Dependence of PM10 particles concentration on aerosol optical thickness value from the MODIS data

Jan Harbula[1]

[1]Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc, Tr. Svobody 26, 771 46, Olomouc, Czech Republic
j.harbula@centrum.cz

Abstract. The quality of the environment has an undeniable impact on public health. One of the most important indicators of environmental quality is environment air purity, in which the individual occurs daily. Mainly air quality influences the respiratory diseases above all. PM10 (particular matter) particles are one of the most dangerous pollutants, which get to the lower respiratory tract and can cause serious health complications.

Obtaining the air pollutant concentration values is limited by the number of ground measuring stations and their spatial location. Their network is often very sporadic and estimated or modeled situation of air quality, especially in areas free of these stations, can be highly inaccurate and unreliable on large scales. Thus it is necessary to seek other sources of data to obtain air pollution. Remote sensing data are one of these sources. Satellite and aerial multi-spectral data contain information about aerosol optical thickness values, for which there is evidence of correlation on the concentration of air pollution. Derivation of this correlation leads to the possibility of obtaining more accurate and better data about air pollution in areas where data from the ground measuring stations do not exist. To derive regression of the concentration of PM10 particles on the value of aerosol optical thickness were used MODIS satellite imagery, product MOD04, Level 2, and data from ground measuring stations.

Linear regression was found between measured concentrations of particular matter PM10 and detected AOT values. Angstrom exponent values corresponding to the size of particular matter to 10 µm were determined. Data from the background type of measuring stations were found as the best for retrieving of PM10 concentration relationship from remote sensing imagery. Next work for the improvement of model is necessary.

Keywords: aerosol optical thickness, PM10, air pollution, MODIS

1 Introduction

Air quality is still a huge problem influencing environment around each of us. Many studies demonstrating direct (or indirect) impact of environment on public health have been published, particularly its relevancy as the cause of respiratory diseases. Some authors point to the fact that the particular matter to 10 µm (PM10) concentration has stronger relationship to three respiratory disease categories among children, namely, pneumonia, acute bronchiolitis and the category which includes bronchitis, emphysema and other chronic obstructive pulmonary diseases [7]. According to other authors, increased values of PM10 affect patients with asthma. For emergency visits, the elevation of PM10 concentration, which occurred two days before the visits, had the most significant influence on this type of patient visit with an increase of 0.14% in the four pollutants model [6].
The possibilities of obtaining of air pollutant concentration data are limited depending on the position and the number of ground measuring stations. Their network is not very numerous to allow reliable estimation or modelling of data for high resolution, especially in areas without these stations. Therefore searching for other sources of data that could be used to obtain air pollution information is necessary. One of the ways is using of Remote Sensing imagery. Such studies have already been published by many authors and PM2.5 and PM10 airborne particles were one of the most frequently monitored pollutants.

Studies by [9], [1], [8], [5] and [3] show that the satellite data contains atmosphere PM10 concentration information and there is also correlation between this concentration and the aerosol optical thickness (AOT) and it can be traced up. There was reached the highest value of correlation R = 0.92 [3]. This article deals with the relationship between the PM10 concentration and AOT value in the area of the Moravian-Silesian Region in the Czech Republic.

2 Data

The remote sensed data from MODIS Level 2 products – MOD04 were used. It is represented by the grid of 10×10 km pixel size containing the AOT values in the three bands of 0.47 μm wavelengths (band 1 – *B1*), 0.66 μm (band 3 – *B3*) and 0.55 μm (band 2 – *B2*, derived from B1 and B3 using Angstrom exponent value). In addition, three “Scientific Data Set flags” (SDS) files (the Angstrom Exponent Land, the Quality Assurance and the Cloud Mask) have been used. The data for period of August - September 2007 were selected, then the summer time when the data are the most readable. Prior to the processing, only relatively cloudless days were picked out, their number was 30. For 4 days, two images were available and values for these days were averaged. The time of taking image over the area corresponds to morning hours (cca from 8 a.m. to 11 a.m.).
To the corresponding dates were obtained PM10 concentration data from ground measuring stations (they fall within the AERONET) available on the website of Czech Hydrometeorological Institute. Data represents average daily values of PM10 concentration in μg/m^3. The total number of stations is 24 and the most of them is the background type (18), followed by 4 industrial and 2 traffic type.

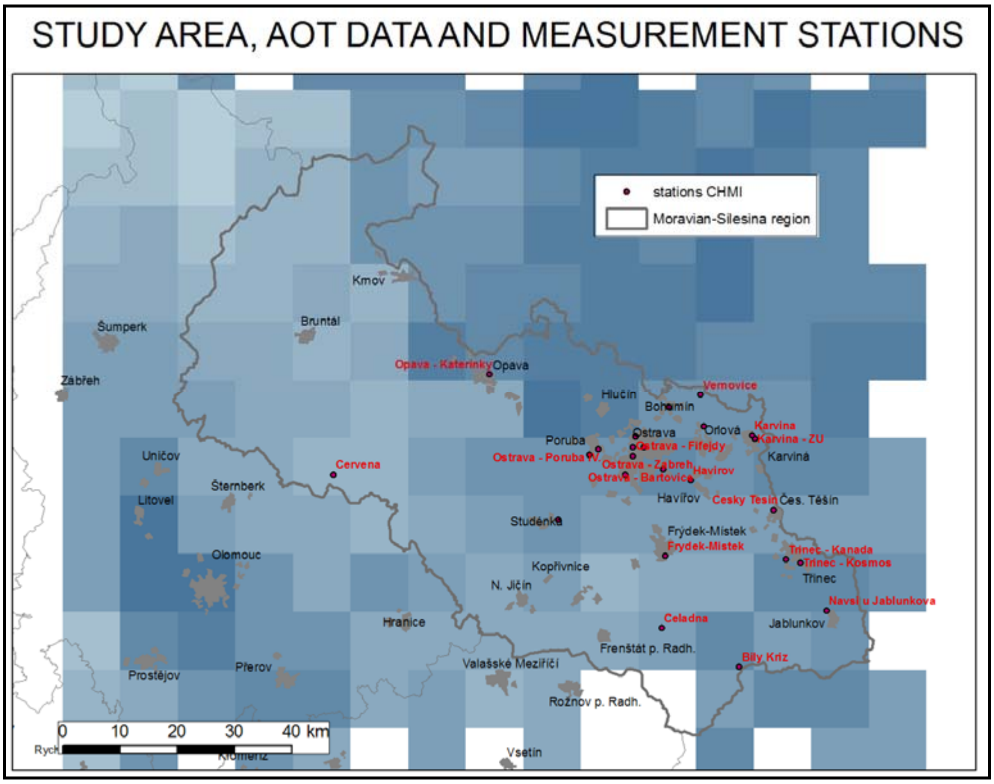

Figure 1. Study area, measuring stations and AOT data raster for 2nd August 2007

3 Methodology of proccessing

Preproccessing of MODIS product data was composed of two substeps: the rectification and the export of required thematic layers. For proccessing the HECtool software was used, that had been developed for this work with MODIS data. Data was rectified into the WGS 84 reference system and the thematic layers SDS flags: the Aerosol Optical Thickness (AOT), the Quality Assurance Land (QA), the Cloud Mask (CM) and the Angstrom Exponent Land (AE) were chosen for export. All output layers were converted to GeoTIFF raster format.
To find the suitable data of the both measured concentration and derived AOT dataset was needed for confirming of potencial linear regression existence between AOT values and PM10 concentration values, the strength of the correlation and retrieving of the dependence.

3.1 Proccessing of AOT data

The Quality Assurance SDS provides data quality information in the pixel. In this SDS (data are for both of different bands B1 - wavelength 0.47 μm and B3 - wavelength 0.66 μm) both information about the overall usability of the pixel and the pixel data quality rate are contained. Considering the similarity of values for the two bands, in the next process only one value for both the original bands (B1 and B3) and also for the derived band B2 was counted.

SDS Quality Assurance and Cloud Mask values

The Cloud Mask SDS provides cloud information in each pixel. Two values were important for the analysis: Determination (yes / no) of the pixel and percentage cloud coverage of the pixel.
The Quality Assurance and Cloud Mask SDS's values are listed with their meanings in Table 1. For the processing only "useful" data were used with Product Quality value greater than or equal to 2 and with maximum of 60% cloud amount (then the Cloud Mask value less than 2).

Table 1. SDS values and their meanings [4]

Flag name	Bit value	Description
Cloud Mask QA		
Cloud mask	0	Undetermined
summary flag	1	Determined
Cloud mask	0	0-30% cloudy pixels
quality flag	1	30-60% cloudy pixels
	2	60-90% cloudy pixels
	3	>90% cloudy pixels
Product Quality QA		
Summary quality flag	0	Not useful data
for AOT	1	Useful
Estimated quality	0	Poor
flag of AOT	1	Marginal
	2	Good
	3	Very Good
	4 – 7	Not Used (TBD)

SDS Angstrom Exponent values

The Angstrom Exponent Land SDS indicates the prevailing Angstrom Exponent (AE) value in the pixel. The AE is connected with the particle size in the atmosphere. That is true that the higher the AE value, the smaller particle diameter. But concrete value can not be specified in general, it always depends on a given territory and on a case study. Therefore it was necessary to determine the range of AE corresponding to the size of searched PM10 in the area of interest.

All of AE data values were between 0.5 and 1.7. In order to trace in detail the AE values, which determine PM10 values, this values range was fractionated to 12 groups, according to the Table 2. Then simple linear regression analysis with groups was computed, the independent variable was the AOT and dependent variable was measured PM10 concentration.

Table 2. Groups and Angstrom Exponent ranges and numbers of points

Group	Range	Number of points	Number of points (%)
A	0.50 – 0.59	9	1,63
B	0.60 – 0.69	132	23,91
C	0.70 – 0.79	13	2,36
D	0.80 – 0.89	8	1,45
E	0.90 – 0.99	7	1,27
F	1.00 – 1.09	11	1,99
G	1.10 – 1.19	0	0,00
H	1.20 – 1.29	43	7,79
I	1.30 – 1.39	11	1,99

J	1.40 – 1.49	1	0,18
K	1.50 – 1.59	14	2,54
L	1.60 – 1.69	19	3,44
M	1.70 – 1.79	284	51,45

Although the Product Quality and the Cloud Mask data were known, missing or erroneous AOT value data may occur in the dataset. Such data are equal to deliberately senseless value -9999 or values less or equal to 0 (zero is the ideal state of absolute atmosphere permeability, the unrealistic in the nature). Therefore the regression analysis was computed only for AOT data greater than 0.

3.2 Proccessing of measured PM10 concentration data

Measured PM10 concentration data from measuring stations of Czech Hydrometeorological Institute are relative to one concrete point with concrete coordinates. In digital form they are represented by point layer with attributes of measured values. First these point layers were adjusted. In several cases there were more than one measuring stations in the area of one pixel in the raster. So in addition there was number of artificially created "averaged" stations. The value of each station is equal to the mean of values from stations, which lie in the area of a pixel from the AOT values raster, thus to the square of 10 $\times$ 10 km. Their position is located to the center of real stations from which the averages were computed. There were created four stations completly, marked "X". Overview of all stations is included with new artificially created stations in the Table 3.

Table 3. Overview of all station with new artificially created stations (B – background, I – industry, T – transport, X – artificial averaged)

Station	Type	Station	Type
Karviná-ZÚ	T	Český Těšín	B
Orlová	B	Návsí u Jablunkova	B
Věřňovice	B	Frýdek-Místek	B
Bohumín	B	Čeladná	B
Karviná	B	Třinec-Kosmos	B
Havířov	B	Studénka	B
Ostrava-Zábřeh	B	Opava-Kateřinky	B
Ostrava-Českobratská (hot spot)	T	Červená	B
Ostrava-Mariánské Hory	I	Bílý Kříž	B
Ostrava-Fifejdy	B	Třinec-Kanada	B
Ostrava-Přívoz	I	X-Orlová-Věřňovice-Karviná	X
Ostrava-Přívoz ZÚ	I	X-Ostrava-Přívoz	X
Ostrava-Poruba IV.	B	X-Ostrava-západ	X
Ostrava-Bartovice	I	X-Ostrava-východ	X
Ostrava-Poruba/ČHMÚ	B		

Selection of suitable concentration data

Corresponding AOT grid value and the corresponding values of selected SDS flags (the Quality Assurance Land, the Cloud Mask and the Angstrom Exponent Land) have been assigned to each point of stations. This procedure was repeated for each day and then all data from all stations and days were transferred to a common database.

To select stations with data the most coresponding to character of AOT data, detected by remote sensing methods, was needed because measured data from measuring stations are very various and they come from various types of stations It means find source of values, which most correlates with data corresponding to the mean of values from square 10 $\times$ 10 km. Type of stations was basic point for division of them. Original (background, industry and trasport) stations are completed by artificially created stations with mark X. Overview of the stations types shown in the Table 3.

4 Results

Due to regression analysis the data character and the regression between detected AOT value and real measured concentration of PM10 were found. First, there was investigated which AE values correspond to particular matter PM10 in the study area, let us say which correlate with them. For better statistical conclusions the significance test with 5 % significance level was computed. And the coefficient of determination was computed for the information about the goodness of fit of a model.

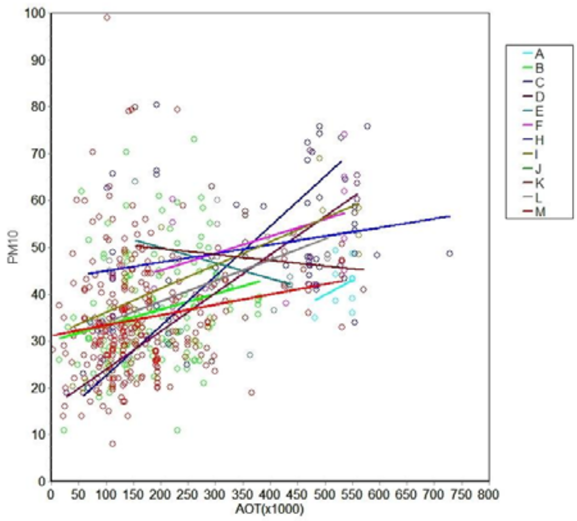

Figure 2. Linear regression according to AE groups

The best correlation coeficient results were achieved in the AE value range of 0.70 – 0.89, it means C and D groups followed by groups F and I. By computing regression analysis only on the data from this groups stronger correlation were proved and correlation coeficient was equal 0.78 in the first band (coefficient of determination R^2 equals to 0.61). These points comprised only 7.79 % of all data and they are not enough as a source of next added data. More points were included in groups B (23.91 %) and M (51.45 %). Because the group B, in the oposite of M, directly concurs with the values of best correlated groups C and D, group B was added into the range. AE values in this group, in the range of 0.60 – 0.69 correlates with the correlation coeficient 0.67 in the first band (R^2 = 0.45).

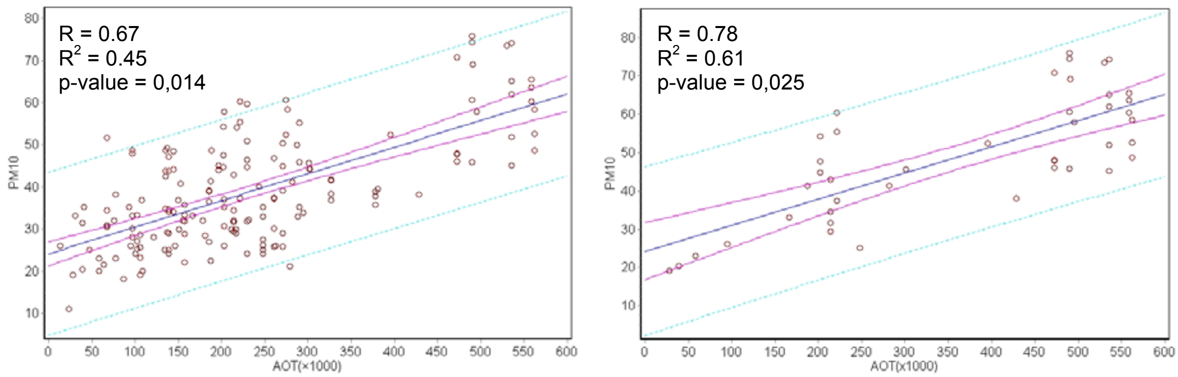

Figure 3. Plots of groups BCDFI (left) and CDFI (right)

Groups B, C, D, F and I were determined the best and the most adequate for the calculation of PM10 concentration. These groups corresponded to the AE values with range of 0.60 – 0.89, 1.00 – 1.09 a 1.30 – 1.39 and include 31.70 % all available points. The results for other two bands were similiar and selected ranges of AE values were same. It is necessary to remember, that using B range values is better to model with other than simple linear regression, or it is important to find other characteristics of them and select the best values to modeling.

Next regression analysis helped to find relationship between detected AOT values and measured concentrations values from various types of stations. According to linear regression which distinguishes types of stations, data (for the first band) from the average stations (R = 0.73, R^2 = 0.54) and background stations (R = 0.72, R^2 = 0.51) correlate the most strongly. Transport and industry stations are similar and they both correlate less significantly, correlation coeficient 0.43, or 0.49.

If are considered only points with AE values corresponding to the groups of CDFI, which had higher correlation coeficient, there is found more significant difference in correlation of various types of stations. In this case the most depended data are data from the background stations, with high correlation coeficient (R = 0.88) and coefficient of determination (R^2 = 0.77). Values from average and industry stations are very similar and they had significantly less regression (R = 0.44 – background and R = 0.41 – industry). In addition, with data from average stations the null hypothesis was not able reject against the alternative in both the BCDFI group and the CDFI group ($p_{(BCDFI)}$ = 0.626, $p_{(CDFI)}$ = 0.193). This means, these data are not usable for building the model. The result for transport stations regression was not possible to calculate in the CDFI range of AE value because of low number of points.
The trend for band 2 and band 3 was very similar and all statistical coefficients values have only small differencies (shown only correlation coefficients):

- *Band 2* – background stations: R = 0.86 (CDFI), R = 0.65 (BCDFI);
  Average stations: R = 0.44 (CDFI), R = 0.73 (BCDFI)
- *Band 3* – background stations: R = 0.84 (CDFI), R = 0.64 (BCDFI);
  Average stations – R = 0.45 (CDFI), R = 0.72 (BCDFI).

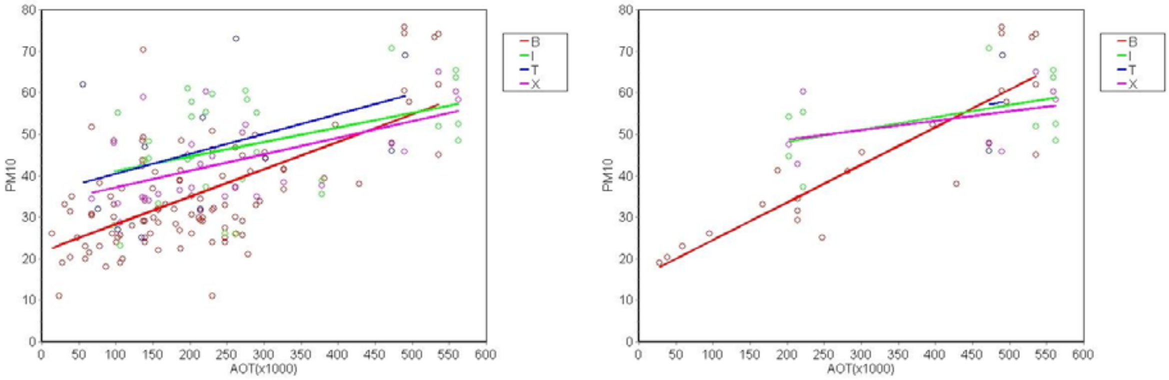

Figure 4. Plots according to the type of stations

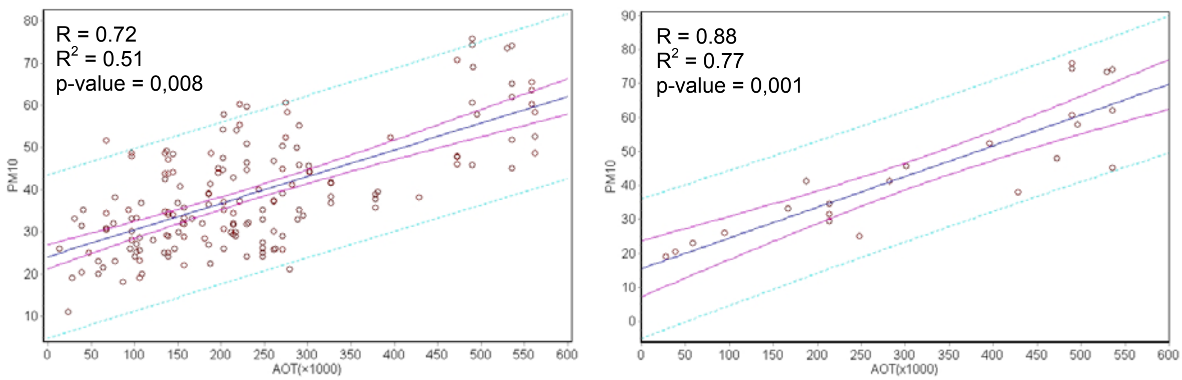

Figure 5. Plots of groups BCDFI (left) and CDFI (right) for background stations only

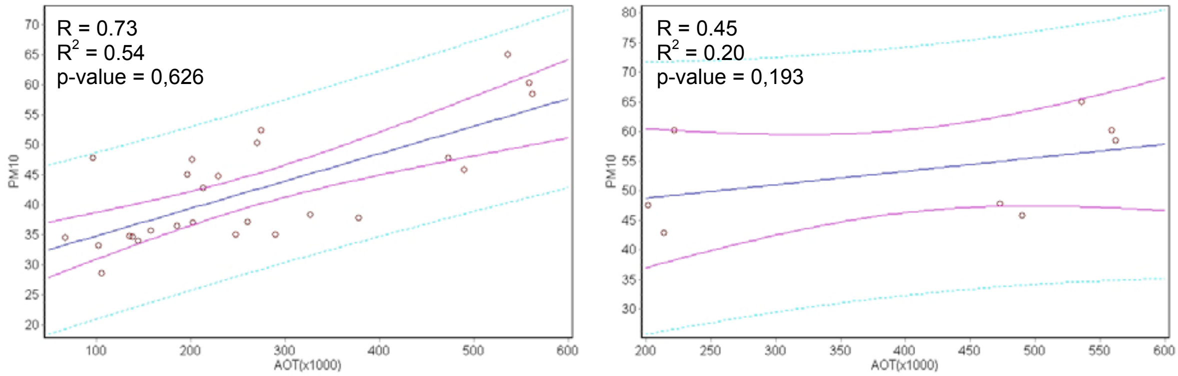

Figure 6. Plots of groups BCDFI (left) and CDFI (right) for averaged stations only

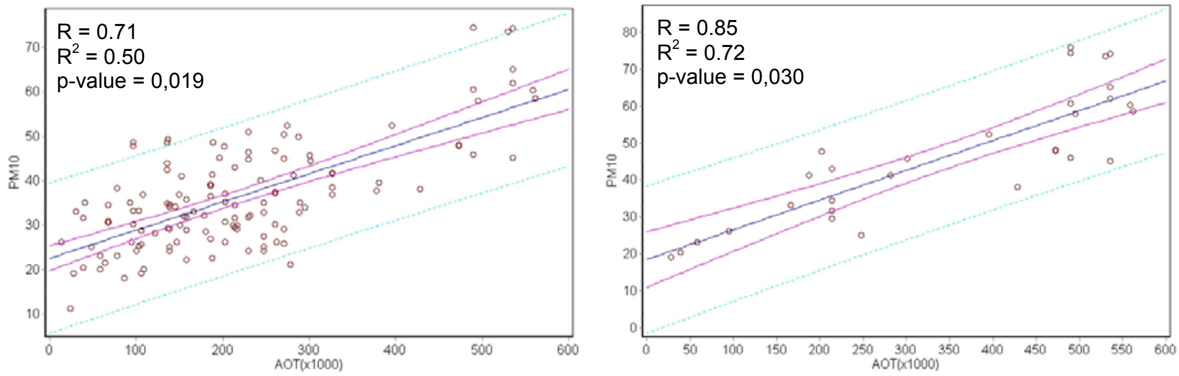

Figure 7. Plots of groups BCDFI (left) and CDFI (right) for averaged and background stations

5 Conclusion

In the study area there was found linear regression between measured concentrations of particular matter PM10 and detected AOT values. AE values were determined, they corresponded to the size of particular matter to 10 μm. Data with AE value from range of 0.60 – 0.89, 1.00 – 1.09 or 1.30 – 1.39 were selected as the most appropriate MODIS AOT product. It is important to explore data from range of 0.60 to 0.69 before the model will be built. Consequently examined dependence was clarified by selection of station type. Using data from background station type is the most suitable for retrieving of relationship of linear regression with real measured values of PM10 concentrations. This fact conforms with the original assumption because compared values correspond to the value of square 10 x 10 km. Then background and mean values of measured concentrations in the square correspond to such a value more than value measured by the local pollution source. Data from the stations artificially created with average values for the pixel are not reliable for retrieving the PM10 concentration values. So it is useless to create these stations.

Derived equation may be used for computing the value of PM10 concentration in an air and it could add points to the sporadic network of points with data from measuring stations. Regression line for the data from band1 has form $PM10 = 20.6 + 68 \times AOT$, for data from band2 $PM10 = 22.68 + 67.3 \times AOT$ and for data from band3 $PM10 = 22.4 + 76.6 \times AOT$. It was confirmed that using the remote sensed data for retrieving the PM10 concentration value is avalaible. Considering the statistical coefficients values it is necessary to improve the model. Next work will be concerned with retrieving of concentrations considering environmental characteristics.

This paper was supported by the grant project GA CR 005/09/1159 "The Intelligent System for Interactive Support of Thematic Map Design".

References

[1] Hutchison, K. D., Faruqui, S. J., Smith, S. Improving correlations between MODIS aerosol optical thickness and ground-based PM2.5 observations through 3D spatial analyses, *Atmospheric Environment vol. 42, pp 530–543,* 2008, ISSN: 1352-2310

[2] King, M. D., Kaufman, Y. J., Tanré, D., Nakajima, T. Remote sensing of tropospheric aerosols from space: Past, present, and future, *Bulletin of the American Meteorological Society, vol. 80, pp. 2229-2259*. 1999

[3] Péré, J.-C. Pont, V., Mallet, M., Bessagnet, B. Mapping of PM10 surface concentrations derived from satellite observations of aerosol optical thickness over South-Eastern France, *Atmospheric Research, vol. 91, pp. 1 – 8*, 2009, ISSN 0169-8095

[4] Remer, L. A., Tanré, D., Kaufman, Y. J.,: Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 005, 2009, [Online] available: http://modis-atmos.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf

[5] Saleh, N. M. Air Quality Derivation utilizing Landsat TM image over Penang, Malaysia, *Recent Advances in Space Technologies, 2007. RAST '07. 3rd International Conference on*, *Istanbul, pp 259 – 26*3, 2007, ISBN: 1-4244-1057-6

[6] Ta-Chien Chan, Mei-Lien Chen, I-Feng Lin, Cheng-Hua Lee, Po-Huang Chiang, Da-Wei Wang, Jen-Hsiang Chuang. Spatiotemporal analysis of air pollution and asthma patient visits in Taipei, Taiwan, *International Journal of Health Geographics, 8:26, 1476-072X*, 2009, [Online] available: http://www.ij-healthgeographics.com/content/8/1/26

[7] Thishan Dharshana, K.G. and Coowanitwong, N., Ambient PM10 and respiratory illnesses in Colombo City, Sri Lanka, *Journal of Environmental Science and Health, Part A, vol. 43, issue 9, pp. 1064–1070 .* 2008, ISSN: 1532-4117

[8] Weber, C., Hirsch, J., Perron, G, Kleinpeter, J., Ranchin, T., Ung, A., Wald, L. Urban morphology, remote sensing and pollutants distribution: An application to the city of Strasbourg, France, *International Union of Air Pollution Prevention and Environmental Protection Associations (IUAPPA) Symposium and Korean Society for Atmosplieric Environmient (KOSAE) Symposium. 12th World Clean Air & Environment Congress, Greening the New Millenniun. 26-31 August*, *Séoul, Korea*, 2001, [Online] available: http://www-cenerg.cma.fr/Public/themes_de_recherche/teledetection/title_tele_air/title_tele_air_pub/paper_urban_morpho.

[9] Zhiyong Hu: Spatial analysis of MODIS aerosol optical depth, PM2.5, and chronic coronary heart disease, *International Journal of Health Geographics, 8:27, 1476-072X*, 2009, [Online] available: http://www.ij-healthgeographics.com/content/8/1/27