Environment Protection Engineering

Vol. 34 2008 No. 4

PRZEMYSŁAW SZCZYGŁOWSKI*, MARIAN MAZUR*

APPLICATION OF BOOT STATISTICAL PACKAGE IN CALCULATING POLLUTANT SPREADING IN AIR

The issue of modelling the pollutant spreading in air requires a number of decisions to be taken in order to ensure a proper selection of the model option in terms of meteorology or dispersion coefficients. The difference in the results of the calculations of the pollution conditions is observed after each change of option.

Negative and positive components of Fractional Bias (FB) or Geometric Mean Bias (MG) may prove to be especially useful, relatively new statistical indicators. They allow the overestimation or underestimation of the results in the case of calculating the pollution distribution within the specified area to be evaluated.

1. BASIC STATISTICAL INDICATORS

The procedure used to compare the individual options of models is based on statistical indicators developed for the models of pollutant spreading. These indicators were gathered in the BOOT Statistical Model Evaluation Software Package, Version 2.0 [1]. The basic statistical indicators used in the comparisons were recommended by EPA (Environmental Protection Agency), and then collected in the BOOT package [2], [3]. The basic statistical indicators are: *FB* – Fractional Bias, *MG* – Geometric Mean Bias, *NMSE* – Normalized Mean Square Error, *VG* – Geometric Variance, *R* – Correlation Coefficient, *FAC*2 – indicator characterising model quality. Aside from the six basic indicators defined above, the following additional indicators were determined as well: FB_{FN} – negative component *FB* (false-negative,), FB_{FP} – positive component *FB* (false-positive), negative (MG_{FN}) and positive (MG_{FP}) Geometric Mean Bias – MG component.

* Department of Management and Environmental Protection, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland.

$$
\begin{aligned}
FB &= \frac{\sum_i (C_{Oi} - C_{Pi})}{0.5 \cdot \sum_i (C_{Oi} + C_{Pi})} \\
&= \frac{0.5 \cdot \sum_i [|C_{Oi} - C_{Pi}| + (C_{Oi} - C_{Pi})]}{0.5 \cdot \sum_i (C_{Oi} + C_{Pi})} - \frac{0.5 \cdot \sum_i [|C_{Oi} - C_{Pi}| + (C_{Pi} - C_{Oi})]}{0.5 \cdot \sum_i (C_{Oi} + C_{Pi})} \\
&= FB_{FN} - FB_{FP}. \qquad (1)
\end{aligned}
$$

FB_{FN}, negative component FB (false-negative), that is only these pairs (C_O, C_P) are considered, in which $C_P < C_O$:

$$FB_{FN} = \frac{0.5 \cdot \sum_i [|C_{Oi} - C_{Pi}| + (C_{Oi} - C_{Pi})]}{0.5 \cdot \sum_i (C_{Oi} + C_{Pi})}. \qquad (2)$$

FB_{FP}, positive component FB (false-positive), that is only these pairs (C_O, C_P) are considered, in which $C_P > C_O$:

$$FB_{FP} = \frac{0.5 \cdot \sum_i [|C_{Oi} - C_{Pi}| + (C_{Pi} - C_{Oi})]}{0.5 \cdot \sum_i (C_{Oi} + C_{Pi})}, \qquad (3)$$

where:

C_P – the concentration determined by the model,

C_O – the actual concentration.

We may determine negative (MG_{FN}) and positive (MG_{FP}) geometric mean bias components in the same way:

$$MG_{FN} = \exp\left[\frac{1}{2N} \sum_i [|\ln C_{Oi} - \ln C_{Pi}| + (\ln C_{Oi} - \ln C_{Pi})]\right], \qquad (4)$$

$$MG_{FP} = \exp\left[\frac{1}{2N} \sum_i [|\ln C_{Oi} - \ln C_{Pi}| + (\ln C_{Pi} - \ln C_{Oi})]\right]. \qquad (5)$$

2. CALCULATION OPTIONS

The wind field is determined using the Calmet meteorological pre-processor that transports pollutant puffs over the area surface. For calculating dispersion of air pollutants the Calpuff model was used (one of the Gaussian puff models). Information on the topographic profile and land use was obtained from U.S. Geological Survey [4]. Calculations were performed in the 40 km × 40 km network with a resolution of 1 km × 1 km. Two types of meteorological data were used for the calculations: those from ground meteorological stations – Katowice (12560), Kraków (12566), Bielsko-Biała (11111) and those from 4 aerological stations – Legionów (12374), Wrocław (12425), Prague (11520) and Poprad (11952).

The emission source was a smokestack in a heat and power station located in Bielsko-Biała, at the coordinates (360.0,5518.0), elevation of the base of 420 m above sea level. The SO_2 emission was 100 g/s. Smokestack height – 100 m, and outlet diameter – 5.0 m. The outlet velocity of exhaust gas was 15.0 m/s at a gas temperature of 400 K. Calculations have been performed based on the meteorological data for a period of 744 hours in March 2005 from which one-hour maximum concentration was chosen [5].

The following options corresponding to the meteorological data used to calculate the distribution of the pollutant concentrations were selected:

1) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 4 aerological stations (Wrocław, Legionów, Prague, Poprad) – 4up3surf,

2) 2 ground stations (Kraków, Katowice), 4 aerological stations (Wrocław, Legionów, Prague, Poprad) – 4up2surf,

3) 1 ground station (Bielsko-Biała), 4 aerological stations (Wrocław, Legionów, Prague, Poprad) – 4up1surf,

4) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 2 aerological stations (Wrocław, Poprad) – 2pow3surf,

5) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 2 aerological stations (Legionów, Prague) – 2prl3surf,

6) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 1 aerological station (Legionów) – leg3surf,

7) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 1 aerological station (Wrocław) – wroc3surf,

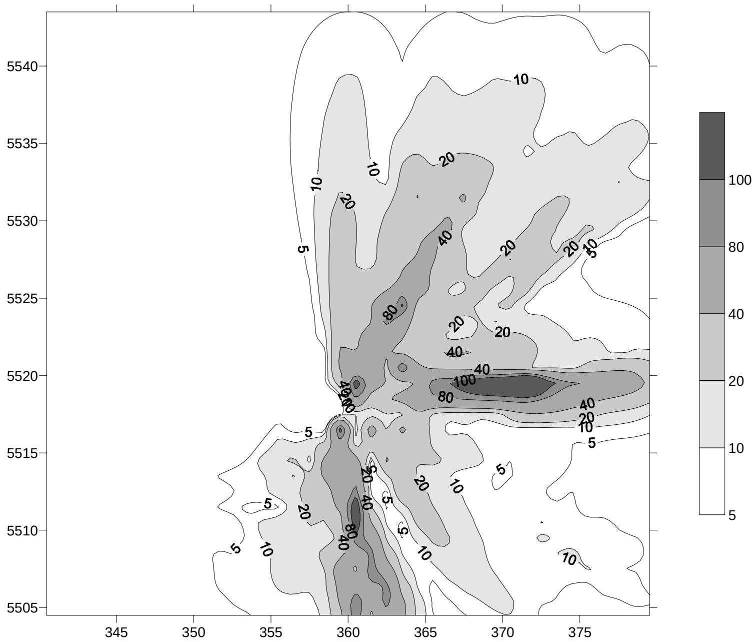
8) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 1 aerological station (Prague) – pra3surf,

9) 3 ground stations (Bielsko-Biała, Kraków, Katowice), 1 aerological station (Poprad) – pop3surf.

As a real distribution of the pollutant concentrations, the calculations for option number 1 were selected (3 ground and 4 aerological stations).

3. COMPUTATION RESULTS

Maps showing the distribution of one-hour maximum concentrations were made for each of the aforementioned options. Figure 1 shows an example of a pollution distribution map.



Distribution of one-hour maximum concentrations
(meteorological data – 4up 3surf, spot emitter height of 100 m)

The table presents a list of statistical indicators determined on the grounds of the distribution of one-hour SO_2 pollutant concentrations in the function of the volume of available meteorological data.

The highest compliance with the standard accepted for the statistical error was observed for 4up1surf option (FB = –0.01), and it was additionally confirmed by the low values of its negative and positive components (FB_{FN} = 0.01, FB_{FP} = 0.02). A high FB indicator value was also observed for 4up2surf option (FB = 0.24, FB_{FN} = 0.32, FB_{FP} = 0.08). In the case of the other options, a relatively low FB indicator was not confirmed by the negative and positive components (e.g. for leg3surf option FB = 0.11, but FB_{FN} = 0.58, FB_{FP} = 0.47). As for the correlation coefficient R, the values higher than 0.7 were reached only by the following options: 4up1surf (R = 1.00), 2pow3surf (R = 0.73). Also for these options the $FAC2$ indicator was higher than 0.5, that is for 4up1surf $FAC2$ and for 4up2surf $FAC2$ it reached the value of 1.00 and 0.85, respectively.

Statistical analysis of the pollutant concentration distributions proved that using the highest possible volume of meteorological data is very important to determine the properimmissible level at a given area. At the same time, lack of meteorological data from ground stations is still acceptable, although it changes the distribution of the pollutant concentration, whereas missing data concerning the vertical atmosphere profile may have a considerable effect on the results of the calculations. This happened in the case of options 6 to 9, which used information on the vertical profile of atmosphere condition obtained only from 1 aerological station. Both the form of the pollution distribution and the statistical indicators were completely different than the standard accepted.

Table

List of statistical indicators determined from distribution of one-hour SO_2 pollutant concentrations, depending on volume of meteorological data (spot emitter height h = 100 m)

Indicator	Computational options selected depending on type of meteorological data used for computing air pollution conditions								
	4up 3surf	4up 2surf	4up 1surf	2pow 3surf	2prl 3surf	leg 3surf	wroc 3surf	pra 3surf	pop 3surf
FB	0	0.24	–0.01	–0.46	–0.18	0.11	–0.69	0.09	–0.55
MG	1	1.22	1.01	3.87	0.51	0.62	0.35	0.43	3.50
NMSE	0	1.66	0.02	3.38	5.94	4.85	12.35	3.70	6.25
VG	1	1.34	1.01	2.43E+04	4.24E+05	2.93E+04	106.00	2.70E+04	3.36E+04
R	1	0.73	1.00	0.61	0.18	0.20	0.28	0.16	0.42
*FAC*2	1	0.85	1.00	0.43	0.27	0.33	0.27	0.33	0.36
Mean	11.5	9.0	11.5	18.2	13.7	10.3	23.6	10.4	20.2
St. dev.	18.7	14.5	19.3	32.3	27.7	19.1	58.8	13.0	40.8
Max	178	147	187	382	300	295	784	145	528
FB_{FN}	0	0.32	0.01	0.14	0.47	0.58	0.23	0.58	0.16
FB_{FP}	0	0.08	0.02	0.60	0.65	0.47	0.92	0.49	0.71
MOE_{FN}	1	0.72	0.99	0.82	0.49	0.45	0.66	0.45	0.78
MOE_{FP}	1	0.91	0.98	0.52	0.41	0.50	0.32	0.49	0.44
MG_{FN}	1	1.29	1.02	5.55	2.61	2.40	1.40	2.11	5.58
MG_{FP}	1	1.05	1.01	1.43	5.14	3.87	4.01	4.93	1.60

As a result of the completed calculations and analyses it can be stated that the use of only basic statistical indicators does not give a complete image of the modelling results. For example, a low value of the *FB* systematic error may result in a wrong evaluation by the researcher, but the selected model gives a good approximation of the real air quality condition. However, completely different conclusions may be drawn if its negative and positive components are determined as well. It may turn out that some results are much overrated (large value of negative component *FB*) or underrated (high value of additional component *FB*) within the selected area.

ACKNOWLEDGEMENT

The work was completed in the scope of AGH-UST statutory research, No. 11.11.150.171.

REFERENCES

[1] CHANG J.C., HANNA S.R., *Technical Descriptions and User's Guide for the BOOT Statistical Model Evaluation Software Package*, *Version 2.0*, 2005.

[2] *ASTM, 2000, Standard guide for statistical evaluation of atmospheric dispersion model performance*, American Society for Testing and Materials, Designation D6589-00. ASTM, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959.
[3] COX W.M., TIKVART J.A., *A statistical procedure for determining the best performing air quality simulation model*, Atmospheric Environment, 1990, Vol. 24A, No. 9, 2387–2395.
[4] http://edc.usgs.gov/
[5] SZCZYGŁOWSKI P., *Assessment of usability of selected Gaussian models in computing of air pollution condition*, Doctoral Dissertation, Faculty of Mining Surveying and Environmental Engineering AGH-UST, Kraków, 2007.

ZASTOSOWANIE PAKIETU STATYSTYCZNEGO BOOT W OBLICZENIACH ROZPRZESTRZENIANIA SIĘ ZANIECZYSZCZEŃ W POWIETRZU

Zagadnienie modelowania rozprzestrzeniania się zanieczyszczeń wymaga podjęcia wielu decyzji mających na celu właściwy dobór opcji modelu związanych z meteorologią czy współczynnikami dyspersji. Każda ich zmiana powoduje różnicę w wynikach obliczeń stanu zanieczyszczeń. Zastosowane w pracy nowe wskaźniki statystyczne pozwalają dokładniej ocenić wpływ zmiany poszczególnych parametrów na wyniki modelowania. Szczególnie przydatnymi, stosunkowo nowymi wskaźnikami statystycznymi, mogą okazać się ujemne i dodatnie komponenty błędu systematycznego (*FB*) bądź geometrycznego błędu średniego (*MG*), które umożliwiają ocenę przeszacowania lub niedoszacowania wyników obliczeń w przypadku obliczeń rozkładu zanieczyszczeń na określonym obszarze.