions can be obtained by considering specific FP components (i.e. sulfates and vehicular aerosols).

TABLE 3.18 COEFFICIENTS FOR THE EQUATION $FP = 1.1\ SO_4^= + 11\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$.

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)			
NATIONAL:	0.14 ± .005			
SITE-TYPE:				
Metropolitan	0.14 ± .005			
Suburban	0.12 ± .012			
Nonurban	0.20* ± .021			
REGIONAL:		REGIONAL/SEASONAL:		
		Summer		Winter
Cal.-San Francisco Area	0.11 ± .027	0.05 ± .023	**	0.16 ± .047
Cal.-Central Valley	0.10 ± .029	0.04 ± .034	**	0.16 ± .042
Cal.-Los Angeles Area	0.16 ± .016	0.20 ± .023	**	0.13 ± .021
Pacific Northwest	0.10 ± .021	0.07 ± .022	**	0.22 ± .038
Arid Southwest	0.12 ± .020	0.13 ± .031		0.11 ± .026
North Central	0.11 ± .013	0.10 ± .014	**	0.17 ± .025
Northeast	0.15 ± .007	0.14 ± .009		0.16 ± .010
Southeast	0.13 ± .009	0.13 ± .010	**	0.19 ± .023

*Differs from national aggregate value at 95% confidence level.

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

TABLE 3.19 COMPARISON OF ERRORS FOR THE NATIONAL EQUATION TO ERRORS IN EQUATIONS SPECIFIC TO SITE-TYPE, REGION, AND REGION/SEASON. PREDICTIVE EQUATION OF THE FORM $FP = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$.

	PERCENT ERROR FOR EQUATION SPECIFIC TO DATA SUBSET (PERCENT ERROR FOR NATIONAL EQUATION APPLIED TO DATA SUBSET)		
SITE-TYPE:			
Metropolitan	37.8 (37.8)		
Suburban	46.3 (46.8)		
Nonurban	35.4 (37.2)		
REGIONAL:		REGIONAL/SEASONAL	
		Summer	Winter
Cal.-San Francisco Area	70.7 (71.6)	56.3 (70.3)	69.0 (69.2)
Cal.-Central Valley	68.0 (69.9)	67.6 (85.3)	61.6 (62.1)
Cal.-Los Angeles Area	45.9 (46.6)	39.5 (44.3)	48.4 (48.4)
Pacific Northwest	47.7 (49.8)	56.8 (71.7)	34.6 (37.8)
Arid Southwest	62.3 (62.9)	70.1 (70.2)	55.4 (56.9)
North Central	40.8 (41.7)	41.2 (44.9)	38.1 (38.7)
Northeast	31.2 (31.3)	28.0 (28.0)	32.8 (33.1)
Southeast	24.6 (24.7)	22.1 (22.5)	28.7 (30.2)

TABLE 3.20 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM $FP = 1.1\ SO_4^= + 11\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$.

	PERFORMANCE VERSUS ALL 930 DAILY DATA POINTS		PERFORMANCE VERSUS THE 21 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.804	39.7%	0.852	16.6%
3 SITE-TYPE EQUATIONS	0.806	39.6%	0.847	16.8%
8 REGIONAL EQUATIONS	0.810	39.2%	0.856	16.4%
16 REGIONAL/ SEASONAL EQUATIONS	0.821	38.2%	Not Applicable	

3.2.4 FP Versus TSP, $SO_4^=$, Pb, and NO_3^-

As a final step in deriving predictive equations for FP, we assume that data are available for all four Hi-Vol parameters. A multiple, zero-intercept regression is run of the form:

$$FP = 1.1\ SO_4^= - 11\ Pb = b_1(TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2 NO_3^- \tag{22}$$

The final predictive model is

$$FP = 1.1\ SO_4^= + 11\ Pb + b_1\ (TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2 NO_3^- \tag{23}$$

As was the case with IP, we will not present tables of the coefficients "b_1" and "b_2" because the model represented by Equation (23) does not perform well. One major problem is the instability of the "b_2" coefficient. Among the regions and seasons, "b_2" varies erratically from -1.5 to +4.5. Part of this instability stems from the colinearity between NO_3^- and "TSP - 1.4 $SO_4^=$ - 15 Pb" (see earlier discussion in Section 3.1.4). Because of the colinearity problem and because of the relative unimportance of the "TSP - 1.4 $SO_4^=$ - 15 Pb" term in predicting FP, the "b_1" coefficient also becomes unstable, varying from -0.8 to +0.14 among the regions and seasons. A second drawback involves the statistical significance of "b_2". Although the NO_3^- term is more important in predicting FP than in predicting IP, the term is still statistically insignificant for several regions and seasons. Related to the lack of a high statistical significance level, we find that Equation (23) does not reduce the error in FP predictions very much compared to Equation (21) (the equation including only TSP, $SO_4^=$, and Pb). A third major reason for rejecting this model is that Equation (23) cannot be readily interpreted on physical grounds.

3.2.5 Summary of Predictive Relationships for FP

Table 3.21 presents matrices summarizing the overall performance of the various FP predictive models described in the previous sections. The simplest model (the single national equation, FP = 0.30 TSP) has an error of 56.3% in predicting daily values of FP and 29.9% in predicting annual mean values of FP. Table 3.21 shows that the predictive errors decrease substantially when $SO_4^=$ or $SO_4^=$ and Pb are added to the equations. For example, the single national equation FP = 1.1 $SO_4^=$ + 11 Pb + 0.14(TSP - 1.4 $SO_4^=$ - 15 Pb) has an error of 39.7% for daily values of FP and 16.6% for annual mean values of FP. We highly recommend that $SO_4^=$ or $SO_4^=$ and Pb be included in the predictions of FP whenever $SO_4^=$ and/or Pb data are available. Because the addition of the NO_3^- variable produces little error reduction, and because the NO_3^- variable lacks credibility due to the unstable regression coefficients, we recommend not including NO_3^- data in predictions of FP.

Table 3.21 shows that the 3-equation site-type model produces negligible reduction in prediction errors. We recommned against using the site-type model. The regional and regional/seasonal models each produce a slight to moderate reduction in predictive errors. The use of the regional/seasonal model is recommended in certain cases, <u>especially in California</u>, where we found significant regional/seasonal variations in the coefficients.

If the reader employs a single national equation (rather than the regional/seasonal equations), he can either use the single overall national error estimate or use the errors in the national equation as determined

TABLE 3.21 PERFORMANCE OF VARIOUS MODELS IN PREDICTING FP.

Table 3.21a 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 3.21b 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
$FP = 1.1 + b(TSP - 1.4\ SO_4^=)$	[illegible]9.7%	20.1%	18.2%	Not Applicable
$FP = 1.1\ SO_4^= + 11\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb)$	[illegible]6.6%	16.8%	16.4%	Not Applicable
$FP = 1.1\ SO_4^= + 11\ Pb + b(TSP - 1.4\ SO_4^= - 15\ Pb - 1.3\ NO_3^-) + b_2 NO_3^-$	[illegible]5.8%	16.5%	18.1%	Not Applicable

individually for specific regions and seasons. Tables 3.13, 3.16, and 3.19 summarized the errors in the single national equation (for 1, 2, and 3 variables respectively) specific to the various regions and seasons. We do not have a strong recommendation with respect to this choice of error measures.

3.3 ESTIMATION OF SSIP

This section develops and evaluates equations for translating routine Hi-Vol data into estimates of SSIP (IP measurements taken using Hi-Vols with size-selective inlets). Because of certain limitations in the data base for simultaneous SSIP and Hi-Vol recordings, and because TSP alone turns out to be an excellent predictor of SSIP, this section has a different organization than the previous two sections. Subsection 3.3.1 uses the rather limited data base with simultaneous recordings of SSIP and all four Hi-Vol parameters to examine the relationship between SSIP and hybrid equations involving TSP, $SO_4^=$, Pb, and NO_3^-. Subsection 3.3.2 uses the very extensive data base involving simultaneous SSIP and TSP readings to examine the relationship between SSIP and TSP in greater detail.

3.3.1 SSIP Versus TSP, $SO_4^=$, Pb and NO_3^-

In order to examine the relationship between SSIP and all four Hi-Vol parameters (TSP, $SO_4^=$, Pb, and NO_3^-), we assembled a data base involving simultaneous readings for all five of these variables. This data base, summarized in Appendix B, differs from our IP/FP/Hi-Vol data base in the sense that we lost many sites and gained a few sites. In total, the resulting data base involves 64 sites with 741 data points. This data base has been subjected to the same data quality screen as our IP/FP/Hi-Vol data base (see Section 2.1.2).

Because of the more limited nature of the SSIP/Hi-Vol data base, we only had three sites that met our criteria for determining annual means. Accordingly, in this subsection, we are not able to evaluate the prediction errors for annual means. Rather, this subsection will deal only with prediction errors for daily values.

As before, we use a hybrid approach (physico-chemical and statistical) in deriving predictive equations for SSIP. Also, as before, the four Hi-Vol variables are added in a stepwise progression. The four final predictive equations are as follows:*

$$SSIP = b.TSP \tag{24}$$

$$SSIP = 1.4\,SO_4^{=} + b(TSP - 1.4\ SO_4^{=}) \tag{25}$$

$$SSIP = 1.4\,SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb) \tag{26}$$

$$SSIP = 1.4\,SO_4^{=} + 15\ Pb + b_1(TSP - 1.4\ SO_4^{=} - 15\ Pb - 1.3\ NO_3^{-}) + b_2NO_3^{-} \tag{27}$$

The coefficients "b" in Equations (24), (25), and (26) are determined from zero-intercept, one-parameter regressions. The coefficients "b_1" and "b_2" in Equation (27) are determined from a zero-intercept, multiple regression.

We find that these hybrid equations perform just as well as best-fit statistical equations in predicting SSIP from the four Hi-Vol parameters. In fact, the degree of correlation in predicting SSIP using any of the hybrid models is within .001 of the degree of correlation achieved by a multiple regression involving best-fit coefficients for all the Hi-Vol parameters.

Table 3.22 presents the final hybrid equations determined from the 741 nationwide data points. Table 3.22 also presents the degree of correlation and percentage error in predicting daily values of SSIP using each of the equations. The simplest formula, SSIP = .74 TSP, performs very well, with an overall correlation of .967 and an error of 17.0% in predicting SSIP. This performance is much better than any of our models for predicting IP or FP (see previous Tables 3.11 and 3.21). The error in predicting SSIP can be reduced even further by adding the other Hi-Vol parameters ($SO_4^{=}$, Pb, and

*For SSIP, we assume a 1.4 coefficient for sulfates to account for the ammonium ion. No adjustment is needed relative to the Hi-Vol data because SSIP involves the same artifact sulfate as the Hi-Vol data and because virtually all the sulfate is less than 15 μm in size. The lead coefficient of "15" is also the same for SSIP and the Hi-Vol data because in both cases we are dealing with just the ratio of suspendable vehicular emissions to suspendible vehicular Pb emissions.

NO_3^-). However, because the simplest formula works so well, and because data are available for evaluating this formula in greater detail (see next subsection), we recommend using the simplest relationship, SSIP = .74 TSP, in predicting SSIP from routine Hi-Vol data.

TABLE 3.22 HYBRID EQUATIONS FOR ESTIMATING SSIP FROM ROUTINE HI-VOL DATA.

EQUATION	PERFORMANCE IN PREDICTING DAILY SSIP: Correlation	Percentage Error
SSIP = 0.74 TSP. ($\pm$ .004)*	0.967	17.0%
SSIP = 1.4 $SO_4^=$ + 0.69(TSP - 1.4 $SO_4^=$). ($\pm$ .004)*	0.971	15.9%
SSIP = 1.4 $SO_4^=$ + 15 Pb + 0.66(TSP - 1.4 $SO_4^=$ - 15 Pb). ($\pm$ .005)*	0.971	15.9%
SSIP = 1.4 $SO_4^=$ + 15 Pb + 0.57(TSP - 1.4 $SO_4^=$ - 15 Pb - 1.3 NO_3^-) + ($\pm$ .006)* 1.46 NO_3^-. ($\pm$ .039)*	0.980	13.2%

*Standard errors of regression coefficients.

3.3.2 SSIP Versus TSP

A very large data base is available for simultaneous measurements of SSIP and TSP. As summarized in Appendix C, we were able to assemble 2169 such data points at 97 sites nationwide. This data base allows us to examine the relationship between SSIP and TSP in detail.

Table 3.23 summarizes the coefficients for the zero-intercept, one-parameter, regression equation: SSIP = b·TSP. The coefficients are presented on a national, site-type, regional, and regional/seasonal basis. As we

TABLE 3.23 COEFFICIENTS FOR THE EQUATION SSIP = b·TSP.

	COEFFICIENT (± STANDARD ERROR OF COEFFICIENT)	REGIONAL/SEASONAL: Summer		REGIONAL/SEASONAL: Winter
NATIONAL:	0.74 ± .002			
SITE-TYPE:				
Metropolitan	0.74 ± .003			
Suburban	0.74 ± .004			
Nonurban	0.75 ± .009			
REGIONAL:				
Cal.-San Francisco Area	0.73 ± .016	0.68 ± .016	**	0.79 ± .024
Cal.-Central Valley	0.74 ± .008	0.72 ± .016		0.75 ± .009
Cal.-Los Angeles Area	0.79* ± .006	0.78 ± .006		0.80 ± .013
Cal.-Other	0.65* ± .011	0.65 ± .014		0.63 ± .020
Pacific Northwest	0.62* ± .022	0.56 ± .024	**	0.79 ± .026
Arid Southwest	0.74 ± .009	0.71 ± .009		0.76 ± .014
North Central	0.70* ± .007	0.68 ± .009	**	0.72 ± .011
Northeast	0.72* ± .004	0.74 ± .005	**	0.69 ± .006
Southeast	0.72* ± .007	0.73 ± .009		0.70 ± .013

*Differs from national aggregate value at 95% confidence level.

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

found in the previous subsection, the national coefficient is 0.74. There are essentially no differences from this national value among the various site types. There are, however, some noticeable regional and regional/seasonal variations in the coefficient. Although these regional and regional/seasonal variations in the coefficient are statistically significant, they are not of great practical importance. As shown in Table 3.24, we find that a predictive scheme based on the single national equation performs nearly as well as disaggregated models based on nine regional equations or eighteen regional/seasonal equations. Thus, for most applications, we would recommend just using the single national equation for predicting SSIP from TSP.

TABLE 3.24 PERFORMANCE OF THE NATIONAL EQUATION AND DISAGGREGATED MODELS, EVALUATED OVER THE ENTIRE DATA BASE. PREDICTIVE EQUATIONS OF THE FORM SSIP = b·TSP.

	PERFORMANCE VERSUS ALL 2169 DAILY DATA POINTS		PERFORMANCE VERSUS THE 9 ANNUAL MEAN DATA POINTS	
	Overall Correlation	Percent Error	Overall Correlation	Percent Error
NATIONAL EQUATION	0.960	18.4%	0.931	11.4%
3 SITE-TYPE EQUATIONS	0.960	18.4%	0.931	11.4%
9 REGIONAL EQUATIONS	0.964	17.3%	0.965	8.2%
18 REGIONAL EQUATIONS	0.966	16.9%	Not Applicable	

Table 3.24 shows that the error in predicting daily data points with the national equation is 18.4%. This differs slightly from the error listed for the same equation in the previous subsection (see Table 3.22) because we are now considering a different data set (2169 data points rather than 741 data points). The error in predicting annual mean SSIP with the national equation is 11.4%. The reader should be cautioned that this error for annual mean predictions is uncertain because our data base provided only

nine valid annual means.* The error in predicting annual means reevaluated when more annual mean data points become available.

*Very few sites met our criteria for an "annual data set" because S toring was started rather late in the EPA IP Network.

4. APPLICATION TO CALIFORNIA HI-VOL DATA

In this chapter, the empirical formulae developed in the previous chapters are applied to the California ARB Hi-Vol data base. Specifically, five years of California Hi-Vol data for TSP, $SO_4^=$, and/or Pb are translated into estimates of IP and FP. A descriptive analysis is then conducted, comparing and contrasting the geographical/seasonal patterns of TSP, IP, FP, and visibility.

4.1 DATA BASE

The data base for this application consists of all of the Hi-Vol measurements in the ARB computerized files for the period 1976 to 1980. Using five years of measurements provides a robust yet manageable data set. The period 1976-1980 is of greatest interest because it is most recent; also, this period contains a much larger quantity of California Hi-Vol data than any other five year period.

Table 4.1 summarizes the availability of California Hi-Vol data for the period 1976 to 1980. Data are available at 287 locations. For each location, Table 4.1 lists the number of data points for TSP, for simultaneous TSP and $SO_4^=$, and for simultaneous TSP, $SO_4^=$, and Pb.

As described in the previous chapter, we have developed equations for estimating IP and FP from TSP data alone, from both TSP and $SO_4^=$ data, or from the triad of TSP, $SO_4^=$, and Pb data. For each application site, we need to select a given level in this hierarchy of equations. This selection process involves a tradeoff. Predictive accuracy increases (especially for FP) as more Hi-Vol variables are included, but the number of available data points decreases as more Hi-Vol variables are required. Our final selections are indicated by the asterisk notations in the central column of Table 4.1. Generally, we have included the $SO_4^=$ and Pb variables wherever possible, except when their inclusion severely reduces the number of data points.

Table 4.1 also indicates the start and end dates for the selected data sets and the number of data points in summer and winter. Because of insufficient data and/or because of strong seasonal biases, certain sites are eliminated from consideration. These sites are indicated by the "not used"

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 TO 1980.

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	TSP and $SO_4^=$	TSP, $SO_4^=$, and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	Winter	DATES OF TSP OR SELECTED DATA SET Start	End
NORTH COAST AIR BASIN								
Arcata Fire Station	296	0	0	*	149	147	760101	801229
Capella	125	0	0	*	57	68	760101	780219
Cloverdale	201	0	0	*	103	98	760101	790626
Crescent City	77	0	0	*	36	41	760101	781228
Eureka H.D. 6 & I	303	0	0	*	153	150	760101	801229
Eureka Hwy. Dept.	293	0	0	*	151	142	760101	801229
Fort Bragg Central	157	156	155	***	89	66	780228	801217
Fort Bragg So. Main	265	229	229	***	123	106	760512	801217
Healdsburg	199	0	0	*	101	98	760105	790626
Ukian Firehouse	297	0	0	*	152	145	760101	801229
Willets	294	0	0	*	149	145	760101	801228
NORTHEAST PLATEAU AIR BASIN								
Alturas	123	112	112	***	58	54	760407	780611
Burney	290	0	0	*	149	141	760101	801229
Cedarville	175	0	0	*	93	82	760106	801229
Fort Jones	204	0	0	*	95	109	761003	801229
McCloud	120	0	0	*	54	66	760101	801229
Mount Shasta	45	0	0	*	19	26	760202	771208
Tulelake Fairground	126	0	0	*	47	79	760102	790122
Weed	222	10	10	*	105	117	760101	801229
Yreka	262	86	86	***	50	36	760401	780102
Yreka Courthouse	25	0	0	not used	18	7	760417	790427
LAKE COUNTY AIR BASIN								
Hobergs Hwy. 175	52	0	0	*	25	27	760101	770531
Kelseyville Dorn Road	285	0	0	*	144	141	760101	801229
Lakeport Lakeport Blvd.	293	275	274	***	150	124	760406	801129
Middletown	10	5	5	not used	5	0	780701	780725
Upperlake	270	0	0	*	140	130	760309	801229
SACRAMENTO VALLEY AIR BASIN								
Anderson Center City	276	0	0	*	144	132	760113	801229
Buckey Elementary School	163	0	0	*	86	77	780219	801229
Chico	132	2	2	*	58	74	760101	780327
Chico Manzanita	214	1	1	*	120	94	770224	801006
Chico State	201	0	0	*	101	100	770805	801223
Citrus Heights Sunrise Boulevard	48	48	48	***	29	19	800203	801123
Corning	170	0	0	*	88	82	760101	781128
Davis 5th Street	71	0	0	*	30	41	760101	770224
Dunnigan Main Street	300	0	0	*	153	147	760101	801229
Gridley Graylodge	242	0	0	*	116	126	760101	800227
Live Oak	231	0	0	*	118	113	760120	801229
Los Molinos	287	0	0	*	150	137	760107	801229
Marysville	136	0	0	*	69	67	760101	780601
Mountain Gate	60	59	59	***	29	30	780102	781228
Nord	88	0	0	*	45	43	760101	770730
Oroville Bird Street	301	0	0	*	151	150	760102	801229
Oroville 6 WNW	60	32	0	**	18	14	780207	781216
Pleasant Grove	104	0	0	*	53	51	760101	771227
Rancho Cordova	230	0	0	*	119	111	761003	800930
Red Bluff	283	0	0	*	147	136	760101	801229
Red Bluff Ag. Comm. Office	286	0	0	*	147	139	760101	801229
Red Bluff Lincoln	163	141	141	***	70	71	760406	790226
Redding H.D. Roof	296	0	0	*	150	146	760101	801229
Rio Vista	273	0	0	*	143	130	760101	801229

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 to 1980 (Continued).

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	TSP and $SO_4^=$	TSP, $SO_4^=$, and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	Winter	DATES OF TSP OR SELECTED DATA SET Start	End
Rio Vista Army Facility	17	0	0	not used	3	14	770107	770425
Sacramento H.D. Stockton Blvd.	168	0	0	*	93	75	780102	801024
Sacramento 1025 P St.	276	257	255	***	139	116	760403	801123
Sacramento Branch Center Road	275	0	0	*	148	127	760101	800930
Sacramento Del Paso Manor	10	0	0	not used	10	0	800807	800930
Sherman Island	61	58	0	**	30	28	780102	781228
Smartville	276	0	0	*	137	139	760102	801129
Sutter City	233	0	0	*	122	111	760101	801229
Vacaville	111	0	0	*	61	50	760101	771028
Vacaville Merchant	204	0	0	*	102	102	770805	801229
Weaverville Hospital	88	0	0	*	45	43	760101	781228
West Sacramento 15 St.	300	0	0	*	154	146	760101	801229
Wheatland	269	0	0	*	133	136	760101	801129
Wheatland 4 W	61	48	0	*	31	30	780102	781228
Williams	37	0	0	*	17	20	760131	780824
Willows	79	0	0	*	30	49	770904	790227
Willows 5 M West	60	0	0	*	30	30	771022	770220
Willows 8 W	60	30	0	**	15	18	780114	781228
Woodland W. Main St.	296	0	0	*	152	144	760101	801229
Yuba City	122	0	0	*	62	60	781104	801229
MOUNTAIN COUNTIES AIR BASIN								
Auburn Dewitt Center	215	0	0	*	104	111	760101	801229
Camino	103	0	0	*	58	45	760101	780520
Columbia	25	0	0	not used	4	21	760909	770302
Georgetown	96	0	0	*	57	39	760101	780520
Lincoln	112	0	0	*	48	64	760107	780207
Placerville Fairlane Dr.	84	0	0	not used	56	28	760406	780526
Placerville Airport	18	18	18	not used	19	0	780601	780929
Portola	77	0	0	*	38	39	760822	771127
Quincy Hospital	91	0	0	not used	65	26	760325	780625
Rocklin Sierra College	125	0	0	*	57	68	771004	801229
Sierra City	102	0	0	*	48	54	771004	790930
Sonora	89	36	36	*	53	36	760406	801018
Sonora Forrest Road	51	0	0	*	28	23	760101	761113
Sonora 155 S. Washington	3	0	0	not used	0	3	800310	800328
Sonora 105 S. Washington	34	0	0	not used	28	6	800403	801105
Tuolomne City	60	0	0	*	29	31	760101	770206
Weimar	13	13	13	not used	13	0	780625	780923
Yosemite Village	28	0	0	not used	14	14	800820	801217
LAKE TAHOE AIR BASIN								
N. Lake Tahoe USCG Sta.	104	0	0	*	52	52	771203	801129
S. Lake Tahoe Police Department	96	11	11	*	56	40	760101	780526
S. Lake Tahoe Airport	94	4	4	*	59	35	760101	780526
Tahoe City	89	0	0	*	40	49	760101	771127
SAN FRANCISCO BAY AREA AIR BASIN								
Berkeley	155	0	0	*	83	72	780108	800924
Bethel Island	59	51	0	**	27	24	780108	781228
Burlingame Burlingame Avenue	229	98	98	***	49	49	760101	791222
Concord Treat Blvd.	303	160	101	***	58	43	760101	800625
Concord 2976 Treat Blvd.	8	8	8	not used	5	3	800905	801017
Fremont Chapel Way	301	116	116	***	60	56	760101	791222
Gilroy Monterey St.	253	112	112	***	59	53	760101	791222

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 TO 1980 (Continued).

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	 TSP and $SO_4^=$	 TSP, $SO_4^=$, and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	 Winter	DATES OF TSP OR SELECTED DATA SET Start	 End
Livermore Railroad	293	151	121	***	71	50	760101	801128
Millrae Sewage Plant	15	15	15	not used	5	10	780101	781227
Napa Jefferson St.	300	264	263	***	142	121	760101	801228
Oakland	14	0	0	not used	0	14	780105	780327
Oakland Jackson	145	0	0	*	76	69	770301	801005
Pittsburg	308	263	195	***	110	85	760101	801128
Redwood City	280	252	251	***	135	116	760101	801228
Richmond	302	180	126	***	73	53	760113	801128
Saratoga Hwy. 85 & SPRR	115	54	54	***	30	24	780224	791222
San Francisco Ellis St.	261	116	115	***	58	57	760101	791222
San Francisco 23 St.	289	50	50	***	24	26	780101	791222
San Francisco Grove St.	14	0	0	not used	0	14	780108	780327
San Jose 4 St.	474	273	273	***	145	128	760113	801228
San Rafael	305	116	115	***	59	56	760101	791222
Santa Rosa Humboldt St.	296	251	250	***	136	114	760101	801228
Sunnyvale	126	56	56	***	27	29	760101	771208
Vallejo Tuolumne	297	274	270	***	143	127	760101	801128
NORTH CENTRAL COAST AIR BASIN								
Aptos	294	0	0	*	148	146	760101	801229
Bradley CDF Fire Sta.	48	48	48	***	25	23	781204	791030
Gonzales High School	135	5	0	*	74	61	760101	780625
Hollister	254	0	0	*	136	118	760101	800726
Hollister 1979 Fariview	26	0	0	not used	11	15	800801	801229
King City Pearl St.	31	10	0	not used	15	16	760101	760629
Salinas H.D. Natividad Road	10	10	0	not used	0	10	760101	760224
Salinas II	288	276	272	***	148	124	760406	801129
San Ardo Water District Office	51	51	51	***	27	24	781204	791024
SAN JOAQUIN VALLEY AIR BASIN								
Avenal Fresno St.	2	0	0	not used	0	2	760131	760307
Bakersfield H.D. Flower	287	179	62	**	83	96	770119	801229
Bakersfield Chester St.	345	328	291	***	145	146	760403	801129
Bakersfield Fruitvale	59	58	58	not used	10	48	770718	780323
Bakersfield Foothill High	56	55	55	not used	9	46	770718	780323
Bakersfield Fairview Avenue	47	47	47	not used	10	37	770718	780321
Bakersfield Golden State	29	27	27	not used	0	27	771103	780323
Bakersfield Armory	41	41	41	not used	0	41	771103	780327
Bakersfield Kern City Golf Course	38	38	38	not used	0	38	771103	780323
Bakersfield Federal Building	23	23	23	not used	0	23	780102	780323
Bakersfield College	26	26	26	not used	0	26	771203	780327
Bakersfield Health Department	38	37	37	not used	0	37	771103	780327
Bakersfield Mt. Vernon School	38	38	38	not used	0	38	771103	780323
Coalinga	219	19	0	*	109	110	770101	801217
Corcoran Chittendon	176	0	0	*	86	90	760125	801229
Five Points	281	20	0	*	143	138	760101	801217
Fresno Cedar St.	296	50	0	*	173	123	770106	801111
Fresno Cal State	66	3	0	*	28	38	760101	770224
Fresno Herndon	289	272	272	***	151	121	760403	801105
Fresno Cal State #2	195	15	0	*	100	95	770302	801223
Goshen	158	1	0	*	158	85	760101	771227

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 TO 1980 (Continued).

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	TSP and $SO_4^=$	$TSP, SO_4,$ and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	Winter	DATES OF TSP OR SELECTED DATA SET Start	End
Hanford	151	0	0	*	81	70	760125	801225
Hanford Berry	5	0	0	not used	0	5	790304	790328
Kellerman City	195	0	0	*	96	99	760125	801229
Kern Refuge	235	152	32	**	73	79	770119	801229
Lemoore	106	1	0	*	55	51	760125	790620
Lodi	14	0	0	not used	5	9	770904	771127
Lodi Ham	15	0	0	not used	5	10	780102	780426
Los Banos	286	7	4	*	144	142	760101	801229
Madera Library	33	19	19	not used	16	3	760406	761201
McKittrick Fire Sta.	274	273	53	**	83	190	780120	791012
McKittrick Hwy. 58 & Hwy. 53	212	212	211	not used	25	186	791021	801128
Merced	274	6	3	*	143	131	760101	801229
Merced 18 & S	187	166	166	***	91	75	760406	790421
Modesto J St.	288	11	6	*	150	138	760101	801123
Modesto Oakdale Rd.	203	4	0	*	104	99	760101	801123
Modesto 1100 I St.	21	0	0	not used	15	6	800515	801123
New Jerusalem	30	1	0	not used	10	20	770904	780426
Oildale Fire Sta. 63	312	307	221	***	94	127	771106	801123
Oildale Manor St.	42	41	41	not used	11	30	800801	801128
Parlier	132	21	0	*	71	61	760125	780607
Patterson	158	3	0	*	74	84	760101	790813
Porterville	96	0	0	*	47	49	760101	771227
Porterville S. Main	61	1	0	*	36	25	770805	780830
Salida	87	2	0	*	43	44	760101	771016
Stockton Hazelton St.	190	105	102	***	53	49	790121	801123
Stockton Pacific Ave.	12	12	12	not used	4	8	800825	801123
Stratford	130	0	0	*	56	74	760125	790527
Taft N. 10 St.	251	161	55	**	74	87	770119	801229
Three Rivers	156	1	0	*	84	72	760101	780830
Turlock	257	3	0	*	127	130	760101	801229
Union Island	5	0	0	not used	0	5	780102	780126
Visalia Old Jail	202	7	2	*	104	98	760101	790702
Visalia Church St.	83	0	0	*	44	39	790720	801123
GREAT BASIN VALLEYS AIR BASIN								
Bishop	102	0	0	*	52	50	781104	801229
Coso Junction	88	0	0	*	43	45	790404	801229
Keeler	154	0	0	not used	48	106	790710	800825
Lee Vining	88	0	0	*	48	40	790503	801229
Lone Pine S. Main	49	0	0	not used	36	13	781122	790915
Lone Pine Visitor Cntr.	41	0	0	not used	0	41	791004	800322
Lone Pine 501 E. Locust	31	0	0	not used	18	13	800602	801229
Mammoth June Lakes Airport	27	0	0	not used	10	17	781204	790527
Mammoth Lakes Fire Sta.	47	0	0	not used	12	35	790920	801229
Mono Lake	35	0	0	*	21	14	790801	800714
SOUTH CENTRAL COAST AIR BASIN								
Camarillo Elm Dr.	134	69	69	*	67	67	760113	780520
Carpinteria	208	0	0	*	116	92	760420	801229
El Capitan Beach	115	106	0	**	61	45	790202	801129
El Rio Rio Mesa School	86	0	0	*	45	41	790614	801229
Goleta	223	0	0	*	127	96	760101	801129
Lockwood Valley	203	0	0	*	88	115	760102	801229
Lompoc G St.	278	223	162	***	88	74	780414	801229
Lompoc Jalama Rd.	291	291	0	**	148	143	780707	801229
Morro Bay	290	58	0	*	150	140	760101	801229
Morro Bay Jr. High	59	59	0	**	31	28	770302	780225
Nipoma	288	0	0	*	147	141	760206	801229

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 TO 1980 (Continued).

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	TSP and $SO_4^=$	TSP, $SO_4^=$, and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	Winter	DATES OF TSP OR SELECTED DATA SET Start	End
Ojai	263	9	9	*	134	129	760107	801018
Oxnard	136	0	0	*	63	73	760101	780625
Paso Robles	278	256	256	***	144	112	760406	801129
Piru	113	0	0	*	69	44	780502	801018
Point Mugu	99	0	0	*	54	45	760101	770829
Port Hueneme	239	9	9	*	121	118	760611	801018
San Luis Obispo	41	0	0	not used	25	16	750101	760828
San Luis Obispo Marsh	254	64	6	*	128	126	760903	801129
Santa Barbara	293	272	272	***	148	124	760406	801129
Santa Maria Library	278	215	158	***	88	70	780414	801229
Santa Maria Briarwood Drive	266	265	0	**	121	144	781017	801229
Santa Paula	148	10	10	*	75	73	760101	780625
Santa Ynez	46	46	0	**	28	18	770302	780126
Simi Valley	282	183	182	***	93	39	760705	801229
Thousand Oaks Windsor	254	10	10	*	132	122	760101	800930
Ventura Telegraph Rd.	113	0	0	*	58	55	760101	771227
Ventura Figueroa	22	0	0	not used	14	8	791117	800923
SOUTH COAST AIR BASIN (Coastal Part)								
Anaheim	705	705	297	not used	492	213	760101	801129
Costa Mesa Harbor	192	192	192	***	92	100	760101	790226
Costa Mesa Placentia	71	71	71	***	36	35	790304	800427
El Toro	271	271	271	***	143	128	760506	801129
Harbor City	110	0	0	not used	0	110	781108	790318
Laguna Beach Broadway	78	77	77	***	34	43	760101	770419
Lennox	497	487	289	***	145	144	760101	801129
Long Beach 2655 Pine	118	0	0	not used	0	118	780104	790318
Los Alamitos Orangewood	294	291	291	***	146	145	760101	801129
Los Angeles Downtown	410	395	224	***	118	106	760101	790912
Los Angeles N. Main	102	100	71	***	36	35	790906	801129
Lynwood	295	295	284	***	150	134	760101	801129
North Long Beach	35	35	35	not used	25	10	800503	801129
Pico Rivera	222	221	220	***	121	99	770401	801129
Reseda	685	681	282	not used	487	194	760101	801129
San Juan Capistrano	136	126	126	***	67	59	760506	780731
Santa Ana Police Sta.	165	0	0	*	91	74	780102	800930
Santa Ana Weir Canyon Road	296	296	296	***	150	146	760101	801129
West Los Angeles	120	117	108	***	57	51	760101	771227
West Los Angeles Robertson	160	160	160	***	91	67	780402	801129
SOUTH COAST AIR BASIN (Inland Part)								
Azusa	796	794	291	***	149	142	760101	801129
Big Bear Lake	142	138	136	***	74	62	760101	780722
Burbank	14	0	0	not used	0	14	780108	780327
Chino Riverside Ave.	148	146	142	***	80	62	760102	780731
Crestline	18	18	9	not used	5	13	760101	760430
Dominguez Cal State	117	0	0	not used	0	117	781104	790319
Fontana Foothill Trailer	268	266	261	***	142	119	760103	801129
Glendora Laurel	9	9	0	not used	0	9	801006	801129
La Habra	299	299	298	***	152	146	760101	801129
Lake Gregory	217	212	211	***	112	99	770302	801123
Mt. Lee	549	547	0	not used	460	87	770415	791031
Ontario Airport	183	33	33	*	111	72	760506	801117
Ontario Archibald Ave.	43	42	37	***	15	22	770606	780327

TABLE 4.1 SUMMARY OF HI-VOL DATA AVAILABLE IN CALIFORNIA FOR 1976 to 1980 (Continued).

AIR BASIN Site	NUMBER OF DATA POINTS TSP Only	TSP and $SO_4^=$	TSP, $SO_4^=$, and Pb	DATA SET SELECTED	SEASONAL SPLIT OF SELECTED DATA SET Summer	Winter	DATES OF TSP OR SELECTED DATA SET Start	End
Pasadena Walnut	294	294	294	***	148	146	760101	801129
Redlands	173	160	160	***	79	81	760101	801129
Redlands Univ. of Redlands, Grove	95	95	95	***	52	43	790304	801129
Rialto Airport	164	163	162	***	82	80	760107	790930
Riverside Trailer	217	213	184	***	117	67	760403	771031
Riverside Rubidoux	284	284	281	***	142	139	760101	801129
Riverside Magnolia	1100	1096	365	**	692	404	760101	801130
San Bernardino	297	287	285	***	145	140	760101	801129
Temple City	565	530	1	**	280	250	770101	780712
Upland Civic Center	177	170	170	***	86	84	760206	790930
Upland Post Office	1107	1046	0	**	646	400	770501	801128
SAN DIEGO AIR BASIN								
Alpine Victoria	210	0	0	*	113	97	770131	801129
Brown Field	101	0	0	*	46	55	760506	780327
Chula Vista	292	0	0	*	149	147	760101	801229
El Cajon	392	273	272	***	145	127	760403	801223
Escondido Valley Pkwy.	290	0	0	*	154	136	760101	801229
Imperial Beach	20	0	0	not used	5	15	760101	760430
Oceanside	295	0	0	*	146	149	760101	801229
San Diego Island Ave.	403	265	262	***	138	124	760403	801229
San Diego Overland	287	0	0	*	147	140	760101	801229
SOUTHEAST DESERT AIR BASIN								
Banning Allesandro	294	289	177	***	95	82	760804	801129
Barstow	281	277	272	***	139	133	760101	801129
Boron Fire Station	280	0	0	*	140	140	760101	801223
Brawley Fire Station	275	0	0	*	144	131	760101	801123
Calexico	276	1	1	*	140	136	760101	801123
China Lake	257	26	26	*	134	123	760101	801229
El Centro Broadway	286	258	256	***	139	117	700406	801123
Indio Oasis St.	293	291	282	***	145	137	760101	801129
Lancaster	292	291	281	***	150	131	760107	801129
Mojave	273	0	0	*	135	138	760101	801229
Needles Bailey	17	17	15	not used	9	6	770525	771203
Palm Springs	294	294	174	***	93	81	760804	801129
Palo Verde San Diego G & E	18	18	18	not used	5	13	770904	771215
Trona Market St.	141	140	140	***	72	68	760301	801129
Twentynine Palms Adobe	152	152	152	***	87	65	780408	801129
Victorville	113	112	112	***	57	55	760101	771227
Victorville Fairground	172	172	172	***	92	80	780108	801129

* TSP Only

** TSP and $SO_4^=$

*** TSP, $SO_4^=$, and Pb

notations in Table 4.1. The total number of useable sites is 226; these are illustrated in Figure 4.1.

4.2 METHODOLOGY

In developing the empirical equations relating Hi-Vol variables to IP and FP (Chapter 3), we used data from three regions in California: the San Francisco area, the Central Valley, and the Los Angeles area. Our results indicated that the coefficients in the equations differed substantially among the regions. Furthermore, within each region, we found notable seasonal differences. Because the regional and seasonal variations are significant in California, we have decided to use the complex, disaggregated, "regional/seasonal" models in the application phase. Table 4.2 lists the equations for these models.

In applying the equations of Table 4.2, the San Francisco area is defined specifically as the San Francisco Bay Area Air Basin, the Los Angeles area is defined as the South Coast Air Basin, and the Central Valley is defined as the San Joaquin Valley and Sacramento Valley Air Basins. All other California air basins are put into the "other locations" category. National average coefficients, without seasonal variations, are generally used for "other locations" (see third footnote to Table 4.2). Fortunately, the national average coefficients are quite close to the coefficients for the Pacific Northwest and Arid Southwest areas (see tables of Chapter 3); these latter two areas should most closely resemble the "other locations" in California.

At each application site, we compute two statistics for IP and FP for the period 1976-1980 -- the <u>annual mean</u> concentration and the <u>yearly maximum</u> concentration. The annual mean is just the average of all data over the entire five years. The yearly maximum is computed by interpolating (or in rare cases extrapolating) the <u>actual</u> frequency distribution of the data. The yearly maximum is determined for an every 6th day sampling schedule (61 samples per year).*

---

*The computation of the yearly maximum is as follows: Let N be the total number of data points at the site. The concentrations are ranked in descending order and indexed by $r = 1, \ldots, N$. The cumulative percentile for the expected yearly maximum for 61 samples per year is $P_{max} = 1/62 = 0.016$. We select r^* such that $r^*/(N+1)$ and $(r^*+1)/(N+1)$ surround P_{max}. The expected yearly maximal concentration C_{max} is then interpolated between C_{r^*} and C_{r^*+1}.

Figure 4.1 Hi-Vol monitoring sites used in the application study.

TABLE 4.2 EQUATIONS USED TO PREDICT IP AND FP FROM HI-VOL DATA IN CALIFORNIA.

EQUATION	COEFFICIENT "b" San Francisco Area		Central Valley		Los Angeles Area		Other
	Summer**	Winter**	Summer	Winter	Summer	Winter	Locations***
$IP = b \cdot TSP$	0.43	0.60	0.43	0.59	0.70	0.62	0.61
$IP = 1.2\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	0.38	0.58	0.39	0.57	0.68	0.60	0.56
$IP^{*} = 1.2\ SO_4^{=} + 15\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	0.33	0.48	0.32	0.50	0.66	0.53	0.50
$FP = b \cdot TSP$	0.19	0.31	0.16	0.29	0.29	0.27	0.25
$FP = 1.1\ SO_4^{=} + b(TSP - 1.4\ SO_4^{=})$	0.11	0.27	0.11	0.25	0.24	0.23	0.21
$FP^{*} = 1.1\ SO_4^{=} + 11\ Pb + b(TSP - 1.4\ SO_4^{=} - 15\ Pb)$	0.05	0.16	0.04	0.16	0.20	0.13	0.14

*The lead (Pb) coefficients listed in these equations are for the year 1980. As explained in Appendix A, the Pb coefficients vary from year to year because of changes in the number of catalytic converter cars and because of variance in the amount of lead in leaded gasoline. Figure A.1 in Appendix A shows how the Pb coefficients change with time. We have included the historical changes in the Pb coefficient as part of our calculation scheme.

**Summer is defined as April - September (second and third quarters), while winter is October - March (first and fourth quarters).

***For the "other locations", the coefficients chosen are the national average coefficients which generally are also representative of average Western conditions. The only exception is that 0.25 has been selected for the FP = b·TSP equation; the national average value (0.30) for this equation is not representative of Western locations (see Table 3.12).

Chapter 3 contained a complete discussion of the errors inherent in using our equations to estimate both annual means and daily data points. Unfortunately, the error discussion of Chapter 3 does not transfer unambiguously to the application phase. First, we have the choice of using either national error estimates or regional/seasonal error estimates. Second, the errors in predicting annual means based on five years of data should be slightly less than the errors in predicting annual means based on a single year of data (in essence, we will be extending the error reduction that was found previously in going from daily data points to annual means). Third, the error in predicting the yearly maximum should be less than the error in predicting a single data point because we are addressing a statistical parameter of a frequency distribution rather than a single isolated data point. Precisely quantifying the errors for this application phase would require a very extensive analysis and might not even be possible with the amount of simultaneous Hi-Vol and dichotomous data currently available. In lieu of a precise quantification, we offer the approximate standard errors listed in Table 4.3; these errors are based on a review of the error tables in Chapter 3 and on a consideration of the new factors discussed previously in this paragraph.

TABLE 4.3 APPROXIMATE STANDARD ERRORS IN PREDICTIONS OF ANNUAL MEAN AND YEARLY MAXIMAL VALUES OF IP AND FP.

HI-VOL VARIABLE USED	ERRORS IN IP		ERRORS IN FP	
	Annual Mean	Yearly Maximum	Annual Mean	Yearly Maximum
TSP	∿ 14%	∿ 25%	∿ 24%	∿ 45%
TSP and $SO_4^=$	∿ 13%	∿ 23%	∿ 18%	∿ 36%
TSP, $SO_4^=$, and Pb	∿ 13%	∿ 22%	∿ 16%	∿ 32%

4.3 TABULAR SUMMARY

Table 4.4 lists the yearly mean and yearly maximal values for TSP, IP, and FP at the 226 Hi-Vol sites in California. The yearly mean is just the average of all data for the period 1976-1980; the yearly maximum is calculated from the frequency distribution assuming every 6th day sampling (61 samples per year). The table also lists the max/mean ratio for TSP, IP, and FP. Furthermore, the last column indicates the predictive scheme used (i.e. TSP alone, both TSP and $SO_4^=$, or the triad of TSP, $SO_4^=$, and Pb).

A scanning of Table 4.4 reveals at least two salient features. First, the max/mean ratios for TSP, IP, and FP are generally in the range of 2 to 4, although a few sites have max/mean ratios significantly greater than 4. Second, two of the air basins, the South Coast and San Joaquin Valley Air Basins, generally show higher levels of TSP, IP, and FP than the other air basins.

4.4 GEOGRAPHICAL PATTERNS

Figures 4.2, 4.3, and 4.4 show annual mean values of TSP, estimated IP, and estimated FP plotted at the locations of the 226 monitoring sites. For the IP and FP maps, the numbers in bold face type represent sites where predictive accuracy is increased by the use of $SO_4^=$ and/or Pb data in addition to TSP data. Approximate isopleths drawn to the data are given in Figures 1.1, 1.2, and 1.3 (pages 11, 12, and 13). In drawing the isopleths for IP and FP, we have assigned somewhat greater weight to the bold faced numbers.

The most notable features in the spatial patterns for TSP, IP, and FP are the high concentrations in the South Coast Air Basin (Los Angeles area) and the San Joaquin Valley Air Basin. The South Coast Air Basin (SCAB) is generally the worst area in the state for IP and FP and the second or third worst area in the state for TSP. Within the SCAB, the most extreme values of TSP and IP are found in the eastern portions of the basin -- specifically the Azusa-Upland-Chino-Ontario-Fontana-San Bernardino-Riverside area -- where mean TSP generally exceeds 125 $\mu g/m^3$ and mean IP generally exceeds 85 $\mu g/m^3$. The most extreme values of fine particles (FP > 40 $\mu g/m^3$) occur in a long belt from Lennox on the coast to Downtown Los Angeles, Lynwood, and Pico Rivera in the center and to Chino, Ontario, and Riverside in the east. The

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP IN µg/m³.

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
NORTH COAST AIR BASIN										
Arcata Fire Sta.	49	112	2.3	29	68	2.3	12	28	2.3	*
Capella	74	192	2.6	45	117	2.6	18	48	2.6	*
Cloverdale	42	105	2.5	26	64	2.5	11	26	2.5	*
Cresent City	46	108	2.4	28	66	2.4	11	27	2.4	*
Eureka H.D. 6 & I St.	57	131	2.3	35	80	2.3	14	33	2.3	*
Eureka Hwy. Dept.	64	146	2.3	39	89	2.3	16	37	2.3	*
Ft. Bragg Central	69	155	2.2	37	81	2.2	14	28	2.0	***
Ft. Bragg S. Main	92	254	2.8	50	132	2.7	18	46	2.5	***
Healdsburg	50	109	2.2	31	66	2.2	13	27	2.2	*
Ukiah Firehouse	66	165	2.5	40	100	2.5	17	41	2.5	*
Willets	67	166	2.5	41	101	2.5	17	42	2.5	*
NORTHEAST PLATEAU AIR BASIN										
Alturas	77	214	2.8	41	113	2.8	14	37	2.7	***
Burney	57	138	2.4	35	84	2.4	14	35	2.4	*
Cedarville	34	131	3.8	21	80	3.8	9	33	3.8	*
Fort Jones	62	161	2.6	38	98	2.6	15	40	2.6	*
McCloud	91	277	3.1	55	169	3.1	23	69	3.1	*
Mount Shasta	49	132	2.7	30	81	2.7	12	33	2.7	*
Tulelake Fairground	59	227	3.9	36	139	3.9	15	57	3.9	*
Weed	50	169	3.4	30	103	3.4	12	42	3.4	*
Yreka	46	257	5.5	26	131	5.0	11	44	4.0	***
LAKE COUNTY AIR BASIN										
Hobergs Hwy. 175	61	128	2.1	37	78	2.1	15	32	2.1	*
Kelseyville Dorn Road	42	194	4.7	25	118	4.7	10	48	4.7	*
Lakeport Lake-port Blvd.	32	84	2.7	18	44	2.5	7	15	2.2	***
Upperlake	30	193	6.4	18	118	6.4	8	48	6.4	*
SACRAMENTO VALLEY AIR BASIN										
Anderson Center City	69	149	2.1	35	79	2.3	15	38	2.5	*
Buckeye Elementary School	61	160	2.6	30	80	2.7	13	38	2.9	*
Chico	78	162	2.1	41	96	2.4	18	47	2.6	*
Chico Manzanita	64	152	2.4	32	89	2.8	14	44	3.1	*
Chico State	65	149	2.3	33	79	2.4	14	39	2.7	*
Citrus Heights Sunrise Blvd.	79	153	1.9	37	84	2.3	14	35	2.5	***
Corning	76	196	2.6	38	116	3.0	17	57	3.4	*
Davis 5 St.	74	191	2.6	39	109	2.8	17	54	3.1	*
Dunnigan Main St.	70	209	3.0	35	120	3.5	15	59	3.9	*
Gridley Gray Lodge	57	155	2.7	29	92	3.2	12	45	3.6	*
Live Oak	99	230	2.3	50	136	2.7	22	67	3.1	*
Los Molinos	64	155	2.4	31	81	2.6	14	39	2.9	*
Marysville	58	144	2.5	28	73	2.6	12	36	3.0	*
Mountain Gate	74	231	3.1	34	119	3.5	12	42	3.5	***

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
Nord	49	201	4.1	24	88	3.6	11	36	3.4	*
Oroville Bird Street	63	156	2.5	32	86	2.7	14	42	3.1	*
Oroville 6 WNW	64	130	2.0	31	75	2.4	12	34	2.7	**
Pleasant Grove	76	228	3.0	39	120	3.1	17	59	3.5	*
Rancho Cordova	95	272	2.9	47	124	2.6	20	58	2.9	*
Red Bluff	77	204	2.6	38	109	2.9	16	52	3.2	*
Red Bluff Ag. Comm. Office	60	142	2.4	30	83	2.8	13	40	3.1	*
Red Bluff Lincoln	65	163	2.5	30	87	2.9	11	36	3.3	***
Redding H.D. Roof	58	128	2.2	30	66	2.2	12	32	2.5	*
Rio Vista	60	179	3.0	31	100	3.2	14	49	3.6	*
Sacramento H.D. Stockton Blvd.	66	207	3.1	33	122	3.6	15	60	4.1	*
Sacramento 1025 P St.	76	192	2.5	38	106	2.8	16	48	3.0	***
Sacramento Branch Center Road	76	200	2.7	38	106	2.8	17	51	3.1	*
Sherman Island	117	999	8.6	59	572	9.7	25	254	10.4	**
Smartville	42	143	3.4	21	82	3.9	9	40	4.5	*
Sutter City	92	279	3.0	46	152	3.3	20	75	3.8	*
Vacaville	65	171	2.6	33	101	3.1	14	49	3.4	*
Vacaville Merchant	51	110	2.2	26	65	2.5	12	32	2.8	*
Weaverville Hospital	36	121	3.3	19	71	3.7	9	35	3.9	*
West Sacramento 15 Street	76	185	2.4	39	108	2.8	17	53	3.1	*
Wheatland	70	179	2.6	35	94	2.7	15	46	3.1	*
Wheatland 4 W	80	457	5.7	41	270	6.6	18	133	7.4	*
Williams	70	185	2.7	36	115	3.2	16	56	3.6	*
Willows	61	127	2.1	32	75	2.4	14	37	2.6	*
Willows 5 M West	58	109	1.9	30	64	2.2	13	31	2.4	*
Willows 8 W	44	207	4.7	22	119	5.4	9	53	5.9	**
Woodland West Main Street	81	202	2.5	41	119	2.9	18	59	3.3	*
Yuba City	99	262	2.7	49	155	3.2	21	76	3.6	*
MOUNTAIN COUNTIES AIR BASIN										
Auburn DeWitt Center	51	101	2.0	31	61	2.0	13	25	2.0	*
Camino	36	90	2.5	22	55	2.5	9	23	2.5	*
Georgetown	32	79	2.5	19	48	2.5	8	20	2.5	*
Lincoln	61	116	1.9	37	71	1.9	15	29	1.9	*
Portola	49	91	1.8	30	55	1.8	12	23	1.8	*
Rocklin Sierra College	51	115	2.2	31	70	2.2	13	29	2.2	*
Sierra City	22	52	2.4	13	32	2.4	5	13	2.4	*
Sonora	60	112	1.9	36	68	1.9	15	28	1.9	*
Sonora Forrest Road	41	74	1.8	25	45	1.8	10	18	1.8	*
Tuolomne City	53	98	1.8	32	60	1.9	13	25	1.9	*

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
LAKE TAHOE AIR BASIN										
North Lake Tahoe USCG Station	19	45	2.4	11	27	2.4	5	11	2.4	*
South Lake Tahoe Police Dept.	55	115	2.1	34	70	2.1	14	29	2.1	*
South Lake Tahoe Airport	30	103	3.5	18	63	3.5	7	26	3.5	*
Tahoe City	45	150	3.3	27	92	3.3	11	38	3.3	*
SAN FRANCISCO BAY AREA AIR BASIN										
Berkeley	45	109	2.4	23	65	2.8	11	33	3.0	*
Bethel Island	119	487	4.1	57	194	3.4	24	96	4.1	**
Burlingame Burlingame Ave.	45	101	2.3	25	64	2.5	13	38	2.9	***
Concord Treat Boulevard	52	186	3.6	28	117	4.2	14	65	4.6	***
Fremont Chapel Way	63	183	2.9	34	99	3.0	16	44	2.7	***
Gilroy Monterey Street	66	136	2.1	33	76	2.3	15	41	2.8	***
Livermore Rail-Road	79	182	2.3	38	99	2.6	16	48	3.1	***
Napa Jefferson Street	59	137	2.3	31	78	2.5	15	37	2.5	***
Oakland Jackson	53	167	3.1	28	93	3.3	14	48	3.5	*
Pittsburg	68	167	2.5	33	93	2.8	15	43	2.8	***
Redwood City	58	149	2.6	31	81	2.6	15	40	2.7	***
Richmond	56	129	2.3	30	77	2.6	15	45	2.9	***
Saratoga Hwy. 85 & SPRR	53	106	2.0	26	68	2.6	12	39	3.3	***
San Francisco Ellis St.	48	133	2.8	31	84	2.7	19	46	2.5	***
San Francisco 23 Street	50	124	2.4	29	75	2.5	16	43	2.7	***
San Jose 4 St.	75	168	2.3	39	108	2.8	19	55	3.0	***
San Rafael	42	117	2.8	26	79	3.0	15	45	2.9	***
Santa Rosa Humboldt St.	46	107	2.4	23	60	2.6	11	28	2.6	***
Sunnyvale	56	169	3.0	31	98	3.1	16	46	2.9	***
Vallejo Tuolumne	52	129	2.5	28	73	2.7	14	38	2.7	***
NORTH CENTRAL COAST AIR BASIN										
Aptos	37	80	2.2	23	49	2.2	9	20	2.2	*
Bradley CDF Fire Sta.	48	87	1.8	28	47	1.7	13	21	1.6	***
Gonzales High School	64	149	2.3	39	91	2.3	16	37	2.3	*
Hollister	54	122	2.3	33	74	2.3	13	30	2.3	*
Salinas II	56	130	2.3	32	71	2.2	14	27	2.0	***
San Ardo Water District Off.	85	182	2.1	47	92	2.0	19	41	2.1	***

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
SAN JOAQUIN VALLEY AIR BASIN										
Bakersfield H.D. Flower	156	457	2.9	82	264	3.2	39	128	3.3	**
Bakersfield Chester St.	159	366	2.3	83	219	2.7	38	109	2.9	***
Coalinga	90	263	2.9	45	146	3.3	20	71	3.6	*
Corcoran Chittendon	173	518	3.0	91	306	3.4	41	150	3.7	*
Five Points	115	322	2.8	58	184	3.2	25	85	3.4	*
Fresno Cedar St.	119	270	2.3	59	153	2.6	25	75	3.0	*
Fresno Cal State	121	323	2.7	65	191	2.9	30	94	3.2	*
Fresno Herndon	134	320	2.4	69	200	2.9	30	96	3.2	***
Fresno Cal State # 2	116	296	2.6	59	175	3.0	26	86	3.3	*
Goshen	156	347	2.2	78	196	2.5	34	96	2.8	*
Hanford	155	405	2.6	78	239	3.1	34	118	3.5	*
Kellerman City	136	402	2.9	70	232	3.3	31	114	3.7	*
Kern Refuge	93	373	4.0	48	218	4.5	22	103	4.6	**
Lemoore	131	419	3.2	67	247	3.7	30	121	4.1	*
Los Banos	89	309	3.5	47	182	3.9	21	90	4.3	*
McKittrick Fire Station	121	255	2.1	65	153	2.4	32	78	2.5	**
Merced	99	244	2.5	50	144	2.9	22	71	3.2	*
Merced 18 & S	103	346	3.4	50	188	3.8	20	75	3.8	***
Modesto J St.	108	267	2.5	56	158	2.8	25	77	3.1	*
Modesto Oakdale Road	85	206	2.4	44	121	2.8	20	60	3.1	*
Oildale	160	414	2.6	83	225	2.7	38	108	2.8	***
Parlier	128	355	2.8	63	176	2.8	27	69	2.6	*
Patterson	89	237	2.7	46	131	2.9	20	64	3.2	*
Porterville	142	340	2.4	72	159	2.2	32	78	2.5	*
Porterville S. Main	123	231	1.9	60	136	2.3	26	67	2.6	*
Salida	84	191	2.3	43	113	2.6	19	55	2.9	*
Stockton Hazelton St.	86	228	2.7	41	120	2.9	17	49	2.9	***
Stratford	140	442	3.2	74	261	3.5	34	128	3.8	*
Taft	118	268	2.3	62	159	2.6	30	79	2.6	**
Three Rivers	73	165	2.2	36	71	2.0	15	35	2.3	*
Turlock	98	284	2.9	51	168	3.3	23	82	3.6	*
Visalia Old Jail	145	384	2.7	74	204	2.8	32	100	3.1	*
Visalia Church Street	133	293	2.2	67	173	2.6	29	85	2.9	*
GREAT BASIN VALLEYS AIR BASIN										
Bishop	41	193	4.7	25	118	4.7	10	48	4.7	*
Coso Junction	42	139	3.3	25	85	3.3	10	35	3.3	*
Lee Vining	26	118	4.5	16	72	4.5	7	29	4.5	*
Mono Lake	140	2395	17.1	86	1461	17.1	35	599	17.1	*
SOUTH CENTRAL COAST AIR BASIN										
Camarillo Elm Drive	81	193	2.4	49	118	2.4	20	48	2.4	*

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
Carpinteria	55	123	2.2	33	75	2.2	14	31	2.2	*
El Capitan Beach	102	237	2.3	61	141	2.3	29	67	2.3	**
El Rio Rio Mesa School	81	251	3.1	49	153	3.1	20	63	3.1	*
Goleta	58	110	1.9	35	67	1.9	14	28	1.9	*
Lockwood Valley	41	133	3.2	25	81	3.2	10	33	3.2	*
Lompoc G St.	81	177	2.2	45	95	2.1	19	34	1.8	***
Lompoc Jalama Road	39	96	2.4	25	57	2.3	13	30	2.2	**
Morro Bay	59	127	2.2	36	78	2.2	15	32	2.2	*
Morro Bay Jr. High	77	205	2.7	46	121	2.6	22	56	2.5	**
Nipoma	62	133	2.2	38	81	2.2	14	33	2.2	*
Ojai	71	136	1.9	43	83	1.9	18	34	1.9	*
Oxnard	76	186	2.5	46	113	2.5	19	46	2.5	*
Paso Robles	77	163	2.1	44	94	2.1	19	40	2.1	***
Piru	79	172	2.2	48	105	2.2	20	43	2.2	*
Point Mugu	59	111	1.9	36	68	1.9	15	28	1.9	*
Port Hueneme	97	210	2.2	59	128	2.2	24	53	2.2	*
San Luis Obispo Marsh	55	129	2.3	34	79	2.3	14	32	2.3	*
Santa Barbara	70	131	1.9	45	82	1.8	23	43	1.9	***
Santa Maria Library	105	288	2.8	60	150	2.5	27	51	1.9	***
Santa Maria Briarwood Dr.	73	220	3.0	45	128	2.9	22	55	2.5	**
Santa Paula	83	169	2.0	51	103	2.0	21	42	2.0	*
Santa Ynez	50	102	2.0	30	62	2.0	15	33	2.2	**
Simi Valley	81	163	2.0	49	98	2.0	24	50	2.1	***
Thousand Oaks Windsor	65	190	2.9	40	116	2.9	16	47	2.9	*
Ventura Telegraph Rd.	64	129	2.0	39	79	2.0	16	32	2.0	*
SOUTH COAST AIR BASIN (Coastal Part)										
Costa Mesa Harbor	75	179	2.4	54	119	2.2	28	64	2.3	***
Costa Mesa Placentia	98	249	2.5	66	153	2.3	30	70	2.3	***
El Toro	81	163	2.0	55	109	2.0	26	57	2.2	***
Laguna Beach Broadway	81	184	2.3	58	124	2.2	30	62	2.1	***
Lennox	97	211	2.2	75	154	2.1	45	94	2.1	***
Los Alamitos Orangewood	105	253	2.4	73	165	2.3	36	84	2.3	***
Los Angeles Downtown	108	253	2.3	78	180	2.3	42	103	2.5	***
Los Angeles N. Main	121	266	2.2	85	189	2.2	43	94	2.2	***
Lynwood	115	241	2.1	83	166	2.0	43	91	2.1	***
Pico Rivera	125	281	2.3	89	204	2.3	45	102	2.3	***
San Juan Capistrano	80	160	2.0	54	108	2.0	25	53	2.1	***
Santa Ana Police Sta.	92	267	2.9	61	166	2.7	26	72	2.8	*
Santa Ana Weir Canyon Road	96	234	2.4	66	148	2.2	32	68	2.1	***
W. Los Angeles	72	172	2.4	51	105	2.1	27	58	2.1	***

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
W. Los Angeles Robertson	73	137	1.9	54	99	1.8	30	59	2.0	***
SOUTH COAST AIR BASIN (Inland Part)										
Azusa	130	296	2.3	88	207	2.4	40	94	2.4	***
Big Bear Lake	52	112	2.2	33	73	2.2	12	36	2.9	***
Chino River-side Ave.	161	358	2.2	108	253	2.3	46	106	2.3	***
Fontana Foothill Trailer	140	337	2.4	93	227	2.4	39	100	2.5	***
La Habra	112	251	2.2	77	158	2.1	37	73	2.0	***
Lake Gregory	62	128	2.1	41	81	2.0	18	37	2.1	***
Ontario Airport	154	375	2.4	103	259	2.5	44	107	2.5	*
Ontario Archi-bald Avenue	115	205	1.8	75	126	1.7	32	64	2.0	***
Pasadena Walnut	103	204	2.0	76	149	2.0	41	84	2.1	***
Redlands	103	323	3.1	70	223	3.2	30	86	2.8	***
Redlands Univ. of, Grove	122	299	2.4	81	193	2.4	34	83	2.5	***
Rialto Airport	121	336	2.8	81	212	2.6	34	93	2.7	***
Riverside Trailer	117	218	1.9	82	158	1.9	38	74	1.9	***
Riverside Rubidoux	167	397	2.4	109	271	2.5	44	111	2.5	***
Riverside Magnolia	148	199	1.3	100	96	1.0	43	43	1.0	**
San Bernardino	123	365	3.0	84	240	2.9	37	101	2.7	***
Temple City	109	234	2.2	73	146	2.0	33	72	2.2	**
Upland Civic Center	123	317	2.6	83	198	2.4	37	91	2.4	***
Upland Post Office	126	250	2.0	85	167	2.0	38	79	2.1	**
SAN DIEGO AIR BASIN										
Alpine Victoria	53	112	2.1	32	69	2.1	13	28	2.1	*
Brown Field	56	119	2.1	34	73	2.1	14	30	2.1	*
Chula Vista	63	116	1.9	38	71	1.9	16	29	1.9	*
El Cajon	86	160	1.9	54	103	1.9	27	53	1.9	***
Escondido Valley Pkwy.	86	159	1.9	52	97	1.9	21	40	1.9	*
Oceanside S. Cleveland	91	174	1.9	55	106	1.9	23	44	1.9	*
San Diego Island Ave.	82	191	2.3	51	123	2.4	27	62	2.3	***
San Diego Overland	61	179	2.9	37	109	2.9	15	45	2.9	*
SOUTHEAST DESERT AIR BASIN										
Banning Allesandro	83	191	2.3	47	108	2.3	21	46	2.2	***
Barstow	96	306	3.2	54	164	3.1	22	60	2.7	***
Boron Fire Sta.	77	362	4.7	47	221	4.7	19	90	4.7	*
Brawley Fire Sta.	205	479	2.3	125	292	7.3	51	120	2.3	*

TABLE 4.4 ANNUAL MEAN AND YEARLY MAXIMUM CONCENTRATIONS OF TSP, ESTIMATED IP, AND ESTIMATED FP (Continued).

AIR BASIN	TSP			IP			FP			DATA SET SELECTED
Site	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	Yearly Mean	Yearly Max	Max/Mean Ratio	
Calexico	220	513	2.3	134	313	2.3	55	128	2.3	*
China Lake	63	234	3.7	38	143	3.7	16	59	3.7	*
El Centro Broadway	121	320	2.6	65	166	2.5	24	54	2.2	***
Indio Oasis St.	104	346	3.3	57	178	3.1	22	57	2.6	***
Lancaster	96	304	3.2	53	158	3.0	21	58	2.7	***
Mojave	78	190	2.4	48	116	2.4	20	48	2.4	*
Palm Springs	68	192	2.8	38	99	2.6	16	33	2.1	***
Trona Market St.	125	379	3.0	70	204	2.9	30	77	2.6	***
Twentynine Palms Adobe	57	130	2.3	31	68	2.2	12	23	1.9	***
Victorville	89	172	1.9	50	92	1.9	20	41	2.1	***
Victorville Fairgrounds	82	250	3.0	46	130	2.8	19	46	2.4	***

* TSP Only

** TSP and $SO_4^=$

*** TSP, $SO_4^=$, and Pb

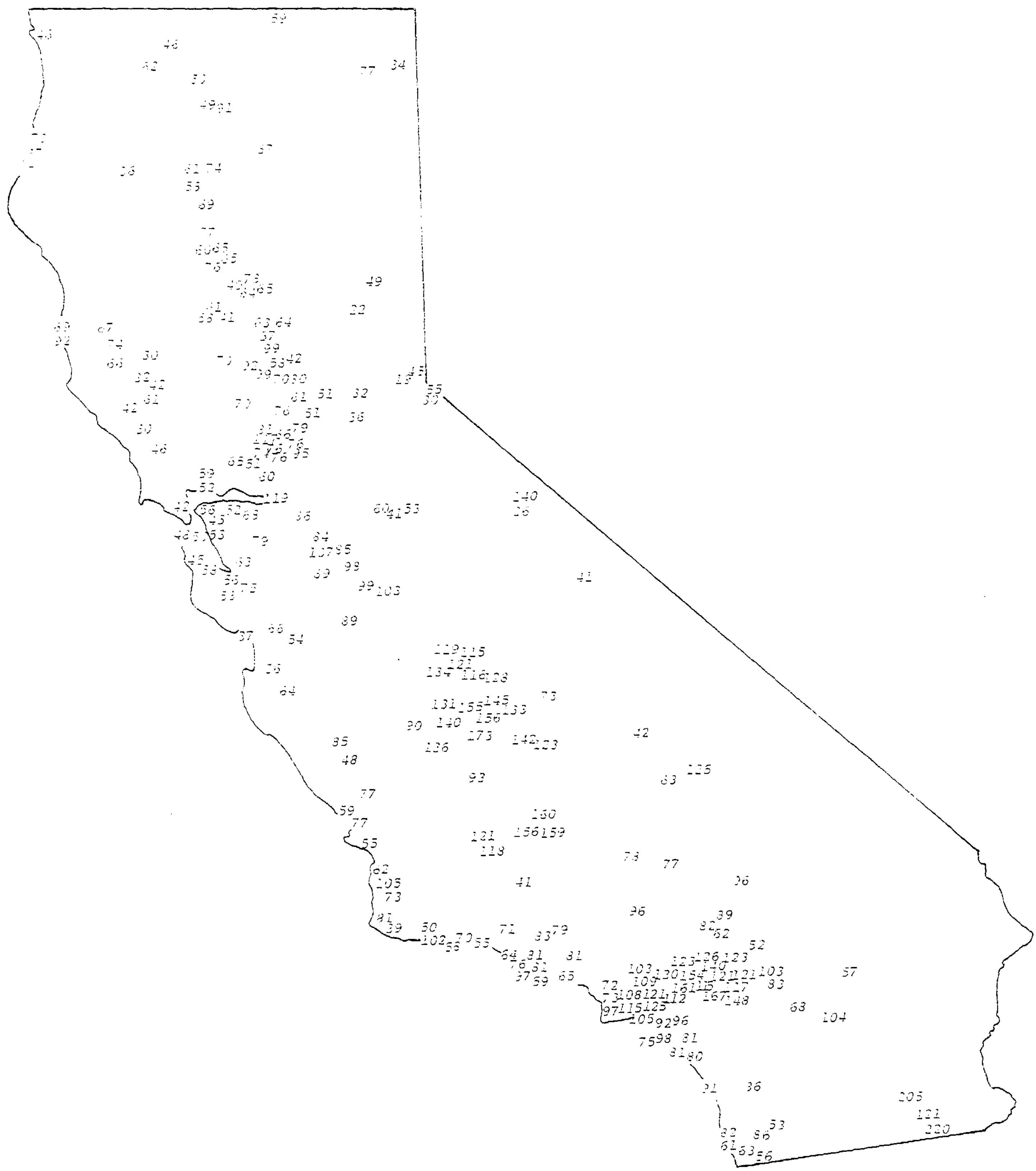
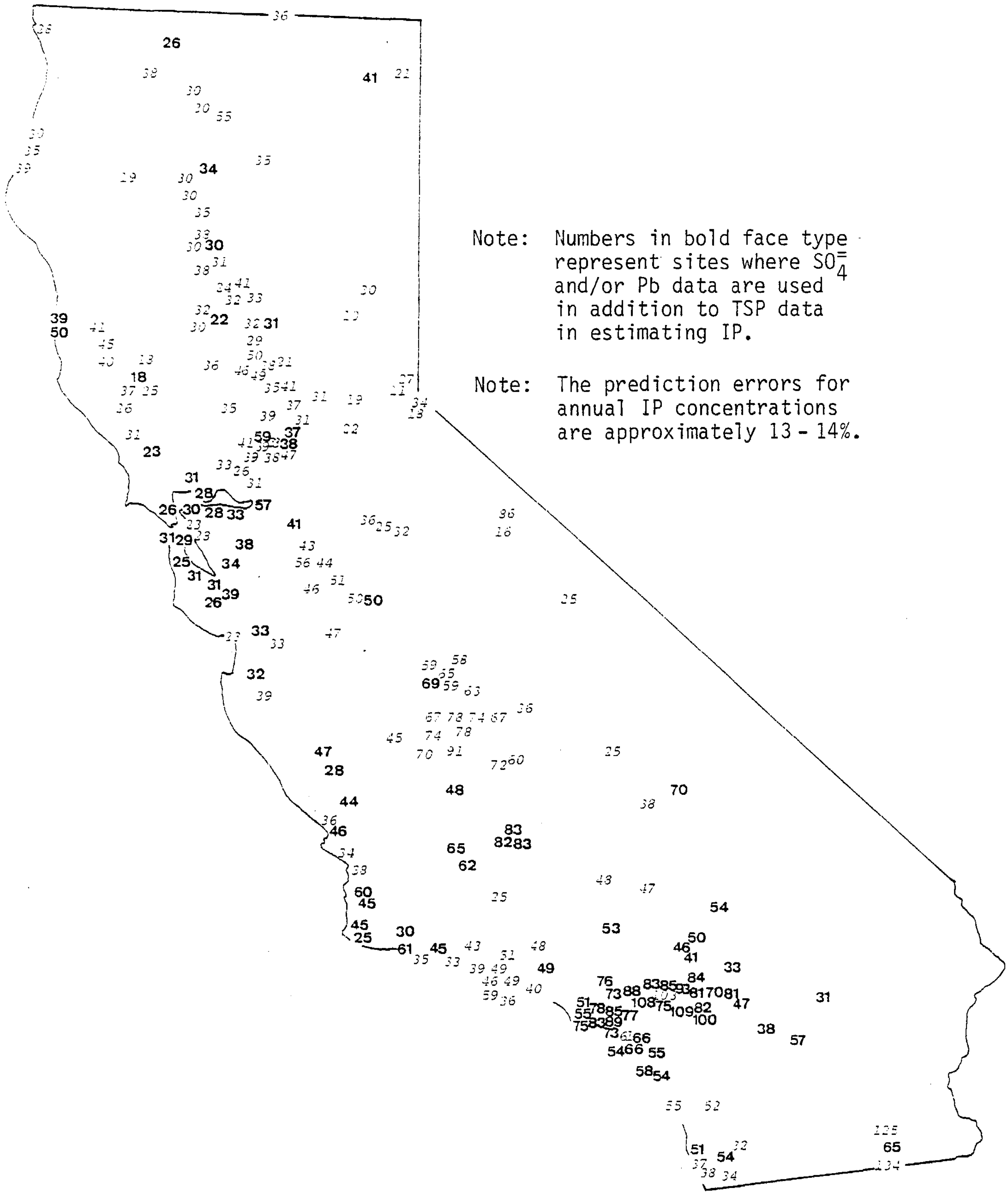


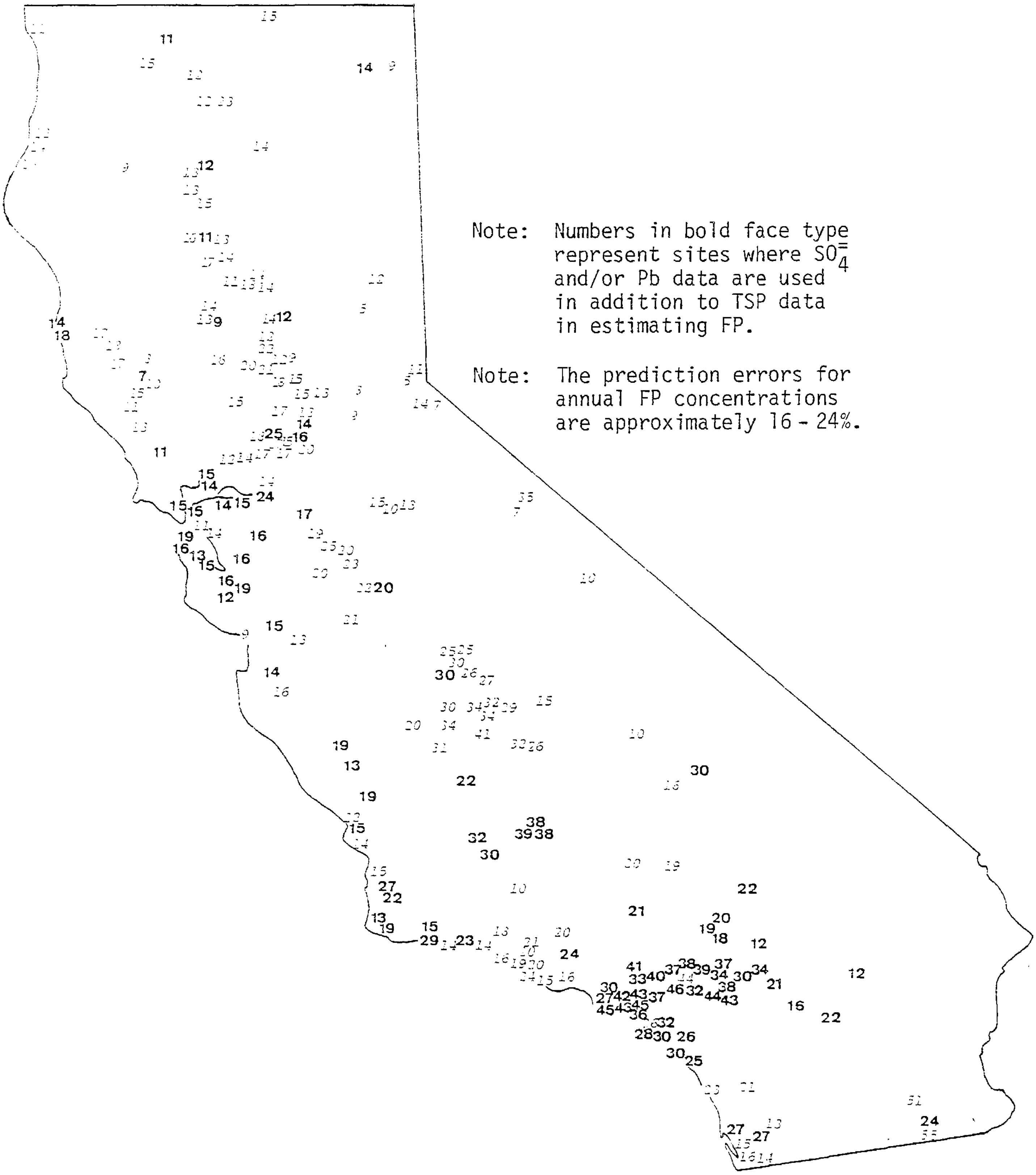
Figure 4.2 Annual mean values of TSP ($\mu g/m^3$) in California during 1976 - 1980.



Note: Numbers in bold face type represent sites where $SO_4^=$ and/or Pb data are used in addition to TSP data in estimating IP.

Note: The prediction errors for annual IP concentrations are approximately 13 - 14%.

Figure 4.3 Annual mean values of predicted IP ($\mu g/m^3$) in California during 1976 - 1980.



Note: Numbers in bold face type represent sites where $SO_4^=$ and/or Pb data are used in addition to TSP data in estimating FP.

Note: The prediction errors for annual FP concentrations are approximately 16 - 24%.

Figure 4.4 Annual mean values of predicted FP ($\mu g/m^3$) in California during 1976 - 1980.

San Joaquin Valley is generally the worst area of the State for TSP (with the possible exception of the Imperial Valley) and second only to the SCAB in IP and FP. Within the San Joaquin Valley, the highest mean concentrations (TSP > 150 $\mu g/m^3$, IP > 70 $\mu g/m^3$, and FP > 30 $\mu g/m^3$) occur in the southern parts of the Valley, from Hanford (just south of Fresno) down to Bakersfield.

The limited data available for the southeast corner of California suggest that the Imperial Valley may possibly be the worst area of the State for TSP and the third worst area for IP and FP. These conclusions are very tenuous because there are only three Hi-Vol monitoring sites in the Imperial Valley, and because we did not have dichotomous sampler data with which to derive area-specific predictive formulae for the Imperial Valley. In the future, it may be very worthwhile to add dichotomous samplers and/or expand the Hi-Vol network in the Imperial Valley.

Two anomalous sites in eastern California deserve brief mention. Both Mono Lake and Trona exhibit unusually high TSP levels compared to other locations near the Nevada border. The high TSP levels are likely due to wind blown dust from dry lake beds, Mono Lake and Searles Lake, respectively.

As emphasized in the isopleth maps (Figures 1.1 to 1.3), the lowest particulate concentrations in California occur in the eastern edge of the state along the Nevada border. In this area, TSP, IP, and FP generally average less than 50 $\mu g/m^3$, 25 $\mu g/m^3$, and 10 $\mu g/m^3$, respectively. There also appears to be a band of low concentrations in the northwest part of the state, from Trinity County down to Lake County.

Many of the above geographical patterns in particulate concentrations make sense in terms of the spatial distributions of man-made emissions in California. Figure 4.5 presents the spatial distribution of primary particulate emissions, while Figures 4.6 to 4.8 present the spatial distribution of emissions for the gaseous precursors of secondary aerosols. The Los Angeles area and the southern San Joaquin Valley both stand out as hot-spots for particulate and SO_x emissions, while the Los Angeles area alone stands out for NO_x and reactive organics. Thus, the high particulate concentrations found in the South Coast and southern San Joaquin Valley Air Basins are reasonable in terms of general statewide emissions patterns. The most

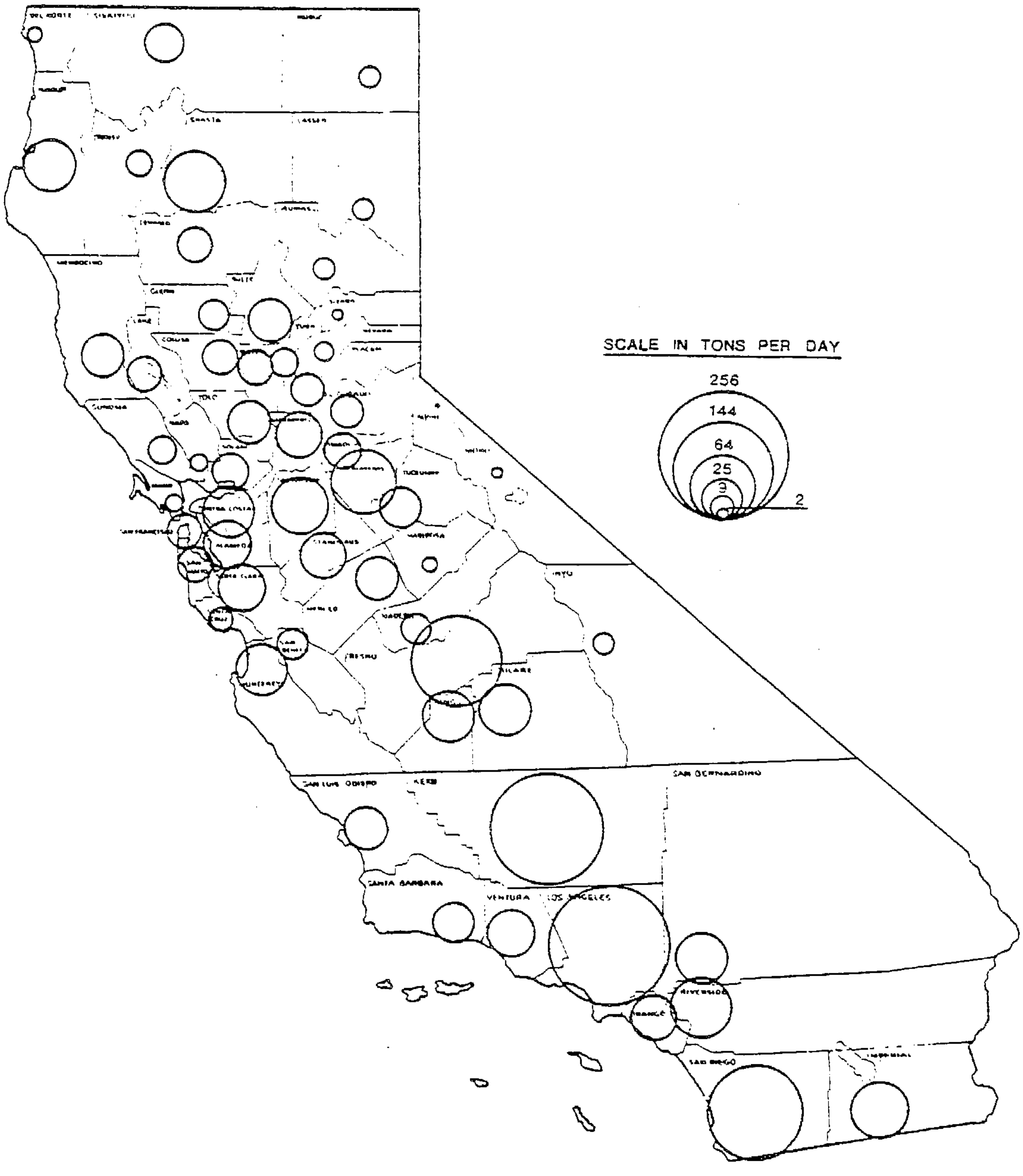


Figure 4.5 Spatial distribution of particulate emissions in California (ARB, 1978).

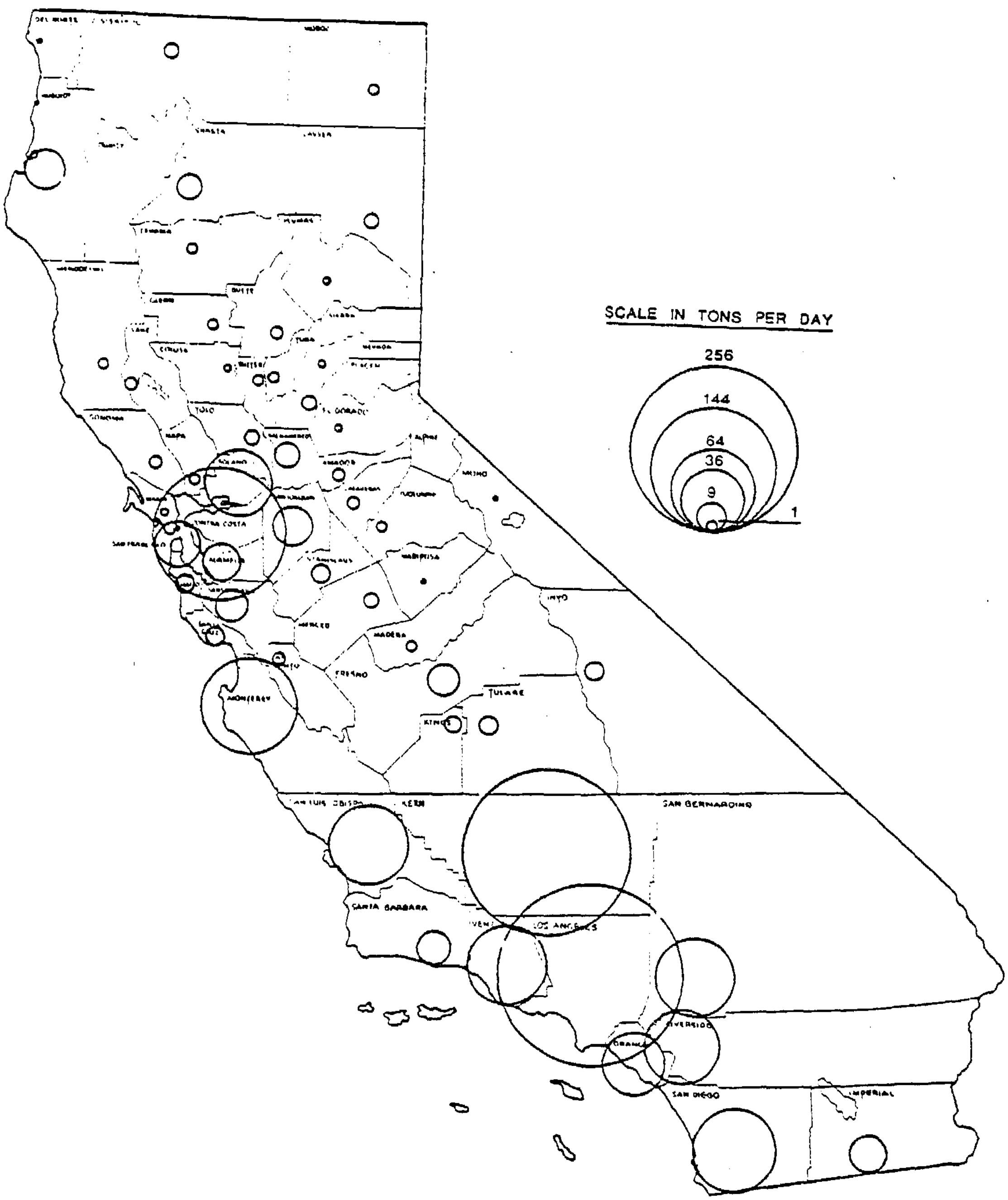


Figure 4.6 Spatial distribution of SO_2 emissions in California (ARB, 1978).

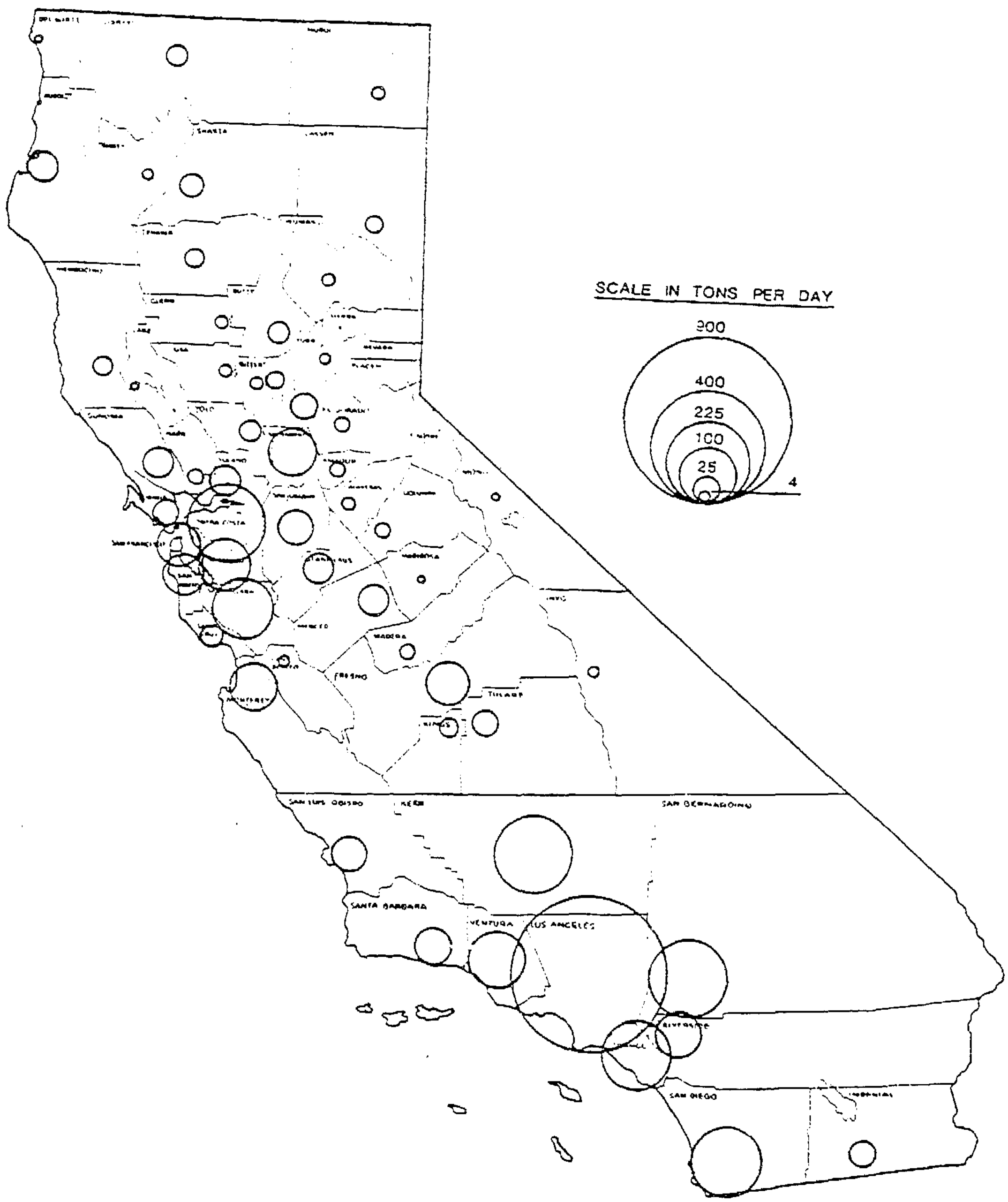


Figure 4.7 Spatial distribution of NO_x emissions in California (ARB, 1978).

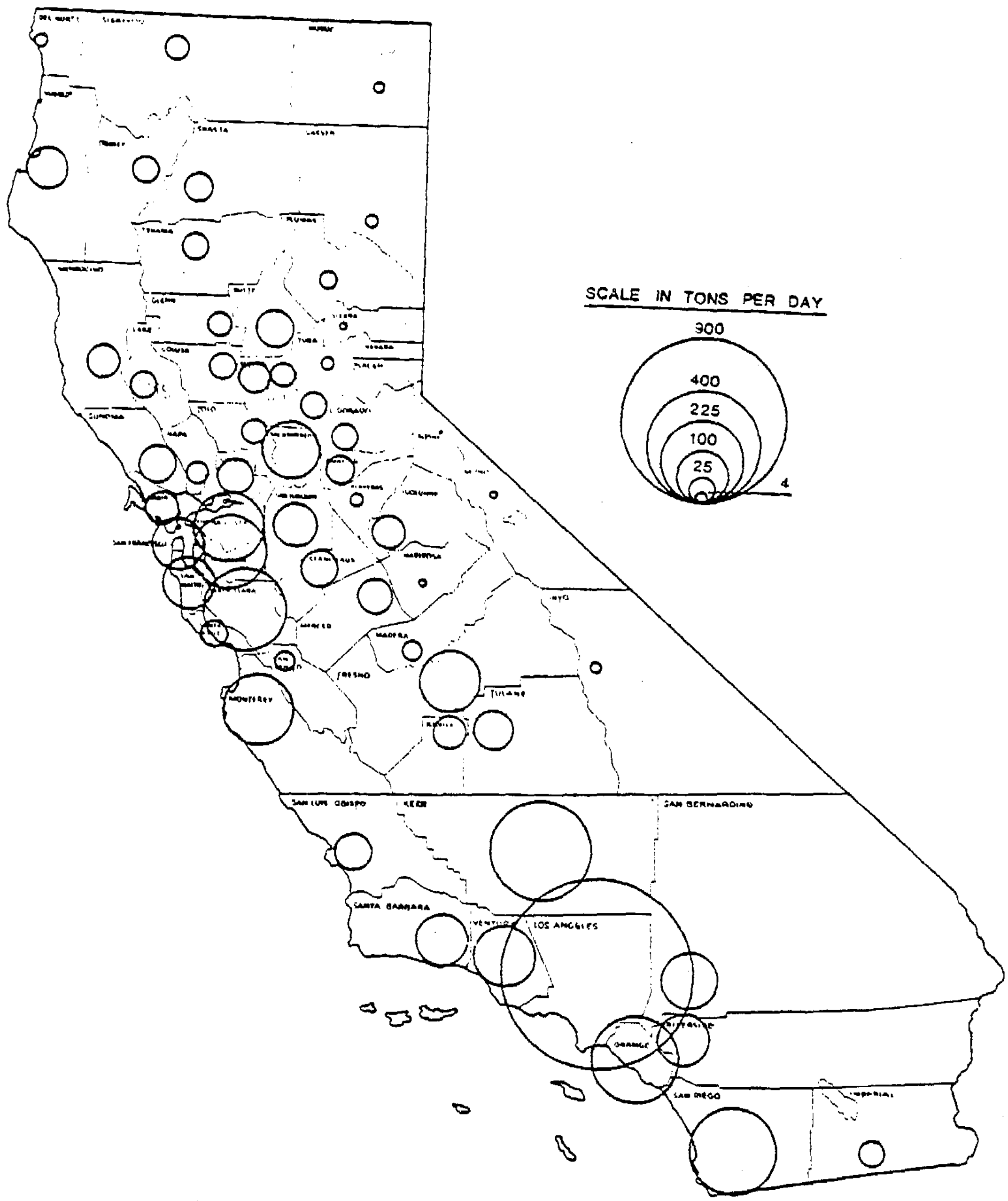


Figure 4.8 Spatial distribution of reactive organic emissions in California (ARB, 1978).

likely cause for the especially high concentrations of IP and FP in the SCAB is the particularly high level of fine secondary aerosols there (Trijonis 1980). The most likely reason for the relatively high TSP concentrations in the San Joaquin Valley is the relatively greater fugitive dust level there (e.g. from agricultural fields). It is also noteworthy that the lowest emission density is along the Nevada border, the area of lowest particulate concentrations.

In a previous report (Trijonis 1980), we examined the geographical patterns of visibility throughout California (see Figure 4.9). That report emphasized that visibility is basically governed by fine particle concentrations and meteorology (e.g. relative humidity). Several analyses in the visibility report indicated that the pockets of low visibility in the South Coast and San Joaquin Valley Air Basins are essentially caused by high levels of anthropogenic fine aerosols. The present study supports that conclusion in the sense that the Los Angeles area and San Joaquin Valley have been shown to be distinct pockets of high fine aerosol concentrations. We also note several other correspondences between the visibility report and the present study:

- The best visibility in California occurs along the Nevada border. Similarly, the lowest fine particle concentrations occur along the Nevada border.
- In the visibility study, we concluded that the west to east gradient of visibility in far northern California (along the Oregon border) is partly due to meteorological factors. This conclusion is supported by the fact that far northern California exhibits only a very slight east to west gradient in fine particle concentrations.
- In the visibility report, we noted that the relatively high density of aerosol precursor emissions in the San Francisco Bay Area produces neither a very great perturbation in local visibility nor very large concentrations of secondary aerosols (sulfates and nitrates). Correspondingly, we now find that San Francisco does not exhibit high concentrations of fine particles. The paradox of relatively low fine aerosol concentrations compared to the emission density might be

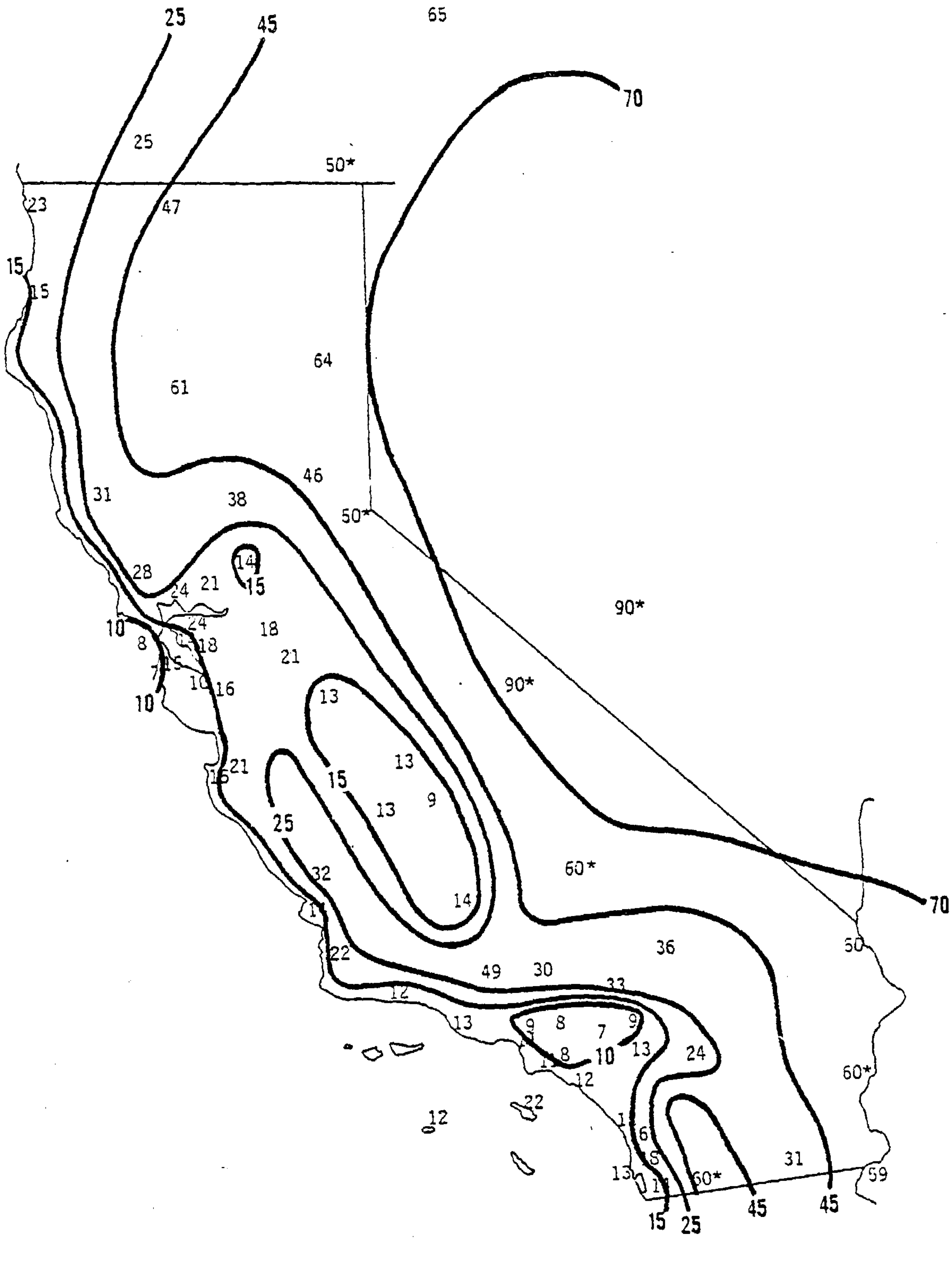


Figure 4.9 Median 1 PM visibilities (in miles) and visibility isopleths for California (Trijonis 1980).

explained by two factors: (1) sunlight intensity is relatively low compared to southern California (NOAA 1977), leading to slower formation rates for photochemical aerosols in the Bay Area and (2) wind speeds are relatively high in the Bay Area (NOAA 1977; Bell 1958), tending to move the precursor emissions into the Central Valley rather quickly. The shape of the visibility and aerosol isopleths east of the Bay Area suggests that the principal visibility impact of Bay Area emissions may tend to be a diluted effect occurring in the Central Valley rather than a concentrated effect occurring locally.

4.5 SEASONAL PATTERNS

Figures 4.10a through 4.10k summarize the seasonal patterns (quarterly averages) of TSP, IP, and FP in various California air basins. These figures are restricted only to those sites that have $SO_4^=$ and/or Pb data in addition to TSP data, so that the IP and FP estimates contain some implicit information on the fine aerosol fraction. The right hand side of each figure shows the seasonal pattern of median extinction coefficient. Extinction (B) is computed from visual range (V) using airport visibility data reported by Trijonis (1980) and using the Koschmeider formula, $B = 24.3/V$, where the units of B are $[10^{-4}m^{-1}]$ and the units of V are [miles].

Scanning Figures 4.10a through 4.10k, several general tendencies can be noted:

- The seasonal patterns of FP often diverge significantly from the seasonal patterns of TSP (see for example the San Francisco Bay Area Air Basin, the inland South Coast Air Basin, and <u>especially</u> the Sacramento Valley and San Joaquin Valley Air Basins). As is reasonable considering the relative particle size cut-offs of TSP, IP, and FP, the seasonal patterns of IP are intermediate to those of TSP and FP.

- The seasonal patterns of extinction usually correspond more closely to the seasonal patterns of FP than to the seasonal patterns of IP or TSP (see in particular the Sacramento Valley, San Francisco Bay Area, San Joaquin Valley, and inland South Coast Air Basins). This makes sense because total light scattering by all particles tends to

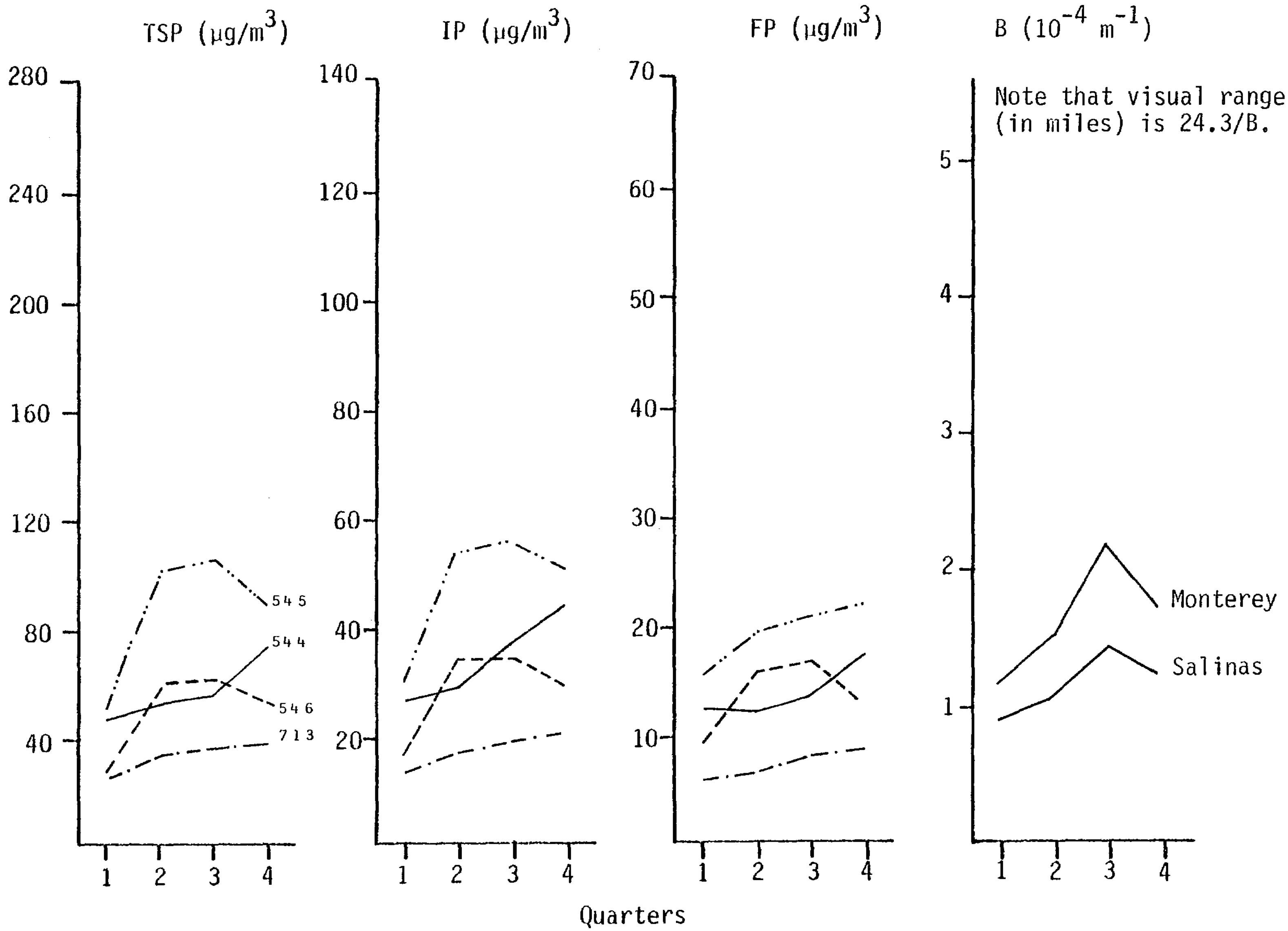


Figure 4.10a North Central Coast and Lake County Air Basins.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction. Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

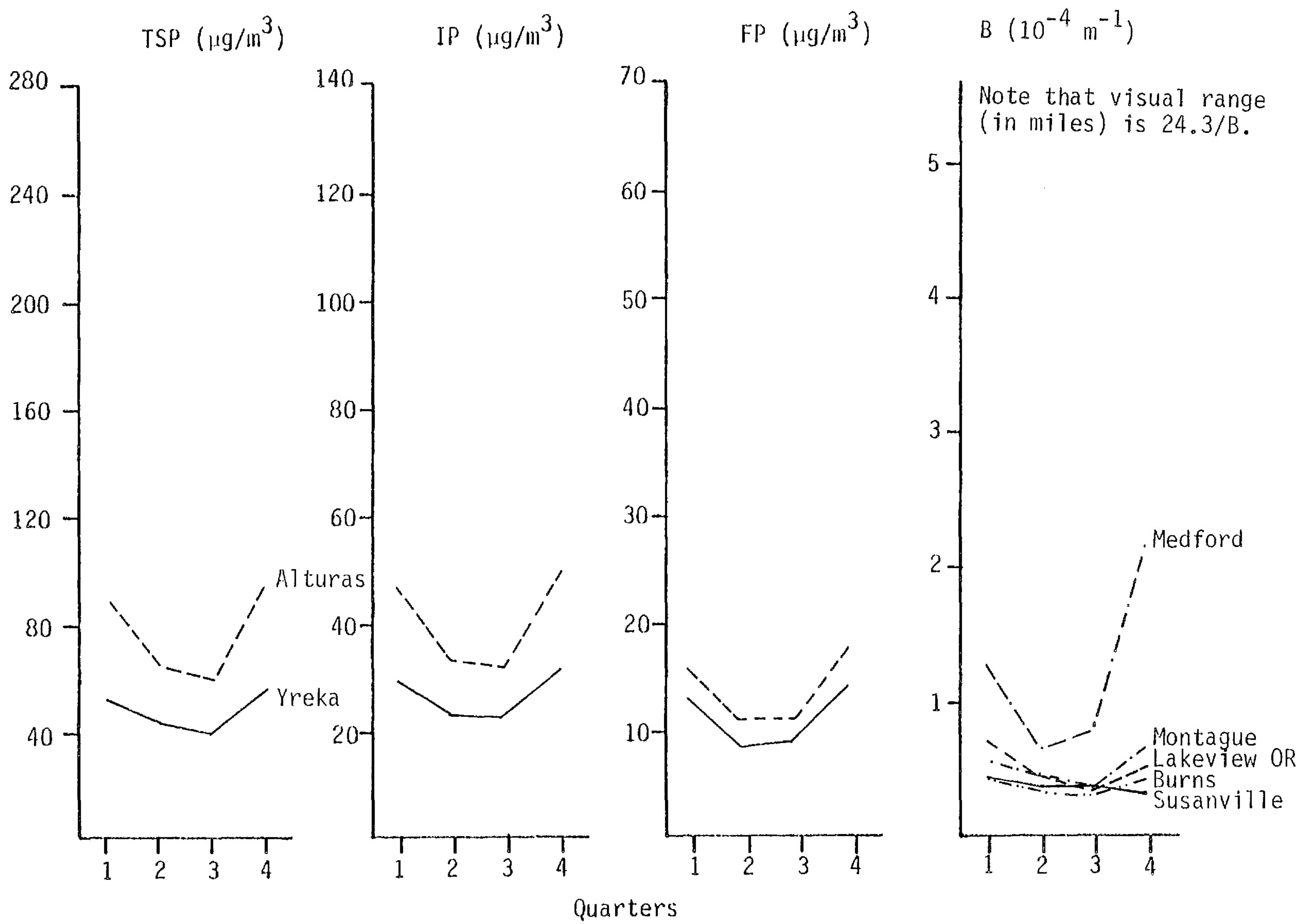


Figure 4.10b Northeast Plateau Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

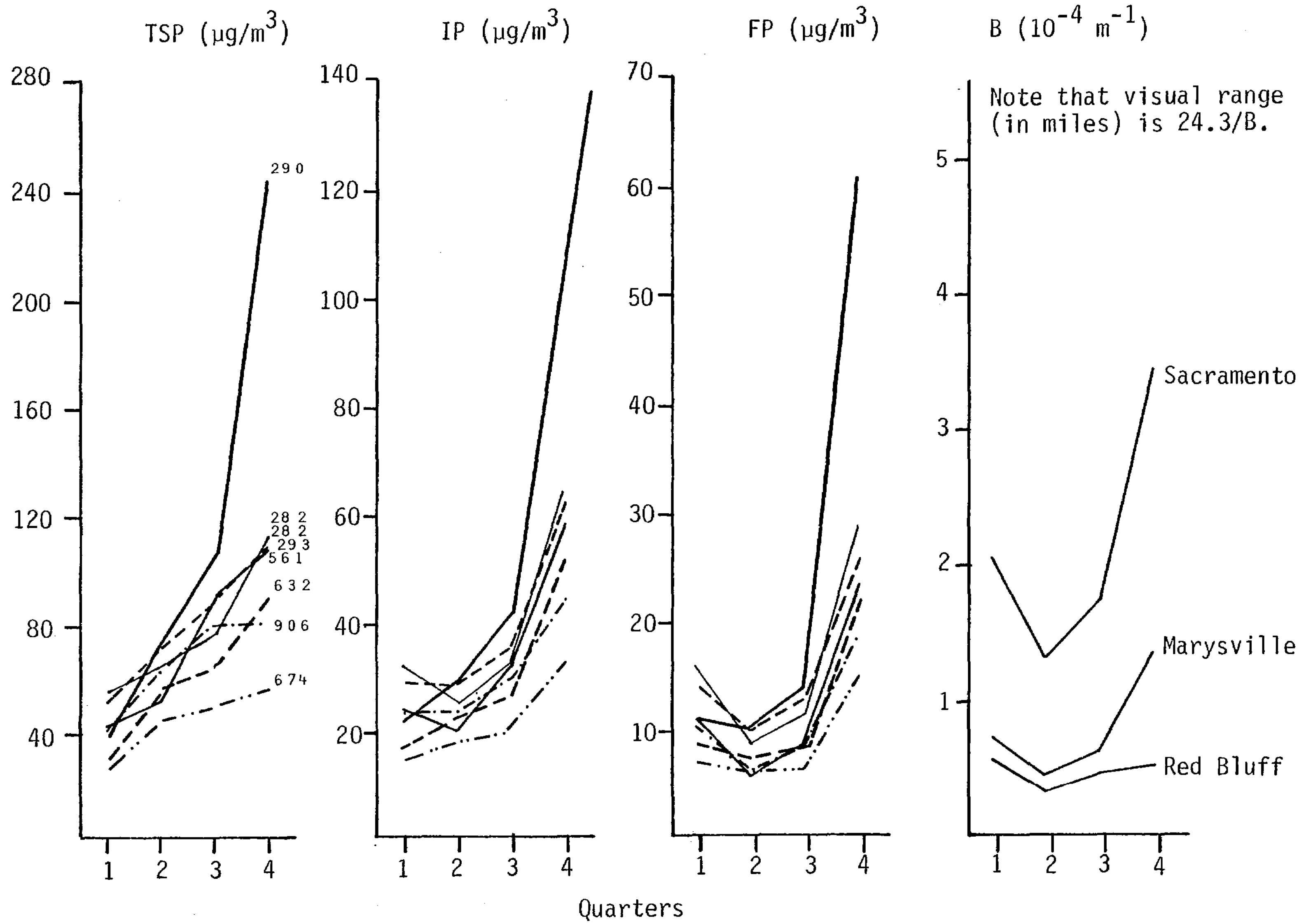


Figure 4.10c Sacramento Valley Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

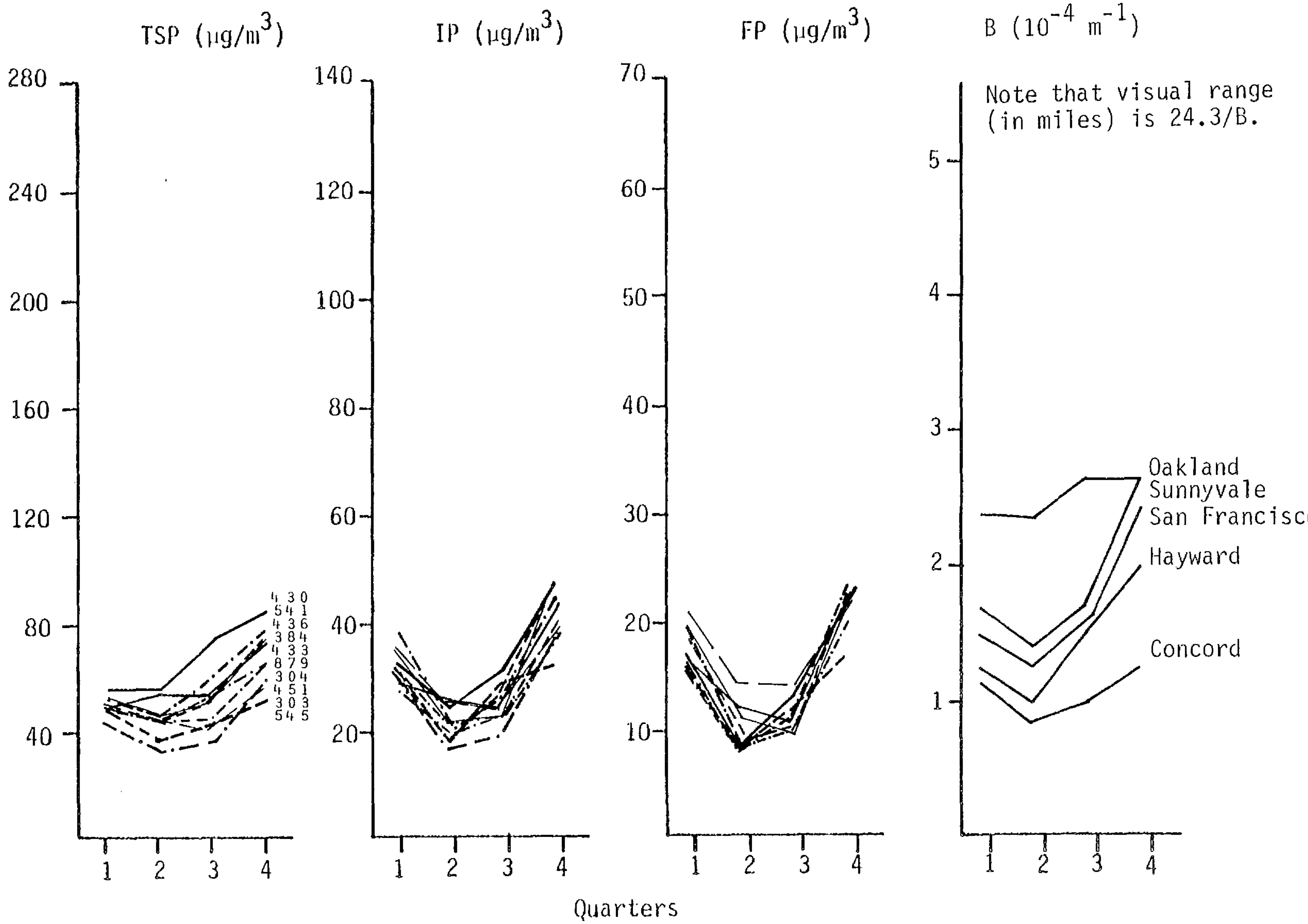


Figure 4.10d San Francisco Bay Area Air Basin (central).

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

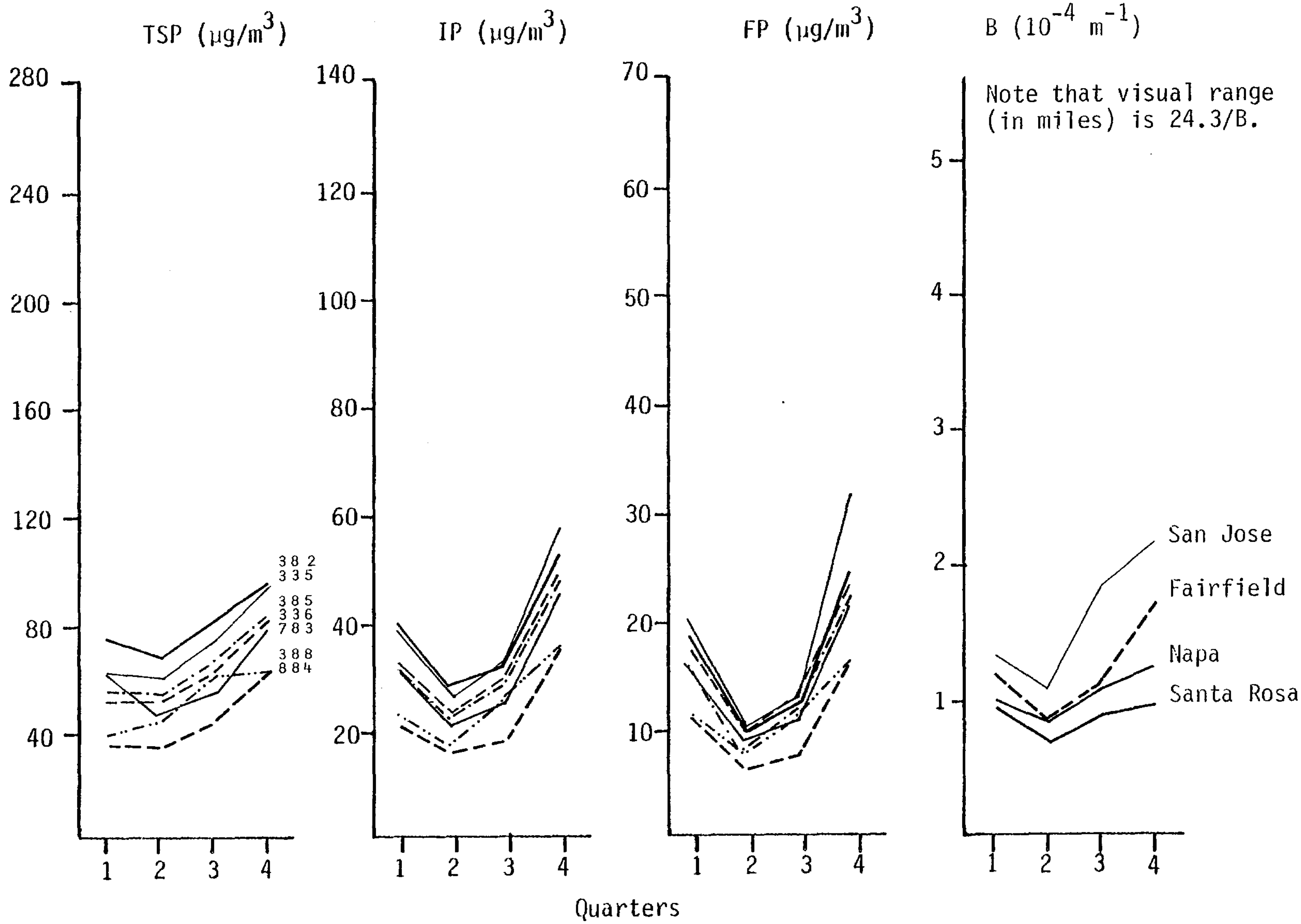


Figure 4.10e San Francisco Bay Area Air Basin (outlying).

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

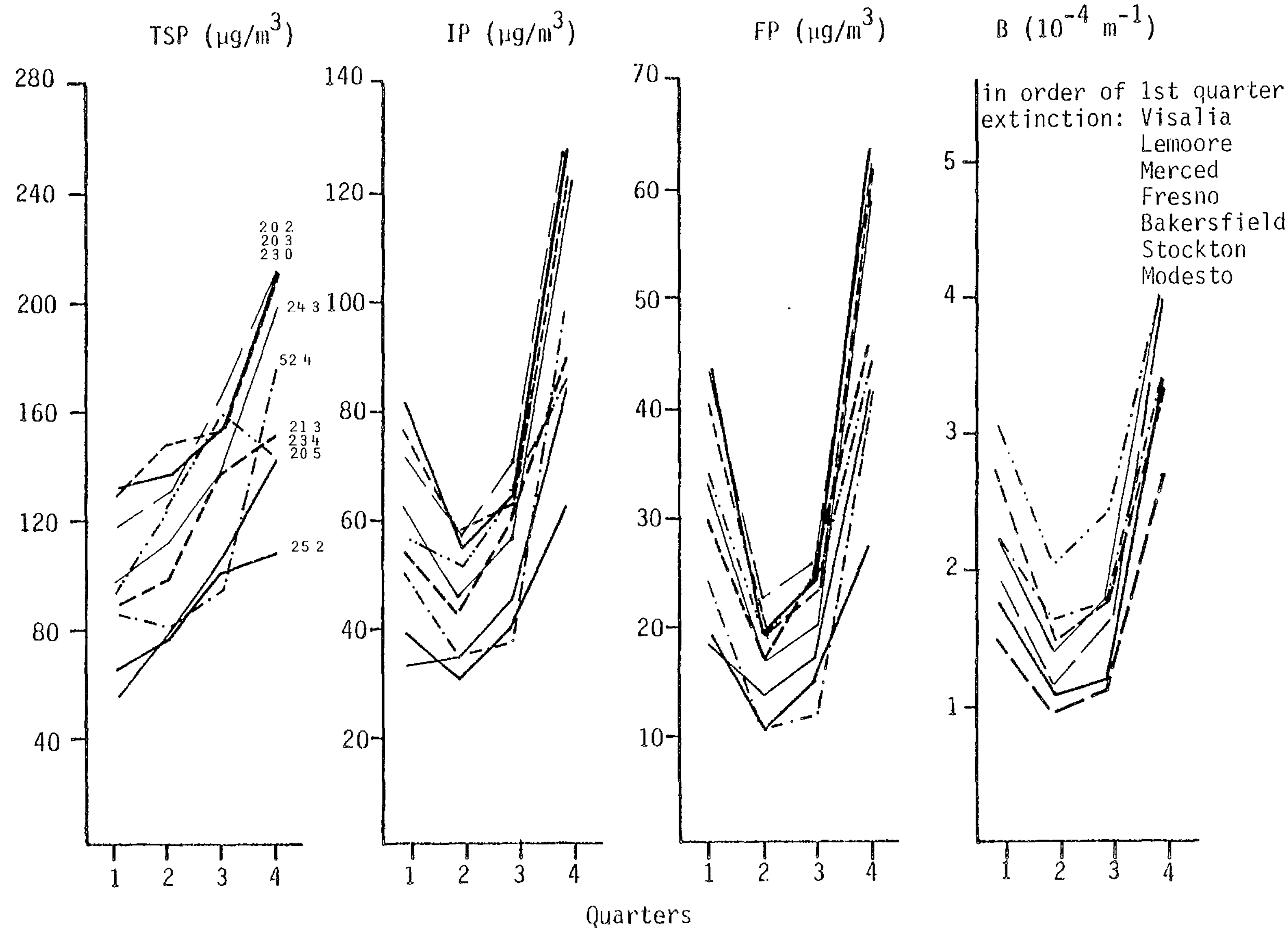


Figure 4.10f San Joaquin Valley Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

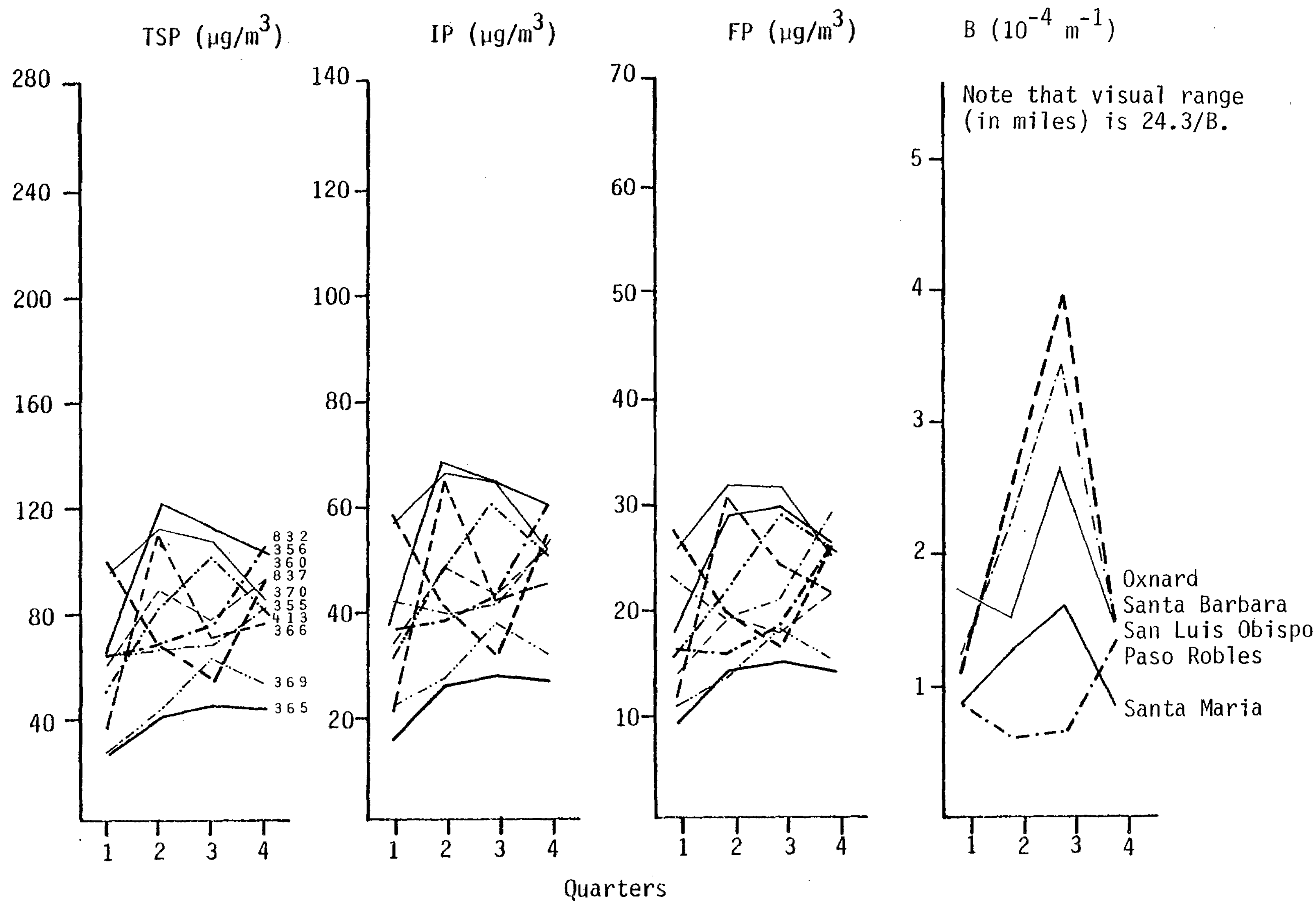


Figure 4.10g South Central Coast Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

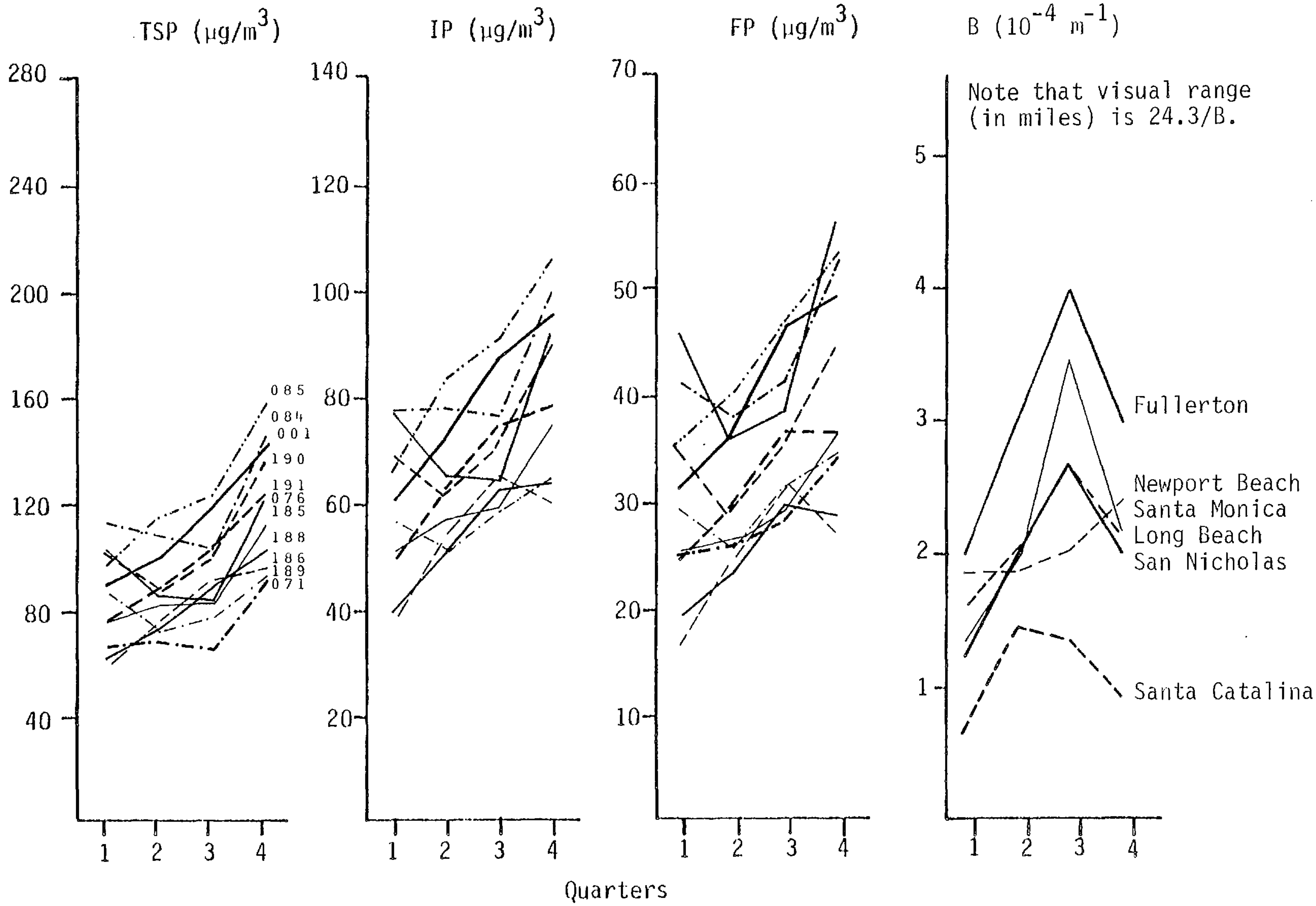


Figure 4.10h South Coast Air Basin (coastal part).

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

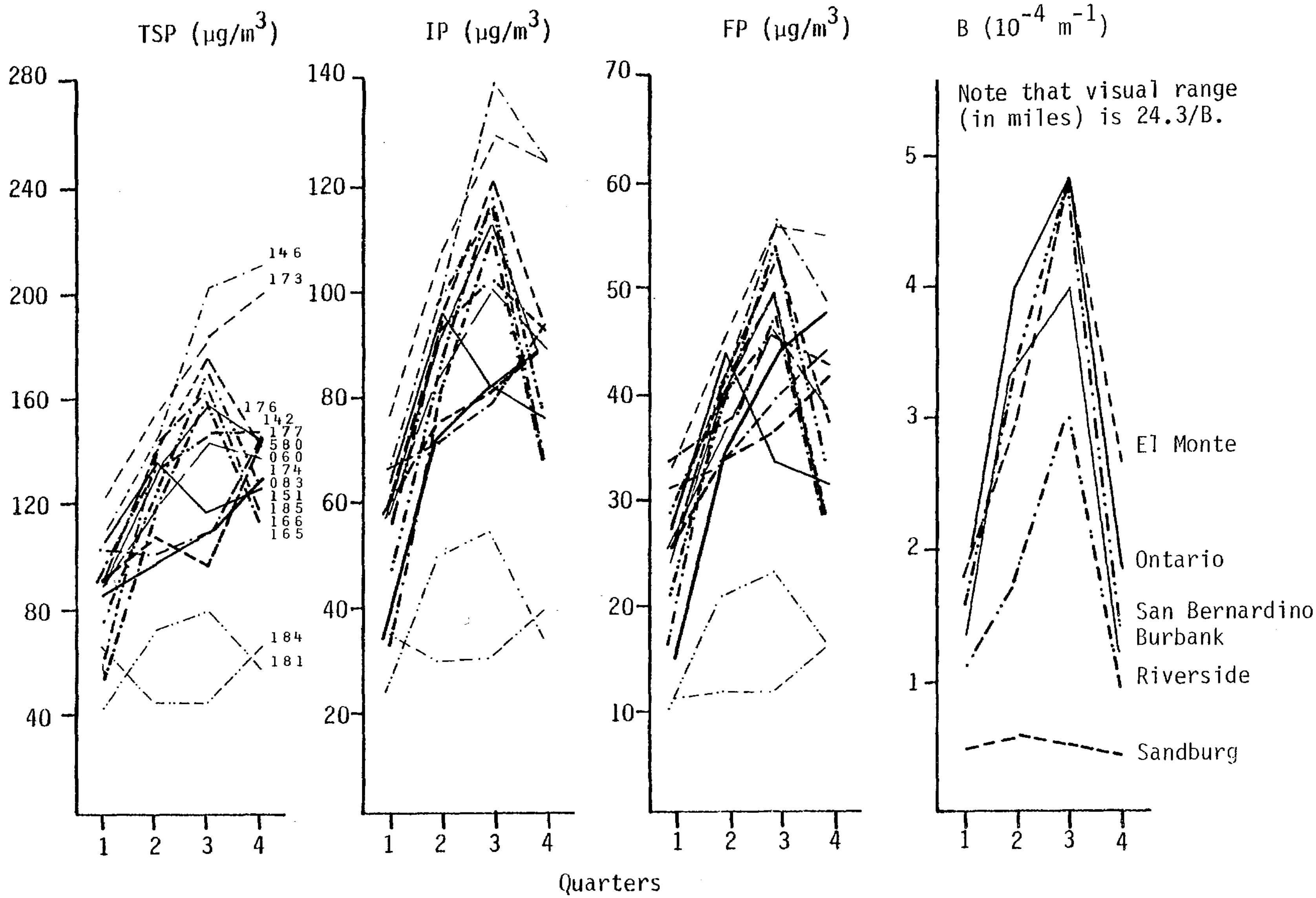


Figure 4.10i South Coast Air Basin (inland part).

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

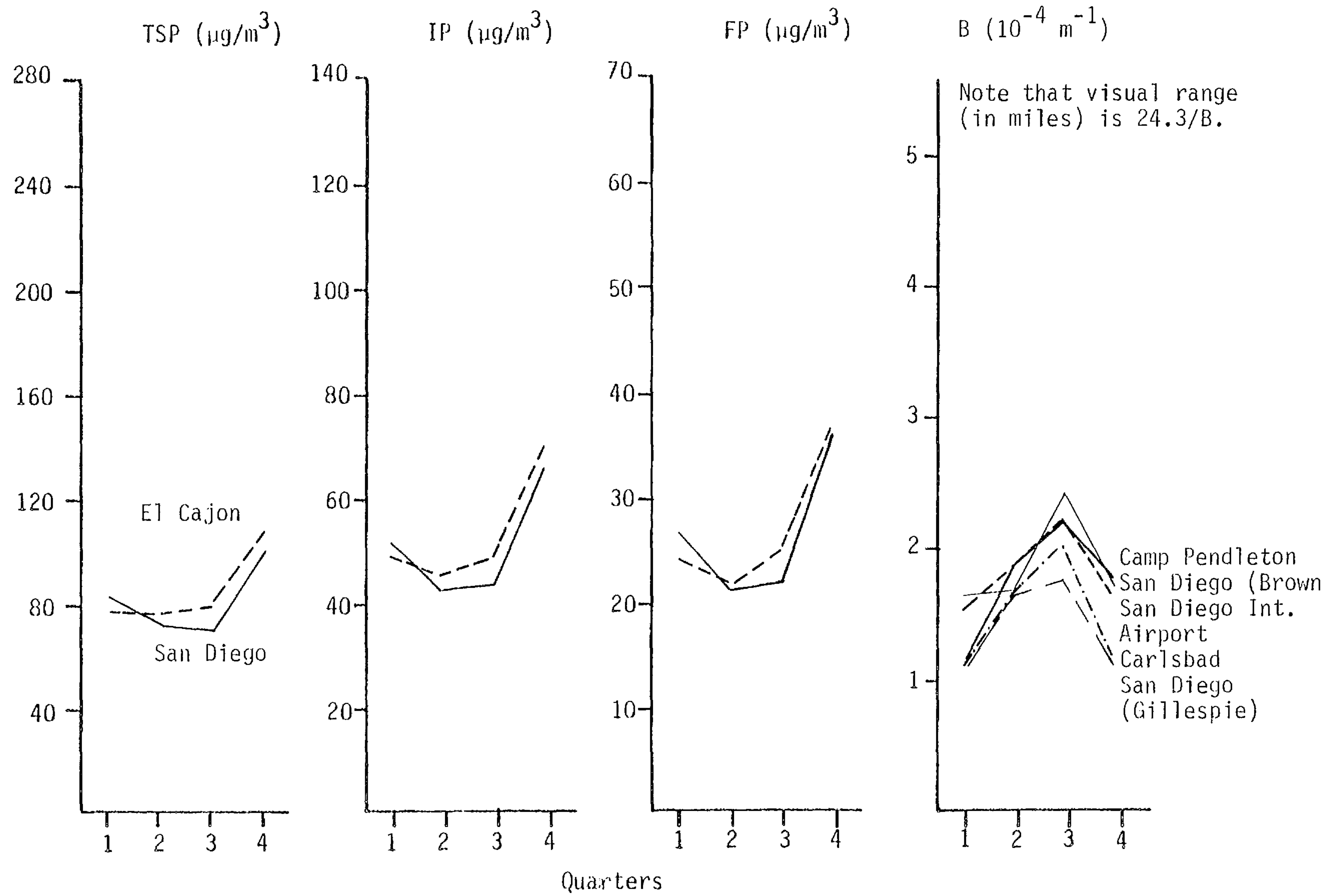


Figure 4.10j San Diego Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

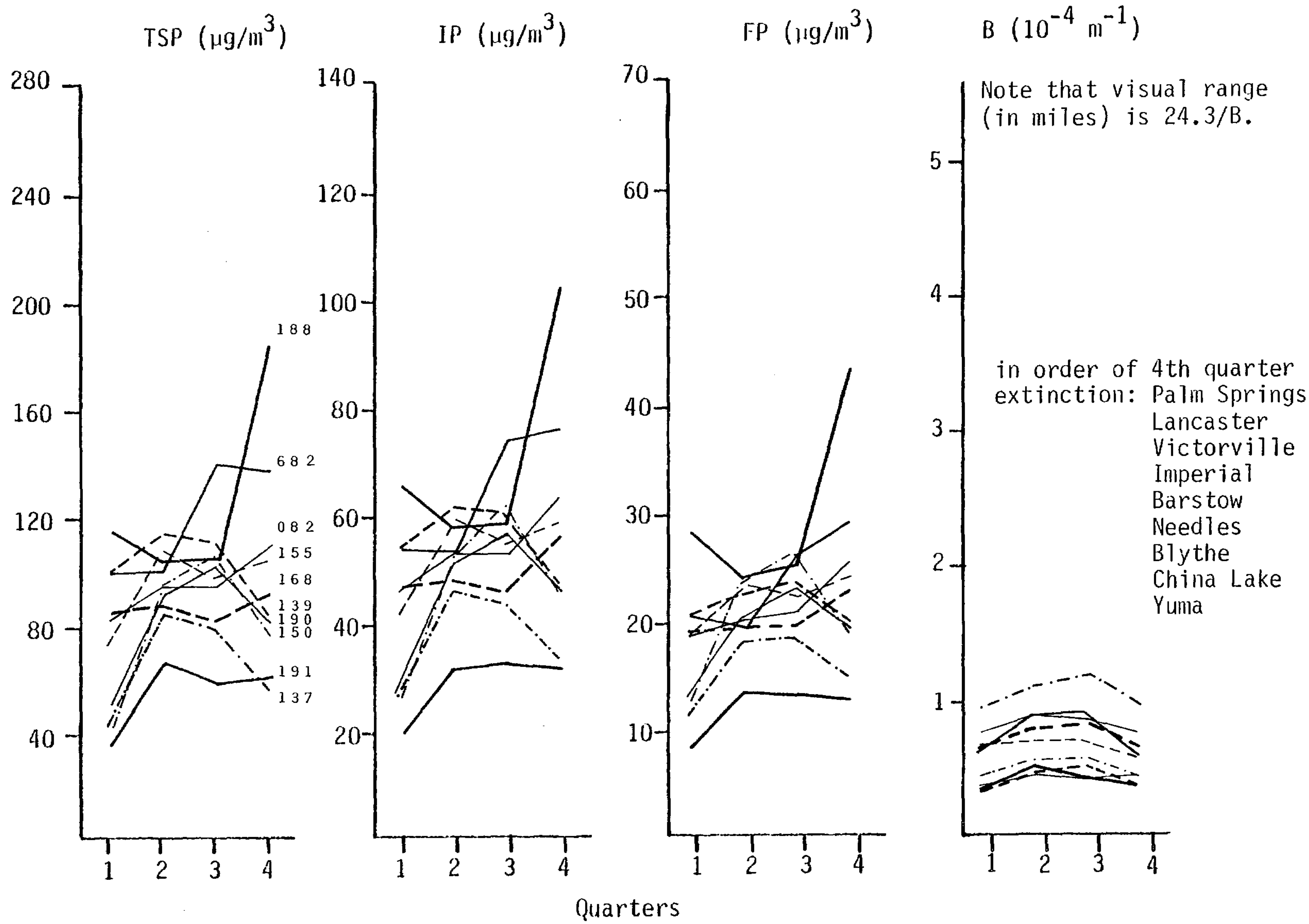


Figure 4.10k Southeast Desert Air Basin.

Figure 4.10 Seasonal patterns in TSP, estimated IP, estimated FP, and extinction (continued). Note that the errors for the quarterly values should be approximately 15% for IP and 20% for FP.

KEY TO SITE CODES FOR FIGURE 4.10

NORTH CENTRAL COAST AND LAKE COUNTY AIR BASINS

546 Bradley CDF Fire Station
544 Salinas II
545 San Ardo Water District Office
713 Lakeport Lakeport Boulevard

SACRAMENTO VALLEY AIR BASIN

282 Sacramento 1025 P Street
290 Sherman Island
293 Citrus Heights Sunrise Boulevard
561 Mountain Gate
906 Red Bluff Lincoln
632 Oroville 6 WNW
674 Willows 8 W

SAN FRANCISCO BAY AREA AIR BASIN (Central)

545 Burlingame Burlingame Ave.
436 Concord Treat Boulevard
430 Pittsburg
541 Redwood City
433 Richmond
303 San Francisco Ellis
304 San Francisco 23rd St.
451 San Rafael
384 Sunnyvale
879 Vallejo Tuolumne

SAN FRANCISCO BAY AREA AIR BASIN (Outlying)

439 Bethel Island
336 Fremont Chapel Way
385 Gilroy Monterey St.
335 Livermore Railroad
783 Napa Jefferson St.
388 Saratoga Hwy 85 & SPRR
382 San Jose 4th St.
884 Santa Rosa Humboldt St.

SAN JOAQUIN VALLEY AIR BASIN

202 Bakersfield H.D. Flower
203 Bakersfield Chester St.

SAN JOAQUIN VALLEY AIR BASIN (Con't.)

243 Fresno Herndon
205 Kern Refuge
234 McKittrick Fire Station
524 Merced 18th & S
230 Oildale
252 Stockton Hazelton St.
213 Taft North 10th St.

SOUTH CENTRAL COAST AIR BASIN

370 El Capitan Beach
360 Lompoc G St.
365 Lompoc Jalama Road
837 Morro Bay Jr. High School
832 Paso Robles
355 Santa Barbara
356 Santa Maria Library
366 Santa Maria Briarwood
369 Santa Ynez
413 Simi Valley

SOUTH COAST AIR BASIN (Coastal Part)

185 Costa Mesa (Harbor & Placentia)
186 El Toro
189 Laguna Beach
188 San Juan Capistrano
076 Lennox
190 Los Alamitos Orangewood
001 Los Angeles (Downtown & N. Main)
084 Lynwood
085 Pico Rivera
191 Santa Ana Weir Canyon Road
071 W. Los Angeles (and Robertson)

SOUTH COAST AIR BASIN (Inland Part)

060 Azusa
184 Big Bear Lake
173 Chino Riverside Avenue
185 Ontario Archibald Avenue
176 Fontana Foothill Trailer
166 Rialto Airport
177 La Habra
181 Lake Gregory

KEY TO SITE CODES FOR FIGURE 4.10 (Continued).

SOUTH COAST AIR BASIN
(Inland Part, Cont'd.)

083 Pasadena Walnut
580 Temple City
165 Redlands (and Univ. of)
142 Riverside (Trailer & Magnolia)
146 Riverside Rubidoux
151 San Bernardino
174 Upland (Civic Center & Post Office)

SOUTHEAST DESERT AIR BASIN

150 Banning Allesandro
155 Barstow
682 El Centro Broadway
139 Indio Oasis St.
082 Lancaster
137 Palm Springs
188 Trona Market St.
191 Twentynine Palms Adobe
168 Victorville
190 Victorville Fairgrounds

be dominated by the contribution from fine particles, specifically those particles in the 0.1 to 1.0 micron size mode.

- Trijonis (1980) has noted that seasonal patterns of visibility should most closely follow seasonal patterns in fine particle concentrations in those air basins where man-made visibility impacts are most severe, but that seasonal patterns in visibility may more closely follow seasonal patterns in meteorology (e.g. relative humidity) in those air basins with lesser air pollution levels. The results of Figure 4.10 are very consistent with this hypothesis. For example, the seasonal patterns of extinction and FP correspond very closely in those air basins where Trijonis (1980) found man-made impacts dominating visibility -- specifically, the Sacramento Valley, the San Joaquin Valley, and inland South Coast Air Basins, and to a partial extent the San Francisco Bay Area Air Basin. Seasonal patterns of FP and visibility do not correspond as well in some of the cleaner areas, e.g. the North Central Coast, Lake County, South Central Coast, San Diego, and Southeast Desert Air Basins.
- Noting that the vertical scales are the same in Figures 4.10a through 4.10k, one can see a fairly obvious correlation among the air basins in overall extinction and fine particle levels. Extinction and fine particle levels are generally the highest in the South Coast and San Joaquin Valley Air Basins.

The following more specific comments regarding seasonal patterns are also noteworthy:

- The Northeast Plateau, Sacramento Valley, San Francisco Bay Area, and San Joaquin Valley Air Basins tend to experience their highest FP and extinction levels in the fourth (fall) quarter. The first (winter) quarter tends to be the second highest season for FP and extinction, while the best air quality occurs during the spring (or sometimes during the summer) quarter. Most remarkably, 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. Trijonis (1980) attributed the fall maximum of FP and extinction in the San Joaquin Valley to very high sulfate and nitrate concentrations during that season; the high sulfate and nitrate concentrations are promoted by stagnant air and elevated relative humidity during the fall.

- The South Central Coast, South Coast, and (to a lesser extent) Southeast Desert Air Basins tend to undergo maximum FP and extinction levels during the third (summer) quarter. Extinction and FP are also relatively high in the second (spring) quarter but reach relative minimums during the first and fourth quarters. The worst air quality occurs in the valleys and eastern inland areas of the South Coast Air Basin during the summer, when FP averages 40 to 60 $\mu g/m^3$ and visibility averages 5 to 6 miles. Trijonis (1980) attributed this summer maximum of FP and extinction in Los Angeles to high levels of sulfates and other photochemical aerosols promoted by the intense sunshine, stronger inversion ceilings, and greater inland penetration of moist air during the summer.

5. REFERENCES

Appel, B.R., S.M. Wall, Y. Tokiwa, and M. Haik, "Interference Effects in Sampling Particulate Nitrate in Ambient Air," Atmospheric Environment, 13, pp. 319-325, 1979.

ARB, "Revision to the State of California Implementation Plan -- Chapter 2. Statewide Perspective," California Air Resources Board, October 1978.

Bell, G.B., "The Uses of Meteorological Data in Large-Scale Air Pollution Surveys," SRI Project No. SU-2238, Prepared for State of California Department of Public Health, Bureau of Air Sanitation, Berkeley, CA 1958.

Coutant, R.W., "Effect of Environmental Variables on Collection of Atmospheric Sulfate," Environmental Science and Technology, 11, pp. 873-878.

Delaware, W., Personal Communication, New York State Department of Environmental Conservation, Albany, NY, 1981.

Dzubay, T.G., "Chemical Element Balance Method Applied to Dichotomous Sampler Data," Annals of the New York Academy of Science, pp. 126-144, 1980.

EPA, "Position Paper on the Regulation of Atmospheric Sulfates," EPA-450/2-75-007, EPA Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina, 1975.

EPA, Protecting Visibility, EPA-450/5-79-008, EPA Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina, 1979.

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 $\leq$ 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.

Harker, A.B., L.W. Richards, and W.E. Clark, "The Effect of Atmospheric SO_2 Photochemistry Upon Observed Nitrate Concentrations in Aerosols," Atmospheric Environment, 11, pp. 87-91, 1977.

Lioy, P.J., J.G. Watson, and J.D. Spengler, "APCA Specialty Conference Workshop on Baseline Data for Inhalable Particulate Matter," J. of the Air Pollution Control Association, 30, pp. 1126-1130, 1980.

Lundgren, D.A., "Atmospheric Aerosol Composition and Concentration as a Function of Particle Size and Time," J. of the Air Pollution Control Association, 20, pp. 603-608, 1970.

Kolak, N.P., J. Hyde, and R. Forrester, "Particulate Source Contributions in the Niagara Frontier," EPA-902/4-79-006, 1979.

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

NOAA, Climatic Atlas of the United States, National Oceanic and Atmospheric Administration, National Climatic Center, Asheville, North Carolina, 1977.

Rodes, C., Personal Communication, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1981.

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.

Spengler, J.D., Personal Communication, Harvard School of Public Health, Cambridge, MA, 1980.

Spicer, C.W. and P.M. Schumacher, "Particulate Nitrate: Laboratory and Field Studies of Major Sampling Interferences," Atmospheric Environment, 13, pp. 543-552, 1979.

Stevens, R.K., T.G. Dzubay, R.W. Shaw, W.A. McClenney, C.W. Lewis, and W.E. Wilson, "Characterization of the Aersol in the Great Smoky Mountains," Environmental Science and Technology, 14, pp. 1491-1498, 1980.

Stevens, R.K., T.G. Dzubay, G. Russwurm, and D. Rickel, "Sampling and Analysis of Atmospheric Sulfates and Related Species," Atmospheric Environment, 12, pp. 55-68, 1978.

Tanner, R.L., J. Forrest, and L. Newman, "Determination of Atmospheric Gaseous and Particulate Sulfur Compounds," Sulfur in the Environment, J.O. Niagu, Ed., John Wiley and Sons, Inc., New York, NY 1978.

Tanner, R.L., "Sulfur and Nitrogen Compounds in Urban Aerosols," Presented at the New York Academy of Science Conference on Aerosols, New York, 1979.

Trijonis, J. and K. Yuan, "Visibility in the Northeast: Long-Term Visibility Trends and Visibility/Pollutant Relationships," EPA-600/3-78-075, 1978.

Trijonis, J. and D. Shapland, "Existing Visibility Levels in the United States," EPA-450/5-79-010, 1979.

Trijonis, J., "Visibility in the Southwest: An Exploration of the Historical Data Base," Atmospheric Environment, 13, pp. 833-843, 1979.

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.

Whitby, K.T. and G.M. Sverdrup, "California Aerosols: Their Physical and Chemical Characteristics," Particle Technology Laboratory Publication No. 347, University of Minnesota, Minneapolis, Minnesota, 1978.

White, W.H. and P.T. Roberts, "On the Nature and Origins of Visibility-Reducing Aerosols in the Los Angeles Air Basin," Atmospheric Environment, 11, p. 803, 1977.

APPENDIX A

ESTIMATING PRIMARY VEHICULAR PARTICULATE CONCENTRATIONS FROM LEAD CONCENTRATIONS

In the hybrid models for predicting IP and FP, lead is used as a tracer for the contribution from primary vehicular particulate emissions. The lead concentration, [Pb], measured on a Hi-Vol is assumed to come from "suspendable" ($\lesssim$ 20 μm in size) Pb aerosols emitted from vehicles. In order to estimate total inhalable (or total fine) aerosol concentrations from vehicles, we need to multiply [Pb] by the ratio:

$$R_{ip} = \frac{\text{total inhalable vehicular particulate emissions}}{\text{suspendable lead vehicular particulate emissions}}$$

or

$$R_{fp} = \frac{\text{total fine vehicular particulate emissions}}{\text{suspendable lead vehicular particulate emissions}}$$

The emission terms in these ratios depend on the year because of changes in the number of catalytic converter cars and because of variance in the fraction of lead in leaded gasoline. We have estimated the emission terms on a national basis for two years, 1972 and 1980. The ratios were approximately constant before 1975 and changed in an approximately linear way after 1975; thus, knowing the 1972 and 1980 ratios allows us to estimate the ratio for other years as well.

Tables A-1 and A-2 present nationwide estimates of inhalable and fine particulate emissions from vehicles in 1972 and 1980, respectively. The emissions include tire wear, brake wear, and diesel exhaust as well as gasoline exhaust. As indicated by the footnotes, each of these tables is based on a very thorough review of the relevant literature.

Based on data compiled by Ter Haar et al. (1972), Habibi (1973), Trijonis et al. (1974), Cass et al. (1981), and Pierson (1981), the national suspended lead emission factors for leaded gasoline vehicles is assumed to be .025{Pb} in [grams/mile], where {Pb} is the concentration of lead in leaded gasoline in units of [grams/gallon]. In 1972, with 96% of the 1.25 x 10^{12} miles per year of traffic accounted for by leaded gasoline vehicles,

TABLE A-1 NATIONAL ESTIMATES OF INHALABLE AND FINE PARTICULATE EMISSIONS FROM VEHICLES IN 1972.

	INHALABLE EMISSIONS		FINE EMISSIONS	
	IP Emission Factor (g/mi)	Nationwide IP Emissions* (Metric tons per year)	FP Emission Factor (g/mi)	Nationwide FP Emissions* (Metric tons per year)
Non-Catalyst Vehicles[a]	0.23	276,000	0.18	216,000
Diesels[b]	1.8	90,000	1.5	75,000
Tire Wear[c]	0.04	50,000	0.02	25,000
Brake Wear[d]	0.02	25,000	0.01	12,000
TOTAL		441,000		328,000

(a) Based on data and calculations from Ter Haar et al. (1972), Habibi (1973), Trijonis et al. (1974), and Cass et al. (1981). A value of 2.2 gm/gal of Pb in gasoline is assumed based on data in BOM (1972). The size distribution is based on Ter Haar et al. (1972), Habibi (1973), and Pierson (1981).

(b) Based on data reported by Baines et al. (1979) and Pierson (1978, 1981). The size distribution is based on Taback et al. (1979) and Pierson (1981).

(c) Based on data reported by Subramini (1971), Pierson and Brachaczek (1974), and Pierson (1981). The size distribution is based on Pierson (1981).

(d) Based on data reported by Anderson et al. (1973), Jacko et al. (1973), and Taback et al. (1979).

* Assuming 1.25×10^{12} miles per year with a mix of 96% gasoline vehicles and 4% diesels.

TABLE A-2 NATIONAL ESTIMATES OF INHALABLE AND FINE PARTICULATE EMISSIONS FROM VEHICLES IN 1980.

	INHALABLE EMISSIONS		FINE EMISSIONS	
	IP Emission Factor (g/mi)	Nationwide IP Emissions* (Metric tons per year)	FP Emission Factor (g/mi)	Nationwide FP Emissions* (Metric tons per year)
Non-Catalyst Gasoline Vehicles[a]	0.20	120,000	0.16	96,000
Catalyst Gasoline Vehicles[b]	0.014	12,000	0.013	11,000
Diesels[c]	1.8	108,000	1.5	90,000
Tire Wear[d]	0.04	60,000	0.02	30,000
Brake Wear[e]	0.02	30,000	0.01	15,000
TOTAL		330,000		242,00

(a) Based on data and calculations from Ter Haar et al. (1972), Habibi (1973), Trijonis et al. (1974), and Cass et al. (1981). A value of 1.5 gm/gal of Pb in leaded fuel is assumed based on data in DOE (1980). The size distribution is based on Ter Haar et al. (1972), Habibi (1973), and Pierson (1981).

(b) Based on data reported by Laresgoiti and Springer (1977), Muhlbaier and Williams (1981), and Pierson (1981). The size distribution is based on Pierson (1981), Cass et al. (1981), and Miller et al. (1976).

(c) Based on data reported by Baines et al. (1979) and Pierson (1978, 1981). The size distribution is based on Taback et al. (1979) and Pierson (1981).

(d) Based on data reported by Subramini (1971), Pierson and Brachaczek (1974), Pierson (1981). The size distribution is based on Pierson (1981).

(e) Based on data reported by Anderson et al. (1973), Jacko et al. (1973), and Taback et al. (1979).

* Assuming 1.5×10^{12} miles per year with a mix of 56% catalyst vehicles, 40% non-catalyst gasoline powered vehicles and 4% diesels.

and with 2.2 gm/gal of lead in leaded gasoline, national suspendable lead emissions were

$$.025 \times (2.2) \times (0.96) \times (1.25 \times 10^{12}) \times \frac{1 \text{ metric ton}}{10^6 \text{ gm}} = 66{,}000 \text{ metric tons/year}$$

In 1980, with 40% of the 1.5×10^{12} miles per year of traffic accounted for by leaded gasoline vehicles, and with 1.5 gm/gal of lead in leaded gasoline, national suspendable lead emissions were

$$.025 \times (1.5) \times (0.40) \times (1.5 \times 10^{12}) \times \frac{1 \text{ metric ton}}{10^6 \text{ gm}} = 22{,}500 \text{ metric tons/year}$$

Based on the emission data in Table A-1, Table A-2, and the preceeding paragraph, the appropriate ratios are

	1972	1980
R_{ip}	7	15
R_{fp}	5	11

The historical changes in the ratio are approximately as shown in Figure A.1.

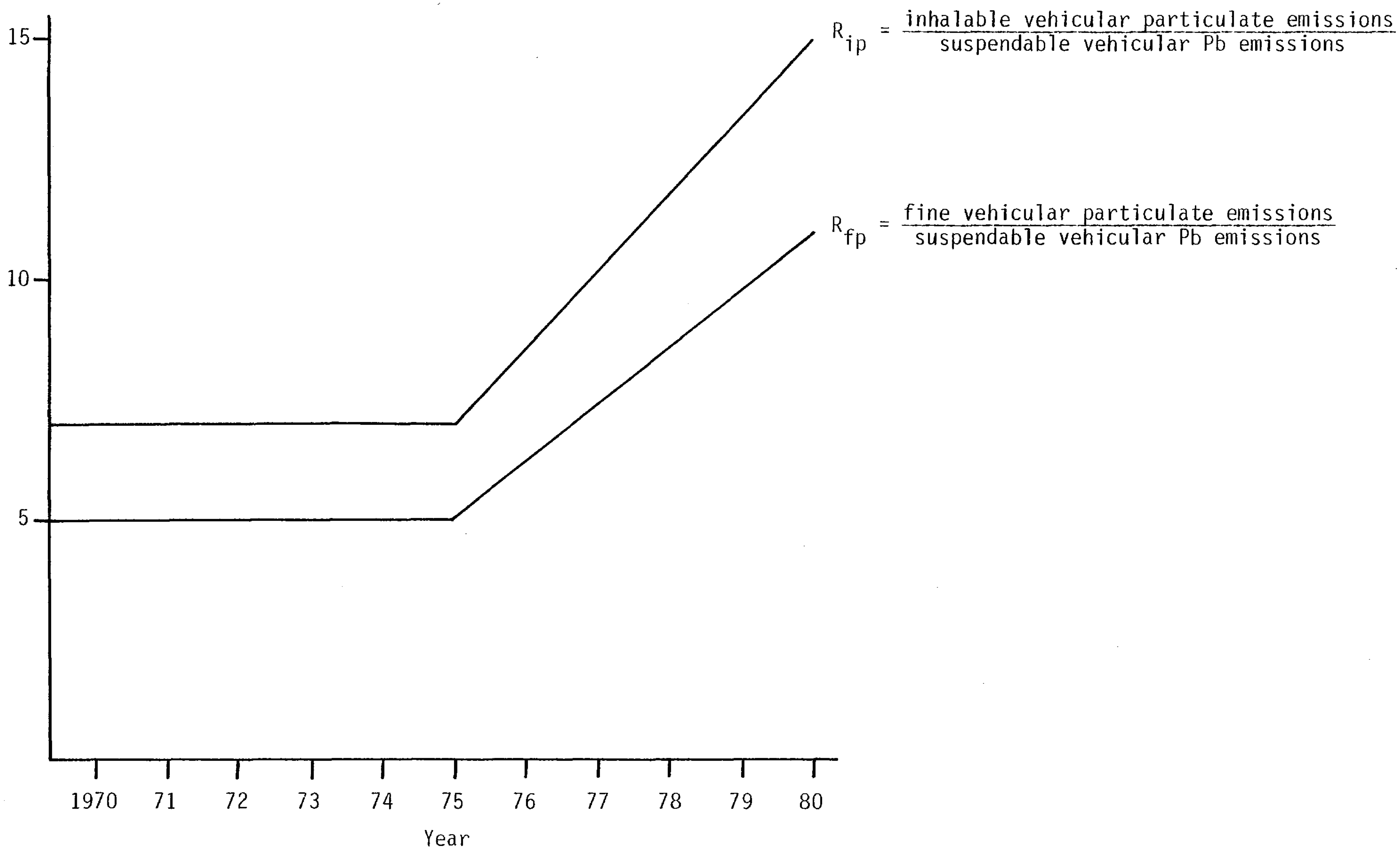


Figure A.1 The ratio of inhalable vehicular particulate emissions and fine vehicular particulate emissions to suspendable vehicular lead emissions.

REFERENCES FOR APPENDIX A

Anderson, A.E., R.L. Gealer, R.C. McCune, and J.W. Sprys, "Asbestos Emissions from Brake Dynamometer Tests," Presented at the Automobile Engineering Meeting, Detroit, Michigan, May 14-18, 1973.

Baines, T.M., J.H. Somers, and C.A. Harvey, "Heavy Duty Diesel Particulate Emission Factors," J. Air Pollution Control Association, Vol. 29, No. 6, pp. 616-621, June 1979.

BOM (Bureau of Mines), Motor Gasolines, Winter 1971-1972, U.S. Department of the Interior, Mineral Industry Surveys, Petroleum Products Surveys No. 75, 1972.

Cass, G., P.M. Boone, and E.S. Macias, "Emissions and Air Quality Relationships for Atmospheric Carbon Particles in Los Angeles," Internal Working Paper, Environmental Quality Laboratory, California Institute of Technology, Pasadena, California, 1981.

DOE (Department of Energy), "Mineral Industry Surveys, Petroleum Products Surveys, 1971-1980."

Habibi, K., "Characterization of Particulate Matter in Vehicular Exhaust," Environmental Science and Technology, Vol. 7, No. 3, pp. 223-234, March 1973.

Jacko, M.G., R.T. DuCharme, and J.H. Somers, "Brake and Clutch Emissions Generated During Vehicle Operation," Presented at the Automobile Engineering Meeting, Detroit, Michigan, May 14-18, 1973.

Laresgoiti, A. and G.S. Springer, Environmental Science and Technology, Volume II, pp. 285-292, 1977.

Miller, D.F., D.A. Trayser, and D.W. Joseph, "Size Characterization of Sulfuric Acid Aerosol Emissions," Presented at the Automotive Engineering Congress and Exposition, Detroit, Michigan, February 23-27, 1976.

Muhlbaier, J.L. and R.L. Williams, "Fireplaces, Furnaces and Vehicles as Emission Sources of Particulate Carbon," in G.T. Wolff and R.L. Klimisch, eds., Particulate Carbon: Atmospheric Life Cycle, Plenum Press, New York, 1981.

Pierson, Wm., Particulate Organic Matter and Total Carbon from Vehicles on the Road," in Proceedings: Carbonaceous Particles in the Atmosphere, Lawrence Berkeley Laboratory, University of California, June 1979.

Pierson, Wm., Scientific Laboratory, Ford Motor Company, Dearborn, Michigan, Personal Communication, 1981.

Pierson, Wm. and W.W. Brachaczek, "Airborne Particulate Debris from Rubber Tires", in Rubber Chemistry and Technology, Vol. 47, No. 5, December 1974.

Subramini, J.P., "Particulate Air Pollution from Automobile Tire Tread Wear," University of Cincinnati, thesis, 1971.

Taback, H.J., A.R. Brienza, J. Macko, and N. Brunetz, "Fine Particle Emissions from Stationary and Miscellaneous Sources in the South Coast Air Basin," KVB Inc., document No. KVB 5806-783, Tustin, California, 1979.

Ter Haar, G.L., D.L. Lenane, J.N. Hu, and M. Brandt, "Composition, Size and Control of Automotive Exhaust Particles," J. Air Pollution Control Association, Vol. 22, No. 1, January 1972.

Trijonis, J.C., G. Richard, K. Crawford, and R. Wada, "An Implementation Plan for Suspended Particulate Matter in the Los Angeles Region," prepared for EPA under Contract 68-02-1384, March 1975.

APPENDIX B

LIST OF SITES CONTAINING SIMULTANEOUS DATA FOR SSIP, TSP, $SO_4^=$, AND Pb.

SITE	NUMBER OF DATA POINTS
CALIFORNIA-SAN FRANCISCO BAY AREA	
Richmond	9
CALIFORNIA-CENTRAL VALLEY	
Bakersfield	10
Citrus Heights	10
Fresno, 1st and Olive	6
Fresno County	5
Fresno, Oliver Street	1
CALIFORNIA-LOS ANGELES AREA	
Azusa	1
Pasadena	10
Riverside	60
Rubidoux	20
West Los Angeles	6
CALIFORNIA-OTHER	
El Centro	9
PACIFIC NORTHWEST	
Deschutes County OR	1
Portland OR	8
ARID SOUTHWEST	
Maricopa County AZ	4
Denver CO	4
Lakewood CO	4
Magna UT	2
Salt Lake City UT	1
NORTH CENTRAL	
Will County IL	1
Topeka KS	3
Minneapolis MN	4
Afton MO	2
Kansas City MO	10

SITE	NUMBER OF DATA POINTS
NORTHEAST	
Hartford CT	10
Washington D.C. (017)	1
Washington D.C. (019)	2
Acadia National Park ME	6
Boston MA	8
Detroit MI (015)	1
Detroit MI (020)	1
Lackawanna NY	3
Akron Summit OH	10
Cincinnati OH	6
Cleveland OH (013)	2
Cleveland OH (041)	3
Medina OH	5
Middletown OH (006)	30
Middletown OH (007)	32
Middletown OH (009)	31
Middletown OH (010)	29
Steubenville OH	8
North Braddock PA	4
Philadelphia PA (003)	131
Philadelphia PA (019)	38
Philadelphia PA (023)	1
Philadelphia PA (024)	5
Philadelphia PA (032)	3
Philadelphia PA (036)	5
Philadelphia PA (038)	47
Philadelphia PA (040)	25
Philadelphia PA (041)	35
Arlington VA	1
Hopewill VA	1
Reston VA	1
Richmond VA	2
SOUTHEAST	
Birmingham AL	6
Centerpoint AL	11
Tarrant AL	8
Atlanta GA	2
Durham NC	13
Dallas TX	12
El Paso TX	5
Houston TX	6

APPENDIX C

LIST OF SITES CONTAINING SIMULTANEOUS DATA FOR SSIP AND TSP.

SITE	SITE TYPE	ANNUAL DATA SET	NUMBER OF DATA POINTS
CALIFORNIA-SAN FRANCISCO BAY AREA			
Richmond	M	•	37
CALIFORNIA-CENTRAL VALLEY			
Bakersfield	S		18
Chino	N		3
Citrus Heights	S		10
Fresno, 1st and Olive	S		6
Fresno County	S	•	26
Fresno, Oliver Street	S		11
CALIFORNIA-LOS ANGELES AREA			
Azusa	M		20
Pasadena	M		14
Riverside	M		81
Rubidoux	M		41
West Los Angeles (500)	M		15
West Los Angeles (541)	M	•	35
CALIFORNIA-OTHER			
El Centro	N		15
Lompoc	N		18
PACIFIC NORTHWEST			
Deschutes County OR	N		10
Eugene OR	S		13
Portland OR	M	•	29
Spokane WA	S		2
ARID SOUTHWEST			
Phoenix AZ	M		3
Maricopa County AZ	M		20
Denver CO	M		18
Lakewood CO	M		8
Reno NV	S		8
Albuquerque NM	S		2
Magna UT	S		18
Salt Lake City UT	M		17

SITE	SITE TYPE	ANNUAL DATA SET	NUMBER OF DATA POINTS
NORTH CENTRAL			
Will County IL	N		2
Indianapolis IN	M		6
Jeffersonville IN	S		2
Marshalltown IA (003)	N		14
Marshalltown IA (004)	N		8
Topeka KS	S		22
Duluth MN	S		9
International Falls MN	N		8
Minneapolis MN	M		26
St. Paul MN	M		10
Afton MO	M		10
Kansas City MO	M		30
St. Louis MO	M		3
Omaha NE	S		17
NORTHEAST			
Hartford CT	M	•	36
New Castle DE	S		11
Washington D.C. (017)	M		9
Washington D.C. (019)	M		17
Ashland KY	S		7
Acadia National Park ME	N	•	25
Baltimore MD (001)	M		4
Baltimore MC (009)	M		1
Boston MA	M	•	23
Detroit MI (015)	M		9
Detroit MI (020)	M		6
Camden NJ	M		11
Jersey City NJ	M		7
Lackawanna NY	M		18
Akron Summit OH	M	•	47
Cincinnati OH	M		24
Cleveland OH (013)	M		9
Cleveland OH (041)	M		12
Columbus OH	M		1
Medina OH	S		14
Middletown OH (006)	S		49
Middletown OH (007)	S		53
Middletown OH (009)	S		50
Middletown OH (010)	S		55
Steubenville OH	S	•	36

SITE	SITE TYPE	ANNUAL DATA SET	NUMBER OF DATA POINTS
Youngstown OH	S		6
Avalon PA	S		13
Bethlehem PA	S		7
North Braddock PA	S		22
Philadelphia PA (003)	M		158
Philadelphia PA (019)	M		110
Philadelphia PA (023)	M		24
Philadelphia PA (024)	M		44
Philadelphia PA (032)	M		30
Philadelphia PA (036)	M		41
Philadelphia PA (038)	M		133
Philadelphia PA (040)	M		34
Philadelphia PA (041)	M		37
Arlington VA	M		9
Hampton VA	S		13
Hopewell VA	S		16
Norfolk VA	S		11
Reston VA	S		6
Richmond VA	S		14
Wheeling WV	S		12
SOUTHEAST			
Birmingham AL	M		37
Centerpoint AL	S		28
Tarrant AL	M		33
Atlanta GA	M		5
Durham NC (006)	S		4
Durham NC (101)	N		36
Chattanooga TN	S		7
Nashville TN	M		4
Dallas TX	M		30
El Paso TX	M		23
Houston TX	M		13