2

Air pollution calculations for urban areas

2.1 Air pollution calculation—or modelling? a semantic discussion

> 'For the sake of persons of different types of mind, scientific truth should be presented in different forms and should be regarded as equally scientific whether it appears in the robust form of the physical illustration or the beauty and paleness of a symbolical expression.'
>
> CLERK MAXWELL

> 'Perhaps the most fascinating aspect of Maxwell's genius is that, as soon as he had worked out the mathematical formulae of his theory, he discarded his model by means of which he reached it. . . . all that remained were "fields" of an abstract non-substantial nature, and the mathematical formalism which described the propagation of real waves in an apparently non-existent medium.'
>
> A. KÖSTLER

It is necessary first of all to define the words 'model' and 'modelling' clearly and definitively, as the colloquial use of technical jargon frequently causes the meanings of words to deviate from their original sense.

Firstly, the word model can mean a tangible copy of the real object but at a reduced scale. In this way, we have models of buildings, landscapes, airplanes, ships, harbours, etc. These models are used for experiments under given conditions and limits, when experiments on the real object would be impossible or uneconomical. Landscape and building models are also used in pollution engineering: they are outside the scope of this book which is limited to calculations.

The second meaning of the word model refers to imaginary structures in

physics which, it is hoped, are governed by the same mathematical laws as the real object. In physics, the model is expressly *not* a miniaturised version of the phenomenon to be modelled: model and phenomenon are of basically different natures. Bohr's atom with elliptically orbiting electrons is an obvious example of such a model; the Kant–Laplace concept of the formation of the solar system is a model; the 'explanation' of electric current and potential by analogy with hydraulic flow and pressure is another model; the lines of force in electromagnetism are models; and so was the ether, before being supplanted by other concepts.

In the following discussion, the first type of model will be called a model object, while the second type will be referred to as a model concept.

Model objects are used to simulate a given occurrence, provided a given set of dimensional and analogous laws are adhered to. In this case, the set of laws is called the similitude laws. For both model objects and model concepts, it is the experimenter that must prove that the model is correct and must show that the conditions under which the model is used allow extrapolation to a real scale, as this is the purpose of the operation. For example, when a relief map placed in a water channel under laminar flow conditions is used to simulate the turbulent atmospheric boundary layer, adequate proof may be lacking. However, as simulation by model objects is outside our scope, we shall end our discussion here, with the clear definition of what model objects are. We have indicated, but not clearly defined yet, what simulation actually is: the definition will be provided later.

Although there is a difference between the ideas of model concept and theory, these two words are often wrongly used as equivalents. For example, Bohr's atomic model is often called Bohr's theory of the atom; the Kant–Laplace model of the solar system is called the Kant–Laplace theory; and so on. We can elucidate the difference between these two terms best by giving an example. Let us consider the kinetic theory of gases. The model used in the kinetic theory is that of a perfectly elastic spherical molecule, flying and bouncing around. However, other concepts and observations make it extremely unlikely that the molecule is a hard perfect ball. Nevertheless, a great number of equations have been worked out by describing the behaviour of molecules as identical with that of billard balls; by using these equations the kinetic theory works very well. The difference between a theory and a model is the following: the theory gives an idea of how objects *are* and the model (the conceptual model) tries to describe how objects *behave*. The molecules of the kinetic theory *behave* as elastic spheres, but no one believes that they actually are elastic spheres. However, phenomena such as molecular diffusion show that molecules do in fact move around, and so the idea of their flying around roughly approximates reality. While a theory tries to explain phenomena, a model concept is a frame of thought to which a mathematical equation may be tied.

If the dispersion of smoke from a point source is described by a so-called gaussian plume, may we call this description a mathematical model, at least in the two precise senses defined above? Certainly not.

The initial assumption (or hypothesis) was to describe this phenomenon by an equation such as Fick's equation. By a sequence of purely mathematical manipulations, we arrived at a gaussian plume. Experimental data supply its spreading characteristics, and thus the concentration at a given point can be computed. Provided that enough time and computing facilities are available, this can be done for a great number of point sources, and, by assuming one of several hyoptheses, the concentrations can be added up.

By proceeding in this way, in principle the same steps are taken as when calculating the area S of a rectangular field by multiplying the length a of one side by the length b of the other. In this case, the assumption is that the field is rectangular. Next, the convenient mathematical expression $a \times b = S$ is applied, by using a second assumption, that is, that a and b are experimentally accurate. The formula $a \times b = S$ is unconditionally correct; so is, for example, the integration of the differential equation of diffusion under given boundary conditions to give a gaussian concentration distribution. However, whether the result is correct depends on how far the basic hypothesis meets reality and on the errors that are made in the process of measurements. For example, this applies to the field: if the angles of the field are not right angles or if the sides are measured with a large error, then the result will be false. The same is true for the concentration calculation: perhaps the basic process is not governed by a fickian equation or perhaps the parameters for the lateral or other spread are far from real. Then, the surface (or the concentration) will also have values that are very far from their real values; but the mathematics will not be at fault. In both cases, the mathematics are absolutely correct; only the assumptions made during their application are erroneous.

Although the concentration calculation is analogous to the surface calculation, it is even less clear why the concentration calculation should be called plume modelling. We did not 'model' the surface of the rectangle: we calculated it. Similarly, for the plume, no conceptual model was involved; only a given formula was applied. For the rectangular surface, a model object could perhaps be used, by cutting an analogous form out of cardboard and by weighing this model; however, we did not use this process: the concentration *and* the area were calculated without the aid of modelling. If the same routine (hypothesis; the application of a mathematical formula; the introduction of experimentally measured parameters) is used for concentration calculations in an urban area, there are no grounds at all for calling this process urban modelling.

Engineering sciences adopted the term to model from such subjects as economics, demography and linguistics and so on and used it to mean to calculate or to compute. Nowadays this usage has become commonplace in atmospheric physics and engineering, and the term modelling is used interchangeably with the term computing. In the following chapters we shall mainly comply with this usage, but, when model object or model concept is meant and not simply calculations, we shall emphasise this by using these fuller terms.

Similar comments apply to the use of the word simulation. A simulation means an experiment performed on a model. It is obvious that experimental manipulations can be made only on model objects. The model concept is used to find a suitable mathematical expression; afterwards the 'theoretical' results are compared with the observations. Sometimes it is more convenient to proceed stepwise by trial and error instead of by deriving a theoretical formula. Engineers have used this approach for centuries and quite often apply it graphically. Most problems of the great nineteenth-century engineering projects were solved graphically, on the drawing board. However, the expression the simulation of stresses by rule and compass was never used. It only appeared when calculations were made on computers, when the term computer simulation was coined. So we shall use the term computer simulation to mean calculation.

2.2 Outline, systematisation and purpose: engineering calculations

The choice of an engineering calculation method for urban pollutant concentrations depends on how the results are to be used. For a purely scientific endeavour, when the aim is to gain knowledge for knowledge's sake, the practical use is of no importance. When he was asked why he climbed Mount Everest, Sir Edmund Hillary answered, 'Because it is there.' However, an engineer would not choose between the projects of constructing a rack railway, a ski lift, a helicopter platform, etc., to the top of Everest without first enquiring as to the purpose of the enterprise. The design will be decided on a cool cost effectiveness basis, and the project could well be abandoned, given the actual number of visitors to the peak per year.

Sir Edmund Hillary's approach is that of a sportsman and equivalent to that of a pure scientist. On the contrary, ours must be a strictly engineering approach, with calculations made either to avoid the unrealistically high cost of real-scale experiments (it is easier to calculate the resistance of a beam than to shatter some real beams several times) or to estimate some future situation not accessible by direct testing (for example, the next day's sulphur dioxide concentration).

In this way, engineering decisions and calculations are made by asking two questions.

(1) Why are these calculations being made?
(2) Which are the best methods to use?

The main purpose of this book is to indicate the problems involved; this means finding out *what* must be done, but it provides very little information about *how* to make the computation itself. Programme listings are mainly provided by the references and are not included in this book as there was not sufficient space.

Before undertaking any urban air pollution calculations, the following

questions should be answered.

(1) Who needs the information?
(2) What purpose has to be attained?

Table 2.1 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 labour 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 is whether the quality and the quantity of the available input does in fact justify the computational burden. However, some computationally simple methods need high-grade input information (Hameed, 1974; Benarie, 1975). An obvious example is the persistence model which without any calculations, when projected a short time forwards, 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.

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 in section 2.3 will be concerned with the search for such limits.

Table 2.1 is an attempt to systematise computations on the basis of their purpose. Another classification can be based on the amount and the nature of external (mostly meteorological) information needed as input for the model (figure 2.1).

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. In the following discussion we shall use the term 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 and Egan (1971) have coined the word driven for this kind of input.

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

Table 2.1 Classification of urban air pollution calculations by their purpose and accuracy requirement

Category of use	*Purpose*		*Time available to perform calculations*	*Meteorological input*	*Output*	
	General	*Special*			*Accuracy of results*	*Probability of materialisation forecast*
To further basic knowledge	To show whether concept or mechanism is correct as basis of phenomena; to prove a theory		Not limited	Any kind	Any systematic error will disprove 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 urban background concentration of long-term (annual, seasonal, etc.) averages (2) Calculation of effects of changes in emission patterns on long-term averages (3) Geographical localisation of effects of changes in emission patterns	Years	Climatological frequency tables or averages	Great accuracy of estimate required as cost of any error can be very high	**As long-term weather statistics (averages) when used as input are quite stable, probability of accurate materialisation of forecast is quite good**

Air pollution control (short-term use)	To activate emergency control procedures	Short-period forecast and warning	Calculation of pollutant concentrations based on given (assumed or forecast) set of meteorological conditions	**From 3 to 24 h**	**Meteorological forecast**	**No high numerical accuracy required; stratification by classes will do**	High probability required as cost-preventive control can be quite high; limited by reliability of meteorological forecast
		Implementation of real-time strategies for episode attenuation	Minimisation of deleterious effects by purposeful rapid changes in emission inventory patterns, by considering evolution of meteorological conditions	Few hours at most		**Relatively good accuracy warranted by continuous input of observed data**	**Relatively high probability warranted by short time span involved**

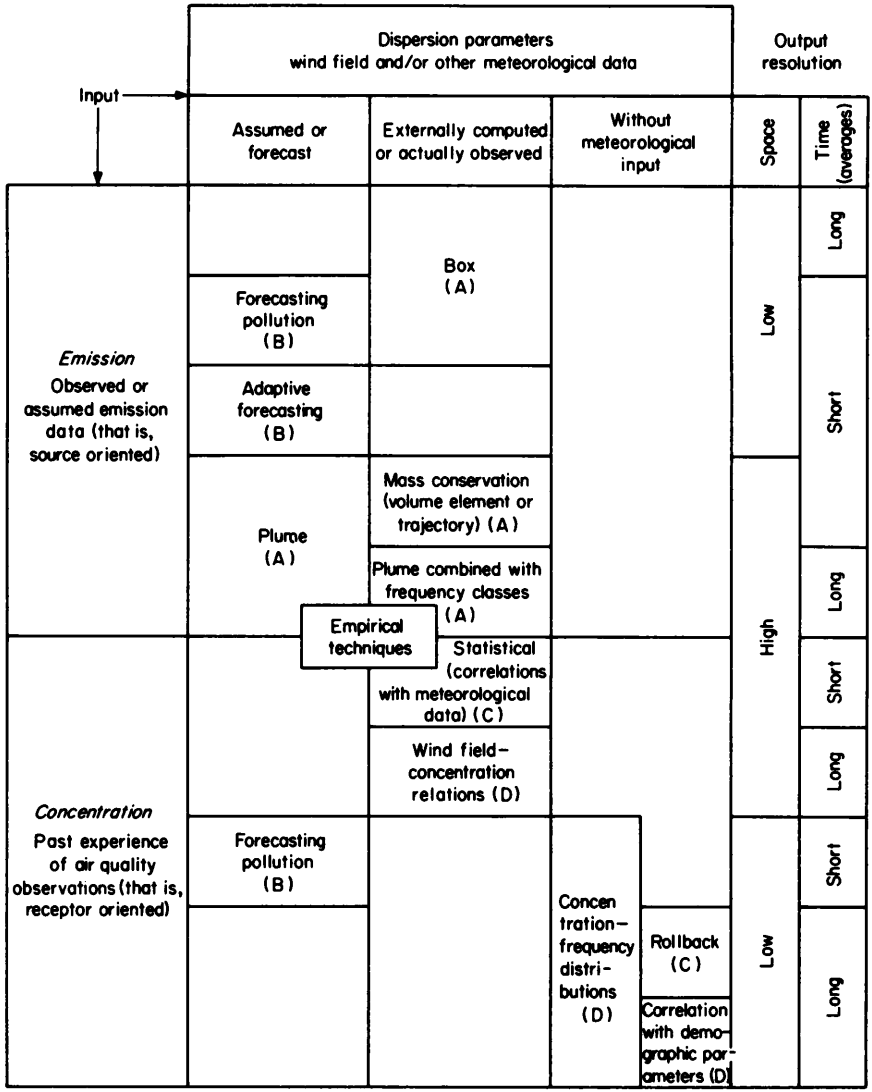

Figure 2.1 The systemisation of urban air pollution models

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.

This is pure idealisation, as in most statistical calculations the independent variables are still entities such as wind direction, wind velocity, etc., that are presumed to be the causal link between source and receptor. In reality, some causality and randomness exist in almost any calculation scheme; but for

argument's sake we shall assume that both ends of the scale are found in a pure form. The causal models are used to explain the underlying physical phenomena; the statistical models merely describe them numerically.

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 for controlling the system and for producing the desired changes in performance. Receptor-oriented models are generally descriptive and less directed towards establishing cause-and-effect relationships.

Figure 2.1 also distinguishes between short- and long-term *objectives*, that is, 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 the 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 h. Anyway, the basic ideas in the 30-min averaging time computation and in that for the seasonal or yearly average show enough difference for recognition of two quite distinct classes for short- and long-time mean calculations.

These principles enable us to classify the main types of urban air pollution models, as shown in figure 2.1.

Examination of figure 2.1 shows that beyond the main classification criteria (for example, by short- or long-term averages or by whether they are 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 the letters A to D in figure 2.1.

(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, that is, 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 materialise. 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 the

content of a reactor is computed from reactants, stirring, temperature, etc. The output is primarily a numerical value assigned to a space and a time coordinate.

(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. Although valuable for predicting trends or cycles, it is of little use for a true forecast (for example, a certain day's estimate of the next day's pollution).

(D) Finally, we have the description (which may also be termed statistical) or summarisation of the data already on record, mostly in the 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.

2.3 The bounds: what can be attempted and the limits of accuracy

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 8 balloons in the atmosphere. These 'mesh particles' will undergo a rapid change in 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, as bounded by squares at $t = 0$. 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, 1971) found that a particle with mesh size equal to 300 km should cease to exist within the period $12\text{ h} < t < 75\text{ h}$. Egger (1973), using the data of Kao and Al-Gain (1968) and Kao and Powell (1969) on large-scale dispersion of clusters of particles in the atmosphere, suggests $45\text{ h} < t < 72\text{ h}$, while, using the data of Eole (Morel, 1970; Larcheveque, 1972), he arrives at $t \approx 45$ h.

This estimate is 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 factor arises from the finite representation of the atmospheric fields in the models, which makes it impossible to describe scales of motion below grid scale. Owing to the non-linearity of the hydrodynamic equations, the parts of the turbulent energy which are contained in the subgrid range will appear under an alias in the larger scales, thus limiting the predictability of these

scales. Fleming (1971) states, 'It is this last type of uncertainty that is generally felt to be responsible for the limit of predictability of various scales.' Another factor is 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) A knowledge of the source strength.
(2) An adequate definition of the meteorological parameters.
(3) A reliable method for the calculation of the dispersion from inputs 1 and 2.
(4) An 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 10 to 20 elementary processes have to be estimated. Only a few of these can be calculated free of error. Many can only be estimated approximately, 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) The wind velocity has no influence on the source strength.
(2) The main meteorological parameters (the 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. The literature on turbulent motion in the atmosphere is unusually scarce on the topic of the directional variance of very light winds. This arises because 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 have 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 add very little to the already plentiful evidence. In the book by Lumley and Panofsky (1964, p. 151) which contains extensive information about atmospheric turbulence, there is only 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 min 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: firstly, 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 just as important when considering light winds. Optimistically, we can only hope that these deficiencies in knowledge will not increase the error under calm conditions.

Thus, even if a low wind velocity did not influence questions 1 and 4, its effect through 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 air flow and urban heat islands that, during conditions favourable to the formation of an urban heat island, 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 emphasised 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 it is the causal chain between the pollutant source and the ambient concentration, becomes meaningless when the wind velocity falls below a certain value.

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

Obviously, this is impossible to attain. However, by the same logic, multi-nodal models with the introduction of random noise at every step *will not* indicate with accuracy the next day'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, to

determine the categories or to observe trends, but it seems of no use for improving the accuracy of forecasts.

This statement should not be interpreted as saying that all sophistication is by definition to be rejected. Some very simple one- or two-step schemes show a creditable, if not outstanding, performance mainly in forecasting. However, if very sophisticated long-chain arguments *must* 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 centre 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 centre 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 when the axis approaches the centre of gravity the chain of causality has broken down and has been replaced by a probability situation. We shall not discuss the fundamentals here, as these are well known from probability calculus; we shall only emphasise that the situation that occurs in urban air pollution is similar to that of the beam example. 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 calculating the faces on which 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 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}, has already been well emphasised in the literature.

In so far 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 generalised and protracted pollution episodes occur mostly during moderate and light winds. On the contrary, plume concepts can be quite useful to localise 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 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, characterised by an unchanged emission rate and by approximately 40 meteorological parameters (such as wind direction and intensity, thermal gradient, cloudiness, the situation of a given air parcel relative to a front, etc.), if *all* these parameters were equal, *then* pollutant concentrations would also be the same for both time intervals.

By the same logic, it could be expected that, if 40 to 60 appropriate parameters were identical, then the same form of cumulus cloud would hover over the same quarter of the city. Of course, no one 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 emphasise 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 6 min during 1973 (personal communication). Flights cancelled before departure, as well as delays due to technical or commercial reasons, are not included in this statistic. If we take the average flight time as about 3 h, this means that the estimation was made with 4 % error. Now, these aircraft are driven by thousands of horsepower, are guided by exceptionally skilled crew and are 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 air flow only (instead of by a jet engine) should perform any better (Benarie, 1976)?

Finally, in considering the accuracy of our engineering calculations, we should not outperform the engineers who seek the technically best estimate within an economic frame. If an engineer conceives girders to support 10 tons, he does not mean that each girder will shatter when loaded at 10.01 tons. In the same way, when computing a given concentration at a given stability and wind speed, our ambition should not go beyond that of the engineers.

2.4 How do we choose a model?

Much has already been published on how to choose a model, without achieving unanimity. The reason is that this question of choice is mixed up with two fundamentally different but distinct questions.

(1) How do we formulate, develop or improve a model?
(2) Which of the available models performs the best?

The first is a question of research and better understanding of the atmospheric process. As model-user engineers, we do not look for a deep understanding but rather a logical use. Our problem is to choose a model from those already formulated.

The answer to the second question is very simple: all models perform equally in the sense of mathematical calculations, from the simplest to the most expensive. The formulae $S = a^2$ and $S = \pi r^2$ are equally effective, provided the first is not applied to compute the surface of a circle and the second is not applied to the surface of a square.

Now to revert to the original question, the following concise method can be recommended.

(1) Make an inventory of your requirements regarding time and space resolution. Look for the model yielding directly the required information without synthesising it out of fragmentary bits. If you need the annual mean, do not start computing 30 000 000 instantaneous values first, and make your yearly average afterwards.

(2) Evaluate how many man hours are (or how much money is) available to obtain the input data; their cost outweighs computation expenses (or time) by 20 to 1 for the very simple models. For the complicated models, an adequate emission inventory alone is easily 1000 to 10 000 times more expensive than just the computer time to treat the punched card batch (for the estimation of computer time and expenses, see chapter 16). As a general rule consider that, the more complex the model formulation is, the more complex are the data and parameterisation requirements also. The cost of non-source parameters, such as, for instance, the continuous determination of the three wind components at each grid point of a $100 \times 100 \times 10$ grid, may be beyond any reasonable or unreasonable spending. Other examples, such as the exact wind direction at some given point at 12:00 next Saturday, might be utterly unattainable.

(3) Consider that the time needed to gather adequate input data (this means mainly but not exclusively emission inventories to an accuracy, say, of $\pm 10\%$) is such that, at the time you will use them, socioeconomic or other factors may have changed them already by $\pm 30\%$ in an unknown direction. The only exception to this is pure assumption. If it is assumed that the source strength of a stack (still unconstructed) is 1 ton h^{-1}, then no more need be said, apart from the fact that the power plant may never be constructed, and, even if it is, then its

instantaneous genuine emission will be quite different from that assumed. All this applies also to receptor-oriented models.

(4) Be certain that the output resolution can never be better than the input resolution. The utmost that computational mathematics can accomplish is to display information in some novel more useful way. This differs from deductive mathematics as in theoretical physics, for example. Apart from a reasonable assumption of space and time continuity which allows interpolation between individually computed points, most smooth output plots are either computational artefacts or else some generalised assumptions lie hidden somewhere.

(5) With all this in mind plus a clear knowledge of the output requirements (see figure 2.1 and the corresponding sections where the topic is discussed in detail), a choice of model should be made.

2.5 The verification of calculations: validation techniques

Information on the merits of a chi-square test, skill score, correlation, root-mean-square error (RMSE) and absolute error, etc., can be found in the book by Panofsky and Brier (1968). Other references are by Brooks and Carruthers (1953), by Godske (1962) and by Eriksson (1962).

Validity and usefulness are sometimes confused. Let us try to clarify the difference between them. The agreement of a model with observation is usually referred to as validity (Brier, 1973). If the agreement is good, the model (theory) is considered true, although it is generally recognised that the expression true may be misleading since any model is at best an approximate description of reality. If the numerical performance of a model concept is poor, this is not a decisive reason for rejecting it. It might be that the experimental data used for the check were not as good as they seemed to be. In first trials it will be difficult to distinguish between conceptual errors and errors of physical parameters used during the computation.

Lack of perfect agreement between prediction and observation might raise questions of validity but does not preclude the possibility of usefulness. It seems a reasonable point of view to interpret validity in a relative rather than in an absolute sense and to consider usefulness in relation to available alternative methods and their cost. Statistics might be helpful in providing the investigator with a measure of confidence in the result of a particular procedure, but he should still go ahead with attempts to develop better models.

In statistical models, numerical performance must be the ultimate test. The very reason for its existence is its numerical performance. We shall discuss next the type of numerical performance that should be chosen.

One of the pitfalls in the verification procedure arises when attempting to compare the relative merits of models at two different locations. As no two strictly identical cities exist, the difficulties of modelling or calculating vary from site to site: a simple model might be adequate for one and invalid for

another. This is also true for the seasons, the time of day and the length of intervals over which the averaging is considered. When the merits of two calculation methods are compared, it should be done for the same events and on the same site.

Another danger is the comparison on a pure performance basis of two calculation methods or concepts by making use of very different amounts of input information. By input information we do not mean the involvement or the length of the computation (as pure calculation may be assumed errorless) but the data we start with. As an example, we shall consider the most elementary persistence model, that is, the assumption that concentrations at any given point will be the same at the time $t+\Delta t$ as they were at the time t. If Δt is relatively very short compared with the sampling interval, say, a few minutes, and if the sampling is taken over 1 h, then our extrapolation (with no computation involved) forwards in time will provide an almost perfect fit to the values actually measured. Obviously, this is not a merit of the model's ingenuity but is due to the enormous amount of up-to-date information used as input.

Thus it can easily be seen that the main problem of validation is to find acceptable significance tests for obtaining the confidence limits of the results of a model, particularly when these results are stated numerically. For forecast scores a scale of goodness has to be defined.

Before any satisfactory verification scheme is adopted, it is necessary to determine the primary purpose or purposes to be served by the verification. Usually what occurs in one part of a range will seem more important than what happens in other parts of the range, and this must be considered when specifying the purpose of verification. For example, the most important characteristic in alert period forecasts are the episodes of exceptionally high concentrations when only the percentage of correct values, in the high range, may be taken into account.

In general, if each purpose of verification is exactly specified in advance, in the form of a hypothesis, not only will it be much easier to select verification scores to satisfy each purpose, but there will also be no doubt as to what action is indicated by any numerical value which the verification score may have. It will often be desirable to select the purpose and the score in such a way that the result will either support or reject an *a priori* hypothesis.

Different verification statistics will be required for different purposes. The danger of using any goodness-of-fit index will be illustrated by means of an example.

Table 2.2 shows the results of a model calculation in its second column (Lamb, 1968); this is based on the concept of mass transport balance and takes 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 frequently quoted as a reference. Together with the calculated values, a random estimate (fourth column) and a constant (average) estimate (fifth column) are presented in table 2.2. To obtain the random

Table 2.2 Computed carbon monoxide concentrations (ppm) (17-h day averages) compared with the observed concentration (ppm) for 23 September 1966 for the Los Angeles basin

Station	*Observed concentration*	*Computed concentration (Lamb, 1968)*	*Random*	*Mean 13.5*
Downtown Los Angeles	16	22	9	14
Azusa	13	3	7	13
Pasadena	17	15	15	14
Burbank	16	7	8	13
East Los Angeles	12	5	10	14
West Los Angeles	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

$$\text{computed concentration} = a \times \text{observed concentration} + b$$

estimate, monotonically increasing values from 0 to 17 ppm were assigned in alphabetical order to each station. In 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 RMSE are a caution against validation by just one statistical criterion. The model shows a *higher* RMSE than the (almost) random or the constant value guesses. The correlation coefficient entry rectifies this situation. The constant estimate (a line parallel 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 generalised 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 that the span of the estimation limited. A third point, also linked to a limited span of possible values, is that, when judged by the RMSE, the mean is often a very good bet, in fact better than most calculations. Finally, no validation should be presented without a comparison with the random estimate (the skill score does just this).

The correlation coefficient that measures the association between two sets of values, such as between the observed and computed concentrations, has the advantage (in forecast verification scores) of not influencing the forecaster in an undesirable way. However, it is insensitive to any bias or error in scale, so that it should *not* be used as a calibration procedure.

The standard linear regression procedures for the respective formulae (for example, those of Snedecor and Cochran (1969)) are based on a number of assumptions, the most important of which are the following.

(1) The regression is linear.
(2) The distribution of y for a given x is at least approximately gaussian.
(3) The variance of the departures from the regression line is constant.
(4) The sample observations must be statistically independent for valid significance tests and reliable estimates of the confidence intervals.

Assumption 1 is usually made for reasons of simplicity and convenience; standard methods of curvilinear regression are available.

In regard to assumption 2, it is not necessary that the errors (measurement or computational) are normally distributed in order that the computed sample variance should be an unbiassed estimate of the population variance. Normality is only important for the standard tests of significance.

Instead of assumption 3 other assumptions can be made, such as the variance can be taken to be proportional to the value of the abscissa. However, in this case, different estimating procedures must be used.

Assumption 4 is important, and unfortunately it is very often neglected in air pollution calculation validations. The reliability of statistical estimates depends upon the size of the sample. If the observations are not independent because of spatial or temporal correlation, then the estimated standard errors will tend to be too low. This will produce too many 'significant' results and will make the confidence bands appear narrower than they actually should be.

It should be stressed, however, that graphical checks on the performance of the models are also absolutely necessary. Anscombe (1973) has emphasised this in a striking manner using specific numerical examples. He showed that pure regression analysis can be extremely misleading. He gave a very impressive analysis of what can happen for a purely numerical analysis of data. In table 2.3, 5 artificial data sets of x and y values are given; the sets have the same regression line, $y = 0.5x + 3$, and the same correlation coefficient which is equal to 0.66.

Table 2.3 Sample of data sets; sets from 1 to 4 given by Anscombe (1973)

Set 1		*Set 2*		*Set 3*		*Set 4*		*Set 5*	
x	y	x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58	0	4.4
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76	1	0.7
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	2	3.9
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	2.5	7.1
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	4	2.9
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	7	7
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	10	11
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	12	3
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	14	10
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	15	18
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89	17	8

Figure 2.2a to e shows the extreme contrasts of the actual data characteristics that can be revealed by graphical presentation but which are totally concealed by the above results of standard analytical procedures. Figure 2.2a shows a typical scatter plot, but figure 2.2b to figure 2.2d reveal regularities that may be missed by linear regression analysis, for example, a smooth curve, a different straight line with an outlier or a constant x value with an outlier.

Although these examples can be applied generally by any user of statistical tests, the example of set 6 (figure 2.2f) is that most frequently required by an air pollution modeller. Sometimes the concentration range in space or time is predicted correctly at the ends of a relatively wide range, but there is no event-by-event correlation. This happens, for instance, when there is a station that records very little pollution, while another reports frequent high concentrations; another example is of a minimal average summer concentration which contrasts with a heavy wintertime pollution. Provided that the averages of the extremes are well estimated, even a random value of these means will show a good correlation coefficient. This is illustrated in figure 2.2e where correlation at either end of the scale is non-existent and yet where the overall correlation coefficient is the same as in Anscombe's sets. This point is stressed even more in figure 2.1f where two groups of values were chosen at random with only the limits set beforehand. These limits were $1 \leqslant x \leqslant 3$, $3 \leqslant y \leqslant 6$ and $18 \leqslant x \leqslant 20$, $11.5 \leqslant y \leqslant 12.5$. The correlation coefficient is 0.98.

We may conclude that, if measurements are confined to the usual statistical tests, a large amount of information can be lost. This loss of information tends to emphasise trends in the calculations which might otherwise be eliminated. Therefore we should be cautious about using only a correlation analysis by itself.

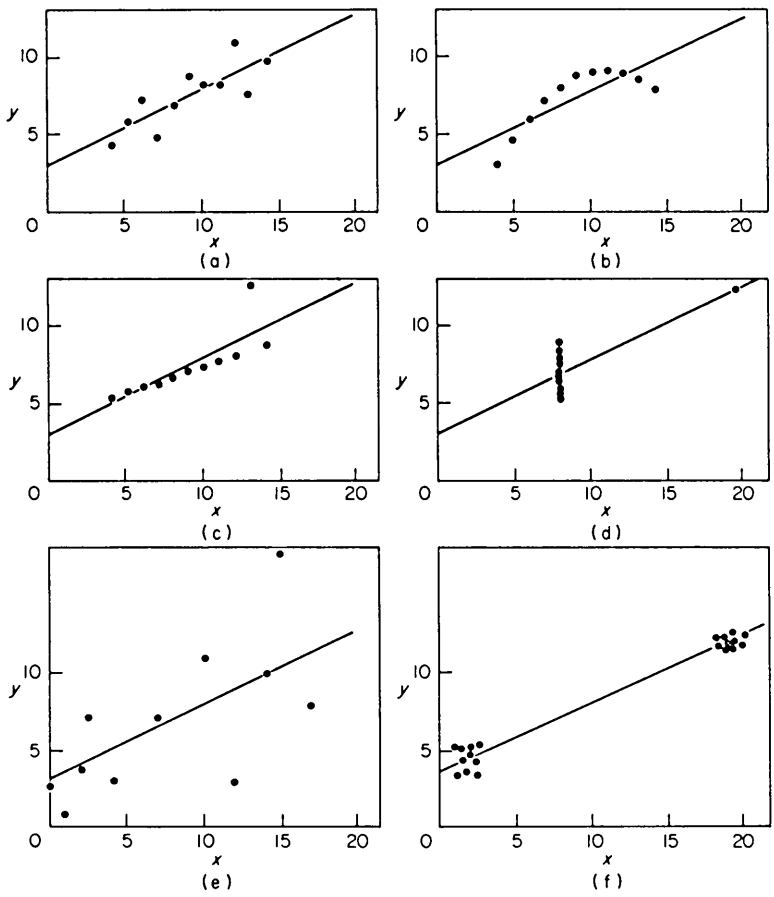

Figure 2.2 Artificial data sets; a to d are from Anscombe (1973); a to e have the same correlation coefficient of 0.66; f has a correlation coefficient of 0.98

References

Anscombe, F. J. (1973). Graphs in statistical analysis. *Am. Stat.*, **27**, 17–21

Benarie, M. (1975). Modelling urban air pollution. *Atmos. Environ.*, **9**, 552–3, discussion

—— (1976). Urban air pollution modelling without computers. *US Environ. Prot. Agency, Publ.*, No. EPA-600/4-76-055, 72 pp. (*Natl Tech. Inf. Serv.*, No. NTIS-PB-262393)

Brier, G. W. (1973). Validity of the air quality display model calibration procedure. *US Environ. Prot. Agency, Rep.*, No. EPA-R4-73-017, 28 pp.

Brooks, C. E. P., and Carruthers, N. (1953). *Handbook of Statistical Methods in Meteorology*, HM Stationery Off., London

Egger, J. (1973). On the determination of an upper limit of atmospheric predictability. *Tellus*, **25**, 435–43

Eriksson, B. (1962). Simple methods of statistical prognoses. *Statistical Analysis in Meteorology, World Meteorol. Organ., Tech. Note*, No. 71, pp. 87–114

Fleming, R. J. (1971). On stochastic dynamic prediction, II, predictability and utility. *Mon. Weather Rev.*, **99**, 927–38

Godske, C. L. (1962). Methods of statistics and some applications to climatology. *Statistical Analysis in Meteorology, World Meteorol. Organ., Tech. Note*, No. 71, 9–86

Hameed, S. (1974). *Atmos. Environ.*, **8**, 555–61

Kao, S. K., and Al-Gain, A. (1968). Large-scale dispersion of clusters of particles in the atmosphere. *J. Atmos. Sci.*, **25**, 214–21

Kao, S. K., and Powell, D. (1969). Large-scale dispersion of clusters of particles in the atmosphere, II, stratosphere. *J. Atmos. Sci.*, **26**, 734–40

Lamb, R. G. (1968). An air pollution model for Los Angeles. *Univ. Calif., Los Angeles, Calif., M.S. Thesis*

Larcheveque, P. (1972). Turbulent dispersion—Eole experiment. *Comm. Space Res. XV, Madrid, Publ.*

Lumley, J. L., and Panofsky, H. A. (1964). *The Structure of Atmospheric Turbulence*, Interscience, New York, 239 pp.

Mahoney, J. R., and Egan, B. A. (1971). A mesoscale numerical model of atmospheric transport phenomena in urban areas. *Proc. 2nd Int. Clean Air Congr., Washington, D.C., 6 to 11 December 1970* (eds H. M. Englund and W. T. Beery), Academic Press, New York, pp. 1152–7

Morel, P. (1970). Satellite techniques for automatic platform location and data relay, *Comm. Space Res. XV, Madrid, Publ.*

Panofsky, H. A., and Brier, G. W. (1968). Some application of statistics to meteorology. *Pennsylvania State Univ., Univ. Park, Pa, Publ.*, 224 pp.

Robinson, G. D. (1967). Some current projects for global meteorological observation and experiment. *Q. J. R. Meteorol. Soc.*, **93**, 409–18

—— (1971). The predictability of a dissipative flow. *Q. J. R. Meteorol. Soc.*, **97**, 300–12

Snedecor, G. W., and Cochran, W. G. (1969). *Statistical Methods*, 6th edn, Iowa State Univ. Press, Ames, Iowa