PB 262 393

EPA-600/4-76-055
November 1976

Environmental Monitoring Series

URBAN AIR POLLUTION MODELING WITHOUT COMPUTERS

Environmental Sciences Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U S Environmental Protection Agency, have been grouped into five series These five broad categories were established to facilitate further development and application of environmental technology Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ENVIRONMENTAL MONITORING series This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

EPA-600/4-76-055
November 1976

URBAN AIR POLLUTION MODELING WITHOUT COMPUTERS

by

Michael M. Benarie
Institut National de Recherche Chimique Appliquée,
B.P.1 - 91710 Vert-le-Petit, France

Visiting Scientist
Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
ENVIRONMENTAL SCIENCES RESEARCH LABORATORY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Environmental Sciences Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

The modeling of urban air pollution by mathematically oriented techniques has for some time provided the primary quantitative basis for the development of air quality management concepts and strategies. As such, the topic of air quality modeling has high priority in the research program of the EPA.

The material of the present report formed the basis for a series of three lectures given by Dr. M. Benarie, Chief of the Atmospheric Pollution Service, Institut National de Recherche Chimique Appliquèe, Vert le Petit, France. The lectures were first given on 15-17 Sept. 1976 in Raleigh, N.C., as part of the "Continuing Seminar on Air Quality Research", which is a joint activity of the Meteorology and Assessment Division, EPA and the North Carolina State University. They were also repeated on 20-22 Sept. at the Pennsylvania State University, University Park, Pa., under support of an EPA grant with the Select Research Group on Air Pollution Meteorology at P.S.U. The publication of this material as an EPA report is made in the interests of wide dissemination of the information to air quality modelers.

Kenneth L. Calder
Chief Scientist
Meteorology and Assessment Division
Environmental Sciences Research Laboratory

PREFACE

The lecture material that is consolidated in this report represents an abridged version of some selected chapters of a monograph entitled Urban Air Pollution Calculation, by the same author. As the title indicates the selection was oriented towards simple but nevertheless efficient methods. The present discussion stresses the reasons why these methods are useful, which are their recommended fields of application, and when they are to be preferred to other kinds of calculation. As compared with the monograph this abridgement differs primarily in the completeness of the coverage. Here only about one tenth of the references are included and discussed. The primary aim is to stress principles and only to present examples that are really necessary. In contrast the monograph tends to be as complete as possible, by providing comprehensive current information on references, formulas, applications and validations.

The author would like to express his thanks to Mr. K. L. Calder who kindly provided editorial assistance on the first draft of this report.

ABSTRACT

This report was the basis for a series of three lectures by the author on urban air pollution modeling, and represents a condensed version of selected topics from a recent monograph by him. The emphasis is on simple but efficient models, that can often be used without necessitating a high-speed computer. It is indicated that there will be many circumstances under which such simple models will be preferable to more complex ones. Some specific topics included in the discussion are the limits set by atmospheric predictability, forecasting pollution concentrations in real time as for pollution episodes, the simple box model for pollution concentrations, the frequency distribution of concentration values including the log-normal distribution and averaging-time analysis, the relationships between wind speed and concentration, and lastly the critical question of model validation and the need to consider several indices of goodness-of-fit if pitfalls are to be avoided.

CONTENTS

FIGURES

TABLES

SECTION 1
INTRODUCTION

It is not in the role of some latter-day follower of the machine-smashers of the 1840's that I chose the title. At that time tailors fearing unemployment destroyed the first sewing machines. I would like to speak about "computer-less" methods and about "unsophisticated methods of modeling." However, this will certainly not be done with any sense of glee that "see, here the abacus outperformed the computer." The abacus will never do that. My point of view is strictly that of the engineer. I consider the test of the engineer to be the attainment of a given practical goal by the most economical of the means available. The scientist and research worker on the other hand seek to improve knowledge without consideration to cost, effort and time, while the inventor labors to increase the available means. The good engineer will happily use the output from the research worker or that of the inventor, but his purpose is practical such as to build a bridge, house, highway or gadget. He has to deliver the bridge, etc., on time and meet specifications, while avoiding any suspicion of a gamble. On the other hand, all research projects contain some element of a gamble. They verify, validate, or prove some hypothesis; they compare or attempt to test or they search for something as yet undisclosed by nature. At the outset, we always hope for a positive answer, but a hope is only a gamble. On the other hand, an engineer who "hopes" is truly a bad engineer. An engineer <u>must</u> deliver his product in the same way as a manufacturer must do.

In our case, the product to be delivered mostly takes the form of information. It is the result of a calculation or of some process akin to calculation. These calculations are usually undertaken in order to avoid the unrealistically high cost of of full scale experiments. In the same way, it is cheaper to calculate the strength of a beam than to shatter a real one; it is also less expensive to calculate the impact of a power plant on an urban area than first to build one and then see what happens.

Before undertaking any urban air pollution calculations, the following questions should be answered:

1. who needs the information and
2. what purpose has to be attained.

Table 1 presents a schematic outline of the answers to these questions.

The user has first to define his operational needs: If an annual arithmetic mean is requested simply to check conformity with an air quality standard expressed as an annual mean, it would obviously be foolish to obtain it from 8760 hourly estimates, when more direct and cheaper methods are available. High resolution--here as everywhere--costs money. This money and labor are spent first in gathering the high-resolution input data, and then in working out the fine details of the output. High resolution data do not necessarily mean accurate or true data. Thus the first consideration we must face before adopting some computationally sophisticated method--which, by the way, rather often coincides with computer vs computer-less methods--, is whether the quality and the quantity of the available input does in fact justify the computational burden. On the other hand, some computationally simple methods need high-grade input information. An obvious example is the "persistence model" which without any calculations, when projected a short time forward, will give a fantastically good fit, because it already contains a tremendous amount of accurate information. Needless to say, extrapolation for a few years, or even a few days ahead, would be pointless.

The juxtaposition of "simple" box models, requiring little computing effort, with mathematically and physically "sophisticated" models, that probably necessitate the use of large digital computers, does not necessarily mean that one is better than the other. First of all it should be emphasized, as Gifford (1973) pointed out, that "simple" is not the opposite of "sophisticated," but of "complex." The antonym of "sophisticated" is "naive." Simple urban air pollution models are not always naive, and may in fact be quite sophisticated. Conversely, complex models can be quite naive, or can contain naive assumptions.

Furthermore, if a complex urban pollution model cannot estimate more accurately than a simple one, say like the use of persistence, then its development is not profitable in applied studies (Gifford and Hanna, 1975). This does not mean that the theoretical importance of a complex model is diminished if it provides a better understanding of the underlying phenomena.

This leads directly to another question: Is there some fundamental limit to the accuracy of the model computations? If there is such a limit, then it is clearly pointless to use computational methods of much greater precision as these only contribute to the proliferation of non-significant figures in the estimate. Much of the discussion of the following section will be concerned with the search for such limits.

Table 1 is an attempt to systematize computations on the basis of their purpose. Another classification can be based on the amount and nature of external - mostly meteorological - information, needed as input for the model.

Meteorological parameters have an overwhelming influence on the behaviour of pollutants in the urban air. Among them, wind parameters (direction, velocity and turbulence), and thermal properties (stability) are the most important. A classification of the models can be based on the method in which this kind of input is generated. The following discussion will use "wind field" as shorthand for all the dynamical and thermal properties associated with the wind.

In some models, the wind field is assumed to be known or has to be fed in by forecasting techniques. Mahoney (1971) has coined the word "driven" for this kind of input.

In a second category of computations, a consistent wind pattern (in the vicinity of the urban area), is either calculated from a full set of meteorological model equations, or an actually observed wind field is used as input

Finally, there are representations - also called models - that provide statistical information on the occurrence of pollutant concentrations, and which do not make use of wind or other meteorological parameters as input.

A further distinguishing feature, different from any of those discussed above, appears in the nature of the model as to whether it is source - or receptor-oriented. The distribution and the emission rate of pollutant sources are assumed to be known in "source-oriented models." Pollutant concentrations are then calculated from this source distribution over the entire region of the model.

The opposite is true for the "receptor-oriented models": In their pure form no assumptions are made about emissions, and only ambient concentration is monitored at a number of receptor sites. Statistical or other inferences, which may or may not be linked to meteorological information, are then drawn - and possibly extrapolated - from the observed data.

Source-oriented models tend mostly to be explanatory and involve causal relationships between the pollutant emissions and concentrations. Only explanatory models can provide the necessary means to control the system and produce desired changes in performance. Receptor-oriented models are generally descriptive and less directed toward establishing cause and effect relationships.

Table 1 also distinguishes between short- and long-term objectives, i.e. whether the result of the calculation is needed in the next few hours, or in a few years. This aspect generally coincides with the distinction made between computation of short-time concentration values or forecasts, and the request for long-time averages. A typical, but rather arbitrary, separation between these two classes, would be 24 hours. Anyway, the basic idea in the 30-minute averaging time computation and that for the seasonal or yearly average, show enough difference for recognition of two quite distinct classes for short- and long-time mean calculations.

The foregoing principles enable us to classify the main types of urban air pollution "models," as shown in Table 2.

Examination of Table 2 shows that beyond the main classification criteria (e.g. of short- or long-term averages; source- or receptor-oriented) four kinds of models may be distinguished in terms of the type of information they provide. This distinction may be termed "model character" and is denoted by letters from A to D in Table 2.

A. This letter denotes models which use either assumed or actually observed values for the meteorological parameters.

With assumed parameters, the plume and volume-element models give numerical results, i.e., concentration values as a function of the space coordinates. It is beyond the scope of the model to consider whether or when the assumed set of meteorological descriptors will materialize. The results are only as good as the input data. This category of models provides ambient concentration from inputs, and is analogous to the situation in chemical engineering where content of a reactor is computed from reactants, stirring, temperature, etc. The output is primarily a numerical value asssigned to a space and a time coordinate. Usually, this kind of calculation justifies the use of a computer, mainly when a larger set of computations is being made. This is often the case in long-term calculations (plume combined with frequency classes).

On the other hand, the "box model," which can conceptually be derived from the principle of mass conservation, is an example of a model where the use of a computer can often be avoided.

B. In forecasting pollution, the output from the calculation is expressed sometimes numerically, but more often by categories, by probabilities or in some other convenient way as used in meteorology. The quality of the forecast is limited by the atmospheric predictability.

C. Statistical description, in either of its forms, is a summary of data already on record. Valuable for predicting trends or cycles, it is of little use for a true forecast (i.e., today's estimate of tomorrow's pollution). For data-management, compilation and computation, the computer is almost a necessity; for search and exploitation, it is only an advantage.

D. Finally, we have the description (which may also be termed "statistical") or summarization of data already on record, mostly in form of graphs or tables, and intended mostly for long-term inferences. The output is in terms of a frequency not assigned to any time coordinate. Although the preliminary tabulation is facilitated by use of a computer, the use of the resulting tables or graphs is mostly computer-less.

The use of computer-less methods may be advisable, mainly in the following situations:

1. When relatively low resolution information (e.g., just an annual mean) is sought.

2. When the available input information does not justify a complex algorithm; when the range of error of a high precision calculation would still be large because of observational inaccuracies. This case is rather more frequent in air pollution than generally assumed.

3. When the predictability of the atmospheric motions sets an upper limit to the predictions of the models.

SECTION 2

THE LIMITS SET BY ATMOSPHERIC PREDICTABILITY

We shall distinguish between the conservation of the identity of air parcels and our ability to simulate or compute the trajectory of these air parcels.

Let us suppose that at a certain instant of time volume elements of air can be marked by tracers, which are "ideal" balloons that are able to follow every motion of the surrounding air, and the tracks of which can be observed. Thus, each mesh cube is determined by eight balloons in the atmosphere. These 'mesh particles' will undergo a rapid change of their shapes during the following days, long bands will be stretching, and finally the development will proceed to a chaotic state where the 'particles' have lost their identity.

All particles have the shape of a cube, i.e., as bounded by squares at $t = o$. A particle will be said to have ceased to exist if one of the corner points of the quadrilateral crosses one or both opposite sides during the course of time.

Robinson (1967) found that a particle with mesh size equal to 300 km should cease to exist within the period $12h < t < 75h$. Egger (1973) using the data of KAO (1968, 1969) on large-scale dispersion of clusters of particles in the atmosphere, suggests $45h < t < 72h$ while using the data of EOLE (Morel 1972; Larcheveque, 1972) he arrives at $t \approx 45h$.

This estimate is one on an upper limit for atmospheric predictability. No numerical forecast model however it is designed, can do better than this. Our ability to predict is further limited by the following factors:

One of these arises from the finite representation of the atmospheric fields in the models, which makes it impossible to describe scales of motion below grid scale. Due to the nonlinearity of the hydrodynamic equations, parts of the turbulent energy which is contained in the subgrid range, will appear under an alias in the larger scales, thus limiting the predictability of these scales. "It is this last type of uncertainty that is generally felt to be responsible for the limit of predictability of various scales" (Fleming, 1971). Another factor is the insufficient knowledge of the initial conditions, such as errors in the raw data.

For planning purposes, we want to be able to calculate the influence of different sources at specific sites, on specific locations of the urban area. By using source-oriented models we attempt to establish a cause - to - effect chain, between the emissions of a number of sources and the ambient concentration at given locations. The main links of this chain are the following:

1. Knowledge of the source strength.

2. Adequate definition of the meteorological parameters.

3. A reliable method for the calculation of the dispersion from inputs 1. and 2.

4. Adequate knowledge of the pollutant losses (or formation) by chemical or photochemical reactions.

Almost all these requirements can be subdivided into many parts. Therefore in passing from source to ambient concentration a total of ten to twenty elementary processes have to be estimated. Only a few of these can be calculated free of error. Many can only be estimated roughly so that each estimate may be tainted by large instrumental or theoretical uncertainties. Almost all of these errors increase with decreasing wind velocity. A brief summary of the facts is as follows:

1. Source strength : wind velocity has no influence on this factor.

2. The main meteorological parameters--wind velocity and direction--are not monitored by currently available instruments when the wind velocity sinks below 1 or 2 m s^{-1}. However, this is not only an instrumental difficulty that could be remedied in the future. For the literature on turbulent motion in the atmosphere is unusually scarce on the topic of the directional variance of very light winds. This arises from the fact that the stability of high building structures and the safety of aircraft is not affected by such winds, and specialists in these areas of research have more immediate

problems at the upper end of the scale. Theoreticians are embarrassed by the lack of an exact approach and prefer to pass on to other topics. In contrast the synoptic meteorologists are fully aware that light winds most frequently are variable and with poorly-defined directions. The common observation of a weathervane or of sailboats under such conditions, makes it unnecessary to cite in detail the few available tetroon-flight experiments. These only add very little to the already plentiful evidence. In the monograph of Lumley and Panofsky (1964, p 151) which contains extensive information about atmospheric turbulence, one only finds the following brief statement on the subject of the standard deviation of the wind azimuth: "The unexpected feature is the tremendously large scatter and the frequently considerable values of standard deviations in stable air. Further analysis of the observations of inversions indicates that the largest standard deviations of azimuth occur in light-wind conditions.......gradual azimuth drifts with periods of the order of 20 minutes were observed in light-wind inversions. The origin of these drifts is unknown. Their occurrence adds two difficulties to the estimation of the lateral diffusion: first, they make lapse rate and wind speed poor indicators of lateral wind fluctuations; secondly, even if the standard deviation of azimuth is known to be large, it is not known whether these large, but more or less local, standard deviations produce rapid spreading of air pollution."

3. All known plume dispersion equations have a singularity near zero wind velocity, and therefore their use at very low velocities becomes suspect.

4. The incomplete knowledge available about pollutant transformations and sinks, is certainly not made any less important in the case of light winds. Optimistically, we can only hope that these deficiencies in knowledge will not increase the error under calm conditions.

Thus, even if low wind velocity did not influence the questions 1 and 4 above, its effect through the questions 2 and 3 would be so overwhelming that source-oriented models would break down completely during light winds. It can be conjectured from evidence concerning urban airflow and urban heat islands that during conditions favorable to the formation of an urban heat island, the source-oriented models will be of no use. In numerical terms, this limit might be expected when the geostrophic wind diminishes to less than

$3\ m\ s^{-1}$. Very probably, this is not a rigid limit, but varies with the city size.

It should be emphasized that because of these arguments, we do not speak about the usefulness of calculations based on plume-dispersion formulae at very low wind speeds. The whole model concept, as being the causal chain between pollutant source and ambient concentration, becomes meaningless when the wind velocity falls below a certain value.

To express these considerations in the terminology of operational research, one would say that we are dealing with a multi-nodal chain. At each node, along with some information, we introduce more or less random noise. Yet, just such a multi-nodal chain with noisy input could be used to simulate the outcome of a throw at roulette. For suppose the torque applied to the roulette wheel could be electronically monitored, and assume the same for the velocity and the angle of the roulette ball. Then apply the known accurate equations of the mechanics of rigid bodies. Do a few more steps of computations and you have the final definitive system to beat Las Vegas.

Obviously, you will never be able to do that. But by the same logic, multi-nodal models with the introduction of random noise at every step <u>will not</u> indicate with accuracy tomorrow's pollutant concentration. On the contrary, the more steps (nodes) that are used, the less accurate will be the forecast of the outcome of any one individual occasion (calculation). Sophistication may be a way to improve the precision of averages, of findings about categories or to observe trends, but it seems of no use for improving the accuracy of forecasts.

The strong statement should not be interpreted as saying that all sophistication is definitively to be rejected. As the body of this paper will show, some very simple one- or two-step schemes show an honorable, if not outstanding performance mainly in forecasting. On the other hand, if very sophisticated, long-chain arguments <u>must</u> be bad, then there is some intermediate length of operational chain which might give optimum results. Research should be oriented towards methods which are intermediate between utmost simplicity and noisy sophistication.

The roulette wheel is an example of a mechanical system beyond the reach of mechanical cause-to-effect calculations. But we shall try to develop this concept gradually, by considering a heavy beam supported by an axis of low friction situated near to its center of gravity. If the latter is below the axis, the device becomes a sensitive balance. Any perturbation of the balance can be described analytically, in terms of oscillation and equilibrium positions. If, however, the center of gravity and rotation axis are made to coincide (they never actually do), the angular position at which the beam will stop, can no longer be predicted analytically and the problem becomes one of probability. Somewhere in the process of approaching the axis to the center of gravity, the chain of causality has broken down and has been replaced by a probability situation. Of course, I do not wish to discuss the fundamentals, as these are well-known from the probability calculus; my purpose is only to emphasize that a similar situation occurs in urban air pollution as for the example of beam. When the chain of governing equations between cause and effect becomes too long, and at each step rather unknown perturbations are introduced then the use of calculus should be abandoned and a new probabilistic approach should be attempted.

This is what occurs in urban air pollution, when the wind velocity sinks below approximately 3 m s^{-1}; above this lower limit, atmospheric aerodynamics is a powerful tool, but below or close to it, hydrodynamical equations are of as much use as classical mechanics would be for calculting the face on which a dice will fall. There are two distinct regimes in urban air pollution: one for strong to moderate winds and another for light winds during calm con-conditions.

The difference between urban air flow conditions with moderate and strong winds and those during light winds, and also the fact that street ventilation changes character when rooftop wind speeds fall between 2 and 5 m s^{-1}, is already well-stressed in the literature.

Insofar as source-oriented models rely on classical analytical equations and on a cause-to-effect chain, they will behave very poorly in warning systems or in episode control strategies, because generalized and protracted pollution episodes occur mostly during moderate and light winds. On the contrary, plume-concepts can be quite useful to localize pollution effects due to point sources or groups, when the winds are above 3 m s^{-1}.

By the same argument, source-oriented models, when used as a basis for long-term averages, may be useful if treated with circumspection and provided that light winds and calm periods only happen infrequently. However, when meteorological tables of the urban area of interest indicate that even only 5 to 10% of the winds are below 2 m s^{-1}, then the validity of the concentration distribution as computed by a source-oriented plume model, should be questioned. Numerically, these concentrations will be in gross error at the higher levels, which--even if they occur with low frequencies--are the most important ones, as regards effects.

Receptor oriented models, sometimes with some empirical keying to the source inventory, can be used for warning systems, provided that meteorological parameters are correctly forecast. The vital question is, what can be reasonably expected from this kind of forecast.

Though nowhere clearly stated, a widespread belief prevails in air pollution circles. It seems to say that for any two time intervals, characterized by unchanged emission rate and by approximately two score meteorological parameters (such as wind direction and intensity, thermal gradient, cloudiness, the situation of a given air parcel relative to a front, etc., etc.), if <u>all</u> these parameters were equal <u>then</u> pollutant concentrations would also be the same, for both time intervals.

By the same logic, it could be expected that if two or three score appropriate parameters were identical, then the same form of cumulus cloud would hover over the same quarter of the city. Of course, nobody would dare to assert this as fact. Continuing in this vein, we should not expect that pollution concentration forecasts will be fully accurate, all the time.

The following example may also emphasize what can be reasonably expected, as regards the accuracy of air pollution concentration computation. The average deviation from scheduled arrival times at Paris airport, due to weather conditions, was only six minutes, during 1973 (personal communication). Flights cancelled before departure, as well as delays due to technical or commercial reasons, are not figured in this statistic. Considering that the average flight times were about three hours, this means that the "estimation" was done with 4% error. Now, these aircraft are driven by thousands of horsepower, guided by exceptionally skilled crew, assisted on the ground by other most competent people and the most powerful computers ever built. If all this complex system results in a 4% relative error, then how can we expect that the calculation of an air parcel's trajectory, driven by its own buoyancy and some turbulent airflow only - instead of a jet engine - should perform any better?

SECTION 3

FORECASTING POLLUTION

Perhaps the least computerized branch of air pollution engineering relates to the forecasting of pollution. We have to distinguish quite clearly between forecast and calculation. The latter term is used to denote the operation of taking some formula - e.g. plume, statistical time series or anything else - and then substituting into this formula some assumed (e.g. for the next winter season...) or meteorologically forecast parameters. The forecast on the other hand is a process which used knowledge that is available today (e.g. past statistical record, to-day's pollution concentration, to-day's meteorological forecast, etc.) to predict (a) a time (e.g., tomorrow; or even a given hour...) and (b) a pollutant concentration for that time. The upper limit of the time span is that for which a forecasting skill can be demonstrated, and might be for a few hours or a few days in advance. We exclude from "forecasting" the climatological estimate of long-term averages, although it is implied that they are often taken into account by the forecaster. In order to be termed a "pollution forecast," the pollution concentration estimate must refer to a specific day or hour and not to a probability of occurrence within a given time span.

Almost synonymous with "air pollution forecasting" are: "episode forecasting" and "alert announcements." By "episode" and "alert," everybody means a spell with above-average pollutant concentrations. However considerable confidence must exist as to the likely duration of the spell and how high the concentration may rise before calling an "alert." If by "episode" we understand a relatively high, and not too-frequent pollution level, and not just the fact of exceeding some hygienic or legal limit, then by definition an episode is a statistically rare event or an extreme occurrence. Former experience about such events is generally scarce, and difficulties increase when the episode level is set very high. If the level is set unreasonably low, so that it is often attained through stochastic fluctuations, then all predictive qualities will be suppressed by the noise. In terms of the cumulative concentration frequency, it would not be sound practice to attempt to forecast the upper 0.1 percentile of the concentrations. On the other hand, concentration forecasts of either below or above the 50% percentile would scarcely be considered as episode forecasts, but rather as being air pollution concentration forecasts limited to two classes. Thus the success or the skill in episode forecasting depends in very large measure on the definition of an episode. Often instead of using the concentration of a single pollutant to set the episode criterion, an "air pollution index" is defined. This may be considered as a scheme that transforms the weighted concentrations of several individual pollutants into a single number. OTT and THOM (1976) found that 35 U.S. metropolitan air pollution agencies use some form of air pollution index and no two indices are exactly the same. Thus, an index value of 100 reported in Washington, D.C. means something entirely different from a value of 100 reported in Cleveland, Ohio.

The goal of forecasting a numerical value for the concentration in air pollution episodes is statistically more elusive than that of the day-to-day prevision of the "episode." With stringent definition of the episode intensity, 90% or more days are non-episode days even in winter. Hence, a random guess, based on the average frequency of the non-episode days, will yield 80% or more correct forecasts. If instead of a random guess, one decides to forecast "non-episode" for each day, then 90% or more of the days will be "correctly" forecast. As during non-episode days the concentration is relatively low, a constant-value estimation near the ensemble average will minimize the RMS error, without demonstrating any real skill of the method. If concentration episodes are predicted exclusively, such spurious effects cannot perturb the judgement on the method.

The difference between air pollution potential forecasting and episode forecasting is that the first does not take into account pollutant emissions, while the second is concerned with the emission-dispersion interaction. The first is a weather forecast that is oriented towards ventilation-forecasting, while the second is an air pollution forecast.

The difference is also one of scale. Air pollution potential forecast concerns a large area on a regional scale, covering perhaps 200,000 km^2 or more, and extending in time over 24 or 48 hours. These scales are related to the controlling synoptic event, the warm core stagnating anticyclone which, with the associated light winds and subsidence, reduces atmospheric dispersion and enhances the accumulation of pollutants. Thus this forecast requirement is primarily dependent on meteorological variables (W.M.O. 1972).

Urban episode forecast has a space-time scale where local circulations (e.g. land-sea breeze, drainage winds, heat island effects, etc.) become very important, especially during large-scale stagnation situations. The length scales are from 10 to 100 km, the latter for large megalopolis areas. The time scales of the prediction requirements range from a few hours to about two days.

All forecasting represents a correlative evaluation of the post hoc - ergo propter hoc type. It was observed that calm winds and restricted vertical exchange are conducive to high pollution episodes. How many times was it observed? Perhaps nobody cared to count and to make up a contingency table. It is not necessary to be a scientist, nor even a grown up human in order to link together two simultaneous or subsequent events into a predictive correlation. Animals build up conditioned reflexes in the same way.

Air pollution levels expressed in such terms as "insignificant", "near average" or "high" can be surprisingly well predicted by a skilled observer who is familiar with the air pollution record of a site, has access to the current weather forecast, and who is able to look out through the window. Such "no cost, no computer" forecasts provided from 75% to 85% correct answers (depending on the forecaster and the season) for the Rouen and Strasbourg (France) urban areas. Such performance equals or even surpasses that of much more elaborate schemes described in the literature.

The opponent of the sophisticated would now exclaim with relish, that simplicity has finally triumphed! This however, is not really the case. The skilled observer here performed with a sophistication not equaled by any computer program available at the present time: he recognized a pattern. We do not have a computer able to identify a handwritten numeral, a signature or a face. Almost any human being can do these things. A computer cannot even recognize the form "circle", unless pre-coded in color, contrast, dimension, perspective, etc. The human is able to do these things automatically. Thus the human outperforms the computer in a specific skill called pattern or gestalt recognition. We do not necessarily have to relinquish the computer. We must simply chose the pattern elements, or in other words the predictors, and the series or the program which will enable the computer to perform objectively the same forecast which was made subjectively and unconsciously by the skilled observer.

The pattern elements or prescriptors must be chosen in an economic fashion. For not only is their observation expensive, but when the number of prescriptor class combinations becomes large, the number of cases included in each category may be undesirably small, even for the longest stable records that are available.

When the prescriptors are of qualitative or discontinuous nature, they can easily be divided into classes, and contingency tables become very useful. The number of classes or groups should be small, and usually less than five for each variable. Contingency tables can strictly only be used if it can be assumed that the data are independent of each other, which is rarely the case in air pollution climatology. The best way to test stability is to generate a new multiple contingency table of similar size and then compare the relative frequencies in the two tables. From such a comparison it should be possible to determine how the system will work when used in actual forecasting. When using a contingency table in forecasting, the actual predictor combination is determined first, and then the forecast will be the predictand group that occurred with the highest frequency in the past. The probability of the forecast can also be estimated.

MOSES (1969, 1970) has described the implementation of this method at the Argonne National Laboratory. Called "the Tabulation Prediction Scheme," the places of the prescriptand values and the frequencies were reversed. The column headings in the Table are the minimum, the 25, 50, 75, 90, 98, 99 cumulative percentiles and the maximum. The entries for each column then give the respective pollutant concentrations: this can be seen in Table 3. Also shown are the inter-quartile range, the difference in SO_2 concentrations between 75 and 95 percentiles, the mean, the standard deviation and the number of cases for each entry. A computer is not an absolute necessity for this compilation, although it certainly makes it easier.

As is the case for any other empirical statistical model, the tables of the tabulation prediction scheme must be continuously updated. The use of this scheme in an air pollution incident control test in Chicago, Ill. was demonstrated by CROKE and BOORAS (1970).

Although the tabulation prediction scheme is easy to use - it is possible to look up any set of meteorological conditions just as one would look up a word in dictionary - a considerable amount of insight is necessary in order to develop an effective set of tables. For fuller details on their selection of variables, and construction of the tabulations and application, see CROKE and ROBERTS (1971, p. 169-184).

This selection may also be performed by a computer. By choosing a set of predictors, which may initially contain useless or redundant ones, an adaptive pattern classifier, which is a device whose actions are influenced by its past experiences, can be used to assign each pattern to a category that has been a priori characterized by a set of parameters. First the pattern is digitalized by a "preprocessor." If any of the dependent variables prove to be misleading or irrelevant, then a technique must be devised for their deletion.

RUFF (1974) used the adaptive pattern classification for the forecast of ozone levels above 0.1 ppm at San Jose, California.

The following list of trial inputs were used as ozone predictors for San Jose:

1. NO_2 - Selected because it reacts with sunlight to ultimately form O_3.

2. CO - While there may be significant photochemical reactions involving CO, evidence indicates that CO is a good indicator for automotive exhaust pollutants, which are known to play an important role in photochemical smog formation.

3. The time-rate of change of CO in the atmosphere as a measure of the degree to which the primary pollutants are being dispersed.

4. O_3 - The concentration of ozone in the early morning hours is indicative of the amount of photochemical activity.

5. Percent sunshine, used along with temperature, to represent radiation intensity.

6. Ventilation index computed at 0400 and 1600 hours daily. Low index values imply that the temperature inversions exist with low wind speeds that inhibit dispersion. High values indicate that the condition of the

atmosphere is more conducive to thorough dispersion.

7. The daily average surface wind.

The time of prediction is another variable that must be considered. If an accurate prediction is made in the early morning hours, such as at 4 a.m., then abatement action can be taken if necessary to reduce the amount of emissions. On the other hand, a later prediction, such as one at 9 a.m. is not as effective in curtailing the sources but can serve to warn the general public to modify their physical activity. The approach will be to optimize the predictor over a time period ranging from 2 a.m. to 10 a.m. in one hour increments. Therefore, the model is predicting from twelve to three hours in advance of the normal daily ozone peak.

Since the degree of photochemical smog is strongly dependent upon time of year, the specific model is optimized over a limited period of three months. Implied in this approach, is the fact that the time of year is itself one of the variables. In an attempt to hold this variable constant, the training set consists of August to October data. The model is then subsequently evaluated for September data. One further restriction is that only week days are considered in the analysis. The rationale is that week days are generally characterized by similar source emission patterns.

A special program was developed so that various combinations of input variables could be tried in rapid succession. Results of the application of this program showed that no single parameter exhibited a pronounced effect on the classification accuracy.

The next step involved eliminating groups of variables. The best results correspond to the case where all NO_2 inputs were deleted. The prediction distribution for 9 a.m. with all NO_2 inputs deleted, is shown in Table 4.

It should be emphasized that the results of Table 4 were obtained on a development sample, the accuracy of the meteorological variables being 100%. When used on independent data (September), prediction accuracies between 65% and 95% were obtained, depending on the hour of prediction. The 7 a.m. prediction, perturbed by the high pollutant peaks during the morning rush hour, exhibits the lowest accuracy. The total number of cases - 7 days with $\geq$ .1 ppm O_3, 11 days below this limit - does not allow a significant statistical analysis of the results.

If this is what the utmost sophistication and resort to a computer can accomplish, one should perhaps look at the other end of the scale for a set of predictors as simple as possible (Benarie, 1971).

As an episode severity criterion any 24-hour concentration larger by a factor of at least 2.0 than the running average over the 30 previous days, was selected. The 30-day running average was considered as a fair approximation of the actual seasonal average.

An essential condition in defining the predictors was their general availability through broadcast, i.e., not limited to the users connected by teletype to the National Meteorological Service. This condition limits not only the number and kind of predictors, but also the time of their availability. The latter if determined by the radio or the press, is several hours behind the information dispatched to the teletype user. Finally, to establish an air pollution forecast it should also be meaningful to the non-meteorologist.

The rationale behind the choice of the main predictor was that 24-hour calm inversion conditions are seldom conducive to pollution episodes. According to the theory of BOUMAN and SCHMIDT (1961), confirmed by the Dec. 1952 episode of London, England and the Dec. 1959 episode of Rotterdam, Netherlands, and by LAWRENCE's (1967) analysis of several London episodes and KOLAR's (1967) discussion of several others, concentrations increase proportionally with time at the beginning of an episode, and grow proportionally to the square root of time afterwards. During the first 24-hour period of calm winds, twice the value of the running seasonal (previous 30-days) mean is seldom attained, although it occurs with high probability if calm conditions persist for a second day or more. In this way, the predictor set becomes of utmost simplicity:

1. The observation of an elapsed period of 24-hours during which the mean ground level wind velocity was less than 3.0 m s^{-1}.

2. A forecast of a similar situation for tomorrow.

This simple rule was checked at Rouen, France, during two additional winters (BENARIE and MENARD, 1972) from October to March, i.e. 540 forecasts, and is shown in Table 5. It appears from this Table that of the average 13 episode-days per winter, on the average 8 are correct even when the occasions of incorrect meteorological forecasts are included. If the latter are excluded in order to reflect more closely the real merits of the method, then out of 28 episode-days only 3 were incorrect, that is around 10%.

Table 5 shows a total of 18 false alerts, of which roughly one half are due to incorrect meteorological forecasts, while the other half are due to the prediction method itself. Research is in progress to determine predictors which will optimize the ratio between the two kinds of errors. Obviously a "broad predictor" would hit all the real episode-days, although additionally set off a great number of false alerts; a "sharp-predictor" would avoid false alerts, but miss a number of real episode-days. But as things now stand, meteorological error is the cause of 17 incorrect episode forecasts, while methodological ("choice of predictor") error is responsible for only 14 failures. Until meteorology becomes a much more accurate science, the search for more accurate predictors would quickly become one of steadily diminishing returns.

This situation appears even more clearly, if Table 5 is re-interpreted as a contingency table, that contains not only the (necessarily restricted) number of episode-days, but the whole forecast period. Table 6 is such a display for 523 days when no meteorological error occurred, indicating the skill score of the predictor choice. Table 7 contains the whole period of 540 days and provides the skill score for the effectiveness of the forecasting program when the meteorological error is included.

The forecast method was further tested during the 1970-71 and 1971-72 winters, in Strasbourg, France (BENARIE, unpublished) and produced the same skill score. Its application to other sites seems possible, since the forecast criterion is not the absolute value of wind velocity or some other locally influenced value, but the duration of the stagnation spell.

Furthermore, a semi-quantitative relation can be found between the duration of the calm and the concentration increase. For the winter season, at Rouen, France, the following concentration factors were obtained by regression analysis of the 1968/69 data (= development sample): Table 8. In each case these factors multiply the 30-day running average concentration recorded at the same station.

The factors relate to surface wind and its duration and thus may be suitable for general application: they were foreseen in the theory of BOUMAN and SCHMIDT (1961) and confirmed by the Rouen data; they also check well with observations from other cities (LAWRENCE, 1967; KOLAR, 1969). BRINGFELT (1971) found in Stockholm, Sweden, that during stagnation periods lasting 3-5 days, the average SO_2-level becomes 2.3 times the winter mean. The factors figuring in Table 8. were further checked and found adequate for two more winter seasons in Rouen and Strasbourg, France. The temperature factor is less general, as it is an emission factor that is related to climate and space heating habits. These factors which depend on wind direction are completely local and linked to the source-receptor configuration (Rouen).

The factors provided by Table 8 were further utilized to obtain the results shown in Table 9 for the 12 forecast episode-days of the test-set represented by the 1968/69 winter season. Four of the 12 forecast episode days did not materialize.

Meteorological forecasts not being sufficiently detailed to justify the breakdown into factors on the day before the episode forecast, an average multiplying factor of 2.0 was applied in each case.

The RMS error of the values using the actual meteorological data observed during the episode-days is 82 $\mu g\ m^{-3}$; the mean for 7 stations and 12 forecast episode-days is 187 $\mu g\ m^{-3}$, i.e. a relative RMSE of 0.43. For the true forecast values (third line in Table 9), the RMS error is 87 $\mu g\ m^{-3}$, corresponding to a relative RMSE of 0.46. These figures are comparable and rather below those generally attained by other mathematical models in the more favorable case of day-to-day pollution calculations. It should be kept in mind, that most models break down just in the stagnation-episode conditions where this simple method seems applicable.

Continuing to test two more winters (BENARIE and MENARD, 1972) which were outside the development sample, the relative RMSE obtained was 0.66, somewhat larger than previously, but still below that for some elaborate forecasting systems.

SECTION 4

THE BOX MODEL

SHORT-TERM AVERAGES

This concept is much too well known to require discussion of the details which will already be familiar to most modelers. Only a few general ideas and some results will be mentioned. A very thorough discussion of the box model was given by LETTAU (1970).

1. Firstly we remember that the box model may be considered as derived from the idea of the continuity of mass of a volume element, as used in the advective transport equation. If the volume element becomes large enough so as to include the whole urban area, or at least a major part of it, and if diffusion can be neglected, then we are concerned with the so-called box model represented by the simple formula:

$$\chi = c\,Q_A/u \qquad (1)$$

where Q_A is the source strength per area unit and u the local wind speed.

This appraoch almost coincides with the intuitive idea which is to assume that pollution coming from an area source is completely mixed within a box, which has its base at the ground, and its top at the limit of vertical mixing L; in this case $L^{-1} = c$ (SMITH, 1961). Here we have the basic form of the box model. The product uL is the ventilation rate, i.e., the flushing rate per unit width of the box. In fact, the general idea of using a simple proportionality between emissions and concentrations goes back at least to the Leicester survey (UNITED KINGDOM, 1945). SHELEIKOVSKII (1949, p. 97) also derived an equation of the form of Eq. (1). He took as the source strength for particulates emitted by domestic space heating, the product of the population density by an emission factor. Thus his concept belongs to the long-term box models.

2. Another way of considering the problem was developed by GIFFORD (1970, 1972, 1973), GIFFORD and HANNA (1970, 1973), and HANNA (1971, 1973a).

It involves the integration in the upwind direction of a cross-wind infinite line-source diffusion formula. GIFFORD and HANNA used the Gaussian version, but other formulations, such as those based on Lagrangian similarity theory, could be used just as well. The results tend to be not very sensitive to the particular diffusion model employed, since according to the simplifications that are made, only vertical diffusion is involved. The usual simple power law

$$\sigma_z = Bx^b \tag{2}$$

where σ_z is the standard deviation, x is the downwind distance, and B and b constants, is used to represent the standard deviation of the concentration distribution in the vertical. The receptor point is assumed to be located at the center of a source square. The lateral dispersion is neglected, so that the concentration at a receptor at any time can only be influenced by the (sum) of the upwind sources (GIFFORD 1959, CALDER 1969: "narrow plume hypothesis").

Based on this line of reasoning, GIFFORD and HANNA concluded that the area-source component of a stable, non-reacting pollutant species would be adequately described by the following formula:

$$\chi = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{(\Delta x/2)^{1-b}}{B(1-b)u} \left\{ Q_0 + \sum_{i=1}^{N} Q_A \left[(2i+1)^{1-b} - (2i-1)^{1-b} \right] \right\} \tag{3}$$

where χ is the pollutant concentration at ground level, u is the mean wind speed, x is the source inventory grid spacing, and Q_A the source strengths in the (N + 1) upwind source boxes, i = 0, 1, ..., N. The total ambient air quality then follows by combining the contributions from Eq. (3), together with the point-source contribution and Q_o , the background concentration. Eq. (3) is closely related to several area source formulas based on the Gaussian model, particularly the study by CLARKE (1964).

In fact, Eq. (3) actually takes into account N advective steps, and so by itself does not correspond to the simple box, but rather to the multiple box category. At present we are concerned with the application of Eq. (3) to a series of upwind boxes. However, if the wind direction changes, 1...N may also be interpreted as the contribution of the 1...Nth direction multiplied by its class frequency.

On a statistical basis and without considering what happens upwind (here we depart from the source oriented, deductive argument) HANNA (1971) concluded that Eq. (1) is a good approximation to Eq. (3) with:

$$c = \left(\frac{2}{\pi}\right)^{1/2} \left[(2N+1)\,\Delta x/2\right]^{1-b} \cdot \left[B(1-b)\right]^{-1} \tag{4}$$

Equation (1) may be considered as a box model whose lid height increases downwind, according to Eq. (4). Since the quantity (1 - b) is quite small and the product B(1 - b) only varies slowly in the stability range ordinarily encountered over cities, the assumption that C = constant for a given stability condition is quite reasonable. According to GIFFORD and HANNA, C can be assigned the approximate values 50, 200 and 600 for unstable, neutral and stable conditions. Since $\sigma_z = Bx^b$ in Eq. (2), it follows that

$$c \approx x/\sigma_z \tag{5}$$

It should also be pointed out that by combining Eqs. (1) and (5) we obtain

$$\chi/Q_a = (1/\sigma_z)\,(x/u) \tag{6}$$

The quantity u/x is essentially what LETTAU (1970) terms the "flushing frequency" in his exposition of the role of the box model of urban diffusion.*

3. The statistical argument put forward by GIFFORD and HANNA - a third way to come to the box model - changes the nature of Eq. (1) from source-oriented into receptor-oriented, and at the same time the formula becomes a description rather than a deduction. Nevertheless, it is included in this section so as not to disrupt the discussion of the box model.

4. MILLER and HOLZWORTH (1967), HOLZWORTH (1972) also treated the city source as a continuous series of infinitely long cross-wind sources. Vertical concentrations follow a Gaussian distribution and average unstable conditions are assumed. The normalized concentration χ/Q_A in these circumstances is given by HOLZWORTH as

$$\chi/Q_A = 4.0\,(x/u)^{0.115} \tag{7}$$

for $x/u \geq 0.47\, L^{1.13}$ (L = mixing height) and if no pollutants achieve uniform vertical distribution. However,

$$\chi/Q_A = 3.61\, z^{0.13} + \frac{x}{2Lu} - \frac{0.088\, uz^{1.26}}{L} \tag{8}$$

*See addendum

for $x/u \geqslant 0.47$ and if some pollutants achieve uniform vertical distribution. In most cases, the term with 0.088 coefficient is very small and can be neglected.

HOLZWORTH (1972) presents tabulated values of χ/Q_A as a function of L, u and x (= + the distance the wind travels across the city), for the two city sizes of 10 and 100 km. The smaller the values of L and u, and the larger the value of x, the smaller are the relative difference between the χ/Q_A values from HOLZWORTH's model and those from the box model. Thus HOLZWORTH's approach may be considered as a fourth way to converge towards the box model. The approximation depends here also (as with GIFFORD and HANNA's deduction) on the vertical mixing.

5. Finally, we may note that a test of the theoretical approaches will depend on the empirical determination of

a. the vertical pollutant profiles and

b. the horizontal mass balance of the box.

These were experimentally checked by HALPERN et al (1971); sulfur dioxide concentrations were obtained from helicopter soundings in the New York City area and vertical wind profiles by using pilot balloon observations. The data of HALPERN et al confirm that the mass balance expressed by Eq. 1 between emissions and ambient concentrations is a fair approximation in urban areas.

Also, we may note the calculation of CO-concentrations by HANNA (1973a) and their comparison with the observations at eight stations in the Los Angeles basin, for the period 5 a.m. - 4 p.m., on Sept 29, 1969. These are presented in Table 10. The concentrations are set initially equal to the 5 a.m. observed concentrations.

The correlations at each station are of the same order as for other more sophisticated models. The overall correlation coefficient from 6 a.m. to 4 p.m. for all eight stations, is 0.43. Obviously, in the computation of correlation coefficient the 5 a.m. figures -- where identity of the calculated and observed concentrations was assumed -- were omitted. The correlation at any given hour over the geographical extent of the Los Angeles basin is uncertain. It seems that in this case, with eight "boxes" centered on the eight stations, the theoretical basis of the box model has been stretched too far. In the first place, the single box model is not meant to provide spatial resolution and in the second, when a multi-box is used, as here, advection must also play a role. It seems that we have here an illustration of the principle, that lengthening the deductive chain by introducing more sophistication (Here in the quest for spatial resolution) does not necessarily improve the results, because at the same time we introduce an increased amount of noise. When hourly means are taken at each hour over the entire basin (last column of Table 10), it is seen that the calculated concentrations are mostly too high, by a factor of two or even greater. The correlation in this column is 0.74, significant at the 1% level.

Eq. 1 has been validated by HANNA (1971), and GIFFORD (1972, 1973) on short-time concentration values for several urban areas. Table 11 is a summary of these validation results. For the comparison of the box model performance with other models, the references should be consulted. It is a personal opinion of the author's, that only models with similar amounts of detail, as regards the input and the output, should be compared (BENARIE, 1975).

The discussion, as so often happens between protagonists of (simple) box models and others pleading for more elaborate ones yielding a finer space-time resolution, somehow misses its point. An analogy with road maps may perhaps help to clarify the situation. A general map, say of 1 : 1,000.000 scale will give the distances among various points almost as exactly the ten sheets of 1 : 100,000 maps pasted together. This is because the higher precision of the latter will probably be destroyed by matching errors. Thus the low resolution map may win, because it achieves the desired purpose (= distance measurement) by far more economical means. But when it comes to the intermediate resolution -- the 1 : 1,000.000 map has nothing comparable to offer. In the same way, it is not fair to compare calculations by simple box (or other space or time average) models with average values given by higher resolution models. If the user only needs averages then it would be a waste of money to search for high-resolution input and compute expensively detailed estimates, only then to lump them together. On the other hand, if the user asks for high-resolution data, even the best quality average estimates will not satisfy his needs.

Notwithstanding the attempts to mathematicize it on a Gaussian or other basis, the box model is really rooted in the statistical validations advanced by HANNA and GIFFORD. Only further testing and comparison with observed data can show just how "good" the concepts are.

HANNA (1973 b,c) proposed the extension of the simple box model to chemically reactive pollutants.

The model seems fairly successful in the prediction of hourly variations of CO, hydrocarbons and NO. However, it is inconsistent in its prediction of NO_2.

SIMPLE BOX MODEL FOR LONG-TERM AVERAGES

Integral Application of the Box Concept to an Entire Urban Area

It was pointed out by GIFFORD and HANNA (1973, where further references may be found) that if Eq. 1 is applied to yearly or seasonal, i.e., long-term averages, the estimates for the pollutant concentrations compare favorably with those obtained from other models. Writing Eq. 1 as

$$\chi = \underline{c}\,\frac{Q_{TOT}}{A\bar{u}} \qquad \text{-(1.a)}$$

where Q_{TOT} is the total yearly, seasonal . . . pollutant emission of the source

A is the area within which the pollutant is being emitted

$\bar{u}$ is the yearly, seasonal . . . average wind velocity; and using published average urban pollutant concentrations, GIFFORD and HANNA obtained Table 12.

The average value of $\underline{c}$ from the particle data is 202, and that for SO_2 is 50. The authors believe that a large part of this difference is caused by the fact that Q_{TOT} for SO_2 generally contains a much larger fraction of emissions from tall-stack "point" sources, such as steam-electric power plants, than does Q_{TOT} for particles. Eq. 1.a. is designed to account for point-sources only to the extent that they are low enough to be considered part of the distributed area-source component. If all the contributions from strong, elevated point-sources were removed, the two $\underline{c}$-values would probably be closer together. However, the necessary source data to make this correction is not included in the published reports.

While this reason, as advanced by GIFFORD and HANNA, certainly contributes to enhance the difference between the c-values, there may also be others. Thus SO_2-averages used do not seem to be representative. For example, it is not likely that the average SO_2 concentration in Detroit would be 16 ug m^{-3}. Even the averages listed in Table 12 for Denver, Buffalo, Kansas City, Milwaukee, seem below the usual rural averages. Another reason for the discrepancy may be the transformation of the SO_2 into sulfates, thus reducing the monitored SO_2-concentrations. It is possible that if all these input errors could be removed, Eq. 1.a. would perform even more generally.

GIFFORD and HANNA (1.c.) tested Eq. 1.a. on the particulate pollution of 15 additional U.S. cities, obtaining the same range of c values as previously. BENARIE (1975) used Eq. 1.a. in a reverse way, calculating from it the yearly average mixing height, since c can be interpreted as the ratio of the transport distance from the city's edge to the average mixing height. Very plausible values were obtained for SO_2, for the French cities of Rouen, Paris, Strasbourg, Lyon, Bordeaux and Marseille, as well as for the Japanese cities of Tokyo and Osaka.*

The relative success of the simple box model for long-time averages must be attributed at least partially to the fact that all box or cell concepts are based on the idea of the infinite vertical diffusivity inside the cell. Taking long-term averages, the diffusion time may be neglected when compared to the averaging time. Thus at least this requirement is satisfied.

Multiple Application of the Box Concept to an Urban Area

It was pointed out earlier that the box concept is the extension of a single volume element over an entire urban area. As such, it is not meant to provide resolution, and its specific advantages are linked to the properties of low resolution, when this is all that is required. GIFFORD and HANNA (1970) applied the box concept to a source pattern in the form of a 1 by 1 km grid. This is obviously a hybrid case lying between the multi-cell and box concept, and with the inherent difficulty of allowing for the influence of the upwind and neighboring boxes. GIFFORD and HANNA (1970) adopted a

*See Addendum

simple step-wise scheme for wind directions other than the cardinal ones, while THUILLIER (1973) in a pragmatic way made a subjective estimation of the relative contributions of the upwind boxes for each county-size box and each of the possible patterns. These relative contributions were then multiplied by the annual recurrence frequencies of the patterns to obtain a set of annual average contribution weighing factors.

A comparative (model-to-model) validation of this multi-box model (with some modifications) and some plume models, was made in a very relevant paper by TURNER, ZIMMERMAN and BUSSE (1972). A subsequent model-to-model and model-to-observed comparison was made by STROTT (1974) for Frankfurt, Germany. As with the previous model-to-model comparisons, the computed concentration isopleths exactly reflect the emission inventory for the area sources -- a rather obvious result, for which no modeling should be needed. Quantitatively, the multi-box model strongly overestimated the concentrations. It seems that the box model should not be overstretched to yield simplicity and resolution at the same time.

SECTION 5

CORRELATION WITH DEMOGRAPHIC PARAMETERS

The basic box model Eq. 1 requires proportionality between pollutant concentration and the source strength per area unit. If the wind speed is averaged over a whole year or even a longer period, this mean value is subject to only relatively slight changes. Hence Eq. 1 may be written:

$$\chi = \frac{c}{\underline{u}} Q_A = c' Q_A \qquad -(1.b)$$

where Q_A is the source strength per area unit

$\underline{u}$ is a long-term average of the wind speed.

Relation 1.b. was found graphically by PEMBERTON et al (1959) to hold between the smoke, in Sheffield, Engl., and the number of electors per acre in the district surrounding the sampling site. The authors considered as an index of population density the number of electors per acre. On the other hand, they assumed that domestic heating accounts for the major part of smoke pollution and hence, the source strength per area unit should be represented by the "elector per acre" index. It should be noted that although PEMBERTON et al did not express the graphical regression algebraically, the "box model" was proven in this way two years before its first mention by SMITH, in 1961. In the analogous graph for sulfur dioxide, PEMBERTON et al did not find a correlation. This may be explained by the fact that sulfur dioxide showed higher average concentration near heavily industrialized areas with low or moderate population density. Consequently, the population density in Sheffield was not a good index for sulfur dioxide source strength per area unit.

The situation is somewhat different in Paris, France, where the contribution of industrial sources to area source strength, is relatively smaller than for Sheffield. As PELLETIER (1967) pointed out -- Figs. 1. and 2. -- both smoke and sulfur dioxide show a strong correlation with population density. It may be noted that population density, as obtained from census figures, is just a first approximation to source strength, but it works out well within reasonable limits, inside comparable demographic entities.

Thus, PEMBERTON's and PELLETIER's linear regressions are not transposable directly to differently structured areas. The per capita pollutant emission may vary between wide limits, depending on the fuel use, nature of heating devices and other factors. Population density may have a different meaning in different countries, because census survey is done within administrative limits that are not well defined emissive entities at the time of origin.

PEMBERTON's and PELLETIER's approaches are regression methods. Instead of meteorological parameters an index of source strength per area unit was considered as an independent variable, linked to population density.

Another good predictor of pollutant concentrations, is the total population of the urban area. In the United States, concentrations of suspended particulate matter were associated with urban population and this is presented in Table 13. Although this table does not fit a linear regression, portions of it may be approximated linearly, as for example:

$$\chi\,(\mu gm^{-3}) = 45 \log P - 145 \qquad (9)$$

where log P is the common logarithm of the population.
This seems to represent fairly well the important 50,000 to 500,000 population range. Thus we may have an extremely cheap way to estimate, within a range of 50%, the yearly average of the particulate concentration of an urban area, provided that climatology, fuel use, traffic conditions and population density are about the same as in the Continental U.S. BOLIN et al (1971) reported a similar relationship to Eq. 9 for Swedish cities.

Statistical correlations for 23 cities, ranging in population from 100,000 to 2 million, were run by CARTER (1973) and by CARTER and NELSON (1973). Using the correlations between the pollutant emissions and the demographic factors of population, number of passenger vehicles registered by county, and the percentage of the work force employed in manufacturing, a linear modeling technique describes the future air pollution emissions of a city by size and the growth of the emissions.

SECTION 6

CONCENTRATION FREQUENCY DISTRIBUTION

THE LOG-NORMAL REPRESENTATION OF CONCENTRATIONS

The histogram of urban air pollutant concentrations sampled over any given time span (1 minute, 1 hour, 24 hours and so on) is quite skew. There are only a few near-zero values, but afterwards the frequency increases sharply, only to decrease again gradually towards the higher concentrations. A large number of skew distribution functions known in statistics can be fitted to such data: Poisson (WIPPERMAN, 1966), negative binomial (PRINZ and STRATMANN, 1966), Weibull (BARLOW, 1971), exponential (BARRY, 1971, SCRIVEN, 1971), gamma (= Pearson IV), beta (= Pearson I) and Pearson IV (LYNN, 1972). None of these has enjoyed the practical success and the wide acceptance of the lognormal distribution. POLLACK (1973, 1975) demonstrated that there is a fundamental similarity among these distributions utilized to fit air quality data.

As early as 1958, it was empirically shown that cumulative frequency distributions of suspended particulates at CAMP (urban) sites fit remarkably well a straight line when plotted on log-normal paper (U.S.D.H.E.W., 1958). Pronounced tendency towards log-normalcy of particulate concentrations was also observed by ZIMMER et al (1959) and by GOULD (1961). LARSEN (1961) extended this representation to carbon monoxide and ZIMMER and LARSEN (1965) to carbon monoxide, hydrocarbons, nitric oxide, nitrogen dioxide, oxidant and sulfur dioxide and to the main urban areas of the U.S.A. From this point on, the lognormal plotting gained almost exclusively amongst the graphical and functional representations of air pollution concentrations, and the number of papers and reports that make use of it is in the hundreds.

Considerable theoretical (GIFFORD, 1972; KNOX and PCLLACK, 1972; KAHN, 1973) and empirical (BENARIE, 1970) support exists for the lognormal distribution, as the most appropriate for characterizing both reactive and inert pollutant concentrations for a wide range of averaging times. These arguments were systematized and publicized by POLLACK (1973, 1975), although overwhelming acceptance of the log-normal representation was a fact long before theoretical proofs became available. Some of the reasons for this acceptance are:

1. The lognormal distribution is a relatively simple two-parameter distribution. Both parameters have easy-to-grasp physical meaning.

2. Convenient plotting paper and methods are available; the user does not have to resort to lengthy numerical calculations. The two parameters are easily read off the graphs.

3. The lognormal function has some mathematical properties (see AITCHINSON and BROWN, 1969) which make its use very easy. Standard statistical tests, mostly requiring a normal distribution of the population, may readily be applied after the logarithmic transformation which is automatically provided by the plotting.

Inhomogeneous source distribution around the measuring site may lead to deviations from the lognormal behavior (BENARIE, 1970). JOST et al (1974) have attributed this reason for the occasional departures from lognormalcy observed in Frankfurt, Germany.

Objections of a theoretical nature may be raised against the log-normal representation of pollutant concentrations (BARLOW, 1971; MILOKAY, 1972; MARCUS, 1972). These arguments mostly consider the extreme values, like zero values of the pollutant concentrations. It should be realized that such (theoretically important) concentrations are ordinarily below the sensitivity threshold of the measuring instruments, which therefore introduces a threshold parameter. The practical advantages of the log-normal representation are full justification for its wide-spread use in air pollution engineering.

AVERAGING-TIME ANALYSIS

LARSEN (1964), ZIMMER and LARSEN (1965), LARSEN et al (1967), LARSEN (1969, 1973, 1974) plotted by computer -- the first paper for a period of one year, and the last for up to a seven-year period -- the concentration frequencies as a function of averaging time for: carbon monoxide, hydrocarbons, nitric oxide, nitrogen oxide, nitrogen oxides ($NO + NO_2$), oxidants and sulfur dioxide for the CAMP sites in downtown Chicago, Cincinnati, Denver, Los Angeles, Philadelphia, St. Louis, San Francisco and Washington. These plots have been called "arrowhead diagrams" by STERN (1969). They may be characterized by the following properties:

1. Concentrations are lognormally distributed for all averaging times.

2. Plotted on a (log averaging time) - (log concentration) diagram, the points representing a given percentile (frequency) are aligned almost on a straight line.* Hence, at constant frequency, concentration is proportional to averaging time raised to a constant power.

3. The 30 percentile is close to the arithmetic mean concentration. The exponent (see above) is only a little different from zero, so that the 30 percentile and the arithmetic mean only vary slightly for all averaging times.

4. For the longest averaging time calculated (usually one year), the arithmetic mean, geometric mean, maximum concentration, and minimum concentration are all equal (and thus plot as a single point).

5. For averaging times of less than one month, maximum concentration is approximately inversely proportional to averaging time raised to an exponent. The maximum concentration is that corresponding to the 1/n frequency point (n = number of samples, e.g., 8760 hourly samples per year) on the linearly extrapolated cumulative diagram.

Potential reasons for characteristic 1 above were cited by BENARIE (1970), GIFFORD (1972), KNOX and POLLACK (1972) and KAHN (1973). Properties 2 and 3 are the most important experimental findings based on the analysis of the CAMP-results mentioned above. Property 4 is a necessary consequence of the averaging process. SINGAPURWALLA (1972) has cited possible reasons for the property 5. McGUIRE and NOLL (1971) verified the relationship between maximum concentration and averaging time, for five different air pollutants at 17 California sites in Los Angeles and San Francisco. The exponents are within the range observed by LARSEN.

---

* The f and the 1-f frequency loci are in fact asymptotes of parabolae. The vertices of these parabolae are located at the one-year arithmetic average point. The nearer to this point, the greater the deviation from a straight line (Personal communication of Dr. LARSEN).

The main consequence of this concept is to interrelate short- and long-averaging times in a descriptive, statistical, and receptor-oriented way. Therefore, in the systematics embodied by Table 2, the concentration frequency distribution model belongs under both the headings short- and long-term (however, receptor-oriented). But there may exist a possible source-oriented extension, as noted by STERN (1969): If we were able to separate the source factors subject to human control, from the weather factors beyond such control, we would be able to synthetize the distributions of air quality data that would result from the application of specific control strategies. We would also be able to compare them with air quality objectives, expressed in like format, to determine which strategy comes closest to effecting a match. The concentration versus averaging time and frequency diagram might have as its components the weather factors and the source factors. The analysis of the source factor arrowhead chart for its individual components would be the converse of the emission inventory approach, in that the latter seeks to arrive at the same result through synthesis, whereas the approach just outlined seeks to arrive at it through analysis. The two approaches should tend to check and reinforce each other, and thus improve our chances of determining the relative influence of various source categories across the averaging-time spectrum. This should give us useful leads to control strategies.

In the papers cited (mainly those of 1969, 1973 and 1974), LARSEN provides examples for interrelating air pollutant effects, air quality standards, air quality monitoring, diffusion calculations, source reduction calculations, and emission standards.

In the same papers, LARSEN published extensive tables -

1. interrelating the ratio of expected annual maximum pollutant concentrations to arithmetic mean concentrations for various averaging times and standard geometric deviations, and

2. the slope of the annual maximum concentration line for various standard geometric deviations of the one-hour frequency plot.

With the experimental arrowhead diagram at hand, the expected annual maximum, or the slope of the line linking it to various averaging times, or to any other parameter of interest, can be read off at least as easily as would be their readout from tables or numerical calculation.

SECTION 7

WIND AND CONCENTRATION RELATIONSHIPS

It is evident that pollutant concentration distributions are only the footprints of the windfield. At the same time, it has been observed that the logarithmic normal function is a convenient empirical representation of the wind velocity distribution (BENARIE, 1969). The fact that we are concerned at this point with a rough approximation appropriate to the argument that follows below -- without pretending to describe general physical properties of the wind -- was stressed in the Appendix and a subsequent discussion of a paper by BENARIE (1972).

Using numerical simulation, for area sources represented by n point sources, BENARIE (1971) obtained fair approximations to the log-normal distribution, provided that $n \geq 10$ and that the geometric means and standard geometric deviations of the component log-normal functions were randomly distributed. The conditions under which the sum of n lognormal variates is approximately lognormal for a limited number of variates in the sum, have been previously formulated by MITCHELL (1968). BENARIE's (1971) paper supports the empirical observations that pollutant concentration for all cities and for all averaging times is approximately lognormally distributed (Section 6) as a consequence of the (approximately) lognormal windfield, but the paper does not quantitatively link the lognormally distributed windspeeds to urban pollutant concentrations.

This link was accomplished in an elegant way by KNOX and LANGE (1974), who noted that the basic box model Eq. 1 suggests that frequency distribution of the wind speed determines the frequency distribution of the normalized concentration χ/Q_A where χ is the surface air concentration of the pollutant and Q_A is the unit area source strength provided that c (the proportionality constant) is nearly independent of frequency. In principle, for the cases of good frequency correlation between χ and U^{-1} at a given sampling station, the constant $c' = cQ_A$ can be found through graphic superposition of the observed and predicted distributions. In this case, no direct knowledge of the source strength Q_A is required.

The normalization constant c' for the 50 percentile point, is obtained from the superposition of c'/U distributions on the corresponding observed χ distribution curves. For the CO-values observed at the building of the San Francisco, California Bay Area Air Pollution Control District during the year 1966, c' was found to equal to 7.5 ppm m s^{-1} and with 1970 data, 7.4 ppm m s^{-1}. The average value c' = 7.45 was used as an experimental normalization factor in Fig. 3.

Instead of using observed χ values, c' may also be estimated by the values calculated on basis of the means of some model. For this purpose, KNOX and LANGE used McCRACKEN et al's (1971) multi-box model. Since characteristic times in this model are of the order of one hour (box dimension/ wind speed), the concentration values can be interpreted as hourly averaged values.

This model was used to predict the CO concentration for a 48-hour period, July 10-11, 1968. Fig. 4 shows the observed wind speed frequency distribution for this period as obtained by the U.S. Weather Bureau on top of the 11-story San Francisco Federal Office Building. This frequency distribution is quite closely lognormal. Fig. 5 shows the observed and the model predicted CO concentration frequency distributions. The predicted concentration distribution χ_{PR} has the same slope as the observed distribution χ_{OB}, and a geometric mean 20% above that observed. The normalization factor c', as derived from this model with the data from Figs. 4 and 5 is $c' = \chi_{PR}(50\%)U(50\%) = 1.6 \times 5.7 = 9.1$ ppm m/sec. Fig. 6 shows how the distribution curves shift when we use this Bay Area model derived normalization factor c' = 9.1 on the date of Fig. 3.

It is of interest to note, that an experimental value of c' could also be computed from the observed wind speeds and the observed concentrations χ_{OB} used as a basis for the 48-hour period in the Lawrence Livermore Laboratory air pollution model study. This factor obtained from Figs. 4 and 5 is $c' = \chi_{OB}(50\%)U(50\%) = 1.3 \times 5.7 = 7.4$ ppm m/sec, which agrees well with the experimental c' = 7.45 discussed previously for the 1966, 1970 annual distributions. The fact that the mean annual normalization constant can be determined so well by considering only a two-day period indicates that the 48-hour period may be sufficiently long to give a good average of the CO source variation in the city. This is not surprising, if one remembers that the main source of CO is the daily automobile traffic. In other words, to find a regional normalizing factor c' for an annual mean concentration frequency distribution, a model need only cover the longest basic time period of any time-dependent sources or sinks involved, just so long as this period is typical.

SECTION 8

VALIDATION - OR THE WAYS TO DELUDE ONESELF

It is the proper function of the statistician (and I am not one) to pronounce on the merits of chi square, skill score, correlation, RMS and absolute error, and a host of other measures of goodness-of-fit. I am sure that everybody knows all about the computation of these indices and so, in principle, nobody needs my advice about the fact that any one goodness-of-fit index may be misleading. Nevertheless, I cannot resist the temptation to illustrate this point by just one example.

Table 14, in its second column, shows the results of a model calculation, (LAMB, 1968, based on the concept of mass transport balance taking into consideration chemical reactions) although the nature of the model and the method of calculation does not concern us here. This model was taken as an example, since it is rather often quoted as a reference. Along with the calculated values, a "random" estimate (Column 4) and a "constant" (average) estimate (Column 5), are presented in Table 14. To obtain the random estimate, monotonically increasing values from 0 to 17 ppm were assigned, in alphabetical order to each station. As for the last column, values of 14 and 13 (to avoid fractional values as the true mean is 13.5 ppm) were assigned alternately. Incidentally, 13.5 ppm is not only the average of the first column but also a very likely average figure for many urban areas with automobile traffic anywhere in the world.

The entries that give the root mean square error (RMSE) are a caution against validation by just one statistical criterion. The model shows a <u>higher</u> RMSE than the (almost) random or the constant value guesses. The correlation coefficient entry rectifies this situation. The constant estimate -- a parallel line to the abscissa axis -- shows as expected, no correlation with the observed values. The model's correlation attains the 5% significance level for 11 degrees of freedom. However, even the random guess presents a correlation which could not be entirely rejected. It should be noted that this "guess" is not completely at random, but rather an educated guess, since the lowest and highest values are linked to some knowledge about the concentrations which might actually be observed.

This simple and somewhat superficial example can be generalized and provides a warning against some of the pitfalls. The lack of representativeness for any single goodness-of-fit index has already been mentioned. A second point is that pure chance can frequently produce a fit which is not too bad, provided that the series to be fitted is short and the span of the estimation limited. A third point, also linked to a limited span of possible values, is that judged by the RMS error, the mean is often a very good bet -- better than most calculations. Finally, no validation should be presented without a comparison with the random estimate (the skill score does just this).

SECTION 9

CONCLUSION

The guiding idea in the present paper is to recommend the use of models which correspond with the utmost parsimony to the end result which has to to attained. Thus, high resolution, sophisticated models should only be used when high resolution output information is really needed. For long-term, low-resolution purposes, we frequently have fully adequate, low-cost models.

It is not a safe procedure to obtain long-term, low-resolution information by integration from short-term, high-resolution estimates. As the chain of reasoning lengthens, unavoidable noise is being introduced at each step. The end result often is that the long-term estimate obtained in this fashion is less reliable than one obtained by some "computer-less" shortcut. Also, it has been shown that, for forecasting purposes, models that involve a very large number of modeling steps must perform less well than those involving simpler chains.

Finally, brief warning was given against validations based on a restricted number of narrow-span values and the use of a single goodness-of-fit index.

REFERENCES*

Aitchinson, J. and J.A.C. Brown, 1969. The lognormal distribution. Cambridge University Press, 176 pp.

Barlow, R.E., 1971. Average time and maxima for air pollution concentration. Univ. of California, Berkeley, Cal., Operations Res. Center Rept. ORC-71-17, NTIS AD-729-413.

Barry, P.J., 1971. Use of argon-41 to study the dispersion of stack effluents. Proc. of Symp. on Nuclear Techniques in Environmental Pollution, Internat. Atomic Energy Agency, Vienna, Austria, p. 241-253.

Benarie, M., 1969. Le calcul de la dose et de la nuisance du pollutant émis par une source ponctuelle (Text in French). Atm. Env. 3, p. 467-473.

Benarie, M., 1970. About the validity of log-normal distribution of pollutant concentrations (Text in French). Second Clean Air Congress, Dec. 6-11, Washington, D.C., Ed. H. M. Englund and W.T. Beery, Academic Press, New York 1971, p. 68-70.

Benarie, M., 1971. About the validity of the log-normal distribution of pollutant concentrations (Text in French). Proc. of the 2nd Internat. Clean Air Congress, Dec. 6-11, 1971, Washington, D.C., Academic Press, Ed. H.M. Englund and W.T. Beery, p. 68-70.

Benarie, M., 1971. Essai de prévision synoptique se la pollution par l'acidité forte dans la region rouennaise (Text in French). Atm. Env. 5, p. 313-326.

Benarie, M., 1972. The use of the relationship between wind velocity and ambient pollutant concentration distributions for the estimation of average concentrations from gross meteorological data. Proc. of the Symp. on Statistical Aspects of Air Quality Dat, Nov. 9-10, Chapel Hill, N.C., U.S.E.P.A., Research Triangle Park, N.C., EPA-650/4-74-038, p. 5-1 to 5-17.

Benarie, M., 1975. Modelling urban air pollution. Atm. Env. 9, p. 552-553, discussion to a paper of S. Hameed, Atm. Env. 1974, 8 p. 555-561.

Benarie, M., 1975. Calculation of the mean yearly mixing height over urban areas, from air pollution data. Sci. Tot. Env. 3, p. 253-265.

*See addendum

Benarie, M., and T. Menard, 1972. Vérification, pour les hivers 1969-1970 et 1970-1971 de la prévision de la pollution par l'acidité forte dans la région rouennaise (Text in French). Atm. Env. 6, p. 65-67.

Bolin, B. et al, 1971. Sweden's case study for the United Nations conference on the human environment. Royal Ministry for Foreign Affairs - Royal Ministry of Agriculture, 96 pp., see p. 21 and ref. 4 by Brosset C.

Bouman, D.J. and F.H. Schmidt, 1961. On the growth of ground concentration of atmospheric pollution in cities during stable atmospheric conditions. Beitr. Phys. Atm. 33, p. 215-224.

Bringfelt, B., 1971. Important factors for the sulfur dioxide concentration in Central Stockholm. Atm. Env. 5, p. 949-972.

Calder, K.L., 1969. A narrow plume simplification for multiple urban source models. Unpublished, ref. No. 20 in Gifford and Hanna (1970).

Carter, J.W., Jr., 1973. An urban air pollution prediction model based on demographic parameters. Thesis, Oklahoma Univ., 171 pp.

Carter, J.W., Jr. and R.Y. Nelson, 1973. An urban air pollution prediction model based on demographic parameters. Preprint, 66th Annual Meeting, Air Poll. Contr. Assoc., Chicago, Ill., June 24-28, 20 pp.

Clarke, J.F., 1964. A simple diffusion model for calculating point concentrations from multiple sources. J. Air Poll. Contr. Assoc. 14, p. 347-352.

Croke, E.J. and S.G. Booras, 1970. Design of an air pollution incident control plan. J. Air Poll. Contr. Assoc. 20, p. 129-138.

Croke, E.J. and J.J. Roberts, 1971. Chicago air pollution systems analysis program - Final Report. Argonne Nat. Lab., Argonne, Ill., ANL/ES-CC-009, 393 pp.

Egger, J., 1973. On the determination of an upper limit of atmospheric predictability. Tellus 25, p. 435-443.

Fortak, H.G., 1970. Numerical simulation of the temporal and spatial distribution of air pollution concentrations. Proc. Symp. Multiple-Source Urban Diffusion Models, Ed. A. Stern., U.S.E.P.A. AP-86, p. 9-1 to 9-33.

Gifford, F., 1959. Computation of pollution from several sources. Int. J. Air Poll. 2, p. 109.

Gifford, F.A., Jr., 1970. Atmospheric diffusion in an urban area. Paper presented at the 2nd IRPA Conf., Brighton, Engl., May 5, 5 pp.

Gifford, F.A. and S.R. Hanna, 1971. Urban Air pollution modelling. Second Clean Air Congress, Dec. 6-11, Washington, D.C., Ed. H.M. Englund and W.T. Beery, Academic Press, N.Y., p. 1146-1151.

Gifford, F.A., 1972. The form of the frequency distribution of air pollution concentrations. Proc. of the Symp. on Statistical Aspects of Air Quality Data, Nov. 9-10, Chapel Hill, N.C., U.S.E.P.A., Research Triangle Park, N.C., EPA-650/4-74-038, p. 3-1 to 3-7.

Gifford, F.A., 1972. Applications of a simple urban pollution model. Proc. Conf. on Urban Environment and Second Conf. on Biometeorology, Philadelphia, Pa., Oct. 31-Nov. 2, p. 62-63.

Gifford, F.A., 1973. The simple ATDL urban air pollution model. Paper presented at the 4th Meeting of NATO/CCMS Panel on Modeling, Oberursel, Germany, May 28-30, p. XVI-1 to XVI-18.

Gifford, F.A. and S.R. Hanna, 1973. Modeling urban air pollution. Atm. Env. 7, p. 131-136.

Gifford, F.A. and S.R. Hanna, 1975. Modeling urban air pollution. Atm. Env. 9, p. 267-275, discussion to Hameed S. (1974). Atm. Env. 8, p. 555-561.

Gould, G., 1961. The statistical analysis and interpretation of dustfall data. Preprint. Proc. 54th Annual Meeting Air Pollution Contr. Assoc., New York, N.Y.

Halpern, P., C. Simon and L. Randall, 1971. Source emission and the vertically integrated mass flux of sulfur dioxide across New York City Area. J. Appl. Meteor. 10, p. 715-724.

Hanna, S.R., 1971. A simple method of calculating dispersion from urban area sources. J. Air Poll. Contr. Assoc. 21, p. 774-777.

Hanna, S.R., 1973a. Urban air pollution models--why? Paper presented at the Nordic Symp. on Urban Air Pollution Modeling, Oct. 3-5, Vedbaek, Denmark, 19 pp.

Hanna, S.R., 1973b. Application of a simple model of photochemical smog. Proc. of the 3rd Clean Air Congr. Dusseldorf, Germany, Oct. 8-12, VDI-Verlag, Dusseldorf, p. B72 to B74.

Hanna, S.R., 1973c. A simple dispersion model for the analysis of chemically reactive pollutants. Atm. Env. 7, p. 803-817.

Holzworth, G.C., 1972. Mixing heights, wind speeds, and potential for urban air pollution throughout the contiguous United States. U.S.E.P.A., AP-101, 118 pp.

Jost, D., R. Kaller, H. Markush and W. Rudolf, 1974. Analysis of six years continuous air pollution surveillance. In: Automatic Air Quality Monitoring Systems, Ed. T. Schneider, Elsevier, Amsterdam, p. 251-260.

Kahn, H.D., 1973. Note on the distribution of air pollutants. J. Air Pollut. Contr. Assoc. 23, p. 973.

Kao, S.K. and A. Al-Gain, 1968. Large-scale dispersion of clusters of particles in the atmosphere. J. Atm. Sc. 25, 214-221.

Kao, S.K. and D. Powell, 1969. Large-scale dispersion of clusters of particles in the atmosphere. II. Stratosphere. J. Atm. Sc. 26, 734-740.

Kolar, J., 1969. The increase in the SO_2-concentration during long-term weather situations with poor diffusion. Staub (Engl.) 29, No. 12, p. 32-35.

Knox, J.B. and R. Lange, 1974. Surface air pollutant concentration frequency distributions: implications for urban modeling. J. Air Pollut. Control Assoc. 24, p. 48-53.

Knox, J.B. and R.I. Pollack, 1972. An investigation of the frequency distributions of surface air-pollutant concentrations. Proc. of the Symp. on Statistical Aspects of Air Quality Data, Nov. 9-10, Chapel Hill, N.C., U.S.E.P.A., Research Triangle Park, N.C. EPA-650/4-74-038, p. 9-10 to 9-17.

Lamb, R.G., 1968. An air pollution model for Los Angeles. M.S. Thesis. Univ. of California, Los Angeles, Cal.

Larcheveque, 1972. Turbulent dispersion--EOLE experiment. COSPAR XV, Madrid.

Larsen, R.I., 1961. A method for determining source reduction required to meet air quality standards. J. Air Poll. Contr. Assoc. 11, p. 71-76.

Larsen, R.I., 1964. United States Air Quality. Arch. Env. Health 8, p. 325-333.

Larsen, R.I., 1969. A new mathematical model of air pollutant concentration, averaging time and frequency. J. Air Pollut. Contr. Assoc. 19, p. 24-30.

Larsen, R.I., 1973. An air quality data analysis system for interrelating effects, standards and needed source reduction. J. Air Pollut. Contr. Assoc. 23, p. 933-940.

Larsen, R.I., 1974. An air quality data analysis system for interrelating effects, standards and needed source reductions - Part 2. J. Air Pollut. Contr. Assoc. 24, p. 551-558.

Larsen, R.I., C.E. Zimmer, D.A. Lynn and K.G. Blemel, 1967. Analyzing air pollutant concentration and dosage data. J. Air Poll. Contr. Assoc. 17, p. 85-93.

Lawrence, E.N., 1967. Atmospheric pollution during spells of low-level air temperature inversion. Atm. Env. 1, p. 561-576.

Lettau, H.H., 1970. Physical and meteorological basis for mathematical models of urban diffusion processes. Proc. Symp. on Multiple-Source Urban Diffusion Models, Ed. A.C. Stern, U.S.E.P.A. AP-86, p. 2-1 to 2-26.

Lynn, D.A., 1972. Fitting curves to urban suspended particulate data. Proc. of the Symp. on Statistical Aspects of Air Quality Data, Nov. 9-10, Chapel Hill, N.C., U.S.E.P.A., Research Triangle Park, N.C., EPA-650/4-74-038, p. 13-1 to 13-28.

McCracken, M.C., T.V. Crawford, K.R. Peterson and B. Knox, 1971. Development of a multi-box air pollution model and initial verification for the San Francisco Bay Area. Lawrence Livermore Lab. - Univ. of California, UCRL-733 48, 96 pp.

McGuire, T., and K.E. Noll, 1971. Relationship between concentrations of atmospheric pollutants and averaging time. Atm. Env. 5, p. 291-298.

Mahoney, J.R. and B.A. Egan, 1971. A mesoscale numerical model of atmospheric transport phenomena in urban areas. Second Internat. Clean Air Congress, Dec. 6-11, Washington, D.C., Ed. H.M. Englund and W.T. Beery, Academic Press, p. 1152-1157.

Marcus, A.H., 1972. A stochastic model for estimating pollutant exposure by means of air quality data. Proc. of the Symp. on Statistical Aspects of Air Quality Data, Nov. 9-10, Chapel Hill, N.C., U.S.E.P.A., Research Triangle Park, N.C., EPA-650/4-74-038, p. 7-1 to 7-15.

Miller, M.E. and G.C. Holzworth, 1967. An atmospheric diffusion model for metropolitan areas. J. Air Poll. Contr. Assoc. 17, p. 46-50.

Milokay, P.G., 1972. Environmental applications of the Weibull distribution function: oil pollution. Science 176, p. 1019-1021.

Mitchell, R.L., 1968. Permanence of the lognormal distribution. J. Opt. Soc. Am. 58, p. 1267-1272.

Morel, P., 1970. Satellite techniques for automatic platform location and data relay. COSPAR XV, Madrid.

Moses, H., 1969. Mathematical urban air pollution models. Argonne Nat. Lab., Argonne, Ill., ANL/ES/RPY-001, 69 pp.

Moses, H., 1970. Tabulation techniques. Proc. of the Symp. on Multiple-Source Urban Diffusion Models, Research Triangle Park, N.C., Ed. A. Stern, U.S.E.P.A. - A.P.C.O. Publication No. AP-86, p. 14-13 to 14-15.

N.A.P.C.A., Air Pollution Control Administration 1969: Air quality criteria for particulate matter. Publication No. AP-49, 211 pp.

Ott, W.R. and G.C. Thom, 1976. Air Pollution index systems in the United States and Canada. J. Air Poll. Contr. Assoc. 26, p. 460-470.

Pelletier, J., 1967. Enquêtes de pollution atmosphérique dans l'environment. (Text in French) Poll. atm. 36, p. 240-252.

Pemberton, J., M. Clifton, J.K. Donoghue, D. Kerridge and W. Moulds, 1959. The spatial distribution of air pollution in Sheffield, 1957-1959. Int. J. Air Poll. 2, p. 175-187.

Pollack, R.I., 1973. Studies of pollutant concentration frequency distributions. Thesis, Univ. of California, Livermore, Cal., 82 pp.

Pollack, R.I., 1975. Studies of pollutant concentration frequency distributions. U.S.E.P.A., Research Triangle Park, N.C., Report EPA-650/4-75-004, 82 pp. This paper is a reprint of the previous reference.

Prinz, B. and H. Stratman, 1966. The statistics of propagation conditions in the light of continuous concentration measurements of gaseous pollutants. Staub (Engl.) 26, p. 4-12.

Ruff, R.E., 1974. Application of adaptive pattern classification to the derivation of relationship between air quality data. In: "Automatic Air Monitoring Systems", Ed. T. Schneider, Elsevier Amsterdam, p. 145-166.

Scriven, R.A., 1971. Use of argon-41 to study the dispersion of stack effluents. Proc. of the Symp. on Nuclear Techniques in Environmental Pollution. Internat. Atomic Energy Agency, Vienna, Austria, p. 254-255.

Sheleikhovskii, G.V., 1949. Smoke pollution of towns. Translation by Israel Program for Scientific Trans. for U.S. NSF and U.S. DOC (1961), 203 pp.

Singapurwalla, N.D., 1972. Extreme values from a lognormal law with applications to air pollution problems. Technometrics 14, p. 703.

Smith, M.E., 1961. The concentration and residence time of pollutants in the atmosphere. Intern. Symp. for Chem. Reactions of the Lower and Upper Atmosphere. San Francisco, Stanford Res. Inst. Advance Papers, p. 273-286.

Stern, A.C., 1969. The systems approach to air pollution control. Proc. of the Clean Air Conf. on the Clean Air Soc. of Australia and New Zealand, Vol. 2; p. 2.4.1 to 2.4.22.

Strott, J.K., 1974. Application of the AQDM model and the ATDL model and the comparison of the results. Fifth Meeting NATO/CCMS Expert Panel on Air Pollution Modeling, 4-6 June, Roskilde, Denmark, p. 12-1 to 12-19.

Thuillier, R.H., 1973. A regional air pollution modeling system for practical application in land use planning studies. Preprint, Bay Area Air Poll. Control District, San Francisco, Cal., 25 pp.

Turner, D.B., J.R. Zimmerman, A.D. Busse, 1972. An evaluation of some climatological dispersion models. Proc. 3rd Meeting of the Expert Panel on Air Poll. Modeling, NATO/CCMS, Paris, France, Oct. 2-3, p. VIII-1 to VIII-25.

United Kingdom, 1945. Atmospheric pollution in Leicester. Dept. Sci. and Industr. Res., Techn. Paper No. 1, 161 pp.

U.S.D.H.E.W., 1958. Air pollution measurements of the National Air Sampling Network - Analyses of suspended particulates, 1953-57. PHS Publication No. 637, p. 245.

Wipperman, F., 1966. On the distribution of concentration fluctuations of a harmful gas propagating in the atmosphere (unpublished MS) 17 pp.

Zimmer, C.E. and R.I. Larsen, 1965. Calculating air quality and its control. J. Air Poll. Contr. Assoc. 15, p. 565-572.

Zimmer, C.E., E.C. Tabor and A.C. Stern, 1959. Particulate pollutants in the air of the United States. J. Air Poll. Contr. Assoc. 9, p. 136-140.

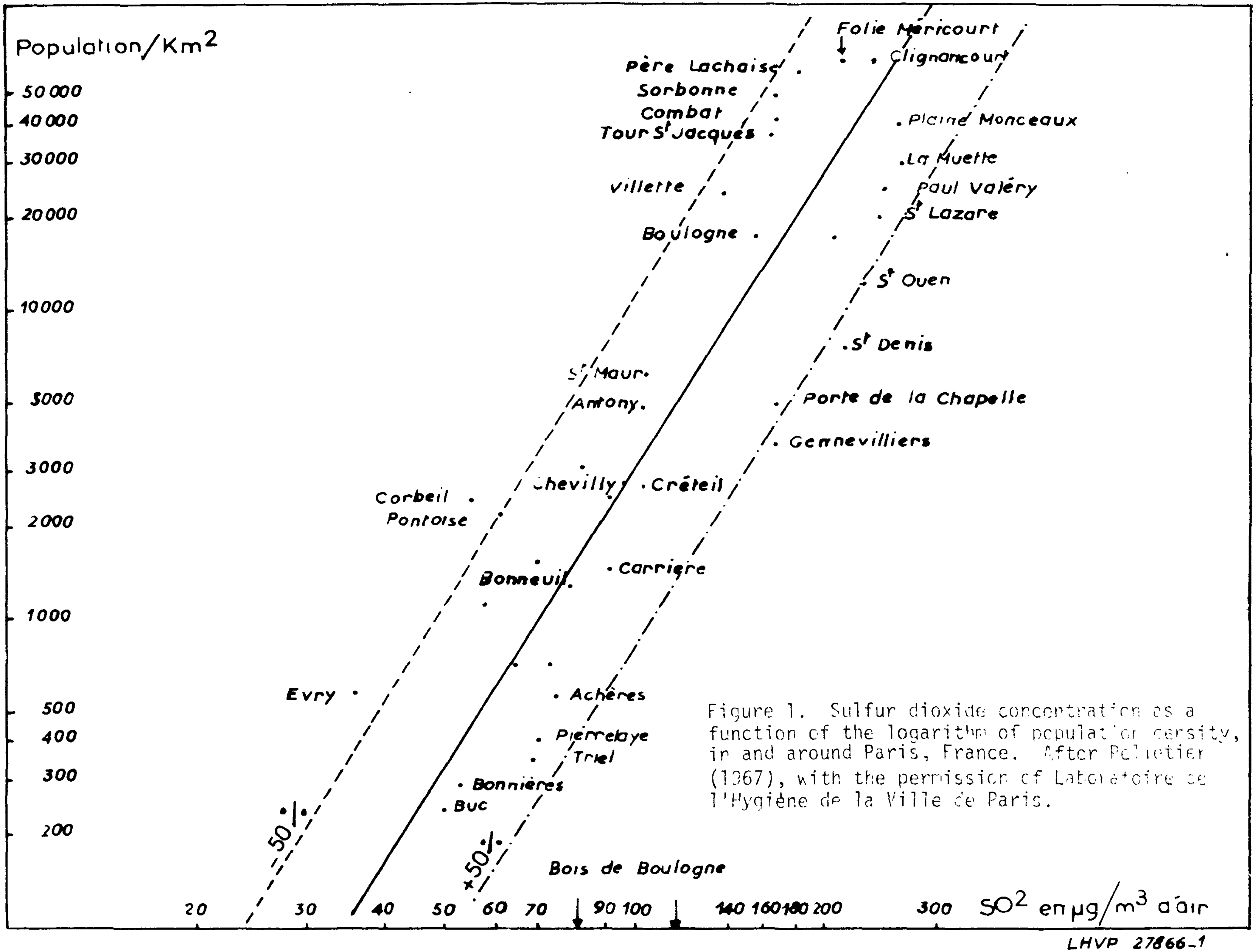

Figure 1. Sulfur dioxide concentration as a function of the logarithm of population density, in and around Paris, France. After Pelletier (1967), with the permission of Laboratoire de l'Hygiène de la Ville de Paris.

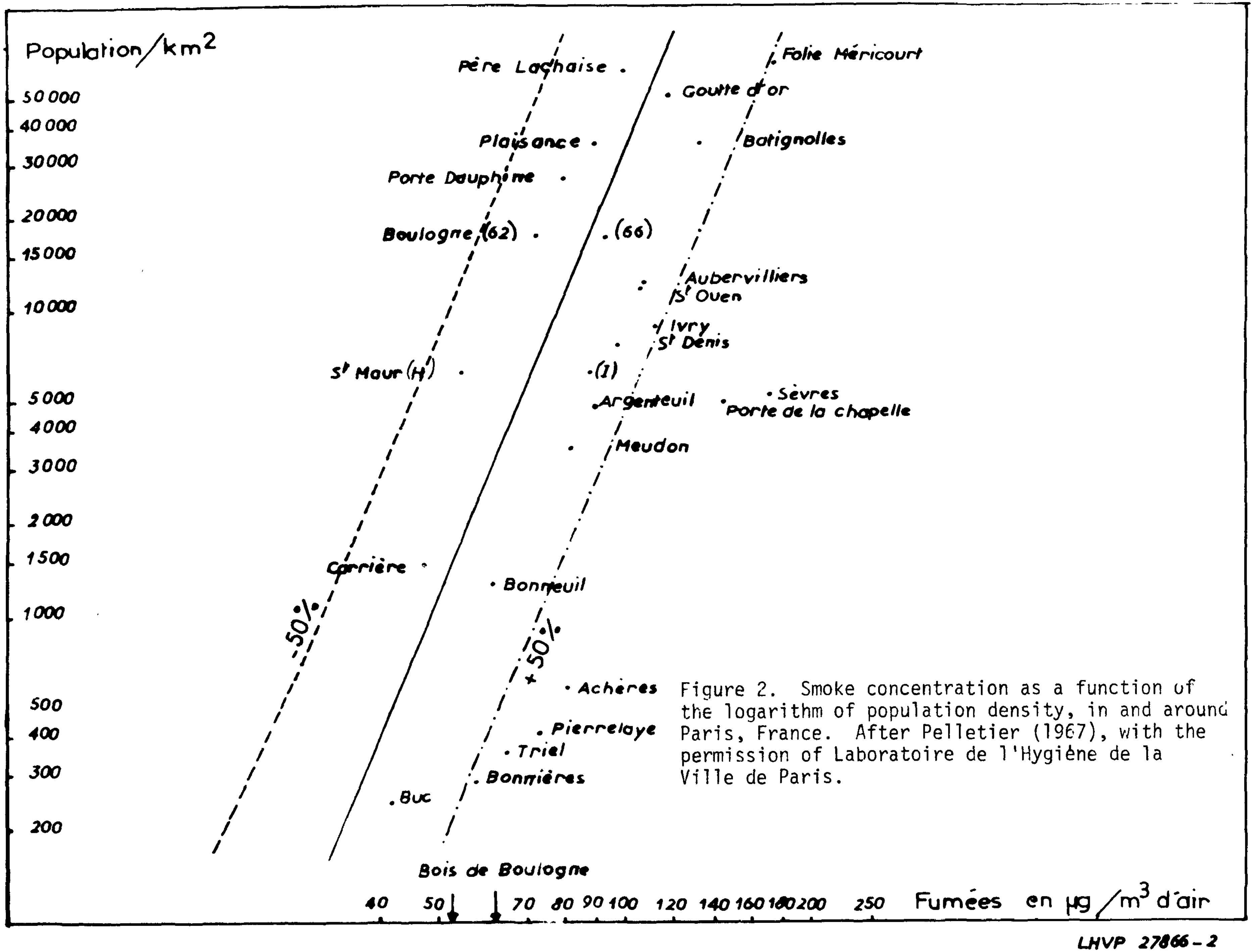

Figure 2. Smoke concentration as a function of the logarithm of population density, in and around Paris, France. After Pelletier (1967), with the permission of Laboratoire de l'Hygiène de la Ville de Paris.

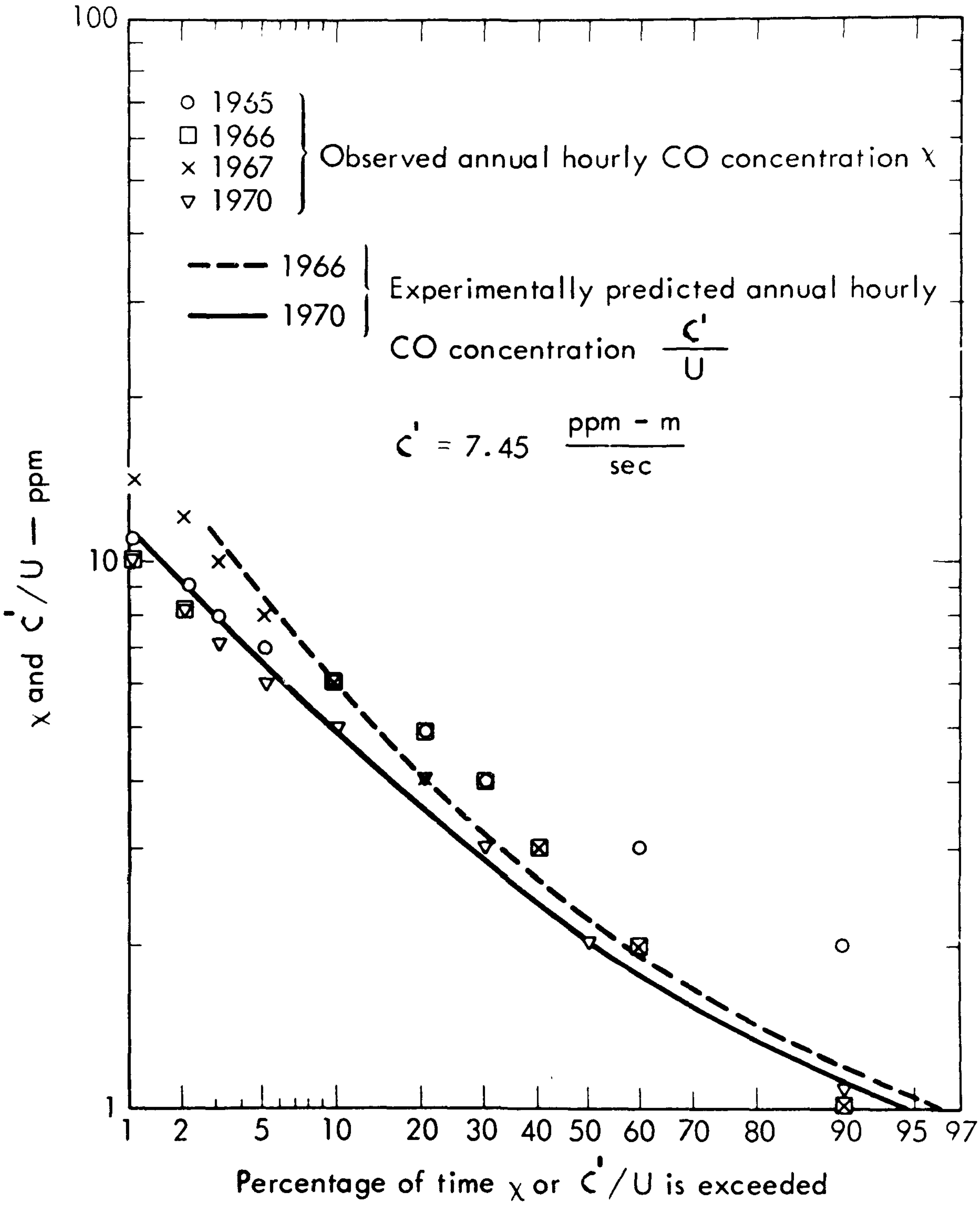

Figure 3. Observed and experimentally predicted annual hourly CO concentration distribution for San Francisco.

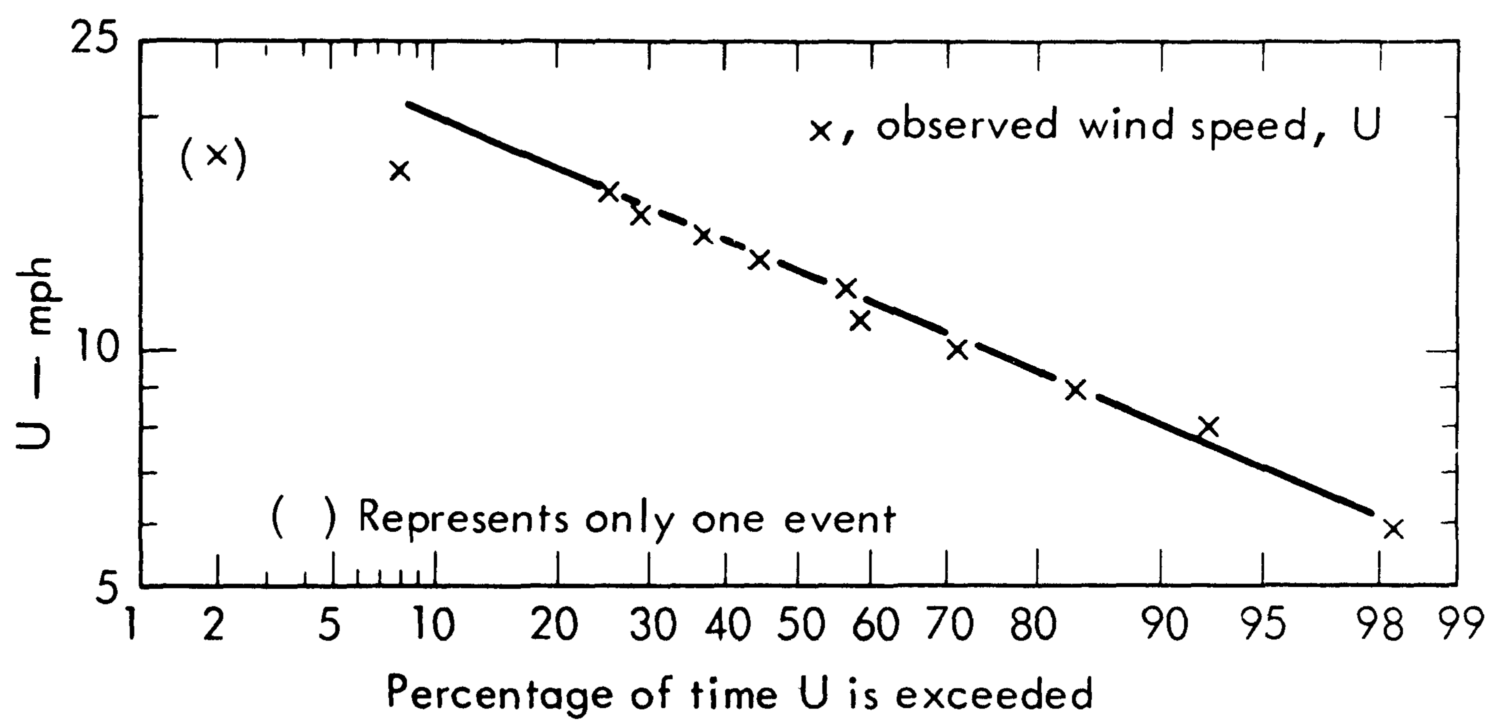

Figure 4. Observed hourly wind-speed distribution for San Francisco Federal building, July 10–11, 1968.

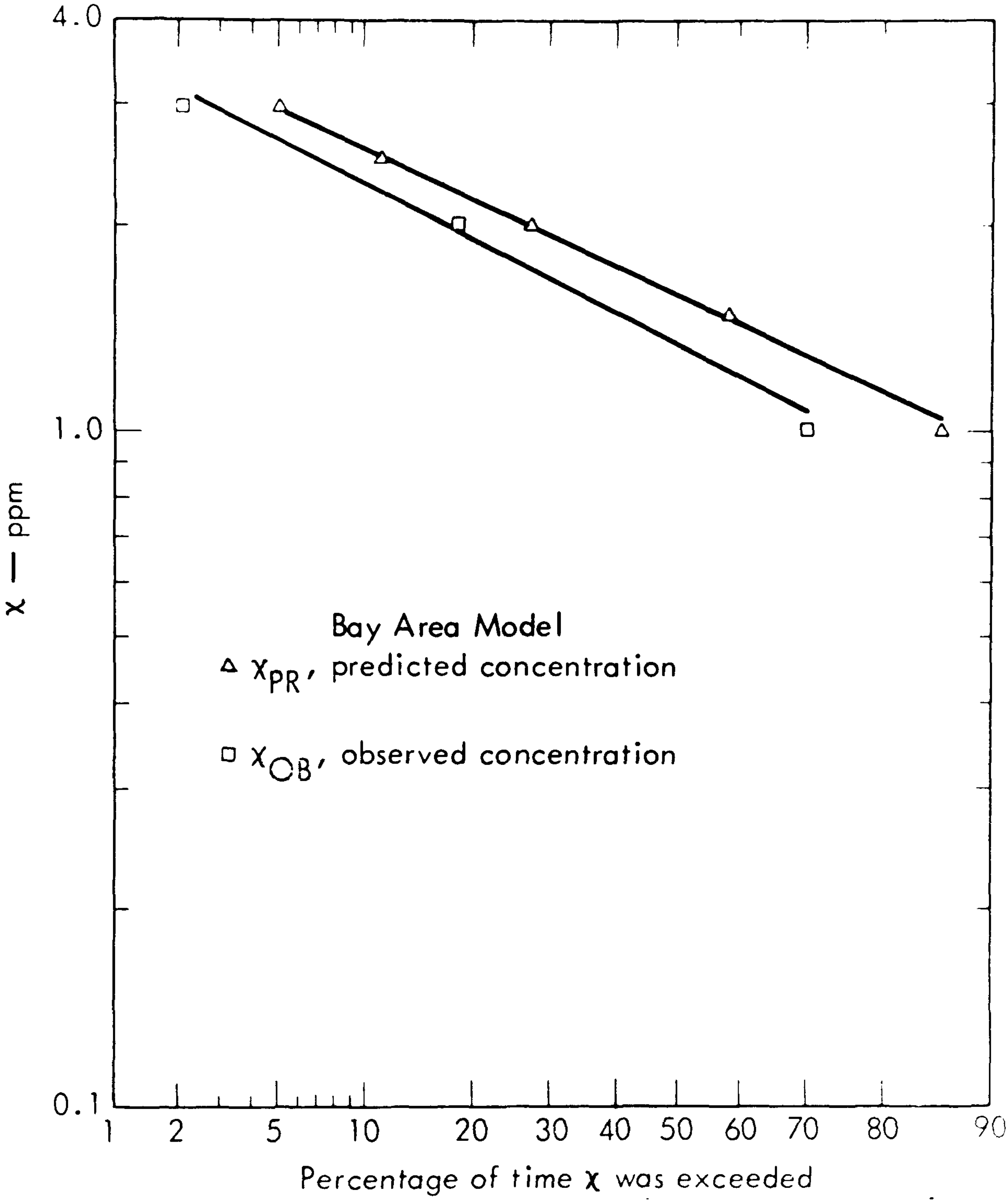

Figure 5. Observed and predicted hourly CO concentration distributions for San Francisco, July 10–11, 1968.

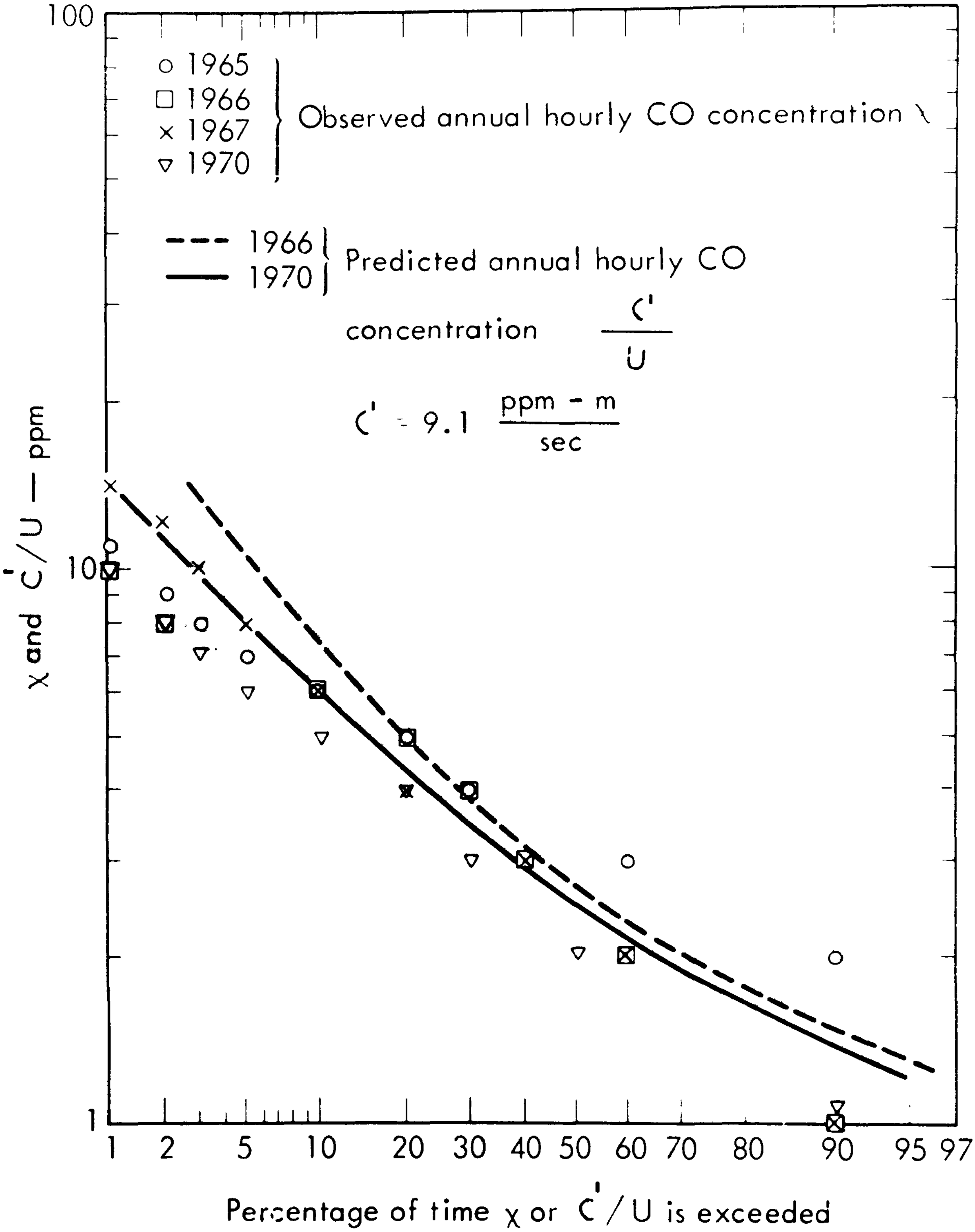

Figure 6. Observed and model-predicted annual hourly CO concentration distribution for San Francisco.

Table 1

Classification of urban air pollution calculations, following their purpose and accuracy requirement

Category of use	Purpose: General	Purpose: Special	Time available to perform the calculations	Meteorological input	Output: Accuracy of the results	Output: Probability of the materialization of the forecast	Appropriate method(s) see Column x in Table 2
Furthering basic knowledge	To show the rightness of some concept or "mechanism", as the basis of the phenomena. To prove a theory.		Not limited	Any Kind	Any systematic error will disprove the underlying theory or assumption		
Air management; urban planning: land use, transportation = long term strategy	Achievement of air quality goals. Estimation of future emission standards to obtain given ambient standards. To assess trends.	1. Calculation of the urban background concentration of long term (annual, seasonal...) averages. 2. Calculation of the effects of changes in emission patterns on long term averages. 3. Geographical localization of the effects of changes in the emission patterns.	Years	Climatological frequency tables or averages	Great accuracy of the estimate required, as the cost of any error can be very high	As long-term weather statistics (averages) used as input are quite stable: the probability of accurate materialization of the forecast is quite good.	2, 4, 6, 7, 8
Air pollution control = short term	To activate emergency control procedures: Short period forecast and warning	Calculating pollutant concentrations based on given (assumed or forecast) set of meteorological conditions.	From 3 to 24 hours	Meteorological forecast	No high numerical accuracy required: stratification by classes will do.	Required: high probability, as the cost preventive control can be quite high. Limited by the reliability of meteorological forecast	1, 3, 5
	To activate emergency control procedures: Implementation of real time strategies for episode attenuation.	Minimizing deleterious effects by purposeful, rapid changes in the emission inventory patterns, considering the evolution of the meteorological conditions.	A few hours at most.	Meteorological forecast	Relatively good accuracy warranted by continuous input of observed data.	Relatively high probability warranted by the short time span involved.	1, 3, 5

TABLE 2. SYSTEMATICS OF URBAN AIR POLLUTION MODELS BASED ON THE INPUT PARAMETERS

Input: Data on dispersion, wind field and/or other meteorological parameters	Input : emission				or concentration			
	Observed or assumed emission data (= Source oriented)				Past experience of observed ambient concentration (= Receptor oriented)			
Assumed or forecast	Plume A							
			Forecasting pollution B				Forecasting pollution B	
Externally computed or actually observed	Conservation of mass (volume element or trajectory) A	Plume combined with frequency classes A	B o x A		Statistical (correlations with meteorological data: C	Windfield-conc. field relations D		
Without meteorological input						Statistical (time series)	Concentration - frequency distribution D D	Correlation with demographic data
Output: Space resolution	H i g h		L o w		H i g h		L o w	
Output: Time averages (resolution)	Short	Long	Short	Long	Short	Long	Short	Long
Column number	1	2	3	4	5	6	7	8

Table 3. TABULATION PREDICTION TECHNIQUE

Wind dir degrees	Ceiling, feet	Wind sp, kt	Temp, deg F	Hour, CST	Percentile values of SO_2 (ppm) concentrations Min	25	50	75	90	95	98	99	Max	75-25	95-75	Mean	St dev	Freq
0–140	7000–****	0–3	70–79	0–3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
0–140	7000–****	0–3	70–79	4–15	0.01	0.01	0.03	0.04	0.06	0.06	0.06	0.06	0.06	0.03	0.02	0.03	0.0183	6
0–140	7000–****	0–3	70–79	16–23	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.0	0.0	0.01	0.0	1
0–140	7000–****	0–3	80–89	0–3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
0–140	7000–****	0–3	80–89	4–15	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.01	0.0	0.02	0.0050	2
0–140	7000–****	0–3	80–89	16–23	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.0	0.0	0.02	0.0	1
0–140	7000–****	4–7	10–19	0–3	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.0	0.02	0.0047	3
0–140	7000–****	4–7	10–19	4–15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
0–140	7000–****	4–7	10–19	16–23	0.08	0.08	0.10	0.12	0.12	0.12	0.12	0.12	0.12	0.04	0.0	0.10	0.0200	2
0–140	7000–****	4–7	20–29	0–3	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.0	0.0	0.02	0.0	1
0–140	7000–****	4–7	20–29	4–15	0.09	0.10	0.15	0.20	0.21	0.21	0.21	0.21	0.21	0.10	0.01	0.15	0.0460	4
0–140	7000–****	4–7	20–29	16–23	0.05	0.05	0.07	0.10	0.27	0.27	0.27	0.27	0.27	0.05	0.17	0.12	0.0864	4
0–140	7000–****	4–7	30–39	0–5	0.01	0.01	0.03	0.04	0.05	0.06	0.08	0.08	0.08	0.03	0.02	0.02	0.0175	17
0 140	7000–****	4–7	30–39	4–15	0.01	0.02	0.04	0.09	0.13	0.15	0.28	0.29	0.30	0.07	0.06	0.06	0.0650	24

Table 4

Contingency table of the ozone prediction results on the training set by RUFF (1974)

	Predicted O_3		Total
Measured O_3	1.p.p.m	1 p.p.m	
1 p.p.m	25	4	29
1 p.p.m	1	10	11
Total	26	14	40

Skill score with the training set, NO_2 deleted : 0.75

Table 5

Breakdown of the episode forecasts for 540 forecadts (3 winters) for Rouen, France, after BENARIE (1971) and BENARIE and MENARD (1972)

Winter	Number of episode days	Out of these forecasts			An episode was forecast, but did not materialize	
		Correct	Incorrect	due to incorrect meteorological forecast		due to incorrect meteorological forecast
68/69	8	6	2	0	1	3
69/70	11	7	1	3	3	3
70/71	19	12	0	7	7	1
Total	38	25	3	10	11	7

Table 6

Contingency table of the forecast results obtained by BENARIE (1971) and BENARIE and MENARD (1972) with apparent meteorological forecast error removed : 523 cases out of the total of 540.

		N° episode	Episode	Total	Per cent correct
		Predicted			
N° Episode	Observed	484	11	495	97.5
Episode		3	25	28	≈ 90
Total		487	36	523	

Skill score : 0,76

Table 7

Contingency table of the forecast results obtained by BENARIE (1971) and BENARIE and MENARD (1972), overall results for 540 cases

		N° episode	Episode	Total	Per cent correct
		Predicted			
N° episode	Observed	484	18	502	94.5
Episode		13	25	38	67
Total		497	43	540	

Skill score : 0.60

Table 8

Multiplying factors to be applied to the 30-day running average to obtain winter episode concentration in ROUEN France.

Condition		Factor
Mean surface wind	3 m s^{-1} for 24 hours	1.5
- " -	1 m s^{-1} - " -	2.0
- " -	0.5m s^{-1} - " -	2.5
- " -	0.5m s^{-1} second 24 hour	3
- " -	0.5m s^{-1} third 24 hour	3.5
Mean temperature below - 3°C		1.5
Wind blowing from 00° to 80° direction		
Factor for downtown		1.3
Factor for "Petit Couronne" station		0.7
Wind blowing from 260° - 280° direction, at 4-7 m s^{-1}, only for the "Petit Couronne" station.		2.0

Table 9

Comparison of observed (first of each group of three lines) data, calculated from the virtually occuring meteorological data (second lines) and forecast, using forecast meteorological data (third lines) for seven sampling stations in ROUEN, France winter 1968/69

		Factors for Wind speed	Temp.	Direction	SS 8	SS 15	Préf.	Sott.	Fac.	Mar.	Pt-Cour.
Mean 9Nov-8Déc.					102	102	52	101	86	82	66
9Dec.	Observed				103	137	56	133	106	131	97
	Calc.a posteriori	1.5	0.5		76	76	39	76	65	62	50
	Forecast				102	102	52	101	86	82	66
10Dec	Observed				119	137	42	111	52	80	82
	Calc.a posteriori	1.5	0.5		76	76	39	76	65	62	50
	Forecast				204	204	104	202	172	164	132
11Dec	Observed				212	242	193	377	127	136	255
	Calc.a posteriori	1.5			153	153	78	151	128	123	99
	Forecast				204	204	104	202	172	164	132
12Dec	Observed				261	274	126	218	286	323	148
	Calc.a posteriori	2	1.5	0.7	214	214	109	212	180	172	138
	Forecast				204	204	104	202	172	164	132
Mean 30Nov-29Dec.					100	99	46	121	75	92	96
30Dec	Observed				169	143	102	157	205	175	86
	Calc.a posteriori	2.5		0.7	175	173	81	212	131	161	167
	Forecast				200	198	92	242	150	184	192
31Dec	Observed				367	352	271	(360)	(214)	(236)	256
	Calc.a posteriori	2	1.5		300	297	138	362	235	276	288
	Forecast				200	198	92	242	150	184	192
1Jan.	Observed				346	358	296	(360)	(214)	(236)	188
	Calc.a posteriori	2			200	198	92	242	150	184	192
	Forecast				200	198	92	242	150	184	192
Mean 24Dec-22Jan.					117	110	67	131	30	103	105
23Jan	Observed				211	202	104	225	150	94	128
	Calc.a posteriori	2			234	220	134	262	160	206	210
	Forecast				234	220	134	262	160	206	210
24Jan	Observed				145	148	42	101	65	66	284
	Calc.a posteriori	2			234	220	134	262	160	206	210
	Forecast				234	220	134	262	160	206	210
Mean 5jan-3Fev.					97	88	54	102	73	84	161
4 Feb	Observed				207	226	192	306	110	190	214
	Calc.a posteriori	2			194	176	108	204	146	168	322
	Forecast				194	176	108	204	146	168	322
5 Feb	Observed				171	208	149	261	112	130	170
	Calc.a posteriori	2			194	176	108	204	156	168	322
	Forecast				194	176	108	204	146	168	322
Mean 18jan-16feb					130	126	78	145	82	95	186
17feb	Observed				198	180	109	(298)	(108)	(115)	226
	Calc.a posteriori	2			260	252	166	290	164	170	372
	Forecast				260	252	166	290	164	170	372

Note. The bracketed observed values correspond to 3 days' sampling and were not taken into consideration for the computation of the RMS error.

Table 10

Predictions of hourly values of CO - concentration in ppm, in the Los Angeles Basin, on September 29 1969, by HANNA (1973). The predictions are initialized with 5 a.m. value.

Time		Downtown LA	Azusa	Burbank	West LA	Long Beach	Reseda	Pomona	Whittier	Correlation coefficient	Mean values for the 8 stations
5	Obs.	2	9	12	6	3	9	7	6		
	Calc.	2	9	12	6	3	9	7	6		
6	Obs.	3	10	11	5	7	10	7	7	-0.32	7.5
	Calc.	2	2	3	17	10	6	7	4		6.4
7	Obs.	7	10	14	9	9	12	8	11	0.44	10.0
	Calc.	7	6	35	46	18	25	12	17		20.6
8	Obs.	17	10	14	17	13	16	9	14	0.47	13.8
	Calc.	14	13	7	46	27	50	17	12		23.2
9	Obs.	18	11	13	17	11	11	9	11	0.06	12.6
	Calc.	7	13	35	46	18	50	17	24		26.2
10	Obs.		10	12	8	9	8		11	-0.39	9.7
	Calc.		13	35	46	13	50		24		30.2
11	Obs.		8	9		6	6	6	13	0.03	8.0
	Calc.		26	35		8	50	17	24		26.7
12	Obs.	6	6	9	5	6	4	5		0.02	5.7
	Calc.	3	5	35	11	10	50	8			17.5
13	Obs.	3	7	4	4	6	5	4	2	-0.01	4.3
	Calc.	3	3	17	9	7	25	8	8		10.0
14	Obs.	3	9	3	4	5	3	5	1	-0.38	4.1
	Calc.	3	3	11	11	5	25				9.2
15	Obs.	2	6	3	4	5	3	6	1	-0.22	3.8
	Calc.	3	3	8	15	5	25		8		9.1
16	Obs.	6	6	4	4	5	5	6	1	-0.51	4.6
	Calc.	3	3	8	11	4	7	8	8		6.5
Correlation Coefficient		0.84	0.33	0.34	0.81	0.97	0.19	0.68	0.69		0.74

T a b l e 11

Validation of the simple box model, according to HANNA (1971) and GIFFORD (1972, 1973).

City	Concentration value modelled	Period	Pollutant	Correlation coefficient
CHICAGO	Point	24 hours (for 18 days)	SO_2	.67
CHICAGO	Point	6 hours (for 4 days)	SO_2	.66
LOS ANGELES	Point	1 hour (for 17 hours	CO	.89
SAN FRANCISCO	Point	1 hour (for 48 hours)	CO	.74
LONDON	Area average	24 hours average	SO_2	.76

TABLE 12. DATA RELATED TO PARTICLE AND SO_x POLLUTION FOR U.S CITIES*

City	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Washington	247	35	7.5	775	72	412	90	58
New York	1795	243	8.2	2330	98	265	346	128
Chicago	1780	586	7.3	2590	139	154	221	81
Philadelphia	1168	231	7.8	4400	124	634	217	218
Denver	29	29	6.3	260	117	227	18	35
Los Angeles	187	101	5.1	1035	114	205	—	—
St. Louis	662	176	6.5	595	161	122	132	27
Boston	428	74	8.0	775	83	240	—	—
Cincinnati	349	73	6.2	905	122	323	44	24
San Francisco	174	102	5.4	1035	73	138	—	—
Cleveland	818	304	7.4	650	106	57	78	16
Pittsburg	934	387	7.1	4815	140	426	93	117
Buffalo	410	140	7.6	620	116	135	25	10
Kansas City	125	60	7.3	390	97	158	12	10
Detroit	786	240	7.3	1035	143	155	16	5
Baltimore	255	104	7.5	200	133	67	107	22
Hartford	337	56	8.1	465	81	188	62	24
Indianapolis	164	78	6.8	415	146	182	54	32
Minneapolis–St. Paul	215	46	7.5	775	88	384	44	41
Milwaukee	242	100	7.5	775	95	191	28	23
Providence	118	23	8.3	155	113	218	125	47
Seattle–Tacoma	225	33	5.5	260	79	117	35	7
Louisville	303	128	6.6	620	121	133	—	—
Dayton	186	174	7.0	1035	116	166	49	66
Houston	144	158	6.6	620	67	60	—	—
Dallas–Ft. Worth	16	52	7.0	570	96	253	—	—
San Antonio	2	129	6.5	630	68	75	—	—
Birmingham	33	205	5.9	520	128	66	—	—
Steubenville	638	155	6.1	520	173	121	—	—
Average	441	146	7.0	1027	111	202	90	50

* Column legends:

(1) Q_{TOT}, SO_x, 10^3 tons yr^{-1}, FENSTERSTOCK *et al.* (1969); 10^3 tons yr^{-1} = 28·73 g s^{-1};
(2) Q_{TOT}, particles, 10^3 tons yr^{-1}, FENSTERSTOCK *et al.* (1969);
(3) u, m s^{-1}, FENSTERSTOCK *et al.* (1969);
(4) Approximate area, km^2, enclosed by 0.1 tons day^{-1} mi^{-2} urban particulate source, estimated from *Reports on Consultation*, U.S. DHEW (1968–1969);
(5) Observed average concentration, X, of particles, $\mu g\ m^{-3}$, FENSTERSTOCK *et al.* (1969);
(6) c (dimensionless) for particles, using equation (2) and cols. (2, 3, 4 and 5);
(7) Observed average concentration, X, of SO_x, $\mu g\ m^{-3}$, from U.S. DHEW (1968);
(8) c (dimensionless) for SO_x, using equation (2) and cols. (1, 3, 4 and 7).

Table 1 3. DISTRIBUTION OF SELECTED CITIES BY POPULATION CLASS AND PARTICLE CONCENTRATION, 1957 TO 1967.

[Avg. particle concentration μg/m³]

Population class	<40	40 to 59	60 to 79	80 to 99	100 to 119	120 to 139	140 to 159	160 to 179	180 to 199	>200	Total cities in table	Total cities in U.S.A.
>3 million	----	----	----	----	----	----	1	----	1	----	2	2
1-3 million	----	----	----	----	----	----	2	1	----	----	3	3
0.7-1 million	----	----	1	----	2	----	4	----	----	----	7	7
400-700,000	----	----	----	4	5	6	1	1	1	----	18	19
100-400,000	----	3	7	30	24	17	12	3	2	1	99	100
50-100,000	----	2	20	28	16	12	6	5	1	3	93	180
25-50,000	----	5	24	12	12	10	2	1	2	3	71	--------
10-25,000	----	7	18	19	9	5	2	3	1	----	64	[a] 5,453
<10,000	1	5	7	15	11	2	1	2	----	----	44	--------
Total urban	1	22	77	108	79	52	31	16	8	7	401	--------

[a] Incorporated and unincorporated areas with population over 2,500.

Table 14

Computed CO-concentrations (17 hour day-averages, ppm) compared with the observed values for Sept. 23, 1966, of the Los Angeles Basin.

1 Stations	2 Observed concentration	3 Computed (LAMB, 1968)	4 Random	5 Mean 13.5
Downtown LA	16	22	9	14
Azusa	13	3	7	13
Pasadena	17	15	15	14
Burbank	16	7	8	13
East LA	12	5	10	14
West LA	16	13	17	13
Long Beach	14	8	14	14
Hollywood	17	7	11	13
Pomona	13	3	16	14
Lennox	13	11	13	13
Anaheim	9	7	6	14
La Habra	6	3	12	13
RMSE		6.8	4.7	3.2
Correlation Coefficient		0.55	0.25	0.00
a		0.32	0.23	0.00
b		10.7	10.8	13.5

a and b refer to the coefficients of the regression equation :

Computed conc. = a (Observed conc.) + b

ADDENDUM

P 25 After Lettau reference add:

Instead of the integration of the Gaussian infinite line source, GOUMANS and CLARENBURG (1975) considered a large number of randomly distributed point sources over the area and a Sutton-like plume formula. Their computational formula is equivalent to that proposed in 2. above (GIFFORD and HANNA 1976)

P 29 After line 19 add:

Calculation of seasonal means by GOUMANS and CLARENBURG (1975) for The Hague and Amsterdam (Netherlands) show a very good fit with the observed values

P 46 Add to References:

Goumans, H.H.J.M. and L. A. Clarenburg, 1975: A simple model to calculate the SO_2-concentrations in urban regions. Atmos. Env. 9, pp 1071-1077

Gifford, F. A. and S. R. Hanna, 1976: Discussion to the paper of Goumans and Clarenburg, Atmos. Env. 10, p 564

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/4-76-055	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE URBAN AIR POLLUTION MODELING WITHOUT COMPUTERS		5. REPORT DATE November 1976
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Michael M. Benarie		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Environmental Sciences Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711		10. PROGRAM ELEMENT NO. 1AA009
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Sciences Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Inhouse
		14. SPONSORING AGENCY CODE EPA-ORD
15. SUPPLEMENTARY NOTES Prepared by Visiting Scientist		

16. ABSTRACT

This report was the basis for a series of three lectures by the author on urban air pollution modeling, and represents a condensed version of selected topics from a recent monograph by him. The emphasis is on simple but efficient models that often can be used without resorting to high-speed computers. It is indicated that there will be many circumstances under which such simple models will be preferable to more complex ones. Some specific topics included in the discussion are the limits set by atmospheric predictability, forecasting pollution concentrations in real time as for pollution episodes, the simple box model for pollution concentrations, the frequency distribution of concentration values including the log-normal distribution and averaging-time analysis, the relationships between wind speed and concentration, and lastly the critical question of model validation and the need to consider several indices of goodness-of-fit if pitfalls are to be avoided.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
* Air pollution		13B
* Meteorological data		04B
* Mathematical modeling		12A
* Model tests		14B

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 82
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)