ATTACHMENT A
Critical Design Value Estimation and Its Applications

Shao-Hang Chu
US Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Strategies and Standards Division (MD-15)
Research Triangle Park, NC 27711

ABSTRACT

The air quality design value is the mathematically determined pollutant concentration at a particular site that must be reduced to, or maintained at or below the National Ambient Air Quality Standards (NAAQS) in order to assure attainment. The design value may be calculated based on ambient measurements observed at a local monitor in a 3-year period or on model estimates. The design value, however, varies from year to year due to both the pollutant emissions and natural variability such as meteorological conditions, wildfires, dust storms, volcanic activities etc. In order to investigate certain policy options related to pollution controls it would be desirable to estimate a critical design value above which the NAAQS is likely to be violated with a certain probability.

In this paper, a statistical technique has been developed to estimate a critical design value that is based on the average design value and its variability in the past. The critical design value could be used as a planning tool for regulatory agencies because it is an indicator of the likelihood of future violations of the NAAQS given the current average design value and its variability. The approach is general and could be applied to estimate the critical design value for any pollutant.

As an example, eleven years (1989-1999) of PM10 data nationwide were extracted from the US EPA AIRS database to estimate the PM10 critical design values. The analyses indicate that PM10 design values in the West have much larger inter-annual variability than those in the East as reflected in their much lower critical design values. This, in turn, suggests that the inter-annual variability in meteorology, wildfires, and dust storms may have played a more significant role in the West, and also this larger variability could be partly explained by the once every six days sampling schedule at most PM10 monitoring sites.

INTRODUCTION

The air quality design value is the mathematically determined pollutant concentration at a particular site that must be reduced to, or maintained at or below the National Ambient Air Quality Standards (NAAQS) in order to assure attainment[1]. The design value may be calculated based on ambient measurements observed at a local monitor in a 3-year period or on model estimates. The detailed calculation of the design values for various criteria pollutants is described in the Appendices of the Code of Federal Regulations[2]. In certain cases, the design value has been used for regulatory purposes to determine whether the local pollutant concentration has violated the National Ambient Air Quality Standard (NAAQS). Most often, however, the design value is used to determine the level of control needed to reduce the pollutant concentration to the NAAQS[3,4,5].

The design value, however, varies from year to year due to both the pollutant emissions and natural variability such as meteorological conditions, wildfires, dust storms, volcanic activities etc. In order to investigate certain policy options related to pollution controls it would be desirable to define a critical design value above which future violations of the air quality standard are likely to occur with a certain probability.

In this paper, an effort has been made to statistically estimate a critical design value based on the average of these yearly design values and their variability in the past. This critical design value is defined in such a way as it is the highest average design value any monitoring site could have before it runs a risk of violating the NAAQS in the future at a certain probability. The technical basis of this estimation approach and its applications will be discussed in the following paragraphs.

CRITICAL DESIGN VALUE ESTIMATION

Our intention is to find a critical design value (CDV) that is the highest possible average design value (ADV) any site could have before it risks a future violation of the standard at a certain probability. First, we try to formulate a relationship among a set of variables involved: such as the CDV, NAAQS, the ADV, the standard deviation of the design values in the past, and a desirable risk factor. We find that if we assume that the design values are normally distributed and the coefficient of variation (CV), which is the ratio of the standard deviation versus the mean of the design values, does not change in the near future, then we can write the relationship as:

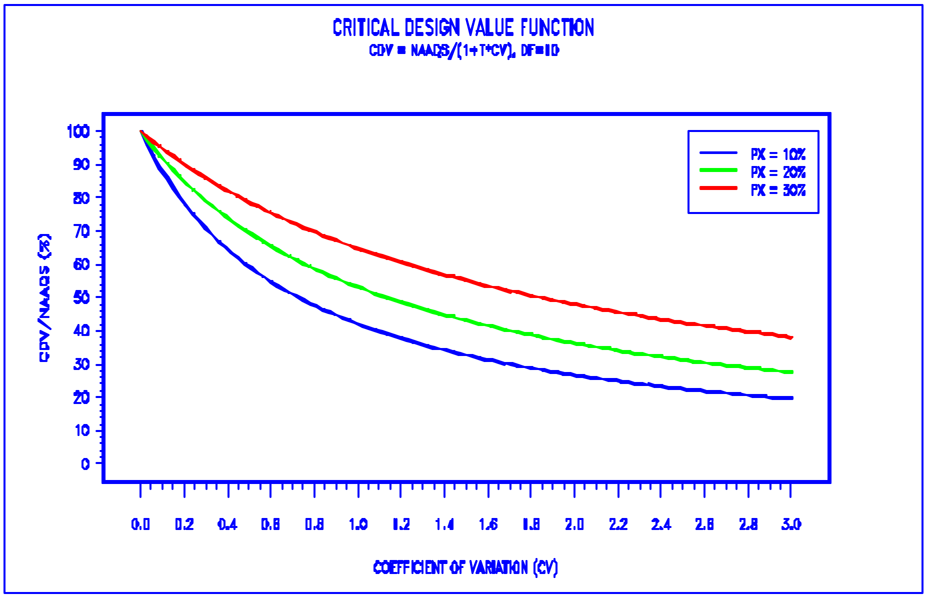
$$CDV = NAAQS/(1+t_c*CV) \qquad (1)$$

Where CDV is the critical design value, CV is the coefficient of variation of the annual design values (the ratio of standard deviation divided by the mean design value in the past), and t_c is the critical t-value corresponding to a probability, c %, of exceeding the NAAQS in the future and the degree of freedom in the estimate to the CV. Equation (1) says that based on the variability of the design values in the past, the probability of any monitoring site with an ADV less than or equal to the CDV to exceed the NAAQS in the future would be no more than c % given the same CV. In other words, the CDV is the highest ADV any monitoring site could have before it may record a future violation of the NAAQS with a certain probability. The percent probability, c, is the chosen risk factor. One can choose either a more, or less, conservative c value depending on how much risk one is willing to take.

The inter-annual variability of the air quality design values at a monitoring site can be estimated from historical data at that station. Using the air quality data in the past, one can calculate the design values for each year. With these design values one can calculate the ADV and its variability in terms of the coefficient of variation (CV). Thus, one can calculate the CDV for any site with a minimum of five years of data.

CHARACTERISTICS OF THE CRITICAL DESIGN VALUE

From equation (1) we see that the CDV is a nonlinear function of the NAAQS of the pollutant, the critical t-value, t_c, and the coefficient of variation, CV, of the design values. The normalized



relationship of the CDV to the product of t_c and CV is shown in figure 1.

Figure 1.

The dependency of CDV on the other two variables can be summarized as:

1. The larger the variability (CV) of the design values in the past, the smaller the CDV will be;
2. The lower the probability of risk for future violations (PX), the lower the CDV will be;
3. If CV=0, i.e., no variability in the design values in the past, then from Figure 1 and Equation (1) we find the highest CDV equal to the NAAQS;
4. As CV increases, the CDV approaches zero;
5. If CV is not zero but $t_c = 0$, then we will also have a CDV equal to the NAAQS, but it will have a 50% chance of violating the standard in the future because $t_c = 0$ corresponds to a probability of 50%.

In Figure 2 we have chosen a risk factor of 10% probability of future violation and plotted two examples using generated data with significantly different variability in the annual PM10 design values. It is intended to illustrate the relationship among design values, ADV, CDV, and the PM10 annual NAAQS of 50 ug/m3. In this example we see that the CDV depends strongly on the inter-annual variability of the design values rather than on their means. Also, from the upper panel of Figure 2 we see that once the ADV is higher than the CDV, the probability of violating the standard will be higher than the risk we have chosen (in this case, it is one out of ten).

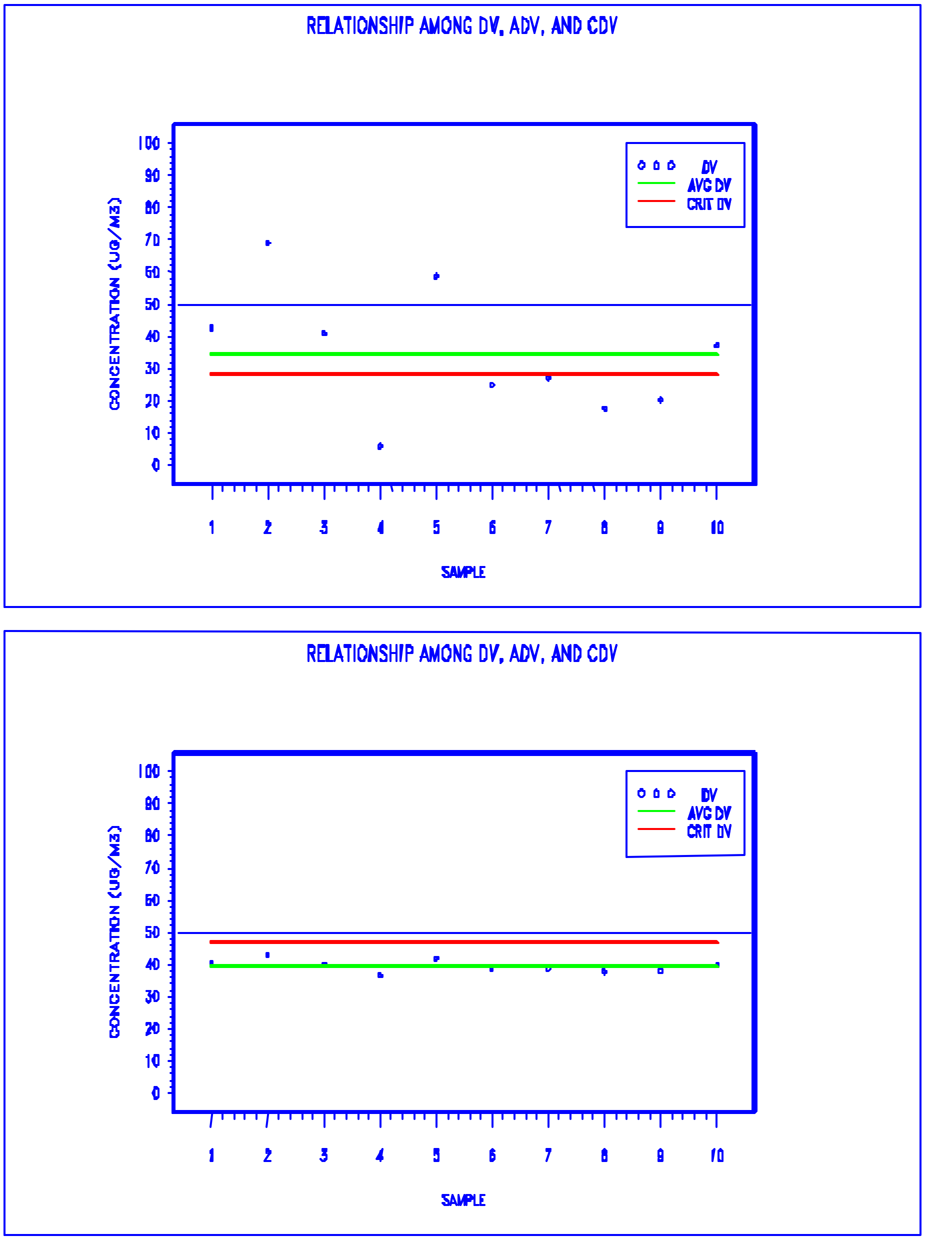


Figure 2.

Contrasting the two panels of Figure 2, we see that whether a site will have a higher or lower risk of violating the NAAQS in the future depends on how much higher or lower the ADV is to the CDV. Thus, unless some drastic change in emissions occurred in the past or should occur in the future, the CDV can be used to assess the likelihood of violating the NAAQS in the future in that area based on normal probability predictions. For this reason, this technique and the estimated CDV could be used as a planning tool for regulatory agencies to decide whether more or fewer pollutant controls are needed in a specific area.

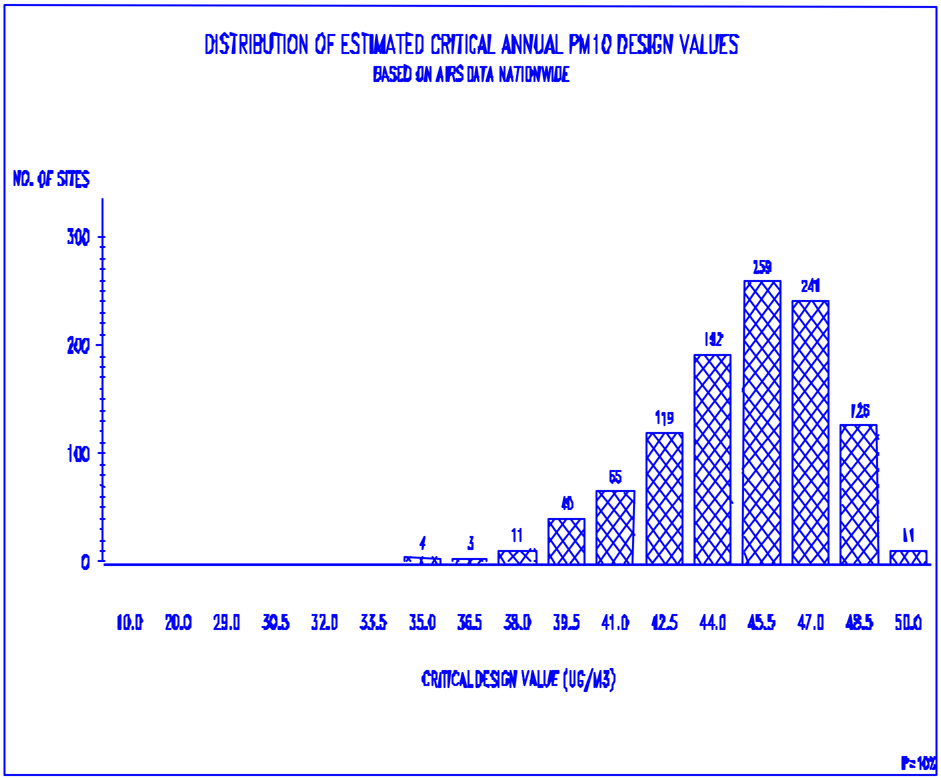
PM10 CRITICAL DESIGN VALUES AND DISCUSSIONS

To demonstrate this approach, eleven years (1989-1999) of PM10 data nationwide were extracted from the United States Environmental Protection Agency AIRS database. The annual and 24-hr PM10 design values were calculated following the US EPA Guidance[1]. Then the methodology described in the previous section was applied using a tolerable risk factor of 10% probability of future violation of the NAAQS to calculate the CDVs for all monitor sites with more than five years of valid data. The analyses are discussed and presented in the following figures.

Figure 3 is a frequency distribution of these calculated annual and 24-hr CDVs. We see that the distributions of both the annual and the 24-hr CDVs are skewed to the left with a median annual CDV of 45.3 ug/m3 and a median 24-hr CDV of 123.2 ug/m3. The long tails to the left (low values) suggest that there are places where the inter-annual variability of the design values are quite large. It also suggests that these areas are likely to have a higher probability of violating the standards if they are already in a major PM10 source region with relatively high PM10 concentrations.

In Figure 4 a longitudinal scatter plot of both the ADVs and the CDVs at all sites spanning from Maine to California, was produced to see whether there is a difference from the East to the West. Comparing the differences between these overlaid ADVs and CDVs we see clearly that most of the higher risk areas (i.e., the areas where the ADVs are greater than the CDVs) are in the West and Midwest. The geographical distribution of the CDVs and the actual ADVs are shown in Figures 5 and 6 respectively. For comparison purposes, the ADVs in Figure 6 are color coded to show their probability of future violation of the NAAQS. The probability of future violation of the NAAQS at each site is calculated by inverting the t-values using equation (1).

The East-West difference in CDVs can be explained largely by the fact that the West, in general, has a much larger inter-annual variability of the design values than the East. However, since the anthropogenic emissions in a region usually do not change very much from year to year, the large variability in the inter-annual PM10 design values in the West may be largely attributable to the inter-annual variation in natural conditions such as meteorology, wildfires, dust storms, and volcanic emissions, etc. The higher occurrences of wildfires and dust storms in the West are known to be associated with its much drier climate, meteorological conditions, and topography. Another influencing factor on the inter-annual variability could be related to the sampling frequency of the PM10 data, which for many sites is only once every six days. However, this is more likely in the East because fewer sites are in non-attainment status and thus not required to sample more frequently than once in six days.


DISTRIBUTION OF ESTIMATED CRITICAL ANNUAL PM10 DESIGN VALUES
BASED ON AIRS DATA NATIONWIDE
NO. OF SITES
300
200
100
0
4
3
11
40
65
119
192
259
241
126
11
10.0
20.0
29.0
30.5
32.0
33.5
35.0
36.5
38.0
39.5
41.0
42.5
44.0
45.5
47.0
48.5
50.0
CRITICAL DESIGN VALUE (UG/M3)
P=102

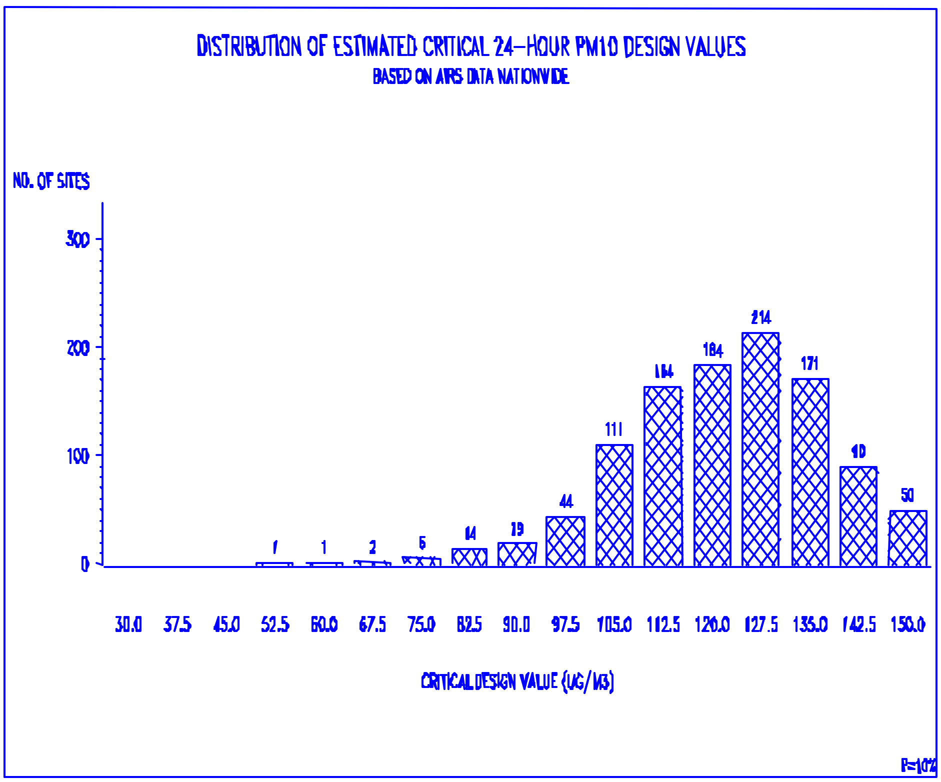
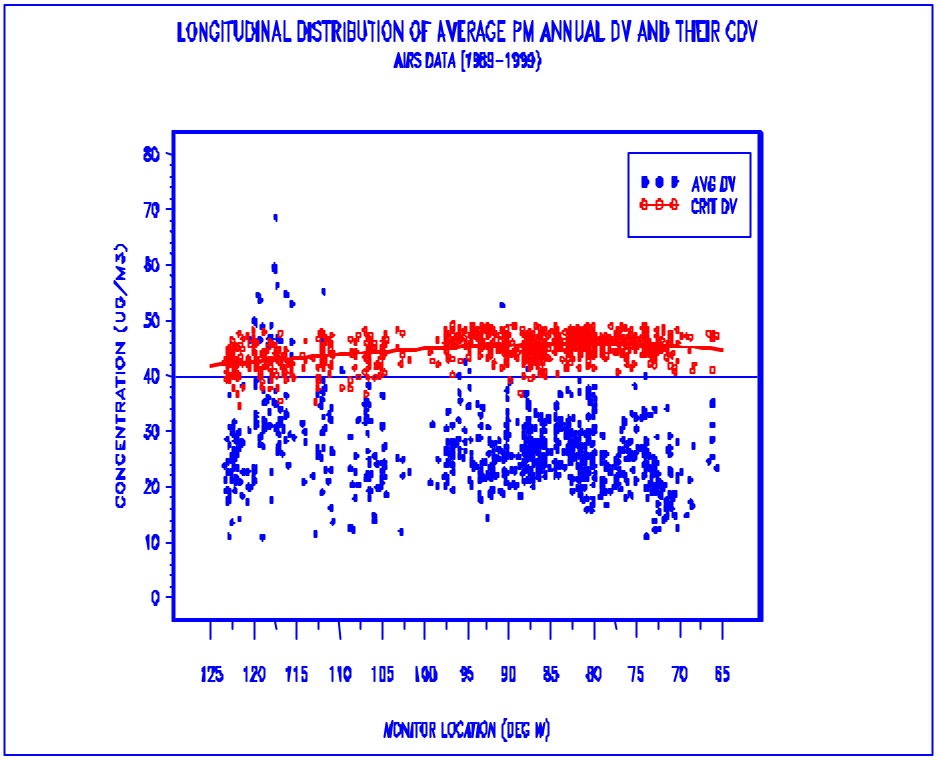

DISTRIBUTION OF ESTIMATED CRITICAL 24-HOUR PM10 DESIGN VALUES
BASED ON AIRS DATA NATIONWIDE
NO. OF SITES
300
200
100
0
1
1
2
5
14
19
44
111
164
184
214
171
90
50
30.0
37.5
45.0
52.5
60.0
67.5
75.0
82.5
90.0
97.5
105.0
112.5
120.0
127.5
135.0
142.5
150.0
CRITICAL DESIGN VALUE (UG/M3)
P=102

Figure 3.


LONGITUDINAL DISTRIBUTION OF AVERAGE PM ANNUAL DV AND THEIR CDV
AIRS DATA (1989-1999)
AVG DV
CRIT DV
CONCENTRATION (UG/M3)
MONITOR LOCATION (DEG W)

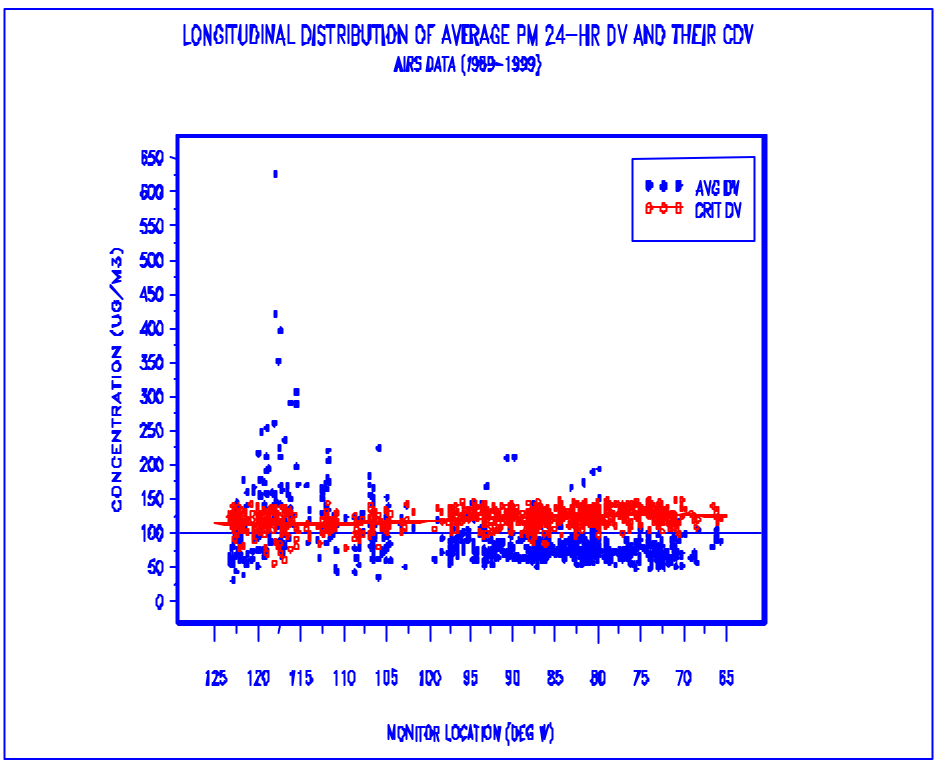
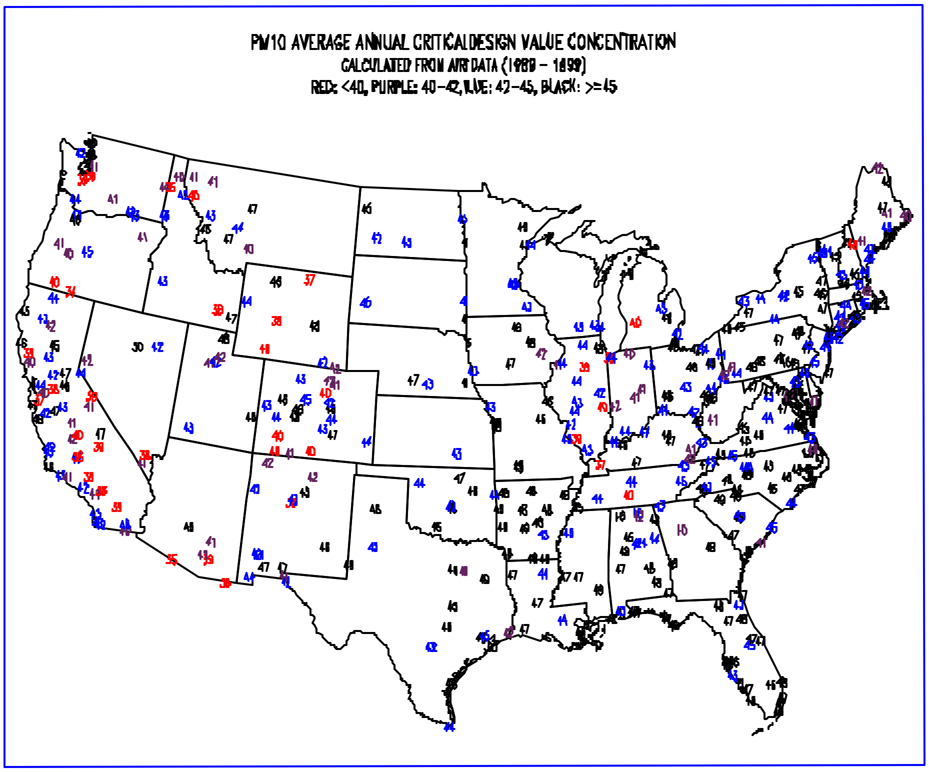

LONGITUDINAL DISTRIBUTION OF AVERAGE PM 24-HR DV AND THEIR CDV
AIRS DATA (1989-1999)
AVG DV
CRIT DV
CONCENTRATION (UG/M3)
MONITOR LOCATION (DEG W)

Figure 4.


PM10 AVERAGE ANNUAL CRITICALDESIGN VALUE CONCENTRATION
CALCULATED FROM AIR DATA

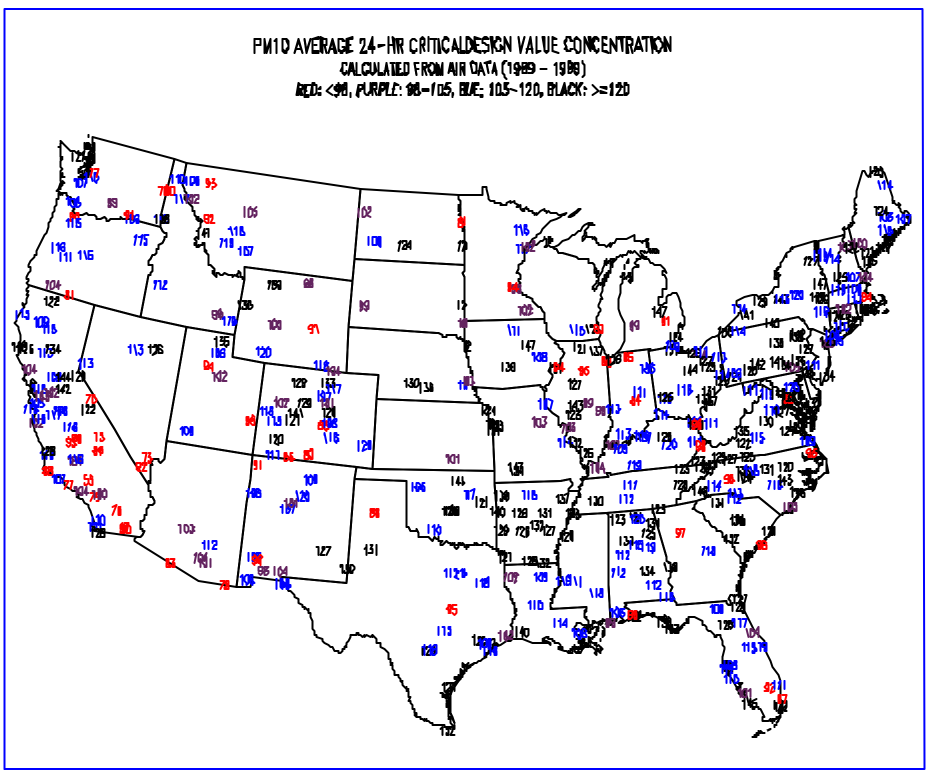

PM10 AVERAGE 24-HR CRITICALDESIGN VALUE CONCENTRATION
CALCULATED FROM AIR DATA

Figure 5.

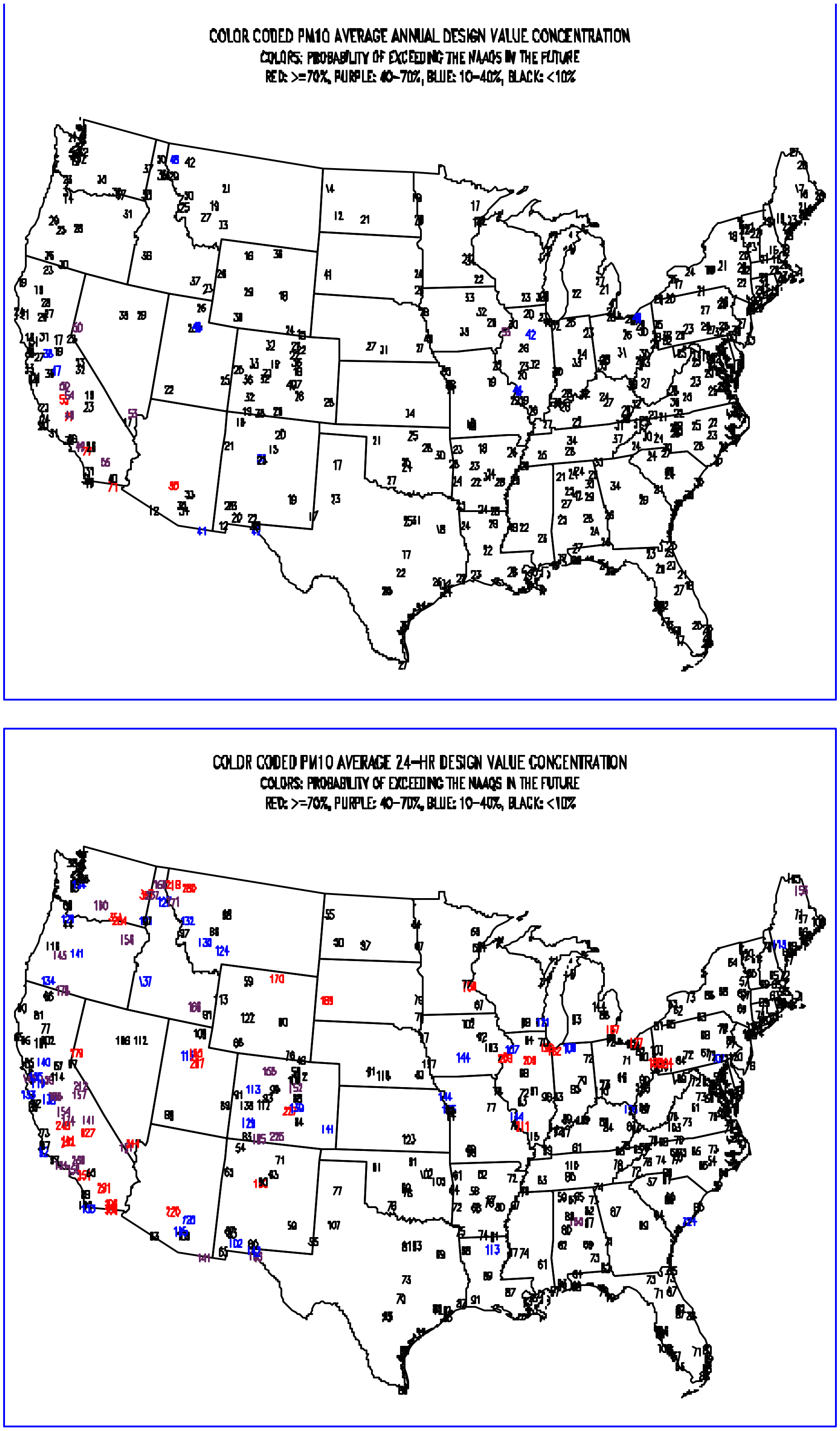

COLOR CODED PM10 AVERAGE ANNUAL DESIGN VALUE CONCENTRATION
COLORS: PROBABILITY OF EXCEEDING THE NAAQS IN THE FUTURE
RED: >=70%, PURPLE: 40-70%, BLUE: 10-40%, BLACK: <10%
COLOR CODED PM10 AVERAGE 24-HR DESIGN VALUE CONCENTRATION
COLORS: PROBABILITY OF EXCEEDING THE NAAQS IN THE FUTURE
RED: >=70%, PURPLE: 40-70%, BLUE: 10-40%, BLACK: <10%

Figure 6.

CONCLUSIONS

In this paper a statistical technique has been developed to determine the CDV which is the highest possible average design value any monitoring site could have before it may record a future violation of the NAAQS with a certain probability. The critical design value is calculated based on the average design value and its variability in the past, and it also involves a risk factor of our choice in the estimation. The difference between the ADV and CDV is a good indicator of whether the site is running a higher or lower risk of violating the NAAQS in the future than one is willing to take. Using this approach, one can even predict the probability of violating the NAAQS in the near future at any given site with adequate data length. Thus, this technique could be used as a planning tool for regulatory agencies to assess the risk of future violation of the NAAQS at any monitoring site and to make decisions about emissions controls. Further, since this technique is very general, it can be applied to any pollutant with a minimum of five years of valid data.

As an example, 11 years (1989-1999) of PM10 data were analyzed using this technique. The results suggest that the inter-annual variability of the design values in the West is, on the average, much larger than that in the East, which is reflected in the calculated CDVs. Since anthropogenic emissions in a region usually do not change very much from year to year, the large variability in the inter-annual PM10 design values in the West may be largely attributable to the inter-annual variation in natural conditions such as meteorology, wildfires, dust storms, and volcanic activities, etc. The higher occurrences of wildfires and dust storms in the West are known to be associated with its much drier climate, meteorological conditions, and topography. The once every six days sampling practice of PM10 monitoring may also have some influence on the inter-annual variability of PM10 design values.

FUTURE WORK

Some further studies have been planned which include applying the same technique to other pollutants, and searching for a better estimate of CV in case when significant trend exists in the yearly design values. Since the variance estimate could be affected by an underlying trend and that a better estimate could be made of the CV if the trend and/or serial correlation could be removed from the estimate.

ACKNOWLEDGMENTS

The author would like to give thanks to Dr. Terence Fitz-Simons and Mr. Bill Cox for reading the draft manuscript and their helpful discussions.

REFERENCES

1. US EPA, *PM10 SIP Development Guideline*. EPA-450/2-86-001, US Environmental Protection Agency, Research Triangle Park, NC 27711, 1986.

2. Code of Federal Regulations, *Protection of Environment*, 40 CFR part 50, 1998.
3. US EPA, *Guideline on Air Quality Models (Revised)*. EPA-450/2-78-27 R. U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, 1986.
4. US EPA, *Guideline for the Interpretation of Ozone Air Quality Standards*. EPA-450/4-79-003, US Environmental Protection Agency, Research Triangle Park, NC 27711, 1979.
5. Curran, T.C. and W.M. Cox, *Data Analysis Procedures for the Ozone NAAQS Statistical Format*. J. Air Pollution Control Association, 1980.

KEYWORDS
Critical design value, design value, inter-annual variability, PM10, probability