Handbook on

PROCEDURES FOR ESTABLISHING THE UNCERTAINTIES OF EMISSION ESTIMATES

ARB Contract No. A5-184-32
VRC Project No. 1055

by:

Yuji Horie

Valley Research Corporation
15904 Strathern Street, Suite 22
Van Nuys, California 91406

Prepared for:

AIR RESOURCES BOARD
Research Division
1800 15th Street
Sacramento, California 95814

February 1988

DISCLAIMER

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, the source of their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

1.0 INTRODUCTION

This handbook is designed to provide environmental engineers and planners with a guide for assessing and reporting uncertainties associated with various types of emission estimates used in compiling emission inventories. In particular, this handbook is intended to help engineers, when faced with the task of estimating emissions from air pollution sources, to quantify their perceptions of uncertainty in relation to their estimates.

1.1 UNCERTAINTY: THE CONCEPTUAL BASIS

As a matter of simple fact, uncertainty exists in any measurement, observation, engineering estimate, quantitative guess, or mathematical model simulation. When a person estimates the distance to a nearby building "by eyeball", his estimate will be more uncertain than if he uses a measuring device. In contrast, a tool and die machinist should be able to estimate the thickness of a particular piece of paper more accurately than he can measure it with a foot ruler. A plant engineer using mass balance principles together with fuel consumption and sulfur content data may be able to estimate monthly or annual SO_2 emissions from a power generating unit as accurately as (and perhaps more accurately than) by calculating it from data generated by a continuous emission monitor.

An engineering estimate is not always inferior to a measured quantity nor is an expert estimate always inferior to an emission value calculated from some emission model. Uncertainty of an engineering estimate depends on the quality of information available, whereas uncertainty of a measured quantity depends on the precision and accuracy of the measurement device. When a model is used, uncertainty depends on the accuracy of the model and of the data input to the model.

In statistical terms, uncertainty of an estimated quantity may be thought of as analogous to random errors in repeated measurements. "Precision" indicates how close those measured values are to each other. The term "uncertainty" is inversely related to precision. If discrepancies between the individual results in a series of measurements are very

small, the precision of measurement is said to be high, while the uncertainty of the measurement is small.

On the other hand, not all error associated with measurement is referable to defects in precision of measurement. An engineer can easily measure a flat piece of material to a precision of ±0.5 mm using a meter rule; but, if the rule he uses is not accurate, his result will be inaccurate, even though precise. Accordingly, if he is uncertain as to the accuracy of his measuring stick, he must also be uncertain about the accuracy of his result, regardless of its precision. This second type of uncertainty is ascribed to systematic error, also called bias. It is "systematic" in that it is characteristic of the particular tool or system used to generate the estimate; it cannot be accounted for or corrected without a specific investigation of that system.

It should be noted that most emissions estimation systems are not based on repeated measurements, but rather on engineering understanding of the problem, exemplified by an emission model. It has been observed, however, that most, if not all, published studies of uncertainties in emission inventory work have addressed only the precision component of uncertainty (e.g., EPRI 1981; SCAQMD 1982; Mangat et al. 1984; and Chun 1986).

1.2 SUBJECTIVE PERCEPTION OF UNCERTAINTY

When an engineer estimates a quantity by measurement, engineering analysis, or mathematical simulation, he ordinarily has a notion or belief as to the accuracy of his estimate. This subjective notion of uncertainty of the estimated value often finds expression in phrases such as "within ± 10 percent", "about right", "perhaps an underestimate", and so on. This handbook is intended to distill this subjectively perceived uncertainty onto a consistent, measurable probability scale.

Emission estimates are seldom made by direct measurement. Instead, most emission estimates are usually arrived at through emission models or algorithms which are supposed to show how emissions are related to process variables. These commonly consist of emission factor equations relating emissions to activity levels of particular emission sources or source categories. Emission estimates are therefore subject to both random error

(also called "precision uncertainty"), and systematic error (also called "bias uncertainty" or simply "bias"). This handbook discusses these two types of uncertainties and offers some practical guidance on how to quantify them.

1.3 ANALYSIS OF UNCERTAINTY: AGGREGATION OF ELEMENTS

Suppose, for example, that emission rate X is calculated from an emission model that is multiplicative in form and consists of factors X_1, X_2 and X_3. Further suppose that X_1, X_2 and X_3 are all measured accurately (i.e., with no bias) and their means $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$ and standard deviations s_1, s_2, s_3 are known. Then, is it possible to calculate the precision uncertainty of the emission rate? Yes, according to the statistical formula for a product of mutually independent variables, the precision uncertainty of $X = X_1 X_2 X_3$ is given by:

$$(s/\bar{X})^2 = (s_1/\bar{X})^2 + (s_2/\bar{X})^2 + (s_3/\bar{X}_3)^2 \qquad (1\text{-}1)$$

Although Equation (1-1) is based on the approximate formula that is good only when $(s/\bar{X})$ is much less than unity, it is assumed to hold for this particular example. Under this assumption then, does the precision uncertainty calculated from Eq. (1-1) really reflect uncertainty of the estimated emission rate? The answer is yes, provided that uncertainty of the emission estimate arises only from precision uncertainties of the estimated means $\bar{X}_1$, $\bar{X}_2$, and $\bar{X}_3$.

However, in most practical cases, other uncertainties enter because the emission model is only a crude approximation of the relationship between pollutant emission and its contributing factors. Moreover it not infrequently happens that an emission model may be accurate for the situation under which the model was developed, but grossly inaccurate for another situation to which the model is being applied. Uncertainty of the emission estimate due to this model representation inaccuracy cannot be estimated from uncertainties of the model parameters alone.

Since practically no emissions in an inventory are directly measured, the uncertainty due to possible model inaccuracies may be very important. It can be adequately assessed only by experts who are familiar with the

physical principles which operate to cause pollutant emissions and with the models and algorithms which have been used to simulate those principles. Although detailed analysis of this sort can be time-consuming and costly, the same experts who have the necessary understanding to achieve such an analysis will often have useful insights into the possible sources of bias in particular model applications. In evaluating uncertainties of emission estimates for inventory purposes, this type of uncertainty must always be considered and, where possible, the opinions of qualified experts should be obtained.

Therefore, the cruxes of uncertainty evaluation are:

1. How to project subjectively perceived uncertainty onto a uniform and consistent uncertainty scale;
2. How to link this subjectively perceived uncertainty to statistically determinable uncertainty; and
3. How to track uncertainties in individual estimates of emissions and emission model parameters through various multiplicative and additive processes, which are involved in emission inventory calculations.

1.4 A FRAMEWORK FOR EVALUATING INVENTORY UNCERTAINTY

Pollutant emissions from many point, area, and mobile sources have been estimated and compiled in inventories at all geographical levels by air pollution control agencies from the U.S. Environmental Protection Agency (EPA) to local air pollution control districts. These emission inventories, developed at great expense, are probably the most important practical tools available to many air pollution control agencies. Despite this importance, there is no widely accepted, mathematically sound procedure to estimate how accurate individual emissions estimates are or how accurate an inventory is as a whole.

Therefore, this handbook is intended to provide engineers and planners with mathematically sound, yet practical procedures to characterize the uncertainties of both the individual emissions estimates and the total inventory. Since nearly all emissions estimates are performed through engineering calculations instead of direct measurements, there is ordinarily no hard evidence to indicate their accuracy. Under these circumstances, it is particularly important to capture and utilize the analyst's

understanding as to the accuracy of his models and algorithms as well as his estimated values.

This handbook offers a method by which the opinions of experts about bias uncertainty in estimating emissions, from individual sources or from source categories, can be quantified and presented for ready comparison with precision uncertainty and other uncertainty components. Thus quantified, this uncertainty component will be referred to as "subjectively perceived bias uncertainty".

Since subjectively perceived uncertainty varies from analyst to analyst, this document provides a plausible (although unproven) method by which each analyst will undertake a specific mental exercise to distill his perceived uncertainty onto a probabilistic scale. This mental exercise is exemplified in a graphical procedure which requires the analyst to estimate the "probability" or "odds" that a "true value", i.e., the unknown but correct real-world value, would be below each of two thresholds. These thresholds are computed from a set of three plausible emission values: the basic estimate, an upper plausible estimate, and a lower plausible estimate.

Once uncertainties of individual estimates and model parameters are evaluated either statistically or subjectively, the analyst must calculate the overall uncertainty of an aggregated total or that of a source emission estimate from the emission model. Presented in Sections 2.2 and 2.3 are statistical formulas, and their underlying concepts, by which to calculate the overall uncertainty of the final estimate from the individual uncertainties as they propagate through the emission calculations.

Similar to the way in which a good emissions inventory is developed, uncertainty estimates for an inventory should be developed in a strategic manner:

- o First, a preliminary uncertainty is estimated for every source category;
- o Second, important source categories and suspected categories are identified for further analysis;
- o Third, additional efforts are strategically allocated to selected source categories where refinements in the estimates appear most likely to reduce overall uncertainty in the inventory;

- Fourth, task forces are formed to evaluate and refine uncertainty estimates of those selected categories;
- Fifth, uncertainty estimates for individual source categories, aggregated source categories, and for an entire inventory are compiled; and
- Sixth, salient features of inventory uncertainties are summarized and any problem areas are identified for future analysis.

Basic concepts and statistical theories involved in uncertainty estimation are discussed in Section 2.0 while illustrative examples of uncertainty estimation are given in Section 3.0 for point, mobile, and area sources. Section 5.0 discusses a few candidate methods by which the uncertainty of an emission inventory as a whole can be computed and reported.

2.0 METHODS FOR EVALUATING UNCERTAINTY

This section describes a method for establishing the uncertainty in the estimated mean of any randomly varying quantity, such as emission estimates generated for a given source category. Other examples of random variables are the emission factors and activity rates that are typically used in calculating those emissions. The uncertainty associated with these quantities has two components: (1) precision error, which indicates the imprecision in the estimated mean; and (2) bias, which reflects the inaccuracy of the estimated mean with respect to the true mean, which is always unknown.

Section 2.1 discusses a few basic concepts that are useful for determining the uncertainty of any estimated quantity. Section 2.2 presents statistical formulas for calculating individual variable uncertainties and for calculating errors propagated through the additive or multiplicative processes that may be used in estimating emissions or aggregating emissions estimates.

2.1 BASIC CONCEPTS FOR EVALUATING UNCERTAINTY

2.1.1 SUBJECTIVE-TO-OBJECTIVE LINKAGE

Like drivers who can estimate distance well enough to avoid traffic accidents, engineers are capable of estimating the range in which the true answer to an engineering problem is likely to be found. Because of this, it is possible for an individual experienced in a particular subject area to subjectively gauge the probability that the true value will be encompassed by a given range around an estimate of that value. Such a subjectively determined probability becomes increasingly accurate as the individual's knowledge and understanding of the matter increases.

This capability is useful because usually, variables in the emission models employed in inventory work are not measured repeatedly, and their precision is, therefore, not known.

The standard statistical treatment of the result of a series of measurements in which the individual values vary somewhat (from unassignable causes) is as follows: The set of measurements is considered to be a

random sample from a practically infinite set of possible measurements which (conceptually) could have been made, and it is assumed that the distribution of values in that infinite set would be Gaussian or "normal". The mean of the distribution of the conceptual infinite set is referred to as the "true mean". It is, of course, unavailable for study (since the actual set of measurements is never infinite). However, the mean of the values in the actual sample can serve as an approximation of the "true mean". Similarly, the standard deviation of the sample set can be taken (after adjustment for the sample size) as an approximation of the "true" standard deviation of the imagined infinite set. In particular, 95 percent confidence limits on the true mean are, in theory, given by taking a = 0.05 in the following formula:

$$m - t_{a/2,n}\,(sn^{-\frac{1}{2}}) \leqslant u \leqslant m + t_{a/2,n}\,(sn^{-\frac{1}{2}}) \qquad (2\text{-}1)$$

where u = true mean of the quantity

m = sample mean of the quantity

s = sample standard deviation of the quantity

n = sample size

$t_{a/2,n}$ = t-statistic for 100 (1-a) % confidence level for the sample size n.

In short, Eq.(2-1) indicates that the probability that the true mean u is in the range of $m = \pm\, t_{a/2,n}(sn^{-\frac{1}{2}})$ is (1-a). For $n \geqslant 30$ and a = 0.05, equation (2-1) reduces (approximately) to:

$$m - 2sn^{-\frac{1}{2}} \leqslant u \leqslant m + 2sn^{-\frac{1}{2}} \qquad (2\text{-}2)$$

Eq. (2-2) indicates that for a large sample size ($n \geqslant 30$) the true mean will be found in the range $m \pm 2sn^{-\frac{1}{2}}$ in 95 out of 100 chances.

In emissions inventory work, many quantities are deduced from other related quantities or computed through engineering analysis. Some quantities are estimated from survey results or repeated measurements, with sample sizes which are often less than 30. Under these circumstances, uncertainty (strictly speaking, "precision uncertainty") of an estimated quantity depends on such statistical parameters as those noted above.

Furthermore, this type of analysis is not even applicable to many parameters derived from engineering principles.

To alleviate this difficulty of precisely describing uncertainty mathematically, a new parameter, s', is defined such that the true mean, u, is expected, in 95 out of 100 chances, to be within the following range about the estimated value, m, of some emission-related quantity:

$$m - 2s' \leqslant u \leqslant m + 2s' \qquad \text{(m: unbiased estimate)} \qquad (2\text{-}3)$$

Unlike Eq. (2-2), Eq. (2-3) is applicable to any estimated value of a random variable. The new parameter s' is considered to be a quarter-width of the 95 percent confidence interval around the estimated value. It is understood that s' is not a statistical parameter in the conventional sense. It is rather a parameter associated with the probability that the true value u is in the range $m \pm 2s'$ in 95 out of 100 chances.

When a quantity is estimated by repeated measurements or a statistical survey, the new parameter s' is explicitly related to the following:

$$s' = t_{0.025,n}(sn^{-\frac{1}{2}}) \qquad (2\text{-}4)$$

In the equation, s is the sample standard deviation of samples of size n, while t is the t-statistic (whose value can be found in most statistical textbooks) for a = 0.05 and sample size n.

When the estimated value m is arrived at through engineering analysis or manipulation of external data for which the gathering method is unknown, the corresponding parameter s' must be estimated from the analyst's understanding in the estimated value. To minimize any conscious or unconscious cognitive bias on the part of the analyst, he is required to undertake a "debiasing" reasoning process in which his true subjective belief is measured on a probabilistic scale. The result of this process is an evaluation, in quantitative terms, of the analyst's subjectively perceived precision uncertainty.

2.1.2 PRECISION AND BIAS UNCERTAINTY

The uncertainty discussed in the preceding section was limited to precision uncertainty, that is uncertainty caused by random errors. The other type of uncertainty that may be found in emissions estimates is termed "bias uncertainty". This uncertainty is principally caused by systematic errors in repeated observations or by under- or over-estimates of the quantity of interest by some engineering analysis or model simulation. Bias uncertainty is related to "accuracy" in that it refers to how close the measurement is to the true value. On the other hand, precision uncertainty is similar to "precision" which refers to how close the measured values are to each other.

However, both bias and precision uncertainties are different from their statistical counterparts in that the former apply not only to errors in the repeated observations, but also to misestimates and impreciseness of the estimated values that may be arrived at by engineering analysis.

Suppose that the estimated value m is found to contain both the precision uncertainty measured as s' and the bias uncertainty measured as B. Then, what is the range that contains true value u in 95 out of 100 chances? One obvious range is given by:

$$m - |B| - 2s' \leqslant u \leqslant m + |B| + 2s' \qquad (m : \text{biased estimate}) \qquad (2\text{-}5)$$

where $|B|$ is the absolute value of B whose sign is yet unknown.

The magnitude of the above range is $(2|B| + 4s')$. Is there any way to reduce this range so that the true value is trapped in a tighter range at the same confidence level as before? Yes, it is possible. Consider the following:

First, define an unbiased sample mean as:

$$m' = m + B \qquad (\text{for } B \geqslant 0) \qquad (2\text{-}6)$$

$$m' = m - B \qquad (\text{for } B < 0) \qquad (2\text{-}7)$$

Then, the range in which true mean u is expected to be found in 95 out of 100 chances is given by:

$$m' - 2s' \leq u' \leq m + 2s' \qquad (m' : \text{unbiased estimate}) \qquad (2\text{-}8)$$

Using the unbiased estimate m', the range of uncertainty reduces to 4s' instead of (2|B|+ 4s') with the biased estimate m. This reduction in the uncertainty range is brought about by separating the bias component from the overall uncertainty in the estimate. Therefore, in this handbook, precision uncertainty and bias uncertainty are always treated separately. This separate treatment of precision and bias uncertainty also makes it possible to track errors propagated through both multiplicative and additive processes involved in emission calculations and aggregations.

2.1.3 AN ANALOGY: MARKSMANSHIP WITH CANNON

This handbook addresses the uncertainty of an emission estimate in both precision and bias uncertainties. Since systematic errors propagate differently from random errors through individual emission calculations and aggregations of individual source emissions into a higher source category, any uncertainty estimation method based on precision uncertainty alone is doomed to fail in fully addressing uncertainties of emission estimates and those of resulting emission inventories.

Figure 2-1 illustrates an analogy of emission estimation to cannon firing. Based on estimates of the distance to the target and the meteorological condition under the situation, an artilleryman sets the angle and direction of the cannon to fire at the target.

The objective of the artilleryman is to land the cannon shells within a certain area around the target, as shown in the figure. The precision of the cannon is only one of the factors that determine whether the shells fall in the target area. Other determinants would include the correctness of a distance estimate made by a frontman, the accuracy of the meteorological condition predicted by a meteorologist, and the cannon setting made by an artilleryman.

CANNON FIRING	VS.	EMISSION ESTIMATING
Frontman		"Source" Engineer
Provides target information		Provide activity/emission information
Artilleryman		Inventory Analyst
Sets cannon		Sets modeling condition
Fires cannon		Provides emission estimate
Spotter		Evaluator
Determines a hit or miss		Determines accuracy of the estimate

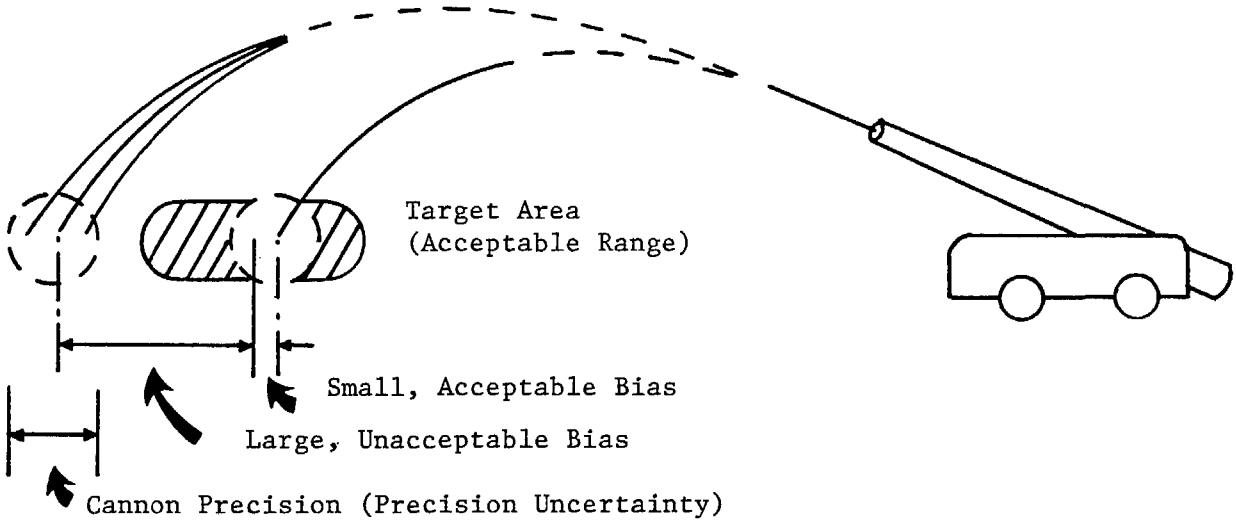


Figure 2-1 Analogy of Cannon Firing to Emission Estimating.

By analogy to this example, it is suggested that uncertainty of an emission estimate can not be fully ascribed only to that associated with the emission model used or with the variables used in the model. The uncertainty would also be dependent upon the accuracy of the input data used and the correctness of the assumptions or the premises under which the analyst calculates the emission. As in firing cannon, the most important concern in estimating emissions should be whether the estimate is a correct estimate; that is, whether, like the cannon shell, it lands within the target area.

In the case of cannon firing, a spotter identifies whether shells are landing in the target area and whether a hit has occurred. According to the spotter's information, the artilleryman corrects the cannon setting to improve the chance of hitting the target. Uncertainty analysis of the emission estimate should be made analogously to the relationship between the spotter and the artilleryman. Specifically, the primary objective of uncertainty analysis of emission estimates should be to determine whether the estimates of various source emissions are in the right range and whether the uncertainties thus determined are acceptable for each use of individual emission estimates and for the emission inventory as a whole.

The crux of the problem of determining uncertainties of emission estimates is how to set up an uncertainty estimation problem so that an emission expert can play the same role in uncertainty determination as does the spotter in cannon firing. Keeping this analogy in mind, users of this handbook are encouraged to consider uncertainty problems in perspective of an entire emission estimation process rather than in isolation.

2.2 OBJECTIVE METHOD FOR COMPUTING UNCERTAINTY

A review of inventory methods and algorithms currently in use shows that many are quite involved. The individual parameters, for instance, range from simple discrete datum points to parameters of complex equations encompassing power terms, summations, ratios, and so forth. Nevertheless, all of these algorithms appear to be reducible to an archetypal emission model of the form:

$$EE = (AR)\ (AF)\ (EF)\ (CF) \qquad (2\text{-}9)$$

where

EE = emission estimate for the category;

AR = activity rate, e.g., material or fuel throughput, VMT, etc.

AF = adjustment or allocation factor (which equals unity if not needed);

EF = emission factor in terms of emissions per unit of AR; and

CF = control factor, which equals one for no control and zero for full control.

When some or all of the factors on the right side of Eq. (2-9) are measured or observed repeatedly to estimate their values, an objective method based on a theory of statistical distribution becomes applicable to the calculation of precision uncertainties of individual variable estimates and their aggregates. Among many statistical distributions, two well-known theoretical distributions have been used almost exclusively for uncertainty analyses of emissions estimates: normal distribution and lognormal distribution. In the latter, logarithms of original data values are distributed normally. This handbook addresses these two distributions only.

In most emission estimates, only one or two factors in the general emission model given by Eq. (2-9) have sufficient numbers of measurements or observations so as to objectively compute their precision uncertainty. Furthermore, the total uncertainty of an emission estimate includes not only precision uncertainty, but also bias uncertainties in individual factors and representation errors in the emission model. Since there are very few statistical methods that are available for determining bias uncertainties and representation errors, there is clearly a need for a method of assessing the subjectively perceived uncertainty and including it in the total uncertainty of any emission estimate.

Nevertheless, users of this handbook are encouraged to compute precision uncertainties of individual factors and emission estimates whenever adequate observational data are available. Although this precision uncertainty alone does not represent, in general, the overall uncertainty of an emission estimate, it does provide a lower bound of the overall uncertainty of an emission estimate. Therefore, precision uncertainty for the emission estimate, computed in this manner, can be used to

check whether a subjectively assessed value for the overall uncertainty of the emission estimate is large enough to be plausible. This type of check will be useful for controlling unrestrained use of subjective methods of assessing overall uncertainty (including both precision and bias uncertainties) of emission estimates.

2.2.1 UNCERTAINTY FOR A NORMALLY DISTRIBUTED VARIABLE

When measured values of a quantity scatter around its sample mean rather symmetrically, the quantity can be said to be normally distributed or its distribution is nearly normal. A normal variant, X, with mean u and variance σ^2 is expressed as:

$$X = u + \sigma U \tag{2-10}$$

where U is a standard normal distribution with mean 0 and variance 1, namely, U(0,1). In a similar notation, the normal distribution of X is expressed as $N(u,\sigma^2)$. Given a set of observational data, (X_1, X_2, ..., X_n), the population mean u and variance σ^2 are estimated by a sample mean, m, and a sample variance, s^2, which are given, respectively, by:

$$m = SUM\ (X_i)/n \tag{2-11}$$

$$s^2 = SUM\ (X_i - m)^2/(n-1) \tag{2-12}$$

where n is the sample size, i.e., the number of observations. The square roots of the variances, namely and s, are called, respectively, population standard deviation and sample standard deviation.

2.2.1.1 Characteristics of Normal Distribution

One characteristic of a normal distribution is the symmetric nature of the distribution around the mean. When actual data values are plotted on a graph, many data points are clustered near the sample mean, with progressively fewer data points as the distance between a data point and the mean increases. Figure 2-2 shows an example of a quantity which

should have a normal distribution. Here, 100 repeated measurements of a ring are plotted in a histogram form.

The data in Figure 2-2 are plotted on "normal probability paper" in Figure 2-3. By definition, a normal distribution is expressed by a straight line on normal probability paper. In the figure, error bounds around the sample mean, m, with the 95 percent confidence level (hereafter called "95 percent confidence interval") are shown for three different sample sizes, n = 100, 10 and 1. The confidence intervals were computed using Eq. (2-1) with the sample mean m = 20.03 mm and the sample standard deviation s = 0.48 mm. It should be noted that the confidence in an estimate increases dramatically from the estimate based on a single measurement to that based on 10 repeated measurements and to the one based on 100 repeated measurements, as exemplified in the figure by the reduced confidence intervals for n = 10 and n = 100.

The above example illustrates the power of repeated measurements for reducing errors in an estimate. Although many emission estimates are arrived at by engineering estimates or with very few measurements, it is highly desirable that at least some factors in an emission estimation formula be measured repeatedly to increase confidence in their estimates.

2.2.1.2 Error Propagation in Normally Distributed Variables

Now, to define precision uncertainty first consider a random variable X with mean u and variance σ^2. Then, standard deviation σ is usually taken to be the precision uncertainty of the mean u. Some may consider the ratio of to u to be a better indicator, however:

$$CV = \sigma/u \qquad (2\text{-}13)$$

where CV is called a "coefficient of variation".

This coefficient of variation is used as a measure of precision uncertainty of an estimated quantity in this handbook. Since CV is a ratio, it is dimensionless and independent of the magnitude of the mean. Thus, CV can be used to compare the relative precision uncertainty of one estimated quantity with that of another quantity even when the magnitudes of the two quantities are vastly different.

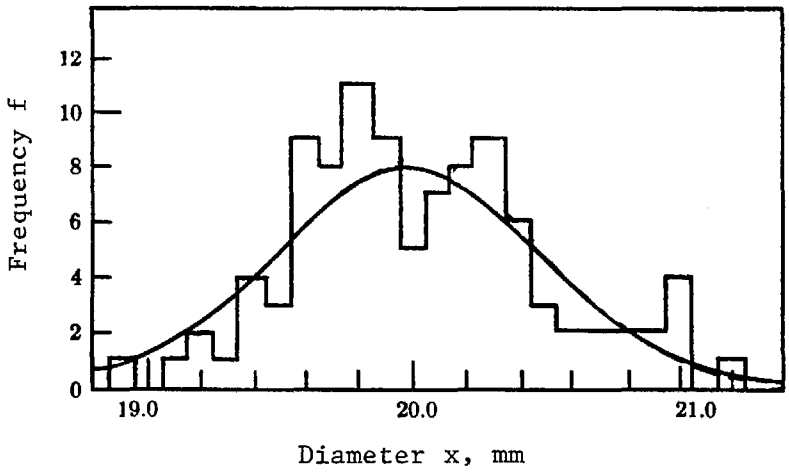


Figure 2-2 Graph of Frequency f vs. Measured Diameter of a Ring. (A student made 100 measurements using a common ruler.)

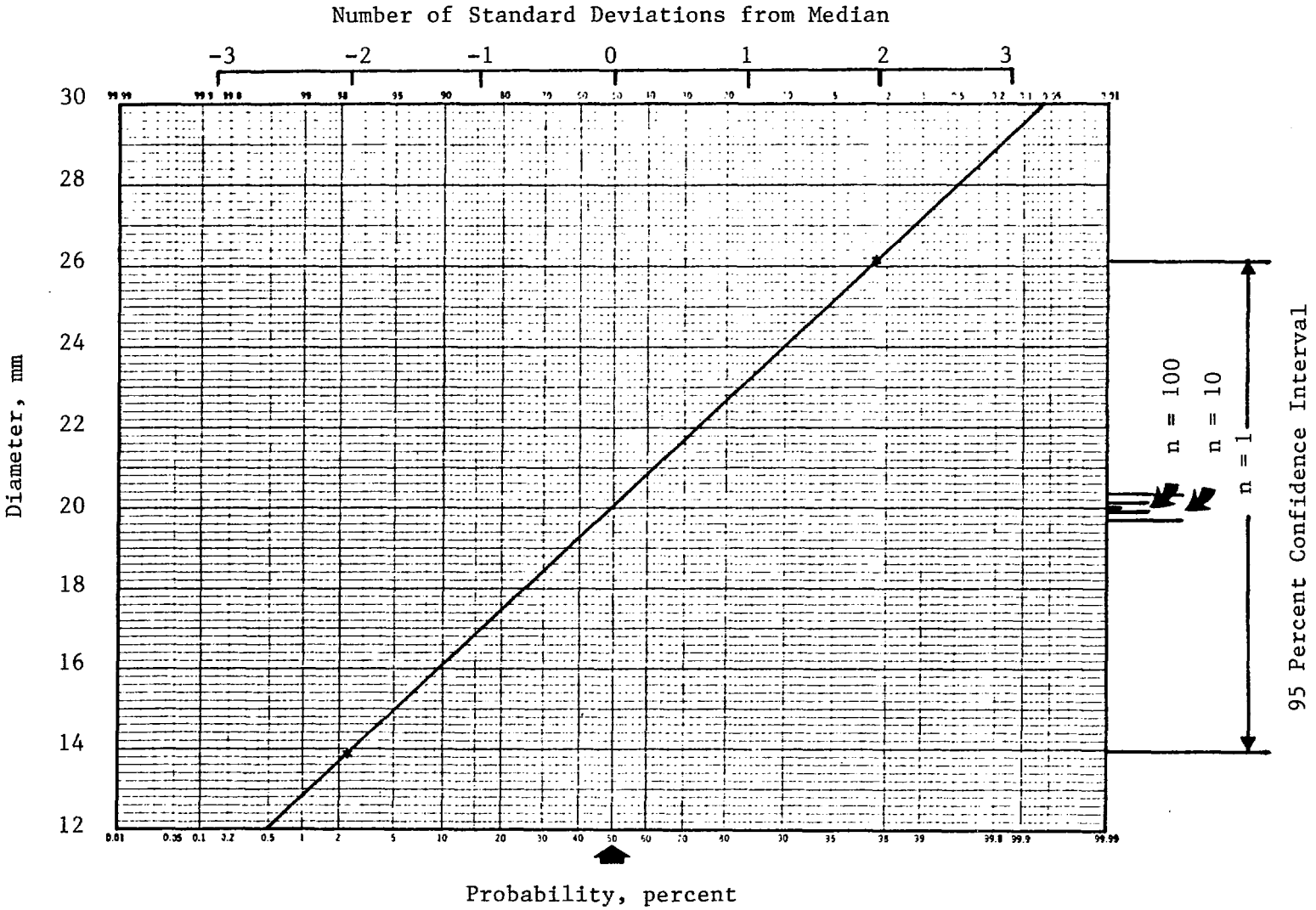


Figure 2-3 Distribution of Ring Diameter and 95 percent Confidence Intervals for the Three Different Numbers of Measurements.

Precision Uncertainty in Multiplicative Process

In Eq. (2-10), the emissions are formulated as the product of the four simple variables: activity rate, adjustment factor, emission factor and control factor. In actual practice, however, emissions are often estimated as products of many variables:

$$X = X_1 X_2 \ldots X_n \tag{2-14}$$

where X is an emission rate at a source or a group of sources, and X_1, X_2, . . ., X_n are pertinent model factors.

In practical terms, precision uncertainty of X can be computed only when X_1, X_2, . . ., X_n are mutually independent. Under this independence assumption, one can write:

$$\begin{aligned} E(X) &= E(X_1 X_2 \ldots X_n) \\ &= E(X_1)\, E(X_2) \ldots E(X_n) \\ &= u_1\, u_2 \ldots u_n \end{aligned} \tag{2-15}$$

$$\begin{aligned} VAR(X) &= E[(X_1 X_2 \ldots X_n)^2] - [E(X_1 X_2 \ldots X_n)]^2 \\ &= E(X_1^2)E(X_2^2) \ldots E(X_n^2)) - (u_1 u_2 \ldots u_n)^2 \\ &= (u_1^2 + \sigma_1^2)(u_2^2 + \sigma_2^2) \ldots (u_n^2 + \sigma_n^2) \\ &\quad - (u_1 u_2 \ldots u_n)^2 \end{aligned} \tag{2-16}$$

$$\begin{aligned} CV(X) &= SQRT\,[VAR(X)]/E(X) \\ &= SQRT\,[(1 + \sigma_1^2/u_1^2)(1 + \sigma_2^2/u_2^2) \ldots (1 + \sigma_n^2/u_n^2) - 1] \end{aligned} \tag{2-17}$$

In the equations, E(X) means mathematical expectation of X and yields the mean of X. VAR(X) is the variance of X; its square root yields σ, the standard deviation of X. CV(X) is the coefficient of variation for X and indicates how large the spread of the distribution is relative to the mean. The letters u_i and σ_i are the mean and the standard deviation of X_i.

Suppose that u_i's and σ_i's are the same for all X_i's. Then by letting $p = \sigma_i/u_i$, Eq. (2-18) reduces to:

$$CV(X) = SQRT\ [(1 + p^2)^n - 1] \qquad (2\text{-}18)$$

The relationship between an aggregate uncertainty CV(X) and individual uncertainty p is plotted in Figure 2-4 by using Eq. (2-18).

Figure 2-4 shows that the relative uncertainty of an aggregate variable X is always equal to or greater than that of an individual factor and becomes much greater as the number of factors increases. For example, for the individual uncertainty p = 0.5 and the number of factors n=7, which is not particularly large for an emission estimating model, the figure yields the aggregate uncertainty CV = 1.9. This is nearly four times as large as the individual factor uncertainty.

Precision Uncertainty in Additive Process

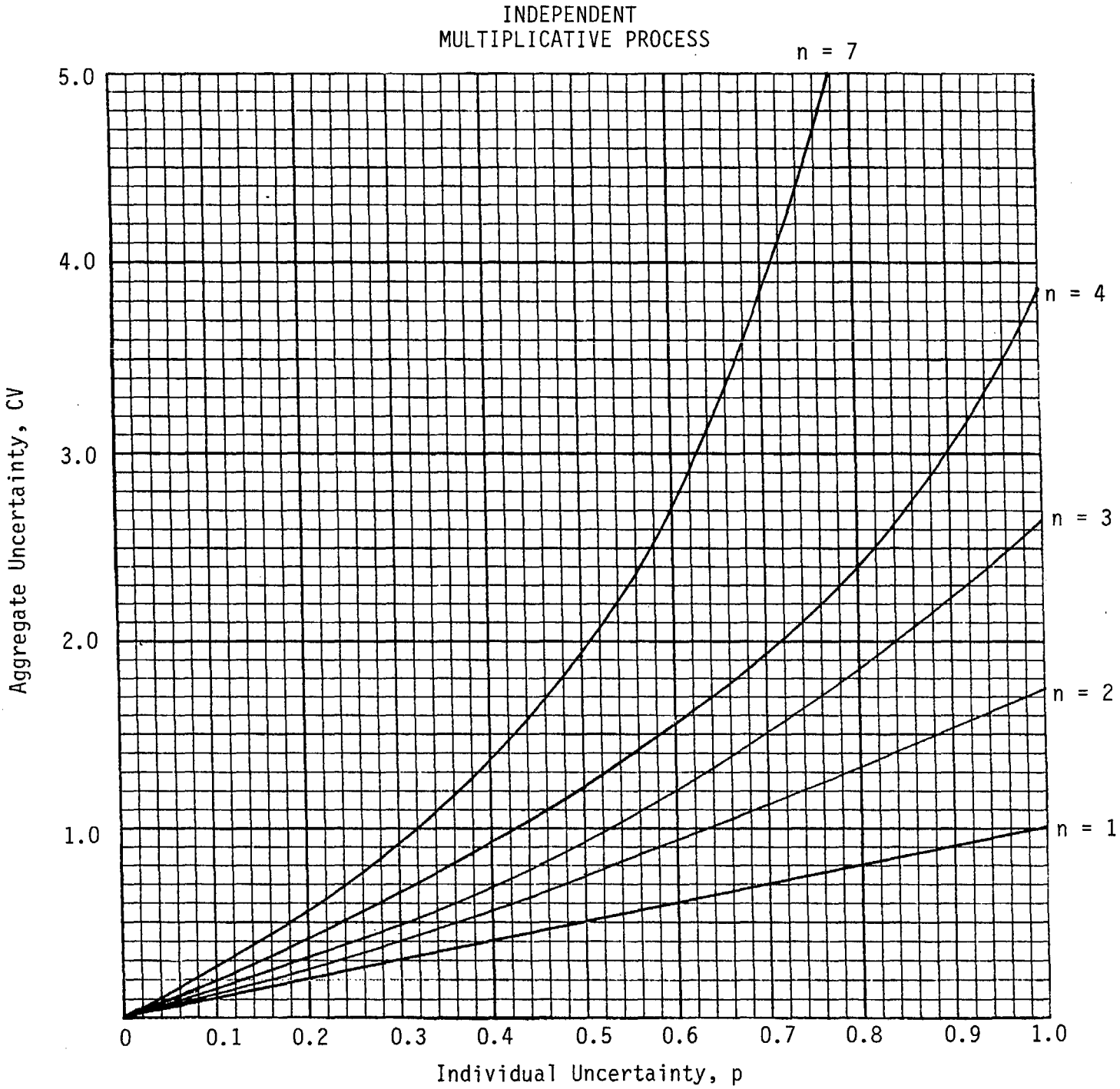
In many source categories, total emissions for the source category are estimated as the sum of individual source emissions or the sum of subregional emissions. In these additive processes, uncertainties of individual emissions estimates propagate differently from those in a multiplicative process. Unlike the multiplicative case, uncertainty of an aggregated emissions estimate in an additive process can be computed for both mutually uncorrelated (i.e., independent) cases and mutually correlated cases:

$$X = X_1 + X_2 + \ldots + X_n \qquad (2\text{-}19)$$

$$E(X) = E(X_1 + X_2 + \ldots + X_n) = u_1 + u_2 + \ldots + u_n \qquad (2\text{-}20)$$

$$VAR(X) = SUM\ (\ \sigma i^2) + 2 \underset{i<j}{SUM}\ (\ \sigma_{ij}) \qquad (2\text{-}21)$$

The second term in the right side of Eq. (2-21) indicates that covariance σ_{ij}'s between variables X_1 and X_j for all i's less than j are summed.



$$CV = [(1 + p^2)^n - 1]^{\frac{1}{2}}$$

Figure 2-4. Aggregate Uncertainty in Independent Multiplicative Process.

Eq. (2-21) can be rewritten by using a correlation coefficient c_{ij} instead of covariance σ_{ij} as:

$$VAR(X) = SUM\ (\sigma_i^2) + 2 \underset{i\ \ j}{SUM} (c_{ij}\ \sigma_i\ \sigma_j) \tag{2-22}$$

To examine some properties of Eq. (2-22), let $\sigma_i = \sigma$ and $c_{ij} = c$ for all i less than j. Then, Eq. (2-22) reduces to:

$$\begin{aligned} VAR(X) &= SUM\ (\sigma_i^2) + 2\ c \underset{i<j}{SUM} (\sigma_i\ \sigma_j) \\ &= SUM\ [(\sigma_i^2) + 2 \underset{i<j}{SUM} (\sigma_i\ \sigma_j)] \quad - 2(1-c) \underset{i<j}{SUM} (\sigma_i\ \sigma_j) \\ &= [SUM\ (\sigma_i)]^2 - 2\ (1-c) \underset{i<j}{SUM} (\sigma_i\ \sigma_j) \\ &= (n\sigma)^2 - 2\ (1-c)\ (\tfrac{n}{2})\ \sigma^2 \\ &= (n\sigma)^2\ (1 - (1-c)(1 - \tfrac{1}{n})) \end{aligned} \tag{2-23}$$

Similarly, let $u_i = u$. Then, a coefficient of variation for the aggregated variable X is given as:

$$\begin{aligned} CV(X) &= SQRT\ (VAR(X))\ /\ E(X) \\ &= (n\sigma/nu)\ SQRT\ [\ 1-(1-c)\ (1 - \tfrac{1}{n})\ ] \\ &= p\ SQRT\ [\ 1 - (\ 1-c)\ (1 - \tfrac{1}{n})\ ] \end{aligned} \tag{2-24}$$

where $p = \sigma/u$ is the uncertainty of each individual variable X_i. From Eq. (2-24), relationships between CV(X) and p can be derived for the following special cases:

CASE I: Mutually independent variables (c = 0)

$$CV(X) = p\ n^{-\frac{1}{2}} \tag{2-25}$$

CASE II: Perfectly correlated variables (c = 1)

$$CV(X) = p \tag{2-26}$$

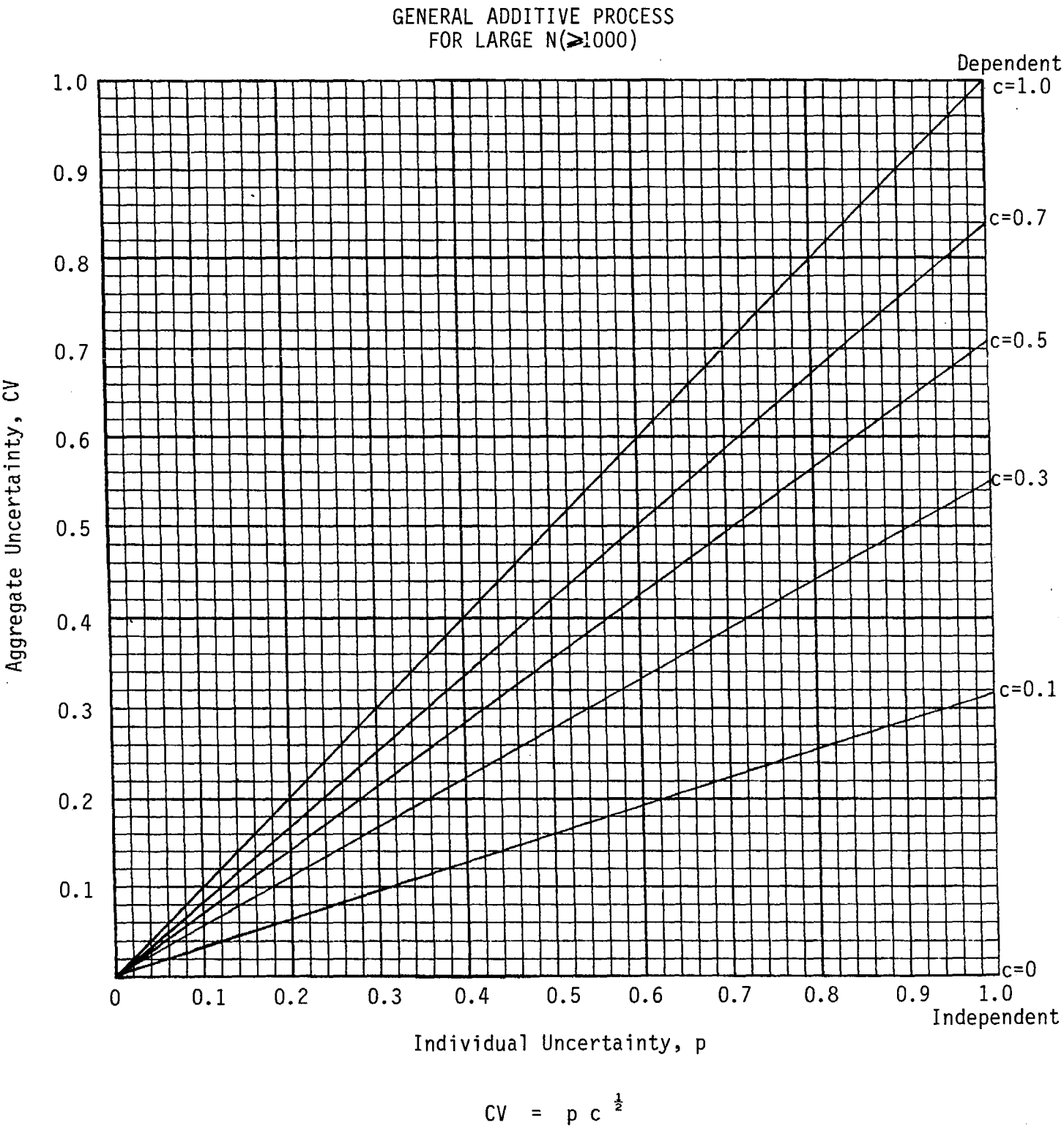
CASE III: Large number of variables ($n \to \infty$)

$$CV(X) = pc^{-\frac{1}{2}} \tag{2-27}$$

Eq. (2-25) indicates that the relative uncertainty of the sum of emissions is less than those of individual emissions estimates and progressively decreases as the number of individual emissions that are added together increases. On the other hand, Eq. (2-26) indicates that if emissions are perfectly correlated with each other, the sum of emissions has the same uncertainty as the individual emissions estimates. This difference between independent variables and correlated variables seems to be crucial in computing uncertainty of an emissions inventory. One needs to know the likely correlation between emissions in different sources and source categories in order to estimate the uncertainty of an aggregate emissions estimate.

Eq. (2-27) links the independent, Eq. (2-24), and dependent, Eq. (2-25), relationships by a correlation coefficient c for a large n (n=1000 is large enough to be accurate). Figure 2-5 shows the relationships between the uncertainty of an aggregate variable and that of an individual variable for various degrees of correlation among the variables. It should be noted that in this graph and in Eq. (2-26), all variables (not just two variables) are correlated with each other with a designated correlation coefficient. It should also be remembered that equal correlation among many variables is possible only for a positive correlation coefficient.

Figure 2-5 shows that if a large number of individual source or source category emissions are uncorrelated with each other (i.e., c=0), uncertainty of the sum of those emissions goes to zero regardless of uncertainty of the individual estimates. On the other hand, if individual emissions are totally correlated with each other (i.e., c= 1), uncertainty of the sum of those emissions remains the same as that of individual emissions. Uncertainty of the sum of partially correlated emissions falls in between these two extremes. It should be noted that even under a lightly correlated case of c = 0.3, uncertainty of the sum of emissions remains as high as 55 percent of those of individual emissions whereas with mutually independent (i.e., uncorrelated) emissions it decreases to zero percent. Since emissions in different source categories or emissions from different sources in one source category are expected to be at least



$$CV = p\, c^{\frac{1}{2}}$$

Figure 2-5. Aggregate Uncertainty for Large Number of Sources under Various Degrees of Correlation.

partially correlated, the use of the assumption of independence may lead to an unrealistically small uncertainty value for an aggregate estimate.

Figure 2-6 provides a relationship between uncertainty in mutually uncorrelated individual estimates and that in aggregated estimates for various numbers of aggregated sources or source categories. For totally uncorrelated emissions, the uncertainty in an aggregate estimate becomes progressively smaller relative to those in the individual estimates as the number of sources to be aggregated increases. As the number grows infinitely large, uncertainties in individual emissions cancel each other, tending to yield zero uncertainty in the aggregated emissions.

2.2.2 UNCERTAINTY OF LOGNORMALLY DISTRIBUTED VARIABLE

The natural logarithm of a lognormally distributed random variable is normally distributed. Since the logarithm of a product of factors is equal to the sum of the logarithms of individual factors, the lognormal distribution may be considered as the multiplicative analog of the normal distribution.

Indeed, many physical processes encountered in emission inventory studies are best described by lognormal distribution. Increasingly, it is being found that the outputs of many physical processes encountered in environmental pollution, particularly in emission of pollutants to the atmosphere, are lognormally distributed or are more nearly lognormal than normal. It is a distribution that engineers and planners in emission inventory studies must familiarize themselves with.

2.2.2.1 Characteristics of Lognormal Distribution

Figure 2-7 shows frequency curves for the standard normal and lognormal distributions. As seen from the figure, the lognormal distribution is characterized by positive values only and a heavy tail having large positive values. In a lognormally distributed variable, the arithmetic mean is always greater than the median, which happens to be equal to the geometric mean. On the other hand, the mode (that is, the value occurring at the greatest frequency) is always less than the median.

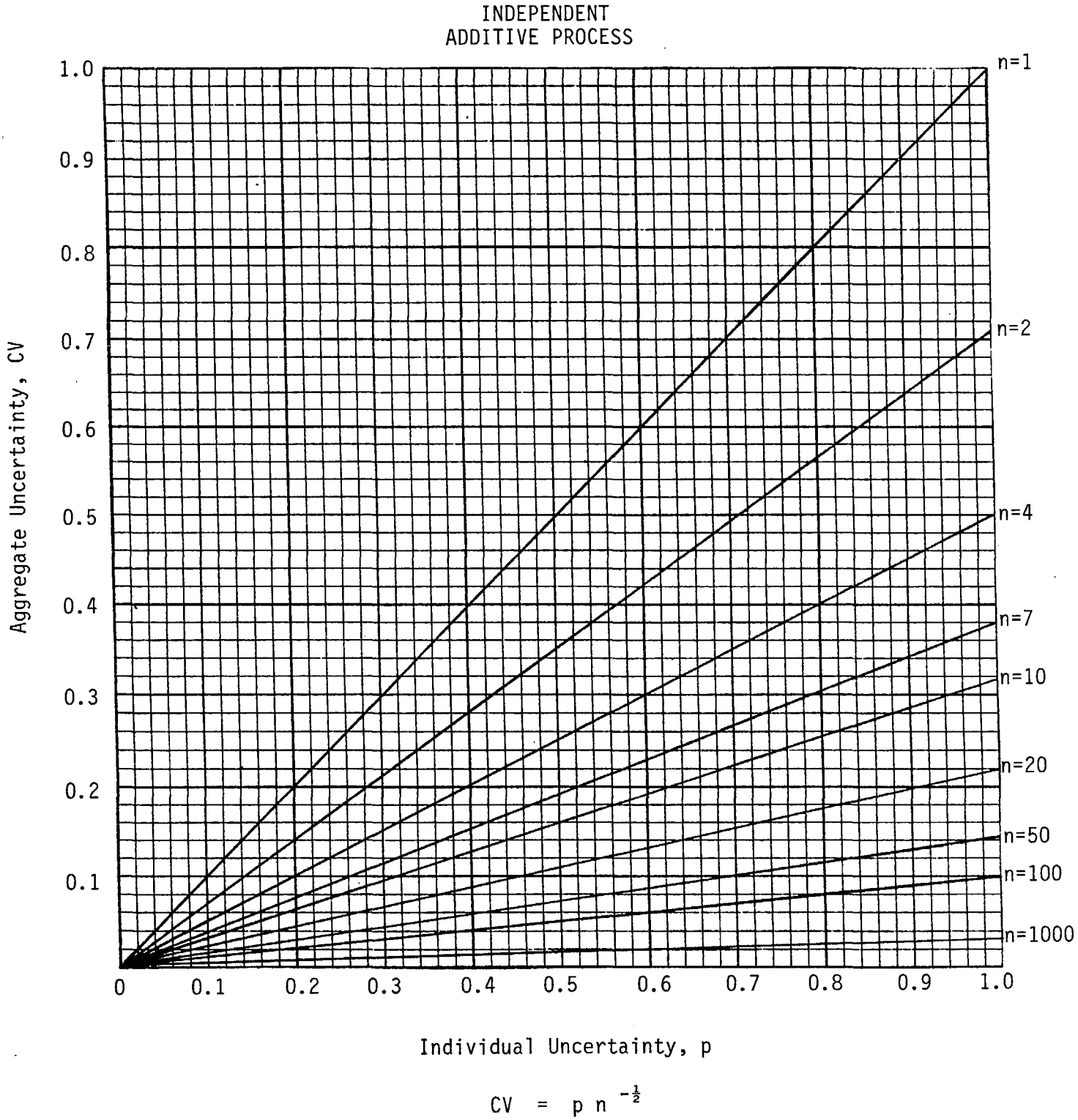


Figure 2-6. Aggregate Uncertainty in Independent Additive Process.

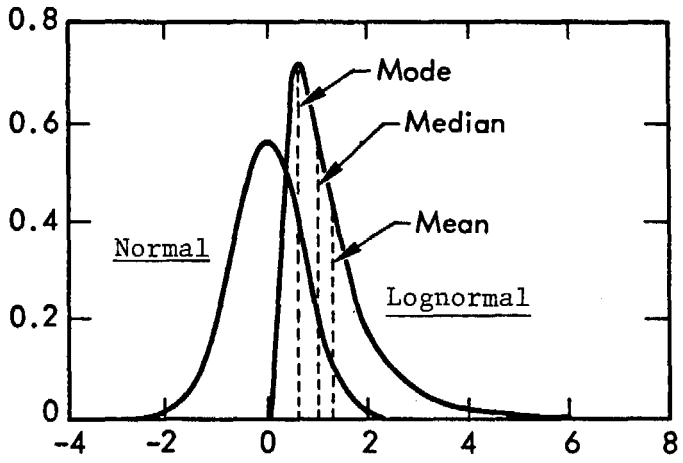


Figure 2-7 Frequency Curves of the Normal and Lognormal Distributions.

The geometric mean, m_g, and geometric standard deviation, s_g, of a lognormal variate, X, are defined as:

$$m_g = \exp [\text{SUM } (\ln X)/n] \tag{2-28}$$

$$s_g = \exp [\text{SUM } (\ln X) - \ln m_g)^2 /n]^{0.5} \tag{2-29}$$

For a lognormal distribution, the arithmetic mean, m, geometric mean, m_g, standard deviation, s, and geometric standard deviation, s_g, are related as follows:

$$s_g = \exp [\ln^{0.5} (s^2/m^2 + 1)] \tag{2-30}$$

$$m_g = m / [\exp (0.5 \ln^2 s_g)] \tag{2-31}$$

Although a lognormal distribution and a normal distribution are, in principle, quite different, the two distributions are actually quite similar if the coefficient of variation is much smaller than unity. Let X_n and X_l be normal and lognormal variates, respectively. Then, using a standard normal distribution U, the distribution of a normal variate, X_n, is expressed as:

$$X_n = u + \sigma U \tag{2-32}$$

where u and are the mean and standard deviation of X_n.

Similarly, the lognormal variate, X_l, is expressed as:

$$X_l = u \exp (\sigma U/u) \tag{2-33}$$

Using a Taylor expansion of the exponent, Eq. (2-33) is rewritten as:

$$X_l = u\{ 1 + (\sigma/u)U + 0.5 (\sigma/u)^2 U^2 + ... \} \tag{2-34}$$

If the coefficient of variation, CV = σ/u, is much smaller than unity, Eq. (2-34) reduces to

$$X_1 = u + \sigma U \tag{2-38}$$

This equation for X_1 is the same as Eq. (2-32) for X_n. This shows that when the coefficient of variation for a lognormal variate, X, is small, the distribution can be approximated by a normal distribution. In other words, it is not important to know whether the distribution is more nearly normal of lognormal, as long as the coefficient of variation is small.

Based on the above analysis, this handbook recommends that the uncertainty calculation formulas for a normal variate be used as long as the coefficient of variation is less than 0.3. At CV = 0.3, the numerical error using Eq. (2-35) instead of the defining equation, Eq. (2-33) is only 5 percent at one standard deviation (i.e., U = 1) from the mean.

2.2.2.2 Error Propagation in Lognormally Distributed Variables

For normal variates, there is a simple algebraic formula for computing errors propagated through sums or products of many variables. For lognormal variates, however, no such formula exists. To overcome this lack, this handbook provides a computer program for simulating distributions in order to compute uncertainties associated with individual variables and their aggregates (Appendix A). The program generates randomly selected values, R, for a normal variate, X, according to:

$$R = A + B * RN \tag{2-36}$$

where A and B are, respectively, the specified mean and standard deviation of X, and RN is a random value drawn from a standard normal distribution.

Similarly, random values, R, for a lognormal variate, X, are generated from:

$$R = \exp (A + B * RN) \tag{2-37}$$

The program generates 1000 random values from one simulation run, repeats the simulation ten times, and generates percentiles and distribution parameters for averages of the ten simulation results. The distribution parameters calculated are:

arithmetic mean
arithmetic standard deviation
geometric mean
geometric standard deviation
minimum
maximum

The basic input parameters to the program are:

A = Mean of a normal variate, N (A, B)

B = Standard deviation of a normal variate, N (A,B)

C = Flag for a desired distribution (0 for a regular distribution and 1 for a positive-only distribution)

D = Distribution type (0 for a normal distribution and 1 for a lognormal distribution)

E = Variable combination (0 for additive combination and 1 for multiplicative combination)

F = The number of random variables.

Since an aggregative combination of two random variables (e.g., $X_1 + X_2$ or X_1X_2) does not yield either a normal or lognormal distribution, the program has a subroutine that describes the resulting distribution by specifying 19 pre-selected percentiles (0.05, 0.1, 0.2, 0.5, 1, 2.5, 5, 16, 30, 50, 70, 84, 95, 97.5, 99, 99.5, 99.8, 99.9, and 99.95). Should the resulting distributions be used for a further combination with another variable, say, X_3, random values for the resulting distribution are generated by repeatedly applying a random pointer selected from a uniform distribution, U(0-1), to the quantile (percentile) data set. These random values are then combined with random values generated from the distribution of X_3 to yield a new distribution for the combination of X_1, X_2, and X_3.

Users of this handbook are cautioned that the program listed in Appendix A is for research use only. It is not designed for routine use as many packaged programs are. To use the program in any other way than as described above, the user must write his own main program. However,

since the program is written in completely modular format, that task should not be very difficult.

2.3 SUBJECTIVE METHOD FOR ASSESSING UNCERTAINTY

Although the uncertainty estimation procedures described in Section 2.2 are mathematically definitive, they are applicable only to precision uncertainties or random errors and not to errors caused by biases. As discussed in the Cannon Firing example (Section 2.1.2), uncertainties of emission estimates can not be fully described without quantifying bias uncertainties in those estimates. Further support for the importance of this statement can be found in the report of the National Acid Precipitation Assessment Program (NAPAP), published by the U.S. Environmental Protection Agency (EPA 1986).

In the NAPAP report, quantitative estimates of emissions uncertainties for the 1980 NAPAP Emission Inventory were developed by applying a conventional statistical concept -- error propagation formulas for mutually independent, normally distributed variables -- to the emission inventory case. Initial uncertainty estimates for individual source categories were provided by a panel of emission inventory experts. Then, uncertainties of various aggregated source categories and for the entire inventory were computed under the following assumptions:

- Emission calculation parameters are independent, that is, they do not covary.
- Emission factors represent true mean values.
- All estimates are unbiased.
- The emission parameters can be treated as random variables which are approximately normally distributed.
- No coding or transcription errors are present.
- The data are complete; no emissions data, emissions sources, or emissions source categories are missing.

Starting with seemingly reasonable uncertainty estimates at a basic source category level, uncertainties of higher source categories decreased steadily as aggregation of source categories progressed. In the end, the study found:

- The values of uncertainty estimates for national levels of aggregation appear to be unreasonably small;

- Although uncertainty values for SO_2 emissions were set to be higher than those for NO_x emissions at a basic source category level, the opposite relationships emerged at national levels; and
- To resolve the above inconsistencies with our common sense understandings of emission estimation, the methodology must be expanded to incorporate bias, coding, and emission errors.

The authors of this handbook have arrived at similar conclusions through extensive review of the literature and through mental experiments on uncertainty estimation. All literature on emission uncertainties reviewed by the authors framed uncertainties in individual emission estimates, as well as in entire inventories, in terms of precision uncertainties or random errors only. In reality, however, errors in emission estimates have a deterministic component (exemplified in biases) as well as a probabilistic component (exemplified in random errors). Since biases and random errors have distinctly different propagation characteristics, any uncertainty assessment efforts based on precision uncertainties alone can only result in incoherent and idiosyncratic results, as reported in the NAPAP study. In the hope of making a useful contribution on this problem, VRC has devised and tested several subjective methods of determining uncertainties in emission estimates on a very limited level. After a few trials and errors, a plausible but unproven method of delineating both precision and bias uncertainties of a given emission estimate, utilizing expert opinion, was devised and is offered in this handbook.

2.3.1 PREPARATION OF SUBJECTIVE ASSESSMENT TOOL

To make a rational judgment as to reliability and validity of an emissions estimate, an evaluator must know what set of assumptions and premises were employed in generating the estimate, as well as what kinds of data were used and what calculations were made. Without such basic information on the estimate, quantification of precision and bias uncertainties would vary greatly from one evaluator to another depending on their perception and knowledge of emission factors, activity levels and control efficiencies for the emission source.

To minimize the variation of subjectively assessed uncertainties from one evaluator to another, this handbook recommends:

1. An analyst who participated in deriving the emission estimate shall document concisely yet comprehensively the emission estimation process, the data used, and the assumptions and premises employed; and
2. An evaluator or evaluators shall be selected from emission inventory experts who have broad knowledge of pollutant emissions but have not participated in preparing the estimates in question.

Although the analyst might be the same person who computed the emissions or made a major contribution in preparing the emission estimate, the evaluator must be a different person. The reason for this is to prevent the recognized cognitive bias known as "anchoring" from influencing an uncertainty assessment by the evaluator. If the evaluator were the same person who made the emission estimate, his assessment of uncertainties of the estimate would be subtly "anchored" to rationalize the estimate regardless of the presence or absence of contrary indications. Such "anchoring" will subvert the purpose of obtaining an unprejudiced, though subjective, uncertainty estimate.

In preparing a subjective uncertainty assessment tool, the emission estimation procedure used for the emission estimate must be described in plain English. Included in the description must be the assumptions and premises employed, the nature of and problems associated with the data used, and the emission estimation method. In addition to the emission estimate used in the inventory, this handbook recommends that every analyst make two additional estimates: the upper plausible estimate and the lower plausible estimate.

As will be seen from examples given in Section 3.0, these two additional estimates can be made without much extra effort. Every emission estimate employs a definite, specifiable set of assumptions and premises, a particular data set as opposed to an alternative data set, or perhaps one estimation method as opposed to another. All an analyst needs to do is to replace the one of those elements used in the original emission estimate with an equally plausible alternative element, in order to obtain a revised emission estimate for comparison with the original.

2.3.2 DELINEATION OF PRECISION AND BIAS UNCERTAINTIES

Let E_o, E_u, and E_l be, respectively, the basic emission estimate, upper plausible estimate and lower plausible estimate for a given source or source category. From these three emission estimates, the analyst calculates the mean, m, and standard deviation, s, of the three estimated values:

$$m = (E_o + E_u + E_l)/3 \tag{2-38}$$

$$s^2 = \sum_{i=1}^{3} (E_i - m)^2/2 \tag{2-39}$$

Using these results, he next calculates an upper threshold level, UL, and a lower threshold level, LL, as:

$$UL = m + s \tag{2-40}$$

$$LL = m - s \tag{2-41}$$

Then, E_o, E_u, E_l, m, UL, and LL are plotted on normal probability paper as shown in Figure 2-8.

With this probability graph and the summary description of his emission estimation methods, including those for the upper and lower threshold estimates, the analyst asks an evaluator to review his emission estimation methods. Upon completion of the review, the evaluator is asked to assess the probability or odds that the true emission value (which is unknown) would, if known, be less than UL. Similarly, he is asked to provide his estimated probability that the true value would be less than LL.

In estimating the probability or odds for UL, the evaluator should first consider whether the true value is more likely to be below UL or above UL. If he thinks that the true value is likely to be below UL, he should assign a probability of greater than 50 percent, say 70. If he thinks the opposite is more likely, he should assign a probability of less

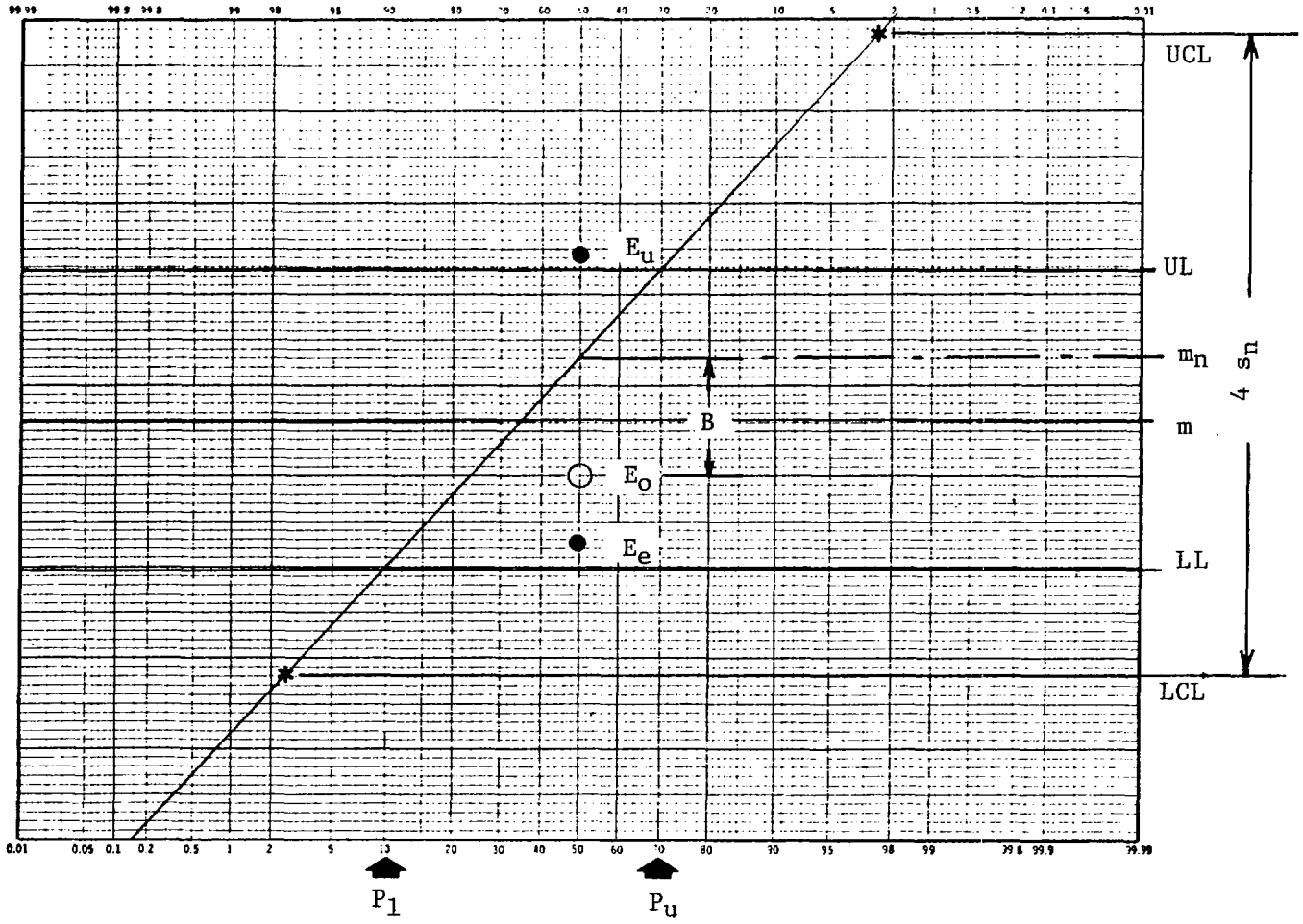


Figure 2-8 Graphical Presentation of Three Emission Estimates and Resulting Precision and Bias Uncertainties.

than 50 percent, say, 10. If he thinks that the true value is equally likely to be either above or below UL, he should assign a probability of 50 percent or very near 50. If he strongly believes that the true value must be below UL, then he should assign a very high probability, such as 99.99 percent. (This probability means that the odds that the true value is below UL are 9,999 to 1.)

Suppose that the evaluator assigns probabilities of 70 percent for UL and 10 percent for LL. (It should be noted that the probability for LL must always be less than that for UL.) Based on these values for UL and LL, a slanted line is drawn in Figure 2-8. By extending the line to the 97.5 and 2.5 percentiles, one can read from the graph both upper confidence level (UCL) and lower confidence level (LCL) for the evaluator's uncertainty assessment. His assessment for the true mean, m_n, is read from the graph at the intersection of the slanted line and the 50 percentile.

The difference between the basic emission estimate, E_o, and the subjectively assessed true mean, m_n, is viewed as a bias assessed by the evaluator:

$$B = E_o - m_n \tag{2-42}$$

The distance between UCL and LCL is four standard deviations. Thus, the standard deviation assessed by the evaluator is given by:

$$s_n = (UCL - LCL)/4 \tag{2-43}$$

where the subscript n indicates that the underlying distribution is assumed to be normal. Should the ratio of this standard deviation to the assessed mean, (s_n/m_n), exceed 0.3, then the procedure should be repeated using lognormal probability paper. (Lognormal probability paper differs from normal probability paper in that the ordinate of the graph is expressed in a logarithmic scale instead of a linear scale.)

The uncertainty assessment procedure described above can be carried out quickly using the worksheet shown in Figure 2-9. Items in the upper half of the worksheet are quantities estimated by the analyst while those

in the lower half are answers by the evaluator or quantities derived from those answers. As stated earlier, the subscripts n and l stand for a normal distribution and lognormal distribution. CV means a coefficient of variation which is given by the following ratio:

$$CV = s_n / m_n \tag{2-44}$$

When a subjective uncertainty assessment is made twice (the first using normal probability paper and the second using lognormal probability paper), the resulting uncertainty parameters (m_n, s_n, B_n) and (m_l, s_l, B_l) should be shown to the evaluator and his preference solicited on. This preference is indicated by marking either "normal" or "lognormal" distribution, as listed in the last line of the worksheet.

UNCERTAINTY WORKSHEET

NAME_____ DATE_____

Three estimates for emissions: TPD

Mean of three estimates: m = TPD
Standard deviation: s = TPD

Upper Level: m + s = TPD
Lower Level: m - s = TPD

Probability for UL _____%
Probability for LL _____%

Upper Confidence Level @ 97.5% _____ TPD
Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_n = _____ TPD
Subjectively evaluated standard deviation: s_n = _____ TPD
Subjectively evaluated bias: B_n = _____ TPD
Coefficient of variation CV = _____

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% _____ TPD
Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_l = _____ TPD
Subjectively evaluated standard deviation: s_l = _____ TPD
Subjectively evaluated bias: B_l = _____ TPD

Prefer: [ ] Normal [ ] Lognormal

Figure 2-9. Sample Worksheet for Subjectively Assessed Uncertainties

3.0 EXAMPLES OF EMISSION UNCERTAINTY CALCULATIONS

To illustrate the uncertainty estimation methods described in Section 2.0, this section presents three real-world examples of emission estimates and associated uncertainties. All three examples reflect actual emission estimates included in the 1982 SoCAB emissions inventory. The first example pertains to NO_x emissions from power plant boilers, the second to total organic gas (TOG) exhaust emissions from light duty vehicles, and the third to TOG emissions from dry cleaners.

These three examples represent typical and clearly defined emission categories for, respectively, point sources, mobile sources, and area sources. Each example first describes how the basic emission estimate and two limiting plausible estimates are arrived at, next presents appropriate questions to solicit judgments of experts on probabilities associated with upper and lower levels computed from the three emission estimates, and finally illustrates how to assess, both subjectively and objectively, the uncertainties associated with the basic emission estimate.

3.1 POWER PLANTS

Pollutant emissions from power plants have been studied more exhaustively than from any other source categories. For years, many power plants have been subject to periodic source tests and rigorous reporting of daily operating conditions, such as fuel burning rate and electricity generated. In recent years, emission estimation for some power plants has been further advanced by the introduction of continuous emission monitoring systems (CEM).

3.1.1 METHOD FOR ESTIMATING POWER PLANT EMISSION

There are two methods for estimating NO_x emissions from power plant boilers: the (old) calculation method and the CEM method. The calculation method has now been phased out in favor of the CEM method. However, in the 1982 emission inventory for the SoCAB, emissions from Southern California Edison Company (SCE) power plants were estimated using the calculation method whereas emissions from Department of Water and Power (DWP) power plants were estimated by the CEM method. Figure 3-1

shows typical NO_x curves, used in the calculation method to compute NO_x emissions as a function of megawatts of electricity generated. Figure 3-2 shows a comparison between monthly emissions estimated by the calculation method and those by CEM. Although the two methods in general yield rather similar emission values, they disagree considerably for some power plants, and especially under start-up conditions.

Appendix B describes how an analyst (i.e., the author) re-computed 1982 basinwide power plant emissions using mean ratios of CEM-based emission estimates to calculation-method-based emission estimates for the SCE and DWP power plants. The basic emission estimate and the analyst's upper and lower plausible estimates are:

Upper Plausible Estimate	39.8 tons per day (TPD)
Basic Estimate	39.4 TPD
Lower Plausible Estimate	38.7 TPD

3.1.2 SUBJECTIVELY ASSESSED UNCERTAINTY OF POWER PLANT EMISSION ESTIMATES

Based on the three estimates arrived at in Section 3.1.1, the analyst computed the upper and lower threshold levels using Eqs (2-40) and (2-41) as:

$$UL = 39.3 + 0.56 = 39.9 \text{ TPD}$$

$$LL = 39.3 - 0.56 = 38.7 \text{ TPD}$$

He then asked of a panel five power plant experts in the South Coast Air Quality Management District the following two questions:

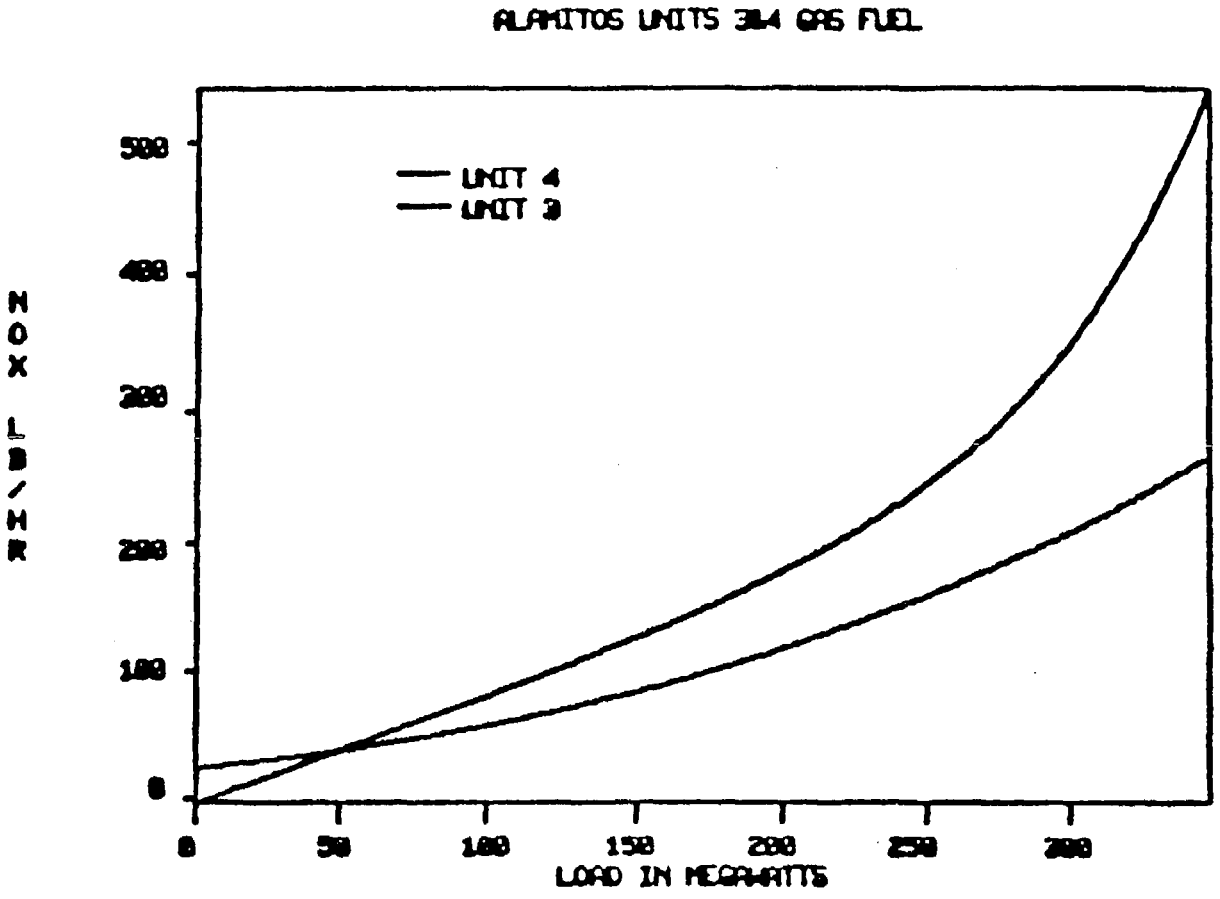
Q1. On a scale of 100, what are the odds that the true level of NO_x emissions from all power plants in this air basin is less than 39.8 TPD?

Answer: _____ chances in 100.

Q2. What are the odds that the true level of NO_x emissions from all power plants in the air basin is less than 38.7 TPD?

Answer: _____ chances in 100.

Answers from the five panelists are summarized in Table 3-1.



Unit 3	Unit 4
A = -939.7	A = -4.8
B = -1.2611	B = .8101
C = 963.22	C = .57
D = .001532	D = .01756

$NO_x = A + BX + Ce^{DX}$

where NO_x in lb/hr and X in MW.

Figure 3-1 Power Plant Boiler NO_x Curves for Alamitos Units 3 and 4 with Gas Fuel

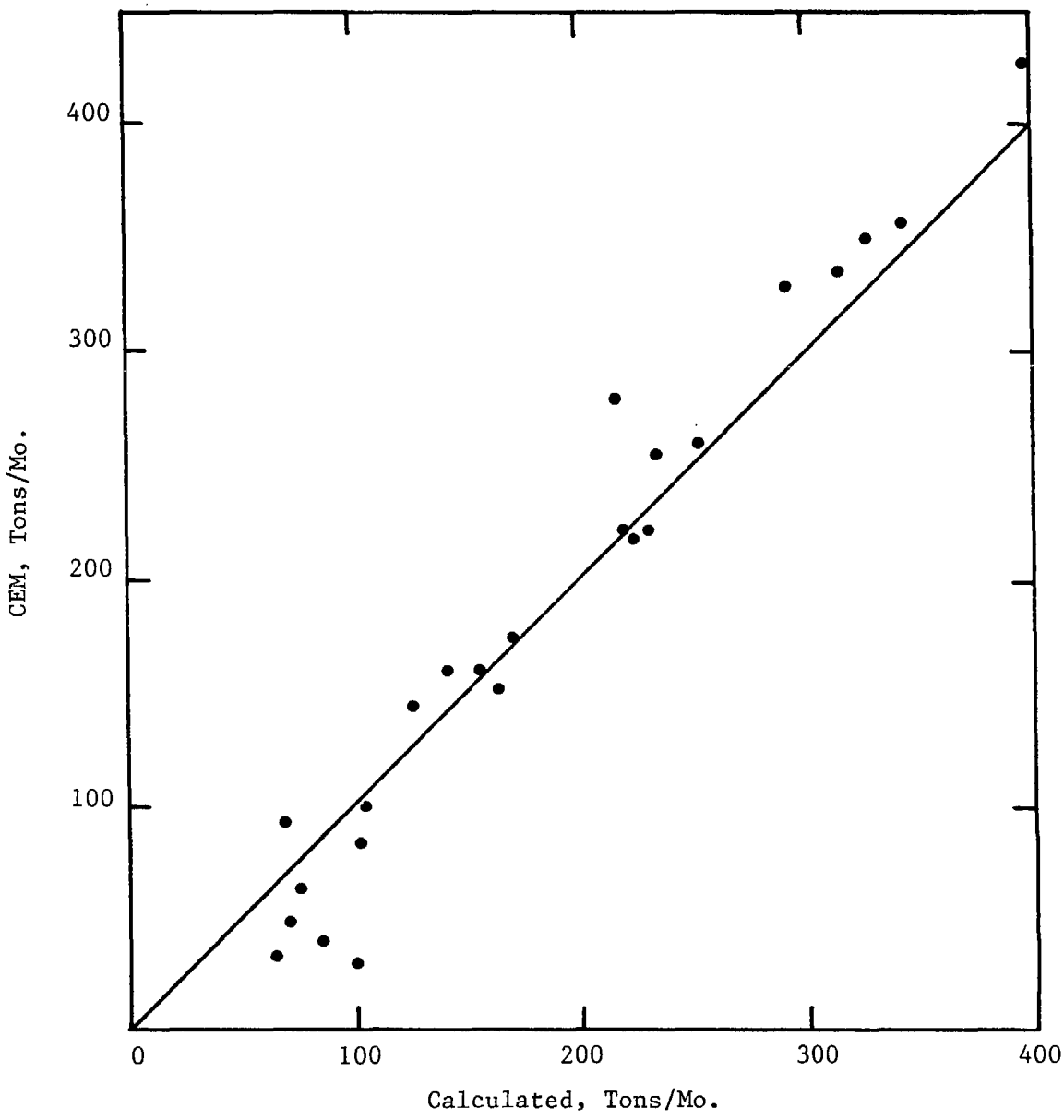


Figure 3-2 Scatter Plot of Monthly NO_x Emissions Computed by Old Calculation Method and CEM for Haynes Power Plant in 1982 and 1983.

TABLE 3-1. SUMMARY OF RESPONSES FROM FIVE EVALUATORS ON THE 1982 BASINWIDE POWER PLANT NO_x EMISSIONS

Evaluator	1	2	3	4	5	Median	Mean
Odds for UL	80	50	70	60	90	70	70
Odds for LL	10	30	50	30	70	30	38

In a Delphi-type survey, the median is generally taken to indicate a consensus opinion. However, in this example, the quantity sought is probability, a continuous and bounded variable ranging from 0 to 1. Therefore, a mean may be more appropriate than a median for indicating a consensus opinion.

If the mean is taken for this purpose, the consensus opinion of the five power plant emission experts is 70 percent for L_1 and 38 percent for L_2. Based on this consensus response, both precision and bias uncertainties of the basic emission estimate for basinwide powerplant NO_x emissions are estimated graphically in Figure 3-3. As seen from the graph, the bias uncertainty is negligible (B = 0.1 TPD) while the precision uncertainty is rather substantial (2s = 2.9 TPD). These results are summarized in a sample work sheet (Figure 3-4) and appear to reflect the experts' confidence that the basic emission estimate is a correct estimate although the estimate is subject to substantial precision uncertainty.

3.1.3 OBJECTIVELY DETERMINED UNCERTAINTY FOR POWER PLANT EMISSIONS

According to the SCAQMD Laboratory, satisfactory performance of a CEM system is specified as follows:

Quarterly Test

- Conduct 9 sets of flue gas tests.
- Maximum relative error ±20%

Daily Calibration

- Maximum allowable error with calibration gas ±5%
- Maximum allowable error in flow rate ±10%

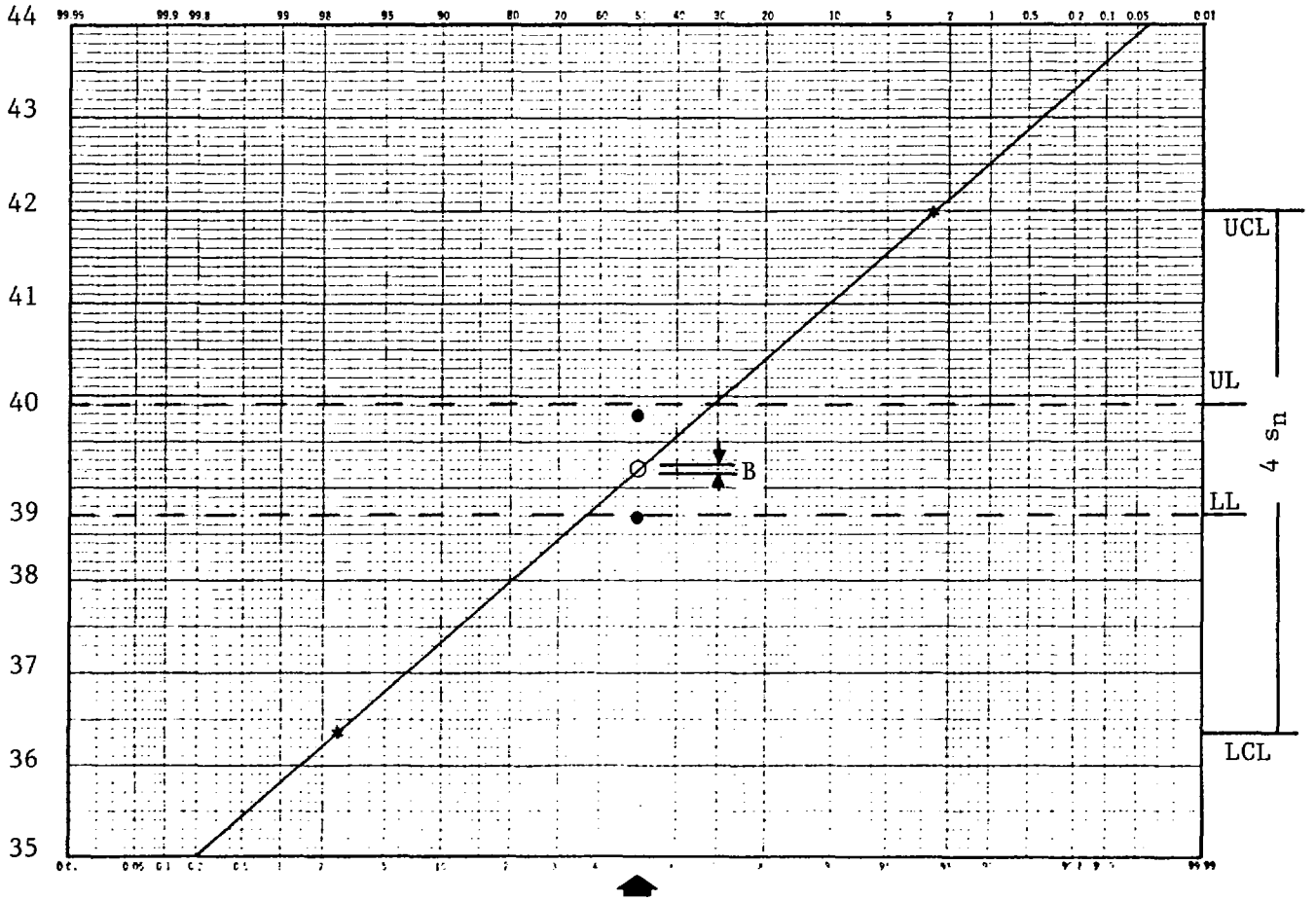


Figure 3-3 Subjectively Assessed Uncertainty of the 1983 Basinwide Power Plant NO_x Emissions.

UNCERTAINTY WORKSHEET

NAME Panel Composite Response DATE January 1988

Three estimates for power plant emissions : 39.8, 39.4, 38.7 TPD

Mean of three estimates: $m = 39.3$ TPD

Standard deviation: $s = 0.56$ TPD

Upper Level: $m + s = 39.9$ TPD

Lower Level: $m - s = 38.7$ TPD

Probability for UL (70) %

Probability for LL (38) %

Upper Confidence Level @ 97.5% 42.00 TPD

Lower Confidence Level @ 2.5% 36.35 TPD

Subjectively evaluated mean: $m_n =$ 39.1 TPD

Subjectively evaluated standard deviation: $s_n =$ 1.4 TPD

Subjectively evaluated bias: $B_n =$ 0.1 TPD

Coefficient of variation CV = 0.04

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% _____ TPD

Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: $m_l =$ _____ TPD

Subjectively evaluated standard deviation: $s_l =$ _____ TPD

Subjectively evaluated bias: $B_l =$ _____ TPD

Prefer: [ ] Normal [ ] Lognormal

Figure 3-4. Sample Worksheet for Uncertainties in Power Plant Emissions.

It is unclear whether this performance specification is any indication of errors in actual CEM measurements. However, for simplicity, we assume that each CEM system exhibits, on average, 20-percent random error in daily measured NO_x emissions (i.e., $CV_d = 0.20$) and that individual daily NO_x emissions are statistically independent of each other. Then, the relative error in annual NO_x emissions at a given power plant can be computed from those of daily emissions as:

$$CV_a = SQRT\ [SUM\ (\sigma_i^2)]\ /\ SUM\ (u_i) \qquad (i=1, 2, \ldots, n)$$
$$= SQRT\ [SUM\ (u_i CV_i)^2]\ /u_a \qquad (3\text{-}1)$$

where

CV_a = coefficient of variation (or relative error) for an annual emission,

CV_i = coefficient of variation for the i-th daily emission,

u_i = mean emissions for the i-th day,

u_a = mean annual emissions, and

n = number of days in the year.

Again, for simplicity, assume that u_i and DV_i are the same for all days, namely, $U_i = U_d$, and $CV_i = CV_d$ $(i = 1, 2, \ldots, n)$.

For $n = 365$, $CV_i = CV_d = 0.20$ and $u_i = u_d$, Eq. (3-1) yields.

$$CV_a = SQRT\ [\ n\ (u_d\ CV_d)^2] \ /\ (n\ u_d)$$
$$= CV_d\ /\ n^{\frac{1}{2}}$$
$$= 0.20\ /\ (365)^{\frac{1}{2}}$$
$$= 0.010$$

Therefore, the relative random error of the annual emission would be only 1 percent, or one twentieth of the daily relative random error.

In the old calculation method, the relative error in daily NO_x emission would be about 20 percent for baseload units and 30 to 50 percent

for intermediate and peaking units. For illustration, assume 50 percent relative error for the latter units.

$$CV_a = CV_d / n^{\frac{1}{2}}$$
$$= 0.50 / (365)^{\frac{1}{2}}$$
$$= 0.026$$

Even with this large relative error in daily emissions, that of the annual emissions is only 2.6 percent.

Uncertainty in basinwide power plant emissions is given by a combination of uncertainties in individual power plant emissions as in the case of annual emissions versus daily emissions. For illustration purposes only, assume that in 1982 there were 4 CEM-equipped power plants accounting for about 30 percent of the basinwide power plant emissions, 3 base load power plants accounting for about 50 percent of the total and 8 intermediate- and peaking-unit power plants accounting for the remaining 20 percent in the SoCAB. Furthermore, insofar as power plants are in the same inventory category (i.e., CEM equipped, base load, or peaking unit), all power plants regardless of size are assumed to emit the same amount of NO_x per Kwh. Then, assuming statistical independence in individual power plant emissions, the relative error in the basinwide power plant emissions is computed using Eq. (3-1) as:

$$CV_{basin} = SQRT\ [SUM\ (u_i\ CV_i)^2] / SUM\ (u_i) \qquad i = 1, 2, \ldots, 15$$
$$= SQRT\ [SUM\ (m_j\ (u_j CV_j)^2)] / SUM\ (m_j u_j) \qquad j = 1, 2, 3$$
$$= SQRT\ [\ 4\ (0.30 \times 0.010/4)^2 + 3(0.50 \times 0.010/3)^2$$
$$+ 8\ (0.20 \times 0.026/8)^2]$$
$$= 10^{-3}\ SQRT[2.25 + 8.33 + 3.38]$$
$$= 3.74 \times 10^{-3}$$

where m_j and u_j are, respectively, the number and mean emissions of power plants in Group j. Thus, according to the statistical formula for a precision uncertainty, the relative error in the basinwide power plant emissions is estimated to be mere 0.4 percent.

The above example demonstrates that as long as there are many emission events, relative random error of aggregate emissions drops to a negligibly small value regardless of the magnitude of relative error in

individual emission events. This finding is consistent with the relationship depicted in Figure 2-6.

It is interesting to note that the power plant emission experts estimated the relative error to be 4 percent (see CV in Figure 3-4) whereas the objective method yielded a relative error of 0.4 percent, i.e., a tenth of the subjectively assessed uncertainty. This may be reasonable, since the objective method provides precision uncertainty only.

3.2 LIGHT DUTY VEHICLES

Emissions from light duty passenger vehicles account for about 60 percent of total mobile source TOG emissions and 45 percent of NO_x emissions. Since mobile sources account for about a half of the basin total emissions for these two pollutants, light duty autos (LDA) should be the most important source category for emission inventories in general and for inventory uncertainty in particular. Unlike power plant emissions, emissions from LDAs are not measured directly at the point of exhaust but are estimated through a complex emission factor model and a transportation model.

3.2.1 SUBJECTIVELY ASSESSED UNCERTAINTY OF LDA EMISSION ESTIMATE

To estimate total organic gas (TOG) emissions from light duty passenger vehicles for the 1983 SoCAB inventory, the latest emission factor model, EMFAC7B, was used for computing a composite emission factor while the CALTRANS transportation model, DTIM, was employed for computing vehicle miles traveled (VMT) and average vehicle speeds during morning peak, afternoon peak, and off-peak hours. The resulting TOG emissions were estimated to be 359 TPD (see Appendix C).

An analyst carefully reviewed the estimation methods and data used for calculating the emissions from light duty autos in the SoCAB. His review identified two apparent errors in the emission calculation:

1. The transportation model appears to have underestimated the VMT value by 3 percent; and
2. The emission factor model appears to have underestimated the emission rate by 14 percent.

Based on these findings, the analyst made the following three plausible estimates (see Appendix C):

Upper Plausible Estimate	397	TPD
Basic Estimate	359	TPD
Lower Plausible Estimate	348	TPD

Using these three estimates, the analyst calculated the upper and lower threshold levels as UL = 394 and LL = 342 TPD. He then asked three mobile source experts (two in the ARB El Monte office and one in SCAG, a planning agency for the SoCAB area) the following questions:

Q1. On a scale of 100, what are the odds that the true level of TOG exhaust emissions from all light duty autos in this air basin is less than 394 TPD?

Answer: _____ chances in 100.

Q2. What are the odds that the true level of TOG exhaust emissions from all light duty autos in the air basin is less than 342 TPD?

Answer: _____ chances in 100.

The three experts read the analysis summary given in Appendix C and made their best guesses based on their knowledge of emission factors and activity levels of light duty autos in the SoCAB. Their answers are listed in Table 3-2.

TABLE 3-2. SUMMARY RESPONSES FROM THREE EVALUATORS ON THE 1983 BASINWIDE TOG EMISSIONS FROM LIGHT DUTY AUTOS.

Evaluator/Odds*	Individual Response 1	2	3	Consensus Response Median	Mean
Odds for UL	90	50	60	60	67
Odds for LL	10	10	30	10	17

*Out of 100 chances

The first evaluator's response yielded bias uncertainty of B = -9 TPD and precision uncertainty of 2s = 40 TPD whereas the second evaluator's response gave B = -35 TPD and 2s = 79 TPD. The reasons for the negative biases (indicating an underestimation in the basic emission estimate) in the two evaluations are as follows:

1. The average speeds predicted by the transportation model appear to be higher than actual average speeds, and the lower the speeds, the higher the TOG emissions; and
2. Ten to fifteen items that have been identified but not yet incorporated into the modeling all tend to suggest an increase in the TOG emissions.

On the other hand, the third evaluator's response yielded B = -18 TPD and 2s = 135 TPD. The reason given for the large precision uncertainty and rather small bias uncertainty assessed by this evaluator is as follows:

3. Simulations of vehicular activities in the SoCAB involve many assumptions and enormous calculations using a large volume of input data. This tends to yield estimates with large uncertainties, but those estimates would be as robust as anyone can hope for.

Documentation of these comments solicited from the evaluators appears to be a useful adjunct of the uncertainty estimation activity under consideration. There are two ways to integrate individual responses to a consensus response: one is to focus on the median or mean of responses (as illustrated in the power plant example); and the other is for the analyst to integrate subjectively both responses and comments of the evaluators. For example, the three comments mentioned above imply the following:

- o The first two evaluators think that the basic estimate is probably too low, based on their knowledge of the emission factor model; and
- o The third evaluator (who is an expert on the transportation model) thinks that uncertainty of any estimate based on the model must be quite large, e.g., ±135 TPD or ±38 percent of the original estimate in this case.

In their comments, the first two evaluators indicated their concern about the accuracy of the emission factor model whereas the third evaluator based his concern on the transportation model. Their responses should be regard as complementary rather than contradictory. Thus, one way that

the analyst might choose to integrate the information, to avoid underestimating the uncertainty, could be:

B = Average of responses of the first two evaluators
= -22 TPD

2s = Response of the third evaluator
= 135 TPD

As shown in the sample worksheet of Figure 3-5, values of m_n, s_n, CV and UCL and LCL were calculated from the basic emission estimate of 359 TPD, the bias of -22 TPD, and the precision uncertainty of 135 TPD. Then, plotting those values on normal probability paper, probabilities for the upper and lower threshold levels were graphically determined as seen from Figure 3-6. These probabilities are 65 percent for UL and 13 percent for LL. It is interesting to note that the two probability values arrived at from the analyst's integrated response are in rather close agreement with the consensus responses given in Table 3-2, particularly those by the mean method (e.g., 65 vs. 67 for UL and 13 vs. 17 for LL).

The good agreement between the consensus probabilities for UL and LL and those derived from the analyst's integrated response could be a mere coincidence. Nevertheless, such a good agreement provides additional credibility for the results obtained by either the integrated method or the consensus method.

3.2.2 OBJECTIVELY CALCULATED UNCERTAINTY OF LDA EMISSION ESTIMATE

According to the ARB report (ARB 1986), the general exhaust emission factor for a given vehicle fleet, calendar year (calyr), and pollutant TOG, is given by:

$$EF = SUM\ (BER * OMTCF * SCF * MISCF * TF_{vmt}) \qquad (3\text{-}2)$$

where

EF = emission factor in grams per mile
SUM = summation from model year (my)=calyr-24 to my=calyr+1
BER = basic exhaust emission rate in grams per mile
OMTCF = operating mode and temperature correction factor
SCF = speed correction factor

UNCERTAINTY WORKSHEET

NAME Analyst Integrated Response DATE January 5, 1988

Three estimates for LDA exhaust emissions: 397, 359, 348 TPD

Mean of three estimates: m = 368 TPD

Standard deviation: s = 25.7 TPD

Upper Level: m + s = 394 TPD

Lower Level: m - s = 342 TPD

Probability for UL (65) %

Probability for LL (13) %

Upper Confidence Level @ 97.5% 44.85 TPD

Lower Confidence Level @ 2.5% 31.35 TPD

Subjectively evaluated mean: m_n = 381 TPD

Subjectively evaluated standard deviation: s_n = 67.5 TPD

Subjectively evaluated bias: B_n = -22 TPD

Coefficient of variation CV = 0.18

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% _____ TPD

Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_l = _____ TPD

Subjectively evaluated standard deviation: s_l = _____ TPD

Subjectively evaluated bias: B_l = _____ TPD

Prefer: [ ] Normal [ ] Lognormal

Figure 3-5. Sample Worksheet for Uncertainties in Basinwide TOG Emissions from LDA Exhaust Gases.

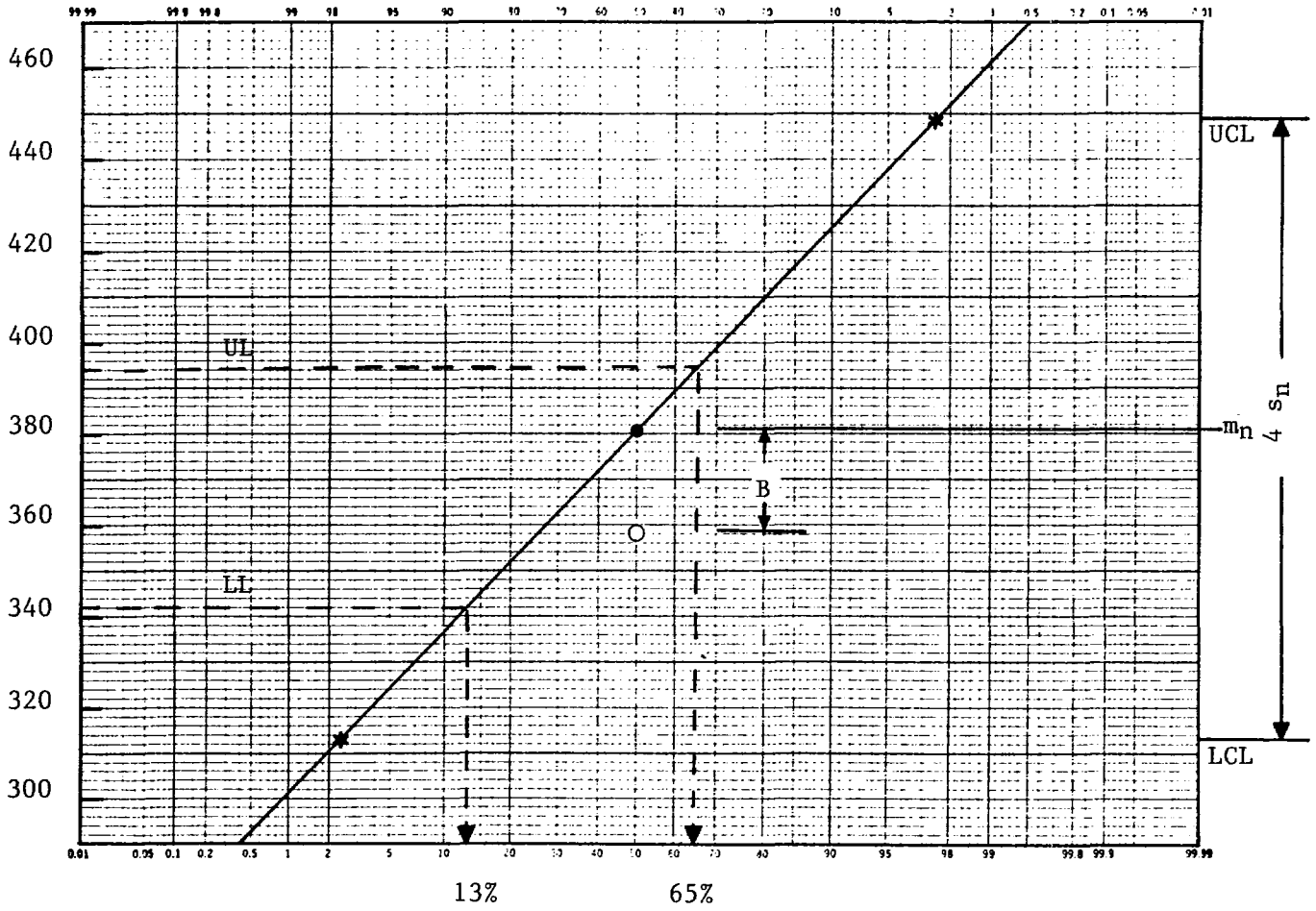


Figure 3-6 Graphical Determination of Probabilities for Upper and Lower Threshold Levels, Based on Integrated Response Method.

MISCF = miscellaneous correction factor

TF_{vmt} = VMT travel weighting fraction

In EMFAC7C, BER is regressed to accumulated miles of each vehicle through a so-called "deterioration factor". This regression equation, however, explains only about 4 percent of the variance of BER. Therefore, it is useless to try to reduce the overall variance of BER by applying the regression equation to individual model years. Instead, this example employs for uncertainty estimation the overall uncertainty of BER, which is estimated to be around ±20 percent of the mean BER.

The author contacted the EPA's Mobile Source Division in Ann Arbor, Michigan to obtain basic data from which the various correction factors are derived. Unfortunately, the EPA does not document the prediction performance of each correction factor. Therefore, in this example, it is arbitrarily assumed that the correction factors carry the following degrees of uncertainty in the estimated factor value:

OMTCF ± 20 percent

SCF ± 30 percent

MISCF ± 15 percent

TV_{vmt} ± 10 percent

For simplicity, we also assume that the most recent ten model years constitute the great majority of motor vehicles in circulation and that the fleet size is about the same for each model year. Under these simplifying assumptions and the commonly employed assumption of mutual independence among the variables, the coefficient of variance for the i-th model year is given by:

$$CV(EF_i) = SQRT[(1+0.20^2)(1+0.20^2)(1+0.30^2)(1+0.15^2)(1+0.10^2) - 1]$$
$$= 0.47$$

If emission factors of individual model years are considered to be statistically independent, then the coefficient of variation for all model years is:

$$CV(EF) = CV(EF_i)/n^{\frac{1}{2}} \qquad (3\text{-}3)$$

where n = 10 years. The result is

$$CV(EF) = 0.15$$

Finally, the uncertainty of the basinwide VMT for light duty autos is assumed to be ±10 percent at the 95 percent confidence level. Then, the overall uncertainty of the basinwide emissions from LDVs is estimated to be:

$$CV(E) = SQRT[(1+0.10^2)(1+0.15^2) - 1]$$
$$= 0.18$$

This "objectively calculated" precision uncertainty of 18 percent is about half of the subjectively assessed precision uncertainty of 38 percent or 135 TPD which was arrived at in Section 3.2.1.

3.3 DRY CLEANERS

Dry cleaners are usually treated collectively as an area source, except for those emitting 25 tons per year or more of any pollutant. Emissions from an area source are generally estimated by taking a product of activity level, A, and emission factor, EF (found in AP-42 or some analogous emission factor document prepared by ARB or a local air pollution control district):

$$E = A \times EF \qquad (3\text{-}2)$$

An example illustrating this method of estimation is worked out for TOG emissions from dry cleaners in Appendix D. However, the author found, in personally interviewing dry cleaning emission experts at the SCAQMD, that this district did not use the emission factor method given by Eq. (3-2) for estimating emissions from dry cleaners. Instead, SCAQMD bases their estimate on permit records. All dry cleaners under the jurisdiction of SCAQMD must apply to and obtain from the district a permit for operation. Therefore, a new example applicable to dry cleaners in the SoCAB was devised and is presented below to illustrate subjective assessment of uncertainties of dry cleaning emissions. However, users of this handbook are encouraged to review the example given in Appendix D as well as the new example, which is disucssed in the next two subsections.

3.3.1 METHOD FOR ESTIMATING DRY CLEANING EMISSIONS

According to the 1983 SoCAB emission inventory, TOG emissions from dry cleaning operations in the basin are estimated to be 18.9 TPD. This estimate is said to be arrived at from the SCAQMD's permit and emission fee records for dry cleaners. The permit system works as follows:

A. If uncontrolled TOG emissions (measured by solvent use) from the dry cleaning operation are estimated to be less than 10 tons per year (TPY), the dry cleaner should report his estimated emissions to the SCAQMD and obtain a permit for operation from the district; and

B. If uncontrolled TOG emissions are estimated to be greater than 10 tons per year, the dry cleaner should obtain a permit and pay an emission fee for the excess emissions over 10 TPY.

Large cleaners of Type B are subjected to a close and frequent scrutiny for their solvent use by SCAQMD enforcement personnel while small cleaners of Type A are subject to once-a-year inspection of their facilities. The district assumes the solvent usage amount indicated in the permit to be the TOG emission rate for a small dry cleaner, whereas for a large dry cleaner the district calculates the emission from the solvent usage and the control efficiency reported in the emission fee application.

According to the SCAQMD's computerized data base (Nazemi 1987), the numbers of dry cleaners and their emissions are as follows:

TABLE 3-3. DRY CLEANING OPERATIONS IN SOCAB
(As of November 1987)

Emission Category	Pre-Control No. of Facilities	Pre-Control Total Emission*	Pre-Control Mean Emission**	Post-Control No. of Facilities	Post-Control Total Emission	Post-Control Mean Emission
10 TPY	1,698	6,351	3.74	1,834	5,364	2.92
10 TPY	224	4,937	22.04	88	1,751	19.90
All Categories	1,922	11,288 (30.9)	5.87	1,922	7,115 (19.5)	3.70

Note: The values in parenthesis are in tons per day (TPD).

* Tons per year (TPY) in TOG emission.

** Tons per year per facility (TPY/Facility) in TOG emission.

Based on the above data, the 1987 basinwide TOG emissions would be reported as 19.5 TPD for all dry cleaners and 14.7 (=5364 / 365) TPD for small dry cleaners. However, since the actual solvent usages of small cleaners could differ from those reported in their permits, TOG emissions from the small dry cleaners could be either lower or higher than the calculated value of 14.7 TPD. Therefore, an analyst (i.e., the author) made the following two assumptions for computing upper and lower plausible estimates:

> For Upper Plausible Estimate: All small dry cleaners, on average, emit half of the threshold level, namely, 5.00 TPY per facility instead of 2.92 TYP; and
>
> For Lower Plausible Estimate: All small dry cleaners, on average, emit half of the emission values reported in their permits, namely 1.46 (=2.92 x 0.5) TPY per facility instead of 2.92 TPY.

Under these two assumptions, the fractional change in the basinwide dry cleaning emissions would be:

Upper Plausible Estimate:

$$\Delta_u = 1834\ (5.00 - 2.92) / 7115$$
$$= 0.536$$

Lower Plausible Estimate

$$\Delta_l = 1834\ (1.46 - 2.92) / 7115$$
$$= -0.376$$

Using the fractional changes computed above, the analyst then figured out the upper and lower plausible estimates that corresponded to the 1983 inventory value of 18.9 TPD:

$$\text{Basic Estimate} = 18.9 \text{ TPD}$$
$$\text{Upper Plausible Estimate} = 18.9 \times (1 + 0.536) = 29.0 \text{ TPD}$$
$$\text{Lower Plausible Estimate} = 18.9 \times (1 - 0.376) = 11.8 \text{ TPD}$$

3.3.2 UNCERTAINTY OF DRY CLEANING EMISSION ESTIMATE

Using the letter three estimates, the analyst (i.e., author) calculated the upper and lower threshold levels as UL = 28.5 TPD and LL = 11.3

TPD. He then asked three dry cleaning emission experts (all three in the SCAQMD's Engineering Division) the following questions:

Q.1 On a scale of 100, what are the odds that the true level of TOG emissions from dry cleaning operations in this air basin is less than 28.5 TPD?

Answer: _____ chances in 100.

Q2. What are the odds that the true level of TOG emissions from dry cleaning operations in this air basin is less than 11.3 TPD?

Answer: _____ chances in 100.

In responding to the above questions, all three experts grumbled that since the district inspected every dry cleaning facility at least once a year, the permit file and emission fee file would be quite complete. To counter their confidence, the analyst pointed out that some dry cleaners might have reported over- or under-estimated values of their solvent uses in the permits, or their operation levels might have changed during the year. After a few exchanges of views regarding the district permit and fee system and resulting emission estimates between the three experts and the analyst, two experts responded to the questions while the third expert declined to state his perceived odds. Their answers are listed in Table 3-4.

TABLE 3-4. SUMMARY RESPONSES FROM THREE EVALUATORS ON THE 1983 BASINWIDE TOG EMISSIONS FROM DRY CLEANING OPERATIONS.

Evaluator/Odds*	1	2	3	Median	Mean
Odds for L	80	60	NA	D	70
Odds for LL	40	30	NA	D	35

Note: NA = No answer
UA = Undeterminable
* Out of 100 chances.

Using the consensus response of UL = 70 and LL = 35, the upper and lower confidence levels were first determined for a normal distribution in Figure 3-7. The results are UCL = 55.1 TPD and LCL = -15.8 TPD. The

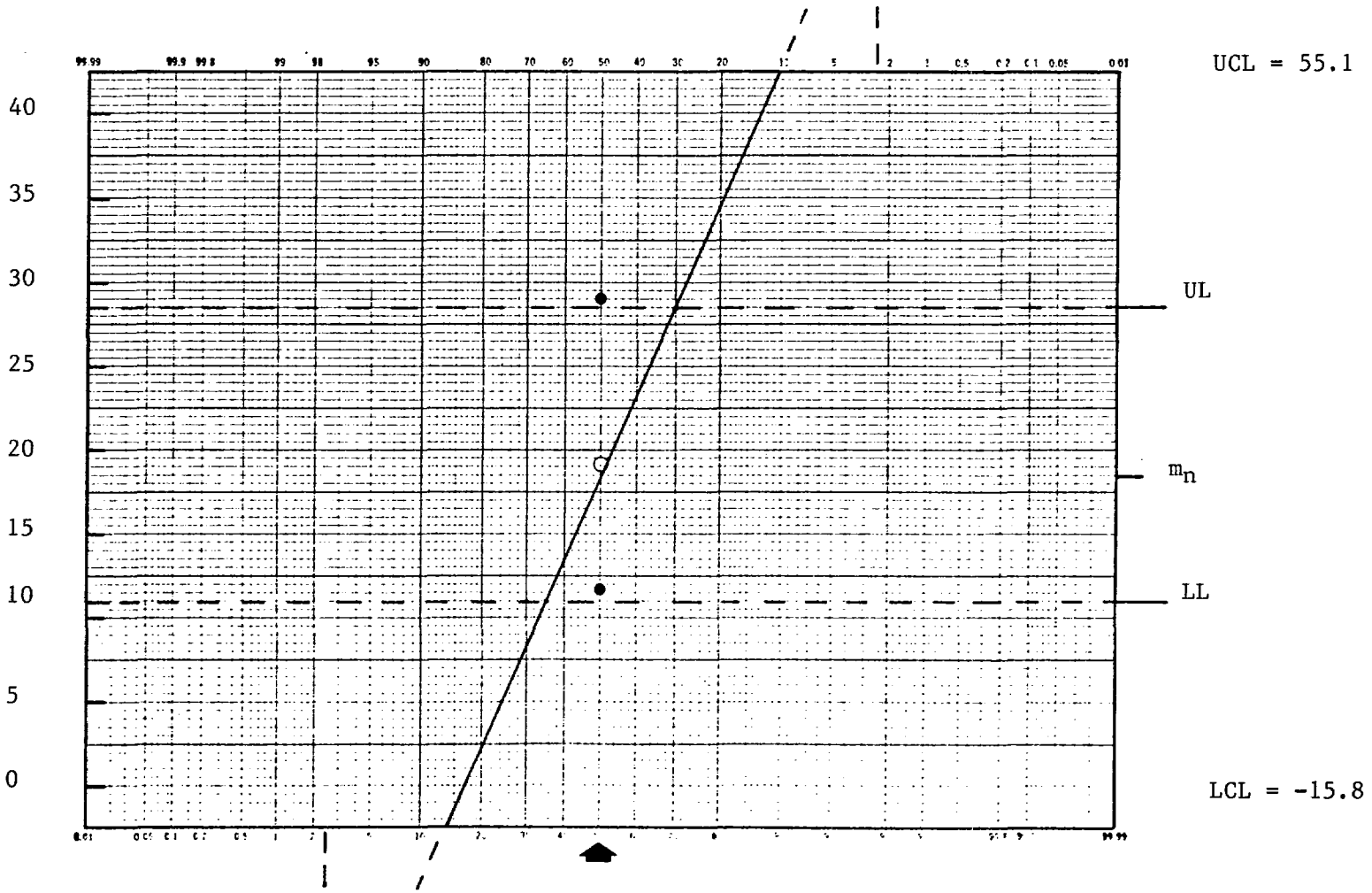


Figure 3-7 Sample Plot of Dry Cleaning Expert Response on Normal Probability Paper.

subjectively evaluated mean and standard deviation were graphically determined:

$$m_n = 18.5 \text{ TPD}$$
$$s_n = 17.7 \text{ TPD}$$

Since these values yielded CV = s_n/m_n = 0.96, exceeding the critical level of 0.3, a new graphical determination was done using lognormal probability paper, as shown in Figure 3-8. The subjectively evaluated upper and lower confidence levels are read from the figure as:

$$UCL = 116 \text{ TPD}$$
$$LCL = 2.3 \text{ TPD}$$

The subjectively evaluated <u>geometric</u> mean is also read from the 50th percentile point in the graph as:

$$m_1 = 16.6 \text{ TPD}$$

With lognormal probability paper, the subjectively evaluated standard deviation is different from (UCL-LCL)/4. Instead, it is given by the following equation:

$$s_1 = \text{SQRT } (UCL/m_1) \qquad (3\text{-}4)$$

Substitution of the subjectively evaluated values of UCL and m_1 into equation (3-4) yields:

$$s_1 = (116/16.6)^{\frac{1}{2}}$$
$$= 2.6 \quad \text{TPD}$$

All subjectively evaluated values using both normal and lognormal probability papers are summarized in the worksheet of Figure 3-9. Finally, the results of normal and lognormal distributions were presented to the two cooperative experts. After reviewing the two sets of results, they indicated a preference for the results of lognormal distribution over

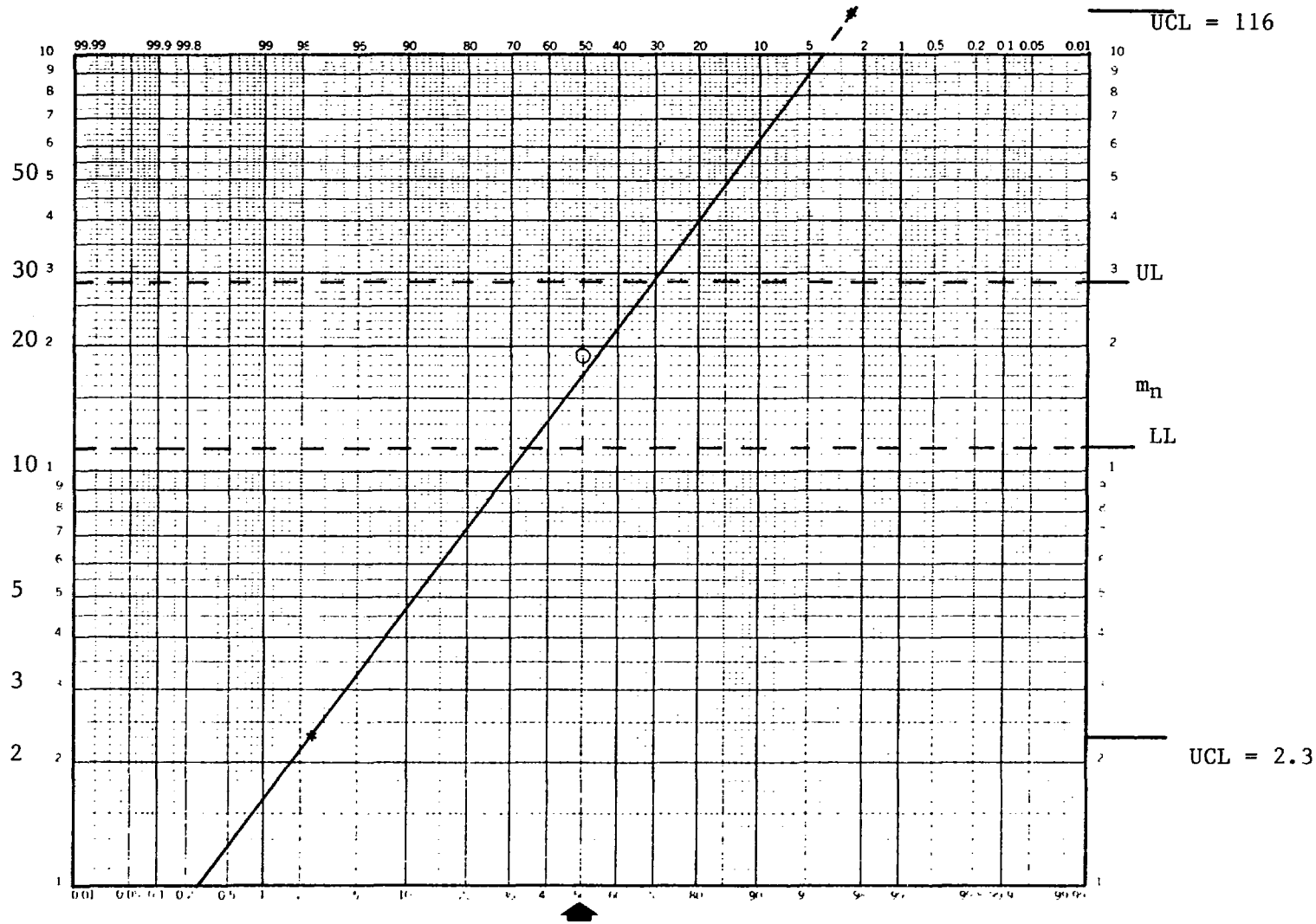


Figure 3-8 Sample Plot of Dry Cleaning Expert Response on Lognormal Probability Paper.

UNCERTAINTY WORKSHEET

NAME Consensus Response DATE February 5, 1988

Three estimates for emissions: 29.0, 18.9, 11.8 TPD

Mean of three estimates: m = 19.9 TPD

Standard deviation: s = 8.6 TPD

Upper Level: $m + s$ = 28.5 TPD

Lower Level: $m - s$ = 11.3 TPD

Probability for UL (70) %

Probability for LL (35) %

Upper Confidence Level @ 97.5% 55.1 TPD

Lower Confidence Level @ 2.5% -15.8 TPD

Subjectively evaluated mean: m_n = 18.5 TPD

Subjectively evaluated standard deviation: s_n = 17.7 TPD

Subjectively evaluated bias: B_n = -0.4 TPD

Coefficient of variation CV = 0.96

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% 116.0 TPD

Lower Confidence Level @ 2.5% 2.3 TPD

Subjectively evaluated mean: m_l = 16.6 TPD

Subjectively evaluated standard deviation: s_l = 2.6 TPD

Subjectively evaluated bias: B_l = -2.3 TPD

Prefer: [ ] Normal [X] Lognormal

Figure 3-9. Sample Worksheet for Dry Cleaning Example

those of normal distribution. Therefore, the subjectively evaluated mean, bias, and confidence interval are:

m = 16.6 TPD
B = -2.3 TPD
UCL = 116.0 TPD
LCL = 2.3 TPD

As to objective estimation of uncertainties in the dry cleaning emissions, the author could find no empirical data which would show how reliable emission estimates based on the permit/emission fee system are. Therefore, no attempt was made to estimate objectively calculated uncertainties of the dry cleaning emissions.

4.0 METHOD FOR ESTABLISHING INVENTORY UNCERTAINTY

In two preceding sections, methods of evaluating uncertainties of individual emission estimates are discussed. Given these emission uncertainties of individual source categories, this section provides procedures for: (1) determining which categories are to be flagged for future studies; (2) aggregating individual source category uncertainties to obtain the uncertainties of broader source categories; and (3) compiling and reporting those uncertainties for the entire emissions inventory. Figure 4-1 depicts diagrammatically a scheme for compiling and reporting uncertainties for the entire emissions inventory (designated as Phase II) as well as the steps used for generating initial source category uncertainties (designated as Phase I).

4.1 OVERALL STRATEGY

In general, the development of an emissions inventory is a dynamic process, reflecting technical improvements in inventory methods and procedures and the changing patterns of emissions within a region (EPA 1981). Of particular importance is the on-going work to eliminate errors through engineering analysis and to increase the confidence in the inventory through full documentation, quality assurance, and ultimately its demonstrated reliability for evaluating control strategies and air quality. The methodology described herein gives an additional dimension to these efforts by providing the means to quantify the reliability of the inventory in terms of precision and bias uncertainties of emission estimates reported at various levels of the inventory.

As indicated in Figure 4-1, the logic flow for this effort requires a two-tiered approach. The objective of the Phase I work effort is to establish the initial uncertainty assessment for each inventory source category (e.g., power plants, on-highway mobile sources, and so forth). To accomplish this, the objective statistical techniques and subjective evaluation methods developed in Sections 2.0 and 3.0 are to be applied to the available information in each category.

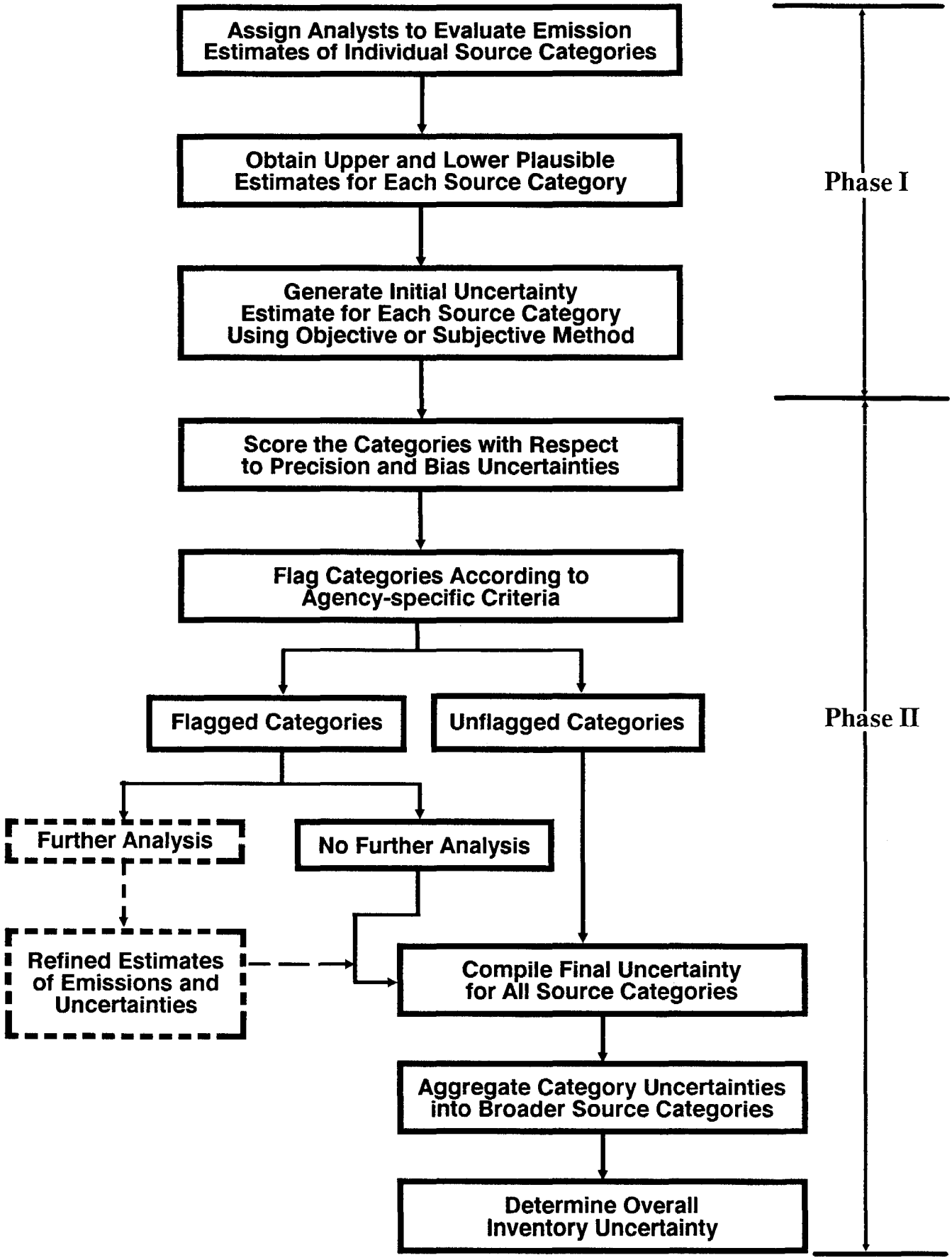


Figure 4-1. Scheme for Quantifying Inventory Uncertainties

Under Phase II, the first thing to be performed is the ranking of the categories by a special scoring system that assigns relative values to the precision and bias uncertainties of each source category. The purpose is to flag those categories that warrant consideration for further work. As indicated in Figure 4-1, these flagged categories may or may not be subjected to further analysis depending upon the timing and the priority designated by the air pollution control agency. The main purpose of flagging categories is the clear recognition that these categories require special attention for enhancing the confidence in the existing emissions inventory.

Independent of the flagging, uncertainties of all source categories will be aggregated to obtain the corresponding uncertainties of broader source categories, using appropriate statistical formulas presented in Sections 2.0 and 3.0 or the computer program listed in Appendix A. How to compute the uncertainty of the aggregated category is discussed in Section 4.3 while in Section 4.2 the flagging method is described. The last subsection, Section 4.4, describes the ways of compiling and reporting uncertainties for the entire emissions inventory.

4.2 FLAGGING PROBLEM CATEGORIES

Once the uncertainties in the emissions estimates for all source categories have been determined, a coordinator of a committee of experts assigned to oversee the uncertainty of assessment effort should tabulate precision and bias uncertainties together with estimated emissions for each of the categories in the inventory. Based on the tabulated results, the coordinator can score the uncertainty of each source category i in any of several different ways:

1. Score category uncertainty according to the precision uncertainty measured by coefficient of variation, CV_i;
2. Score category uncertainty according to the precision uncertainty, CV_i, and the bias uncertainty measured by absolute value of relative bias, b_i where $b_i = B_i/E_i$;
3. Score category uncertainty according to the sum of precision and bias uncertainties $(CV_i + b_i)$; and
4. Score category uncertainty according to the sum of standard deviation and bias, $(s_i + B_i)$;

Each scoring scheme listed above has varying degrees of utility depending on the intended use of the scores. If, say, the purpose were to find problem categories in some sense, then the multiple scoring scheme would be most suitable.

Suppose that the coordinator has picked Scheme 4 for finding problem categories for further investigation. Then, he will set threshold values for standard deviation and bias, say s* and B*, such that K out of N (K N) source categories in the inventory will be flagged for a closer look. In this flagging scheme, a category may be flagged because of s_i s*, because of B_i B*, or both. The choice of K should be made by the coordinator according to his evaluation of the initial uncertainty estimation result and the scope of the agency's uncertainty assessment efforts. It should be noted that he can make a similar flagging effort at any aggregation level of the inventory system.

For each flagged category, the coordinator should examine whether or not the category is really worth further investigation. Some flagged categories may have been previously studied so intensively that there seems to be little room for further improvement in tightening the uncertainty. For such a category, further investigation may not be rewarding. After reviewing all the flagged categories, the coordinator should select a set of the flagged categories for future studies.

4.3 AGGREGATING INITIAL UNCERTAINTIES INTO BROADER SOURCE CATEGORIES

Compilation of initially assessed uncertainties and their aggregation into broader source categories can be made in several different ways, depending on the ultimate use of the uncertainty estimates. Suppose that the purpose of uncertainty estimation is limited to indicating the uncertainty of each of the 5 to 10 major source categories (e.g., on-highway mobile sources and fugitive-dust area sources). The uncertainty estimates of all sub-categories may then be grouped into those of the major categories by following the inventory category hierarchy. For each major group, uncertainties of all the subcategories are aggregated into the total by using an appropriate statistical formula given in Sections 2.0 and 3.0.

In this aggregation process, the analyst must give careful consideration to the following issues:

1. Precision uncertainties and bias uncertainties must be aggregated separately;
2. Before aggregating subcategory precision uncertainties, a decision must be made as to whether the assumption of no correlation (i.e., independence) or full correlation (i.e., dependence) best describes the statistical associations among subcategory emissions;
3. Precision uncertainties must be aggregated by using an appropriate formula for either the independence or dependence assumption employed above; and
4. A partial correlation between emissions in different subcategories should only be considered when there is strong evidence that it more truly reflects the statistical association between the emission variations than either of the two simple assumptions, i.e., independence or dependence.

Example: NO_x Emissions from Fuel Combustion

According to the 1983 SoCAB emission inventory, NO_x emissions from all fuel combustion sources are estimated as summarized in Table 4-1.

TABLE 4-1. 1983 NOx EMISSIONS FROM FUEL COMBUSTION SOURCES IN SoCAB

Source Category	NOx Emissions (TPD)	Estimated Uncertainties Precision,s	Bias, B
Agricultural	0.0	0.5	-0.2
Oil and Gas Production	25.4	3.7	2.2
Petroleum Refining	51.0	4.3	-1.8
Other Manufacturing/Industrial	75.9	13.2	-5.6
Electric Utilities	39.4	1.4	0.1
Other Services and Commerce	33.2	5.3	3.2
Residential	33.5	7.1	-4.5
Other	3.9	0.8	0.5
Total Fuel Combustion	262.3	(17.0)	-6.1

In this table, uncertainty values are all assumed for illustration purposes only, except for electric utilities whose uncertainty values were arrived at by five power plant emission experts at the SCAQMD, applying the subjective assessment method. For simplicity, further assume that emissions from all source categories are normally distributed and mutually independent. Under these assumptions, Eq. (2-23) reduces to:

$$VAR\ (X) = SUM\ (\ _i^2\ ) \tag{4-1}$$

or as an approximate equation

$$s = SQRT\ [SUM\ (s_i^2\ )] \tag{4-2}$$

Substituting s_i values given in Table 4-1 into Eq. (4-2), one finds

$$s = 17.0 \quad TPD$$

(The same result will be obtained using Eq (3-1) instead of EQ. (4-2).)

It should be noted that, in this example, the source category "Other Manufacturing/Industrial" accounts for a particularly large percentage of the overall precision uncertainty.

$$\text{Percentage accounted for by this category} = (13.2/17.0)^2 \times 100$$

$$= 60.3 \text{ percent}$$

For bias uncertainties, the simplest measure for the aggregated category may be given by the sum of individual category biases as indicated in Table 4-1. Another useful parameter for indicating the bias uncertainty of an aggregated category is given by the following equation:

$$b = SUM(\ B_i\ )/SUM(u_i) \tag{4-3}$$

where B_i is the absolute value of the i-th category bias B_i and u_i is the estimated emission of that category.

Substituting the bias values given in Table 4-1 into the above equation, we find

$$b = 0.069 \text{ or } 6.9 \text{ percent}$$

It should be noted that in this example, this relative bias uncertainty, b , is as big as the relative precision uncertainty, CV, which is computed as:

$$CV = s/SUM(u_i)$$
$$= 17.0/262.3$$
$$= 0.065 \text{ or } 6.5 \text{ percent}$$

4.4 COMPILING AND DOCUMENTING INVENTORY UNCERTAINTIES

Potential uses of the assessed uncertainties of inventoried emission estimates range from a quantitative indication of the reliability of the inventory system to a diagnostic tool for identifying problem source categories and to a systematic means of upgrading the existing inventory. In view of potential uses of the uncertainty analysis, it is especially important to employ the best presentation scheme for assessed uncertainties of the inventoried emissions.

Table 4-2 presents a sample format that should be useful for summarizing results of an uncertainty study. In this example, eight major categories and three totals (stationary, mobile, and inventory) are used to characterize uncertainties of the emissions inventory. Each major source category is described by the total emissions estimate, the percentage contribution of that category to inventory total, the precision uncertainty, and the bias uncertainty.

TABLE 4-2. SAMPLE FORMAT FOR SUMMARIZING UNCERTAINTY STUDY RESULTS

Major Source Category	Category Emissions: Estimated Emission (TPD)	Category Emissions: Percent Contribution To Inventory Total	Precision Uncertainty: CV	Precision Uncertainty: s	Precision Uncertainty: % Contr. to Inv. Total	Bias Uncertainty: b	Bias Uncertainty: B	Bias Uncertainty: % Contr. to Inv. Total	Remarks
Stationary Sources									
1. Fuel Combustion									
2. Waste Burning									
3. Solvent Use									
4. Petroleum S & T									
5. Industrial Processes									
6. Misc. Processes									
Mobil Sources									
7. On Road Vehicles									
8. Other Mobile									
All Sources									
9. Stationary Total									
10. Mobile Total									
11. Inventory Total									

The precision uncertainty is described by the coefficient of variation (or relative precision uncertainty), CV, the standard deviation, s, and the percentage contribution of the category's precision uncertainty to the overall inventory uncertainty. The bias uncertainty is described by the relative bias uncertainty, b, the total bias, B, and the percentage contribution of the category's bias uncertainty to the overall inventory uncertainty. In the "remarks" column, any noteworthy remarks such as whether the category is flagged or not will be recorded.

Quantities listed in Table 4-2 permit uncertainty of each category or an entire inventory to be concisely characterized. For example, uncertainty of the overall inventory emission (m) may be expressed as:

95% Confidence Interval = ± 2CV x 100 percent of m
Total Bias = B
Mean Relative Bias = b x 100 percent of m

5.0 REFERENCES

Chun, K. C., "Uncertainty Data Base for Emissions Estimation Parameters: Interim Report," Argonne National Laboratory, Prepared for the U.S. Department of Energy, Office of Fossile Energy, September 1986.

EPA, "Estimation of Uncertainty for the 1980 NAPAP Emissions Inventory", EPA-600/7-86-055, U.S. Environmental Protection Agency, Air and Energy Research Laboratory, Research Triangle Park, NC, December 1986.

EPA, "Procedures for Emission Inventory Preparation, Volume I: Emission Inventory Fundamentals", EPA-450/4-81-026a, U.S. Environmental Protection Agency, Research Triangle Park, NC, September 1981.

EPRI, "Emissions Inventory for the SURE Region", EA-1913 Research Project 862-5, Prepared by GCA Corp. for the Electric Power Research Institute, Palo Alto, CA, April 1981.

Mangat, T. S., L. H. Robinson and P. Switzer, "Medians and Percentiles for Emission Inventory Totals", Paper No. 84-72.3, Presented at the 77th APCA Annual Meeting, San Francisco, CA, June 1984.

Nazemi, M., Personal Communication between Mike Nazemi of the South Coast Air Quality Management District and Yuji Horie of Valley Research Corporation, November 1987.

SCAQMD, "Uncertainty of 1979 Emissions Data, Chapter IV: 1983 Air Quality Management Plan Revision, Appendix IV-A; 1979 Emissions Inventory for the South Coast Air Basin, Revised October 1982", Planning Division of the South Coast Air Quality Management District, El Monte, CA, October 1982.

Additional References Cited in Appendices B, C and D

ARB, "Methodology to Calculate Emission Factors for On-Road Motor Vehicles", California Air Resources Board, Technical Support Division, Sacramento, CA, November 1986.

CALTRANS, "California 1982 Daily Vehicle Miles of Travel for Arterial and Collector Highways Based on HPMS Data", California Department of Transportation, Division of Highways, Sacramento, CA, October 1983.

EPA, "Compilation of Air Pollutant Emission Factors, Volume I: Stationary Point and Area Sources", AP-42 Fourth Edition, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, September 1985.

Horie, Y., and R. Rapoport, "Assessment of Heavy-Duty Gasoline and Diesel Vehicles in California: Population and Use Patterns", PES No. 654, Prepared by Pacific Environmental Services, Inc., for California Air Resources Board, Research Division, Sacramento, CA, July 1985.

Rogozen, M. B., R. D. Rapoport and A. Shochet, "Development and Improvement of Organic Compound Emission Inventories for California", SAIC-85/1531, Prepared by Science Applications International Corporation, for California Air Resources Board, Research Division, Sacramento, CA, January 1985.

SCAQMD, "Draft 1983 Emissions Inventory South Coast Air Basin, Working Paper No. 1: 1987 AQMP Revision", South Coast Air Quality Management District, El Monte, CA, May 1986.

APPENDIX A

NUMERICAL SIMULATION MODEL FOR STATISTICAL DISTRIBUTION OF AN AGGREGATED VARIABLE

```
C
      REAL SERIES(1000,10),FINAL(1000),STATS(6,10),PTILES(19,10)
      REAL MEAN,STDEV,MEANS(6)
      CHARACTER ST1*10,MODES(2)*20,DISTS(2)*20,PSTVS(2)*20
      DATA MODES/ 'ADDITIVE' , 'MULTIPLICATIVE' /
      DATA DISTS/ 'NORMAL'   , 'LOGNORMAL'      /
      DATA PSTVS/ 'REGULAR'  , 'POSITIVE ONLY'  /
C
 11   FORMAT(I4,' - ',A)
  1   FORMAT(' ',A,\)
  2   FORMAT(F6.0)
  3   FORMAT(I6)
C
      WRITE(*,1) 'ENTER MEAN: '
      READ (*,2) MEAN
C
      WRITE(*,1) 'ENTER S.D.: '
      READ (*,2) STDEV
C
      WRITE(*,'(/)')
      WRITE(*,11) (I,DISTS(I+1),I=0,1)
      WRITE(*,1) 'ENTER DIST: '
      READ (*,3) IDIST
C
      WRITE(*,'(/)')
      WRITE(*,11) (I,PSTVS(I+1),I=0,1)
      WRITE(*,1) 'ENTER PSTV: '
      READ (*,3) IPSTV
C
      WRITE(*,'(/)')
      WRITE(*,11) (I,MODES(I+1),I=0,1)
      WRITE(*,1) 'ENTER COMB: '
      READ (*,3) ICOMB
C
      WRITE(*,1) 'ENTER NVAR: '
      READ (*,3) NVAR
C
      OPEN(1,FILE='TESTRN.OUT')
      DO 30 ITIME = 1,10
C
      CALL GETTIM(IHR,IMN,ISC,I100)
      WRITE(*,4) IHR,IMN,ISC,'GENERATING NUMBERS -',ITIME
  4   FORMAT(2X,I2,':',I2,':',I2,' - ',A,I3)
C
      DO 10 I = 1,NVAR
 10   CALL RANDOM(MEAN,STDEV,IPSTV,IDIST,1000,SERIES(1,I))
C
      CALL GETTIM(IHR,IMN,ISC,I100)
      WRITE(*,4) IHR,IMN,ISC,'COMBINING NUMBERS'
C
      DO 20 I = 1,1000
C
           IF ( ICOMB .EQ. 0 ) THEN
                FINAL(I) = 0.0
           ELSE
                FINAL(I) = 1.0
           END IF
C
      DO 20 J = 1,NVAR
C
```

```
            IF ( ICOMB .EQ. 0 ) THEN
                 FINAL(I) = FINAL(I) + SERIES(I,J)
            ELSE
                 FINAL(I) = FINAL(I) * SERIES(I,J)
            END IF
C
 20    CONTINUE
C
       CALL GETTIM(IHR,IMN,ISC,I100)
       WRITE(*,4) IHR,IMN,ISC,'CALCULATING PERCENTILES -',ITIME
C
       CALL PCTILE(IDIST,2,1000,FINAL,
      +            PTILES(1,ITIME),STATS(1,ITIME))
C
       CALL GETTIM(IHR,IMN,ISC,I100)
       WRITE(*,4) IHR,IMN,ISC,'PERCENTILES COMPLETE -',ITIME
C
   30  CONTINUE
C
C      WRITE(1,'(1H
```

```
                )')
      WRITE(1,5) MEAN,STDEV,IDIST,DISTS(IDIST+1),
     +           IPSTV,PSTVS(IPSTV+1),ICOMB,MODES(ICOMB+1),NVAR
   5  FORMAT(/' ENTERED MEAN   =',F10.3,
     +       /' ENTERED S.D.   =',F10.3,
     +       /' ENTERED DIST   =',I10  ,' ',A,
     +       /' ENTERED PSTV   =',I10  ,' ',A,
     +       /' ENTERED COMB   =',I10  ,' ',A,
     +       /' ENTERED NVAR   =',I10  )
C
      DO 60 I = 1,6
      MEANS(I) = 0.0
      DO 50 J = 1,10
   50 MEANS(I) = MEANS(I) + STATS(I,J)
   60 MEANS(I) = MEANS(I)/10.0
C
      WRITE(1,'(/)')
      WRITE(1,61) (STATS(1,I),I=1,10),MEANS(1)
      WRITE(1,62) (STATS(2,I),I=1,10),MEANS(2)
      WRITE(1,63) (STATS(3,I),I=1,10),MEANS(3)
      WRITE(1,64) (STATS(4,I),I=1,10),MEANS(4)
      WRITE(1,65) (STATS(5,I),I=1,10),MEANS(5)
      WRITE(1,66) (STATS(6,I),I=1,10),MEANS(6)
   61 FORMAT( ' A MEAN =',11F9.3)
   62 FORMAT( ' A S.D. =',11F9.3)
   63 FORMAT( ' G MEAN =',11F9.3)
   64 FORMAT( ' G S.D. =',11F9.3)
   65 FORMAT( ' MIN    =',11F9.3)
   66 FORMAT( ' MAX    =',11F9.3)
      WRITE(1,'(/)')
C
      WRITE(1,7)
    7 FORMAT(/'  PERCENTILES')
C
      DO 70 I = 1,19
      MEANS(1) = 0.0
      DO 71 J = 1,10
   71 MEANS(1) = MEANS(1) + PTILES(I,J)
      MEANS(1) = MEANS(1)/10.0
   70 WRITE(1,75) I,(PTILES(I,J),J=1,10),MEANS(1)
   75 FORMAT(I5,4X,11F9.3)
C
      END
```

```
C
      SUBROUTINE RANDOM(MEAN,STDEV,IFLAG,DFLAG,N,SERIES)
C
C     DRAW A SERIES OF RANDOM NUMBERS, RN, FROM N(0,1) AND TRANSFORM
C     THEM TO RANDOM NUMBERS, R, FOR A SPECIFIED DISTRIBUTION
C
C     AUTHOR:  MICHAEL A. THIELE
C              10016 ORION AVENUE
C              MISSION HILLS, CA 91345
C              (818) 892-2201
C
C     DATE:    FEBRUARY, 1987
C
C     PARAMETERS:  -> = PASSED; <- = RETURNED; I = INTEGER; R = REAL
C
C     MEAN   ->  R  MEAN OF DESIRED NORMAL DISTRIBUTION
C     STDEV  ->  R  STANDARD DEVIATION OF DESIRED DISTRIBUTION
C     IFLAG  ->  I  REGULAR OR POSITIVE ONLY FLAG
C                    = 0  - REGULAR DISTRIBUTION
C                    = 1  - POSITIVE ONLY DISTRIBUTION (DEFAULT)
C     DFLAG  ->  I  DISTRIBUTION FLAG
C                    = 0  - NORMAL DISTRIBUTION
C                    = 1  - LOGNORMAL DISTRIBUTION
C     N      ->  I  THE NUMBER OF RANDOM NUMBERS DESIRED
C     SERIES <-  R  VECTOR FOR THE SERIES OF RANDOM NUMBERS
C
C     CALLS:
C
C     RANDU   =  UNIFORN RANDOM NUMBER GENERATOR FUNCTION, U(0-1)
C     ZFUNC   =  INVERSE Z FUNCTION, N(0,1)
C
      INTEGER IFLAG, DFLAG, N
      REAL    MEAN, STDEV, SERIES(1)
      REAL*8  X
C
      REAL*8 RANDU,ZFUNC
C
      IF ( N .LE. 0 ) RETURN
C
      DO 10 I = 1,N
C
C     GENERATE THE NORMAL RANDOM NUMBER, N(0,1)
C
    1 X = RANDU( 0 )
      X = ZFUNC( X )
C
C     TRANSFORM IT TO DESIRED DISTRIBUTION
C
      X = MEAN + STDEV * X
C
C     CHECK FOR POSITIVE ONLY FLAG AND IF IT FAILS GO BACK
C
      IF ( (IFLAG .EQ. 1) .AND. (X .LT. 0.0) ) GOTO 1
C
      SERIES(I) = X
      IF ( DFLAG .EQ. 1 )  SERIES(I) = EXP( SERIES(I) )
C
   10 CONTINUE
C
      RETURN
```

END

```
C
      REAL*8 FUNCTION RANDU( I )
C
C     REAL*8 UNIFORM PSEUDO-RANDOM NUMBER GENERATOR U(0-1)
C
C     AUTHOR: MICHAEL A. THIELE
C             10016 ORION AVENUE
C             MISSION HILLS, CA 91345
C             (818) 892-2201
C
C     DATE:   FEBRUARY, 1987
C
C     PARAMETERS:  -> = PASSED; <- = RETURNED; I = INTEGER; R = REAL
C
C     I      ->  I    DUMMY ARGUMENT, NOT USED
C     RANDU  <-  R*8  DOUBLE PRECISCION RANDOM NUMBER IN
C                     THE RANGE  0.0 <=  RANDU  < 1.0
C
C     THE FORMULA USED IS THE LINEAR CONGRUENTIAL METHOD:
C
C       X(N+1) = ( A*X(N) + C ) MODULO M
C
C       WHERE:
C          A       = 2**26 + 5
C          C       = 2**26 + 5
C          M       = 2**31 - 1
C          X(N)    = PREVIOUS MEMBER OF SEQUENCE
C          X(N+1)  = NEXT MEMBER OF SEQUENCE
C
C       THE SEED, X(0), COMES FROM THE SYSTEM TIME AND IS
C       THE SECONDS SINCE MIDNIGHT
C
C     REFERENCE:
C
C     "THE ART OF COMPUTER PROGRAMMING, VOLUME 2: SEMINUMERICAL
C     ALGORITHMS", BY DONALD E. KNUTH, 2ND EDITION, 1981,
C     ADDISON-WESLEY, PP 9-22
C
C     CALLS:
C
C     MSTIME - DOS SPECIFIC ROUTINE THAT RETURNS THE SYSTEM TIME
C
      REAL*8 M,A,C,X
      INTEGER IFIRST
      CHARACTER*10 ATIME
      DATA IFIRST/ 0 /
C
      IF( IFIRST .EQ. 0 ) THEN
C
C       THE MAGIC NUMBERS
C
        A = 2.0**26.0 + 5.0
        C = 2.0**26.0 + 5.0
        M = 2.0**31.0 - 1.0
C
C       THE SEED VALUE - TIME IN SECONDS SINCE MIDNIGHT
C
        CALL GETTIM(IHR,IMN,ISC,I100)
        X = FLOAT(IHR)*360000.  +  FLOAT(IMN)*6000. +
     +      FLOAT(ISC)*100.     +  FLOAT(I100)
```

```
C
        IFIRST = 1
      END IF
C
      X = A*X + C
      X = DMOD( X , M )
      RANDU =  X / M
      RETURN
      END
```

```
C
      REAL*8 FUNCTION ZFUNC( P )
C
C     INVERSE Z FUNCTION -  CALCULATE THE Z VALUE FOR THE
C     PASSED PERCENTILE VALUE, ABS( ERROR ) < 0.00045
C
C     AUTHOR: MICHAEL A. THIELE
C             10016 ORION AVENUE
C             MISSION HILLS, CA 91345
C             (818) 892-2201
C
C     DATE:   FEBRUARY, 1987
C
C     REFERENCE:
C
C     "HANDBOOK OF MATHEMATICAL FUNCTIONS WITH FORMULAS,
C     GRAPHS, AND METHEMATICAL TABLES", M. ABRAMOWITZ AND
C     I.A. STEGUM, EDITORS, 1972, NATIONAL BUREAU OF
C     STANDARDS, WASHINGTON, D.C., PP 933
C
      IMPLICIT REAL*8 (A-H,O-Z)
C
      REAL*8 P,T,T2,ZERO,ONE,HALF,INFNTY
      REAL*8 C,D,C0,C1,C2,D0,D1,D2,D3
C
      DATA ZERO   / 0.0      /
      DATA ONE    / 1.0      /
      DATA HALF   / 0.5      /
      DATA INFNTY / 1.0E38   /
C
C     INVERSE Z FORMULA CONSTANTS
C
      DATA C0,C1,C2    / 2.515517 , 0.802853 , 0.010328            /
      DATA D0,D1,D2,D3 / 1.000000 , 1.432788 , 0.189269 , 0.001303 /
C
      IF ( P .GE. ONE ) THEN
         ZFUNC = INFNTY
         RETURN
         ENDIF
      IF ( P .LE. ZERO ) THEN
         ZFUNC = -INFNTY
         RETURN
         ENDIF
C
      D = P
      IF ( D .GT. HALF ) D = ONE - D
C
      T2 = D*D
      T2 = ONE / T2
      T2 = DLOG( T2 )
      T  = DSQRT( T2 )
C
      C = C0 + C1*T + C2*T2
      D = D0 + D1*T + D2*T2 + D3*T*T2
      ZFUNC = T -  C / D
C
      IF ( P .LE. HALF ) ZFUNC = -ZFUNC
C
      RETURN
      END
```

```
C
      SUBROUTINE PCTILE(DFLAG,PFLAG,N,SERIES,PTILES,STATS)
C
C     RANK-ORDER THE SERIES OF RANDOM NUMBERS IN ASCENDING
C     ORDER AND COMPUTE SPECIFIED PERCENTILE VALUES AND
C     DISTRIBUTION PARAMETERS
C
C     AUTHOR:  MICHAEL A. THIELE
C              10016 ORION AVENUE
C              MISSION HILLS, CA 91345
C              (818) 892-2201
C
C     DATE:    FEBRUARY, 1987
C
C     PARAMETERS:  -> = PASSED; <- = RETURNED; I = INTEGER; R = REAL
C
C     DFLAG    ->  I  DISTRIBUTION TYPE FLAG
C                      = 1  -  NORMAL DISTRIBUTION
C                      = 2  -  LOGNORMAL DISTRIBUTION
C                      = 3  -  MIXED DISTRIBUTION
C     PFLAG    ->  I  FLAG FOR DESIRED PERCENTILES
C                      = 0  -  ( 5) 2.5, 16 , 50 , 84, 97.5
C                      = 1  -  (11) 1, 2.5, 5, 16, 30 , 50,
C                                   70, 84, 95, 97.5, 99
C                      = 2  -  (19) 0.05, 0.1, 0.2, 0.5, 1, 2.5, 5,
C                                   16, 30, 50, 70, 84, 95, 97.5,
C                                   99, 99.5, 99.8, 99.9, 99.95
C              NOTE:   PFLAG = 2 SHOULD BE USED TO CALCULATE VALUES
C                      TO BE USED BY THE CFDIST ROUTINE
C     N        ->  I  NUMBER OF RANDOM NUMBERS IN THE SERIES
C     SERIES   ->  R  THE SERIES OF RANDOM NUMBERS
C     PTILES   <-  R  THE VECTOR OF CALCULATED PERCENTILES VALUES
C              NOTE:   THE VECTOR MUST BE DIMENSIONED LARGE ENOUGH IN
C                      THE CALLING PROGRAM TO ACCOMODATE THE NUMBER OF
C                      PERCENTILES SPECIFIED BY PFLAG ( 5, 11, OR 19)
C     STATS    <-  R  VECTOR OF SIZE 6 OF CALCULATED STATISTICS
C                      LOC 1  -  ARITHMETIC MEAN
C                          2  -  ARITHMETIC STANDARD DEVIATION
C                          3  -  GEOMETRIC MEAN
C                          4  -  GEOMETRIC STANDARD DEVIATION
C                          5  -  MINIMUM VALUE
C                          6  -  MAXIMUM VALUE
C
C     CALLS:
C
C     HPSORT   -  HEAP SORT ROUTINE
C
      INTEGER  DFLAG, PFLAG, N
      REAL     SERIES(1), PTILES(1), STATS(1)
      INTEGER  NPS(3)
      REAL     PCTLS(19,3)
C
      DATA NPS  /  5, 11, 19 /
      DATA PCTLS/
     +   0.0250 , 0.1600 , 0.5000 , 0.8400 , 0.9750 , 0.0000 ,
     +   0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 ,
     +   0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 ,
     +   0.0000 ,
     +   0.0100 , 0.0250 , 0.0500 , 0.1600 , 0.3000 , 0.5000 ,
     +   0.7000 , 0.8400 , 0.9500 , 0.9750 , 0.9900 , 0.0000 ,
```

```
     +   0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 , 0.0000 ,
     +   0.0000 ,
     +   0.0005 , 0.0010 , 0.0020 , 0.0050 , 0.0100 , 0.0250 ,
     +   0.0500 , 0.1600 , 0.3000 , 0.5000 , 0.7000 , 0.8400 ,
     +   0.9500 , 0.9750 , 0.9900 , 0.9950 , 0.9980 , 0.9990 ,
     +   0.9995 /
C
      IF ( N .LE. 0 ) RETURN
C
C     SORT THE SERIES OF VALUES
C
      CALL HPSORT( SERIES , N )
C
C     DETERMINE THE PERCENTILES TO CALCULATE
C
      IPTR = PFLAG + 1
      IF ( (IPTR .LT. 1) .OR. (IPTR .GT. 3) ) IPTR = 3
      NT = NPS(IPTR)
C
C     CALCULATE THE PERCENTILES
C
      DO 10 I = 1,NT
      R1  = FLOAT(N)*PCTLS(I,IPTR) + 0.5
      IR1 = R1
      IR2 = IR1+1
      IF ( IR1 .LT. 1 ) THEN
          PTILES(I) = SERIES(1)
      ELSE IF ( IR2 .GT. N ) THEN
          PTILES(I) = SERIES(N)
      ELSE
          X1 = SERIES(IR1)
          X2 = SERIES(IR2)
          IF ( IFLAG .EQ. 1 ) THEN
              X1 = LOG( X1 )
              X2 = LOG( X2 )
          END IF
          XT = X1 + (X2 - X1) * (R1 - FLOAT(IR1))
          IF ( IFLAG .EQ. 1 ) XT = EXP( XT )
          PTILES(I) = XT
      END IF
   10 CONTINUE
C
C     CALCULATE THE STATISTICS
C
      SUM = 0.0
      SSQ = 0.0
      STATS(5) = SERIES(1)
      STATS(6) = SERIES(1)
C
      DO 20 I = 1,N
      SUM = SUM + SERIES(I)
      SSQ = SSQ + SERIES(I)*SERIES(I)
      IF ( SERIES(I) .LT. STATS(5) ) STATS(5) = SERIES(I)
   20 IF ( SERIES(I) .GT. STATS(6) ) STATS(6) = SERIES(I)
C
      STATS(1) = SUM/FLOAT(N)
      STATS(2) = SQRT( (SSQ - SUM*SUM/FLOAT(N)) / FLOAT(N-1) )
      X1 = STATS(1) * STATS(1)
      X2 = STATS(2) * STATS(2)
      XT = SQRT( LOG( X2/X1 + 1.0 ) )
```

```
      STATS(4) = EXP( XT )
      STATS(3) = STATS(1) / EXP( XT*XT * 0.5 )
C
      RETURN
      END
```

```
C-------------------------------------------------------------
C
C     HEAPSORT ROUTINE
C
C     THE HEAPSORT ROUTINE IS A QUICK METHOD OF SORTING A SERIES
C     OF VALUES BY TRANSFORMING THE SERIES INTO A HEAP AND USING
C     A TOP-DOWN SELECTION METHOD FOR THE FINAL SORT. THE METHOD
C     IS DESCRIBED IN DETAIL IN THE TWO REFERENCES. THESE ROUTINES
C     FOLLOW THE PROCEDURES PRESENTED IN THE SECOND REFERENCE. THIS
C     REQUIRES THE USE OF A SECOND PROCEDURE TO ADJUST THE VALUES
C     TO SATISFY THE HEAP PROPERTY.
C
C     PERFORMANCE IS OF THE ORDER ( N LOG N )
C
C     AUTHOR:  MICHAEL A. THIELE
C              10016 ORION AVENUE
C              MISSION HILLS, CA 91345
C              (818) 892-2201
C
C     DATE:    FEBRUARY, 1987
C
C     REFERENCES:
C
C     "THE ART OF COMPUTER PROGRAMMING, VOLUME 3: SORTING AND
C     SEARCHING", DONALD E. KNUTH, 1973 ADDISON-WESLEY, PP 145-149
C
C     "FUNDAMENTALS OF DATA STRUCTURES", ELLIS HOROWITZ AND
C     SARTAJ SAHNI, 1976, COMPUTER SCIENCE PRESS, PP 357-361
C
C-------------------------------------------------------------
C
C     HPSORT - HEAPSORT A VECTOR OF SINGLE PRECISCION VALUES
C
      SUBROUTINE HPSORT(R,N)
C
C     ARGUMENTS:
C        R  <=>   VECTOR OF VALUES TO BE SORTED
C        N   =>   SIZE OF ARRAY
C                 IF N < 0 THEN SORT INTO DESCENDING ORDER
C
      REAL*4  R(1),RT
      INTEGER N,J,L,GOBACK
C
      IF ( ABS(N) .LE. 1 ) RETURN
      NT = ABS(N)
C
C     CREATE THE HEAP
C
      ASSIGN 1 TO GOBACK
      L = NT/2
C
      DO 1 I = L,1,-1
      IADJ = I
      NADJ = NT
      GOTO 10
  1   CONTINUE
C
C     DO THE TOP-DOWN SELECTION AND READJUST THE HEAP
C
      ASSIGN 2 TO GOBACK
```

```
C
      DO 2 I = NT-1,1,-1
      RT = R(I+1)
      R(I+1) = R(1)
      R(1) = RT
C
      IADJ = 1
      NADJ = I
      GOTO 10
  2   CONTINUE
C
      IF ( N .GT. 0 ) RETURN
C
C     REVERSE THE ORDER
C
      DO 3 I = 1,L
      I2 = NT - I + 1
      RT = R(I)
      R(I) = R(I2)
  3   R(I2) = RT
C
      RETURN
C
C-----------------------------------------------------------------------
C
C     ADJUST THE HEAP
C
 10   J = 2*IADJ
      RT = R(IADJ)
C
 11   IF ( J .GT. NADJ ) GOTO 12
C
      IF ( J .LT. NADJ .AND. R(J) .LT. R(J+1) ) J  = J + 1
      J2 = J/2
      IF ( RT .GE. R(J) ) THEN
         R(J2) = RT
         GOTO GOBACK
         ENDIF
      R(J2) = R(J)
      J = 2*J
      GOTO 11
C
 12   J2 = J/2
      R(J2) = RT
      GOTO GOBACK
C
      END
```

```
C
      SUBROUTINE CFDIST(DFLAG,IFLAG,PTILES,X)
C
C     A VALUE IS CALCULATED FOR A RANDOMLY SELECTED PERCENTILE FROM
C     A DISTRIBUTION DEFINED BY THE PASSED PERCENTILE VALUES
C
C     AUTHOR:  MICHAEL A. THIELE
C              10016 ORION AVENUE
C              MISSION HILLS, CA 91345
C              (818) 892-2201
C
C     DATE:    FEBRUARY, 1987
C
C     PARAMETERS:  -> = PASSED; <- = RETURNED; I = INTEGER; R = REAL
C
C     IFLAG  ->  I  INTERPOLATION FLAG
C                    = 0  -  LINEAR INTERPOLATION
C                    = 1  -  LOG-LINEAR INTERPOLATION
C     PTILES ->  R  VECTOR OF 19 PERCENTILE VALUES FROM PCTILE
C                   ROUTINE USING PFLAG = 2  ( 0.05, 0.1, 0.2,
C                   0.5, 1, 2.5, 5, 16, 30, 50, 70, 84, 95,
C                   97.5, 99, 99.5, 99.8, 99.9, AND 99.95)
C     X      <-  R  THE CALCULATED VALUE
C
C     CALLS:
C
C     RANDU - UNIFORM RANDOM NUMBER GENERATOR U(0-1) FOR
C             THE RANDOM PERCENTILE
C
      INTEGER IFLAG, N, RANK
      REAL    PTILES(1), X
C
      REAL*8 RANDU
C
      REAL   PCTLS(19)
C
      DATA PCTLS /
     +    0.0005 , 0.0010 , 0.0020 , 0.0050 , 0.0100 , 0.0250 ,
     +    0.0500 , 0.1600 , 0.3000 , 0.5000 , 0.7000 , 0.8400 ,
     +    0.9500 , 0.9750 , 0.9900 , 0.9950 , 0.9980 , 0.9990 ,
     +    0.9995 /
C
C     GENERATE THE RANDOM PERCENTILE
C
      P = RANDU(0)
C
C     CHECK IF ITS ABOVE OR BELOW THE RANGE OF PERCENTILES
C     IF SO, ASSIGN THE MINIMUM OR MAXIMUM VALUE AND RETURN
C
      IF ( P .LE. PCTLS(1) ) THEN
          X = PTILES(1)
          RETURN
      ELSE IF ( P .GE. PCTLS(19) ) THEN
          X = PTILES(19)
          RETURN
      END IF
C
C     FIND THE LOCATION IN THE PERCENTILES
C
      DO 10 I = 1,18
```

```
   10  IF ( (P .GE. PCTLS(I)) .AND. (P .LT. PCTLS(I+1)) ) IPTR = I
C
C      CHECK IF WE HIT A PERCENTILE - IF WE  DID, ASSIGN THE VALUE
C      AND RETURN
C
       IF ( P .EQ. PCTLS(IPTR) ) THEN
           X = PTILES(IPTR)
           RETURN
       END IF
C
C      INTERPOLATE FROM THE 2 NEAREST VALUES
C
       X1 = PTILES(IPTR  )
       X2 = PTILES(IPTR+1)
       P1 = PCTLS (IPTR  )
       P2 = PCTLS (IPTR+1)
C
C      CHECK FOR LOG-LINEAR INTERPOLATION
C
       IF ( IFLAG .EQ. 2 ) THEN
           X1 = LOG( X1 )
           X2 = LOG( X2 )
       END IF
C
       X = X1 + (X2-X1) * (P-P1)/(P2-P1)
C
       IF ( IFLAG .EQ. 2 ) X = EXP( X )
C
       RETURN
       END
```

APPENDIX B

POWER PLANT EMISSIONS

POWER PLANT EMISSIONS

Emissions from power plants are probably the most studied source category in the emission inventory system in most air quality management districts. Emissions of oxides of nitrogen (NO_x) from a power generating unit used to be calculated from generating load and the so-called "load curve". This load curve was constructed for each generating unit and fuel type by applying a regression analysis to source test data of that unit. According to this old calculation method, an emission rate from a power plant is computed by:

$$E = \sum_i \sum_j \sum_k U_{jk} E_{ij}(L_k) \qquad (1)$$

where E is an emission over a given time k, U_{jk} is a unit function taking the value 1 if j-th fuel type is burned at k-th hour and 0 otherwise, $E_{ij}(L_k)$ is an emission rate at load L_k for i-th generating unit burning j-th fuel type.

The old calculation method described above has now been replaced with a continuous emission monitoring (CEM) system. By continuously monitoring NO_x concentrations inside the stack, CEM yield more factual emission readings than the old calculation method. However, CEM has its own limitations: a certain amount of down-time, periodic drifting from the calibrated performance, relatively low precision (said to be ±20%) and decreased accuracy at low concentrations.

The 1983 SoCAB emissions inventory (SCAQMD 1986) reports basinwide NO_x emissions from electric utilities as 39.4 tons per day (TPD) on an annual average base. This inventory estimate is said to be arrived at by summing CEM-based emission values for the Department of Water and Power (DWP) power plants and calculation-based emission values for the Southern California Edison (SCE) power plants.

As a part of the inventory uncertainty determination project, a district engineer was assigned to estimate uncertainty of the basinwide power plant NO_x emissions reported in the 1983 inventory. To do the task, the engineer searched for data and information that indicated how comparable the NO_x emissions estimates arrived at from the CEM data and these

from the calculation method are. Luckily, he found in the district library old computer printouts recording daily and monthly NO_x emissions from all generating units of DWP power plants in the SoCAB. He also found that SCE possessed similar data. From these data, the engineer constructed Table 1 showing monthly emission estimates for the DWP power plants and for the SCE power plants both by CEM and by the old calculation method.

The engineer thought that if he could calculate the 1983 basinwide power plant NO_x emissions separately by the CEM method and by the calculation method, the true emission level would be found in between the two estimated values. To this end, he calculated the following conversion factors using Table 1:

$$CF_D = \text{(CEM-based emission)/(Calculation-based emission)}$$
$$= 317.8/303.7$$
$$= 1.046$$

$$CF_S = 577.5/616.1$$
$$= 0.937$$

where CF_D is the conversion factor for DWP power plants and CF_S is that for SCE power plants.

Although he knew of the existence of a few city-operated power plants beside the DWP's and the SCE's in the SoCAB, the engineer assumed that 35 percent of the basin's power plant emissions came from the DWP's and the remaining 65 percent came from the SCE's. Under this assumption, he re-computed the 1983 NO_x emissions using the calculation method only as:

$$E_{calc.} = (0.35/CF_D + 0.65) \times 39.4$$
$$= 38.8 \text{ TPD}$$

He also re-computed the 1983 NO_x emissions using the CEM method only as:

$$E_{monit.} = (0.35 + 0.65\ CF_S) \times 39.4$$
$$= 37.8 \text{ TPD}$$

TABLE 1. MONTHLY NO_x EMISSIONS FROM DWP AND SCE POWER PLANTS IN THE SoCAB ESTIMATED FROM THE CEM AND CALCULATION METHODS (All values in tons per month)

Mo/Yr	DWP Power Plants[a] CEM	CALC	Mo/Yr	SCE Power Plants[b] CEM	CALC
Aug 82	470.5	449.7	Aug 86	548.4	495.1
Sep 82	435.3	420.5	Sep 86	351.4	332.7
Oct 82	455.9	427.6	Oct 86	621.7	621.7
Nov 82	345.9	331.2	Nov 86	467.7	531.4
Dec 82	404.7	399.7	Dec 86	423.8	489.8
Jan 83	320.0	338.1	Jan 87	639.8	658.9
Feb 83	354.5	321.8	Feb 87	699.4	699.1
Mar 83	443.3	358.8	Mar 87	867.1	862.8
Apr 83	127.0	100.1	Apr 87	594.5	710.8
May 83	150.6	156.5	May 87	468.0	534.7
Jun 83	119.1	141.0	Jun 87	655.0	724.8
Jul 83	186.6	199.6	Jul 87	593.5	731.7
Mean	317.8	303.7	Mean	577.5	616.1
s.d.	135.7	122.5	s.d.	138.4	144.1

[a] Harbor, Haynes, Scattergood and Valley

[b] Huntington Beach, Alamitos, El Segundo, Redondo, Etiwanda, Highgrove and San Bernardino

In comparing the re-computed values with the reported emission values, the engineer wondered why both re-computed values were less than the reported values. After some thinking, he decided to determine a conversion factor that would apply to both DWP and SCE power plants. This conversion factor is calculated as:

$$CF = (\text{CEM-based emission})/(\text{Calculation-based emission})$$
$$= (317.8 + 577.5)/303.7 + 616.1)$$
$$= 0.973$$

Using this conversion factor, he re-calculated the 1983 NO_x emissions as:

$$E_{calc.} = (0.35/CF + 0.65) \times 39.4$$
$$= 39.8 \text{ TPD}$$

$$E_{monit.} = (0.35 + 0.65\ CF) \times 39.4$$
$$= 38.7 \text{ TPD}$$

In summary, the engineer now has three different estimates of the 1983 NO_x emissions from all power plants in the SoCAB:

Upper Plausible Estimate	39.8 TPD
Basic Estimate	39.4 TPD
Lower Plausible Estimate	38.7 TPD

Even with these three estimates, he is unsure how to determine the uncertainty of the reported emission value because he does not know how reliable the CEM and the calculation method are for estimating power plant emissions. After some thinking, he has decided to accomplish this by asking a panel of a few colleagues knowledgeable about this source category to help formulate a judgment about the uncertainty in the basic estimate. They responded to the following questions:

Q1. On a scale of 100, what are the odds that the true level of NO_x emissions from all power plants in this air basin is less than 39.8 TPD?

Answer : _____chances in 100.

Q2. What are the odds that the true level of NO_x emissions from all power plants in the air basin is less than 38.7 TPD?

Answer : _____chances in 100.

APPENDIX C

EXHAUST GAS EMISSIONS FROM LIGHT DUTY AUTOS

EXHAUST GAS EMISSIONS FROM LIGHT DUTY AUTOS

In the South Coast Air Basin (SoCAB) the average annual emissions from light duty autos were estimated for the latest inventory year of 1983 (SCAQMD 1986). In this emission estimation, the latest emission factor model EMFAC7B was used for computing a composite emission factor while the Direct Travel Impact Model (DTIM) was employed for computing vehicle miles traveled (VMT), the number of trips generated and the average speed during morning peak, afternoon peak, and off-peak hours. Because this was a major inventory effort for the SoCAB, the most elaborate emission estimation method was used in computing composite emission factors, VMT, and the number of trips generated in each hour of the day. Details of model input data and options used are not known except that the ARB ran the EMFAC model for composite emission factor calculations and the SCAG, the planning agency for the SoCAB, ran the DTIM model for VMT and vehicle-trip simulations. The actual computer run was made on the 12th of April, 1986. The resulting total organic gas (TOG) emissions are listed in Table 1.

TABLE 1. TOG EXHAUST EMISSIONS FROM LIGHT DUTY AUTOS IN THE SOUTH COAST AIR BASIN (All Values in Tons per Day)

Exhaust Type	NON-CAT	CAT	DIESEL	TOTAL
Running Exhaust & Crank Case	173	54	2	229
Cold Start Exhaust	31	73	N	104
Hot Start Exhaust	13	12	N	26
Total Exhaust	217	139	3	359

N : Negligibly small

An analyst in the Air Quality Management District reviewed the computational procedure used for mobile source emissions and found it quite comprehensive. However, the analyst noted some discrepancy between EMFAC/DTIM-predicted heavy duty truck content and that reported in

Caltrans HPMS Report (1983) and an ARB report (Horie and Rapaport 1985) : 4.2% vs. 6.1%. The latter percentage was arrived at by taking a ratio of the SoCAB total truck daily VMT of 12.3 x 10^6 to the four-county total daily VMT of 200.3 x 10^6 for all vehicles. Assuming that the total VMT were correct, the analyst figured out that the mis-estimate in truck content caused an over-estimate of the TOG emission by 3% [= (100 - 4.2)/(100 - 6.1)] or 11 (= 359 x 0.03) tons per day (TPD). Thus, if the correct truck content of 6.1% were used, the TOG emission from light duty autos would be 348 TPD.

Recently, the analyst has learned that the EPA's Test and Evaluation Branch is working on MOBILE4, an updated version of MOBILE3, the current mobile emission model. A preliminary analysis of the new model indicates that it tends to yield about 14% greater emission factor than the current model (see Figure 1). Therefore, the analyst further adjusted his emission estimate as:

348 X 1.14

= 397 TPD

In summary, the analyst now has three different estimates of the 1983 TOG exhaust emissions from light duty autos in the SoCAB:

Upper Plausible Estimate	397 TPD
Basic Estimate	359 TPD
Lower Plausible Estimate	348 TPD

Even with these three estimates, he is unsure how to determine the uncertainty of the reported emission value. After some thinking, he has decided to accomplish this by asking a panel of a few colleagues knowledgeable about this source category to help formulate a judgment about the uncertainty in the basic estimate. The questions are the following:

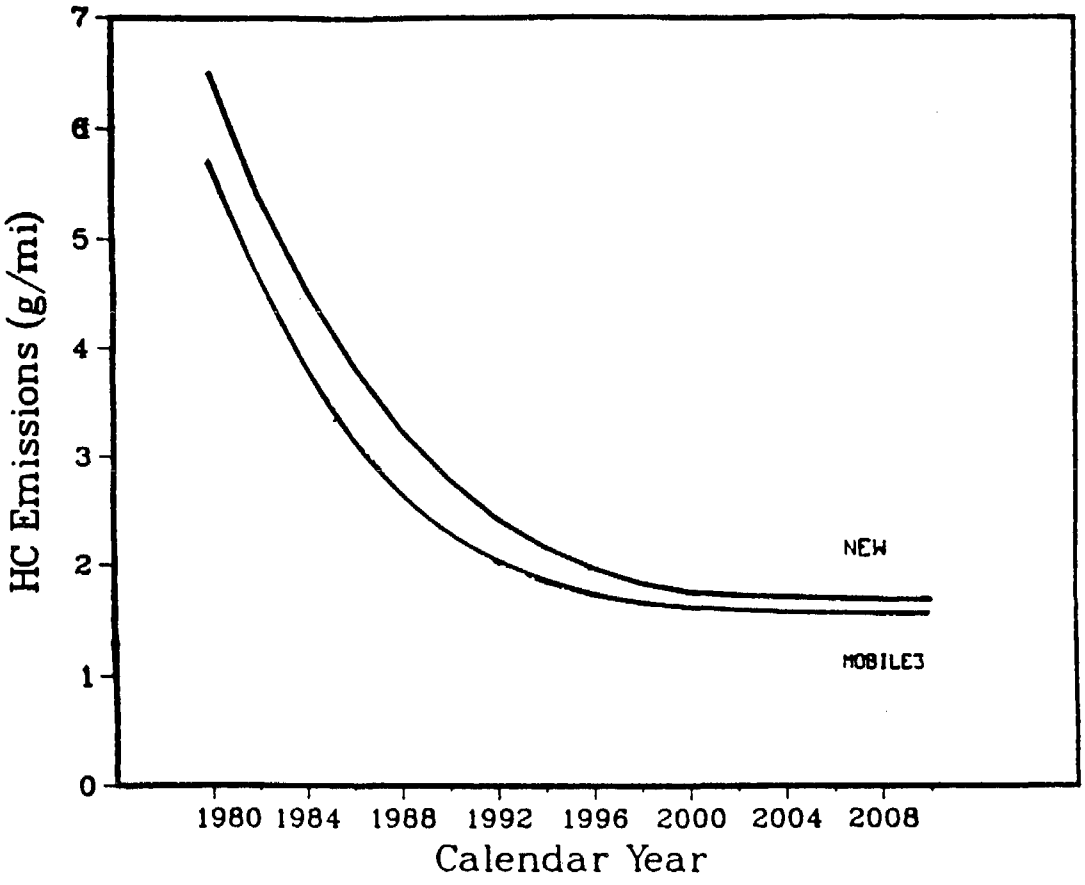


Figure 1. Total Hydrocarbons, Light Duty Gasoline Vehicles: MOBILE 3 vs. NEW

Q1. On a scale of 100, what are the odds that the true level of TOG exhaust emissions from all light duty autos in this air basin is less than 394 TPD?

Answer : _____chances in 100.

Q2. What are the odds that the true level of TOG exhaust emissions from all light duty autos in the air basin is less than 342 TPD?

Answer : _____chances in 100.

UNCERTAINTY WORKSHEET

NAME_____ DATE_____

Three estimates for LDA exhaust emissions: 397, 359, 348 TPD

Mean of three estimates: m = 368 TPD
Standard deviation: s = 25.7 TPD

Upper Level: m + s = 394 TPD
Lower Level: m - s = 342 TPD

Probability for UL _____%
Probability for LL _____%

Upper Confidence Level @ 97.5% _____ TPD
Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_n = _____ TPD
Subjectively evaluated standard deviation: s_n = _____ TPD
Subjectively evaluated bias: B_n = _____ TPD
Coefficient of variation CV = _____

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% _____ TPD
Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_l = _____ TPD
Subjectively evaluated standard deviation: s_l = _____ TPD
Subjectively evaluated bias: B_l = _____ TPD

Prefer: [ ] Normal [ ] Lognormal

APPENDIX D

DRY CLEANING EMISSIONS

DRY CLEANING EMISSIONS

A district engineer was assigned to estimate TOG emissions from small dry cleaning operations in San Bernardino County. Although large industrial dry cleaners with annual TOG emissions of 25 tons or more are already estimated in the district's major point source file, emissions from commercial and coin-operated dry cleaners have yet to be estimated.

He first estimated the emissions using EPA's AP-42 emission factors, namely, 1.3 lb/yr/capita for commercial cleaners and 0.4 lb/yr/capita for coin-operated type. The EPA emission factor rating for both was B on a scale of A through E, with A being the best. In making his estimate, he randomly picked 100 residential telephone numbers from the county telephone directories, and conducted a mini-questionnaire survey on this sample to learn the typical usage of commercial and coin-operated dry cleaners. Results of the survey were:

Commercial only	35
Commercial and coin-operated	10
Coin-operated only	25
Never used	13
No answer	17
TOTAL	100

Assuming that every resident would use one of two types of dry cleaners, he then estimated that 60% would use commercial dry cleaners and the remaining 40% would use coin-operated machines. From this, he calculated the TOG emissions using the January 1987 county population of 1,167,200 as follows:

$$E = (1.3 \times 0.60 + 0.4 \times 0.40) \times 1,167,200/2,000/365$$
$$= 1.50 \text{ tons per day (TPD)}$$

Upon review of this work, a colleague wondered whether the AP-42 emission factors are only for those who use a particular type of dry cleaning service. After reading the AP-42 explanations on the emission factors, the colleague concluded that both emission factors were meant to

be applied to the total resident population. Therefore, the correct emission calculation would be:

$$E = (1.3 + 0.4) \times 1,167,200/2,000/365$$
$$= 2.72 \text{ TPD}$$

The engineer is now in a quandary as to which method is correct. To resolve this puzzling question, he searched for other estimation methods and came across a recent ARB publication in which the county total dry cleaning emissions from small dry cleaners with annual TOG emissions less than 25 tons were given for eight counties in the San Joaquin Valley. Table 1 lists these emission values along with January 1987 urban population in each of these counties. In the table, the per-capita emission value listed in the last column is computed from the county total emissions and the resident population.

By applying the total resident population of 1,167,200 in San Bernardino County to the per capita emissions in the table, the engineer arrived at the following two estimates:

1. Use of the Fresno County per capita emission yields

$$E = 1.59 \times 1,167,200/2,000/365$$
$$= 2.54 \text{ TPD}$$

2. Use of the mean per capita emission over the eight counties yields

$$E = 0.715 \times 1,167,200/2,000/365$$
$$= 1.14 \text{ TPD}$$

Of the two estimates, he thought that the latter estimate would be more believable because the mean emission factor for the eight counties was used instead of that of a particular county, Fresno.

TABLE 1. COUNTYWIDE EMISSION FROM SMALL COMMERCIAL AND COIN-OPERATED DRY CLEANERS IN THE SAN JOAQUIN VALLEY.

County	1979 VOC Emissions (ton/yr)	1987 Resident Population	Per Capita Emission (lb/yr/capita)
Fresno	468	588,300	1.59
Kern	212	496,200	0.85
Kings	28	85,700	0.65
Madera	22	79,300	0.55
Merced	22	166,400	0.26
San Joaquin	150	435,700	0.69
Stanislaus	97	320,600	0.61
Tulare	75	287,900	0.52
Mean			0.715

In summary, the engineer now has three different estimates of TOG emissions from small dry cleaners in San Bernardino County:

Upper Plausible Estimate	2.72 TPD
Basic Estimate	1.50 TPD
Lower Plausible Estimate	1.14 TPD

Even with these three estimates, he is unsure how to provide a quantitative measure of the reliability of his basic estimate. After some thinking, he has decided to accomplish this by asking a panel of a few colleagues knowledgeable about this source category to help formulate a judgment about the uncertainty in the basic estimate.

The panel members will be asked to respond to the following two questions:

Q1. On a scale of 100, what are the odds that the true level of TOG emissions from small commercial and coin-operated dry cleaners in this county is less than 1.79 TPD?

Answer : _____chances in 100.

Q2. What are the odds that the true level of TOG emissions from small dry cleaners in the county is less than 0.83 TPD?

Answer : _____chances in 100.

UNCERTAINTY WORKSHEET

NAME_____ DATE_____

Three estimates for dry cleaning emissions : 2.72, 1.50, 1.14 TPD

Mean of three estimates: $m = 1.79$ TPD

Standard deviation: $s = 0.83$ TPD

Upper Level: $m + s = 2.62$ TPD

Lower Level: $m - s = 0.96$ TPD

Probability for UL _____%

Probability for LL _____%

Upper Confidence Level @ 97.5% _____ TPD

Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_n = _____ TPD

Subjectively evaluated standard deviation: s_n = _____ TPD

Subjectively evaluated bias: B_n = _____ TPD

Coefficient of variation CV = _____

If CV 0.3, proceed to log-probability paper.

Upper Confidence Level @ 97.5% _____ TPD

Lower Confidence Level @ 2.5% _____ TPD

Subjectively evaluated mean: m_l = _____ TPD

Subjectively evaluated standard deviation: s_l = _____ TPD

Subjectively evaluated bias: B_l = _____ TPD

Prefer: [ ] Normal [ ] Lognormal