

PAPER • OPEN ACCESS

Modelling and Forecasting of SO₂ Concentration in Atmospheric Air – A Case Study of the City of Krakow

To cite this article: Zbigniew Skrobacki *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **471** 102057

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Modelling and Forecasting of SO₂ Concentration in Atmospheric Air – A Case Study of the City of Krakow

Zbigniew Skrobacki ¹, Lidia Ewa Dabek ², Piotr Lagowski ¹

¹ Kielce University of Technology, Faculty of Mechatronics and Mechanical Engineering, Al. Tysiąclecia Państwa Polskiego 7, Poland

² Kielce University of Technology, Faculty of Environmental, Geomatic and Energy Engineering, Al. Tysiąclecia Państwa Polskiego 7, Poland

p.lagowski@tu.kielce.pl

Abstract. The paper presents a statistical analysis of SO₂ air concentration in Krakow for the period from 2007 to 2016. The source data were obtained from two measuring stations of the Provincial Environmental Protection Inspectorate in Krakow. SO₂ concentration data had daily average values. The average monthly values and quarterly averages were used for the analysis. The analysis started with the determination of the trend function and the time series of the moving average. Then indices of monthly and quarterly changes were determined. Considering the whole period of ten years, a slight downward trend was noted, even though a somewhat rising trend was observed in the first five years. In Krakow, the level of sulphur dioxide pollution is approximately two to 2.5 times higher than in the least polluted area of the region the capital of which is Krakow. To create a mathematical model, the following two series were identified: cyclic variability and random variation. A multiplicative model composed of four components was obtained. This model interpolates the source data. The methodology of time series modelling was also applied to monthly and quarterly data. The wind directions prevailing in the city centre were also defined for the last four years. Statistical analysis and multiplicative models of SO₂ variability for these directions were made. For each of the seven selected wind directions, the basic concentration statistics were determined and all components of the multiplicative models were identified. No significant effect of the wind direction on SO₂ concentration in Krakow was observed. This allows formulating a hypothesis of the dominating impact of the relatively uniformly distributed local sources of surface emission on the magnitude of air pollution by sulphur dioxide in a large urban agglomeration, which is confirmed by some selected scientific sources. The models presented were used to obtain quarterly and monthly forecasts for SO₂ concentrations in Krakow. The discussed methodology is universal and can be used in the study of variability and forecasting of concentrations of other air pollutants in other cities.

1. Introduction

Atmospheric air is a significant element of environment, directly affecting human and animal health and quality of plants [1-10]. The SO₂ imission is of a considerable importance in air pollution assessment. It is especially significant for Krakow, which is classified as one of the most polluted agglomerations in Poland and EU owing to the state of its air. For this reason, it is necessary to conduct air quality monitoring studies and to select effective methods allowing for a thorough data



assessment, pollutants spread estimation and the forecast of variations. Today, the miniaturization of measuring equipment and lower unit purchase costs permit the development of measurement stations network and the spread of portable stations [11]. Comprehensive analysis of a large number of data obtained from numerous measurement stations requires different statistical methods [12-13]. Majority of researchers merely apply descriptive statistics and correlation studies. The authors of this paper conduct the investigation on air pollution in urban agglomerations with the aim of proposing an effective pollution analysis and forecast methods based on methodologies applied i.e. in the time series analysis. The aim of this paper is to present a part of the authors' research on the SO₂ concentration in the Krakow air over ten years. The paper discusses selected results of the basic statistical data analysis and gives examples of using the multiplicative model to record the variability of the investigated feature as a function of time i.e. for the prevailing wind directions.

2. The analysis of SO₂ concentration from 2007 to 2016

In the area of Krakow, the automatic measurement of air quality is performed by three measuring stations of Provincial Environmental Protection Inspectorate. The paper presents the analysis of SO₂ concentration measurements collected at two stations: in Al. Krasińskiego and in Bujaka Street, which are situated in the closest vicinity to the city centre. The archived data of the station in Al. Krasińskiego provided information on SO₂ concentration up to 31st of January 2013, after which the station ceased measuring SO₂. On 1st of January, the station in Bujaka Street started operating, performing i.a. SO₂ concentration measurements. The paper presents the collected measurement results from both stations, which provided the information for a database covering a period of ten years: 2007-2016 (figure 1).

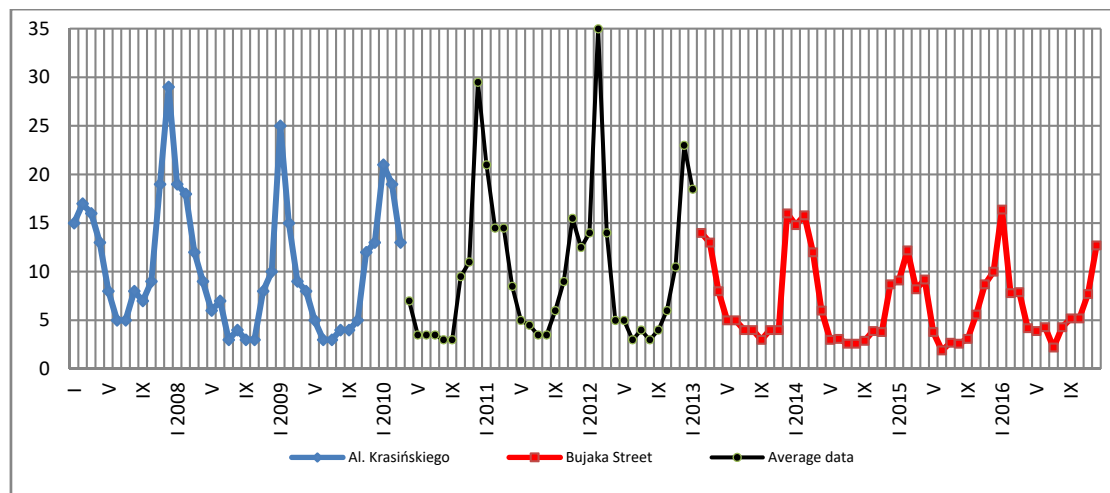


Figure 1. The average monthly SO₂ concentration [$\mu\text{g}\cdot\text{m}^{-3}$] in the city Krakow from 2007 to 2016

The initial data analysis for the whole 120-month period shows that the SO₂ concentration values reveal a slightly decreasing trend. The trend function takes the form of $y = -0.0466x + 11.71$ [$\mu\text{g}\cdot\text{m}^{-3}$] where: x is the next month number from the set of $x = \{1, 2, \dots, 120\}$. It was found that the maximum concentration for the examined period was recorded in February 2012 with the average value of $35 \mu\text{g}\cdot\text{m}^{-3}$ and that the trend was slightly increasing prior to that date and slightly decreasing after that date. The trend function for both periods was determined: prior to February 2012 the function takes the form of $y = 0.0134x + 10.318$ [$\mu\text{g}\cdot\text{m}^{-3}$] where $x = \{1, 2, \dots, 62\}$ and after that date $y = 0.0917x + 15.956$ [$\mu\text{g}\cdot\text{m}^{-3}$] where $x = \{62, 63, \dots, 120\}$. Taking into account the change in trend, it was decided to perform a five-year concentration variability analysis. The data on the SO₂ concentration in the yearly cross-section for the first five years and the next five years are shown in the graphs (figure 2, figure 3).

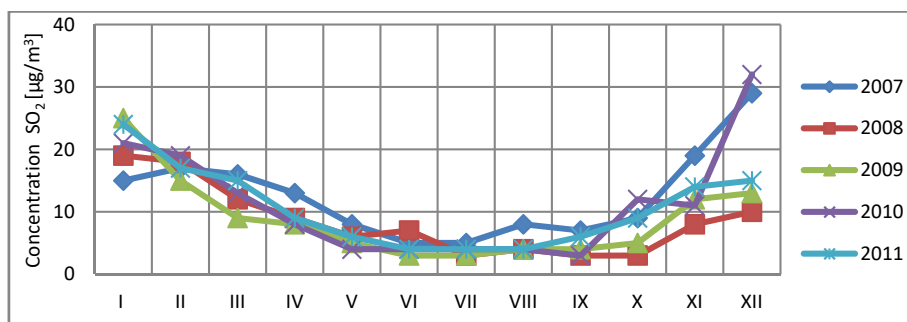


Figure 2. Monthly average concentration of SO₂ [$\mu\text{g}\cdot\text{m}^{-3}$] in the city Krakow during the period from 2007 to 2011

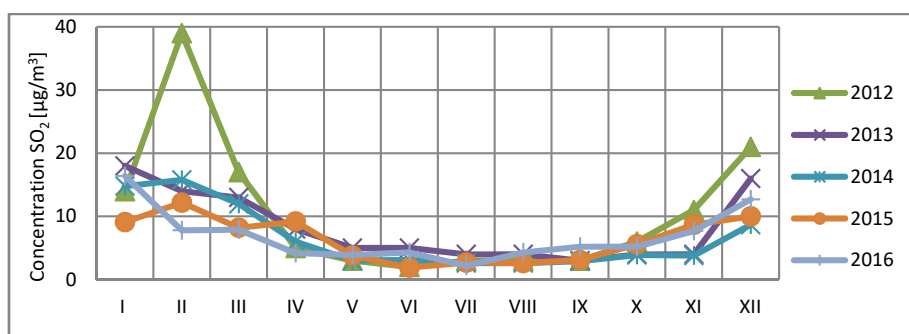


Figure 3. Monthly average concentration of SO₂ [$\mu\text{g}\cdot\text{m}^{-3}$] in the city Krakow during the period from 2012 to 2016

The analysis of graphs (figure 2, figure 3) shows that the monthly data in the successive years are characterized by a high repeatability and little dispersion. Exceptionally high concentrations, however, occurred in the following three months: December 2007, December 2010 and February 2012. At this stage of the analysis, there were no attempts to explain the reasons for the occurrence of those anomalies. The concentration values in the city of Krakow were compared to the concentrations registered during the ten-year period by the Provincial Environmental Protection Inspectorate in Szymbark [9]. This station is app. 100 km away from Krakow in the SE direction, situated in the rural mountainous area. The station measures, i.a. SO₂ concentration, which is regarded as the reference background for the data from other stations in the region and in Krakow. The graph (figure 4) shows the average monthly SO₂ concentrations obtained for the ten-year period in both locations.

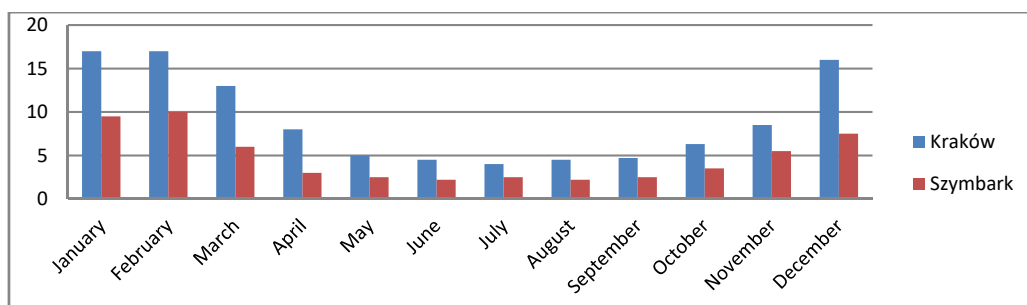


Figure 4. Monthly average concentration of SO₂ [$\mu\text{g}\cdot\text{m}^{-3}$] in the city of Krakow and in the Szymbark station over a period of ten years

On average, the concentration in Krakow is 95% higher than in Szymbark. In the coldest months of the heating season: in December, January and February, the concentrations are 113%, 79% and 70% higher, respectively.

3. Model of SO₂ concentration

The data presented above was used to develop the mathematical models for the average SO₂ concentration in Krakow. The methodology of time series modelling based on recommendations in Aczel's monograph [13] was applied. This methodology permits model development for both original data analysis and forecasting purposes. The model is multiplicative and takes into account the variability recorded in the form of a trend equation as well as seasonal one-year variability, cyclical variability with the cycle length different from one year and random variability. Being the interpolation model, it very accurately reproduces the original data plot as a function of time. This article presents two multiplicative models. The first model represents monthly averages, the second depicts quarterly averages.

$$Z_m = T_m \cdot S_m \cdot C_m \cdot I_m, \quad (1)$$

$$Z_{kw} = T_{kw} \cdot S_{kw} \cdot C_{kw} \cdot I_{kw}, \quad (2)$$

where: Z_m, Z_{kw} – SO₂ concentration [$\mu\text{g}\cdot\text{m}^{-3}$]: monthly average, quarterly average, respectively;

T_m, T_{kw} – the trend function with the equation, respectively: $T_m = -0.0466t_m + 11.71$ [$\mu\text{g}\cdot\text{m}^{-3}$],

where: t_m – the number of the month, $t_m = \{1, 2, \dots, 120\}$,

$T_{kw} = -0.1463t_{kw} + 12.05$ [$\mu\text{g}\cdot\text{m}^{-3}$], where: t_{kw} – the number of the quarter, $t_{kw} = \{1, 2, \dots, 40\}$;

S_m, S_{kw} – seasonality index [-]: monthly, $S_m = \{S_{m,1}, S_{m,2}, \dots, S_{m,12}\}$; quarterly $S_{kw} = \{S_{kw,1}, S_{kw,2}, S_{kw,3}, S_{kw,4}\}$, respectively;

C_m – cyclicity index [-]: monthly $C_m = \{C_{m,7}, C_{m,8}, \dots, C_{m,114}\}$, quarterly $C_{kw} = \{C_{kw,3}, C_{kw,4}, \dots, C_{kw,38}\}$, respectively;

I_m, I_{kw} – random variation index [-]: monthly $I_m = \{I_{m,7}, I_{m,8}, \dots, I_{m,114}\}$, quarterly $I_{kw} = \{I_{kw,3}, I_{kw,4}, \dots, I_{kw,38}\}$, respectively.

The values of the seasonal variation index for Krakow were calculated for monthly and quarterly periods as shown in the graph (figure 5). Seasonal indices were calculated using the general relationship [13]:

$$S \cdot I = Z_{emp}/MA, \quad (3)$$

where: Z_{emp} – SO₂ concentration obtained from measurements, MA – moving average.

This stage should be followed by the elimination of random effects. The cyclical variation index is defined by the general formula [13]:

$$C = MA/T, \quad (4)$$

and the random variation index is calculated from the equation:

$$I = Z_{emp}/(T \cdot S \cdot C), \quad (5)$$

The use of the empirical quantities Z_{emp} in equations is based on the assumption that the theoretical values obtained from the model Z_m must be equal to the empirical ones: i.e. $Z_m = Z_{emp}$. A similar assumption was made for quarterly averages. Such an assumption makes the applied model interpolate empirical values. The calculated values of the components C and I are shown in the graph (figure 6).

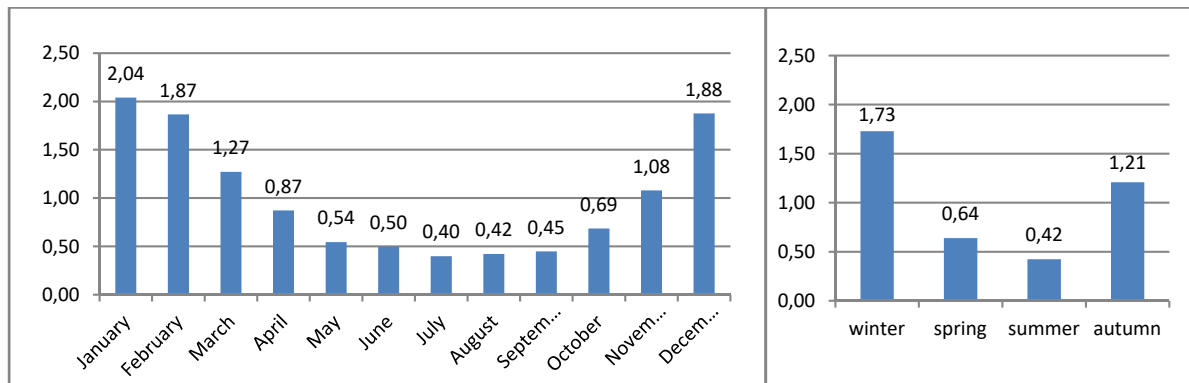


Figure 5. Seasonal variation index for monthly averages S_m [-] (graph on the left) and for quarterly averages S_{kw} [-] (graph on the right).

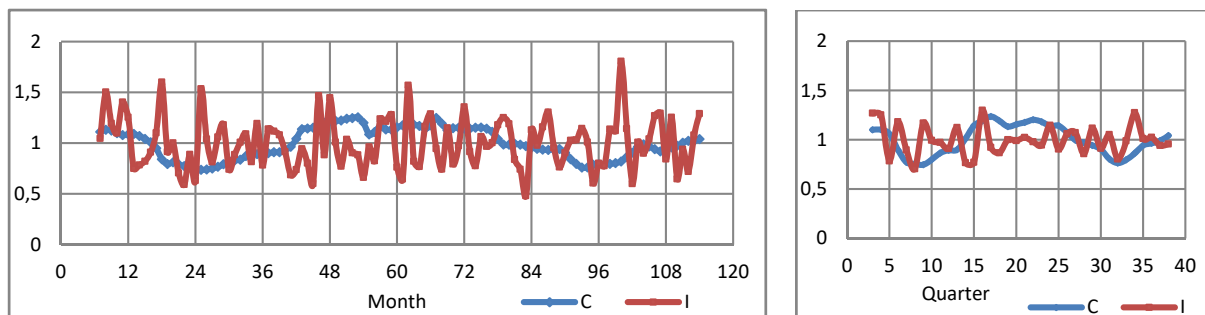


Figure 6. Time series for cyclical variation indices, C_m [-] and random variations indices, I_m [-] for monthly averages (graph on the left); an analogous plot ($(C_{kw} \text{ and } I_{kw}) = f(\text{time})$) for quarterly averages (graph on the right).

4. Models of SO_2 concentration for wind directions

The methodology of time series analysis and modelling was applied to perform a statistical analysis and develop multiplicative models of SO_2 concentration for selected seven wind directions in the city centre of Krakow in the period from 2013 to 2016. The identification of the prevailing wind directions followed the analysis of average daily data from the investigated four-year period. The wind rose plot (figure 7) developed for each year shows the dominance of the wind from the SWW direction (about 25% share each year) and W (app. 25% share in the years 2015 and 2016 and 15% share in the other years). For the other southern directions: S, SSE, SEE and E direction, the shares range from 5% to 10%. A slightly higher share is noted for the SSW direction. Further analysis focused on those wind directions that were observed in 90% of all the days of the investigated 4-year period.

The SO_2 concentration distribution at the time of the wind blowing from a particular direction is presented in seven box plots (figure 8). The initial analysis of the charts shows that the wind direction has a very insignificant effect on SO_2 concentration in a large agglomeration. This allows formulating a hypothesis of the dominating impact of the relatively uniformly distributed local sources of surface emission on the magnitude of air pollution by sulphur dioxide in a large urban agglomeration, as confirmed by Oleniacz and others [10].

A multiplicative model was developed for the quarterly data for each of the seven identified wind directions. The selected data relevant for the development of approximate quarterly forecasts are presented in table 1.

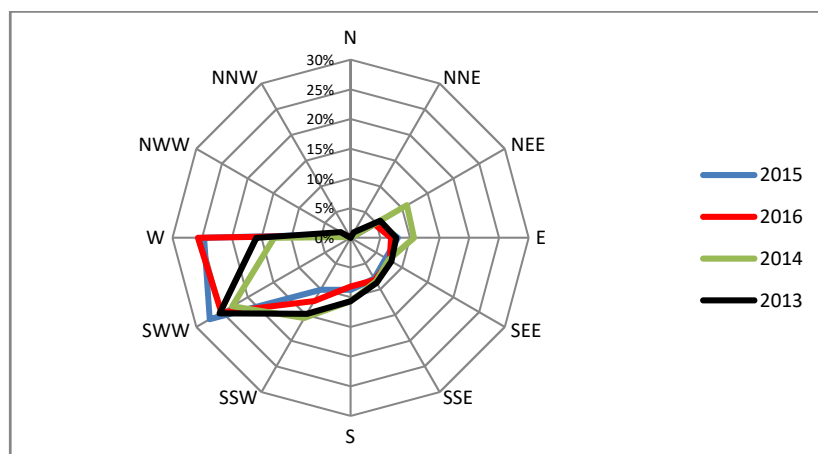


Figure 7. Share [%] of the wind direction in the city centre of Krakow for each year from 2013 to 2016.

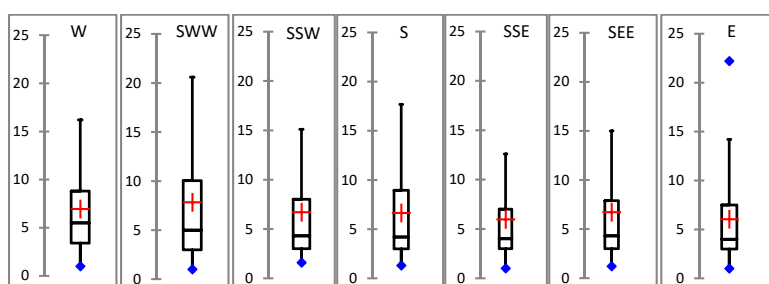


Figure 8. Distribution statistics for the average daily concentration of SO_2 [$\mu\text{g}\cdot\text{m}^{-3}$] in the period from 2013 to 2016 for selected wind directions.

Table 1. Selected data from the multiplicative models.

Wind direction	The average concentration of SO_2 [$\mu\text{g}\cdot\text{m}^{-3}$] for the winter quarter				Trend	Index S for quarterly variations			
	2013	2014	2015	2016		winter	spring	summer	autumn
W	15	13	9	12	$-0.1587x+8.37$	1.754	0.733	0.475	1.038
SWW	16	18	14	12	$-0.2128x+9.94$	1.880	0.640	0.430	1.040
SSW	17	15	12	12	$-0.2956x+10.00$	1.580	0.610	0.390	1.430
S	21	15	8	10	$-0.2738x+9.62$	1.750	0.560	0.420	1.260
SSE	17	16	9	9	$-0.3385x+9.74$	1.800	0.660	0.450	1.080
SEE	22	12	10	12	$-0.2813x+9.54$	1.820	0.570	0.390	1.220
E	13	11	5	11	$-0.0756x+6.47$	1.650	0.660	0.490	1.200

5. The forecast model for monthly data

5.1. Assumption

In order to obtain a forecast based on model (1), the model should be transformed into a form in which each component of the model will be a forecast:

$$\check{Z}_\tau = \check{T}_\tau \cdot \check{S}_m \cdot \check{C}_\tau \cdot \check{I}_\tau, \quad (6)$$

where: \check{Z}_τ – forecast for $Z[\mu\text{g}\cdot\text{m}^{-3}]$, \check{T}_τ – forecast for $T[\mu\text{g}\cdot\text{m}^{-3}]$, \check{S}_m – forecast for $S_m[-]$,
 \check{C}_τ – forecast for $C[-]$, \check{I}_τ – forecast for $I[-]$.

Forecasts for T and S are obtained on the basis of the assumption: (a) determined for historical data $t = 1, 2, \dots, 120$; general change trend for time $\tau > 120$ will not change, (b) the index values $S_m = S_1, S_2, \dots, S_{12}$; do not change. When applying methodological recommendations in monograph [13], it should be assumed: (a) $\check{I}_\tau = 1$ for each moment τ , (b) forecast \check{C}_τ , is obtained by trying to guess the value of \check{C}_τ after analysing the course of historical C_t values. Because these assumptions may cause errors in the forecasts, it was decided to propose an own method in which a forecast for the product $(C_t \cdot I_t)$ is determined using the autoregressive model $AR(p)$.

5.2. Autoregressive model $AR(p)$

Let us consider the case linear process, in which the first p of the weights are nonzero. This model may be written:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t, \quad (7)$$

where: $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}$ – variable values for subsequent months $t, t-1, t-2, \dots, t-p$;
 $\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p$ – set of weight parameters; p – order of autoregressive; e_t – random values.

Equation (7) is a record of the autoregression process of order p , which is written in short $AR(p)$. In order to determine the p value, it is necessary to calculate [14-15]:

- (1) values of autocorrelation coefficients ρ_k ($k = 1, 2, \dots, K$),
- (2) estimation of parameters $\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_K$,
- (3) coefficients of partial autocorrelation $\varphi_{k,k}$.

The source data for the autoregression model is the time series of the product $(C_t \cdot I_t)$, the values of which are shown in figure 9.

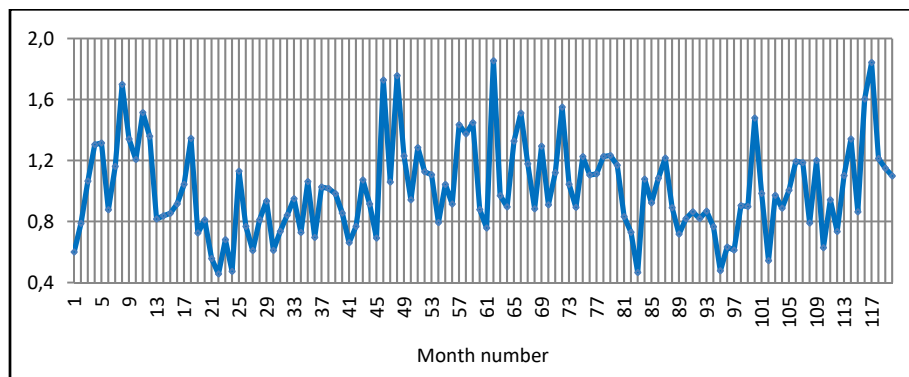


Figure 9. Time series for $(C_t \cdot I_t)$ [-].

Implementing the methodological recommendations [14-15], the value of $p = 1$ was determined and the values of autoregression model coefficients were identified. The equation was obtained:

$$AR(1)=y_t = 0.681+0.327y_{t-1}+ e_t, \quad (8)$$

The AR (1) process is stationary because the value of the coefficient $\phi_1 = 0.327 < 1$.

5.3. Ex post forecast

After entering the equation (8) into the general formula (6), the SO₂ forecast equation was found:

$$\check{Z}_t = \check{T}_t \cdot \check{S}_m \cdot [0.681+0.327(C_{t-1} \cdot I_{t-1})], \quad (9)$$

Model (9) is a combination of a non-linear model with trend and seasonality as well as a linear autoregressive model. After obtaining data on the concentration of SO₂ in the period from January to December 2017, an ex post forecast was made. The formula (9) was used both for forecasting in the last five years (figure 10) and in 2017 (figure 11). The "weakness" of the model (9) is visible from the preliminary assessment of both charts, which consists in the creation of large forecast errors in some winter months. This phenomenon is caused by the fact that the time series models do not take into account causal relationships, and in this case, the impacts of atmospheric factors on the level of SO₂ concentration in the air in the urban agglomeration.

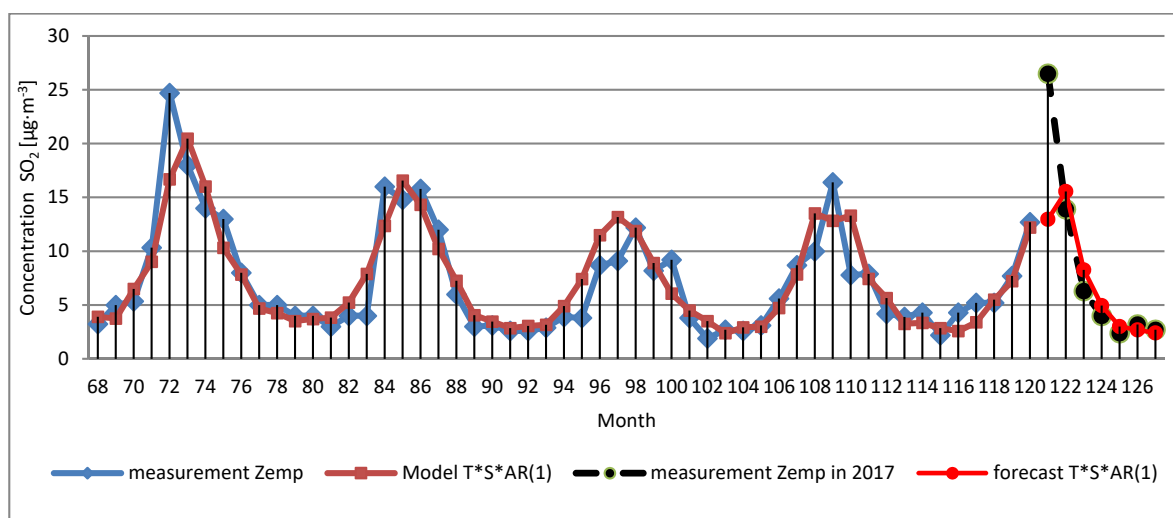


Figure 10. Ex post forecast for the average monthly concentration of SO₂ [μg·m⁻³] in the period from August 2012 to July 2017.

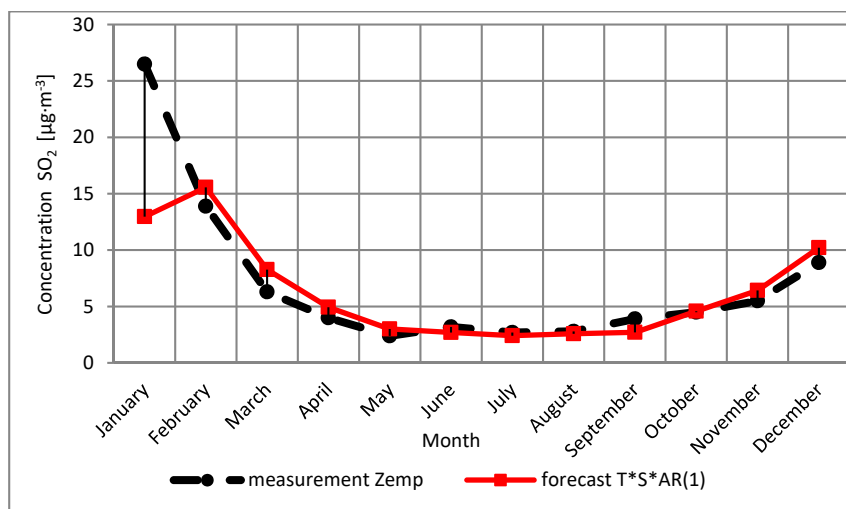


Figure 11. Ex post forecast for the average monthly concentration of SO_2 [$\mu\text{g}\cdot\text{m}^{-3}$] in the period from January 2017 to December 2017.

6. Conclusions

- Considering the whole period (from 2007 to 2016) of the SO_2 concentration analysis, a slight downward trend was noted, even though a somewhat rising trend was observed in the first five years. In Krakow, the level of sulphur dioxide pollution is approximately from twice to 2.5 times higher than in the least contaminated area in the Małopolska region.
- The multiplicative models were developed for the monthly and quarterly SO_2 concentrations, covering the 10-year period and taking into account the trend and the seasonal, cyclical and random variability.
- For each of the seven selected wind directions, the basic concentration statistics were determined and all the components of the multiplicative models for quarterly averages were identified.
- No significant effect of the wind direction on SO_2 concentration in Krakow was observed. This allows formulating a hypothesis of the dominating impact of the relatively uniformly distributed local sources of surface emission on the magnitude of air pollution by sulphur dioxide in a large urban agglomeration.
- A forecasting model was created, which is a combination of a non-linear model with trend and seasonality as well as a linear autoregressive model. After obtaining data on the concentration of SO_2 in the period from January to December 2017, an ex post forecast was made. Forecast values were found to be satisfactory. The "weakness" of the forecasting model is the creation of forecast errors in some winter months.

References

- [1] K. Juda-Rezler, "Impact of air pollution on the environment," *Publishing House of the Warsaw University of Technology*, Second Edition, 2006 (in Polish).
- [2] B. Dębski, A. Olecka, K. Bebkiewicz, I. Kargulewicz, J. Rutkowski, D. Zasina, M. Zimakowska-Laskowska, and M. Żaczek, "Basic report," *KOBIZE, Institute for Environmental Protection – National Research Institute*, Warsaw 2015, (in Polish).
- [3] B. Dębski, A. Olecka, K. Bebkiewicz, I. Kargulewicz, J. Rutkowski, D. Zasina, M. Zimakowska-Laskowska, and M. Żaczek, "Report 2015, Submission under the UN ECE Convention on Long-range Transboundary Air Pollution," *KOBIZE*, Warsaw 2015.
- [4] "Directive 2008/50/EC," *Official Journal L 152*, 11.06.2008.
- [5] "Directive 2010/75/EU," *Official Journal L 334*, 12.12.2010.

- [6] “Air quality in Europe — 2014 report,”*EEA Report No 5/2014*, Copenhagen, 2014.
- [7] M. Markiewicz, “Basic of modeling the diffusion of pollutants in atmospheric air,”*Publishing House of the Warsaw University of Technology*, 2004 (in Polish).
- [8] M. Rzeszutek, and R. Oleniacz, *Engineering and Environmental Protection* 18, 1,5 (2015) (in Polish),
- [9] W. Bochenek, Swietokrzyski Yearbook, Series B, *Natural Sciences*, 36, 9 (2015) (in Polish).
- [10] R. Oleniacz, M. Bogacki, A. Szulecka, M. Rzeszutek, and M. Mazur, V International Conference INFRAEKO 2016 Modern cities, *Publishing House of the Rzeszów University of Technology*, 2016 (in Polish).
- [11] W. Batko, Z. Skrobacki and M. Oliszewski, *Journal of Powertrain and Transport KONES*, Vol. 23, No. 3, pp. 25-32, 2016.
- [12] A. Bąkowski, L. Radziszewski, Z. Skrobacki, *XXI International Polish-Slovak Conference Machine Modeling and Simulations, Procedia Engineering* 177 (2017) pp. 281-288.
- [13] Amir D. Aczel, “Complete Business Statistics,” 2nd edition, *Richard D. Irwin, Inc.*, 1993.
- [14] G.E.P. Box, and G. M. Jenkins, “Time Series Analysis: Forecasting and Control,” *Holden-Day*, San Francisco, 1976.
- [15] P. J. Brockwell and R. A. Davis, “Time series: Theory and Methods,” 2nd edition. *Springer Verlag*, New York, 1996.