http://dx.doi.org/10.7494/geom.2016.10.3.87

Adriana Szulecka*, Marian Mazur*

Application of the Statistical Error and Quantitative Performance Measures in the Evaluation Process of Short-Term Air Quality Forecasts for Krakow (Poland)**

1. Introduction

Currently, the results of mathematical air quality modelling serve many purposes (e.g. global and local pollution prevention strategies, industrial emissions reduction, actions of local governments). Additionally, in the past few years, due to continuous digital technology development, prognostic models have been utilised for drawing up publicly available short-term forecasts of the upcoming air quality. This information is of particular importance in the areas of densely populated cities, which are frequently struggling with poor air quality. Issuing alerts about possible exceedances of the air quality standards [8] helps to mitigate adverse effects on the health of residents [6].

Krakow, located in the south-eastern Poland and with more than 750 000 inhabitants [3], has been recently qualified as one of the most polluted European cities due to excessive concentrations of PM2.5 [11]. In 2011, on behalf of the Marshal Office of Małopolska Voivodeship, an air quality forecasting system based on the results of global multiscale chemical weather modelling system GEM-AQ was developed and fully operational since then [23, 24]. The air quality forecasts for the area of Małopolska Voivodeship and major cities within are prepared on the basis of a modelling system created by the EkoForecast foundation, which utilises a global multiscale chemical weather model – GEM-AQ [9].

In order to ensure the correct functioning of the short-term air pollution forecasting system, an appropriate assessment at every stage of this application is essential. Basically, the validation process of a mathematical model consists of scientific, operational and statistical evaluation. Scientific evaluation requires a thorough knowledge of the model's basis. It examines the accuracy of described physical and chemical processes, as well as other assumptions in the model. Operational

^{*} AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Environmental Management and Protection, Krakow, Poland

^{**} The paper has been prepared within the scope of the AGH UST statutory research no. 11.11.150.008

evaluation considers user-related issues connected with, among others, the user interface, error checking of data and internal model diagnostics. Statistical assessment focuses mainly on the comparison between forecasted and measured observations. This method is rather intuitive and may not provide precise reasons for the divergence between observations, but it gives information about the nature and severity of possible error [2, 7, 14]. Due to its simplicity, this assessment should be carried out regularly to ensure reliable air quality forecasts and to improve general performance of the modelling system.

Basic statistical analysis may be performed with the use of the typical error measurements and correlation coefficients (Pearson, Spearman) for assessing model accuracy, which are applicable in many fields (economics, weather and air quality prediction). Those can be extended with a set of quantitative performance measures suggested by the U.S. Environmental Protection Agency as a basis for air quality model evaluation. To facilitate the interpretation of the values obtained one can begin with plotting the data in different ways (scatter, quantile-quantile, residual or conditional scatter plots) [4, 5].

2. Description of the Short-Term Air Quality Forecasting System for Krakow

Short-term air quality forecasts drawn up for the area of Krakow are based on the deterministic system modelling dynamics and atmospheric chemistry -GEM-AQ (Global Environmental Multiscale - Air Quality). It was developed as a result of extending the operational weather prediction model GEM by implementing air quality chemistry processes (i.e. transport, deposition, emission, limited wet chemistry). The GEM model was originally created by the Meteorological Services of Canada (MSC) and is presently used for weather prediction over Canada. Current mechanism in the GEM-AQ model contains 50 gas-phase compounds, 116 chemical and 19 photolysis reactions. The CAM module (Canadian Aerosol Model) implemented in this model describes 5 aerosol types (sulphate, sea-salt, organic carbon, black carbon and soil dust) and their physiochemical reactions. Gas-phase chemistry is based on the modified ADOM model (Acid Deposition and Oxidants Model). The chemical module is operating "on-line" - advection of the chemical compounds is performed at each timestep. Advection and vertical diffusion processes are computed with Semi-Lagrangian scheme deriving from the GEM model. The variable resolution capability makes it possible to perform high resolution local simulations based on the global runs [9, 12, 15].

Calculations in the short-term air quality forecast system for Małopolska Voivodeship are performed in a two-stage run. Results of the global simulation over Central Europe with variable resolution of 0.135 (3-D meteorological fields, chemical

composition of the atmosphere) are used as initial and boundary conditions for the nested run. Nested simulation covers the territory of Poland with resolution of 0.05° (Fig. 1). The forecast horizon is 75 hours (from 21:00 UTC on the previous day) [24]. The emission input data have been prepared with the EMEP inventories with 0.5° resolution for 2011 and 2012 [19, 20]. The emission rates of PM_{coarse}, PM2.5, nitrogen oxides, sulphur dioxide, carbon oxide and NMVOCs have been used in the calculations.

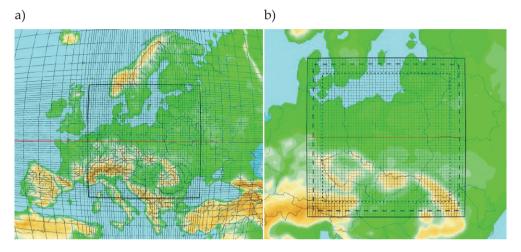


Fig. 1. Visualization of the global (a) and nested (b) run in the GEM-AQ model for air quality forecasting system in Małopolska Voivodeship

Source: [19, 20]

The results of the analysed air quality forecasting system are accessible via the web service Wrota Małopolski of the Marshal Office [24] and available for selected cities within the area of Małopolska Voivodeship. Forecasts are presented in the form of daily averaged concentrations of the following compounds: PM10, PM2.5, nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon oxide (CO) and ozone (O_3) for the next three days. An additional statistical correction of forecast data is performed to obtain better results [19, 20, 24].

In addition, maps of the distribution of the Common Air Quality Index (CAQI) over Małopolska Voivodeship are presented at this website. This indicator was created in the CITEAIR project and has been used at the Air Quality in Europe web service [1] for assessing air quality among more than 100 European cities from 2006. Its value is computed based on the three main pollutants in Europe: PM10, nitrogen dioxide (NO₂) and ozone (O₃), which can be extended by additional substances: PM2.5, sulphur dioxide (SO₂) and carbon oxide (CO) [1, 10]. Therefore, an accurate and reliable forecasting of these concentrations is notably significant, as they are forming the overall air quality in the region.

3. Statistical Methods of Model Evaluation

3.1. Description of the ex post Forecast Error Measurements

The parameters applied in this study indicate differences between the actual and forecasted values and have been selected from a larger set of variables. Those are described below [17–20, 22, 25, 26]:

Mean Bias Error:

$$MBE = \frac{1}{m} \sum_{\tau=1}^{m} (C_p - C_O)$$
 (1)

Mean Percentage Error:

MPE =
$$\frac{\sum_{\tau=1}^{m} \frac{C_{p} - C_{O}}{C_{O}}}{m} [\%]$$
 (2)

Mean Absolute Bias Error:

$$MABE = \frac{1}{m} \sum_{r=1}^{m} |C_{p} - C_{O}|$$
 (3)

Mean Absolute Percentage Error:

MAPE =
$$\frac{1}{m} \sum_{\tau=1}^{m} \left| \frac{C_p - C_O}{C_O} \right| \cdot 100 \, [\%]$$
 (4)

- Root Mean Square Error:

RMSE =
$$\sqrt{\frac{1}{m} \sum_{\tau=1}^{m} (C_P - C_O)^2}$$
 (5)

where C_p and C_O denotes a predicted and observed concentration, respectively, and is a total amount of observations during the time series ($\tau = 1, 2, ..., m$). Ideally, the value of MBE should be equal or close to zero. A positive or negative value indicates overprediction or underprediction of prognostic model, respectively. A significant difference between RMSE and MABE point to presence of high values of discrepancies between observed and predicted concentrations [2, 26].

3.2. Description of Quantitative Performance Measures (US EPA)

In order to assess the general performance of the analysed prognostic model, a set of quantitative measures developed in 1993 by the National Environmental Research Institute (NERI) of Denmark was applied in this study. They have been recommended by the US Environmental Protection Agency for air quality model evaluation tool and later implemented in a software package BOOT (Statistical Model Evaluation Software Package, Version 2.0). Those measures include [4, 5, 16, 21]:

a) Fractional Bias:

$$FB = \frac{\left(\overline{C_O} - \overline{C_P}\right)}{0.5 \cdot \left(\overline{C_O} + \overline{C_P}\right)} \tag{6}$$

b) Geometric Mean Bias:

$$MG = \exp\left(\overline{\ln C_O} - \overline{\ln C_P}\right) \tag{7}$$

c) Normalized Mean Square Error:

$$NMSE = \frac{\overline{\left(C_O - C_P\right)^2}}{\overline{C_O} \cdot \overline{C_P}}$$
 (8)

d) Geometric Variance:

$$VG = \exp\left[\overline{\left(\ln C_O - \ln C_P\right)^2}\right] \tag{9}$$

e) Fraction within a factor of two:

FAC2 = fraction of data that satisfy
$$0.5 \le \frac{C_p}{C_O} \le 2.0$$
 (10)

where C_P and C_O denotes a predicted and observed concentration, respectively, and m represents a total amount of observations during the τ time series. An ideal model would have MG, VG and FAC2 = 1.0, and FB and NMSE = 0.0. Based on the evaluation of many models with many field data sets, model acceptance criteria for those measures were developed [4, 21]:

- fraction within a factor of two should be equal or greater than 50% (FAC2>0.5),
- |FB| < 0.3 or 0.7 < MG < 1.3,
- values of NMSE and VG should be less than 1.5 and 4, respectively.

4. Statistical Evaluation of Short-Term Forecasts for Krakow

In this study a brief comparison of prognostic and observed concentrations of selected air pollutants in Krakow was carried out. Prognostic observations regarding the upcoming daily average concentrations of PM10, PM2.5, SO_2 , NO_2 and O_3 were systematically collected over a period of April 2014 – March 2015 from the Wrota Małopolski web service of the Marshal Office of Małopolska Voivodeship [24]. Those values were subsequently confronted with the corresponding measurements recorded at the urban background measuring station (Bujaka St., Krakow) at the same time. This station belongs to the surveillance grid of the Voivodeship Environmental Protection Inspectorate in Krakow and it has been operational since 2010. Continuous measures at this station are performed for the following substances: PM10, PM2.5, nitrogen oxides (NO_x) , nitrogen dioxide (NO_2) , nitrous oxide (NO), sulphur dioxide (SO_2) and ozone (O_3) . It is located in the southern part of Krakow (geogr. 50.010575; 19.949189) in a residential area. There are no significant emission sources in the area surrounding the station [13].

Below the time series of daily average modelled and measured concentrations of the analysed air pollutants are presented (Figs 2–6). They are supplemented by the values of Spearman correlation coefficients (Tab. 1) and the results of statistical evaluation of the examined modelling system (Tabs 2, 3). Due to the variability of the observed concentrations during the year, the error statistics have been calculated with additional division of the analysed time period to non-heating season (April – September) and heating season (October – March).

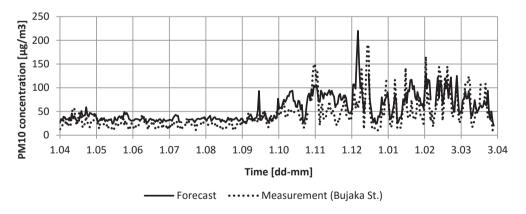


Fig. 2. Time series (April 2014 – March 2015) of forecast and measured daily average concentration of PM10 [μg/m³]

Predicted PM10 concentrations are properly modelled by the evaluated fore-casting system as evidenced by the high correlation coefficient (0.79) and the value of FAC2 = 0.801. However, over the whole analysed period of time the model tends

to slightly overestimate the observed concentrations of PM10 (MBE = $12.70 \mu g/m^3$). Higher rates of errors have been obtained during the heating season with a maximum of $163 \mu g/m^3$ (on December 7^{th} 2014).

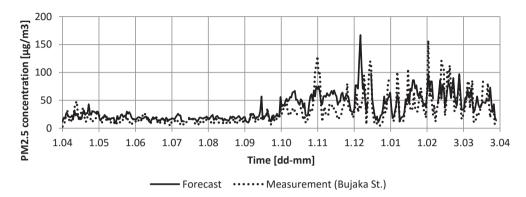


Fig. 3. Time series (April 2014 – March 2015) of forecast and measured daily average concentration of PM2.5 [μg/m³]

In the case of PM2.5 predictions the overestimation tendency of the model is similar to those observed for PM10 forecasts. It is yet consistently lower in value (MBE = $6.97 \,\mu g/m^3$) which provides a slightly higher correlation rate of observations (r = 0.82). It is important to note that the percentage absolute percentage errors are comparable (MAPE_{PM10} = 61.17%; MAPE_{PM2.5} = 63.38%) and better representation of PM2.5 observations may be superficial.

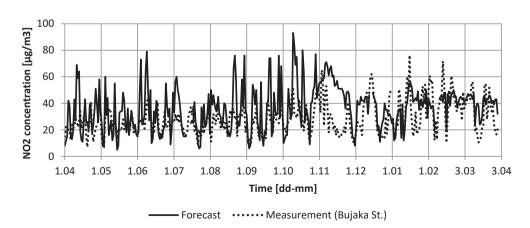


Fig. 4. Time series (April 2014 – March 2015) of forecast and measured daily average concentration of nitrogen dioxide (NO₂) [μg/m³]

Observed variation of the forecasted nitrogen dioxide concentrations during the non-heating season are not coherent with the measured values which affects the lower value of Spearman correlation coefficient (r = 0.49). For most of the observed days the evaluated prognostic model is overestimating NO₂ concentrations. The values of error statistics are yet not significantly high (MAPE = 54.80%) which allows meeting the model acceptance criteria (FB = -0.254; MG = 0.8; FAC2 = 0.793).

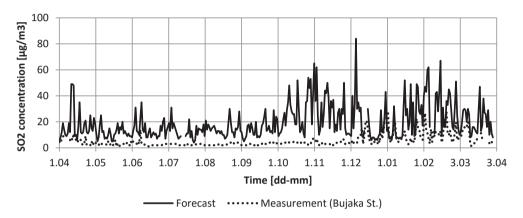


Fig. 5. Time series (April 2014 – March 2015) of forecast and measured daily average concentration of sulphur dioxide (SO₂) [μ g/m³]

Daily average sulphur dioxide concentrations measured at Bujaka St. station only rarely exceed 20 $\mu g/m^3$ whereas model predictions indicates the possible occurrence of significantly higher values during the analysed year of observations. Thus, error statistics are great in value with MBE = 14.13 $\mu g/m^3$ and MAPE = 382.74%. Accordingly, the quantitative performance measures have deficient values (FB = -1.148; MG = 0.250; FAC2 = 0.180) in reference to the assumed model quality criteria.

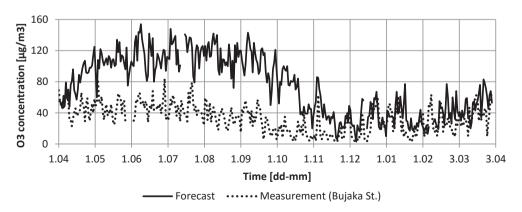


Fig. 6. Time series (April 2014 – March 2015) of forecast and measured daily average concentration of ozone (O_2) [µg/m³]

Pollutant	Correlation coefficient (r)			
PM10	0.79			
PM2.5	0.82			
NO ₂	0.49			
SO ₂	0.42			
O ₃	0.52			

Table 1. Spearman correlation coefficients for GEM-AQ model evaluation (April 2014 – March 2015 observations)

High discrepancies between modelled and measured ozone concentrations are particularly evident during the non-heating season. Over warm months of the year GEM-AQ modelling system produces highly overestimated predictions for this air pollutant (MBE_{NHS} = 63.95 μ g/m³), yet the mean percentage error value remains similar for the heating season (MPE_{NHS} = 191.55% and MPE_{HS} = 194.88%). As a result the correlation coefficient denotes a moderate relationship between observations (r = 0.52). Although the model performs better during the heating season (FAC2_{HS} = 0.524), the overall error statistics indicate insufficient reliability of the short-term forecasts of ozone in the EkoForecast system.

Table 2. Selected error statistics for GEM-AQ model evaluation – daily average concentrations [μg/m³]

Pollutant	Seasons	Error statistics					
		MBE	MABE	MPE	MAPE	RMSE	
PM10	non-heating season	9.43	11.34	52.48	56.72	13.28	
	heating season	16.14	27.59	56.62	65.83	33.95	
	year	12.70	19.25	54.50	61.17	25.53	
PM2.5	non-heating season	4.42	6.37	45.64	52.63	8.02	
	heating season	9.64	20.96	62.82	74.70	26.93	
	year	6.97	13.50	54.03	63.38	19.68	
NO ₂	non-heating season	7.08	12.09	31.62	52.56	16.67	
	heating season	8.72	14.12	45.95	58.08	18.45	
	year	7.89	13.09	38.71	54.80	17.58	
SO ₂	non-heating season	10.99	10.99	392.57	392.57	12.94	
	heating season	17.37	17.85	374.44	372.56	23.05	
	year	14.13	14.34	383.72	382.74	18.58	
O ₃	non-heating season	63.95	64.26	191.55	192.07	68.76	
	heating season	20.23	23.50	194.88	202.77	31.07	
	year	41.96	43.76	193.23	197.45	53.25	

D. II	Year seasons	Quantitative performance measures					
Pollutant		FB	MG	NMSE	VG	FAC2	
PM10	non-heating season	-0.309	0.690	0.200	1.280	0.842	
	heating season	-0.257	0.690	0.270	1.380	0.758	
	year	-0.274	0.690	0.290	1.330	0.801	
PM2.5	non-heating season	-0.254	0.730	0.220	1.260	0.935	
	heating season	-0.212	0.700	0.320	1.480	0.729	
	year	-0.223	0.710	0.380	1.360	0.818	
NO ₂	non-heating season	-0.258	0.850	0.360	1.310	0.808	
	heating season	-0.251	0.760	0.270	1.340	0.776	
	year	-0.254	0.800	0.310	1.330	0.793	
SO ₂	non-heating season	-1.257	0.230	3.630	11.400	0.086	
	heating season	-1.085	0.290	2.950	8.950	0.279	
	year	-1.148	0.250	3.410	10.200	0.180	
O ₃	non-heating season	-0.857	0.380	1.040	3.060	0.292	
	heating season	-0.595	0.480	0.930	3.160	0.524	
	year	-0.775	0.430	1.140	3.110	0.408	

Table 3. Selected quantitative performance measures (US EPA) for GEM-AQ model evaluation – daily average concentrations [μg/m³]

5. Summary and Conclusions

The main objective of the carried out analysis was to perform a statistical evaluation of the results of a short-term air quality forecasting system for the area of Krakow. Available forecasted data derived from the Wrota Małopolski web service [24] were compared to the urban background measurements at the station located in Krakow, Bujaka St., over the period of April 2014 – March 2015 with heating (October – March) and non-heating (April – September) seasons taken into account.

In general, the analysed observations show a tendency of overestimation in reference to the actual concentrations recorded at the urban background station, which is particularly noticeable during the non-heating season (in the case of sulphur dioxide and ozone). The results obtained indicate a good reliability of PM10 and PM2.5 forecasted daily average concentrations as evidenced by the high Spearman correlation coefficient values (0.79 and 0.82, respectively) as they are also meeting the model acceptance criteria [4]. Despite the rather accurate realization of the short-term forecasts of nitrogen dioxide in relation to the error statistics values, the correlation coefficient indicates moderate correlation rate of the modelled and measured

observations (r = 0.49). This might be due to the higher variability of forecasted values compared to the actual observations during the analysed time series. In the case of sulphur dioxide and ozone forecasts the values of the statistical measures indicate high discrepancies in reference to the measurement data, which qualifies the assessed model as insufficiently accurate in predicting the upcoming concentrations of these compounds.

It is important to emphasize that a general evaluation of the analysed modelling system has been published annually for the area of Małopolska Voivodeship since 2010. Those reports [19–20] point to the on-going process of improving the model configuration using more recent input data. The extended statistical evaluation presented in this paper is consistent with those reports in relation to most of the concerned pollutants, showing generally good reliability of predicted values with slight overestimation noticeable during the non-heating season (especially for PM10 and PM2.5). Discrepancies regarding the results of the assessment for ozone predictions may be related to different averaging time of data for the evaluation or with the availability of more accurate modelling data for the location of Kurdwanów urban background monitoring station in Krakow.

The conducted analysis highlights the importance of publicly available air quality forecasting data system and its limitations associated with spatial accuracy and data quality. Furthermore, discrepancies in the forecasted air pollutants concentrations are likely to affect the verifiability of the Common Air Quality Index [1, 10] which is calculated on the basis of these values and published at the Wrota Małopolski website for general information about current air quality. An additional examination of the model characteristics is recommended for adequate identification of the causes and nature of the noticed errors in order to improve the air quality management system in Krakow [22].

References

- [1] Air Quality in Europe, [on-line:] http://www.airqualitynow.eu/ [access: December 2015].
- [2] Armstrong J.S., Collopy F.: *Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons*. International Journal of Forecasting, vol. 8, 1992, pp. 69–80.
- [3] Central Statistical Office (GUS): Size and structure of population and vital statistics in Poland by territorial division in 2014. As of December 31. Warsaw 2015.
- [4] Chang J.C., Hanna S.R.: Technical Descriptions and User's Guide for the BOOT Statistical Model Evaluation Software Package, v. 2.0, 2005.
- [5] Chang J.C., Hanna S.R.: *Air quality model performance evaluation*. Meteorology and Atmospheric Physics, vol. 87, issue 1, 2004, pp. 167–196.

- [6] Davila S., Ilić J.P., Bešlić I.: *Real-time dissemination of air quality information using data streams and Web technologies: linking air quality to health risks in urban areas.* Archives of Industrial Hygiene and Toxicology, vol. 66, issue 2, 2015, pp. 171–180.
- [7] Derwent D., Fraser A., Abbott J., Jenkin M., Willis P., Murrells T.: *Evaluating the Performance of Air Quality Models*. 2010, [on-line:] http://uk-air.defra.gov.uk/ [access: December 2015].
- [8] Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe (CAFE).
- [9] EkoForecast foundation, [on-line:] http://www.ekoprognoza.pl/ [access: January 2016].
- [10] Elshout S., Bartelds H., Heich H., Léger K.: *CAQI Air quality index Comparing Urban Air Quality across Borders –* 2012. Shiedam 2012.
- [11] European Environmental Agency: *SOER* 2015 *The European environment state and outlook* 2015. Copenhagen 2015.
- [12] Kamiński J.W., Neary L., Strużewska J., McConnell J.C., Lupu A., Jarosz J., Toyota K., Gong S.L., Côté J., Liu X., Chance K., Richter A.: GEM-AQ, an on-line global multiscale chemical weather modelling system: model description and evaluation of gas phase chemistry processes. Atmospheric Chemistry and Physics, vol. 8, 2008, pp. 3255–3281.
- [13] WIOŚ Kraków: *System monitoringu jakości powietrza*, [on-line:] http://monitoring.krakow.pios.gov.pl/ [access: December 2015].
- [14] Markiewicz M.T.: Podstawy modelowania rozprzestrzeniania się zanieczyszczeń w powietrzu atmosferycznym. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2004.
- [15] Neary L., Kamiński J.W., Lupu A., McConnell J.C.: *Developments and Results from a Global Multiscale Air Quality Model (GEM-AQ)*. [in:] *Air Pollution Modeling and Its Application*, XVII, vol. 5, Springer US, 2007.
- [16] Oleniacz R., Rzeszutek M.: Assessment of the impact of spatial data on the results of air pollution dispersion modeling. Geoinformatica Polonica, vol. 13, 2014, pp. 57–68.
- [17] Shaocai Y., Mathur R., Kang D., Schere K., Pleim J., Otte T.L.: A detailed evaluation of the Eta-CMAQ forecast model performance for O₃, its related precursors, and meteorological parameters during the 2004 ICARTT study. Journal of Geophysical Research, vol. 112, D12S14, 2007.
- [18] Strużewska J., Zdunek M., Kamiński J.W., Łobocki L., Porębska M., Jefimow M., Gawuc L.: Evaluation of the GEM-AQ model in the context of the AQMEII Phase 1 project. Atmospheric Chemistry and Physics, vol. 15, 2015, pp. 3971–3990.
- [19] Strużewska-Krajewska J., Kamiński J.W., Durka P., Jefimow M.: Sprawdzalność operacyjnej prognozy stanu jakości powietrza dla obszaru województwa małopolskiego w okresie od grudnia 2013 do września 2014. Warszawa 2014.

- [20] Strużewska-Krajewska J., Kamiński J.W., Durka P., Jefimow M.: Sprawdzalność operacyjnej prognozy stanu jakości powietrza dla obszaru województwa małopolskiego w okresie od października 2014 do września 2015. Warszawa 2015.
- [21] Szczygłowski P., Mazur M.: *Application of BOOT statistical package in calculating pollutant spreading in air*. Environment Protection Engineering, vol. 34, no. 4, 2008, pp. 151–156.
- [22] Szulecka A.: *Porównanie prognostycznych i rzeczywistych wartości stężeń zanieczyszczeń powietrza w Krakowie*. AGH, Wydział Geodezji Górniczej i Inżynierii Środowiska, Katedra Kształtowania i Ochrony Środowiska, Kraków 2015 [M.Sc. thesis, unpublished].
- [23] Szulecka A.: *Prognozowanie jakości powietrza w Krakowie*. AURA: pismo Naczelnej Organizacji Technicznej, 1/2016, pp. 7–9.
- [24] Wrota Małopolski: *Jakość powietrza w Małopolsce*, [on-line:] http://www.po-wietrze.malopolska.pl/prognozy/ [access: March 2016].
- [25] Zhang Y., Liu P., Pun B., Seigneur C.: A Comprehensive Performance Evaluation of MM5-CMAQ for the Summer 1999 Southern Oxidants Study Episode, Part I: Evaluation Protocols, Databases and Meteorological Predictions. Atmospheric Environment, vol. 40, issue 26, 2006, pp. 4825–4838.
- [26] Zhang Y., Seigneur C., Bocquet M., Mallet V., Baklanov A.: *Real-Time Air Quality Forecasting. Part I: History, Techniques, and Current Status.* Atmospheric Environment, vol. 60, 2012, pp. 632–655.