Application of DOPPLER SODAR in short-term forecasting of PM10 concentration in the air in Krakow (Poland)

Ewa Krajny1, Leszek Osrodka2 and Marek Wojtylak2

1 IMWM-NR, 40-045 Katowice, Bratkow 10
2 Retired researcher (M.W.)

Correspondence to: Ewa Krajny (E.K.) ewa.krajny@imgw.pl, Leszek Osrodka (L.O.) leszek.osrodka@imgw.pl

Abstract. This paper describes an attempt to use data obtained from SODAR (Sound Detection and Ranging) for short-term forecasting of PM10 concentration levels in Krakow. Krakow is one of the most polluted cities in Central Europe (CE) in terms of PM10 concentration. This is due to the high municipal emissions. Thanks to intensive corrective actions taken by the city authorities, it is being effectively eliminated, but the unfavourable topographic location of the city limits natural ventilation. The article describes all these conditions, focusing on presenting the method of short-term correction of air quality for the time needed to take quick corrective actions by the city authorities in the event of anticipated exceedances of the permissible values. Based on several years of measurements of the physical properties of the atmosphere with SODAR, the authors of the paper suggest that SODAR data could be considered for operational use to generate short-term predictions. The proposed method is based on the use of the spectrum, i.e. the set of amplitudes of signals returning to the SODAR receiver from the reflection of a single-frequency sound transmission and its characteristic properties depending on the physical state of the atmosphere. Similar spectra were parameterized with a single numerical value using statistical methods. It was found that in some cases preceding high concentrations of PM10, the spectral parameters had similar values. This made it possible to develop a forecasting method for such concentrations by using data mining to search for conditions in historical data closest to the state of the atmosphere at the time of forecasting. In this part of the study, data from 2017–2018 were used. In the next step, three methods of using the SODAR data developed in this way for PM10 prediction were proposed, comparing them with the method without SODAR use. The study results were tested on independent material using data concerning the episodes of high concentrations of pollutants from October 2021 to March 2022 in Krakow. The findings were considered encouraging also taking into account the speed and low cost of preparing the forecast.

[Background]

Air pollution is a significant problem for residents of urban agglomerations. The growing size of cities and the increasing number of inhabitants result in both an increase in emissions and a reduction in the efficient removal of pollutants outside the city limits. This then causes a periodic pollution accumulation leading to smog episodes. There is evidence that air pollution has a significant impact on both life expectancy and quality of life. During episodes, an increased number of hospital admissions and deaths are observed, especially among children and the elderly (WHO, 2004; Bell and Davis, 2001; Dockery et al., 1993). The occurrence of high PM concentrations in Krakow is described below.

Increasing emissions as a result of lower air temperatures during the cold season, reducing the demand for heat and, thus, reducing emissions as a result of "warm winters", are well-known examples of emission control mechanisms based on meteorological and climatological factors. Another and equally important role is played by thermal and dynamic conditions in the boundary layer of the atmosphere, where apart from solar radiation and dynamic factors, the degree of urbanization and land cover are also important (Xue et al., 2021; Ji et al., 2020; Xu et al., 2018; Engelbart et al., 2009; Fisher et al., 2006; Prünger et al., 2004; Arya, 1999). These issues have also been studied in Poland, where the problem of high concentrations of particulate pollutants is prevalent, especially in the winter months. January–March and October–December (Toczkó, 2015; Ziemińska and Osrodka, 2012). Among the large Polish cities, Krakow is the most polluted when it comes to PM concentration. Therefore, research in this area is particularly intensive (Bajorek and Zydron, 2016; Bokwa, 2010; Matuszko, 2007; Walczewski, 1994). For example, extensive research on secondary sources on the legacy and contemporary research on climate diversity in Krakow can be found in Bokwa (2019) and Matuszko (2007).

The inspiration for this research was the analysis of the current state of knowledge regarding the atmospheric structure of the border layer over Krakow (Godłowska et al., 2008; Lewinska et al., 1982; Morawicka-Horawicka, 1978). In connection with the introduction in 1993 in Krakow of the first automatic air quality monitoring system in Poland and the delivery of system equipment, including REMTECH Doppler SODAR delivered to the Center Observation Station of the Institute of Meteorology and Water Management National Research Institute in Krakow-Czyzyn, the research became more extensive. Based on data obtained from SODAR, wind roses were developed for Krakow at the altitude of 50 to 550 m at intervals of 100 m separately for different types of atmospheric equilibrium (Rozwoda, 1995) for mixing layers and wind circulating over the city (Fisher et al., 2006; Prünger et al., 2004) and for COST 720 (Engelbart et al., 2009), while the measurement results obtained at that time became a useful knowledge base on the subject at modern research (Godłowska and Kaszowski, 2019; Bajorek, Zydron and Zydron, 2016; Bokwa, 2010; Matuszko, 2007; Walczewski, 1994). For example, extensive research on secondary sources on the legacy and contemporary research on climate diversity in Krakow can be found in Bokwa (2019) and Matuszko (2007).
In recent years, thanks to the use of drones and balloons for civilian applications, as well as miniaturization of measuring instruments, experimental studies of the vertical structure of the atmosphere from the border layer in Krakow have been widely developed, with an emphasis on the impact of meteorological conditions on the vertical profile of air pollutants (Sekula et al., 2021). These studies provide important information on the vertical distribution of PM10 concentrations under different meteorological conditions.

Intensive corrective actions were carried out in Krakow as part of research activities on the causes and effects of excessive air pollution. In Poland, legal solutions have been applied in accordance with the provisions of 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 (2008). Therefore, in areas (zones) with excessive air pollution, it is necessary to establish air quality programs to direct corrective actions (Regulation of the Minister of the Environment, 2012). In addition to an important legislative function (they constitute an act of local law), these programs are of significant preventive and educational importance. Due to the fact that the concentrations of PM10, PM2.5, and benzo(a)pyrene (B(a)P) in Krakow exceeded the limits specified in the applicable legal standards, air quality programs have been implemented for several years. The effects of their implementation include, among other things, the so-called “anti-smog resolution” (Resolution No. XVIII/243/16, 2016), which from 2019 prohibits the combustion of solid fuels for space heating and results in a gradual reduction of PM concentrations. As a result of these legislative actions, the replacement of boilers for space and water heating, a process carried out since 1995, has increased sharply, leading to a significant reduction in PM emissions in the municipal sector (the number of solid fuel boilers decreased by around 15,000 boiler installations in 2017–2019) (AQP Malopolska, 2018). This, in turn, significantly reduced PM10 concentrations, shown in Figure 1 as the number of days when the PM concentration threshold was exceeded. Specific meteorological conditions in recent years (mild and warm winters) have also contributed to the situation but have not diminished the importance of reducing municipal emissions.

![PM10 concentrations](image)

**Figure 1.** Total number of days with PM10 concentrations exceeding 50, 100, and 150 μg/m³ in 2015–2020.

The unfavourable geographical location of the city also causes failure to meet air quality standards in Krakow. In the north and south, the historical center of the city, located in the Vistula Valley, is sheltered by round hills that cross river valleys. The biggest problem of Krakow is that from the west, it is protected by the features of a topographically diverse area called the Krakow Bridge, located at 370 m above sea level (a.s.l.) (German, 2001). The hills of the Krakow Bridge are a barrier for the entire city, reducing the efficiency of ventilation and limiting the clearance of the Vistula Valley in the western part of Krakow, threatening the influx of wind, with western air circulation dominating in Poland. The location of Krakow in the Vistula Valley affects not only the wind conditions but also contributes to temperature inversions, determining a constant type of equilibrium of the atmosphere in harsh conditions and reducing the ability to mix air vertically. Urbanization processes, which result in an increasing density of taller and taller buildings, also contribute to the reduction of wind speed and the ability to remove air pollution outside the city area. However, urbanization often contributes to improving vertical ventilation conditions by reducing the horizontal dispersion of pollutants. This is facilitated by increased anthropogenic heat production, which reduces the frequency of inversions in areas with large communities.
2 Purpose of the study

The episodes of high PM10 concentrations in Krakow prompted the city authorities to develop a system that will provide residents with access to free public transport when the limit of 8 hours of concentration is to be exceeded the next day. The implementation of an air quality forecasting system to control public transport imposes the obligation to exercise the utmost care in modelling. An inaccurate forecast is associated with the risk of the city budget incurring unjustified costs (by multiplying the air quality forecasts) or social (health) costs, which are difficult to estimate for an underestimated air quality forecast. In the specific conditions prevailing in Krakow, mainly local factors of stagnation, such situations are not uncommon. An additional element that makes it difficult to correctly forecast air quality using a deterministic model is the inability to provide an accurate emission inventory. Given this state of affairs, methods are needed to frame air quality forecasts, in particular, time trend forecasts of the absolute concentration of PM10. The long-term activity of Doppler SODAR in Krakow and the analysis of the impact of atmospheric stability conditions on air quality gave the opportunity to use the results of these measurements to improve the air quality forecasting system (Bajorek, Zydnowski, and Wezyk, 2016).

3 Materials and methods

The work uses two data sources:

- Data from the selected Inspectorate for Environmental Protection/National Environmental Monitoring (IEP/NEM) - 1105 automatic air quality monitoring stations based in Krakow, data from 2015 to March 2022.
- Measurements of SODAR tags from 2017–March 2022.

The station characteristics and locations are shown in Table 1 and the station codes are shown in Figure 2.

### Table 1. Characteristics of monitoring stations in Krakow

<table>
<thead>
<tr>
<th>No</th>
<th>Name/Locatio n of the monitoring station</th>
<th>Measured elements</th>
<th>GPS coordinates</th>
<th>Altitude of area/Stat t</th>
<th>Type of area/ Stat t</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Krasins ki</td>
<td>PM10</td>
<td>50°32.77, 19°35.53</td>
<td>202 m</td>
<td>urban</td>
</tr>
<tr>
<td>A</td>
<td>Bulwarowa</td>
<td>PM10</td>
<td>50°03.49, 20°03.13</td>
<td>105 m</td>
<td>natural</td>
</tr>
<tr>
<td>A</td>
<td>Bocek Site</td>
<td>PM10</td>
<td>50°00.28, 19°56.56</td>
<td>223 m</td>
<td>rural</td>
</tr>
</tbody>
</table>

**Figure 2.** Geographical location of the research area showing the locations of meteorological/air pollution monitoring stations. Meteorological stations of the Institute of Meteorology and Water Management – National Research Institute (marked in black): (A) Krakow-Balice, (B) SODAR; Monitoring stations of the National Environmental Monitoring (NEM) (marked in white): (1) Krasinski Avenue, (2) Bulwarowa Street, (3) Bocek Street.
The monostatic Doppler SODAR PCS.2000 and an overview of SODAR is shown in Figure 3. The principle of operation of SODAR is based on the Doppler effect. The SODAR PCS.2000 system was built to measure wind speed and wind direction profile, including elements based on PSC.2000 manual, METEK version 2013, as part of the MONIT-AIR project. The system was manufactured by the German company METEK Meteorologische Messtechnik GmbH (https://metek.de). The SODAR PCS.2000-24 system transmits three audio beams simultaneously, each with a different frequency, which turbulence parameters are determined in the form of atmospheric stability class and indirectly, air temperature inversion. It is a monostatic Doppler SODAR system transmitting antenna that transmits an audio signal of a given frequency and switches to receive and record a feedback signal. The monostatic SODAR PCS.2000-24 system transmits three audio beams simultaneously, each with three antennas: one vertical and two inclined phased antennas (Table 2).

The maximum range of vertical SODAR detection depends on the frequency of the transmitted signal and the current atmospheric conditions. SODAR measures the physical parameters of the atmosphere by analyzing the spectrum of sound waves dispersed by fluctuations in the atmosphere of different degrees, which are the result of heat and dynamic turbulence in the atmosphere (air temperature gradients, wind speed, and hydrometers). The Doppler effect for flat-phased antennas depends only on the wind speed and the distance between the sound transmitting the sound. The technical specifications of SODAR are shown in Table 2, and an overview of SODAR is shown in Figure 3.


<table>
<thead>
<tr>
<th>Parameters</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating frequency</td>
<td>1500...2600 Hz</td>
</tr>
<tr>
<td>Range of measured speeds of horizontal wind elements</td>
<td>±50 m/s</td>
</tr>
<tr>
<td>Range of measured winds directions</td>
<td>±360°</td>
</tr>
<tr>
<td>Range of measured speeds of vertical wind elements</td>
<td>±10 m/s</td>
</tr>
<tr>
<td>Minimum working height</td>
<td>≥15 m (adjustable)</td>
</tr>
<tr>
<td>Minimum vertical resolution</td>
<td>≥5 m (adjustable)</td>
</tr>
<tr>
<td>Height range</td>
<td>=500 m (adjustable)</td>
</tr>
</tbody>
</table>

### 3.1 Measurements of SODAR TAGS

The monostatic Doppler SODAR PCS.2000-24 system has been operating in Krakow since January 2015 as part of the MONIT-AIR project. The system was manufactured by the German company METEK Meteorologische Messtechnik GmbH (https://metek.de). The technical specifications of SODAR are shown in Table 2, and an overview of SODAR is shown in Figure 3. The SODAR PCS.2000-24 system was built to measure wind speed and wind direction profile, based on which turbulence parameters are determined in the form of atmospheric stability class and indirectly, air temperature inversion. It is a monostatic Doppler SODAR system transmitting antenna that transmits an audio signal of a given frequency and switches to receive and record a feedback signal. The monostatic SODAR PCS.2000-24 system transmits three audio beams simultaneously, each with three antennas: one vertical and two inclined phased antennas (Table 2).

The maximum range of vertical SODAR detection depends on the frequency of the transmitted signal and the current atmospheric conditions. SODAR measures the physical parameters of the atmosphere by analyzing the spectrum of sound waves dispersed by fluctuations in the atmosphere of different degrees, which are the result of heat and dynamic turbulence in the atmosphere (air temperature gradients, wind speed, and hydrometers). The Doppler effect for flat-phased antennas depends only on the wind speed and the distance between the sound transmitting the sound. The technical specifications of SODAR are shown in Table 2, and an overview of SODAR is shown in Figure 3.
The analysis of episodes of high concentrations of PM10 in Krakow and the classes of atmospheric stability identified by SODAR led to the use of SODAR data for short-term forecasting of PM10 concentrations in the city. Figure 4 illustrates an episode of high concentrations of PM10 dust against the background of DC determined from SODAR.
To build the PM10 concentration forecast, the basic results of SODAR measurements as a spectrum were used, i.e. a set of spectra at different altitudes. However, the SODAR data concerning only DC is insufficient for a precise diagnosis and prediction of the episodes of high PM10 concentrations. In view of the above, it was decided to use properly processed data of the SODAR spectrum, assuming that this would allow for a more thorough analysis of ventilation conditions.

### 3.2.2 Preparation of SODAR data filtration

To build the PM10 concentration forecast, the basic results of SODAR measurements as a spectrum were used, i.e. a set of spectra at different altitudes. However, the SODAR data concerning only DC is insufficient for a precise diagnosis and prediction of the episodes of high PM10 concentrations. In view of the above, it was decided to use properly processed data of the SODAR spectrum, assuming that this would allow for a more thorough analysis of ventilation conditions.

To build the PM10 concentration forecast, the basic results of SODAR measurements as a spectrum were used, i.e. a set of spectra at different altitudes. However, the SODAR data concerning only DC is insufficient for a precise diagnosis and prediction of the episodes of high PM10 concentrations. In view of the above, it was decided to use properly processed data of the SODAR spectrum, assuming that this would allow for a more thorough analysis of ventilation conditions.

The lack of a clear relationship between the high DC obtained from the visualization of SODAR data and the simultaneous high concentration of PM10 dust in stagnant weather conditions (high pressure system, low wind speed, temperature inversion) suggests that the SODAR data concerning only DC is insufficient for a precise diagnosis and prediction of the episodes of high PM10 concentrations. In view of the above, it was decided to use properly processed data of the SODAR spectrum, assuming that this would allow for a more thorough analysis of ventilation conditions.

The lack of a clear relationship between the high DC obtained from the visualization of SODAR data and the simultaneous high concentration of PM10 dust in stagnant weather conditions (high pressure system, low wind speed, temperature inversion) suggests that the SODAR data concerning only DC is insufficient for a precise diagnosis and prediction of the episodes of high PM10 concentrations. In view of the above, it was decided to use properly processed data of the SODAR spectrum, assuming that this would allow for a more thorough analysis of ventilation conditions.

### Figure 4

PM10 concentrations at selected Environmental Monitoring Stations in Krakow and the average measurements as at 20 January 2019, compared to atmospheric stability classes calculated based on SODAR data.

The example refers to a 5-minute period on 7 October 2017. This is an illustration of the spectrum development method. Only the filtered spectra were used for further study. In addition, in order to slightly reduce the amount of data and adjust the frequency of air volume measurements, the spectrum of interest was averaged to one hour. However, SODAR measures the data at 10-minute intervals.
Figure 5. Example of filtering the SODAR spectrum on 32 channels: a) original spectrum; b) spectrum with subtracted background; c) context.

3.2.3 Spectral properties of SODAR data

Because of the difficulty in comparing full spectra (a string of 32 values at each height) with each other and with spectra from other periods, each spectrum was characterized by a single number (parameter). Thus arose many functions of real values, the argument of which is height. Any such function will be called the atmospheric state profile (ASP). The most commonly observed feature of ASP is a rapid, non-linear decrease in value as the height increases. Attempts were made to determine as many parameters of the spectrum as possible. The numerical characteristics of the spectrum were chosen following a statistical approach, without analyzing their physical interpretations.

The characteristics with which ASP was determined were as follows:

- A. Mean value of the SODAR reflection spectrum;
- B. Maximum value of the spectra with a beam;
- C. Signal-to-noise ratio (SNR);
- D. Modal value (channel number with maximum high spectrum);
- E. Standard deviation;
- F. Median;
- G. Skewness;
- H. Kurtosis;
- I. Similarity to the Gaussian curve.

Analyzing ASP shapes, their similarity to the \( \Phi \) function turned out to be familiar (Eq. 1):

\[
\Phi(h) = \frac{a}{N + b} + c(h)
\]  

The ASP for the parameter \( C \) profile was similar to another feature (Eq. 2):

\[
\Psi(h) = a \cdot h^c \cdot e^{-b_h}
\]

The functions \( \Phi \) and \( \Psi \) were chosen arbitrarily. Parameters \( a, b \) and \( c \) of Equations (1) and (2) were determined for individual ASP (h – height above ground level using the quadratic mean approximation method). The mean square approximation method determines the quality of the \( \Phi \) matching with the measurement data in the form of a coefficient of determination (the square of the correlation coefficient). This coefficient for transparency is marked RF (regularity factor) and ranges from 0 (significant irregularity) to 1 (perfect regularity).
The sample analysis suggested that RF = 0.373 means poor regularity (Figure 6a) and RF = 0.956 means good regularity (Figure 6b). A general feature was that the transformed spectra were close to zero from a certain height. Therefore, the calculations included up to 19 heights (215 m). Figure 7 shows that the ASP at 18:00 was irregular, meaning that the vertical structure of the atmosphere was deformed and irregular. ASP at 23:00 GMT was almost perfect. This led to an investigation of the relationship between the value of ASP and PM10 concentration. For example, the chart in Figure 7 shows that 24–27 January 2018 is the period during which excessive PM10 concentrations occurred. The regularity factor for the characteristic was compared with the average concentration of PM10 from three automatic NEM stations near the SODAR location (Krasinski Avenue, Bulwarowa and Bujaka Streets). The three stations were selected for analysis because they are located in different areas of spatial development. This allowed approximating the average PM10 concentration for Krakow as they had the most complete series of measurements. For example, only one RF was shown in graphs for clarity, but the other regularity factors were similar. In this basis, it was hypothesized that ASP was disturbed a few hours before the increase in PM10 concentration. The hypothesis was developed based on the analysis of many charts, including Figure 7. This happened so often that it became the inspiration for this work.

**Figure 6.** Graph of the regularity coefficient for parameter 4 (mean spectrum — spectrum average) compared to the PM10 concentration: a) irregularity to 18:00 UTC; b) good regularity to 23:00 UTC on 26 January 2018.

**Figure 7.** Example of the regularity factor (RF) for characterization (mean spectrum or spectrum average) between the SODAR data height of 35–205 m a.g.l. and PM10 concentration.

For a complete description of the state of the atmosphere, it was decided that the meteorological RF wind characteristics should be added to the regularity coefficients A to E:

J. Horizontal wind speed (averaged in ASP);
K. Vertical wind speed (difference: min-max);
L. Wind direction (dispersion of direction around the average vector of this wind direction);
M. Temperature measured with SODAR at 2 m a.g.l.

**4 PM10 forecast models**

This article proposes four methods of forecasting PM10 concentration in Krakow. Three use SODAR data, and the fourth is a reference method for a forecast when only pollution data are available. The forecast is for the 12-hour time horizon, and the forecasting methods are based on SODAR and pollution data for the October–March winters of 2015–2019. Reference data from 1 October 2021 to 28 February 2022 were used to compare the methods. Each method was employed to calculate the forecast every hour, and the results were compared with the actual (mathematical statistics: IEA, MAPE, MSE, Thiel’s UHI coefficient) measurements of PM10 concentrations. Thanks to this, a quantitative assessment of each method was made, and a correlation coefficient was derived for the relationship between PM10 measurement and short-term PM10 forecast. Four methods for predicting PM10 concentrations have been developed, with descriptions and application systems of the methods explained below (Figure 8).
4.1 Reference method without SODAR data

Linear regression was used to calculate the coefficients in Equation (3):

\[ PM10(h + 12) = b_0 + b_1 \cos \left( \frac{\pi h}{12} \right) + b_2 \sin \left( \frac{\pi h}{12} \right) + b_3 PM10(h) \]  

where \( h \) is the hour of the day.

4.2 Full regression method

Linear regression was used to calculate the coefficients in Equation (4):

\[ \ln(\text{PM}10(h + 12)) = b_0 + b_1 \cos \left( \frac{\pi h}{12} \right) + b_2 \sin \left( \frac{\pi h}{12} \right) + b_3 \ln(\text{PM10}(h)) + \sum_{c=0}^{\infty} a_c R_F^c(h) \]  

The corresponding transformations led to a weather pattern

\[ \text{PM}10(h + 12) = e^{b_0 + b_1 \cos \left( \frac{\pi h}{12} \right) + b_2 \sin \left( \frac{\pi h}{12} \right) + b_3 \ln(\text{PM10}(h)) + \sum_{c=0}^{\infty} a_c R_F^c(h)} \]  

\( R_F^c \) means the regression coefficient for the characteristic \( c = A, \ldots, M \).

4.3 Hourless regression method

In the full regression method, the arguments (independent variables) removed factors related to daily periodicity since they also characterized SODAR factors. As a result, the following equations were formed:

\[ \ln(\text{PM}10(h + 12)) = b_0 + b_2 \ln(\text{PM10}(h)) + \sum_{c=0}^{M} a_c R_F^c(h) \]  

The corresponding transformations have led to a forecasting pattern:

\[ \text{PM}10(h + 12) = e^{b_0 + b_2 \ln(\text{PM10}(h)) + \sum_{c=0}^{M} a_c R_F^c(h)} \]  

4.4 Search method

There is a method using an algorithm that is more flexible and can accommodate changes and improvements. The main idea of the method is to find the conditions in historical data (data mining, big data, big data form) closest to the state of the atmosphere at the time of forecasting.

Historical data and current state are expressed as a vector of numbers:

\[ X(h) = [\text{PM10}(h), RF_F(h), \ldots, RF_F^n(h)]^T \]  

The historical data for each \( X \) contain the actual values of PM10 concentration after 12 hours, which means that there is a sequence of pairs \( (X(0), \text{PM10}(12)), (X(1), \text{PM10}(13)), \ldots, (X(n), \text{PM10}(n + 12)) \). \( Y \) denotes the vector of the current state. \( Y \) is related to \( X \) and \( X(h) \), and the resulting extended sequence is normalized as follows:

\[ \beta = \frac{p - \min_{\text{max}} - \min_{\text{min}}} {p - \min_{\text{max}} - \min_{\text{min}}} \]  

Where \( \text{max} \) and \( \text{min} \) are the maximum and minimum numbers for the corresponding coordinate. The values of each coordinate are reduced to \( [0,1] \). The sequence \( \{X(0), \ldots, X(n)\} \) is searched for the most similar vectors \( Y \). To avoid the effects of the dimensionality curse, which means poor differentiation of radius in Euclidean spaces greater than 10, a fractional distance is used:

\[ \rho(U, V) = \frac{1}{n} \sum_{i=0}^{n} (U_i - V_i)^2 \]  

where \( k = 1 \).

The most similar vectors are selected \( X(k), \ldots, X(k + 12) \), and the value PM10 \( (k + 12) \). PM10 \( (k + 12) \) is used to generate a forecast for a situation 12 h after the current state. This may be an arithmetic mean, but in practice, the geometric mean is better. The values of \( A \) and \( s \) can be specified at any time. For the purposes of this test, \( k = 0.6 \).

\[ \rho(U, V) = \min_{k=0.1} \rho(U - X(h)) \]  

4.4.1 Forecasting

The most similar vectors are selected \( X(k), \ldots, X(k + 12) \), and the value PM10 \( (k + 12) \). PM10 \( (k + 12) \) is used to generate a forecast for a situation 12 h after the current state. This may be an arithmetic mean, but in practice, the geometric mean is better. The values of \( A \) and \( s \) can be specified at any time. For the purposes of this test, \( k = 0.6 \).

\[ \rho(U, V) = \min_{k=0.1} \rho(U - X(h)) \]  

4.4.2 Verification

The historical data for each \( X \) denote \( \text{PM}10(0), \ldots, \text{PM}10(n + 12) \). To avoid the effects of the dimensionality curse, which means poor differentiation of radius in Euclidean spaces greater than 10, a fractional distance is used:

\[ \rho(U, V) = \frac{1}{n} \sum_{i=0}^{n} (U_i - V_i)^2 \]  

where \( k = 1 \).

The most similar vectors are selected \( X(k), \ldots, X(k + 12) \), and the value PM10 \( (k + 12) \). PM10 \( (k + 12) \) is used to generate a forecast for a situation 12 h after the current state. This may be an arithmetic mean, but in practice, the geometric mean is better. The values of \( A \) and \( s \) can be specified at any time. For the purposes of this test, \( k = 0.6 \).

\[ \rho(U, V) = \min_{k=0.1} \rho(U - X(h)) \]  

4.4.3 Result

Verification of the forecast using the historical data for each \( X \) denotes \( \text{PM}10(0), \ldots, \text{PM}10(n + 12) \). To avoid the effects of the dimensionality curse, which means poor differentiation of radius in Euclidean spaces greater than 10, a fractional distance is used:

\[ \rho(U, V) = \frac{1}{n} \sum_{i=0}^{n} (U_i - V_i)^2 \]  

where \( k = 1 \).

The most similar vectors are selected \( X(k), \ldots, X(k + 12) \), and the value PM10 \( (k + 12) \). PM10 \( (k + 12) \) is used to generate a forecast for a situation 12 h after the current state. This may be an arithmetic mean, but in practice, the geometric mean is better. The values of \( A \) and \( s \) can be specified at any time. For the purposes of this test, \( k = 0.6 \).

\[ \rho(U, V) = \min_{k=0.1} \rho(U - X(h)) \]
For the sample, the values of PM10 (h + 12) were selected for the periods in which the inequality is satisfied. Various descriptive statistics can be calculated based on the sample obtained this way. This may be an arithmetic mean. In practice, the geometric mean was better. In addition, good results were obtained for the median or quantiles (0.35) and (0.6). The values of k and l can be specified at any time. For the purposes of the study, k = 0.6, l = 0.5 was assumed. The sample size was usually around 100.

The undoubted advantage of this method is the constant supplementation of historical data. After 12 hours, when the concentration of PM10 is known, the current state (situation) is considered historical, and these, the latest historical data drive further forecasts. This method also uses the division of modelling results into the geometric mean and the median.

### 4.5 Results

Forecasting methods were verified for the entire population of areas with an average PM10 concentration in Kraków from October 2021 to March 2022 and episodes of high PM10 concentration during which the momentary PM10 concentration was above 100 µg/m³. The initial forecast assessment compared it with a situation in which the forecast would be replaced by an average value (option 0). The first step is to verify the correctness of the forecasting, to determine the forecast errors for the entire data population using the following metrics:

- **MAE** (mean absolute error):
  \[
  \text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t| \tag{11}
  \]

- **MAPE** (mean absolute percentage error):
  \[
  \text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \tag{12}
  \]

- **MSE** (mean square error):
  \[
  \text{MSE} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{13}
  \]

- **IT** - Theil’s coefficient:
  \[
  \text{IT} = \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} y_t^2} \tag{14}
  \]

Where T is the sample size (forecast length), \( y_t \) represents the measurement value, \( \hat{y}_t \) denotes the measurement forecast, and CORR is a correlation factor.

Table 3 Basic statistical characteristics of the differences between forecasts and PM10 measurements for the entire data population from October 2021 to March 2022.

<table>
<thead>
<tr>
<th>Statistical Parameters</th>
<th>Option</th>
<th>No SODAR Method</th>
<th>Fall</th>
<th>Regression Method</th>
<th>Search on</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>(in formula)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MAE (11)</td>
<td>20.8</td>
<td>15.87</td>
<td>16.09</td>
<td>16.15</td>
<td>15.76</td>
<td>16.16</td>
</tr>
<tr>
<td>MAPE (12)</td>
<td>97.1</td>
<td>68.8</td>
<td>67.2</td>
<td>68.2</td>
<td>67.9</td>
<td>69.4</td>
</tr>
<tr>
<td>MSE (13)</td>
<td>727</td>
<td>477</td>
<td>474</td>
<td>475</td>
<td>445</td>
<td>471</td>
</tr>
<tr>
<td>UII (14)</td>
<td>0.326</td>
<td>0.214</td>
<td>0.212</td>
<td>0.213</td>
<td>0.199</td>
<td>0.211</td>
</tr>
<tr>
<td>CORR</td>
<td>0.591</td>
<td>0.657</td>
<td>0.648</td>
<td>0.636</td>
<td>0.626</td>
<td></td>
</tr>
</tbody>
</table>

The results concerning the whole population revealed that the error results were better for all forecasting methods than if the average value was used (Table 3). This means that the forecasts have been made as intended. Analysis of specific model data revealed various statistics on forecast errors, while the MAE (11) achieved the lowest values for the reference method from the no SODAR data (65%). MSE (13) and UII (14) proved that the search method using geometric mean was the best. CORR analysis showed that it...
In the second stage, episodes of high PM10 concentration were distinguished for detailed analysis. Episodes of high PM10 concentration were associated with critical public transport in Krakow (PM10 > 100 µg/m³).

Measurements and forecasts of PM10 concentration were made from October 2021 to March 2022. The trend in the observed concentration from October 2021 to March 2022 was best with the full regression method. The results of such analyses for the whole population did not show a preference for a particular forecasting method. Any four interesting methods are shown in Figures 9 to 12 of the paper. The forecasts of PM10 concentration were made without SODAR data; b) forecast time; c) forecast base; d) forecast time without SO2DAR data; e) forecast base with SO2DAR data; f) forecast base with SO2DAR data; g) forecast base with SO2DAR data; h) forecast base with SO2DAR data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.

The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data. The forecast results for PM10 concentration were compared with the observed data.
Measurements and forecasts for 25 – 30 December 2021:

- a) forecast without SODAR data;
- b) forecast predicted by regression without hours;
- c) full regression forecast;
- d) search method forecast.

**Figure 10.** Measurements and forecasts for 25 – 30 December 2021: a) forecast without SODAR data; b) forecast predicted by regression without hours; c) full regression forecast; d) search method forecast.
Figure 11. Measurements and forecasts for 23–26 January 2022: a) forecast without SODAR data; b) forecast predicted by regression without hours; c) full regression forecast; d) search method forecast.
The above trends in PM10 predictions compared to measurements during PM10 episodes > 100 $\mu g/m^3$ (Figures 9 to 12) revealed that each prediction method underestimated the measured maximum PM10 concentration. It should also be noted that the discrepancy between predictions and measurements changed with episodes. It can be concluded that the meteorological origin of each prediction method underestimated the measured maximum PM10 concentration. It should also be noted that the episode provided a trend result that mimicked the measurements, sometimes with a delay of several hours. The methods fit well under certain circumstances. This was the case, for example, for the forecast (d) for the section 12–16 December 2021 (Figure 9d), forecast (d) for the section 23–30 December 2021 (Figure 10d) and the forecast (c) for the section from 23–26 January 2022 (Figure 11c).

Table 4. Basic statistical characteristics of the differences between PM10 projections and measurements for 100 PM10 episodes $\mu g/m^3$ from October 2021 to March 2022.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Measurement</th>
<th>Forecast without SODAR</th>
<th>Full regression</th>
<th>Search on validation</th>
<th>Median forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE (12)</td>
<td>33.35</td>
<td>26.61</td>
<td>23.46</td>
<td>22.95</td>
<td>24.16</td>
</tr>
<tr>
<td>MAPE (12)</td>
<td>74.1</td>
<td>70.7</td>
<td>62.9</td>
<td>60.8</td>
<td>60.4</td>
</tr>
</tbody>
</table>
The aggregated characterization of the forecast error statistics for episodes (Table 4) revealed that, as with the entire data population, each forecasting method produced better results than the average values, meaning the predictions were useful. As for the statistical parameters studied, the fifth method—regression without hours (with the lowest MAE; UII and MSE and the highest CORR) was the best fit. However, this does not mean this method is universal or suitable for every episode, as shown above.

5 Conclusion

The use of PM10 forecasts for short-term improvement of air quality is becoming more and more frequent in Poland (three-day forecasts and forecasts of the Chief Inspectorate for Environmental Protection, available online: http://powietrze.gov.pl/pl/pjpairPollution). However, air-quality forecasting is rarely used to guide administrative and economic decision-making (e.g. providing free public transport). This is because inaccurate forecasts cause high social costs (disatisfaction of residents) or unjustified financial costs (lack of revenue from public transport tickets). Therefore, applying air quality models to these forecasts must ensure as little loss as possible due to poor decision-making.

The research discussed here shows that using SODAR data to support an air quality forecasting system is reasonable. In particular, the following proposals were made:

- The SODAR model can be complementary to other forecasting methods, as it is highly useful due to its simplicity and speed of calculations.
- The SODAR model does not require emission data, for which temporal and spatial variability are difficult to verify quickly.
- Table 4 shows that, especially at high concentrations, SODAR data provide significant information relative to the model (3) without SODAR.
- The use of simple formulas for regression models in forecasting while maintaining their multivariate (taking into account the four forecast options) facilitates the optimization of the predictive process.
- The model is ready for use, but work is underway to improve it through a different selection of SODAR parameters.

Author Contributions: idea and conception, E.K. and L.O.; methodology, E.K.; software, M.W.; validation, M.W. and E.K.; formal analysis, L.O.; investigation, L.O.; resources, E.K.; data curation, E.K.; writing—original draft preparation, L.O.; writing—review and editing, E.K.; visualization, E.K.; supervision, M.W.; project administration, L.O.; funding acquisition, L.O. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was also developed from the research results produced by the AIR BORDER project, funded by the Ministry of Science and Higher Education of the Republic of Poland, Contract No. 3911/INTERREG VA/2018/2. The research was co-financed from the science fund under INTERREG CE AIR TRITIA, and by the Ministry of Science and Higher Education of the Republic of Poland (Contract No. CE1101). The MONIT-AIR project was co-financed with funds from the European Economic Area’s 2009-2014 Financial Mechanism.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The meteorological data used in this paper (the wind roses) are available to the public: https://danepubliczne.imgw.pl (accessed July 2022). The air quality data used in this paper are available to the public: https://powietrze.gov.pl/pl/pjp/archives (accessed July 2022). The owner of Sodar is the Krakow City Hall, but it is lent to IMWM-PIB. These measurement results are not made publicly available because they constitute raw data and are prepared periodically after validation.

Conflicts of Interest: The authors declare no conflicts of interest.

References


Bokwa, A. Many years of changes in the structure of the mezoklimat city on the example of Krakow; Institute of Geography and Spatial Management of the Jagiellonian University: Krakow, Poland, 2010.


