Environment Protection Engineering

Vol. 21 1995 No. 1–4

HALINA PYTA*, JERZY ZWOŹDZIAK**

A MINIMUM NUMBER OF MEASUREMENTS OF AIR POLLUTANT CONCENTRATION USING QUANTILE–QUANTILE MODEL

Results of experiment aimed at determination of a minimum number of measurements of air pollutant concentration that is required for reliable imission assessment, i.e., within the limits of an error typical of the discussed assessment method and analyzed conditions of concentration changes, have been discussed in the paper. An empirical quantile–quantile model has been used to assess imission recorded in the vicinity of big industrial sources. This model allows estimation of a quantile of 0.998 and maximum of 30-min on the basis of discontinuous series of 30-min measurements at a given point, supported with a more complete (e.g., continuous) measurement series in a similar airshed. The minimum number of measurements has been determined using analysis of variance of the estimation errors of the aforementioned parameters that accompany different sizes of measurement series. The method allowing determination of minimum measurement number has been illustrated by the example for total suspensed particulate (TSP), sulphur dioxide (SO_2) and the total hydrocarbons (THC) whose concentrations have been established based on an automatic monitoring in the region of Kędzierzyn-Blachownia.

1. INTRODUCTION

The imission assessment on the basis of discontinuous measurement series is carried out assuming that the obtained sequence of results is a realisation of a random variable generated by a stochastic process. Pioneer papers of JUDA et al. from the close of the 60-ties [1], [2] and the research of LARESEN [3], [4] and BENARIE [5] carried out simultaneously have shown that the concentration distribution at a given point can be quite well described by a log-normal function. This view was verified in the 70-ties. It has appeared, especially in the vicinity of industrial sources, that the concentration distribution can be of various character and that, in principle, can be

---

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

** Faculty of Environmental Engineering, Technical University of Wrocław, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.

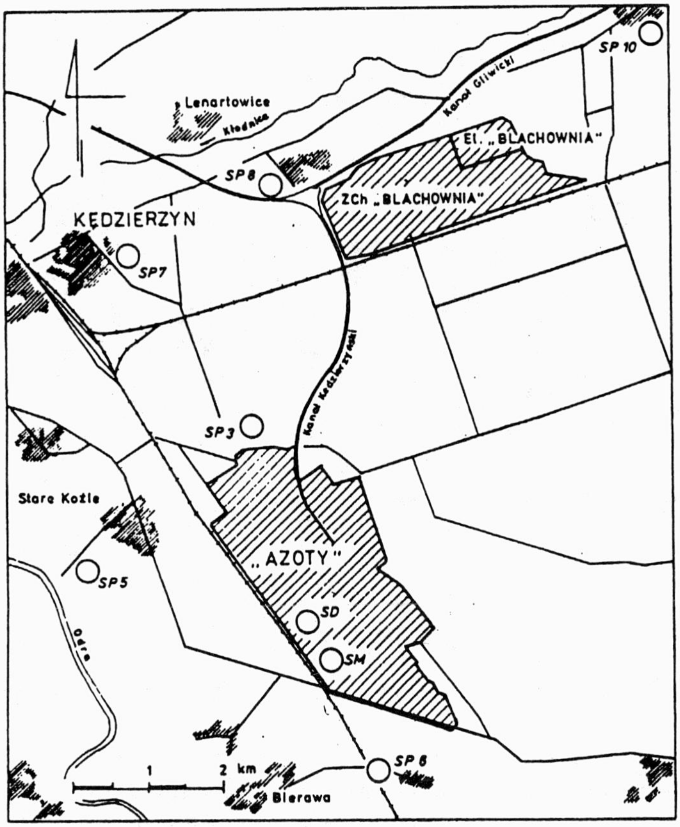
approximated by virtually any distribution density function determined in the range $\langle 0, +\infty)$, most often uni-modal and positively skew. The measurement data has been interpreted either using complicated function systems, consisting of superposition of many elementary distributions, or using a general definition of a functions' family, e.g., Pearson's family, at an extended identification procedure of the specific form of the distribution. The trend related mainly to variation of meteorological factors and the effect of autocorrelation, significant at short averaging time, make that the estimation of imission parameters different from the average annual concentration is sometimes inaccurate. Relatively accurate estimations for higher concentrations are obtained using the methods typical of time series processing, e.g., ARIMA model, Kalman's filter. Common application of these methods to typical analyses was probably limited by the complexity of computing procedures. A simple empirical quantile–quantile model [6] allowing extrapolation of the results of discontinuous series of measurements of pollution concentration into the total measurement period with a support of more complete series of measurements from another point with similar (verified in the model) imission conditions has been used to assess the imission discussed in this paper. Application of this method to determine a representative measurements' number is justified because it requires neither complete identification of the distribution nor *a priori* assumption of its general form. So the estimation of the minimum measurements' number is accurate and we do not deal with the error resulting from misfit of the hypothetical distribution form to the observation results.

2. SCOPE OF INVESTIGATIONS

The considerations are related to conditions occurring in the vicinity of technological sources, i.e., strongly variable imission field, and 30-min averaging time with the highest dynamics of variations. The assessment of imission in the 30-min standard category requires determination of frequency of exceeding an appropriate standard level and the maximum value. The standard is considered to be met if the maximum does not exceed a double value of the permissible concentration and the frequency of occurrence of concentration higher than the standard does not exceed 0.2% of all the time in a year. The quantile–quantile method makes possible determination of any quantile that does not occur among the values of empirical cumulative distribution function at the given number of measurements. The objective of the experiment was to determine the minimum number of random series of 30-min measurements enabling estimation of standard quantile 0.998 and a quantile 1.000 being 30-min maximum. To avoid controversies over the choice of measurement dates (related to daily and seasonal variability) a more universal approach has been chosen and a random scheme of measurements' performance has been assumed. The problem was being solved by partial numerical simulation of selected numbers of observations. The relative error accompanying the given number of observations, expressed by the ratio of the absolute value of quantile estimator deviation to its real value, was used as the criterion limiting the number of random measurements. The analysis of variance with simultaneous testing of contrasts for pairs of objects was applied when the minimum size of the measurement series (model calculations cover only some sector of the population) was established.

3. CHARACTERISTICS OF DATA SET

The data set was composed of the results of measurements of concentrations of TSP, SO_2 and THC (averaged out for 30-min intervals) from six automatic monitoring stations situated in the region of Kędzierzyn-Blachownia. The period of investigations lasted from January 1st to December 31st, 1993. The Automatic



Location of measurement stations (SP3–SP10) of monitoring network in the region of Kędzierzyn-Blachownia

Imission Control System BASKI covers the area affected directly by the Nitrogen Works "Kędzierzyn", the Chemical Works "Blachownia" and the Power Station "Blachownia". The pollutants taken to be analyzed, selected from seven ones recorded in the monitoring network, guarantee the highest completeness of data, wide diversity of physical and chemical properties and are composed of commonly emitted substances. The location of both the network and individual stations is shown in the figure, where SP3–SP10 are imission control stations, SM is the meteorological station, and SD – dispatcher's station. The general characteristics of data, including basic parameters of statistics of annual continuous measurement series of TSP, SO_2 and THC concentrations in individual stations, have been compiled in table 1.

Table 1

Basic statistics of 1-year continuous records of concentrations (30-min averages) for TSP, SO_2 and THC in stations examined

Station label	Mean concentration	Mean standard deviation	Minimum	Maximum	Quantile 0.998	Number of observations
			TSP [μg/m^3]			
SP3	58.80	53.98	0.10	494.50	387.90	17427
SP6	60.23	54.69	0.10	499.60	373.60	17094
SP7	56.75	56.40	0.10	596.80	439.80	17338
SP8	58.71	57.57	0.10	585.00	416.40	17161
			SO_2 [μg/m^3]			
SP3	41.05	50.70	0.10	424.80	315.50	16924
SP5	46.99	60.74	0.10	446.80	366.90	16046
SP6	58.19	61.96	0.10	483.10	379.50	17063
SP7	57.37	63.41	0.10	444.40	365.10	17078
SP10	60.80	68.41	0.10	499.50	418.20	16954
			THC [ppm]			
SP3	0.53	0.38	0.00	2.20	2.00	15221
SP6	0.58	0.42	0.00	2.10	2.00	15283
SP7	0.73	0.50	0.00	3.60	3.30	15438
SP10	0.75	0.47	0.00	2.40	2.29	13825

4. DESCRIPTION OF QUANTILE–QUANTILE MODEL

Two sets of observations are given: $x_i(i = 1, \ldots, n)$ and $y_j(j = 1, \ldots, m)$. Denote the respective quantiles of their empirical distribution function as Q_x and Q_y. Then the function $Q_y(p)$ versus $Q_x(p)$ expresses the so-called empirical relation of the quantile–quantile type for the same fractions of the data p, defining the quantile order in the range of variability from 0 to 1. When the distributions of both series are identical, their distribution functions are also identical and distribution similarity can be directly transferred onto the relation describing a congruence of quantiles of both distributions as $y = x$. CHAMBERS et al. [7] have extended this simple model to a more general form, assuming linear relation between quantiles of the same order:

$$Q_y(p_i) = \alpha Q_x(p_i) + \beta \tag{1}$$

where α, β are the coefficients estimated from the data by means of the least-square error method.

In the case of sets of equal numbers of observations, arranged in ascending order, the empirical quantile–quantile relation describes the relation between consecutive, matching quantiles. In our case, when one set of data (hereafter called "the analyzed" – (y_j)) is definitely less numerous than the other (called "the basis" – (x_i)), the maximum information contained in the set analyzed should by used and all sorted values y_j should be applied in interpolation of matching quantiles x_i from the more numerous basic set. In order to determine the proper quantiles in the basic set of data, the fraction p_i should be estimated. Each of the values of less numerous data sets corresponds to this fraction. Hence:

$$p_i = (i - 0.5)/n \tag{2}$$

where n denotes the number of observations in the sample.

If the analysed sequence of data and the basic sequence we denote as $\{y_j\}$ and $\{x_i\}$, respectively, then the value y_j, which is the $((j - 0.5)/m)$-th quantile of data y, is linked with the interpolated $((j - 0.5)/m)$-th quantile of the set of data x. Hence, the order statistics v in the basic set searched after interpolation is determined in the following way:

$$p_j = p_v . \tag{3}$$

From (2) and (3):

$$(j - 0.5)/m = (v - 0.5)/n, \tag{4}$$

hence:

$$v = (n/m)(j - 0.5) + 0.5 . \tag{5}$$

If v is not an integer, then it should be divided into an integral component i and a decimal one θ. Estimated (interpolated) quantile $Q_x(p_j)$ is calculated according to the following equation:

$$Q_x((j - 0.5)/m) = (1 - \theta)x_i + \theta x_{i+1} . \tag{6}$$

When $Q_y(p_j)$ (ascending sequence of values (quantiles) in the set analyzed) as well as $Q_x(p_j)$ (the sequence of values (quantiles) calculated according to (6)) have been already given, then the slope α and the abscissa β from equation (1) are known from a simple linear regression. Knowing α, β at given $Q_x(p_i)$ (ascending sequence of values (quantiles) from the basic set) we determine the quantile $Q_y(p_i)$ according to (1). The order p_i of the quantile cannot be found among the values of the empirical distribution function of the less numerous series analyzed. The only assumption accepted during model construction is the similarity of concentration distribution in both measurement series, i.e., the analyzed and the basic series. Whatever way of similarity verification we choose, the identification of distributions is not required. When there are given two independent random samples with numbers of observations n and m representing a population of distribution functions (theoretical) $F_n(x)$ and $F_m(x)$, then the following hypothesis H_0 can be formulated:

$$F_n(x) = F_m(x) \quad \text{for every } x. \tag{7}$$

The verification of the null hypothesis takes place as a result of application of one of nonparametric tests for two independent samples, e.g., Kolmogorov–Smirnov test carried out to calculate the following statistics:

$$D_{n,m} = \max |[S_n(x) - F_n(x)] - [S_m(x) - F_m(x)]| \tag{8}$$

where:

$S_m(x)$ – the empirical distribution function for data y_j $(j = 1, \ldots, m)$ from the less numerous analyzed set,

$S_n(x)$ – the empirical distribiution function for data y_i $(i = 1, \ldots, n)$ obtained from the basic set x_i and modified according to equation (1).

5. PLAN OF THE EXPERIMENT

Because of the technical limitations we had to choose the pairs of stations analyzed in the experiment discussed. The first selection has been carried out on the basis of similarity of concentration distributions for complete, annual measurement series in individual points. It has been assumed that among the pairs representing the continuous measurement series, for which fitting of distributions has been proven, it will be easier to verify this fitting for their random discontinuous subsets (analyzed series) and a continuous basic series. Using the Kolmogorov–Smirnov test the hypothesis of similarity of distributions for each two measurement stations (from six) has been verified at the significance level 0.01. Some of pairs of "similar" stations have been taken for analysis. The choice of these pairs (special attention being paid to the basic station) resulted from location of individual points in relation to the emission sources and from taking account of the wind rose.

In the model area, the highest frequency of winds is observed in sectors 5–7 (the angle range of 12-directional wind rose is 150°–120°); winds blow predominantly from the south. For this reason the locations SP3, SP7 and SP8 are greatly affected by the Nitrogen Works "Kędzierzyn", while the location SP10 – by the Nitrogen Works "Kędzierzyn" and the Chemical Works "Blachownia". The highest concentrations are observed in these points. So it has been decided that stations SP3, SP8 and SP10 will be basic stations (continuous observations) for the analyzed sequence of data from other stations. In the case of dust (TSP), the series from SP6 and SP7 stations have been analyzed; as the basic series we used a complete (annual) sequence of measurements carried out in SP3 station and in SP8 station. In the case of SO_2, the series from SP5 station (the basic station SP3) and series SP6 and SP7 (the basic station SP10) have been analyzed. In the case of hydrocarbons, the series from stations SP6 and SP7 have been analyzed (the respective basic stations SP3 and SP10).

Continuous series of 30-min measurements recorded throughout the year allowed obtaining of discontinuous series which were based on the random samples of sizes $N = 8500$, 4500, 1500, 750, 500, 250 and 100. For each series of a given number of observations, quantiles 0.998 and 1.000 of the empirical distribution function and a relative error (BL) of this quantile estimation have been calculated according to the procedure presented in the preceding paragraph. The relative error can be expressed as:

$$BL = \frac{|KW_p - KW_o|}{KW_o} \times 100\% \tag{9}$$

where:

KW_p – predicted quantile 0.998 (1.000), quantile estimator calculated according to the model,

KW_o – quantile 0.998 (1.000) calculated for a given series of measurements (an actual value).

Actual values KW_o of quantiles 0.998 and 1.000 have been compiled in table 1. This procedure has been repeated in 10 replications for each number of observations of the series analyzed; the model assumption was verified according to (8) every time.

6. RESULTS

Values of relative error (BL), averaged for 10 replications, have been compiled in table 2. The analysis of the BL results has not allowed an explicit evaluation of error accompanying the predetermined numbers of measurement series of individual pollutants and estimated parameters. For the given pair of stations the increase in average estimation error with the decrease in number of observations in series has often been disturbed. Therefore the assumption that averaging the values of relative error for a given number of observations include not only the examined pairs of stations, but also the adjoining numbers of observations in measurement series is reasonable. As there was a necessity to test the hypothesis of equality of object means (equality of $BLSR$ for individual numbers of observations) that created the basis for testing the detailed hypotheses concerning the so-called contrast, the application of one-way analysis of variance (ANOVA) seemed obvious. To that end the values of relative error BL recorded in all replications and examined pairs of stations

Table 2

Summary specification of mean estimation errors of 0.998 quantile and 30-min. maximum according to quantile–quantile model for the chosen numbers (*N*) of random 30-min. measurements in stations "analyzed"

N	TSP								SO_2						THC			
	Mean estimation errors of 0.998 quantile for the following pairs of stations				Mean estimation errors of maximum for the following pairs of stations				Mean estimation errors of 0.998 quantile for the following pairs of stations			Mean estimation errors of maximum for the following pairs of stations			Mean estimation errors of 0.998 quantil for the following pairs of stations		Mean estimation errors of maximum for the following pairs of stations	
	SP3_6*	SP3_7	SP8_6	SP8_7	SP3_6	SP3_7	SP8_6	SP8_7	SP3_5	SP10_6	SP10_7	SP3_5	SP10_6	SP10_7	SP3_6	SP10_7	SP3_6	SP10_7
8500	3.14	7.01	4.49	4.79	1.91	12.17	9.11	1.29	1.94	0.97	3.76	12.70	6.54	1.66	3.87	3.02	6.68	5.45
4500	2.97	8.67	5.30	4.90	3.03	13.81	9.99	3.82	1.92	1.32	3.71	12.42	7.57	1.83	6.28	3.97	11.33	7.75
1500	4.34	9.16	7.51	4.07	3.87	13.95	11.54	4.03	2.67	3.49	5.24	10.98	6.06	3.07	6.77	5.73	11.89	7.12
750	5.68	7.09	3.02	7.33	6.23	10.35	6.68	7.29	4.59	3.00	3.65	12.04	5.57	3.86	7.32	7.19	11.49	6.98
500	2.88	11.71	9.37	4.74	4.79	15.88	12.68	6.41	5.05	7.20	4.06	11.38	8.63	4.63	8.78	8.11	14.02	9.72
250	12.88	13.09	10.70	10.32	12.67	14.39	13.38	10.62	11.48	10.67	6.80	16.39	10.11	6.30	10.28	12.25	14.63	14.67
100	21.19	23.60	17.25	21.26	18.91	26.67	20.66	23.37	23.64	13.43	9.60	28.68	12.96	8.30	13.41	17.32	18.89	19.49

* Example of label: SP3_6 – the first number (i.e. 3) in this label is notation of "basic" station, the second number (i.e. 6) – "analyzed" station with discontinuous measurements.

separately for quantile 0.998 and the 30-min maximum of each pollutant have been combined. Six sets of error observations, classified according to number of measurement series, have been obtained. The analysis of variance of one-way classification model, in which numbers of random series of concentration measurements were objects, has been carried out for each of these sets. In the analysis of variance of this model in the standard version, the following assumptions should be satisfied: normal distribution of observations, independence and homogeneity of observation variance for individual objects. In all cases, first two assumptions proved to be reasonable. But Cochran and Bartlett tests have shaken the assumption about the homogeneity of variance. The analysis of variance for classification according to number of random measurements has been carried out following CHUDZIK's and MARKIEWICZ's [8] algorithm and using the relations quoted in the Appendix. The hypothesis of equality of object means has been rejected for all cases at the significance level below 0.0001. Detailed analyses, in which confidence intervals for simultaneous testing of contrast between individual pairs of objects were used, allowed separation of homogeneous groups of objects (groups of measurements' number) at the significance level of 0.05. The differences within these groups are considered to be accidental, while the differences between average values of relative error in adjoining groups are statistically significant. On the basis of the above, the *BL* errors recorded for numbers of random measurements in individual groups have been averaged. It has been assumed that the smallest number of measurement series in a given group provides a sufficient certainty of preserving average (in this group) accuracy of estimation. Table 3 contains the results of variance analysis; homogeneous groups being marked. These results enable determination of a minimum measurement number. The obtained division into groups can be interpreted as follows: the minimum number of measurements (the smallest in a given group) allows us to obtain such estimators of 0.998 quantile and the maximum which after being increased by percentage of relative error characteristics of this number of measurements (average in the group) may be directly compared with the air quality standard. Permissible 30-min concentration D_{30} can be considered as a proper, provided that the sum of 0.998 quantile estimator and the aforementioned correction is smaller than D_{30} and the sum of the maximum estimator and appropriate correction is smaller than $2D_{30}$.

The analysis of table 3 leads to the following conclusion: division into homogeneous groups for estimation error of 0.998 quantile basically coincides with such a division for estimation error of 30-min maximum. This consistence is desirable because it makes possible estimation of both parameters with a similar accuracy at the same minimum number of random measurements. The error of estimation of 30-min maximum for individual numbers of measurement series is higher than that accompanying the estimation of 0.998 quantile. The difference between the aforementioned errors is the slightest in the same homogeneous groups and disappears with decreasing number of observations. The probability of assess-

Table 3

Summary specification of the results of variance analysis for relative errors (*BL*) at one-way classification according to the numbers of measurements (*N*)

TSP				SO_2				THC			
Estimation error of 0.998 quantile											
Number of homogeneous groups	Range of series numbers *N* in the following groups	Range of error *BL* [%] in the following groups	Mean estimation error in the following groups	Number of homogeneous groups	Range of series numbers *N* in the following groups	Range of error *BL* [%] in the following groups	Mean estimation error in the following groups	Number of homogeneous groups	Range of series numbers *N* in the following groups	Range of error *BL* [%] in the following groups	Mean estimation error in the following groups
3	(I) 8500, 4500	4.8–5.5	5.2	3	(I) 8500–1500	2.2–3.8	2.8	4	(I) 8500	3.4	3.4
	(II) 1500–250	5.8–11.4	7.7		(II) 750–250	3.7–9.7	6.3		(II) 4500–750	5.1–7.2	6.2
	(III) 100	20.8	20.8		(III) 100	15.6	15.6		(III) 500, 250	8.5–11.3	9.9
	–	–	–		–	–	–		(IV) 100	15.4	15.4
Estimation error of 30 min. maximum											
3	(I) 8500, 4500	6.1–7.6	6.8	3	(I) 8500–750 (min. numb. 1500)	6.7–7.3	7.0	4	(I) 8500	6.1	6.1
	(II) 1500–250	7.6–12.8	9.7		(II) 500, 250	8.2–10.9	9.6		(II) 4500–750	9.2–9.5	9.4
	(III) 100	22.4	22.4		(III) 100	16.6	16.6		(III) 500, 250	11.9–14.7	13.3
	–	–	–		–	–	–		(IV) 100	19.2	19.2

ment of goodness of fit reduces with decreasing number of observations, especially for higher concentrations of the discontinuous series and complete basic series. Therefore the estimation error of both parameters in the upper concentration interval is significant and similar. This conclusion is related to the comment on the last homogeneous group, identical for all cases and containing only one number of measurements –100. High average relative error (15.6%–22.4%) occurs in this group which suggests that 100 random measurements per year may lead to overestimation of exceeding frequency; therefore we should not use the minimum number of measurements when the archive data reveal frequent occurrence of concentrations exceeding standards.

Deviations from the mentioned uniform division into homogeneous groups for the error of both parameters being estimated, observed for SO_2, needs commenting. The analysis of variance carried out for the error of the maximum estimation made possible separation of three homogeneous groups. The lower limit of the first group is made by the number of observations reaching 750 and not by 1500, as for the 0.998 quantile estimation error. In order to avoid a paradoxical statement that 0.998 quantile estimation requires more numerous minimum series of measurements than the maximum estimation, the number of observations of 1500 has been decided to be the minimum in group I. At the same time the values of estimation errors in groups I and II were unchanged. Such a treatment justifies the result of testing the contrasts between groups I and II of the maximum error. This procedure is nearly off limits of correctness.

Compared with dust and SO_2, hydrocarbons created a special category which was divided into 4 homogeneous groups. It is worth noting that the minimum number of the measurements of hydrocarbon concentrations, which allows us to maintain the accuracy on a level of about 3% for 0.998 quantile and 6% for the maximum (i.e., 8500), is approximately 5.5 times higher than for SO_2 (1500) and twice as high as for TSP (4500). The best relation between the size of series and accuracy of estimation has been observed for hydrocarbons. This example firmly corroborates the opinion that the chemical character of pollutants can be fundamental in planning measurements, because it often causes strong seasonal imissions and hinders imission assessment.

7. SUMMARY

Under conditions characteristic of the vicinity of industrial sources, similar to those analyzed in this experiment (table 1), it is possible to obtain considerably accurate estimations of 0.998 quantile and 30-min maximum using the quantile–quantile model. The minimum numbers of 30-min random measurements, necessary to maintain the relative estimation error below 10%, amount to 250 for the TSP and SO_2 and 750 for THC. For hydrocarbons it is possible to reduce the measurement number to 250, but the fulfilment of the above accuracy condition becomes impossible at 30-min maximum estimation when it is necessary to extend the tolerance range to 15%. The reduction in minimum number of observations to about 100 measurements per year gives the values associated with an error of approximately 20% and should be used cautiously; we should be aware of the consequences derived from interpretation of the result in the regions exposed to concentrations exceeding the standards. However, further reduction of relative error by 30–50% compared with the error in homogeneous group II requires as large number of observations as 1500 for SO_2 and 8500 for hydrocarbons, which is economically unreasonable at usual requirements.

In order to examine the statistical significance of differences between average estimation errors of the parameter at a given number of measurements in individual pairs of stations (table 2), verifying the assumption that linking of observations is reasonable, the variance analysis of errors for classification according to label of stations' pairs has been carried out. The model of one-way analysis in the standard version, i.e., at valid assumption of variance homogeneity [9], [10], has been used separately for each of 7 assumed numbers of random measurements. It has been found that these differences are statistically insignificant at 500 and less 30-min measurements throughout the year, thus on the level of recommended above minimum numbers of measurements. Hence the choice of the basic station (its location in relation to the emission source and the station analyzed) does not affect substantially the magnitude of estimation error if the assumption of the quantile–quantile method is valid. The basic station does not need to be situated in the closest neighbourhood.

8. FINAL REMARKS

The paper presents the methodology of determining the minimum number of random measurements based on the quantile–quantile model which is appropriate for the estimation of occurrence frequency of the highest concentration values. This methodology differs from JUDA's et al. [11], [12] approach who estimated the minimum number of measurements by interval estimation of annual average concentration.

The methodology has been verified under extreme conditions, i.e., in close vicinity of industrial sources and for the shortest averaging time, and has been illustrated with an example of the database from the automatic imission control network in Kędzierzyn-Blachownia. Minimum numbers of random 30-min measurements for SO_2, TSP and THC have been determined. They can make a starting point in planning measurements under similar imission conditions.

Essential advantages of the quantile–quantile model such as high accuracy of estimation at a relatively low (in the conditions discussed) measurement number and extreme simplicity offer sufficient prerequisites for broader application of this model, say, in restructuring the existing monitoring network and in designing the network elements. Due to application of that model we are able to make a spatial interpolation of air pollution data – a problem especially significant in the regions of high saturation with diversified emission sources, e.g., in the Upper Silesian Agglomeration (possibility of using 10 automatic monitoring stations).

The disadvantage of the quantile–quantile model used is its poor functional quality limited to the measurement averaging time. It is not possible, as in Juda's case, to recalculate the distribution parameters into 24-h averaging time. Application of the aforementioned method requires the existence of "similar", more complete basic series with possibly low level of discontinuity because this affects accuracy of estimation.

APPENDIX

The analysis of variance has been carried out using F^* statistics according to BROWN and FORSYTH [13]:

$$F^* = \frac{\sum\limits_{i=1}^{k} n_i (Y_i\cdot - Y\cdot\cdot)^2}{\sum\limits_{i=1}^{k} \left(1 - \frac{n_i}{N}\right) s_i^2} \tag{1}$$

where:

$$Y_i\cdot = \sum_{i=1}^{n_i} y_{ij}, \quad Y\cdot\cdot = \frac{1}{k}\sum_{i=1}^{k} y_i\cdot, \quad N = \sum_{i=1}^{k} n_i,$$

$$s_i^2 = \frac{\sum\limits_{j=1}^{N} (y_{ij} - Y_i\cdot)^2}{(n_i - 1)}, \quad i = 1, 2, \ldots, k, \tag{2}$$

k – the number of objects,

n_i $(i = 1, 2, \ldots, k)$ – the number of i-th object observations (replications),

y_{ij} $(i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, n_i)$ – j-th observation of i-th object.

Statistics (1), provided that the hypothesis of equality of object means is accepted, has an approximate F-Snedecor distribution with $k-1$ and f degrees of freedom, while the number of degrees of freedom f is approximated according to the Satterthwaite formula:

$$f = \sum_{i=1}^{k} \frac{l_i^2}{n_i - 1}$$

where:

$$l_i = \frac{\left(1 - \frac{n_i}{N}\right) s_i^2}{\sum\limits_{i=1}^{k} \left(1 - \frac{n_i}{N}\right) s_i^2}.$$

For simultaneous pairwise comparison of means the T2 method, based on Šidak inequality described by TAMHANE [14], has been used. According to this method $(1-\alpha)\,100\%$ confidence intervals for all possible comparisons $\mu_i - \mu_j$ $(i, j = 1, 2, \ldots, k;\ i < j)$ of pair of means have the following form:

$$\mu_i - \mu_j \in \left(Y_i\cdot - Y_j\cdot - t_{\gamma,v_{ij}} \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}},\quad Y_i\cdot - Y_j\cdot + t_{\gamma,v_{ij}} \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}} \right)$$

where $Y_i\cdot$ and s_i^2 are given by formulae (1) and (2), $t_{\gamma,v_{ij}}$ is the critical value of t-Student distribution on significance level γ:

$$\gamma = 0.5\left[1 - (1-\alpha)^{2/k(k-1)}\right]$$

for $v_{ij} = n_i + n_j - 2$ degrees of freedom, where α is a pre-set significance level for all the test.

REFERENCES

[1] JUDA J., *Planung und Auswertung von Messungen der Verunreiningungen in der Luft*, Staub. Reinhalt. Luft., 1968, 5, 186–192.

[2] JUDA J., JĘDRZEJOWSKI J., NOWICKI M., *Metodyka pomiarów zanieczyszczeń powietrza atmosferycznego*, Prace i Studia Zakładu Badań Naukowych GOP PAN, 1969, zeszyt 1.

[3] LARSEN R., *Analyzing Air Pollutant Concentration and Dosage Data*, J. Air Pollut. Control ASS., 1967, 17, 85–93.

[4] LARSEN R., *A New Mathematical Model of Air Pollutant Concentration Averaging Time and Frequency*, J. Air Pollut. Control ASS., 1969, 19, 24–30.

[5] BENARIE M., *The Use of the Relationship Between Wind Velocity and Ambient Pollutant Concentration Distribution for Estimation of Average Concentrations from Gross Meteorological Data*, Proc. Symp. Statistical Aspects of Air Quality Data, 1974, EPA 650/4–74–038.

[6] TAYLOR J. et al., *Statistical Modelling of Restricted Pollutant Data Sets to Assess Compliance with Air Quality Criteria*, Environmental Monitoring and Assessment, 1987, 9, 29–46.

[7] CHAMBERS J. et al., *Graphical Methods for Data Analysis*, Duxbury, 1983, Boston.

[8] CHUDZIK H., MARKIEWICZ A., *Wielokrotne porównania średnich przy braku jednorodności wariancji*, Roczniki AR w Poznaniu, Algorytmy Biometryczne i Statystyczne, 1984, **11**, 125–135.

[9] AHRENS H., *Analiza wariancji*, PWN, 1970, Warszawa.

[10] OKTABA W., *Metody statystyki matematycznej w doświadczalnictwie*, PAN, 1986, Warszawa.

[11] JUDA J., *Design of Air Quality Monitoring Networks*, Paper No. 18 in Design of Environmental Information Systems. Ann Arbor Science Publishers Inc., Michigan 1974, 347–356.

[12] JUDA J., CHRÓŚCIEL S., *Ochrona powietrza atmosferycznego*, Wyd. Nauk.-Techn., 1974, Warszawa.

[13] BROWN M.B., FORSYTHE A.B., *The Small Sample Behaviour of Some Statistics which Test the Equality of Several Means*, Technometrics, 1974b, **16**, 129–132.

[14] TAMHANE A.C., *A Comparison of Procedures for Multiple Comparisons of Means with Unequal Variances*, Journal of the American Statistical Association, 1979, **74**, 471–480.

MINIMALNA LICZBA POMIARÓW STĘŻENIA ZANIECZYSZCZEŃ POWIETRZA Z UŻYCIEM MODELU KWANTYL–KWANTYL

Omówiono rezultaty eksperymentu, którego celem było określenie minimalnej liczebności serii pomiarów stężenia zanieczyszczeń wymaganej do wiarygodnej oceny imisji, tj. w granicach błędu typowego dla omawianej metody oceny i analizowanych warunków zmian stężenia. Aby w kategoriach normy 30-min ocenić imisję w sąsiedztwie dużych zakładów przemysłowych, wykorzystano empiryczny model kwantyl–kwantyl. Model ten umożliwia oszacowanie kwantyla rzędu 0,998 i maksimum 30-min na podstawie nieciągłej serii pomiarów 30-min w danym punkcie i bardziej kompletnej (np. ciągłej) serii pomiarów w podobnym polu imisji. Minimalną liczbę pomiarów ustalono korzystając z analizy wariancji błędu oceny wspomnianych parametrów towarzyszącemu różnej liczebności serii pomiarowych. Proponowaną metodykę zilustrowano przykładem dla pyłu zawieszonego, dwutlenku siarki i sumy węgolowodorów uzyskanym na podstawie danych z automatycznego systemu monitoringu w rejonie Kędzierzyna-Blachowni.

МИНИМАЛЬНОЕ ЧИСЛО ИЗМЕРЕНИЙ КОНЦЕНТРАЦИИ ЗАГРЯЗНЕНИЙ ВОЗДУХА С УПОТРЕБЛЕНИЕМ МОДЕЛИ КВАНТИЛЬ–КВАНТИЛЬ

Обсуждены результаты эксперимента, целью которого было определение минимальной численности серии измерений загрязнений, требуемой для верной оценки имиссии, т.е. в пределах ошибки типичной для обсуждаемого метода оценки и анализируемых условий изменений концентрации. Чтобы в категориях нормы 30-мин оценить имиссию в близости больших промышленных предприятий, была использована эмпирическая модель квантиль–квантиль. Эта модель дает возможность оценки квантиля порядка 0,998 и максимума 30-мин на основе прерывистой серии измерений 30-мин в данном месте и более комплектной (напр. непрерывистой) серии измерений в похожем поле имиссии. Минимальное количество измерений было установлено с использованием анализа варианции ошибки оценки вышеупомянутых параметров, сопутствующей разной численности измерительных серий. Предлагаемая методика была проиллюстрирована примером для взвешенной пыли, двуокиси серы и суммы углеводородов, полученным на основе данных из автоматической системы мониторинга в районе Кендзежына–Бляховни.