ON THE PROPAGATION OF ERROR
IN AIR POLLUTION MEASUREMENTS

JOHN S. EVANS, DOUGLAS W. COOPER, and PATRICK KINNEY*

Department of Environmental Science and Physiology Harvard School of Public Health,
Boston, MA 02115, U.S.A.

(Received May 3, 1983)

Abstract. Four methods for estimating the uncertainties in air pollution measurements are outlined. The approaches are: analytical solution – approximation; application of distribution theory; experimentation; and simulation. The advantages and disadvantages of each method are illustrated using data from High-Volume air samplers, the instrument most commonly used for monitoring ambient concentrations of airborne particles.

1. Introduction

In air pollution studies, as elsewhere, important variables are actually derived from measurements of more fundamental quantities. Concentration, for example, is often determined by the change in mass of a filter, the volume flow rate, and the duration of sampling. Uncertainties in the measurements of each of these quantities then combine to produce uncertainties in the inferred value of the variable of interest, concentration. This paper presents four approaches to estimating uncertainty in the derived quantity from uncertainties in the measurements of the fundamental quantities:

(a) analytical solution – approximation;
(b) application of distribution theory;
(c) experimentation;
(d) simulation.

The application of each method is illustrated using the problem of determination of the total uncertainty in mass concentration measurements given by the High-Volume air sampler.

We are addressing the use of a single set of measurements to calculate a single value of a derived quantity. Reasons for wanting to know the uncertainty in that derived quantity are partly determined by the use to which that quantity will be put, but also we might use the uncertainty estimate to decide how many replicate measurements to make (the uncertainty diminishing with the square root of the number of independent determinations) or whether the equipment is suitable for the task, or for which component of the derived quantity it is the most cost-effective to improve measurement precision.

* Dr Evans is Assistant Professor of Environmental Science and Mr Kinney is a doctoral candidate in the Department of Environmental Science and Physiology, Harvard School of Public Health. Dr Cooper is at the Manufacturing Research Laboratory, Thomas J. Watson Research Center, IBM, Yorktown Heights, NY.
Statistical uncertainties in a measured quantity, or 'measurement errors', are often thought of as falling into two classes, systematic errors (or bias) and random errors. We restrict our attention in this paper to random errors, those errors which increase uncertainty but do not create bias.

Surprisingly little attention has been paid in the recent air pollution literature to these important aspects of data collection and interpretation. Most often, an analysis of random measurement errors (as opposed to systematic biases) begins and ends with an estimate of the variance of the overall distribution of errors, often termed 'precision' (see e.g., Fein and Bailey, 1978). The distribution of errors is almost always assumed to be normal. A useful refinement which occasionally appears is the attempt to identify and estimate each of several individual sources of error and to evaluate their relative influence on the total error (Jaklevic et al., 1981; ACGIH, 1978; USEPA, 1973). The value of such an approach lies in its ability to indicate those components of the total measurement system upon which additional control efforts will be most effective in reducing overall measurement errors.

The derived quantity we study here is the concentration as determined by a High-Volume air sampler. The High-Volume air sampler draws air through a filter which is almost perfectly efficient (> 99.97%) at capturing particles. The mass concentration of particles in the sampled air is estimated on the basis of the change in weight of the filter during the sampling interval and the volume of air passed through the filter. Mathematically, the correct value, \( f_0 \), is a function of three variables:

\[
 f_0 = x_1 x_2^{-1} x_3^{-1}
 \]

where:
- \( x_1 \) = net change in filter mass during sampling interval (µg);
- \( x_2 \) = average flow rate during sampling interval (m³ min⁻¹);
- \( x_3 \) = duration of sampling interval (min).

However, the observed TSP concentration, \( f \), depends not only upon \( x_1 \), \( x_2 \), and \( x_3 \), but also upon the errors in measurement of each, \( dx_1 \), \( dx_2 \), and \( dx_3 \):

\[
 f = (x_1 + dx_1)(x_2 + dx_2)^{-1}(x_3 + dx_3)^{-1}.
 \]  

The error in measurement of the TSP concentration, \( df \), is then given by:

\[
 df = f - f_0.
 \]

A complete investigation of measurement errors might have as its goals: characterization of the distribution of errors, estimation of the mean and variance of the distribution of errors, generation of confidence intervals for individual values of \( f_0 \) based upon knowledge of \( f \) and \( df \), and determination of the relative importance of each of several sources of error.