STATISTICAL MODELING OF TSP DATA

SEPTEMBER, 1978

Statistical Modeling of TSP Data

Don Feeney

9/7/78

Seasonal periodicities, particularly relating to snow cover, are highly likely, and eighteen months of study is not long enough to establish periodicities of this length. Diurnal cycles can also not be established as the readings are 24 hour averages. In addition, the data set contains a large number of missing observations, while spectral analysis works only with a relatively complete set of observations. Finally, to apply spectral analysis it is necessary to assume that the series is stationary; that aside from regular periodicities, there are no long-term trends in the data (Koopmans, 1974). The mine strike would seem to invalidate the assumption. For these reasons, spectral analysis was not employed.

The approach used is the development of a linear statistical model. This model takes the form

$$Yij = \mu + s_i + d_j + e_{ij}$$

where Y = observed TSP concentration at site i and time j

μ = the overal mean TSP concentration for the region

s_i = the average deviation from the overall mean observed at site i

d_j = the average deviation from the overall mean observed at time j.

e_{ij}= deviations from the overall mean at a particular site and time not accounted for by s_i and d_j.

To better understand the model, assume that there is an average background level of total suspended particulates in the Study Region. If a prediction had to be made for a particulate concentration without knowing the specific location and date for which the prediction were to be made, this average level would be a reasonable guess. Yet with more information we can make a better estimation. Variations from this mean can be placed in three categories. To begin with, it is clear that not all sites are the same. Sites located near particulate sources tend to run higher than the average background level, while those in relatively pristine areas will tend to have consistently lower particulate concentrations. Thus, if we know what location we are asked to make a prediction for, we can improve our guess by calculating the average value for that site rather than the entire region. In the model, the difference (s) between the average for the whole region (μ) and the average at each site is calculated. The average concentration at site i can then be expressed as $\mu + s_i$. S will be negative at sites with little pollution and positive at sites strongly affected by particulate sources.

These estimates can clearly be improved if we take into account temporal variability. It is clear that on certain days particulate concentrations will be higher than average due to a particularly dirty air mass and on other days air over the entire region will be cleaner than usual because of air masses originating in unimpacted areas. We can then adjust the estimate by knowing whether the air on a given day was cleaner or dirtier than average, and then substituting the regional average on the day in question for the regional average over all time (μ). This can again be expressed mathematically as a deviation from the mean, where the average on day j is equal to the overall average (μ) plus the deviation from this average on day j (d_j).

These factors can be combined. To obtain an estimate for a particular place at a particular time, we can start with the overall average (μ). We can then adjust this if the day in question had dirtier or cleaner air than average ($\mu + d_j$). Finally, we can adjust if a site tends to have higher or lower than average particulate concentrations ($\mu + d_j + s_i$).

As an example, suppose we are interested in making a guess at what the TSP concentration was on a particular day when the sampler at a particular site was broken. We know from previous observation that the overall average concentration in the region is 45 $\mu g/m^3$. However, we also know from previous observation that this site tends to run on the average 10 $\mu g/m^3$ higher than the regional average. Our best guess, then, for the missing observations is 45 + 10 = 55 $\mu g/m^3$. However, from observations at other sites, we know that the air quality over the region was 20 $\mu g/m^3$ cleaner than average. The estimate then becomes 45 + 10 - 20 + 35 $\mu g/m^3$.

This is probably the best estimate available under the above circumstances, yet if we were to go through this procedure at a site and date for which a TSP reading was available, we might find a substantial difference between the value predicted by the above procedure and the actual reading. This is due in part to random fluctuation but is also due to a third sort of factor, namely some circumstance that is unique to a particular place at a particular time. Suppose, for example, a highway construction crew happened to be working near the site on that particular day. It is highly probable that the particulate concentration under these circumstances will be higher than normal, yet the effect will be highly localized. These sorts of effects concerning site i and day j are included in the model as e_{ij}.

The purpose of this analysis is to break a particulate reading down into several components. Estimation of the s_i terms enables us to identify which sites are consistently higher or consistently lower than average, and to quantify the magnitude of the difference between any two sites. The d_j terms provide an estimate of the magnitude of events affecting the entire region. Finally, identification of those samples with a high e_{ij} component provides a guide to the location of short term local effects.

Estimation of these effects is a fairly simple procedure. The easiest method would be to use the arithmetic mean of all observations as an estimate of μ, the arithmetic mean of all observations at site i as an estimate for $\mu + s_i$ and the mean of all observations taken on day j as an estimate of $\mu + d_j$. This procedure would indeed yield maximum likelihood unbiased statistical estimates if there were no missing values in the data set (Schleffe, 1959). However, if the data are not complete, this procedure can lead to biased estimates. Suppose, for example, that readings on one day were missing from the three sites when TSP concentrations are usually lowest. An estimate for d taken on that day from the remaining eight locations would clearly be too high. The estimate must be modified to take missing values into account. This is essentially done by estimating the missing values in the manner described above and calculating means using these estimates. The statistical methodology for obtaining these adjusted estimates of μ, s and d, while straightforward, involves development of a matrix notation too cumbersome to be present here. Detailed discussions can be found in Graybill (1961) and Scheffe (1959). Estimates of e_{ij} terms are obtained by taking the difference between the observed TSP concentration at site i and d_j and the predicted value obtained from the equation

$$\text{Predicted TSP} = \mu + s_i + d_j.$$

Estimates of e_{ij}, therefore, cannot be obtained for dates and sites where no TSP reading was taken.

Derivation of the estimates in this way also enables us to summarize results using an analysis of variance table. This has the advantage of permitting tests of the significance of the site and day effects. These tests will determine whether there is any statistically significant difference between sites or if they all behave alike. Similarly, we can test if any days are significantly different from any other or if regional effects tend to be constant. Analysis of this sort requires that certain assumptions be made involving normality of error (e_{ij}) terms and that the e_{ij} terms have the same variance for all sites and dates. A number of studies have shown that lognormal models are often appropriate for the description of air quality data (Larsen, 1971, 1973, 1974; Hunt, 1972; Neustadter and Sidik, 1974). Examination of frequency distributions of our TSP observations and the running of the model with several possible transformations of the data indicate that the lognormal model was indeed appropriate in this case, and that the assumptions outlined above were met under such a model. Accordingly, all analysis was done using log transformed data. All mean values resulting from the model are thus geometric means.

The model was applied to the entire data set, and the analysis of variance table (Table 1) reveals the presence of highly significant spatial and temporal effects. Estimates of the site effects (Table 2) indicate the magnitude of the difference between extreme background sites (such as Fernberg Road) and community and industrial locations. Note that the geometric mean at the highest station (Virginia) is more than five times the mean reading obtained at Fernberg Road. Note also that no site was in violation of either the primary (75 $\mu g/m^3$) or secondary (60 $\mu g/m^3$) annual standard for TSP concentrations. Finally, a graph of the adjusted day means (figure 1) shows the fluctuations observed over time.

Estimates of e_{ij} were computed for each observation. As these estimates are approximately normally distributed with mean 0 and variance 0.05 (from the analysis of variance) the upper 1% and 5% of the distribution can be calculated. Observations with e_{ij} estimates falling above these bounds may represent outliers, those points representing significant, short-term, local events. Note that a certain number of estimates of e_{ij} would be over these limits even if no such events occurred. If none of the events occurred, we would expect to find eight observations over the 1% limit. Fourteen were observed. It is likely, therefore, that some of these were true outliers. A list of observations falling above the 1% and 5% limit is found in table 3.

Conditions in the study region were not constant over the sampling period. In particular, several events took place that had a potential effect on TSP concentrations over periods of several weeks or longer. Most notable were snowcover, which can be expected to reduce particulate concentrations by preventing liftoff, and the strike against taconite mining observations in the second half of 1977. Two questions relating to these events are of interest. First, what was the effect on the regional air quality? Secondly, did these events affect some sites differently than others?

These questions can be answered by running the model separately for each of the time periods in question. A comparison of the regional mean estimated at each time period will provide an answer to the first question. To answer the second, we need to take the ratio between the mean observed at each site and the regional mean. Comparison between site effects obtained from the same site during different events will not reveal if that site was affected differently than the rest of the region. Suppose, for example, that during the mining strike the adjusted mean at a station was 20 $\mu g/m^3$ while the regional mean was 30 $\mu g/m^3$. Suppose, also, that before the strike the mean at the same site was 30 $\mu g/m^3$ while the regional mean was 45 $\mu g/m^3$. Clearly, in this instance, the strike had an effect both on the region and the site. Note, however, that both the site and region decreased by the same percentage (33%) and during both periods, the ratio of the site mean to the regional mean was 2/3. This implies that the drop in TSP concentrations observed at the site during the strike was a reflection of the regional trend. However, if the mean at the site during the strike was 10 $\mu g/m^3$, we would conclude that the strike had a greater effect at this location than over the region as a whole, as the site showed a decrease of 66% as opposed to the 33% drop in the regional mean, and the ratio of the site mean to the regional mean decreased to 1/3.

Table 4 contains a list of the time periods considered and the mean TSP concentrations over the region during the period. It should be noted that the figures for period 1 (startup, no snow cover) may not be reliable and are definitely not comparable with the figures for other time periods. Three sites, including two background sites, were not operational during this period. The regional mean for this interval is probably biased as a result. It should also be noted that the date for the resumption of mining activities is approximate. Not all mining operations resumed at the same time, though most of the larger operations went back to work very close to December 21, 1977. A notable exception was Erie Mining, where activity was sporadic from December 21, 1977 until February 19, 1978, when normal activities resumed.

It appears that both snow cover and mining activity play an important role in determining particulate concentrations. Note that the adjusted geometric mean concentration at the eleven sampling sites increased by approximately 17 $\mu g/m^3$ in the period following snow melt in 1977. Note also the drop of 21 $\mu g/m^3$ following the cessation of mining activities. Only a slight drop (3 $\mu g/m^3$) was noted when snow cover was present during the mining strike, and only a small increase (2.5 $\mu g/m^3$) was noted when mining activity was resumed. It is possible that this last difference might have been greater had all opportunities resumed at the same time.

The analysis of variance results for each time period are found on table 5. These tables show that both differences between sites and temporal differences were highly significant during each period in question. There is some evidence that there was, however, less variability between sites during the mining strike. The variance of site means in the period immediately before the strike (period 3, variance = 392. 49) is significantly greater than the variance seen in the comparable period with no snow cover during the strike (period 4, variance = 110.37) ($F = 3.55$, $p < .05$). By contrast, no difference in the between site variance was found for periods of snow cover and no snow cover. It appears, then, that mining activities play a major role in determining differences between sites.

Table 6 contains the site means for each period expressed as a percentage of the regional mean. These means are also graphed in Figure 2. Several interesting features may be discerned. The figures for periods 3 and 4 show that some sites were indeed disproportionately affected by the mining strike. In particular, the Erie Mining office went from 122% of the regional mean before the strike to 70% during the strike. Another location showing a drop in particulate concentrations greater than that seen over the region is the Hoyt Lakes Police Station (166% to 127%).

A few stations, however, did not show as great a drop as the regional average. Two of them, Kawishiwi and Toimi, were background sites, showing low concentrations throughout the course of the study. It is not surprising that mining activity would be of less importance at these locations than at other sites in the region. The other stations where the decrease in TSP concentrations were less than average were the larger communities, Virginia (284% to 215%) and Hibbing (192% to 150%), suggesting that activities other than mining were of importance at these sites. It should be noted, though, that every station showed a drop in particulate concentrations after the strike (table 7, figure 3), indicating that the air quality in all portions of the region is affected by mining activity.

The effect of snow cover also seems to vary from site to site. With the exception of Mountain Iron, all stations showed an increase in TSP concentrations from period 2 (mining, snow cover) to period 3 (mining, no snow cover). However, from period 6 (no mining, no snow cover) to period 7 (no mining, snow cover) six stations showed changes of less than 2 $\mu g/m^3$. Of the remaining five, four (Dunka Road, Hoyt Lakes Police Station, Mountain Iron and Hibbing) decreased while one (Virginia) showed a substantial increase. It should be noted that the effect of snow cover does seem to be less in the communities. This may reflect an increase in home and business heating during the snow season. An exception to this trend is Ely, where it is likely that activity is substantially increased during the spring and summer. The large increase from period 2 to period 3 observed at Fernberg Road, a popular entry point to the Boundary Waters Canoe Area, may also reflect in increase in activity near this site following snow melt.

In an effort to further explore the relationship between air quality in different portions of the region, correlation coefficients were computed between each pair of sites. The results (table 8) seem to indicate that all stations in the region correlate most closely with the background sites (Fernberg Road, Kawishiwi Lab and Toimi). This suggests that whatever relationships exist between stations are due to regional trends and that those effects causing differences between stations are highly localized. Note, for example, that the highest correlations are found between the three background stations. Fernberg Road and Toimi, located 35 miles apart, have a correlation of .94. Developed sites that are very close together show little correlation. Note, for example, the correlation of .35 between Mountain Iron and Virginia, separated by only three miles. The communities do not correlate at all well with each other, and in fact show stronger relationships with the background sites.

Table 9, for purposes of comparison, shows the correlation coefficients between the study region sampling sites and five locations in the Duluth area. Correlations between the study region sites and those two sites located away from the lakeshore in Duluth (Airport and Cloquet) are surprisingly strong; again, relationships are strongest between these two sites and the study region background sites. Correlations between the study region and the three Duluth sites near the lake are weak, but again seem to be strongest with the background sites. The relationship between these three Duluth sites and the Iron Range cities (Virginia, Mountain Iron and Hibbing) is virtually nonexistent.

However, overall correlations are not sufficient to illustrate the relationship between the study region and Duluth. Table 10 contains correlations between study region sites and Duluth sites when the wind at Hibbing was blowing from the south and southwest, from the Duluth area to the study region. For the purposes of this analysis,only those days when the wind was blowing from an arc between 150^{o} and 240^{o} for four or more daylight hours were considered. Nineteen sampling dates fell into this classification, comprising 21% of the total sample. Of these nineteen dates, only five occurred during the period of snow cover. Average wind speed on these days was 4.42 meters/second, slightly higher than the average wind speed at Hibbing of 3.95 meters/second (Watson, 1978).

The contrast between these correlations and the overall correlations is dramatic. Nowhere is this more apparent than at Mountain Iron. The overall correlation between Mountain Iron and Duluth west end is -.01, effectively non-existent. However, on the nineteen days with prevailing southerly and southwesterly winds, the correlation between these sites rises to .76, a very strong relationship. This pattern is not unique. Of the fifty-five possible correlations between Duluth and Study Region sites, fifty-two were higher when the wind was blowing from the Duluth area to the study region. Many of the increases are substantial. The relationship is most striking at Virginia and Mountain Iron, yet shows in other areas as well. The correlation between Duluth West End and Kawishiwi Laboratory, for example, was .32 overall, but rose to .62 when the wind was from the south. It appears, then, that particulate transport from the Duluth area can play a significant role in determining the air quality of the Study Region.

Relationship to Meteorological Factors

In an attempt to better explain spatial and temporal variations in total suspended particulate observations, statistical models were derived to relate the observations at each site to meteorological parameters. Particular attention was paid to wind direction, as analysis of the relationships between direction and particulate concentrations can suggest possible sources of particulates.

Some researchers (e.g. Samson, Neighmond and Yencha, 1975) have suggested using correlation coefficients as a measure of association between suspended particulate and wind direction. This method utilizes wind frequency distributions and involves the computation of correlation coefficients between 24 hour mean TSP concentrations and the wind frequency $\left(= \frac{\text{\# of hours wind blows from direction } i}{24} \right)$ for each wind direction under consideration. This method if direction-pollution association is viewed as an alternative to the "pollution rose" commonly used for this sort of model.

However, the pollution rose has an ease of interpretability that the displays of Samson et al. seem to lack. The figures plotted on a pollution rose represent the actual particulate concentrations expected when the wind is blowing from a particular direction. Correlation coefficients, while providing a measure of the strength of association between concentrations and wind directions, do not provide any indication of the level of pollution expected. However, most pollution roses do not provide any indication of the strength of the association, or any indication of the possible error in a plotted association.

The methodology presented here attempts to combine the best features of both methods. The method used is multiple regression analysis. Correlation analysis as used by Samson et al. essentially involves the computation of a separate bivariate regression model for each wind direction. Multiple regression results in one model accounting for all wind directions. The form of this model is:

$$TSP = \beta_1 D_1 + \beta_2 D_2 + \ldots + \beta_n D_n$$

where D_i = expected TSP concentration when the wind is blowing from direction i.

The rationale for this model is simple. It states that the mean concen= tration over 24 hours will be an arithmetic average of the concentration observed from each wind direction weighted by the frequency of each wind.

The major computational task is estimation of the βi terms. This can be done using standard regression analysis techniques. (Draper and Smith, 1966). It is also possible to compute standard errors for these co- efficients. By computing both the coefficient and its standard error, we estimate both the expected particulate concentration when the wind is blowing from a given direction and the deviation that might be expected from this estimate.

It is clear that the concentration observed when the wind is blowing from a given direction will not always be that predicted by the model. In fact, it may be very different. This is particularly true if short term local conditions exist that affect pollution readings for one or two sampling dates. An example of such a condition would be a construction project at or near a sampling location. If the wind blows from the construction site to the sampler during construction, pollution levels may well be much higher than would be observed under identical meteorological conditions before or after construction. Identification of these atypical points ("outliers") is necessary for a complete analysis of suspended particulate data, and can easily be accomplished by examination of the residuals (difference between predicted and observed values) arising from the multiple regression models. Outliers can be detected using the Bonferroni criterion (Snedecor and Cochran, 1967; Weisberg, 1977). The models should be redone after outliers are deleted, as it is possible for one or two extreme values to grossly alter a regression estimate.

Models were constructed for all eleven study region sites at which TSP samples were taken. Wind data were obtained from the Hibbing airport. A wind rose for those dates on which TSP samples were taken is attached (Figure 4). It compares quite closely with the ten year wind rose for the Hibbing airport (Figure 5) implying that the wind conditions for the study were typical of long term regional patterns. It must be assumed, however, that the wind data from Hibbing represent conditions throughout the study region; no better information is available.

It was decided to use only daylight hours to determine the wind frequency distribution, as nighttime winds were found to be light and highly variable. For modeling purposes, daylight was defined as the period between 6 AM and 6 PM. Furthermore, a better fit was found if only non-calm hours were used. The independent variables, then, represent the percentage of non-calm daylight hours during which the wind was blowing from each direction. Wind was grouped into twelve thirty-degree intervals. All calculations were done using the computer program MULTREG, developed by the Department of Applied Statistics, University of Minnesota (Weisberg, 1977). Pollution roses were generated by a FORTRAN program utilizing the CALCOMP plotting package on the University of Minnesota Cyber 74 computer. A 95 percent upper confidence limit is plotted along with the pollution rose. This was computed using the formula:

$U.L. = \beta + [s.e.\ (\beta)]\ t_{.05,df}$ (Snedecor and Cochran, 1967)

where U.L. = upper confidence limit

$s.e.(\beta)$ = standard error of estimated β

$t_{.05,\ df}$ = 95^{th} percentile from from a t distribution with n-12 degrees of freedom (n = sample size).

The distance between the upper confidence limit and the estimated concentration was found to vary greatly. This implies that some of the expected concentrations are very accurately estimated. From other wind directions (those for which the differenc e between the upper confidence limit and the estimate is high) the estimates are not very accurate. The reasons for this lack of accuracy are three. First, and most difficult to estimate, is lack of precision in the data, most notably inaccuracies arising from applying Hibbing wind data to other locations. Secondly, pollution levels at a given wind direction may be highly variable. This cause tends to disappear after outliers are deleted. Thirdly, a glance at the Hibbing wind rose

(Figure 5) will show that some winds are quite rare in the study region. In particular, winds from the northeast and southwest were rarely observed for more than one or two hours a day and on most days were not observed at all. There are simply too few observations at these wind directions to permit the derivation of a reliable estimate. In some extreme cases, this may even lead to negative estimates. There are statistical artifacts caused by a lack of data along with high variability at those observations that were made. We would expect those estimates to become positive and stabilize as the number of samples is increased. When a negative estimate was encountered in the models for the study region, the value 1 $\mu g/m^3$ was substituted as a reasonable minimum value.

To assess the importance of a suspended particulate source to a specific location, it is necessary to know both the pollution level that can be expected from the source and the frequency with which the wind blows from the source to the location under consideration. A pollution rose displays the former, a wind rose the latter. It is possible to combine the two by multiplying the expected particulate concentration at a given direction by the probability of the wind blowing from that direction. This number can then be standardized to obtain the expected percentage of annual pollution contributed from each wind direction. Specifically, the formula for the expected contribution from direction D_k is:

$$\text{Exp. Cont. } (D_k) = \frac{(\text{Concentration } 1\ D_k)\ P\ (DK)}{\sum_{i=1}^{n} (\text{concentration } 1D_i)\ P\ (D_i)} \times 100$$

where Concentration 1 D_k = expected TSP concentration when wind is blowing from direction k, and $P(D_k)$ = probability that wind is blowing from direction k.

Results

Pollution roses for each of the eleven study region TSP sites are presented in figures 6 through 41. Two pollution roses are presented for each site, the first calculated from all observations and the second calculated after outliers had been deleted as described above. These outliers are listed in table 11. Expected contribution roses are also included. The expected contribution roses often show peaks to the south and the northwest, reflecting the dominant winds. Sources to the east and northeast of a site are almost never important contributors, though they may cause isolated high readings. A site-by-site summary follows.

Fernberg Road (7001) -- The pollution rose shows a peak to the west-southwest, in the general direction of the town of Ely and the dirt road leading up to the site. Peaks to the southwest and south may indicate contributions from mining areas and more populated regions, as there are no obvious local sources in these directions. The annual contribution rose shows the peak to the south to be the most important, contributing about 25% of the annual pollution.

Ely High School -- The most notable peak on the pollution rose lies to the east-southeast. However, there do not seem to be any apparent local sources in this direction. A smaller peak to the west-southwest may result from emissions from the school heating plant stack. Other peaks are seen to the south, in the general direction of the eastern Iron Range. The annual contribution rose reflects the wind rose. Twenty-five percent of the annual particulate pollution at this site comes from the northwest, indicating the Ely business district. Concentrations when the wind blows from this direction, however, are not high ($<20\ \mu g/m^3$).

Kawishiwi Laboratory (7003) -- Two peaks on the pollution rose are most noticeable, indicating sources to the south (average concentration of ~25 $\mu g/m^3$) and southwest (30 $\mu g/m^3$). These most likely indicate local sources, with the dirt laboratory parking lot to the southwest and the dirt road leading to the laboratory from the south. Long range transport, possibly from the Iron Range, may account for a portion of these peaks, though it should be remembered that this location shows a smaller proportional decrease due to the mining strike than did most other sites. Concentrations when the wind was blowing from the forested areas to the north, northeast and east were quite small. The annual contribution rose is again seen to reflect the wind rose, the bulk of the particulate matter coming from the south (25%) and the northwest (18%).

Dunka Road (7006) -- The pollution rose for Dunka Road shows an area of elevated concentration stretching clockwise from the south to the west. This most likely indicates the nearby dirt logging road as a local source and may possibly indicate Erie Mining and other taconite operations as more distant sources. A peak from the northeast may result from Reserve Mining. The bulk of the total annual pollution comes from the north and northwest, reflecting, perhaps, a consistent low level of dust from Dunka Road. Chemical dust control is practiced on the road but prevailing northwest winds may still make the road an important source for downwind locations.

Toimi (7007) -- The Toimi pollution rose is notable for the lack of distinct peaks. An area of higher concentration is found clockwise from the southeast to the west-northwest, indicating long range transport from populated and industrial areas. Very low concentrations are found when the wind is from the nonpopulated areas to the northeast and east. The annual contribution rose very closely resembles the wind rose, with the lack of any significant contribution from the northeast and east being the most notable factor.

Erie Mining Office (7008) -- Not surprisingly, the largest source indicated by the pollution rose at this site is the open pit mine located just to the west of the site. Other peaks are seen to the northeast and north-northeast, towards the tailings basin and processing plant. This is also in the general direction of the Reserve Mining operation, so longer range transport may be occurring. Some elevation is also seen to the south and southwest, in the general directions of the communities of Hoyt Lakes and Aurora. Annual pollution is seen largely to come from the dominant north-west and south winds and from the mine area to the west. It should be noted, however, that Erie Mining is located at a break in the Iron Range. The wind at this site is channeled more on a north-south axis than at Hibbing. Therefore, the peak indicated to the northwest may well represent sources to the north.

Hoyt Lakes Police Station (7009) -- Peaks at this site seem to indicate residential and industrial areas. The peak to the west indicates the town of Aurora as well as possibly the large Iron Range cities, while peaks to the south and southwest point to the residential areas of Hoyt Lakes. These peaks may also reflect contributions from the Duluth area, as this site showed a fairly strong correlation with several Duluth sites (Table 9). No single source is apparent to account for the peak to the east-southeast, although local traffic and buildings may be the cause. The peak to the northeast is probably due to the Erie Mining operation.

Over 40% of the annual particulate pollution at this site seems to come from the south. Other directions from which major contributions are made are the west and northwest.

Hoyt Lakes Golf Course (7010) -- Very few distinct features can be found on the pollution rose. Higher concentrations are found in a sector running clockwise from the southeast to the west, towards the road and, less immediately, towards residential areas. The rest of the rose shows low concentrations, with the notable exception of a peak to the north towards Erie Mining. Particulates coming from this direction comprise the most important contribution of any direction, accounting for about 25% of the annual particulate pollution at this site.

Mountain Iron (7514) -- Mountain Iron is one of the few locations with elevated concentrations coming from the northeast and north-northeast. These readings almost certainly result from the large Minntac open pit taconite mine. Other notable features include peaks from the southeast, towards Virginia and Eveleth; the south, towards the center of town; and from the west-southwest, possibly resulting from the tailings basin and local traffic. The annual contribution rose shows that the single largest contribution again comes from the northwest, reflecting the dominant wind. The relative infrequency of wind from the northeast minimizes the importance of the high concentrations seen from this direction.

Virginia (1300) -- Virginia recorded the highest mean level of total suspended particulates seen in the study region. The pollution rose suggests the presence of sources in several directions, with the largest found on the dominant northwest and south wind axes. The largest peak lies to the west-northwest. Two potential sources lie on this axis, the municipal power plant and the Minntac operation at Mountain Iron. It is likely, in fact, that the Minntac operation contributes more particulate pollution to Virginia than to Mountain Iron due to the dominance of the northwest wind. Peaks also exist to the south and southeast. A large number of potential sources exist in these directions, among them mining operations, the center of Virginia and the city of Eveleth.

Hibbing (7516) -- Hibbing is the largest city in the study area (excluding Duluth), yet there is very little active mining in the immediate area. It would be expected, therefore, that much particulate pollution would come from general activity. The pollution rose for Hibbing shows a large peak to the west-southwest, towards downtown Hibbing, and to the east towards a heavily traveled Highway 169. Other peaks are from the south. The annual contribution rose reflects the wind rose except for the strong west-southwest component.

Region -- A pollution rose for the region was constructed using results from the statistical model described earlier. Regional TSP readings were calculated as the sum of the overall mean (μ) and the day effects (d_j). The resulting figures enable us to estimate trends in particulate concentrations affecting the entire region. Site differences and local effects have been removed.

The pollution rose shows that the largest regional effects occur when the wind is from the west and west-southwest. This most likely represents regional contributions from the densely populated areas of the Iron Range on the western fringe of the study region. It may also represent long range transport from agricultural areas. Another peak is seen from the south which may represent transport from the Duluth area or possibly more distant sources. The annual contribution rose shows that regional contributions reflect the wind rose with the bulk of particulates coming from the northwest and south and only insignificant contributions from the northeast and east.

Other meteorological factors -- Although wind direction was considered to be the meteorological parameter of primary importance, correlations of total suspended particulate concentrations with wind speed and precipitation were computed. Wind speed does not seem to be important, as a significant correlation between wind speed and TSP concentration was found at only one location (Virginia). Precipitation was seen to have a greater effect. Correlations were computed between TSP and an indicator variable for precipitation. This variable took on the value 1 if precipitation occurred on the date in question and 0 if no precipitation was recorded. Correlations were computed for both precipitation on the day TSP samples were taken and the day before TSP samples were taken. In all cases correlations betweeen TSP and precipitation were seen to be negative, implying that precipitation is associated with lower TSP values and that the surface is a major source of particulates.

Unusual or Aberrant Observations (Outliers) -- Outliers among the TSP observations were detected by two methods. The first was from the statistical model, and the second was from the regression analysis that led to the pollution roses. A list of outliers from the statistical model may be found in table 3 and a similar list from the pollution roses is presented in table 11. These lists are not identical. This is because in practice, each method is detecting a different sort of outlier.

The statistical model detects those observations not explained by the differences between sites or by regional trends. These outliers represent short-term, local effects. These could arise for two reasons. The first is what we hope to detect by this analysis, a short-term, local disturbance such as a forest fire, logging or construction. The second arises from a source almost always present, but wind conditions that will transport material from the source to the site in question are very rare. A town located just west of a mine, for example, may almost never be affected by the mine because of the scarcity of easterly winds.

These latter points, however, will not show up as outliers in the regression (pollution rose) analysis. The regression analysis outliers arise from short-term local effects and from short-term regional effects. Unlike the statistical model, the regression analysis does not separate regional from local effects.

However, both models do detect those outliers resulting from short-term local sources. We can identify these by finding which observations appear as outliers in both models. Those observations that are outliers in the wind model but not in the statistical model represent short-term regional effects, while those outliers resulting from the statistical model but not from the wind regressions may represent high concentrations caused by rare wind patterns.

A list of outliers from both models is found in table 12. Only 11 observations fall into this category, about 1% of the sample. It is worth noting that six of these observations were found in one two-month interval, from April 13, 1977 to June 6, 1977.

Table 6. Site Means Expressed as a Percentage of Regional Mean for each time period.

Site	Overall	2	3	4	6	7	8
		Time Period					
Fernberg Road	45%	29	53	55	40	59	35
Ely High School	105	92	101	97	133	150	126
Kawishiwi Lab	46	33	47	62	43	59	41
Dunka Road	98	112	105	91	131	58	98
Toimi	52	49	55	63	60	68	43
Erie Mining Ofc.	94	100	122	70	73	80	104
Hoyt Lakes Police	134	120	166	127	154	166	115
Hoyt Lakes Golf	72	96	65	65	52	76	65
Mt. Iron	220	365	161	164	215	72	318
Virginia	254	310	215	284	188	302	257
Hibbing	175	166	150	192	232	223	200

Table 7. Site Means during each time period.

Site	1	2	3	4	6	7	8
Fernberg Road	--	5.50	18.59	8.05	6.27	7.68	5.45
Ely High School	23.17	17.45	36.00	14.19	20.85	19.53	19.61
Kawishiwi Lab	28.68	6.26	16.75	9.07	6.74	7.68	6.38
Dunka Road	44.56	21.25	37.42	13.31	20.54	7.55	15.25
Roimi	--	9.29	19.60	9.22	9.41	8.85	6.69
Erie Mining Office	53.27	18.97	43.48	10.24	11.45	10.42	16.18
Hoyt Lakes Police	--	22.76	59.16	18.58	24.15	2;/61	17.89
Hoyt Lakes Golf Crs.	37.90	18.21	23.17	9.51	8.15	9.90	10.11
Mt. Iron	88.10	69.24	57.38	23.99	33.71	9.37	49.48
Virginia	116.27	58.81	76.63	11.55	29.48	39.32	39.99
Hibbing	64.54	31.49	53.46	28.09	36.38	29.03	31.12

References

Draper ".R. and H. Smith, 1966. Applied Regression Analysis, New York, John Wiley and Sons.

Graybill, F.A. 1961. An Introduction to Linear Statistical Models. Volume 1, New York, McGraw-Hill.

Hunt, W.F. Jr., 1972. "The precision associated with the sampling frequency of lognormally distributed air pollutant measurements." J. Air Pollution Control Assoc. 22:4.

Koopmans, L.H., 1974. The Spectral Analysis of Time Series, New York, Academic Press.

Larsen, R.I., 1971. A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards, Publ. AP-89, U.S. Environmental Protection Agency, Research Triangle Park, N.C.

Larsen, R.I., 1973. "An air quality data analysis system for interrelating effects, standards, and needed source reductions." J. Air Pollution Control Assoc. 23:933

Larsen, R.I., 1974. "An air quality data analysis system for interrelating effects, standards, and needed source reductions - part 2." J. Air Pollution Control Assoc. 24:6.

Neustadter, H.E. and S.M. Sidik, 1974. "On evaluating compliance with air pollution levels 'not to be exceeded more than once per year.'" J. Air Pollution Control Assoc. 24:559.

Phadke, M.S., M.R. Gape and G.C. Tiad, 1978. "Statistical Evaluation of Trends in Ambient Concentrations of Nitric Oxide in Los Angeles." Environmental Science and Technology. 12:430.

Rao, S.T., P.S. Samson and A.R. Peddada, 1976. "Spectral Analysis Approach to the dynamics of air pollution." Atmospheric Environment. 10:375.

Samson, P.J., G. Neighmond and A.J. Yencha, 1975. "The transport of suspended particulates as a function of wind direction and atmospheric conditions." J. Air Pollution Control Association. 25:1232.

Scheffe, H., 1959. The Analysis of Variance. New York, John Wiley and Sons.

Snedecor, G.W. and W.G. Cochran, 1967. Statistical Methods. Ames, Iowa, Iowa State Press.

Watson, B.F., 1978. The Climate of the Copper-Nickel Region of Northeastern Minnesota - Part A. Minnesota Environmental Quality Board. (In Press).

Weisberg, S., 1977. MULTREG User's Manual, University of Minnesota, School of Statistics Technical Report #298.

Figure 1

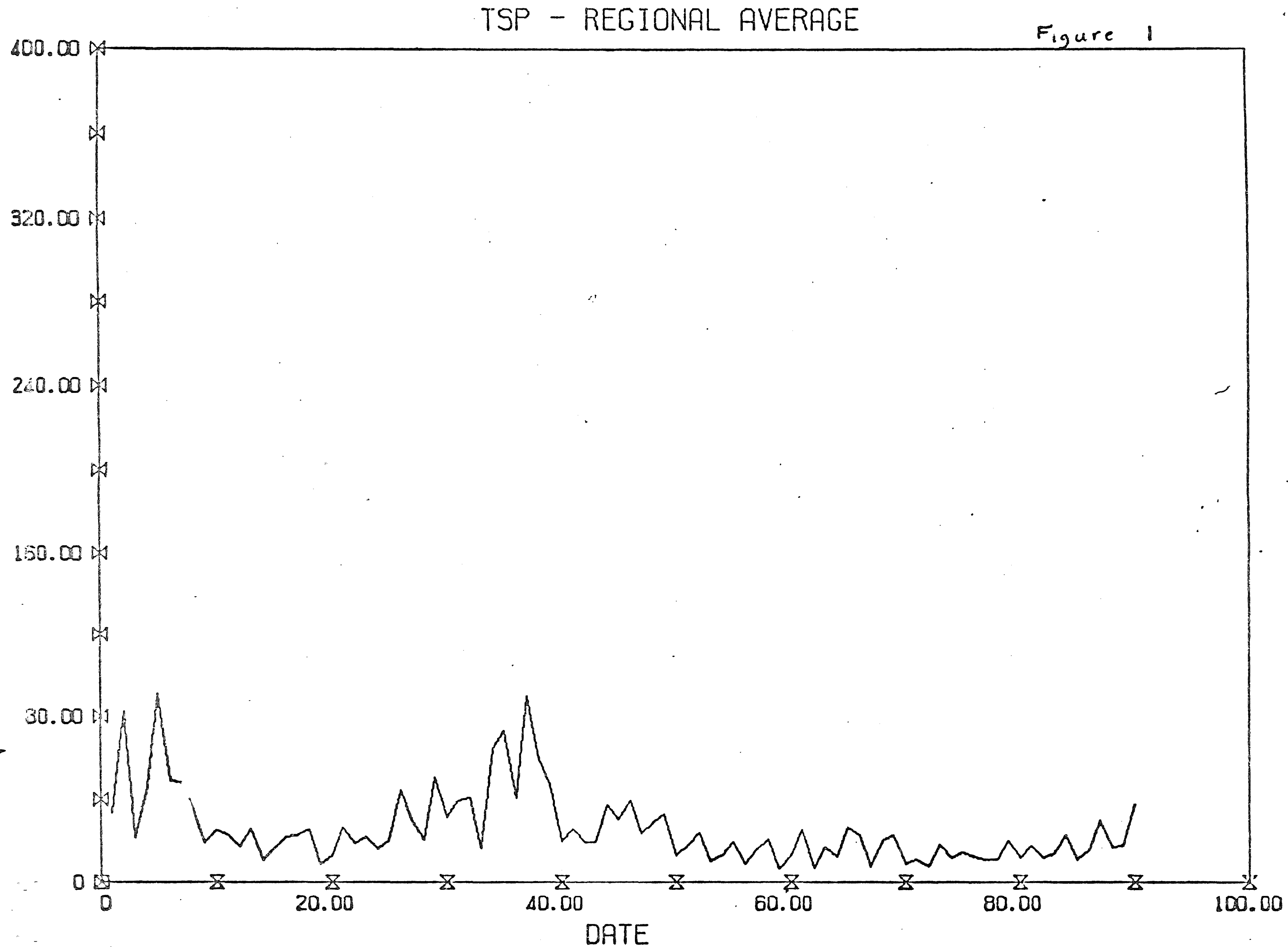
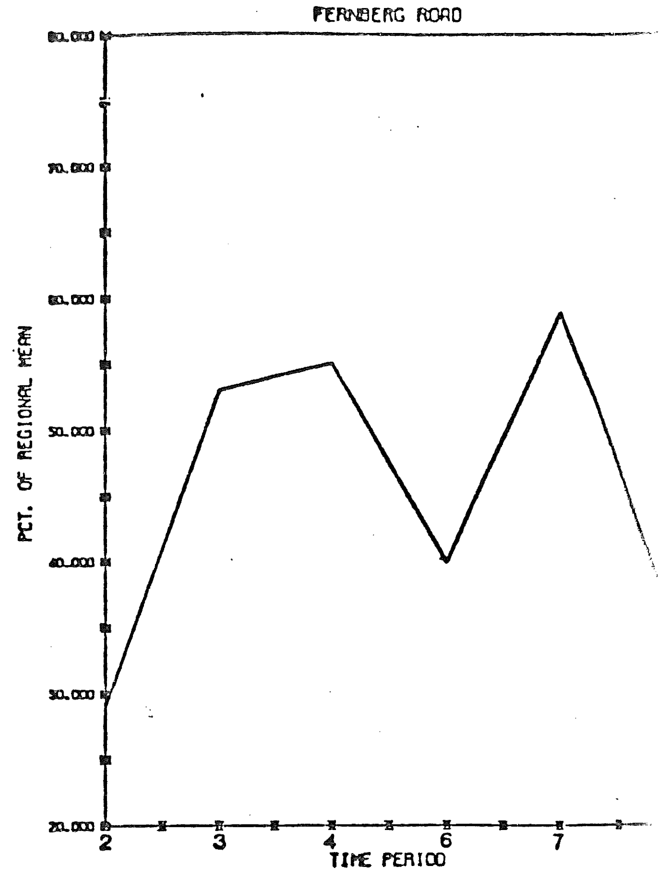
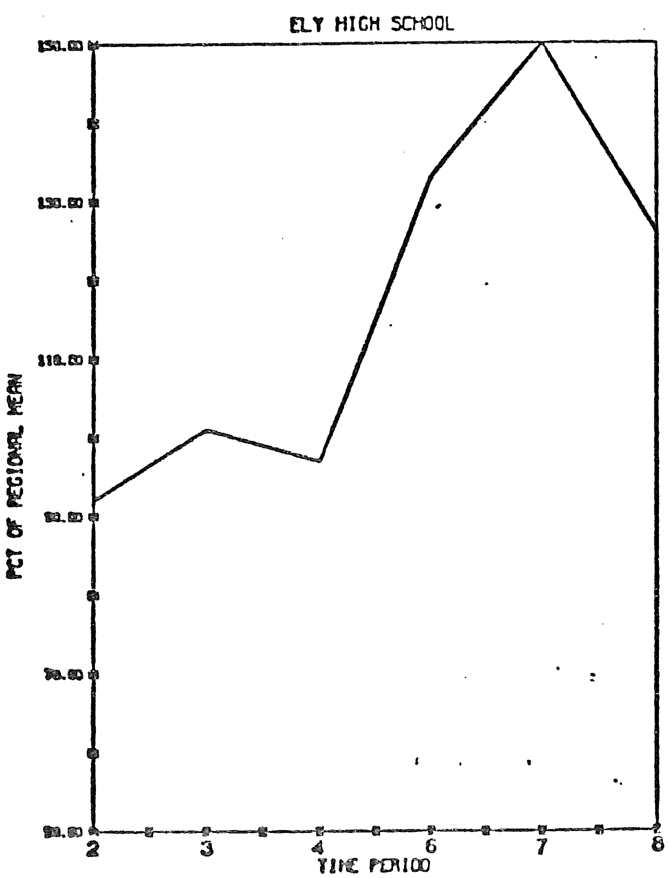

TSP - REGIONAL AVERAGE
400.00
320.00
240.00
160.00
30.00
0
0
20.00
40.00
60.00
80.00
100.00
DATE

FIGURE 2
PERCENTAGE OF REGIONAL TSP MEAN BY TIME PERIO

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





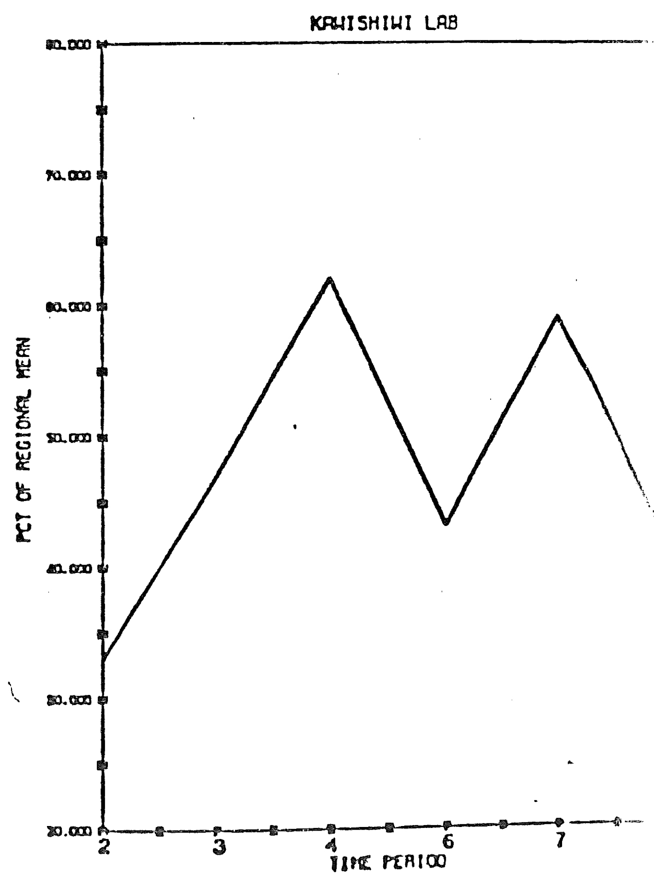
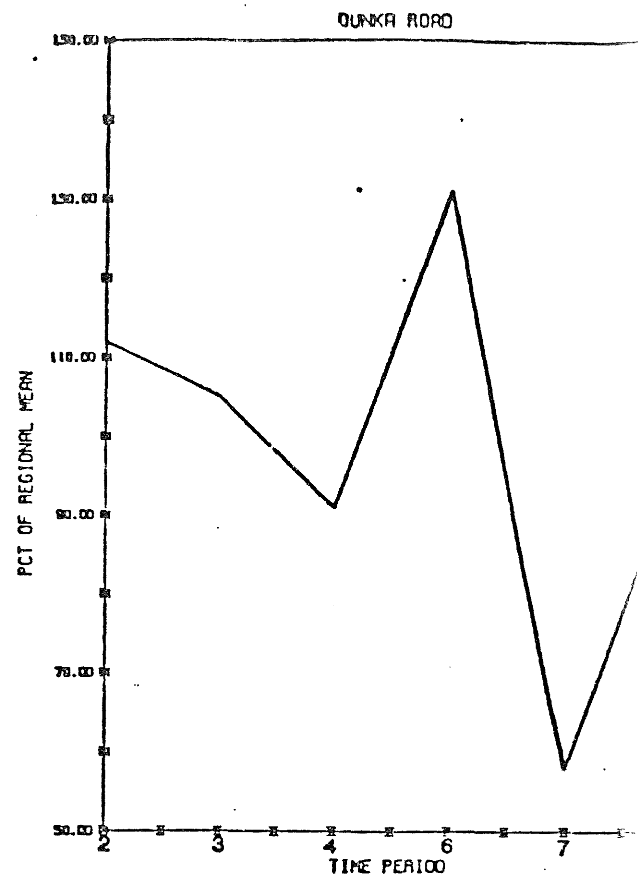
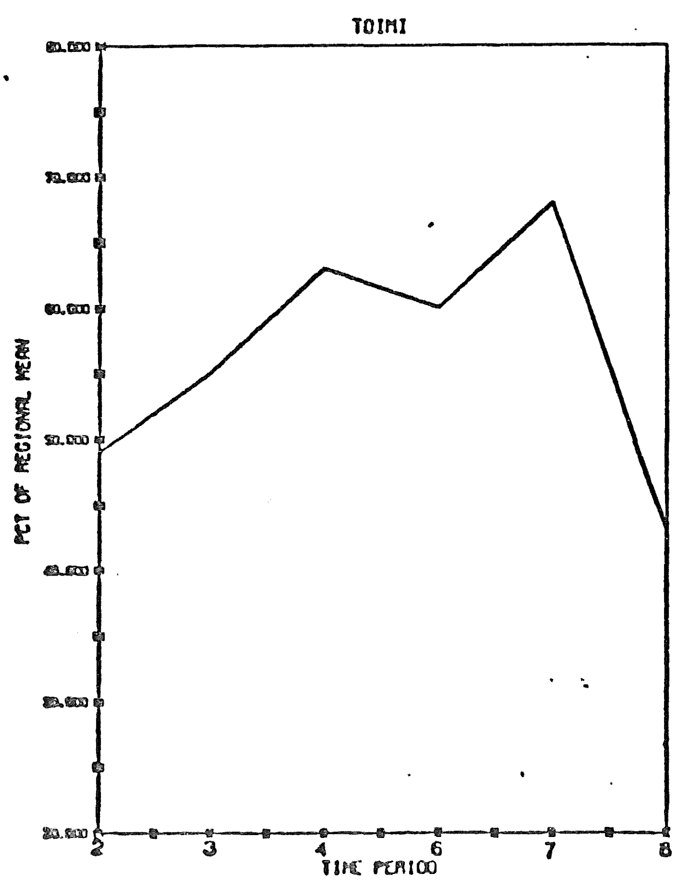


FIGURE 2
PERCENTAGE OF REGIONAL TSP MEAN BY TIME PERIC

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





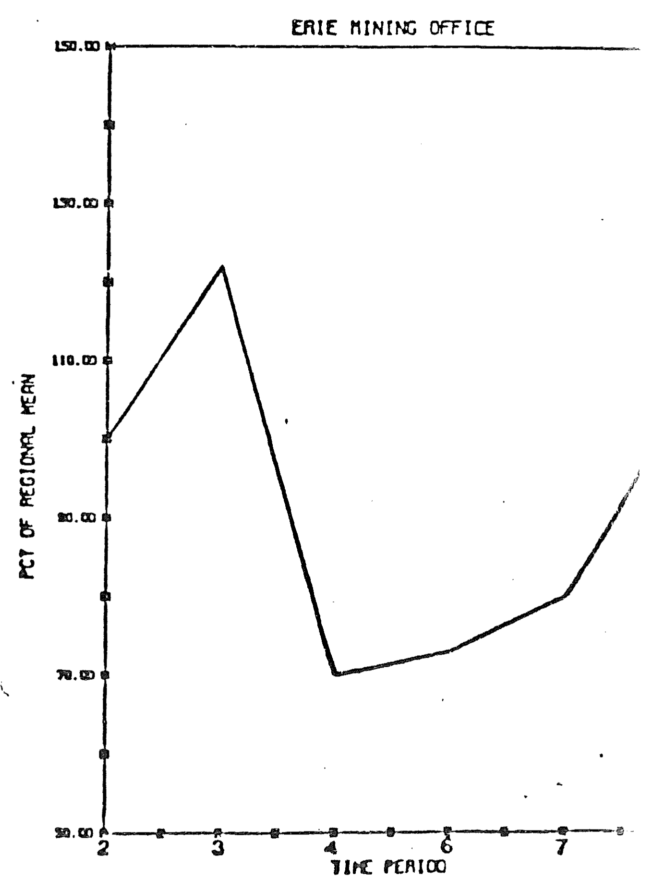
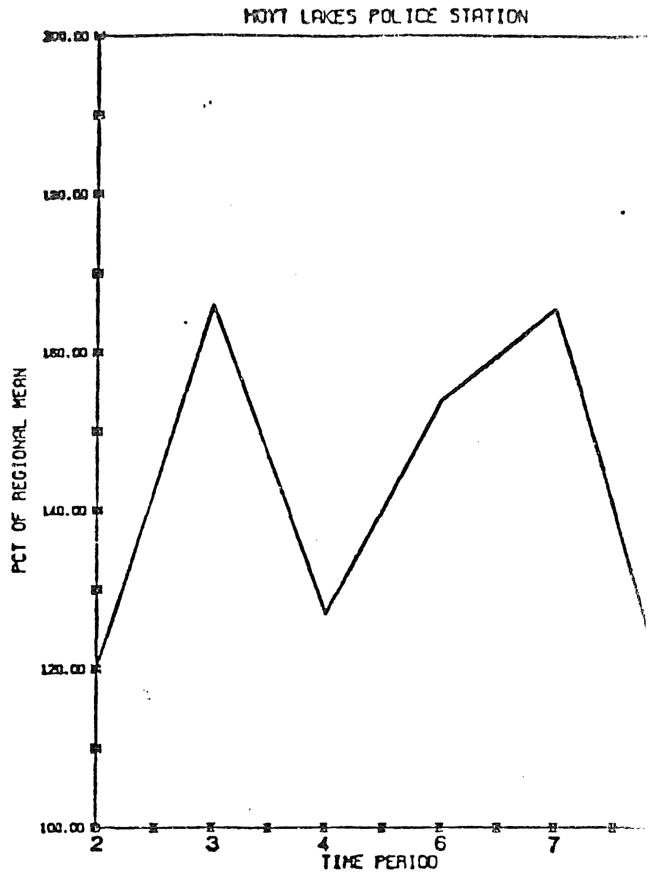
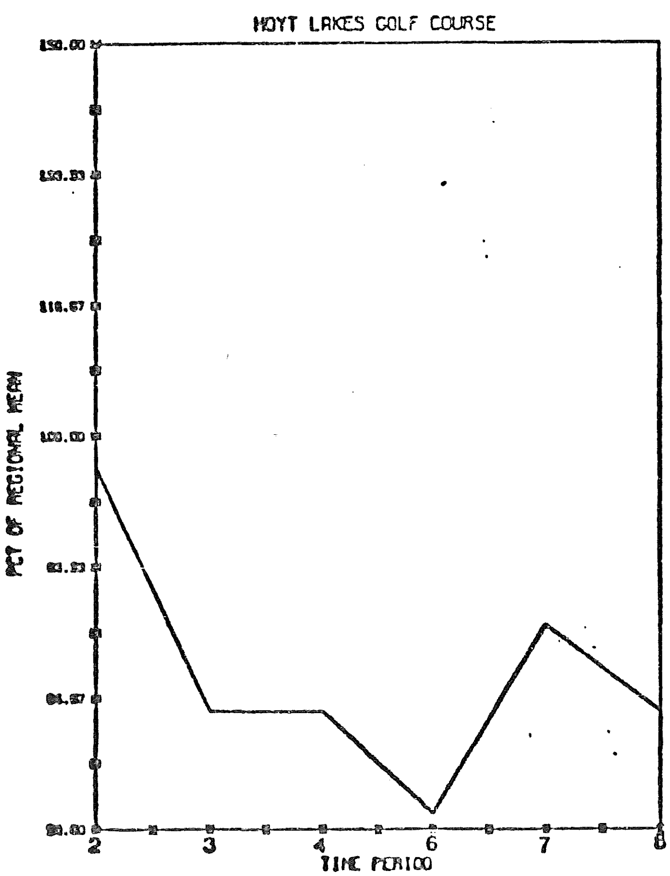


FIGURE 2
PERCENTAGE OF REGIONAL TSP MEAN BY TIME PERIC

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





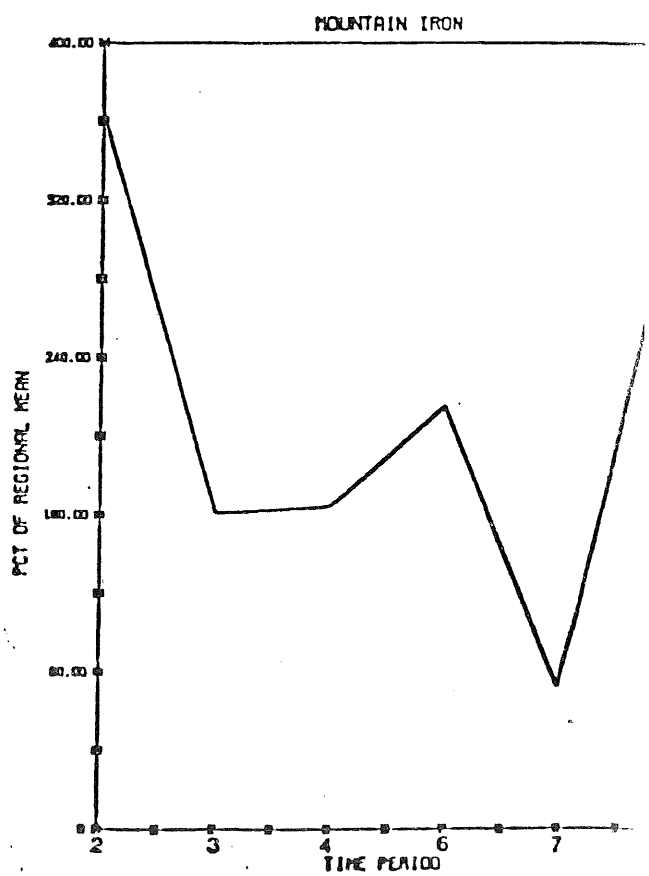
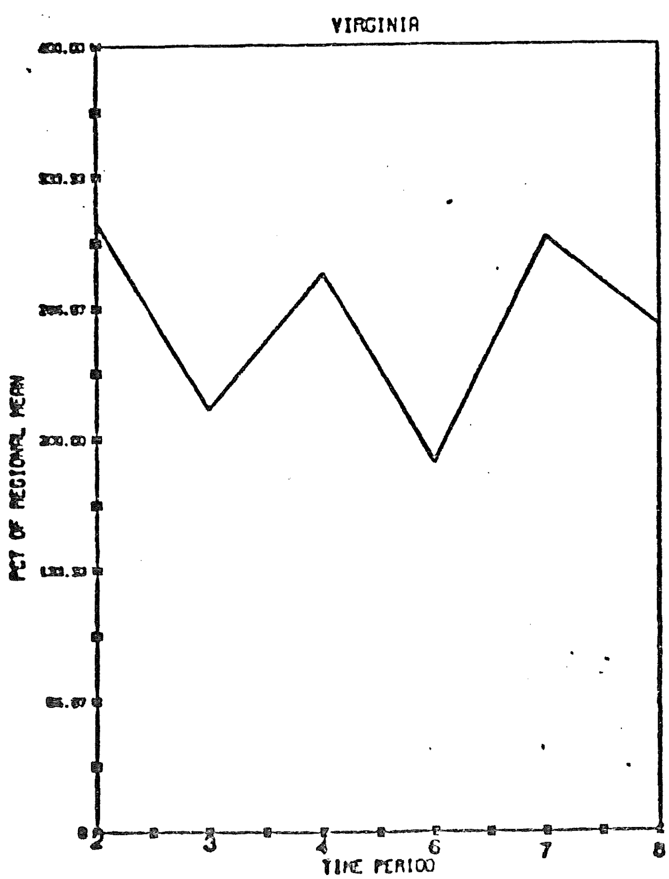


FIGURE 2
PERCENTAGE OF REGIONAL TSP MEAN BY TIME PERIC

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78



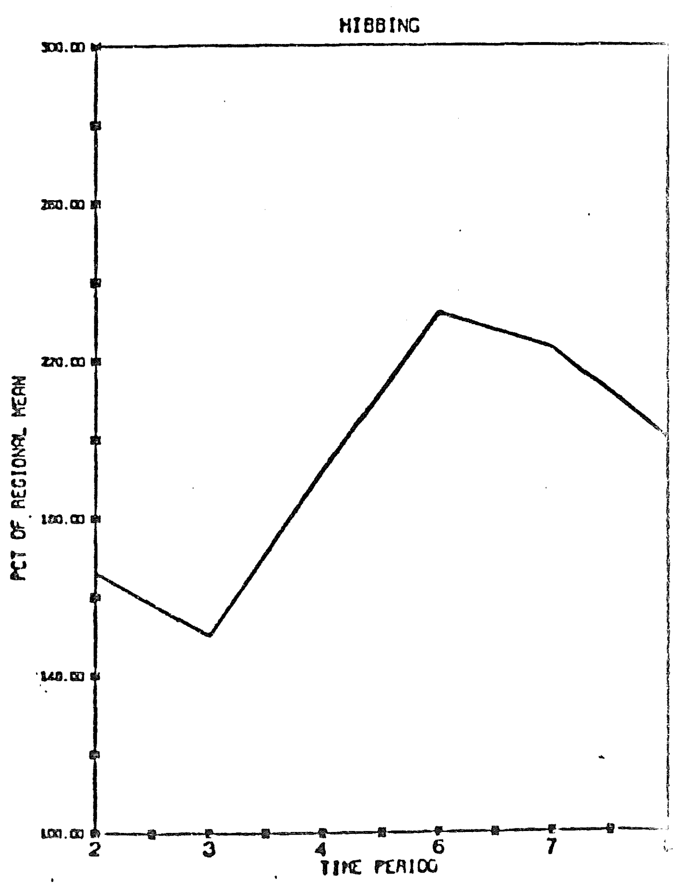
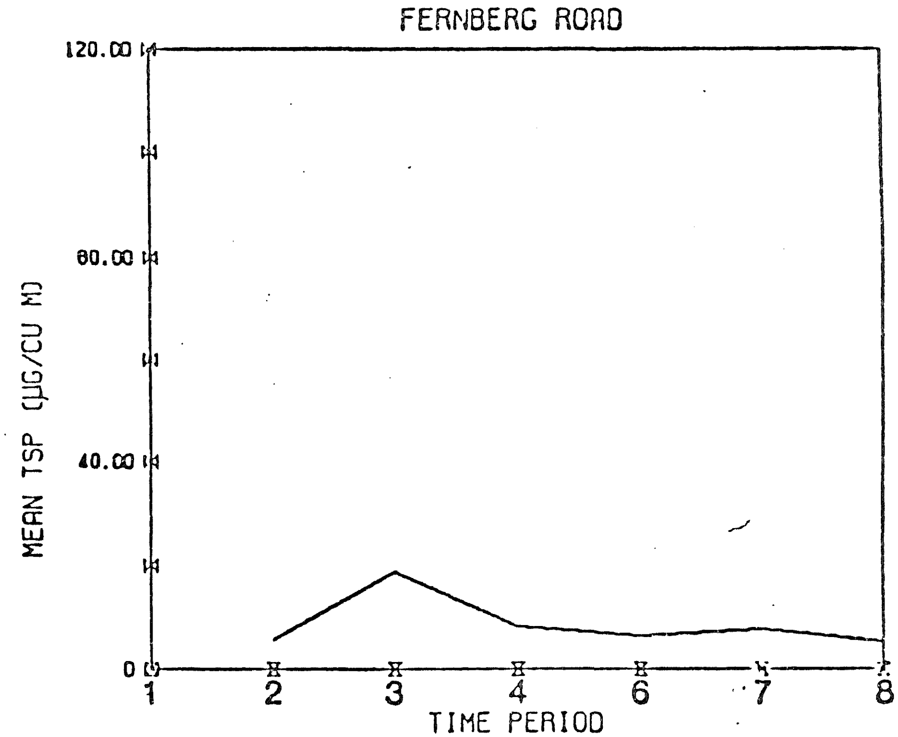
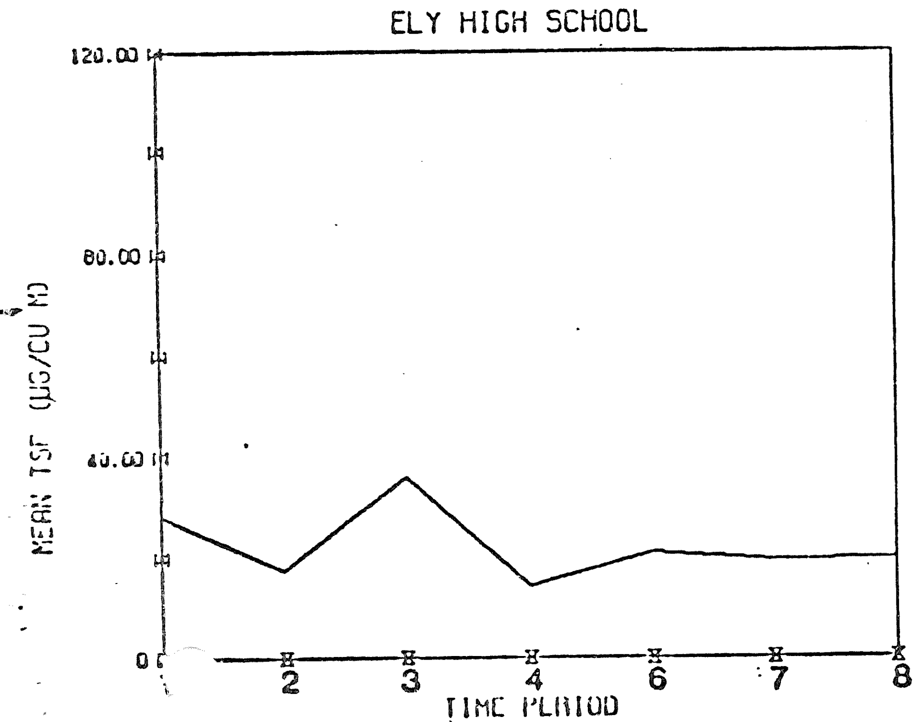


TABLE 3 MEAN TSP BY TIME PERIOD

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





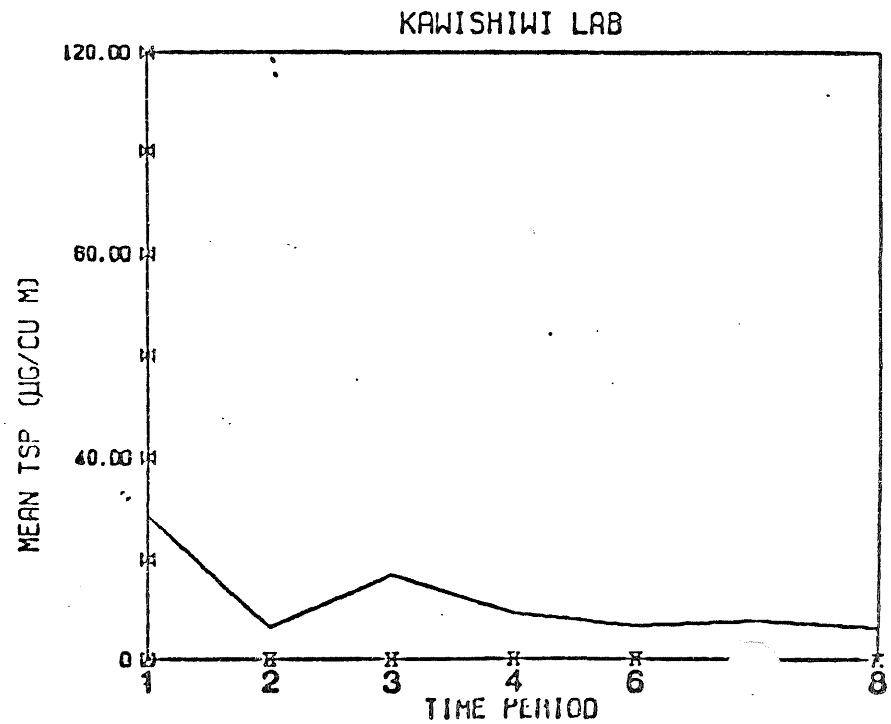
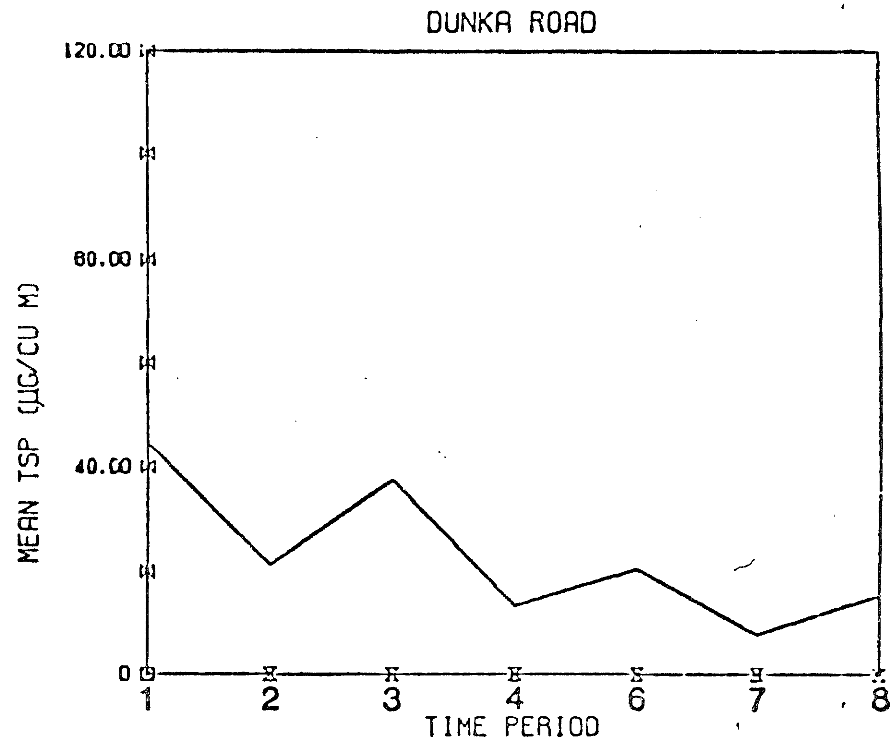
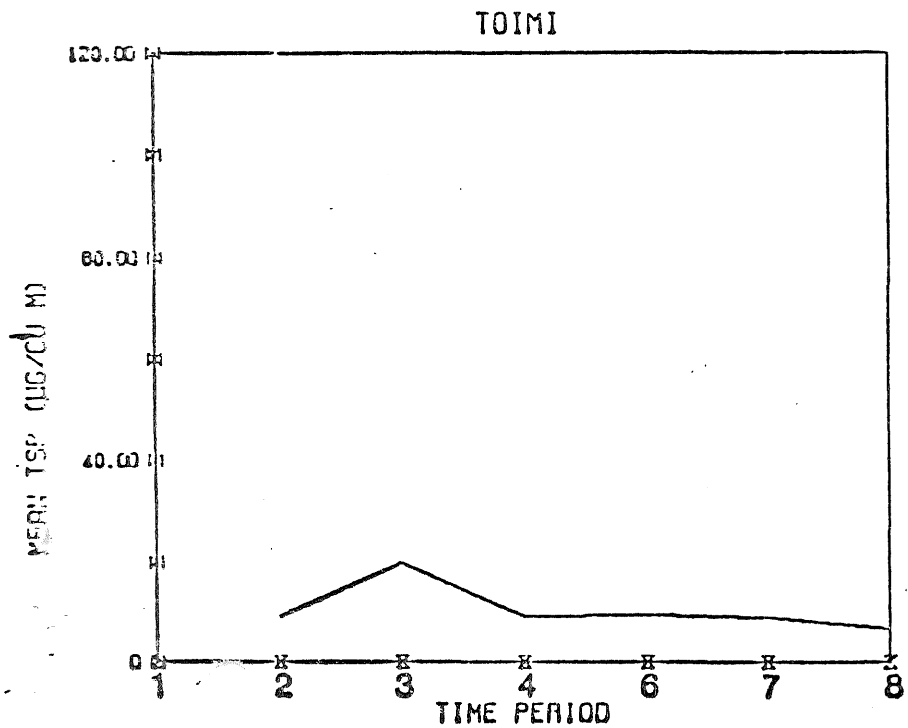


TABLE 3 MEAN TSP BY TIME PERIOD

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





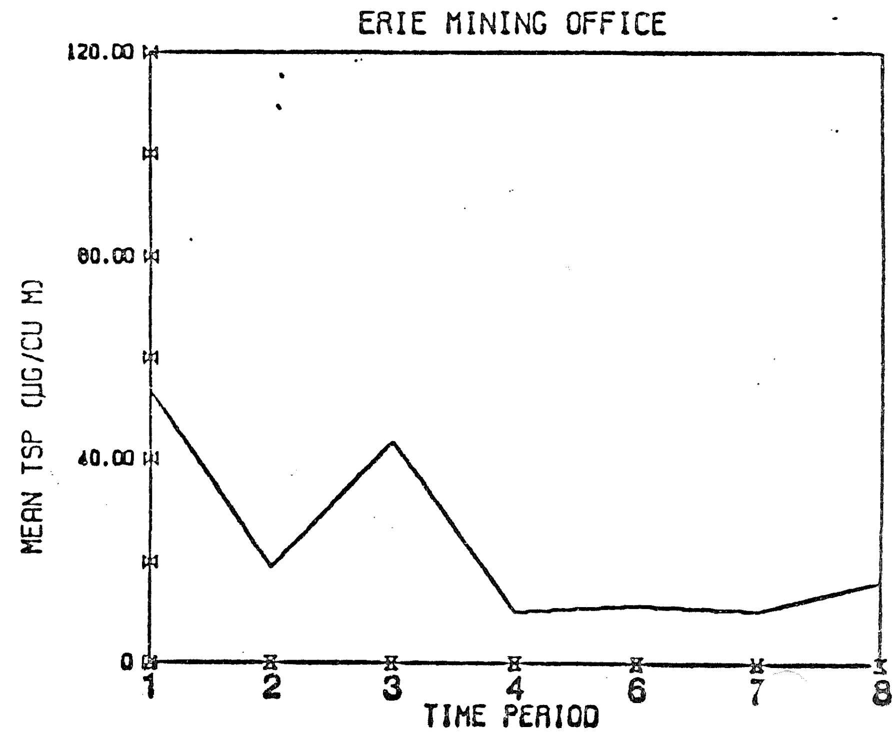
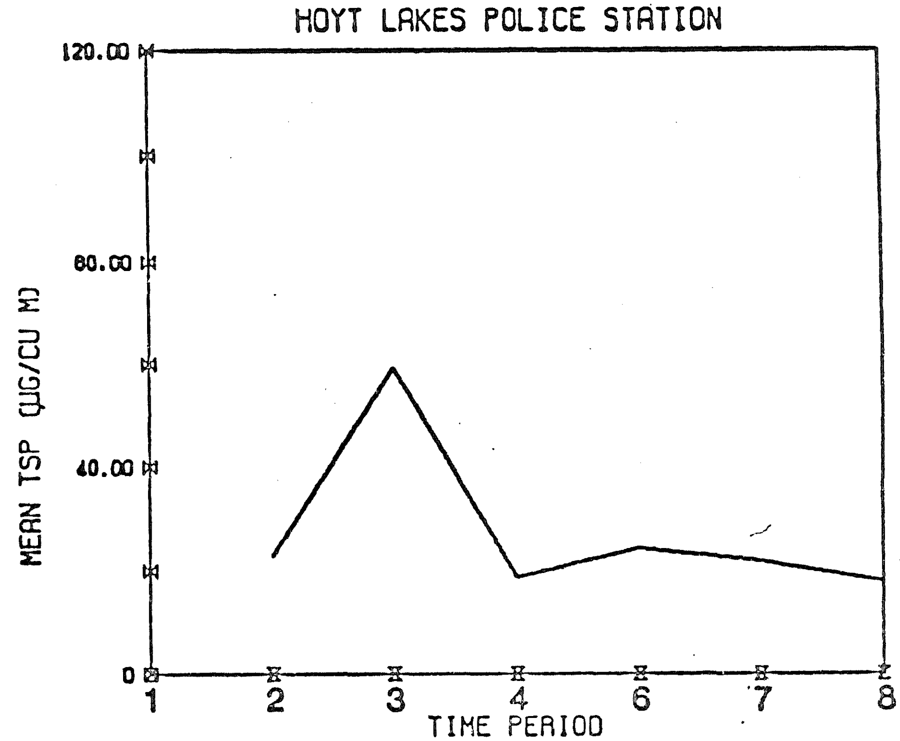
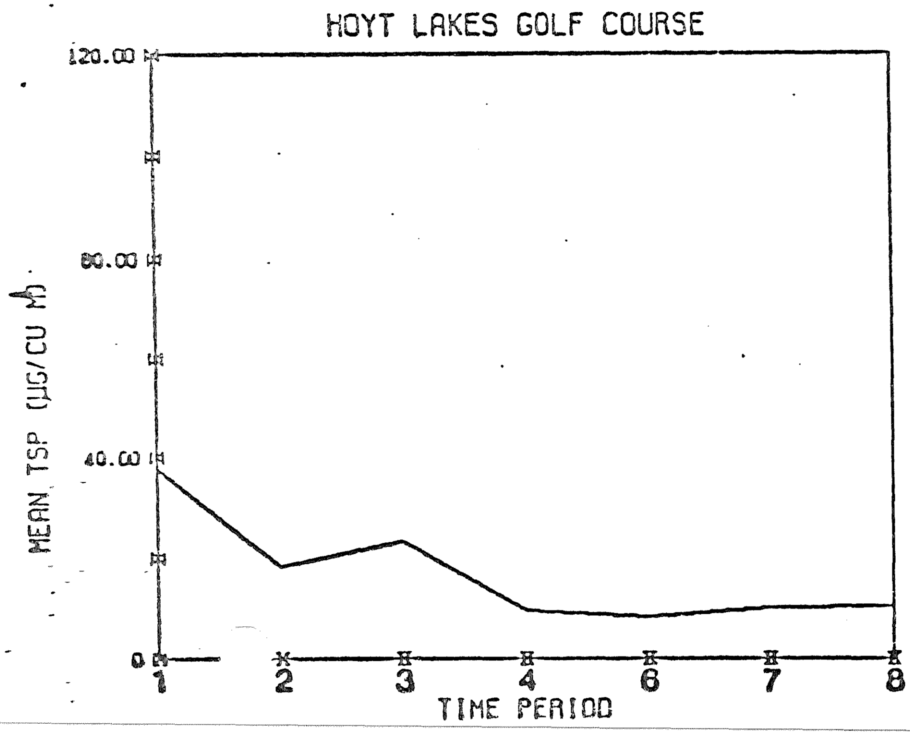


TABLE 3 MEAN TSP BY TIME PERIOD

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78





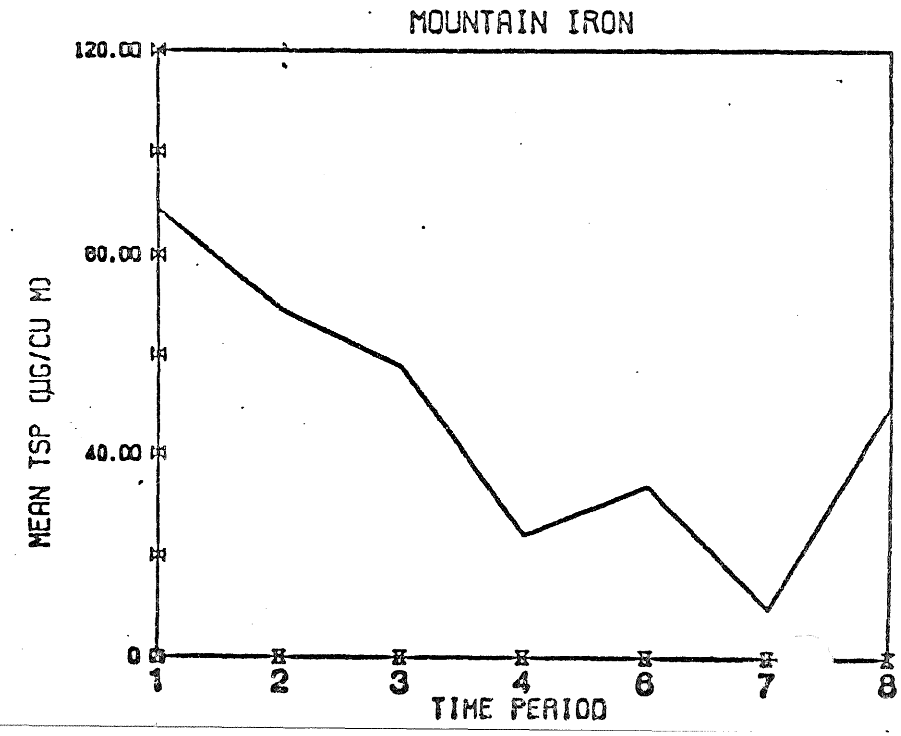
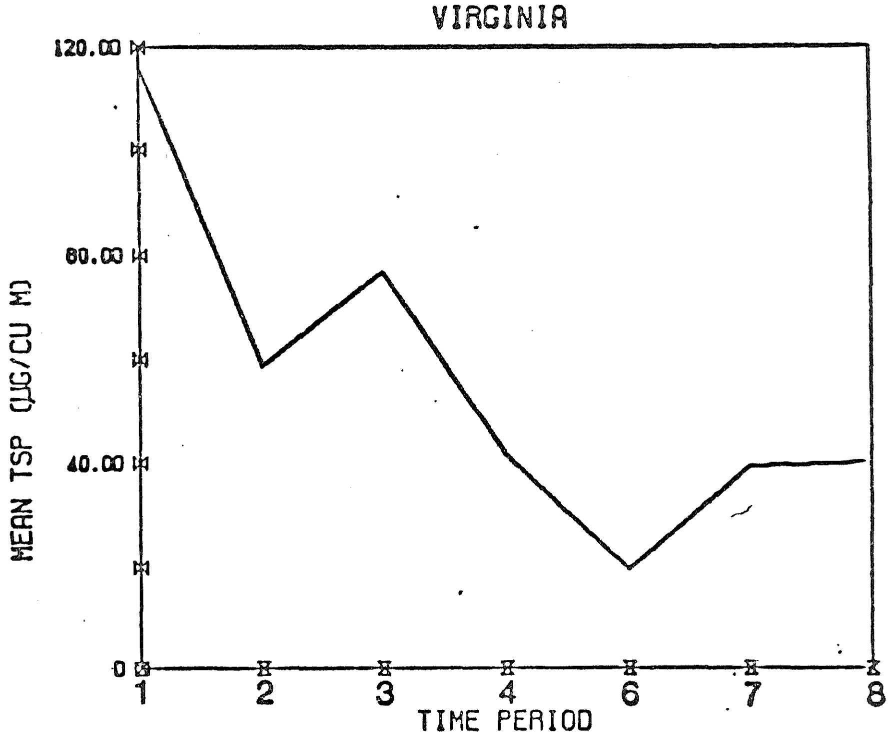
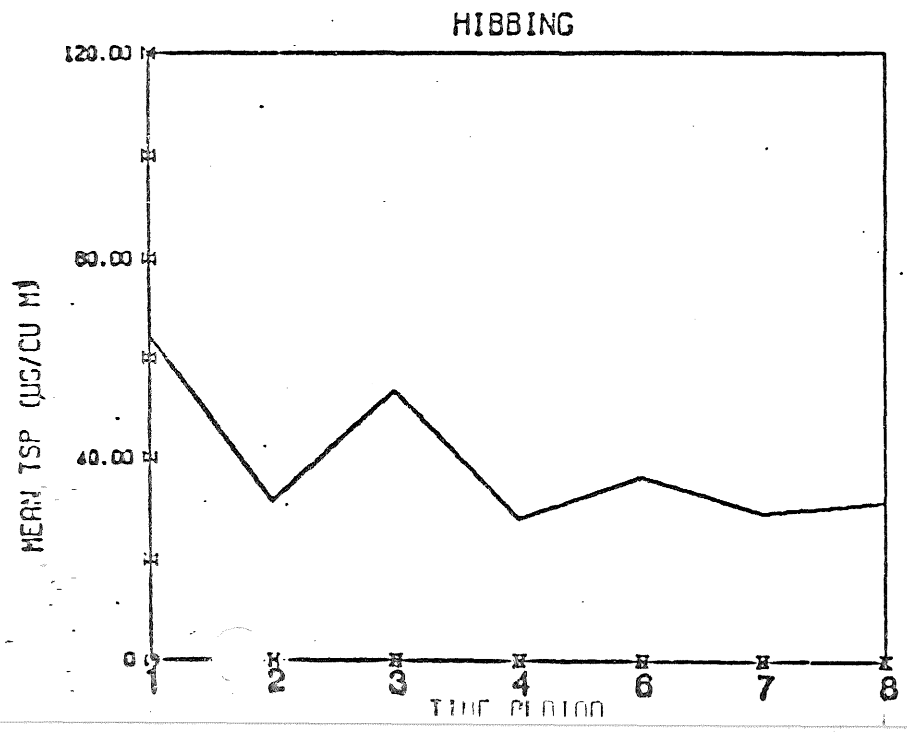


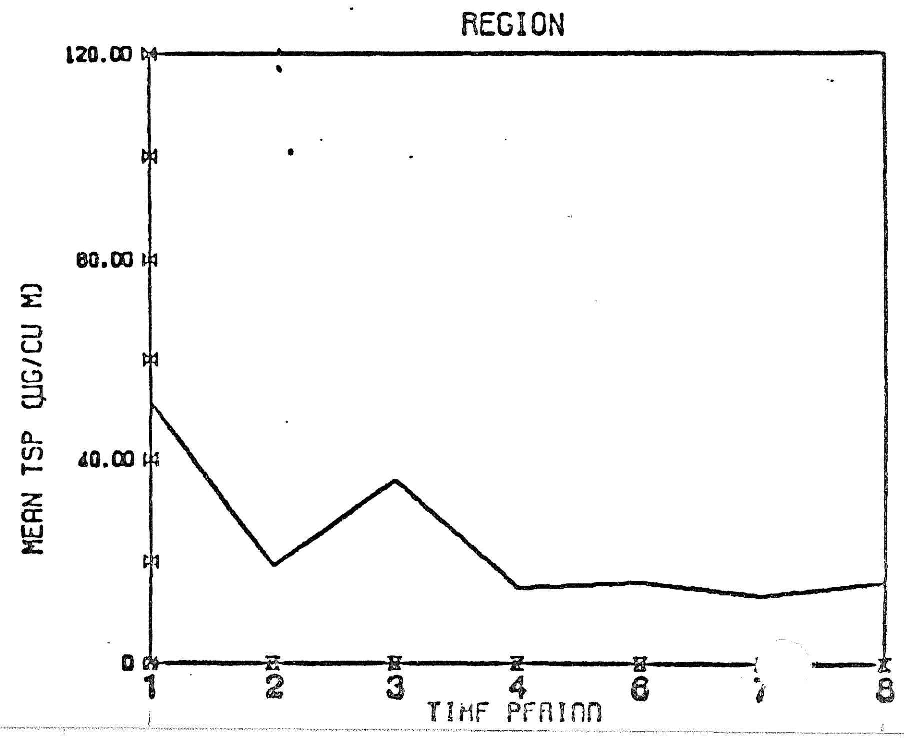
TABLE 3 MEAN TSP BY TIME PERIOD

LEGEND

TIME PERIOD	DATE
1	10/09/76-11/20/76
2	11/26/76-03/08/77
3	03/14/77-07/24/77
4	07/30/77-10/04/77
6	10/16/77-11/09/77
7	11/15/77-12/15/77
8	12/21/77-03/27/78

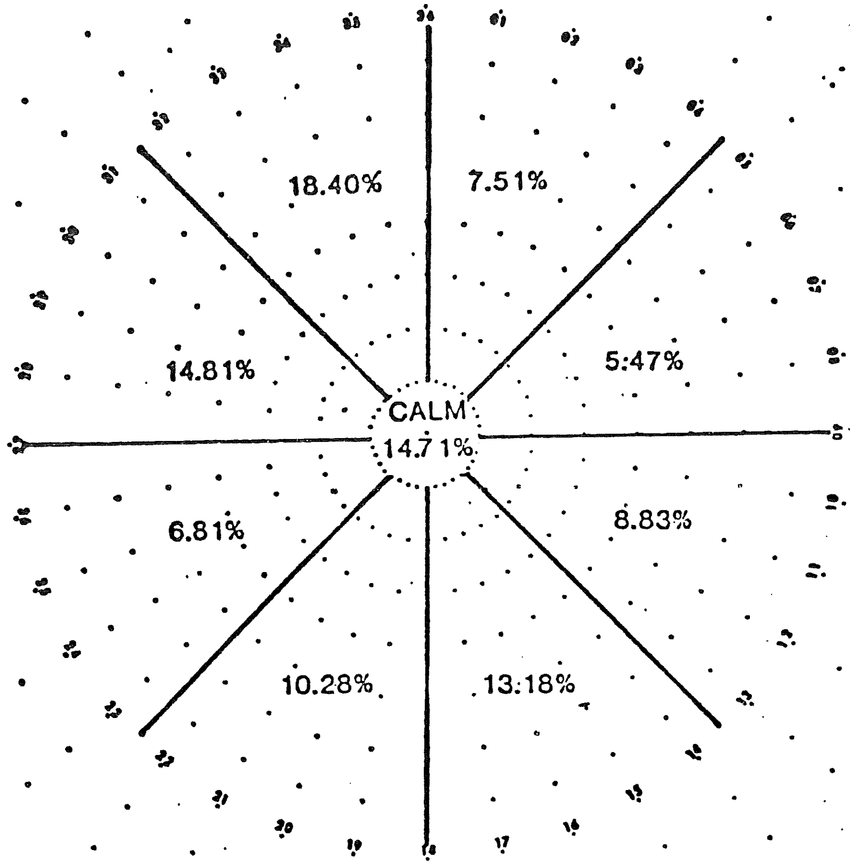




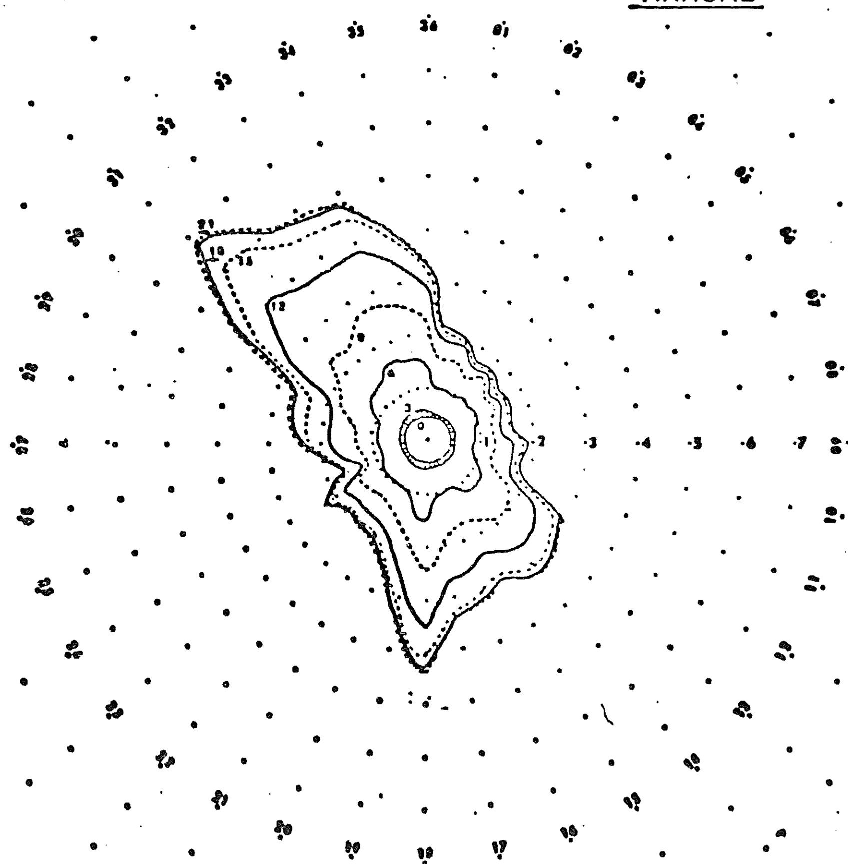


HIBBING 1964-1973

WIND DISTRIBUTION
% FREQUENCY BY OCTANTS

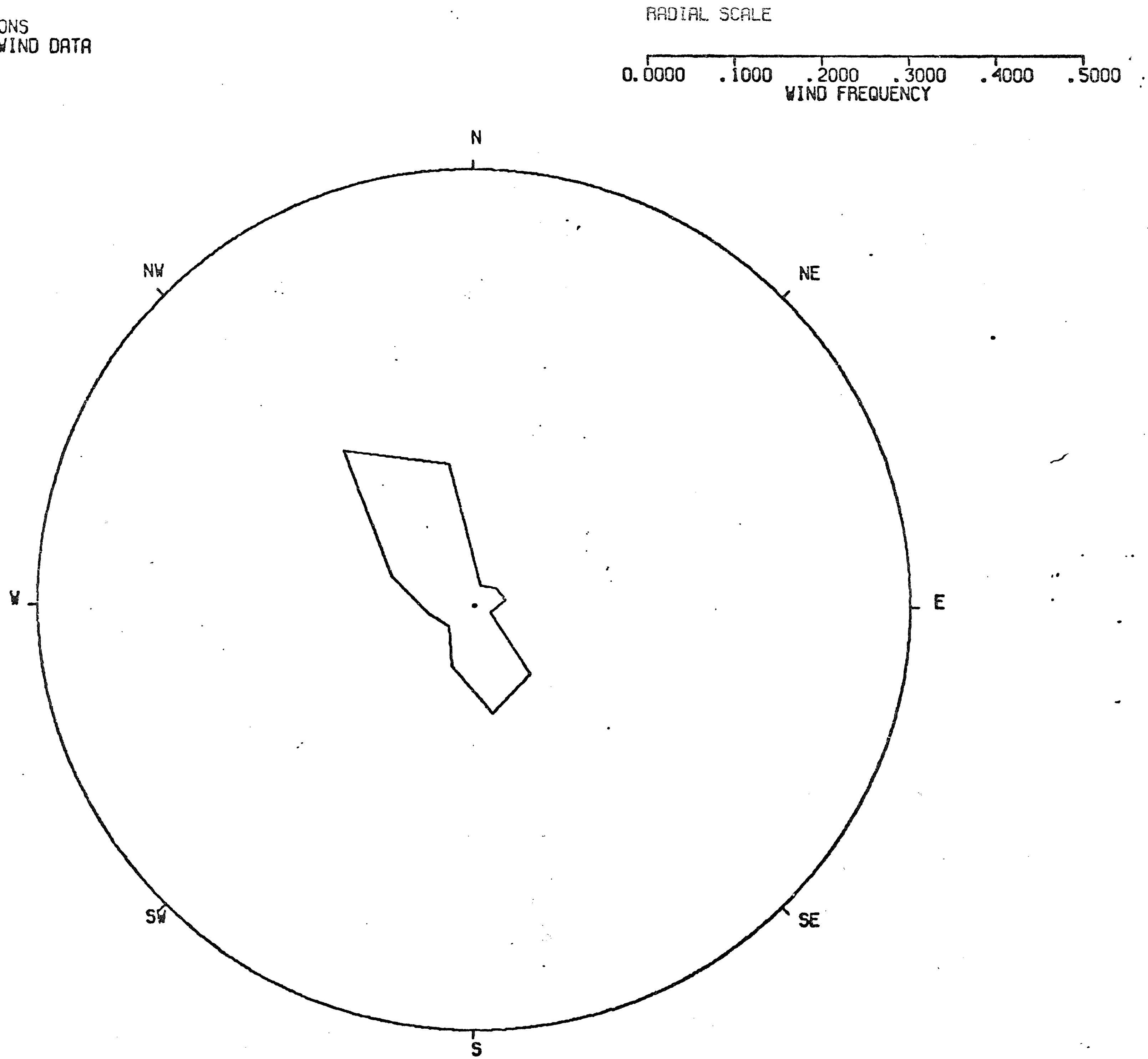


WIND ROSE
ANNUAL



Source: Watson (1978)

WIND ROSE FOR
TSP OBSERVATIONS
1977 HIBBING WIND DATA

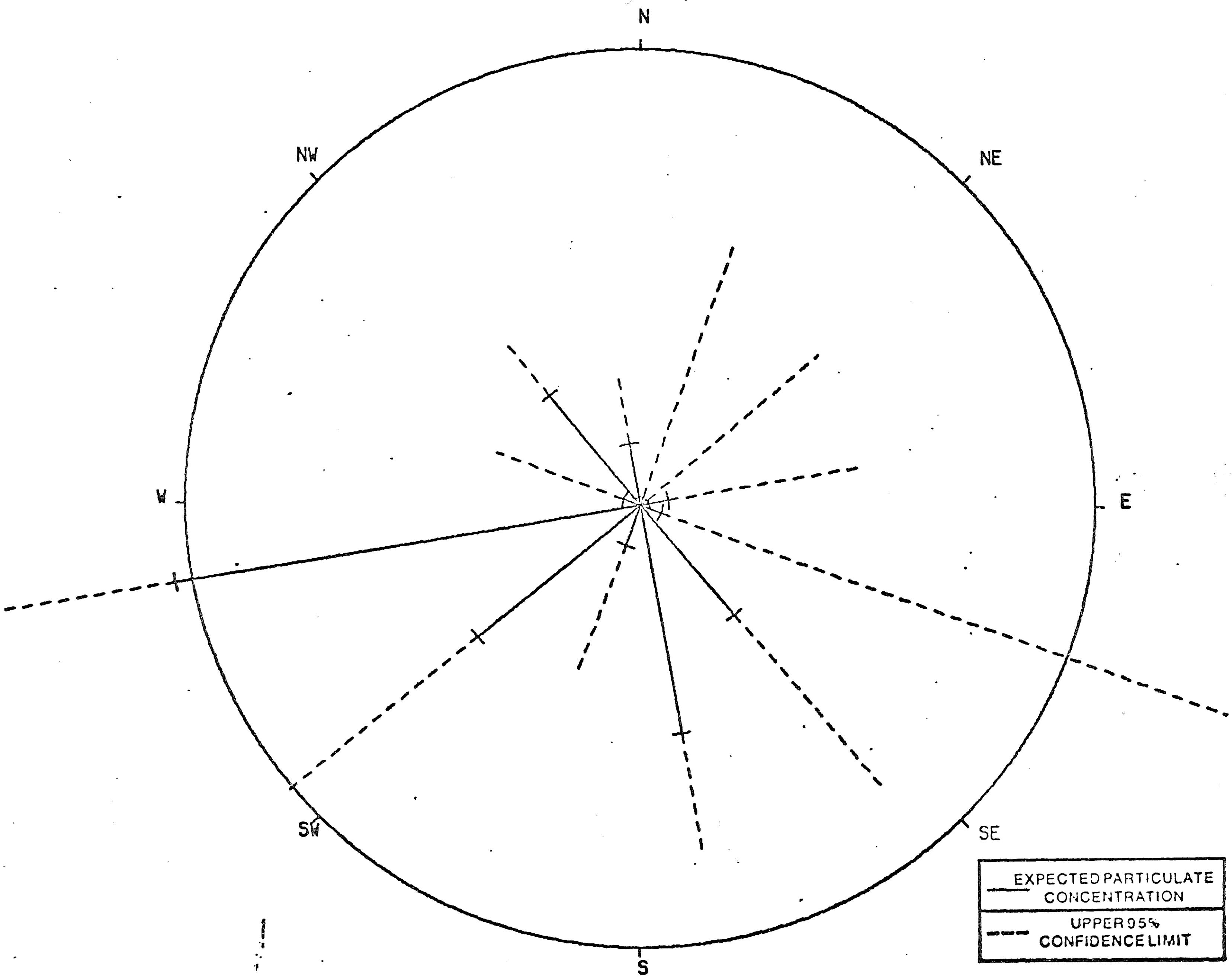


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
FERNBERG ROAD

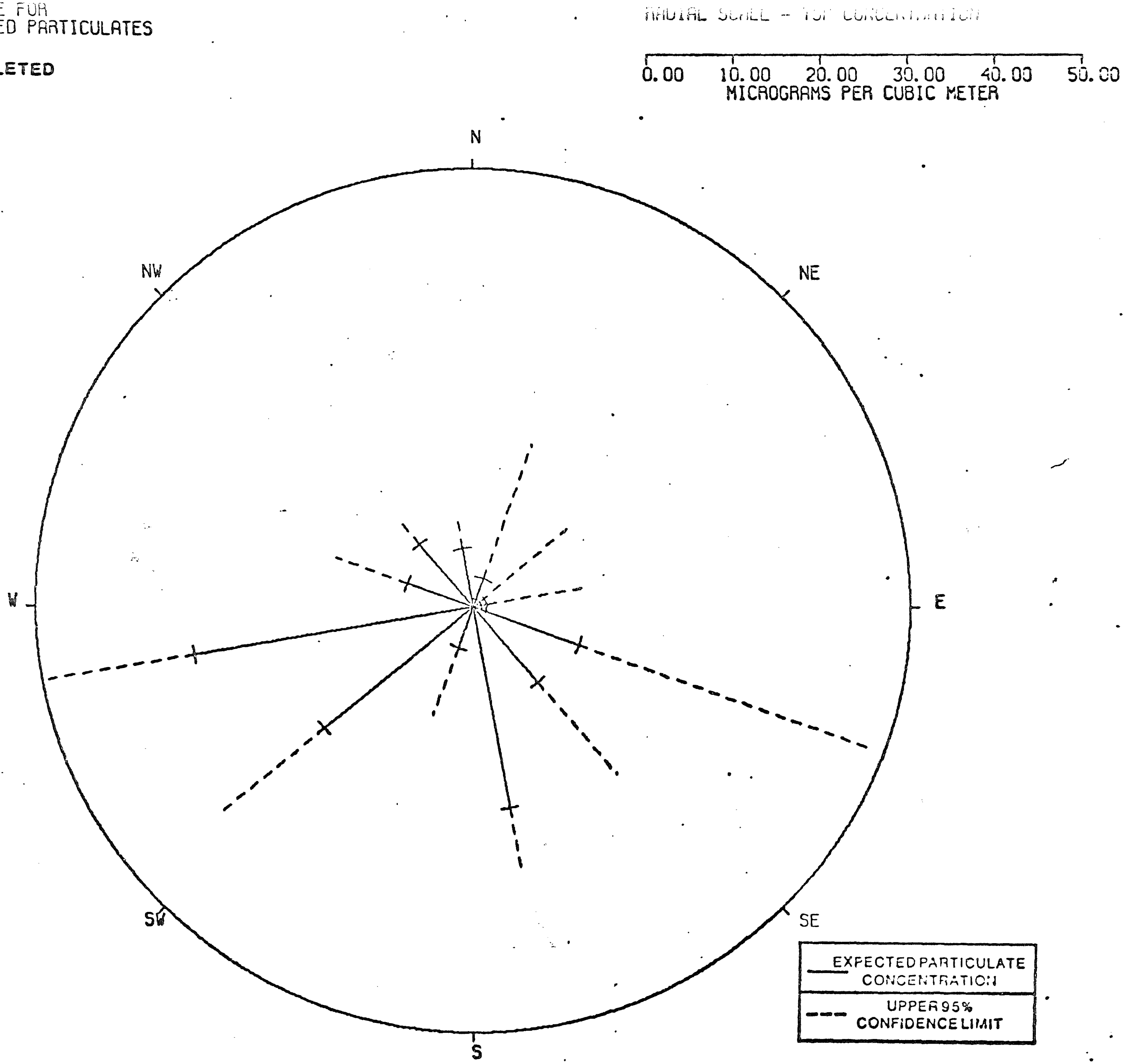
ALL VALUES INCLUDED

RADIAL SCALE - TSP CONCENTRATION

0.00 10.00 20.00 30.00 40.00 50.00
MICROGRAMS PER CUBIC METER

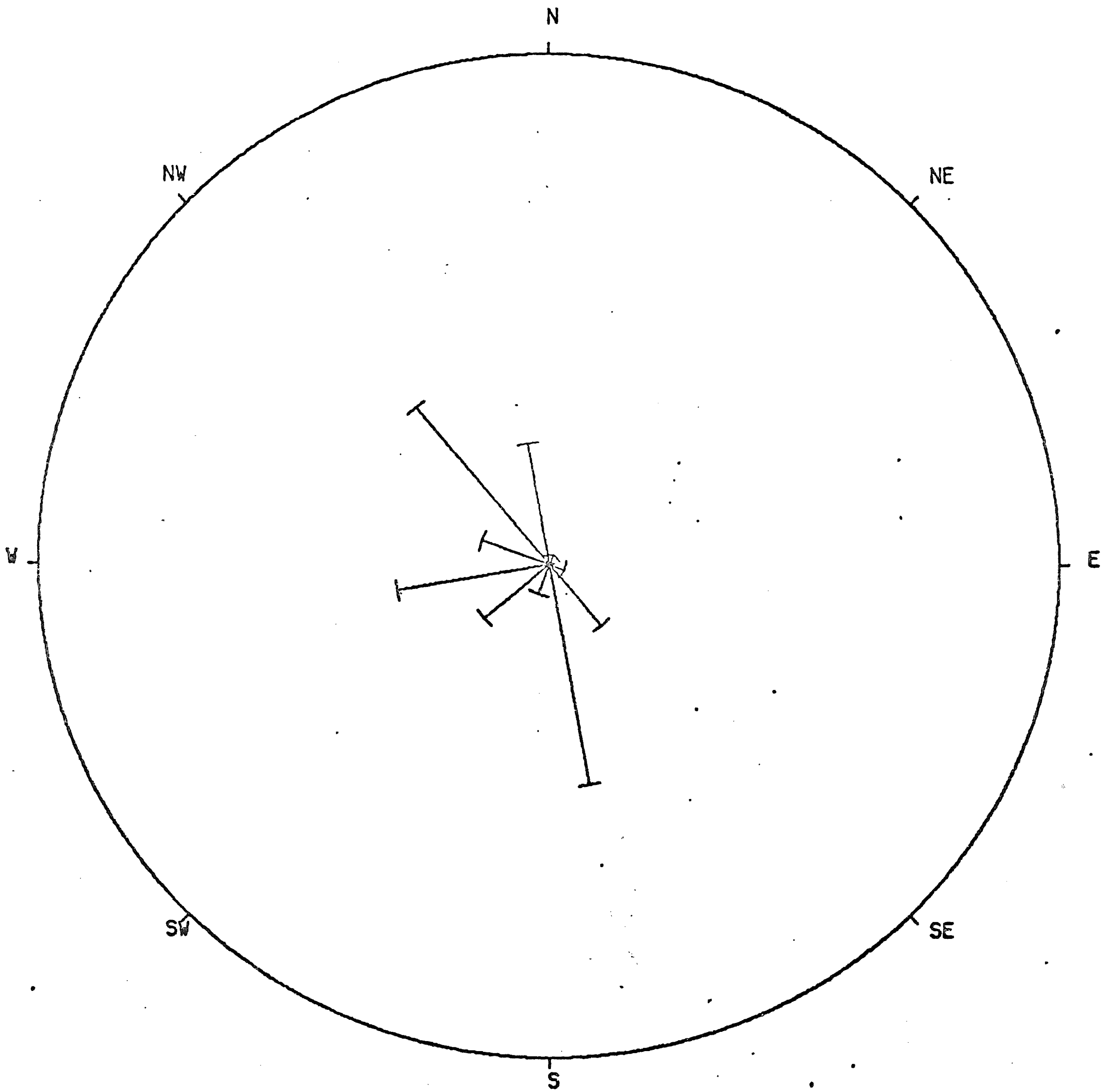


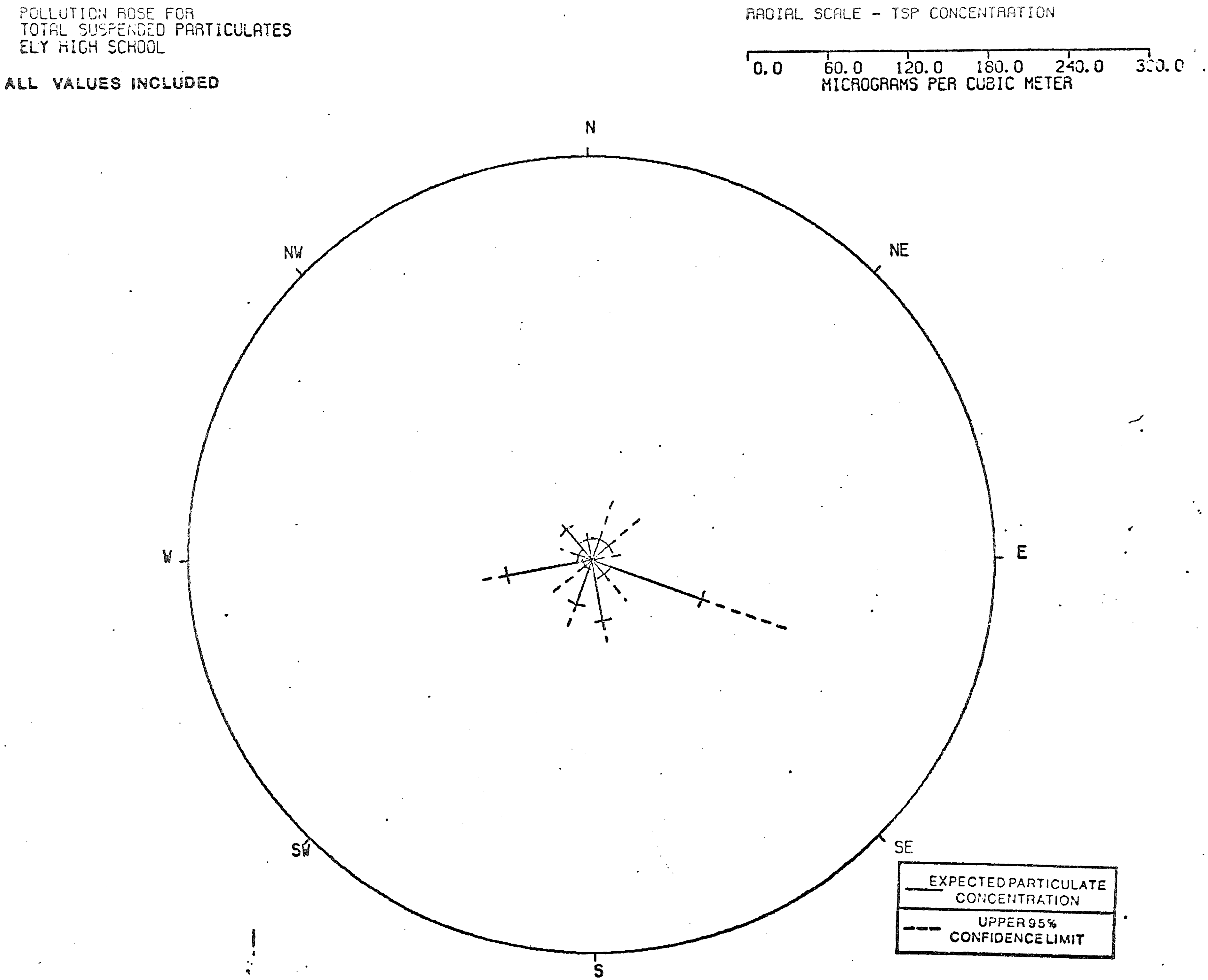
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
FERNBERG ROAD
OUTLIERS DELETED

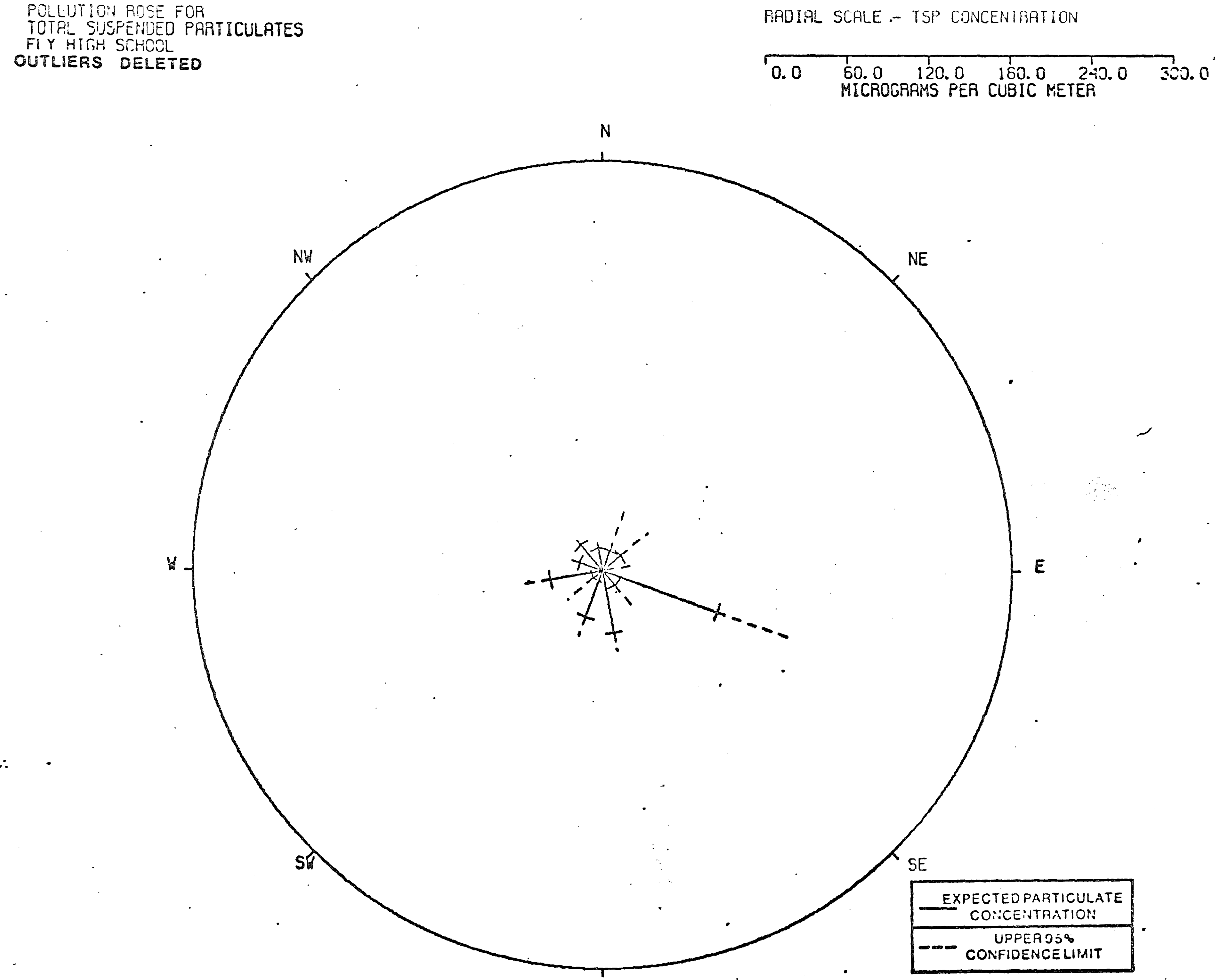


ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
FERNBERG ROAD

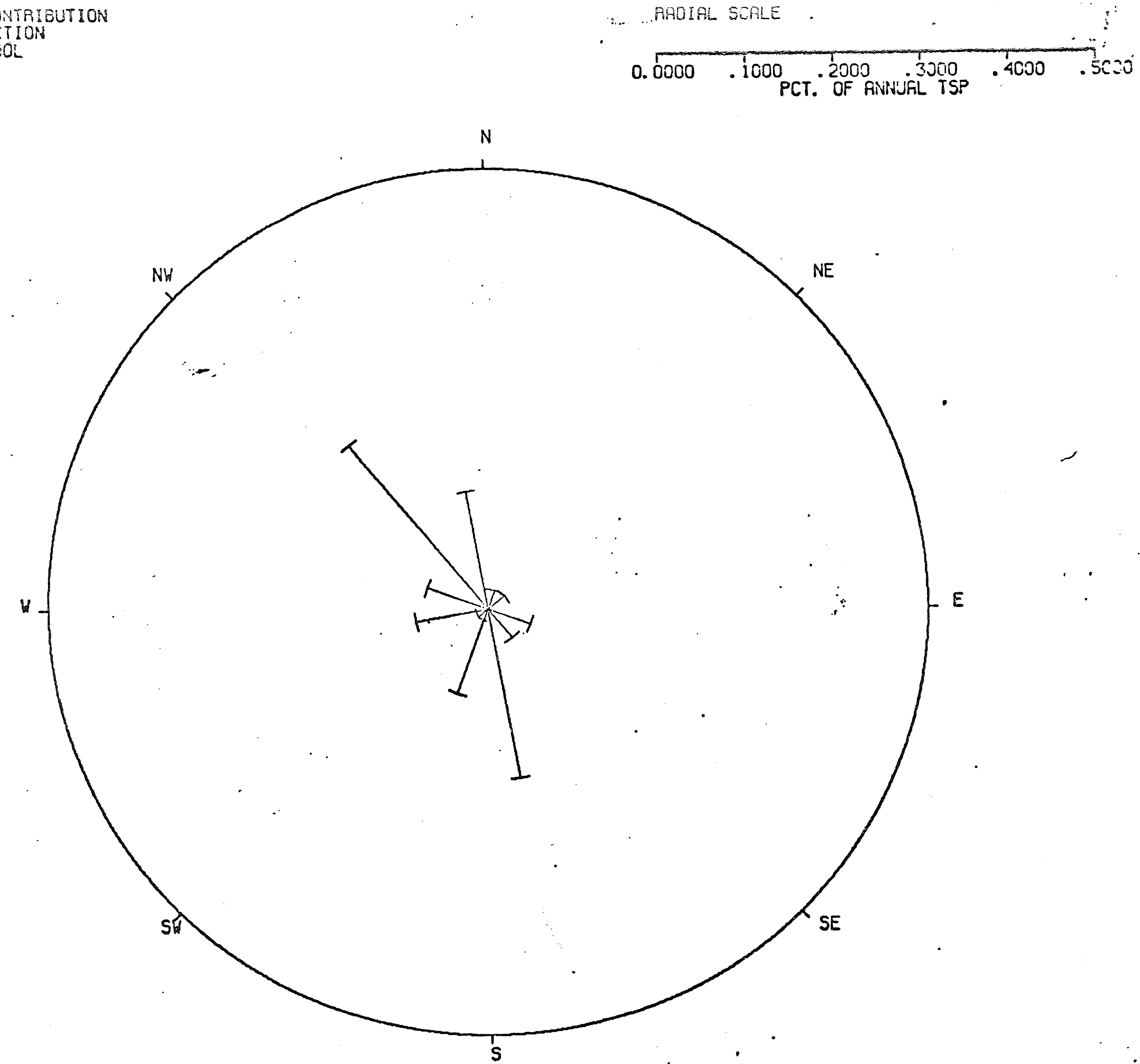
RADIAL SCALE.




POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
ELY HIGH SCHOOL
ALL VALUES INCLUDED
RADIAL SCALE - TSP CONCENTRATION
0.0
60.0
120.0
180.0
240.0
300.0
MICROGRAMS PER CUBIC METER
N
NE
E
SE
S
SW
W
NW
EXPECTED PARTICULATE CONCENTRATION
UPPER 95% CONFIDENCE LIMIT


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
FLY HIGH SCHOOL
OUTLIERS DELETED
RADIAL SCALE - TSP CONCENTRATION
0.0
60.0
120.0
180.0
240.0
300.0
MICROGRAMS PER CUBIC METER
N
NE
E
SE
S
SW
W
NW
EXPECTED PARTICULATE CONCENTRATION
UPPER 95% CONFIDENCE LIMIT

ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
ELY HIGH SCHOOL

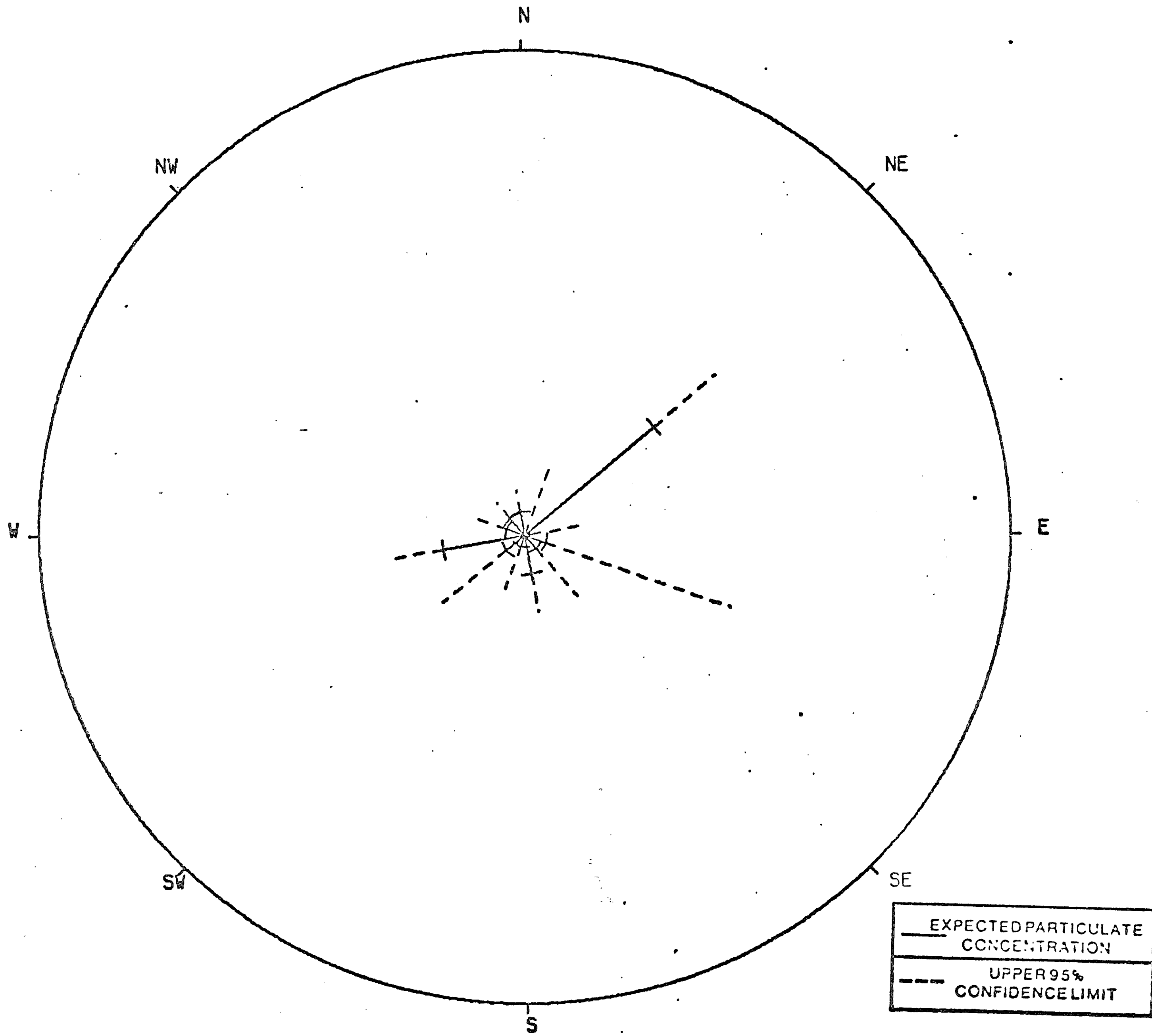


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
KAWISHIWI LABORATORY

ALL VALUES INCLUDED

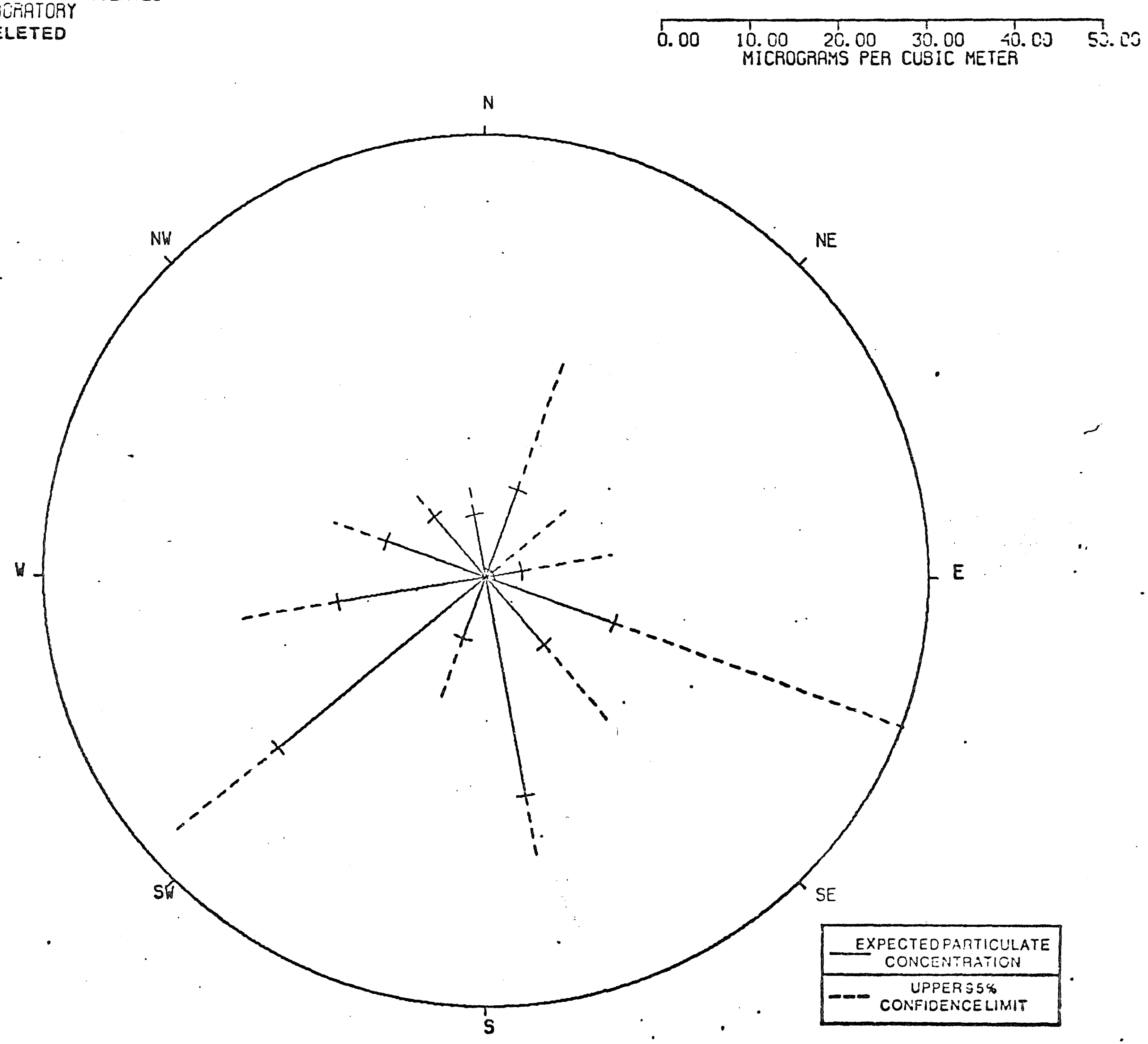
RADIAL SCALE - TSP CONCENTRATION

0.0 60.0 120.0 180.0 240.0 300.0
MICROGRAMS PER CUBIC METER



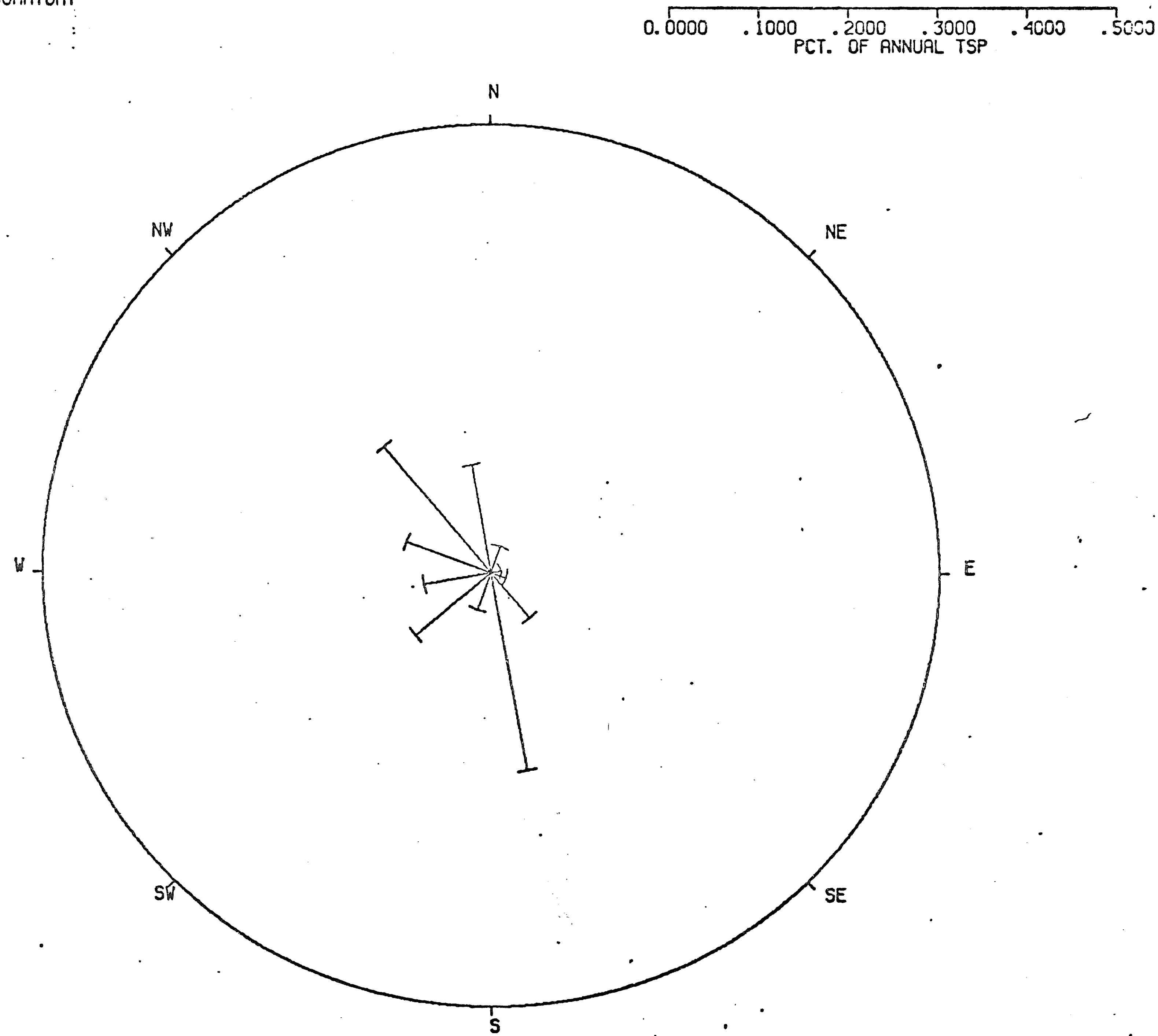
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
KAWISHIWI LABORATORY
OUTLIERS DELETED

RADIAL SCALE = TSP CONCENTRATION



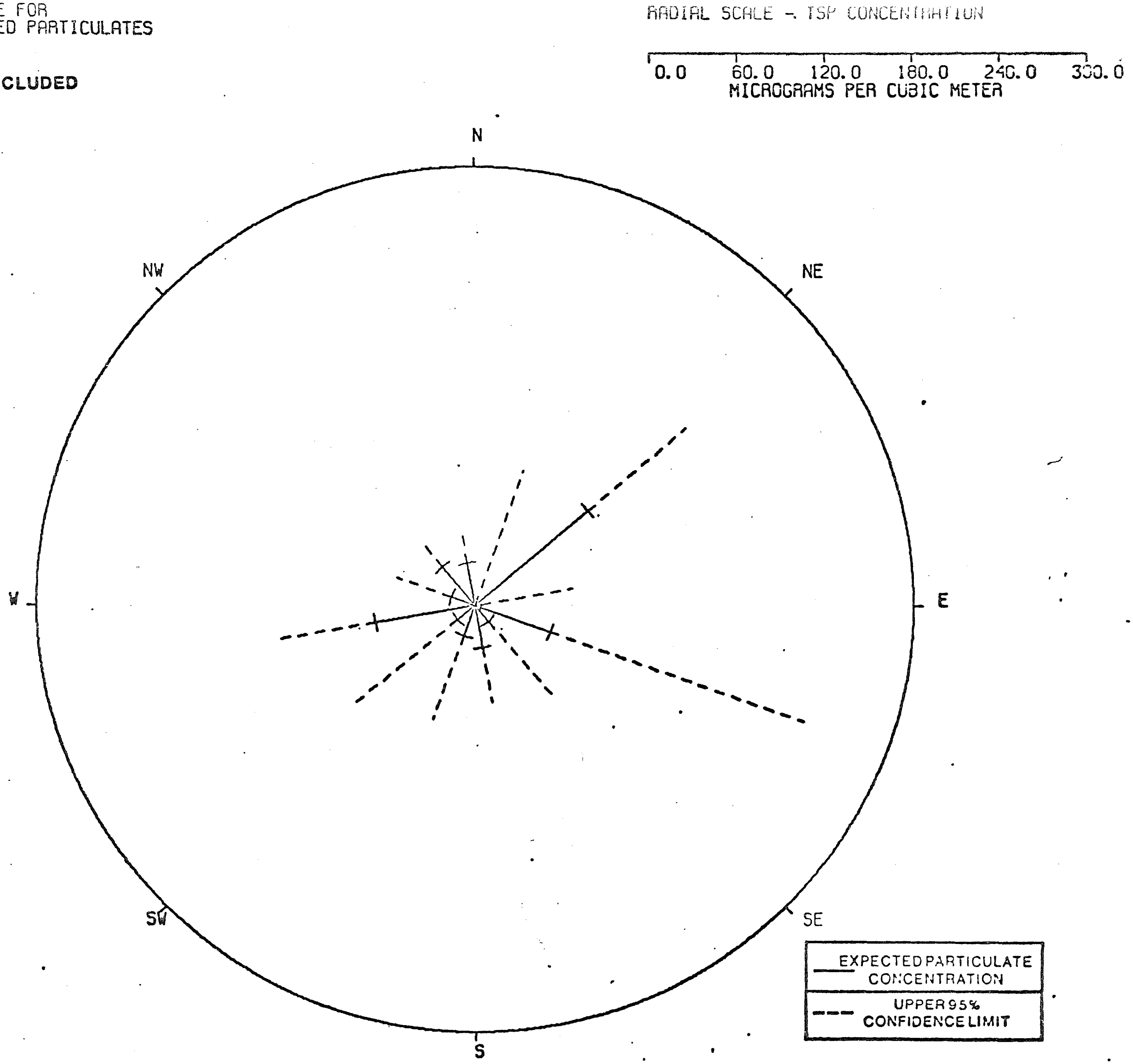
ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
KAWISHIWI LABORATORY

RADIAL SCALE.



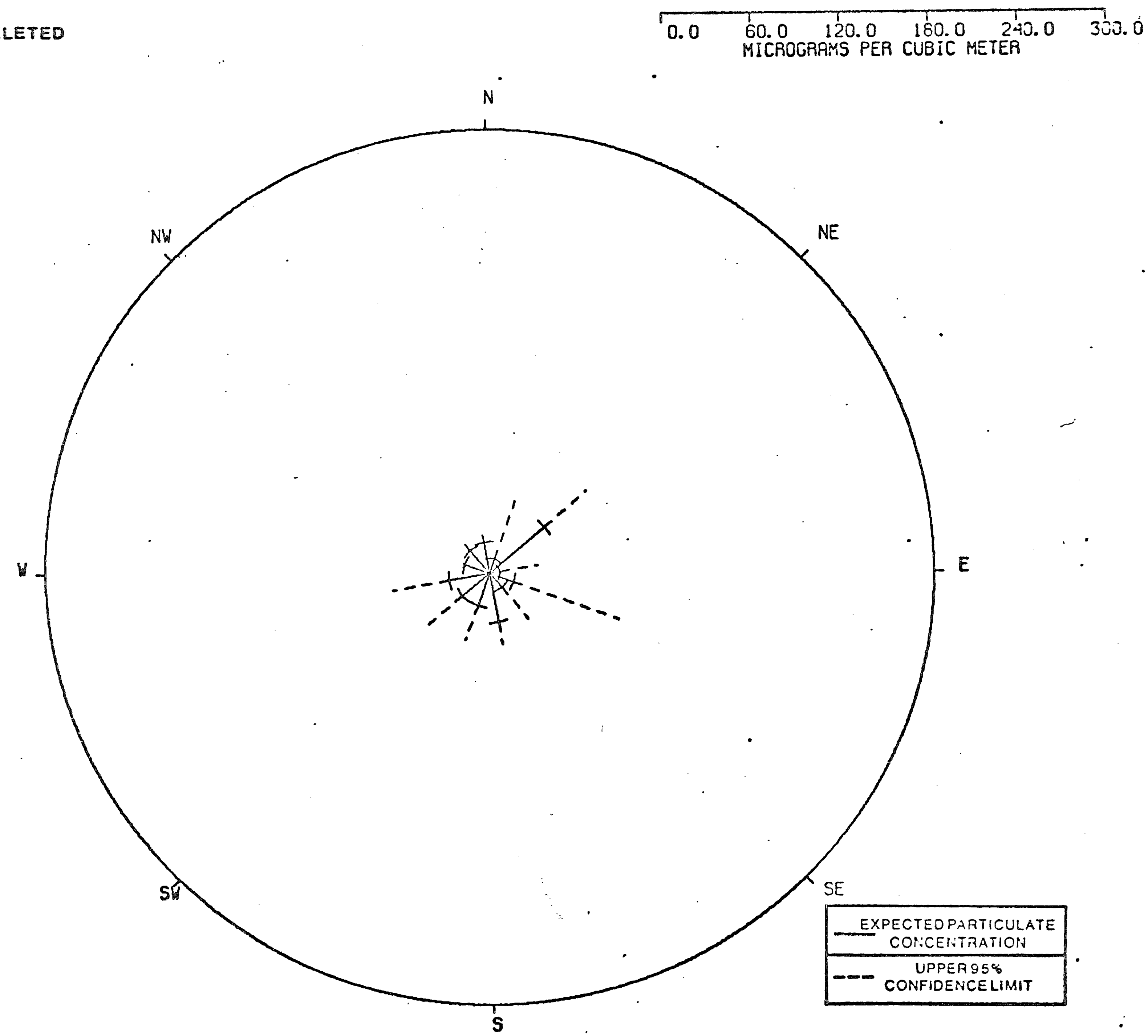
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
DUNKA ROAD

ALL VALUES INCLUDED

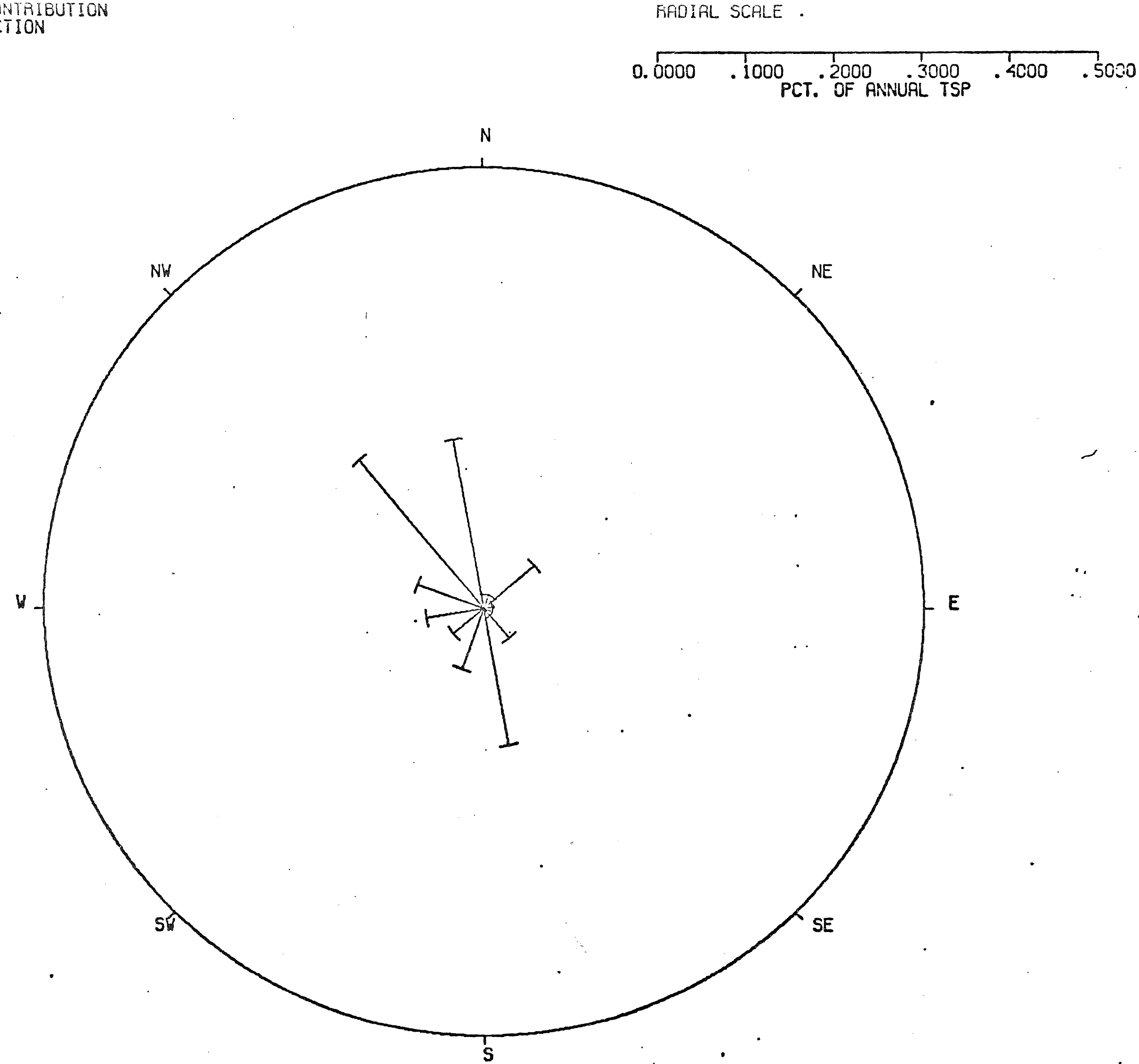


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
DUNKA ROAD
OUTLIERS DELETED

RADIAL SCALE - TSP CONCENTRATION

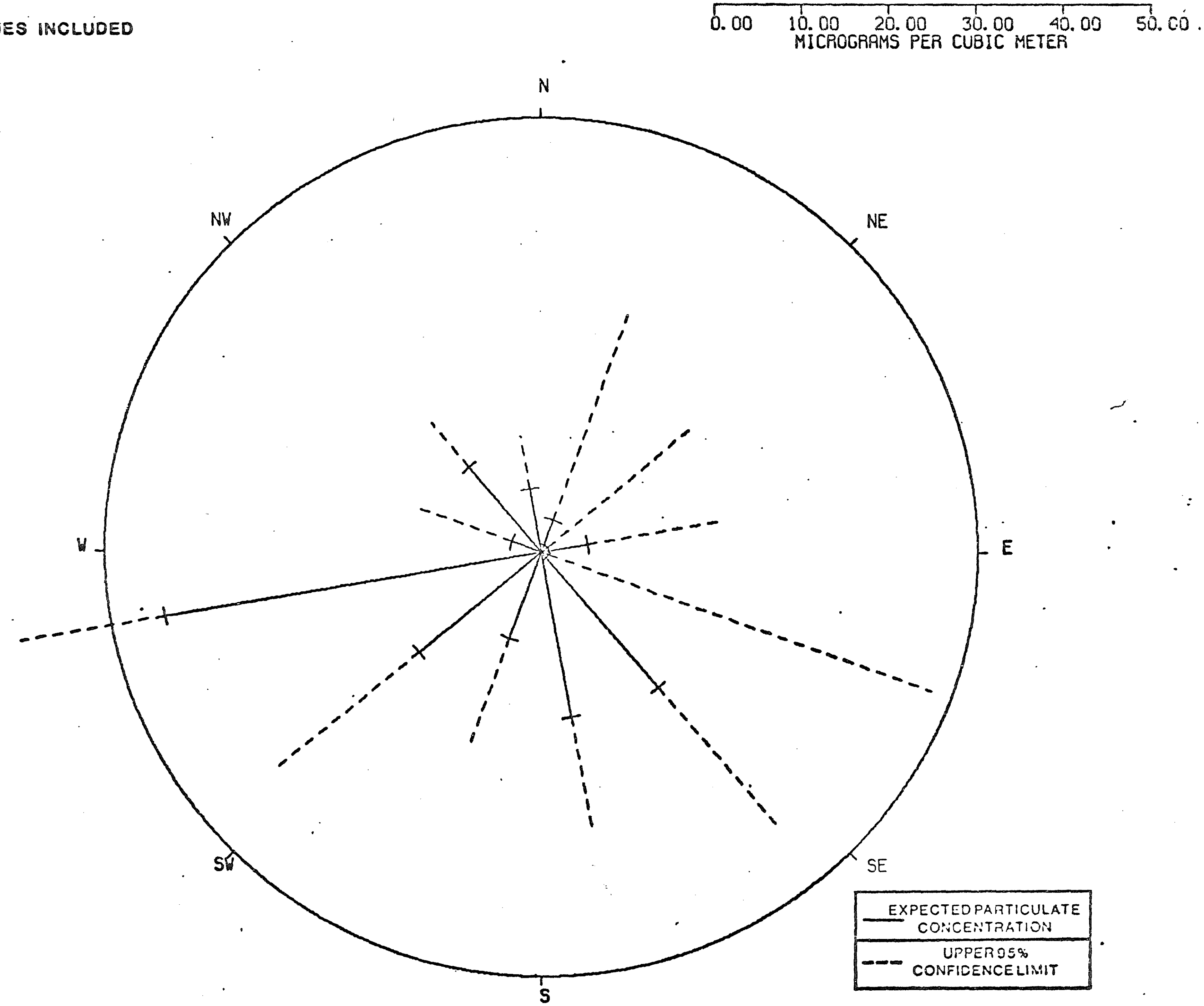


ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
DUNKA ROAD



POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
TOIMI

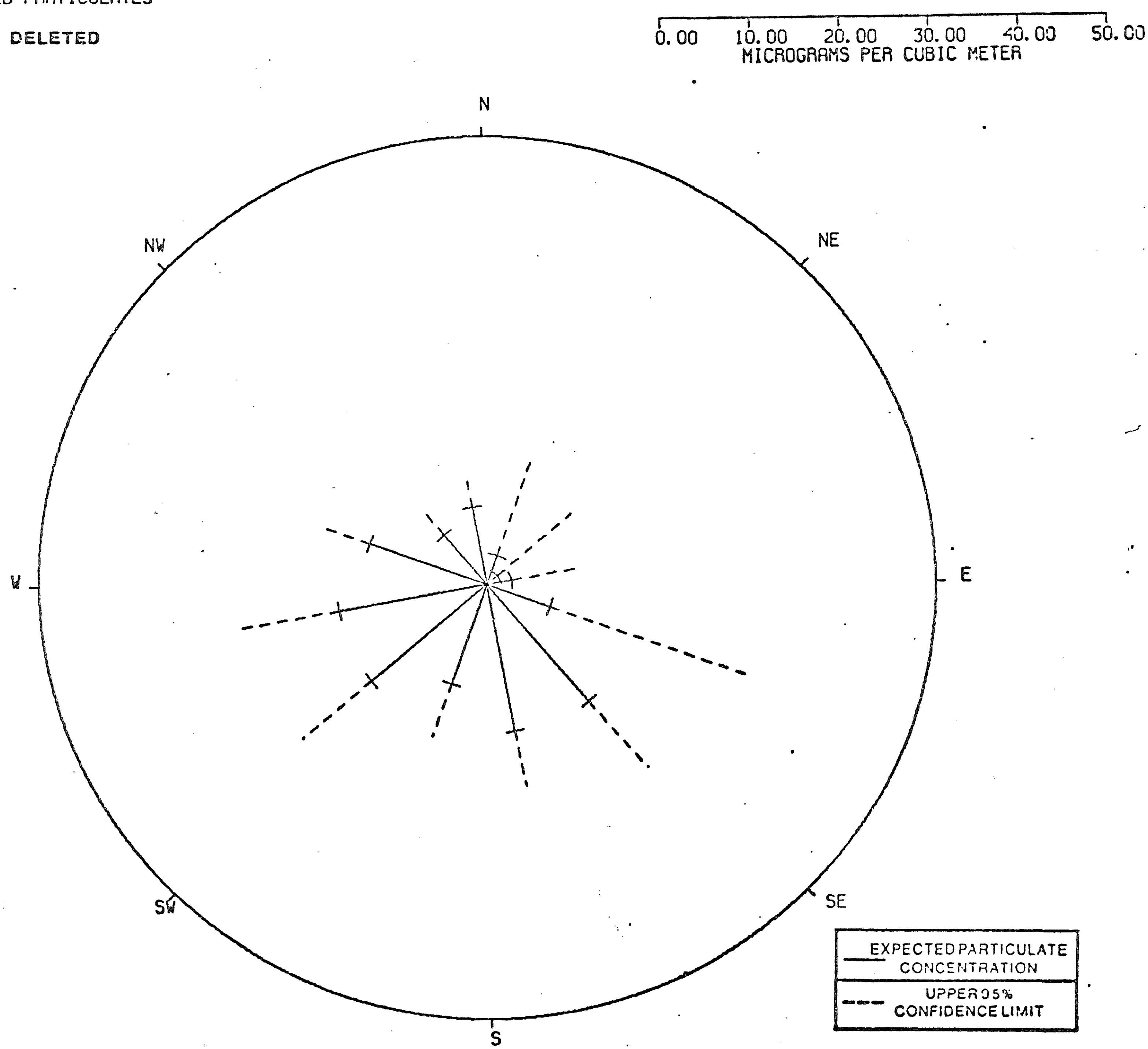
ALL VALUES INCLUDED



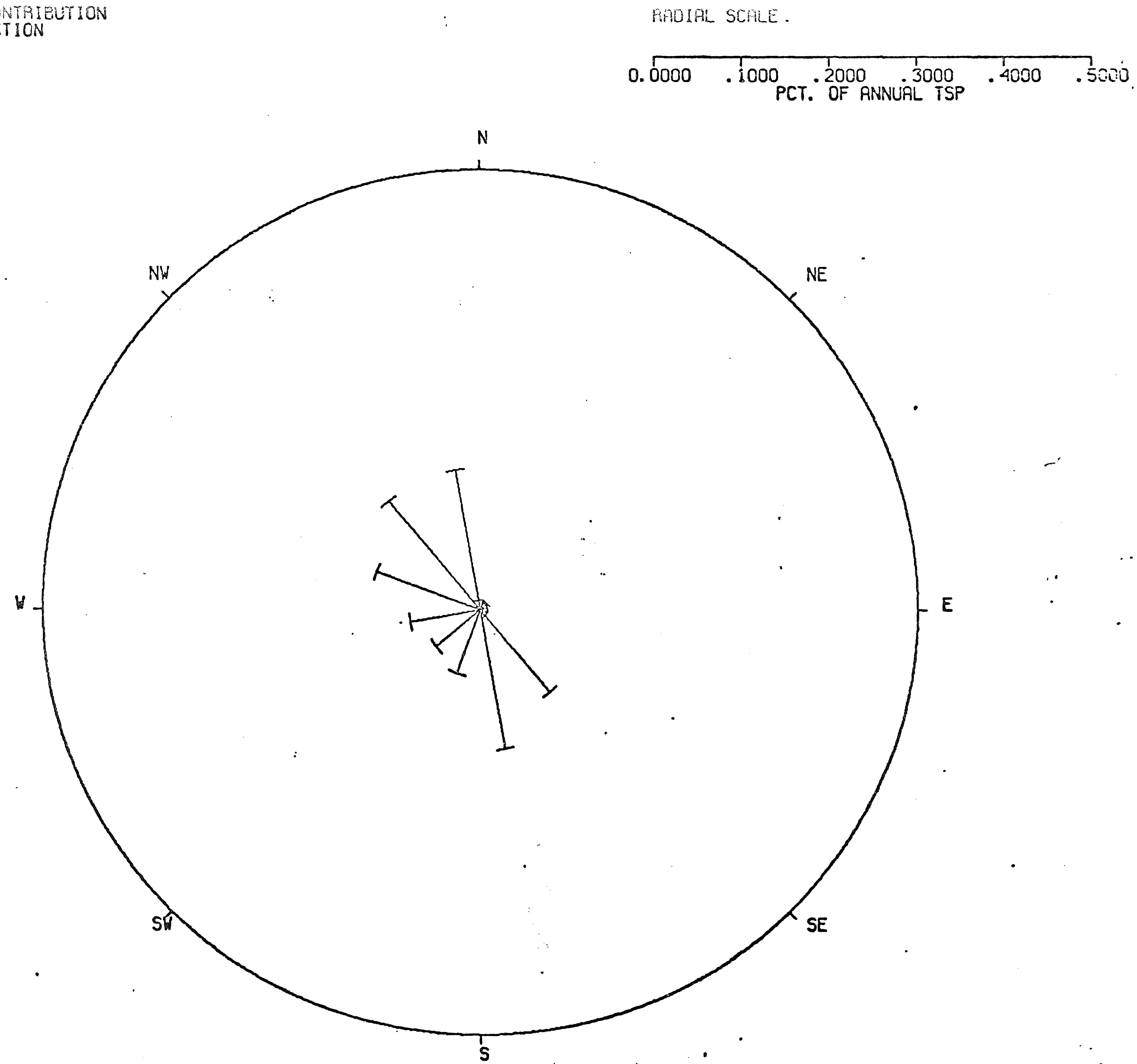
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
TOIMI

OUTLIERS DELETED

RADIAL SCALE - TSP CONCENTRATION

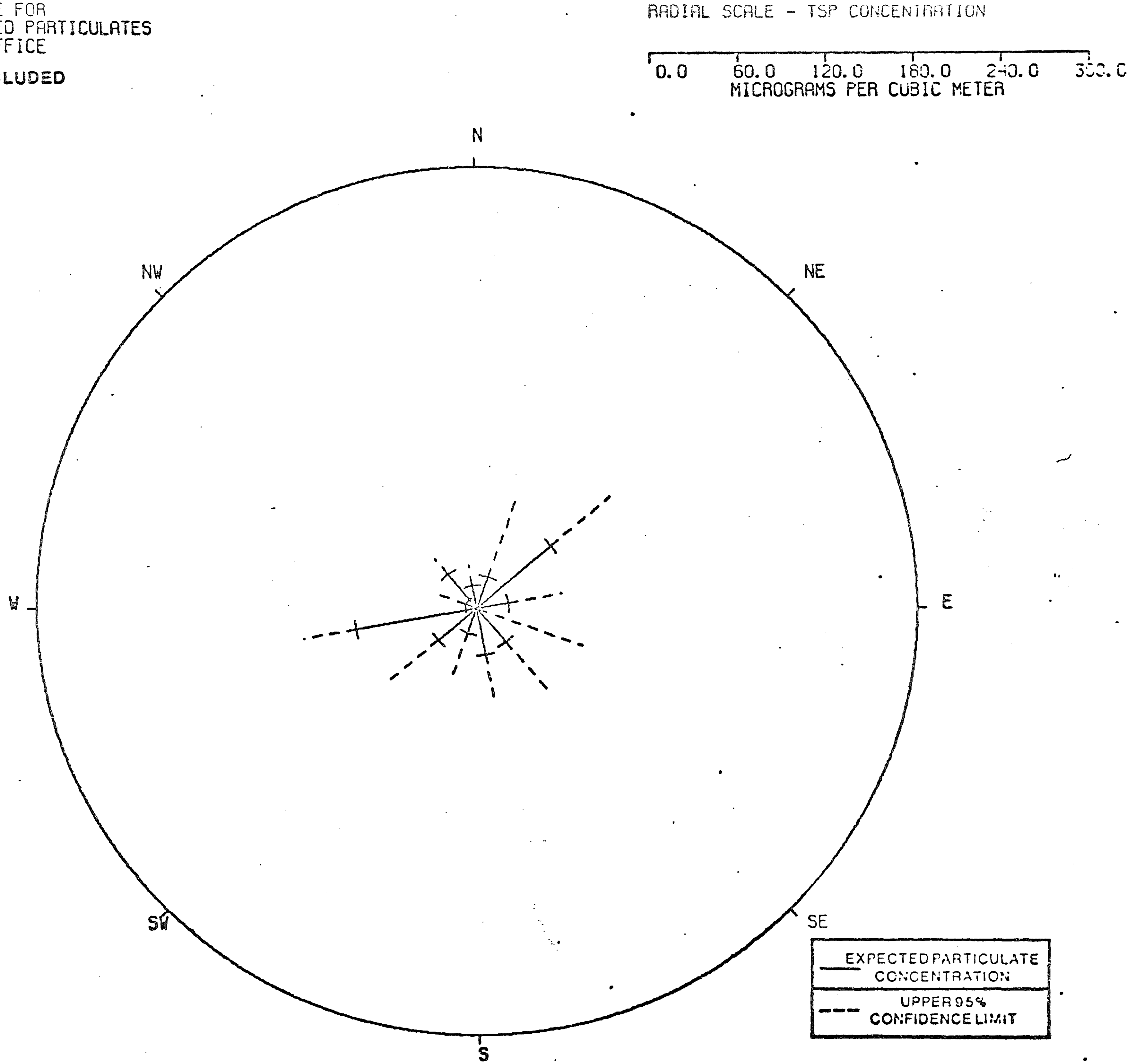


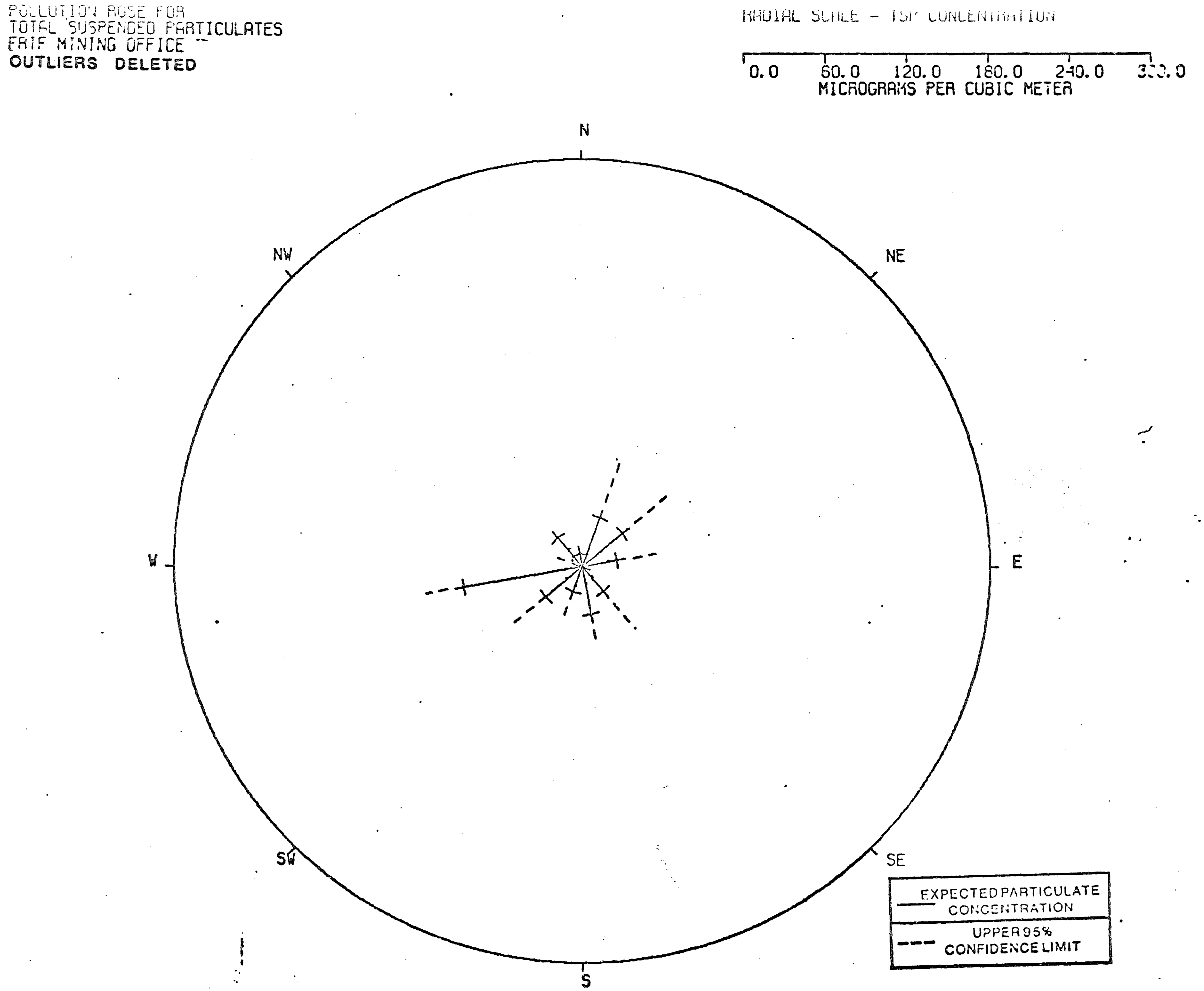
ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
TOIMI



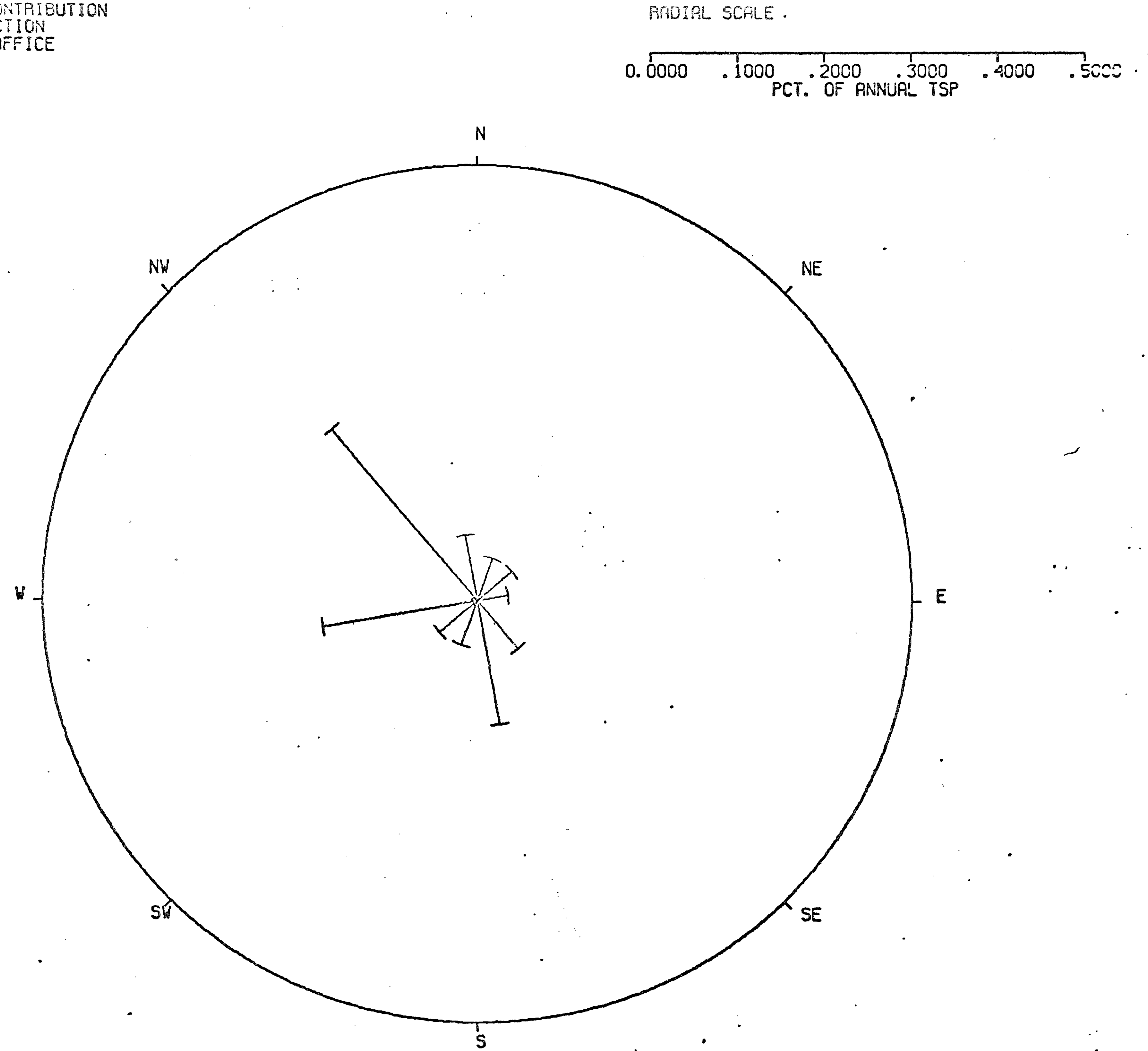
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
ERIE MINING OFFICE

ALL VALUES INCLUDED




POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
ERIE MINING OFFICE
OUTLIERS DELETED
RADIAL SCALE - TSP CONCENTRATION
0.0
60.0
120.0
180.0
240.0
MICROGRAMS PER CUBIC METER
N
NE
E
SE
S
SW
W
NW
EXPECTED PARTICULATE CONCENTRATION
UPPER 95% CONFIDENCE LIMIT

ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
ERIE MINING OFFICE

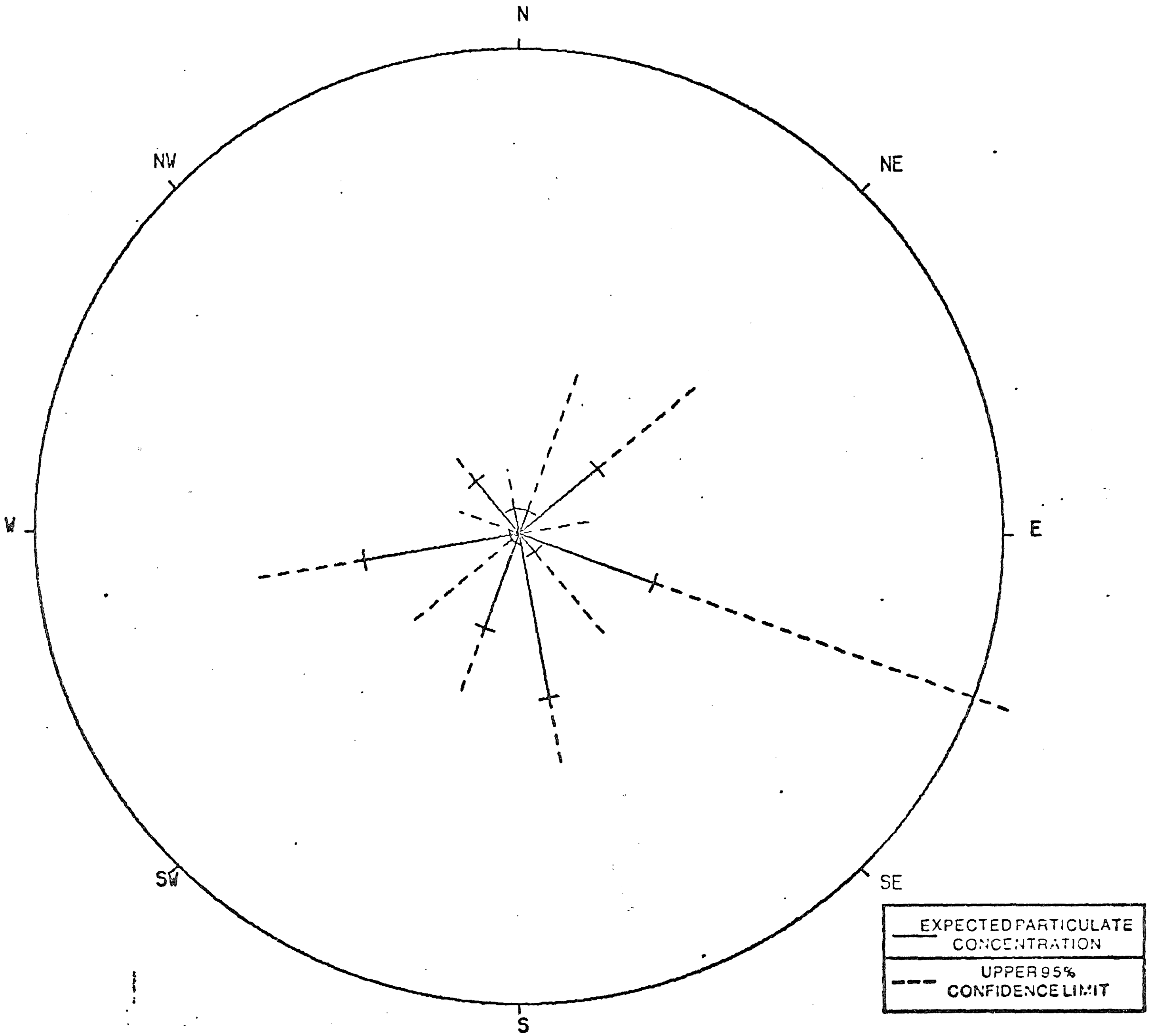


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HOYT LAKES POLICE STATION

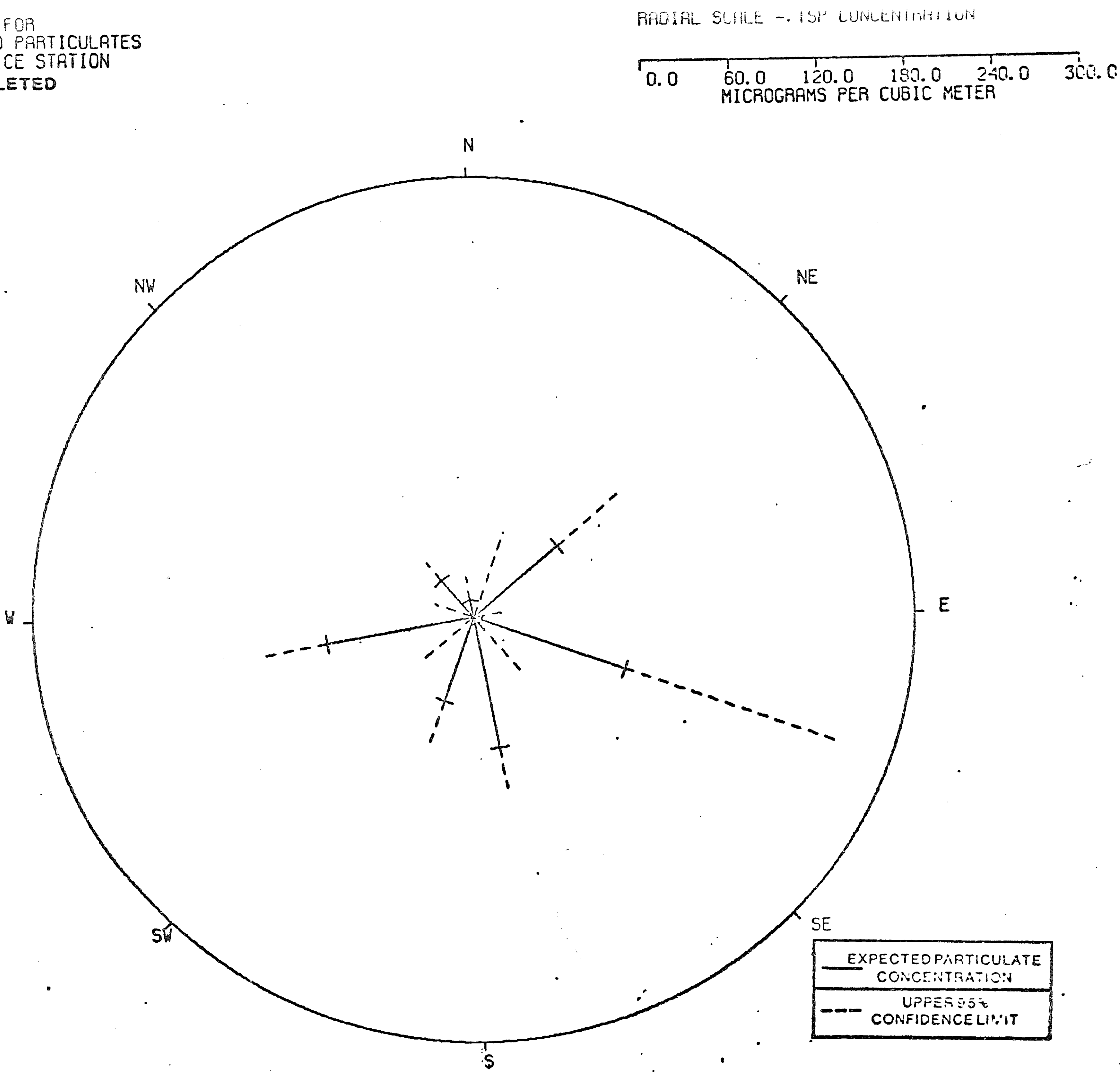
ALL VALUES INCLUDED

RADIAL SCALE - TSP CONCENTRATION

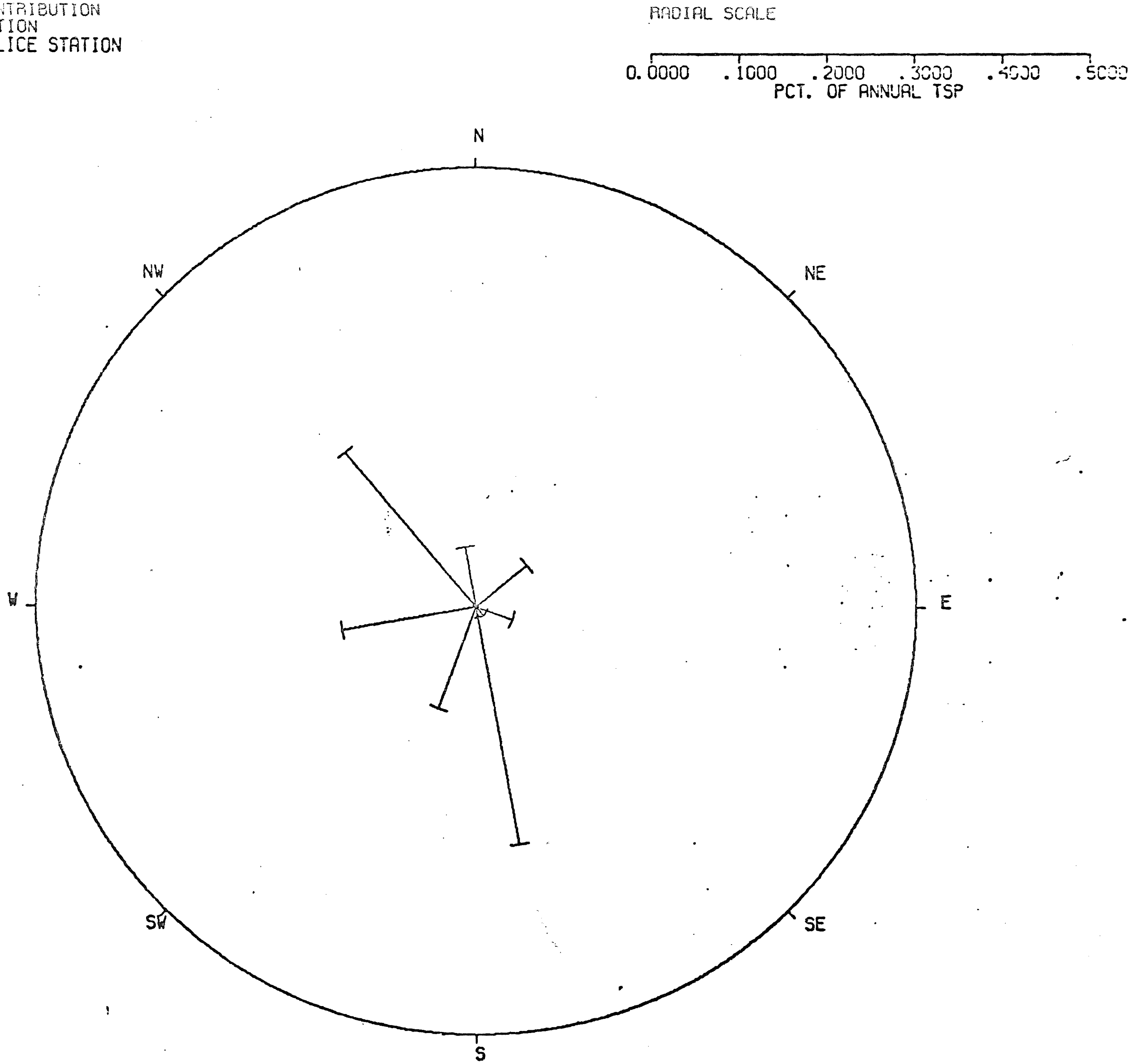
0.0 60.0 120.0 180.0 240.0 300.0
MICROGRAMS PER CUBIC METER



POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HOYT LAKES POLICE STATION
OUTLIERS DELETED

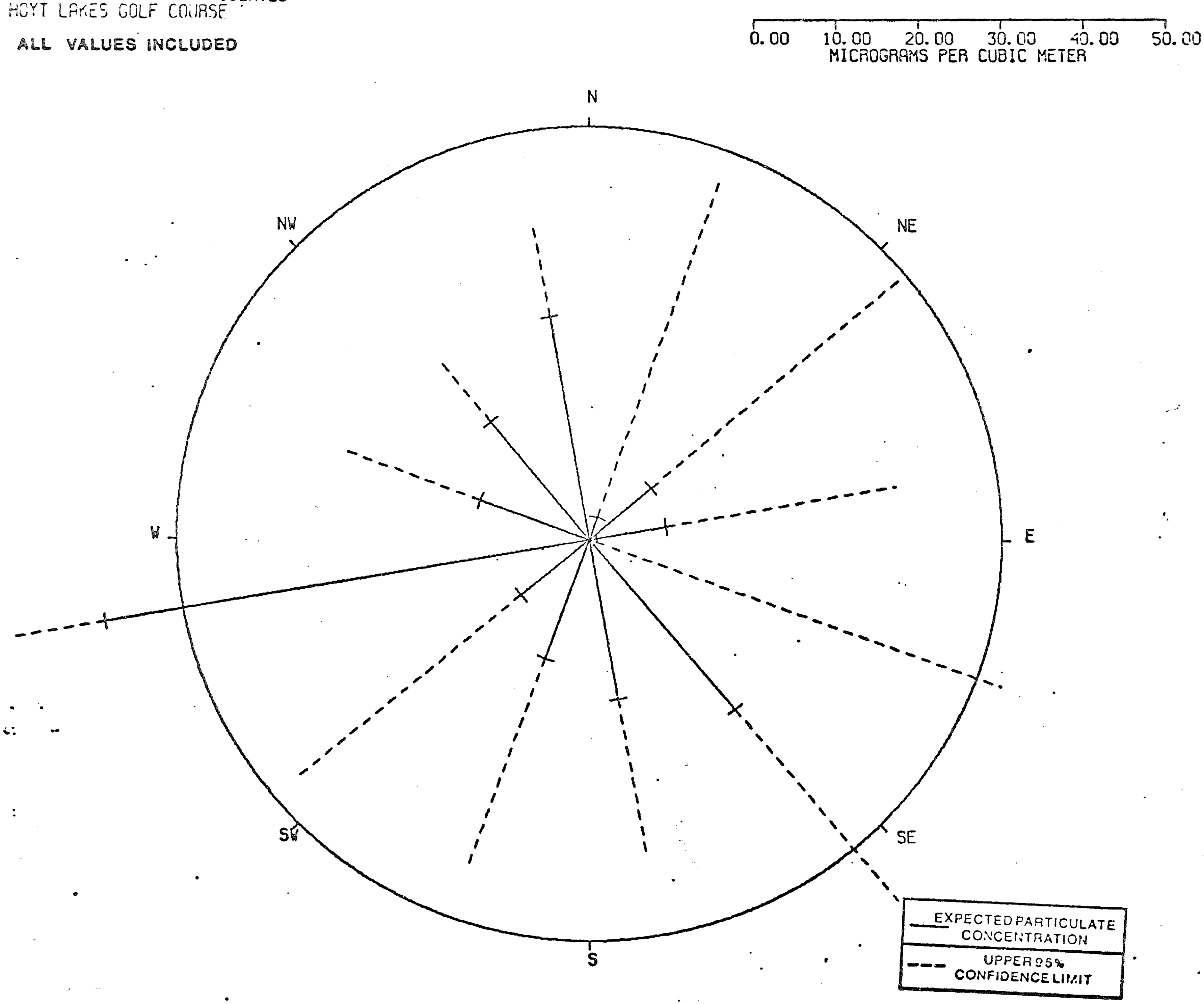


ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
HOYT LAKES POLICE STATION



POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HOYT LAKES GOLF COURSE

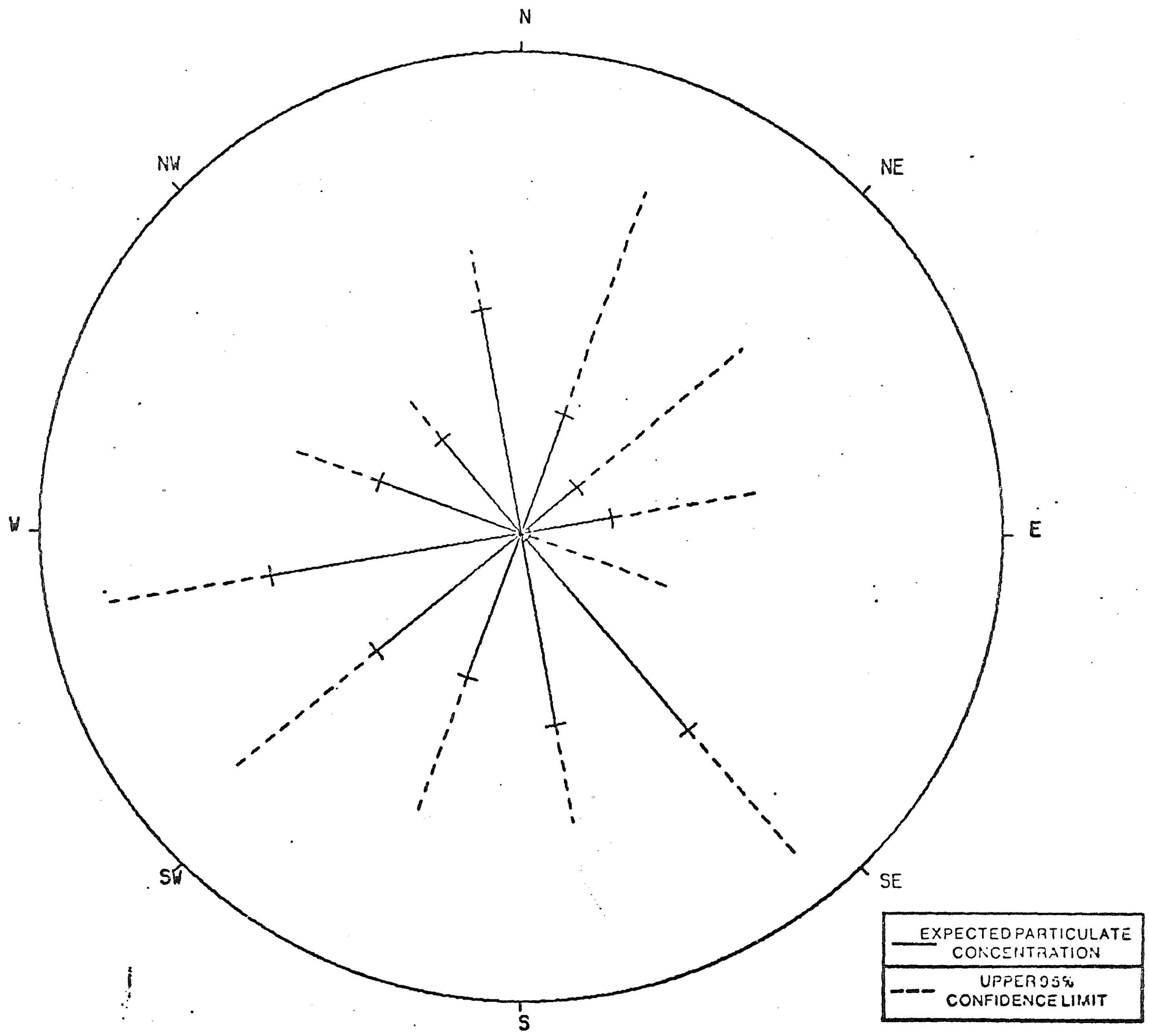
ALL VALUES INCLUDED



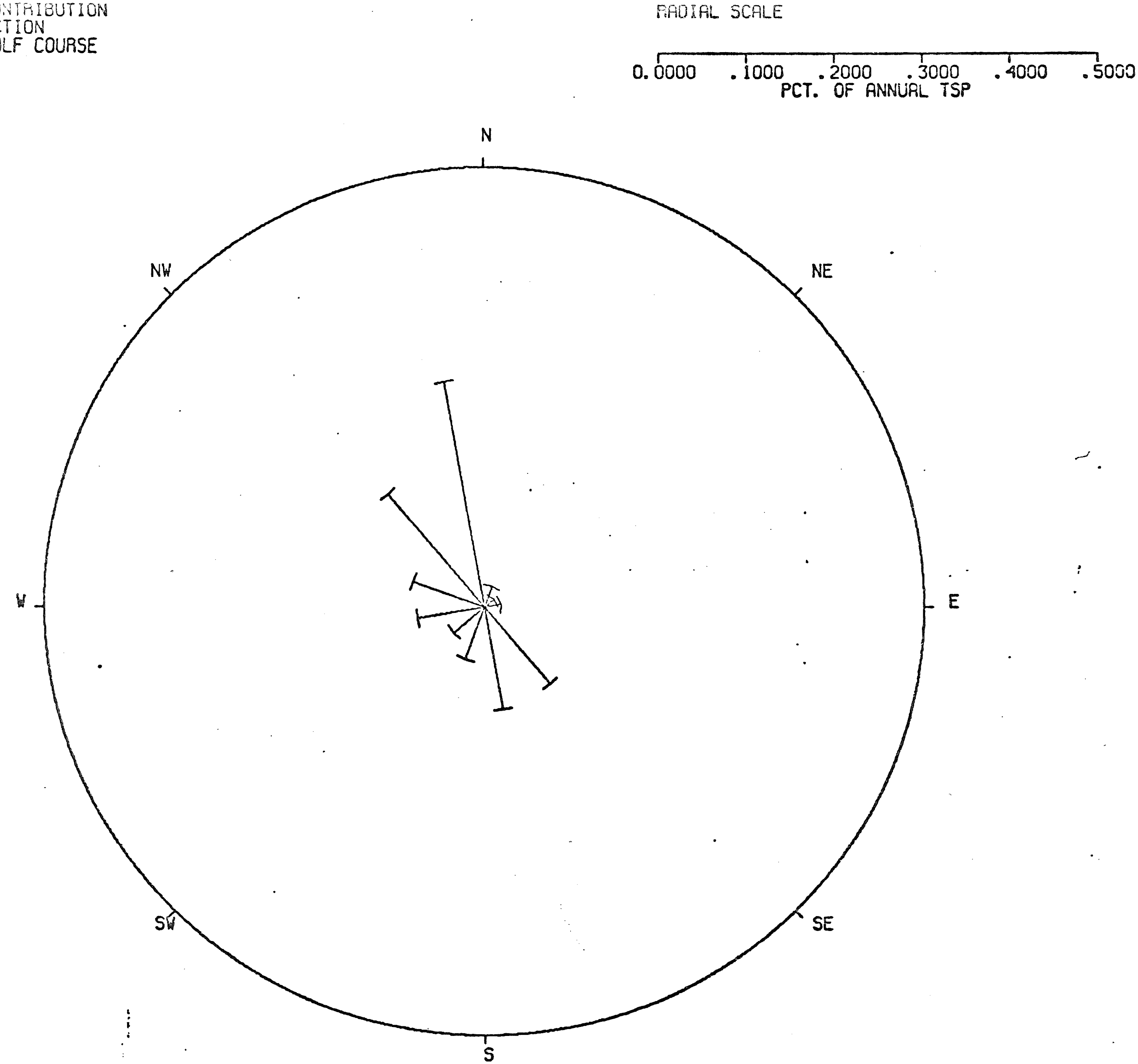
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HOYT LAKES GOLF COURSE
OUTLIERS DELETED

RADIAL SCALE - TSP CONCENTRATION

0.00 10.00 20.00 30.00 40.00 50.00
MICROGRAMS PER CUBIC METER

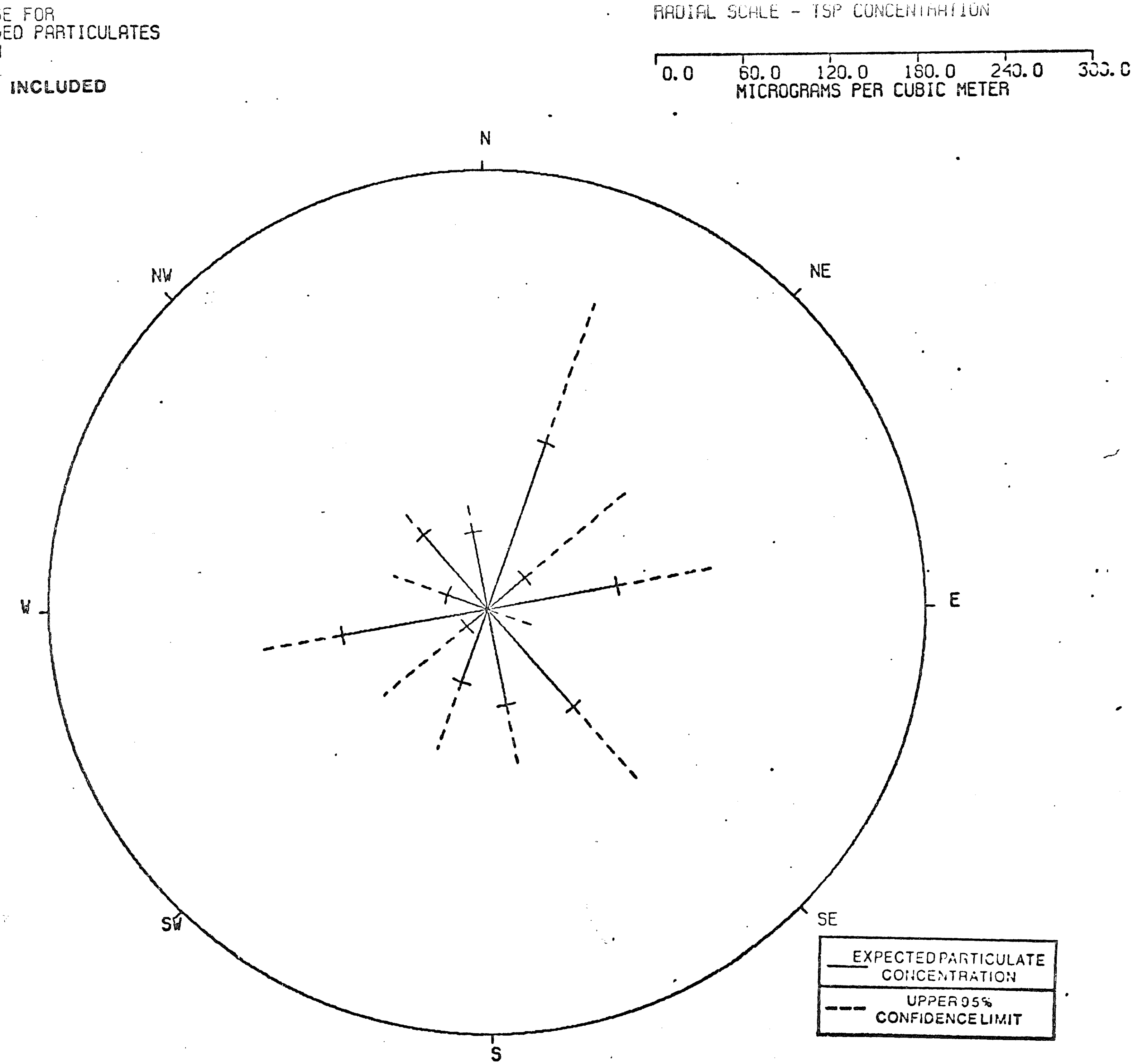


ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
HOYT LAKES GOLF COURSE

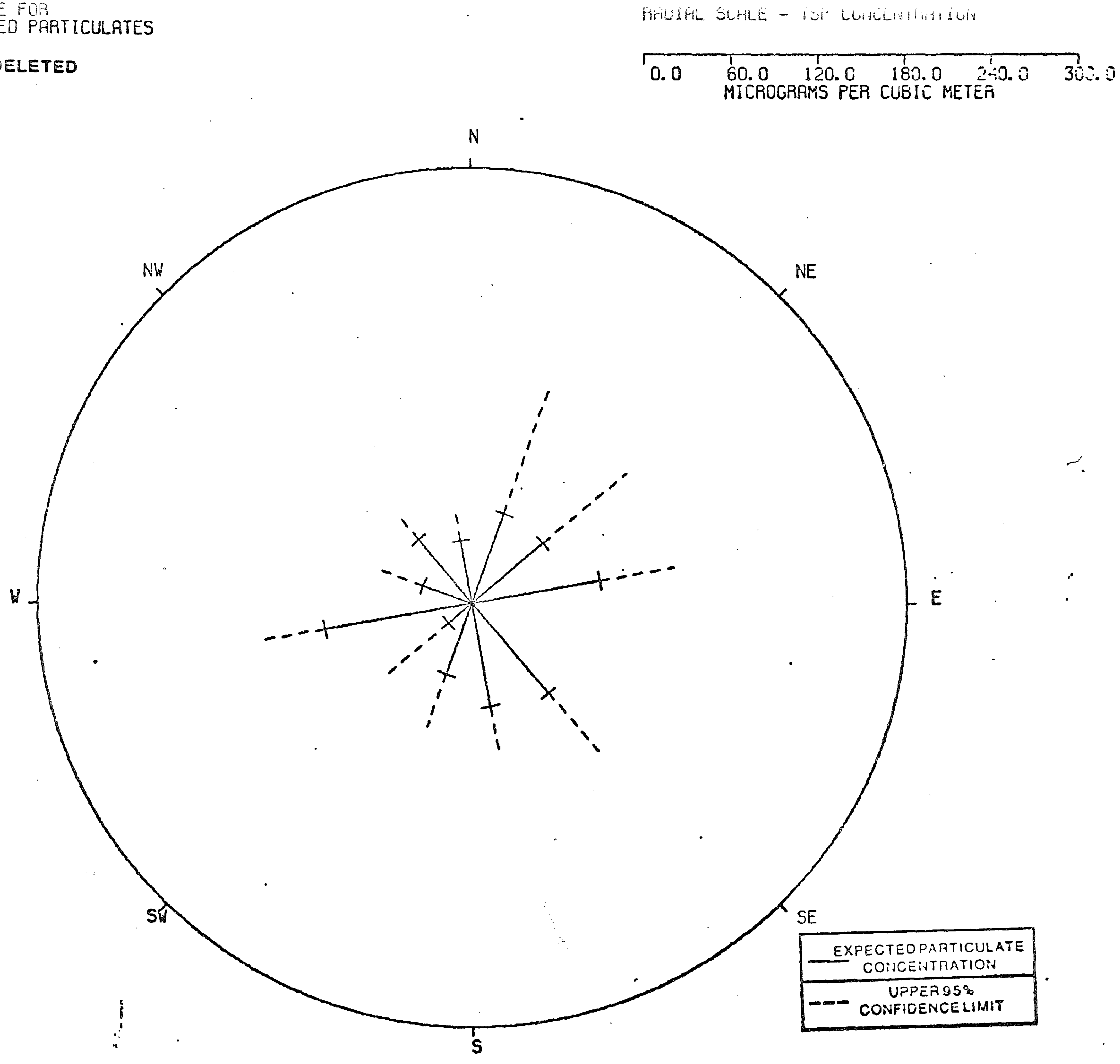


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
MOUNTAIN IRON

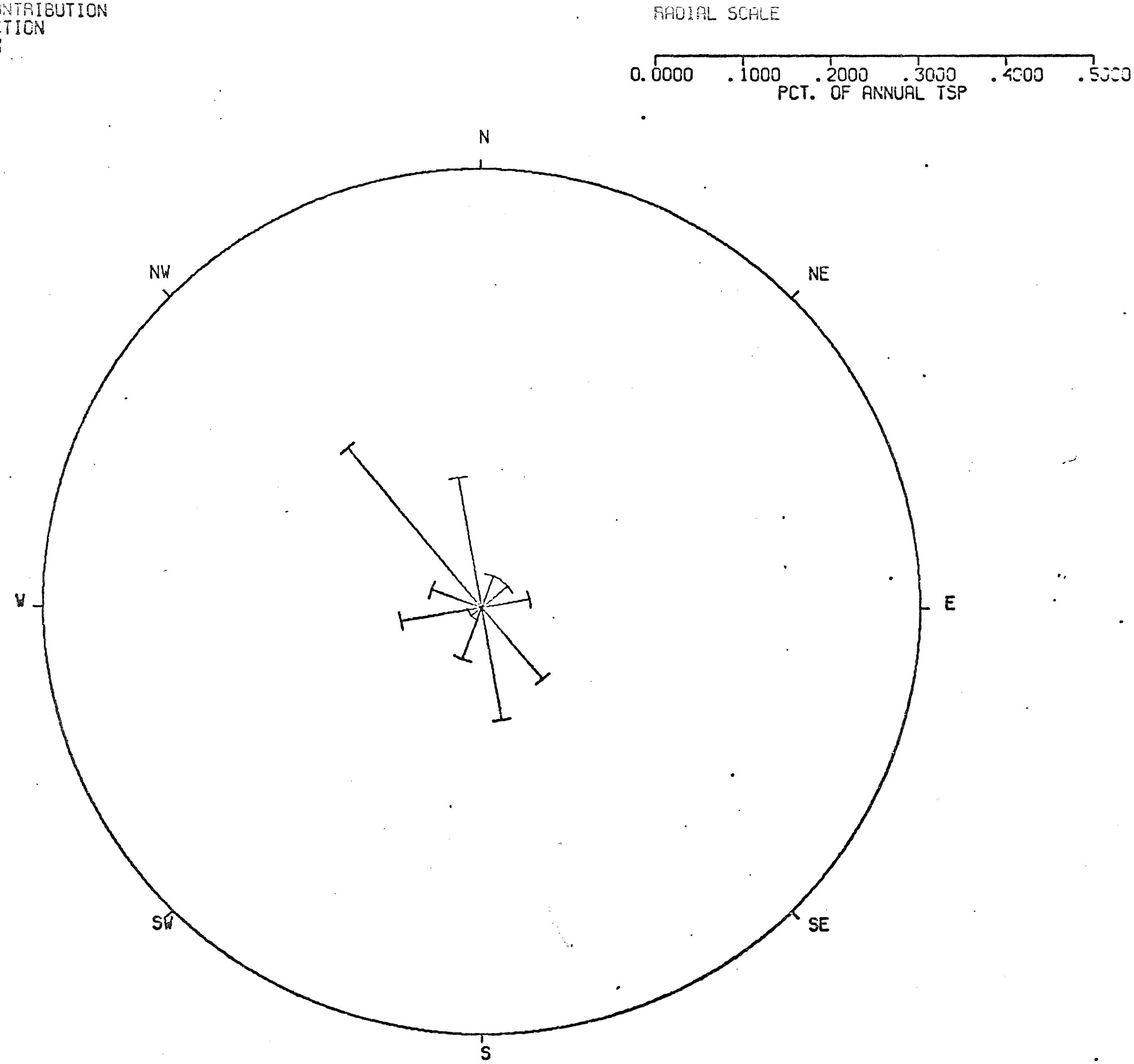
ALL VALUES INCLUDED



POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
MOUNTAIN IRON
OUTLIERS DELETED



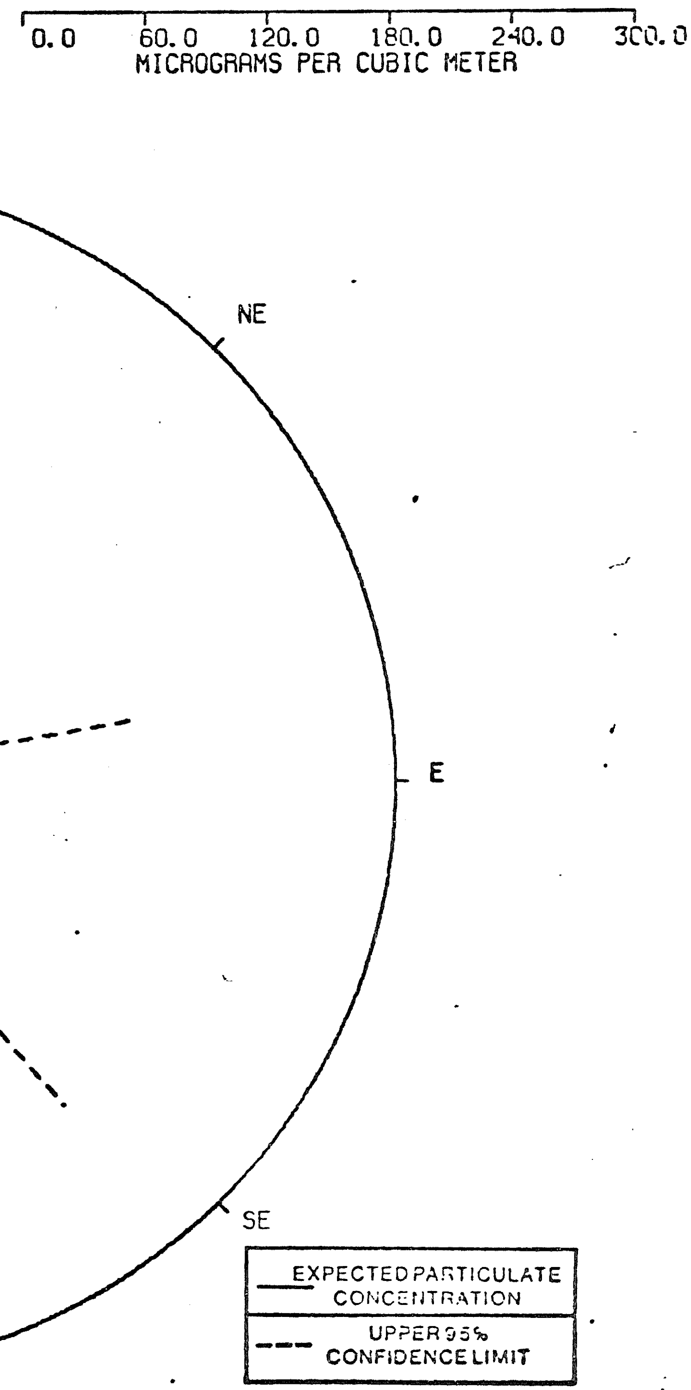
ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
MOUNTAIN IRON



POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
VIRIGNIA

ALL VALUES INCLUDED

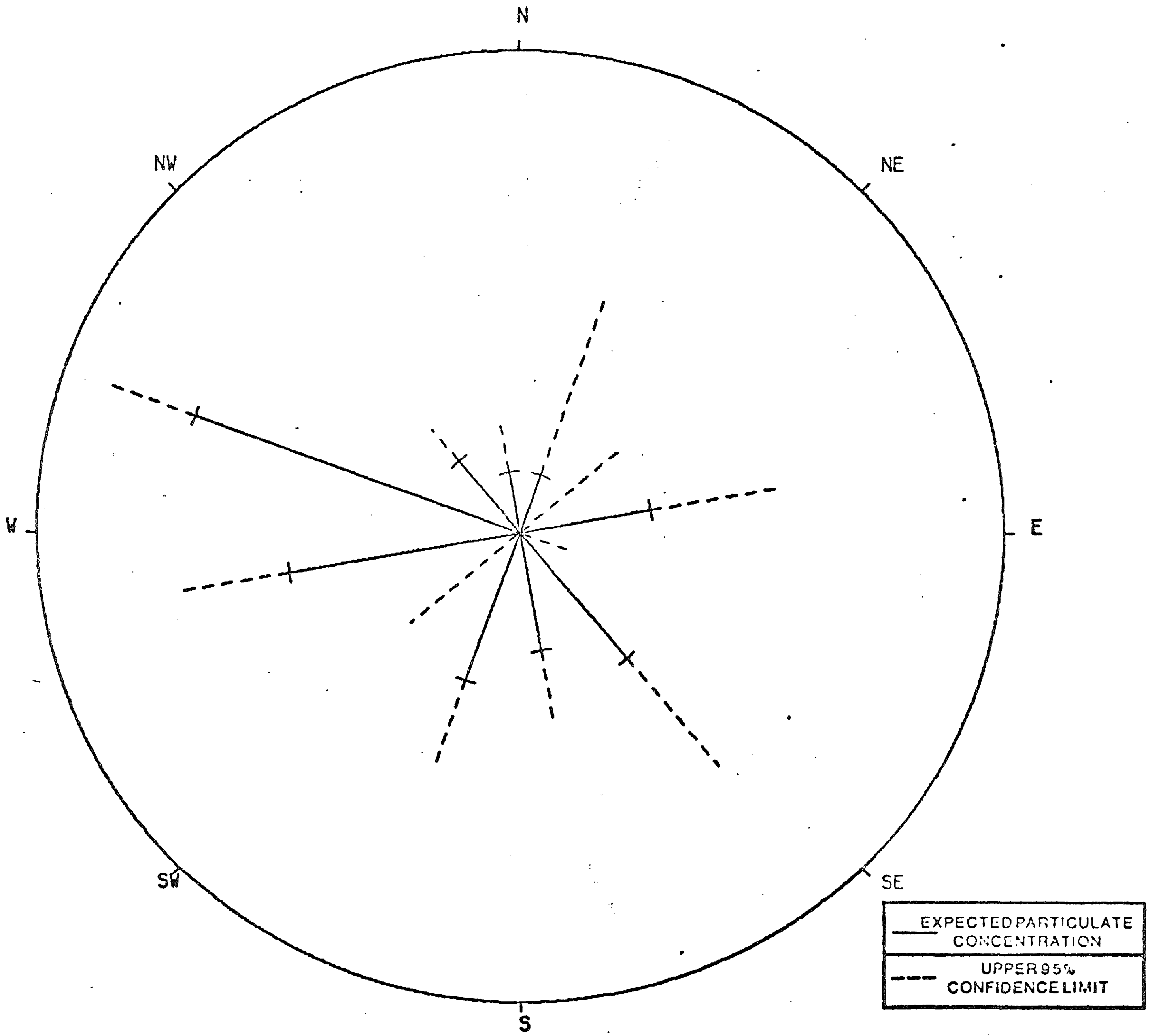
RADIAL SCALE - TSP CONCENTRATION



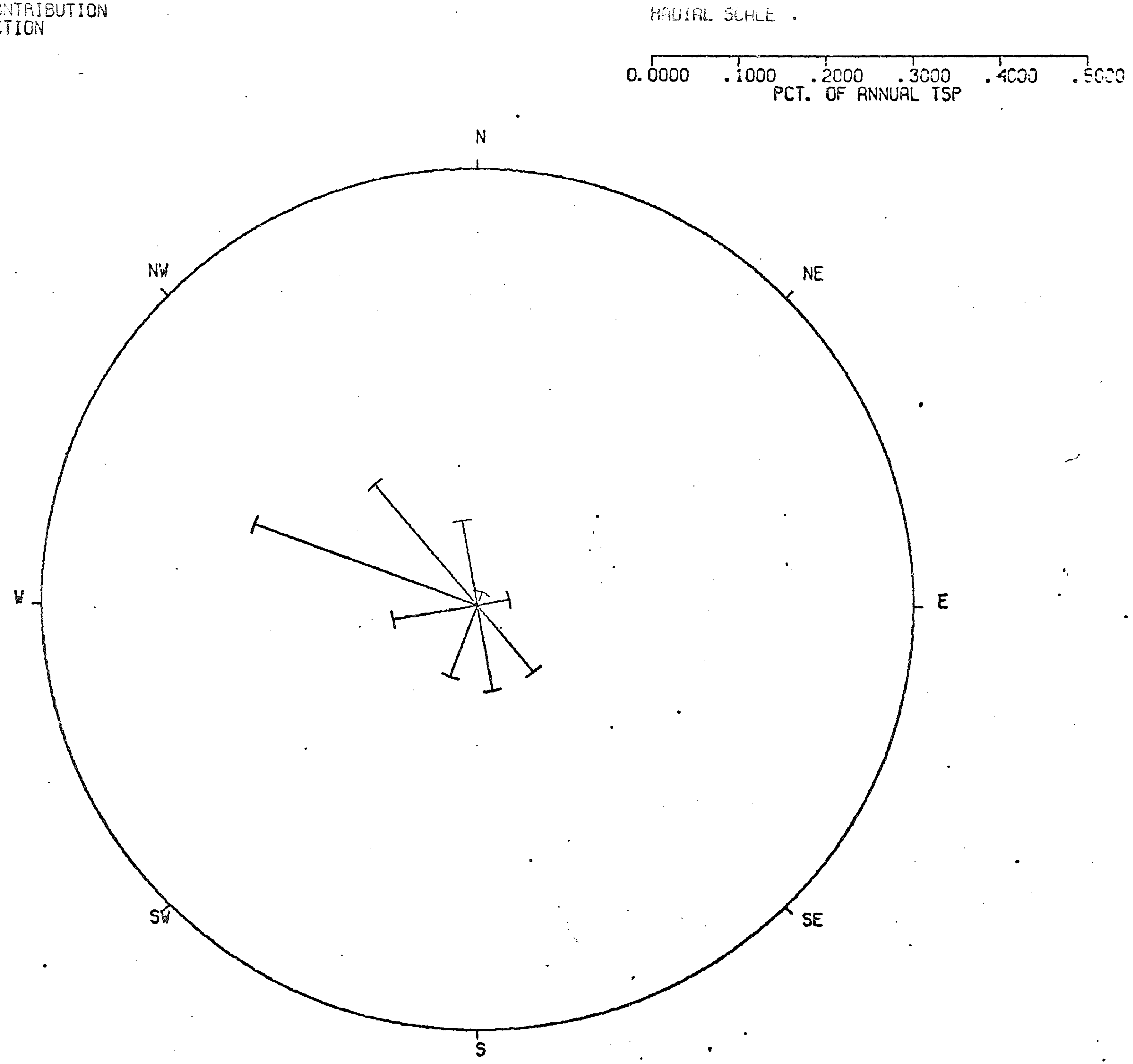
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
VIRGINIA
OUTLIERS DELETED

RADIAL SCALE - TSP CONCENTRATION

0.0 60.0 120.0 180.0 240.0 300.0
MICROGRAMS PER CUBIC METER

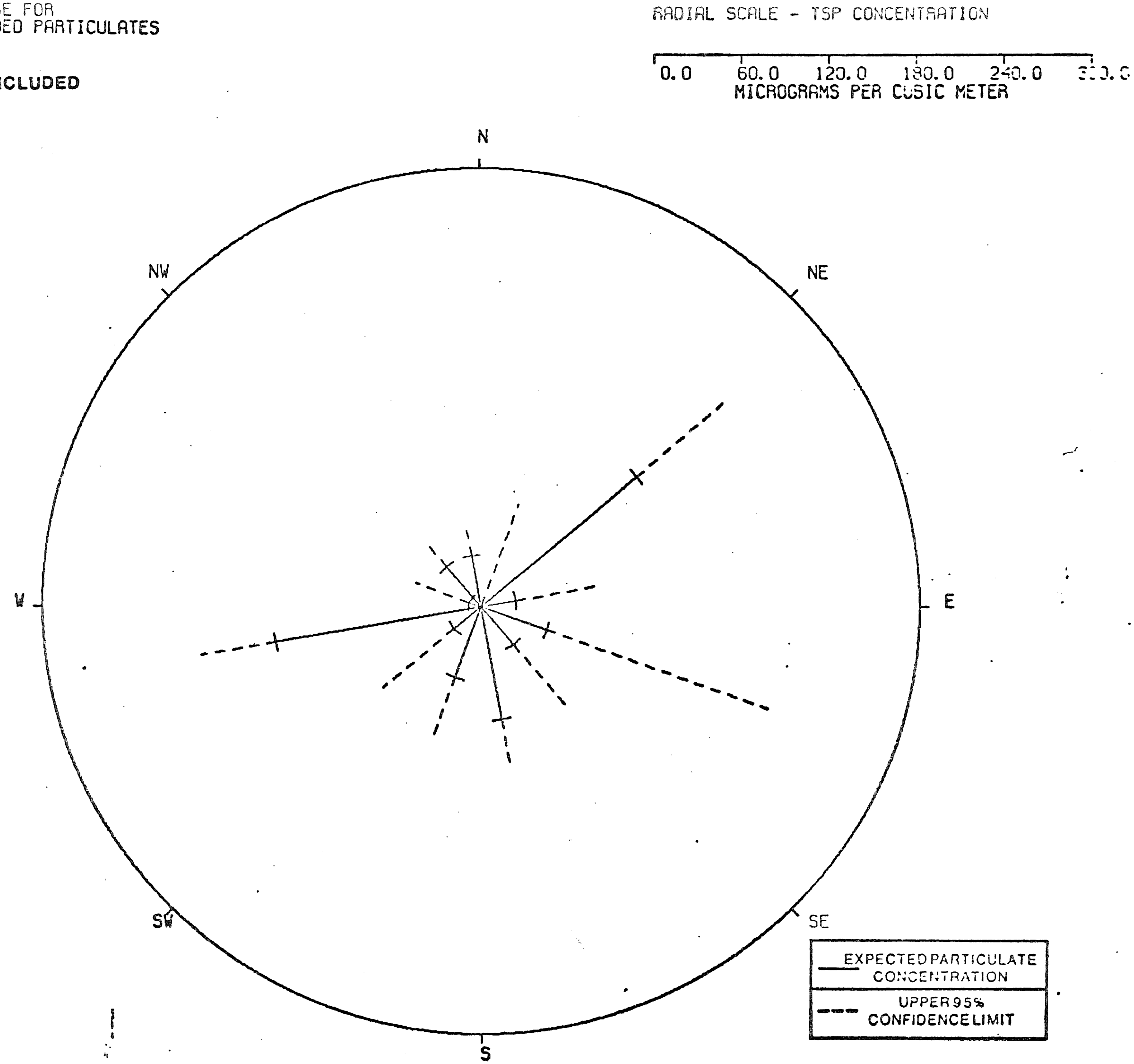


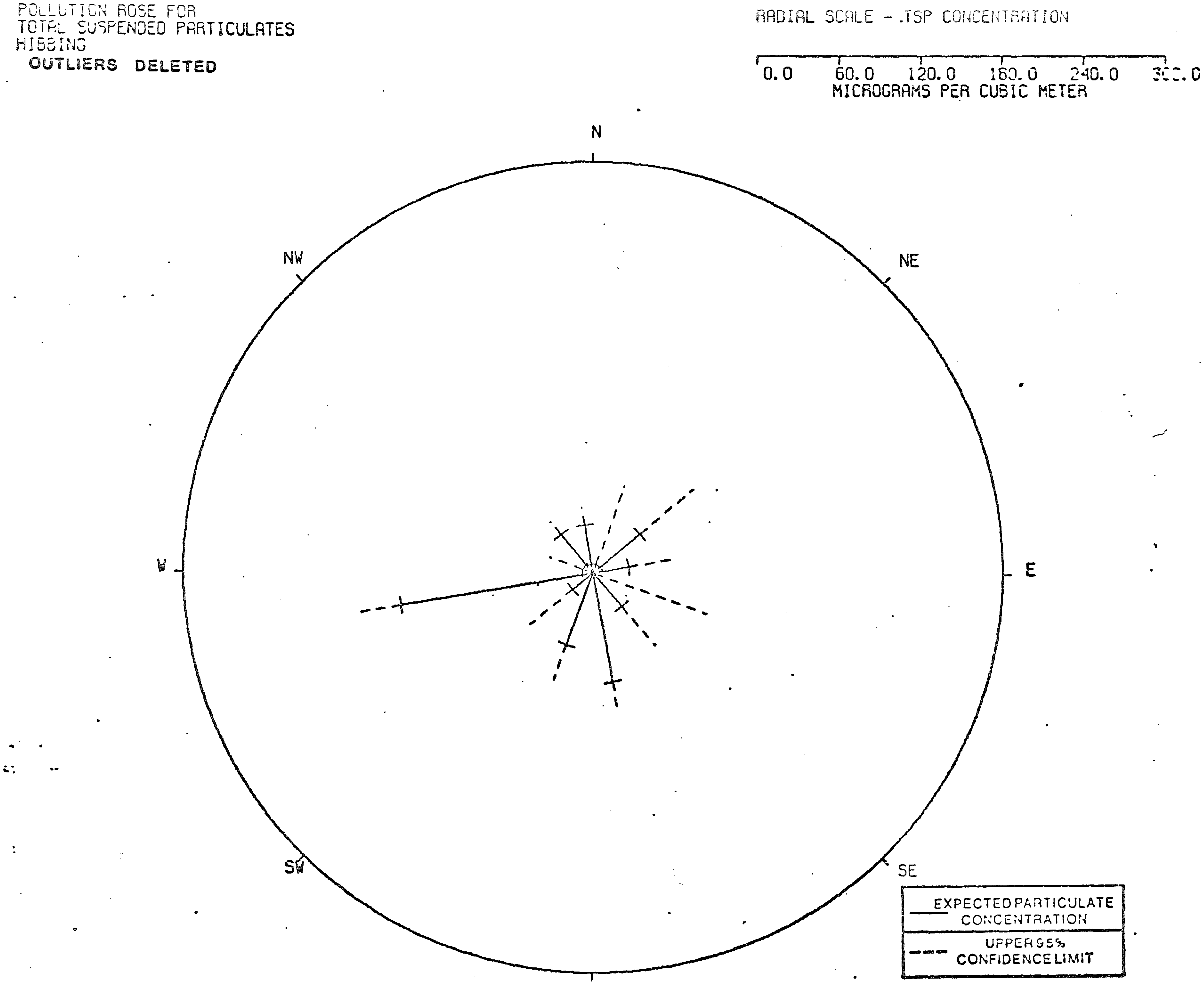
ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
VIRGINIA



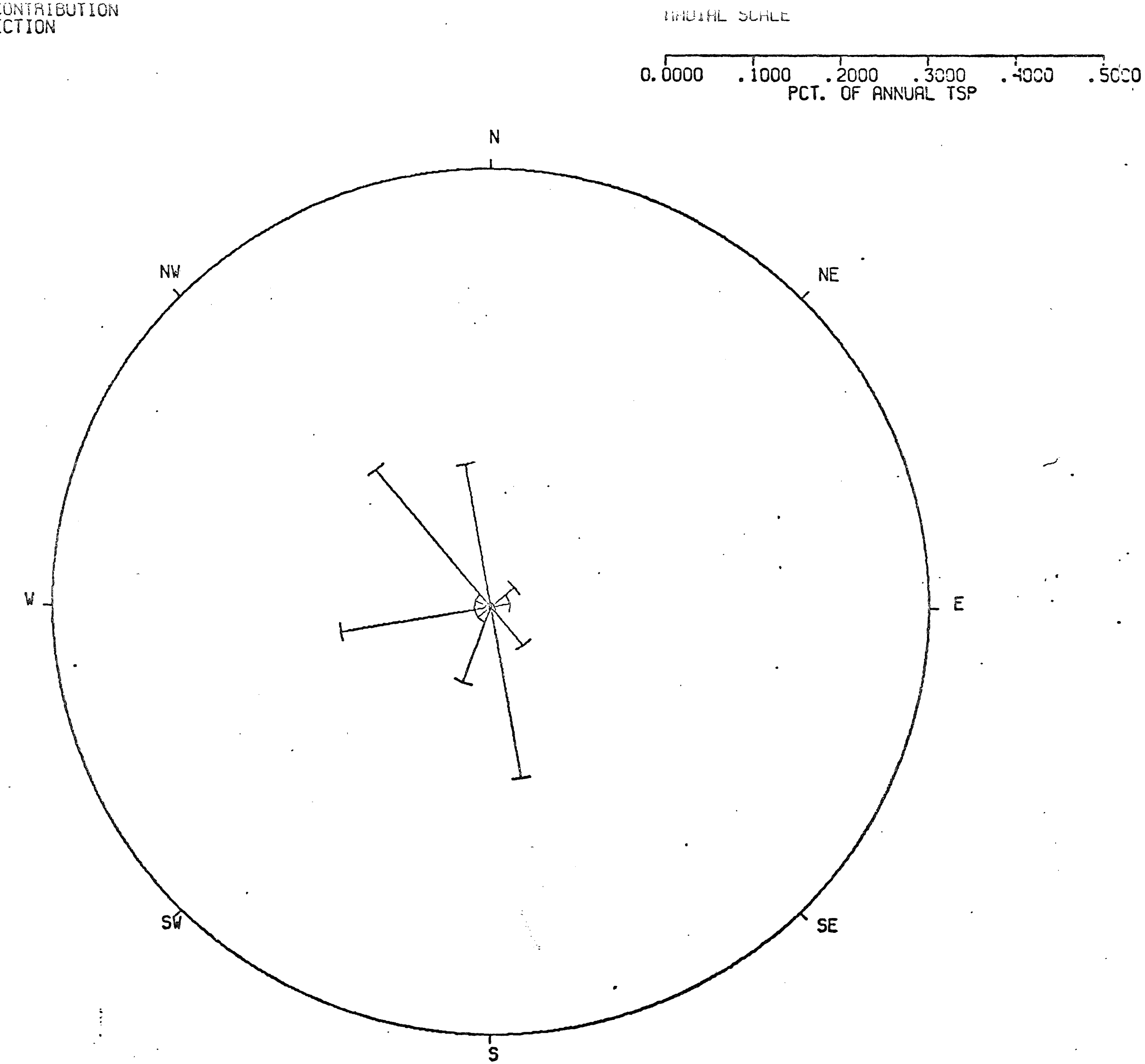
POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HIBBING

ALL VALUES INCLUDED




POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
HIBBING
OUTLIERS DELETED
RADIAL SCALE - TSP CONCENTRATION
0.0
60.0
120.0
180.0
240.0
300.0
MICROGRAMS PER CUBIC METER
N
NE
E
SE
S
SW
W
NW
EXPECTED PARTICULATE CONCENTRATION
UPPER 95% CONFIDENCE LIMIT

ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
HIBBING

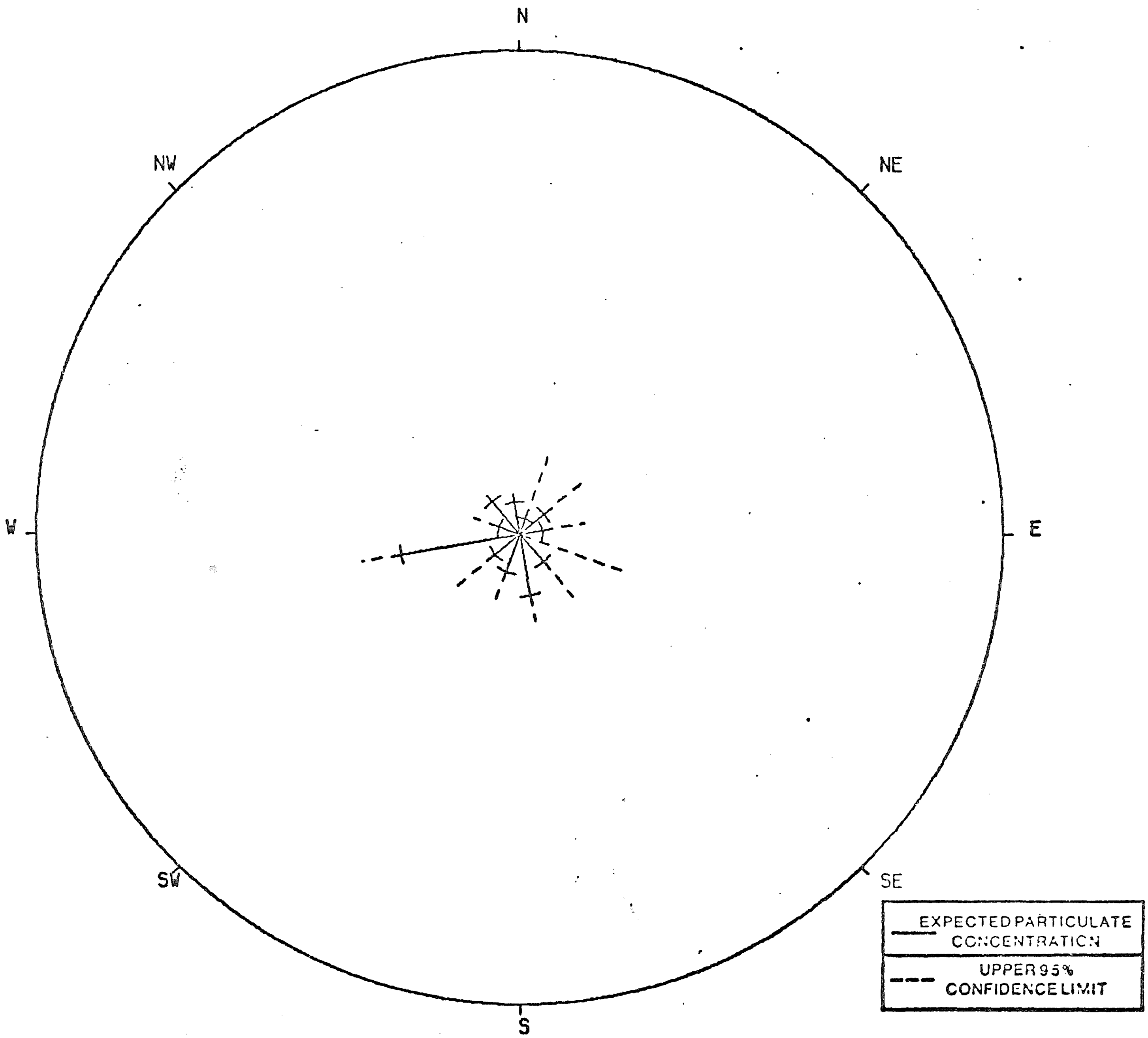


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
CU-NI STUDY AREA

ALL VALUES INCLUDED

RADIAL SCALE - TSP CONCENTRATION

0.0 60.0 120.0 180.0 240.0 300.0
MICROGRAMS PER CUBIC METER

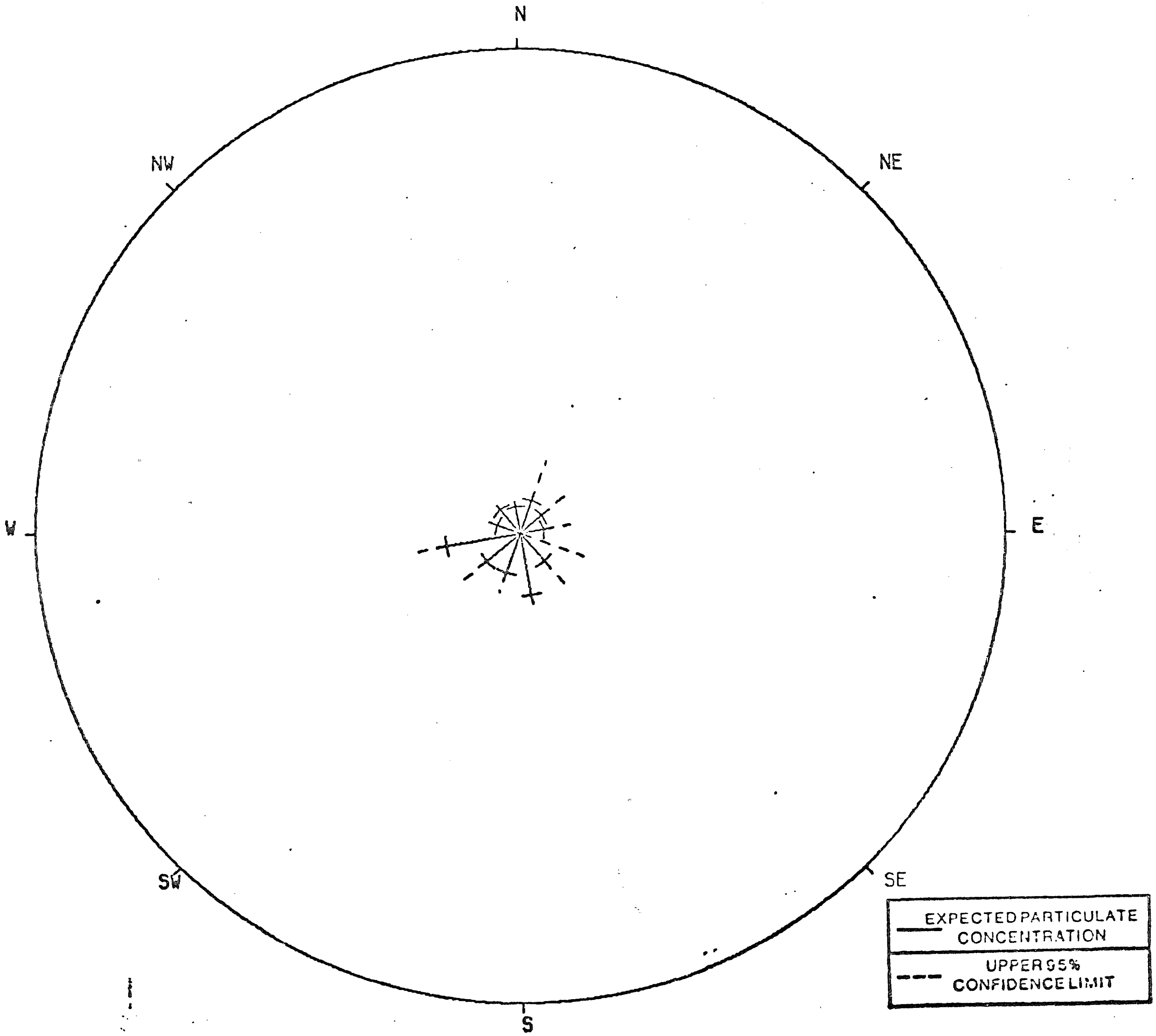


POLLUTION ROSE FOR
TOTAL SUSPENDED PARTICULATES
CU-NI STUDY AREA

OUTLIERS DELETED

RADIAL SCALE - TSP CONCENTRATION

0.0 60.0 120.0 180.0 240.0 300.0
MICROGRAMS PER CUBIC METER



ANNUAL TSP CONTRIBUTION
BY WIND DIRECTION
CU-NI STUDY AREA