Environment Protection Engineering

Vol. 23 1997 No. 1–2

HALINA PYTA*

THE HYBRID MODEL IN AIR QUALITY MONITORING AROUND INDUSTRIAL SOURCES

In this paper, an alternative approach to the description of ambient air quality near industrial sources, i.e. the hybrid model containing deterministic and statistical components, has been presented. The assumptions, advantages, and limitations of the modified hybrid model have been discussed. The results of hybrid model verification for various numbers of measurements of sulphuretted hydrogen concentrations from cokery surroundings have been presented. There has been defined the minimum number of 30-min measurements which under similar conditions makes possible estimation of annual average concentration, 99.8th percentile and 30-min maximum.

1. INTRODUCTION

Ambient air quality according to standards required is described by annual, 24-h and 30-min average concentrations. For 24-h and 30-min averaging times the respective 98th and 99.8th percentiles and 24-h and 30-min maxima are limited. Assessment of air quality can be done indirectly (1) by measurement of emission and meteorological paramerters, or directly (2) by measurement of pollutants' concentration (ambient air monitoring). In the first method, we apply deterministic models that use physical basis of pollutants' dispersion. In this case, an inventory of sources is required. The range of the method applicability is practically reduced to the group of technological sources. Inherent feature of such models is inaccuracy of description in the range of high concentrations that appear under extreme conditions of pollutants' dispersion. It results from simplification of equations describing turbulent diffusion phenomena and accompanying processes. Another reason for inaccuracy is the quality of input data, particularly emission data. In the case of strongly variable, estimated emission and chemically active pollutants, the valuation different from annual average concentration should be supplemented by monitoring. Good projection of ambient air conditions can be gained from dense monitoring network. For economical reasons, the number of sampling sites and sampling frequency

* Institute of Fundamentals of Environmental Engineering, Polish Academy of Sciences, ul. Skłodowskiej-Curie 34, 41-800 Zabrze, Poland.

should be limited. The method of results' treatment chosen adequately makes possible precise estimation of air quality parameters for the given number of measurement series. Reliable results are given by the techniques of time series analysis, but their application is limited to separate experiments. In common use there are less complicated, in terms of calculation, methods in which empirical distribution of concentration is approximated by positively skewed density function of probability determined for non-negative values. In conventional analysis, done by National Inspection of Environmental Protection and Regional Sanitary-Epidemiological Stations, the methods based on assumption of lognormal distribution of concentration are used. In this case, an inconformity of hypothetical distribution form, assumed *a priori*, with empirical distribution is a main source of errors.

The hybrid model – synthesis of deterministic and statistic approaches – seems to be an optimal solution. In such a model, a character of distribution is identified based on measuring data. Estimation of distribution parameters, and then evaluation of standardized air quality parameters are done using a part of deterministic model results (censored distribution – between certain percentiles) concentrated around arithmetic mean and predicted with satisfactory accuracy. A precondition of deterministic model application is its calibration. Concentration at a given point, resultant of many factors, is not in this case a random variable only. Experiences with the hybrid model show that it gives good evaluation for high values of concentration and can be used for the description of ambient air quality in the surronding of industrial sources. The above thesis was confirmed by the verification of the simplified model. It allows utilization of the method for routine examination of air monitoring under Polish conditions in relation to national air quality standards, using deterministic component based on algorithm of MAGTiOŚ [3] with no additional requirements for emission and meteorological data.

2. BASIS OF THE HYBRID MODEL

A common procedure for model calibration involves the ordinary least squares fitting of model predictions (deterministic component) versus observations (statistical component). The calibration function can be stated as:

$$\boldsymbol{Y} = a\boldsymbol{X} + b \tag{1}$$

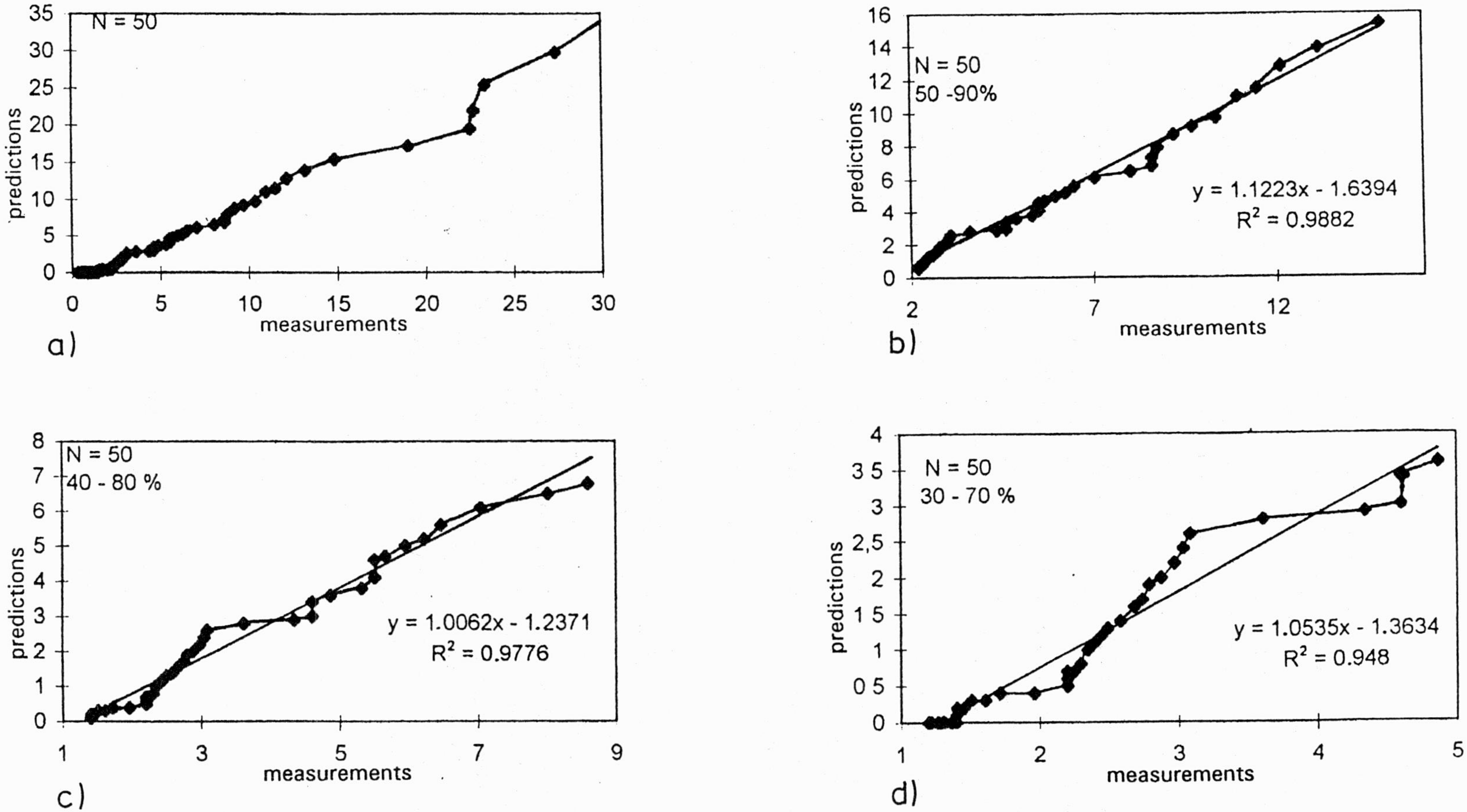
where:

$\boldsymbol{Y}$ – vector of ascending sequence of concentration predictions,

$\boldsymbol{X}$ – vector of ascending sequence of concentration observations,

a and b – coefficients determined by least squares method.

A perfect fit between $\boldsymbol{X}$ and $\boldsymbol{Y}$ for $a = 1$, $b = 1$ is equivalent to linear dependence between predictions and observations. For empirical data, the estimations of a and b give results with error that is a measure of the model adjustment. The first step is to make the vectors $\boldsymbol{X}$, $\boldsymbol{Y}$ comparable. For both sequences, the quantiles from 0.00



Quantile–quantile plots showing a distribution fitting of H_2S concentration, predictions and measurements

$N = 50$ – number of 30-min random measurements in a year

a) in full range, b) in distribution function interval of 50–90%, c) in distribution function interval of 40–80%, d) in distribution function interval of 30–70%

to 1.00 at 0.01 intervals are calculated. Comparison of the successive pairs of quantiles for the same value of empirical cumulative distribution function is shown in the figure (a). The relationship between corresponding quantiles takes linear characteristic in certain interval of variability (calibration interval), where model fitting is the best. In examples quoted in references [4], [5], the calibration interval was assumed either *a priori* or graphically from quantile–quantile plots. In this model, more accurate version, analytical method is proposed. For the given calibration intervals, the coefficients a and b are calculated based on regression. Then the "improved" values of the vector Y of predictions are calculated according to (1). The correlation coefficient R^2 between Y and X is the measure of predictions matching with observations and the basis of calibration interval selection. Basic change in the proposed version of the model applies to its parameter estimation. The tendency towards the simplification of calculating procedure allows us to resign from using information matrix and interval estimation in favour of point estimation with utilization of Bayes' probability elements. The hybrid model due to the calibration coefficients a and b enables us to evaluate the parameters of "improved" censored distribution. Distribution of random variable and calibration equation (1) are known, therefore the transition to "improved" distribution of predictions with linearly shifted parameters is possible. Censored distribution is the particular case of conditional distribution. If A and B are two optional events, the probability $P(A|B)$ with assumption that B has happened is conditional probability of A calculated from:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0. \tag{2}$$

If the random variable X with values from the interval $\langle c, d \rangle$ is a model of the characteristic examined, then transition to censored distribution of X is possible. Censored distribution is a conditional distribution – the event B is equivalent to the reducing variable domain $\langle c, d \rangle$. Let X have cumulative distribution function F, and Z be a censored random variable $(Z = X | c \le X \le d)$ with the cumulative distribution function G. According to (2) the cumulative distribution function of censored distribution is the following:

$$G(z) = P(Z > z | c \le X \le d) = \frac{P(c \le X < z)}{P(c \le X \le d)} = \frac{F(z) - F(c)}{F(d) - F(c)}. \tag{3}$$

If X has the density of probability f then, by virtue of equation (3), the density g of censored distribution Z is equal to:

$$g(z) = \begin{cases} \dfrac{f(z)}{F(d) - F(c)} & \text{for } c \le z \le d \\ 0 & \text{for another } z. \end{cases} \tag{4}$$

Analytical form of censored distribution density can be gained by dividing the primary density distribution f by the difference of cumulative distribution function F at the ends of reduced variable domain. The hybrid model converts censored random variable according to linear transformation. The parameters of "improved" distribution (after calibration) are calculated due to estimated calibration coefficients. The linear shift keeps primary type of the distribution. Therefore, if we use the distribution parameters calculated for given observations' series (statistical component), we are able to make a transition from the censored distribution to the distribution in entire domain. This statement can be justified by the properties of the distribution parameters $E(X)$, $D^2(x)$. The expected value and variance after linear transformation of the variable X are as follows:

$$E(aX+b) = aE(X)+b,$$

$$D^2(aX+b) = a^2D^2(X).$$

3. METHODS AND DATA

The aim of this research was to assess the usability of a modified hybrid model for describing air quality in surrounding of industrial sources and to determine the optimal number of measurements. As a deterministic component, the Pasquill model according to [3] was used, which limits the range of analysis to 30-min standard. For this research, the annual continuous records of H_2S concentration from monitoring station near the Coke Works "Zdzieszowice", averaged for 30-min, have been used. In the Coke Works an inventory of point emitters as well as surface sources, which are difficult to measure because of their changeability over time, was done. Such data base is very useful for verification of the presented model. Basic sequence of predictions was obtained by calculation for all cases of the windrose described with use of data from local anemometer. Two-dimensional windrose (wind speed and direction) was supplemented by occurrence of rates of particular stability classes, averaging for Polish conditions [6]. From complete sequence of 30-min concentration measurements, the random samples of the sizes $N = 50$, 100, 150, 200, 250 and 500 were collected. For each sample size, 30 replications were done. In the Kolmogorov–Smirnov test for a single measurement series, one of the following types of distribution has been determined: exponential, lognormal, Weibull or gamma. Then, using the maximum likelihood method, the distribution parameters, average and variance have been evaluated. In the next step, a calibration inetrval was specified. Concentration predictions were assigned successively to the following cumulative distribution intervals: 50–90%, 40–80% and 30–70%. Then the correlation coefficient R^2 and "improved" values of predictions were calculated. For the calibration, such an inetrval was chosen for which the strongest correlation between prognosis and observations had occurred. Figures b–d (sample consisting of observations) show an influence of calibration interval on correlation coefficient and calibration coefficient values. When coefficients of calibration function and distribution parameters of a sample are known, the parameters of "improved" distribution in entire range of concentration were evaluated. Next, arithmetic mean (measure of annual average concentration), 99.8th percentile, 30-min maximum and relative bias of their estimates b_w were calculated. The relative bias is defined as:

$$b_w = \frac{|x^p - x^*|}{x^*}.$$

The above equation enables us to state what fraction of real air quality parameter x^* from a complete sequence of observations is an absolute bias of its estimate x^p obtained in hybrid model. Mean relative bias $\bar{b}_w$ for 30 replications was chosen as a main criterion of optimum size of measuring series. The measure of an average dispersion of estimates for the given air quality parameter x_i^p close to its real value x^* was the relative root mean square error $\bar{s}$ for $n = 30$ replications:

$$\bar{s} = \left(\frac{1}{n}\sum_{i}^{n} (x_i^p - x^*)^2\right)^{1/2}.$$

This is an additional criterion of optimum size of measuring series.

4. RESULTS

For each air quality parameter estimated, both underestimate and overestimate occurred. Their values depend on variance of the hybrid model (distribution "improved" in entire domain). If a ratio of expected value after calibration to expected value from sample is stable, the variance increases by 2 or 3 times for the sample size $N = 500$ or $N = 50$, respectively. High variance makes impossible a precise estimation of higher concentration values – quantile 0.998 (standard 99.8th percentile) and 30-min maximum (assumed quantile 0.9999 as a measure of maximum) for 50 and 100 measurements during a year. High relative bias values for series size $N = 250$ and $N = 500$ have another reason. An approximate distribution for large number of measurements is close to undetermined distribution of complete measurement sequence (for sample size $N > 500$, some problems with distribution identification occur). The best evaluation of quantile 0.998 and 30-min maximum was obtained for middle size of measuring series, i.e. $N = 150$ and $N = 200$. In general, the mean relative bias for annual average concentration is monotonic, i.e., it decreases with series size increase. For the assumed sizes (N) of measuring series, the mean relative bias and relative root mean square error of annual average concentration, 0.998 quantile and 30-min maximum of concentration estimates are compared (see the table).

Table

Mean relative bias and mean root square error of annual average concentration, 99.8th percentile and 30-min maximum concentration for assumed sizes of measuring series

N	Annual average concentration		99.8th percentile of 30-min concentration		Maximum of 30-min concentration	
	Mean relative bias	Relative mean root square error [$\mu g/m^3$]	Mean relative bias	Relative mean root square error [$\mu g/m^3$]	Mean relative bias	Relative mean root square error [$\mu g/m^3$]
50	0.21	1.22	0.32	40.34	0.31	75.30
100	0.13	0.79	0.39	43.41	0.41	92.18
150	**0.09**	**0.50**	**0.15**	**18.66**	**0.14**	**32.13**
200	0.08	0.46	0.17	19.68	0.15	33.56
250	0.10	0.60	0.36	38.08	0.37	76.34
500	0.05	0.31	0.16	16.40	0.39	102.70

The most precise evaluation of all determined parameters of ambient air quality with least dispersion were gained for the size $N = 150$. The series of 150 30-min measurements during a year is proposed as an optimal sample size.

5. CONCLUSIONS

The proposed modification of the hybrid model is an alternative to traditional methods of description of air quality. It is based upon assumption of lognormal distribution of concentration, particularly in the neighbourhood of technological sources. Under such conditions with characteristic peaks of concentration, this method yields precise estimates of high concentration values. This method makes possible the assessment of ambient air quality for 30-min standard (by evaluation of 99.8th percentile and 30-min maximum) and for annual standard (a measure of annual average concentration is the expected value for the distribution of 30-min predictions). The method does not need determination of background level of pollution. Recalculation of concentration distribution for 24-h averaging time is not possible. Under conditions similar to those analyzed, the hybrid model, which is basd on 150 30-min measurements, allows us to assess the annual average concentration with relative estimation error smaller than 10% and 99.8th percentile as well as 30-min maximum with relative estimation error smaller than 15%.

REFERENCES

[1] Państwowa Inspekcja Ochrony Środowiska, *Zasady projektowania elementów sieci monitoringu zanieczyszczenia atmosfery*, OIKOS, 1991, Warszawa.

[2] Wojewódzka Stacja Sanitarno-Epidemiologiczna w Katowicach, *Zanieczyszczenie atmosfery w województwie katowickim, 1991–1993*, Biuletyn WSSE, 1994, Katowice.

[3] MAGTiOŚ, *Wytyczne obliczania stanu zanieczyszczenia powietrza atmosferycznego*, PZITS, 1981, Warszawa.

[4] TAYLOR J.A. et al., *A hybrid model for predicting the distribution of sulphur dioxide concentrations*, Ecological Modelling, 1987b, 36, 269–296.

[5] JAKEMAN A.J. et al., *Modelling distributions of air pollutant concentrations. III. The hybrid distribution approach*, Atm. Environ., 1988, 22, 163–174.

[6] MAGTiOŚ, *Wytyczne obliczania stanu zanieczyszczenia powietrza atmosferycznego. Katalog danych meteorologicznych*, PZITS, 1979, Warszawa.

MODEL HYBRYDOWY W MONITORINGU POWIETRZA WOKÓŁ ŹRÓDEŁ PRZEMYSŁOWYCH

Przedstawiono alternatywne podejście do opisu imisji wokół źródeł przemysłowych – tzw. model hybrydowy, statystyczno-deterministyczny. Omówiono założenia, zalety i ograniczenia zmodyfikowanej wersji modelu. Zaprezentowano wyniki weryfikacji modelu hybrydowego dla różnej liczby pomiarów stężenia siarkowodoru w sąsiedztwie zakładów koksowniczych. Określono minimalną liczbę pomiarów 30 min., która w podobnych warunkach umożliwia estymację stężenia średniorocznego, percentyl 99,8 i maksimum 30 min.