Correlation Between Surface Visibility and Aerosol Optical Depth Recorded at Geophysical Observatory at Belsk

Jerzy PODGORSKI

Institute of Geophysics, Polish Academy of Sciences
ul. Księża Janusza 64, 01-452 Warszawa, Poland
e-mail: jp@igf.edu.pl

Abstract

Correlation between surface visibility and aerosol optical depth has been analyzed on the basis of results of observations taken at the Geophysical Observatory at Belsk from March 2001 to May 2005. Marginal and conditional distributions of both variables as well as their cumulative distribution functions have been determined. Covariance in two-dimensional empirical distribution of both variables is negative. The values of correlation ratios from the sample ($e_{xy} = 0.37; e_{yx} = 0.31$) indicate that about 13.5% of the total variation of surface visibility can be explained by influence of aerosol optical depth, while about 9.5% of the total variation of aerosol optical depth can be explained by variations of surface visibility.

The value of correlation coefficient $r = -0.28$ for the study period indicates rather moderate linear correlation between surface visibility and aerosol optical depth. Empirical curves of regression of both characteristics in relation to each other indicate non-linearity of the relationships, while curvilinear indexes of interdependence suggest linearity of the relationships. Examination of linearity of linear regression models constructed on the basis of empirical data indicates that the dependence of both characteristics lies at the limit of linearity. At 0.05 significance level, the dependence of aerosol optical depth on surface visibility is linear, while the dependence of surface visibility on aerosol optical depth is nonlinear.

Statistical tests executed at 0.05 significance level indicate statistical significance of relationship between analyzed characteristics, essentially different from zero values of correlation ratios, and the value of correlation coefficient essentially less than zero.
1. Introduction

Recently, several works examining relationship between aerosol optical depth and surface visibility have been presented (Hand et al. 2004, Li and Lu 1997, Qiu and Yang 2000, Sztyler 2005). The authors pointed at considerable correlation between aerosol optical depth and surface visibility. It may be connected with the fact that aerosol particles in the atmosphere are concentrated mostly in near-surface layer. However, most of comparisons concerned observations that were made in urban or industrial regions. In this study an analysis of correlation between aerosol optical depth and surface visibility on the basis of results of observations taken at the Geophysical Observatory at Belsk – which is located in rural area, away from big cities and industrial regions – is presented.

2. Observation data

Data since March 2001 to May 2005 have been analyzed. Measurements of surface visibility at Belsk are made from the terrace about 10 m above the ground level, according to WMO recommendations (WMO 1983). The method requires various fixed landmarks at known distances from the point of observation. The visual range is the distance to the farthest landmark visible. Values of 10 graded scale used at Belsk Geophysical Observatory for surface visibility measurements are given in Table 1.

<table>
<thead>
<tr>
<th>Surface visibility</th>
<th>Visible landmark</th>
<th>Invisible landmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>–</td>
<td>50 m</td>
</tr>
<tr>
<td>1</td>
<td>50 m</td>
<td>200 m</td>
</tr>
<tr>
<td>2</td>
<td>200 m</td>
<td>500 m</td>
</tr>
<tr>
<td>3</td>
<td>500 m</td>
<td>1 km</td>
</tr>
<tr>
<td>4</td>
<td>1 km</td>
<td>2 km</td>
</tr>
<tr>
<td>5</td>
<td>2 km</td>
<td>4 km</td>
</tr>
<tr>
<td>6</td>
<td>4 km</td>
<td>10 km</td>
</tr>
<tr>
<td>7</td>
<td>10 km</td>
<td>20 km</td>
</tr>
<tr>
<td>8</td>
<td>20 km</td>
<td>50 km</td>
</tr>
<tr>
<td>9</td>
<td>50 km</td>
<td>–</td>
</tr>
</tbody>
</table>

Aerosol Optical Depth (AOD) has been registered at the Geophysical Observatory at Belsk since March 2001 by means of POM-01L Sky Radiometer (Nakajima et al. 1996). As AOD depends, and it is rather strong dependence, on wavelength, AOD obtained in 500 nm wavelength, the closest to maximum sensitivity of human eye, has been chosen from all the channels in which measurements are being made.
On the other hand, as hand-made measurements of surface visibility are executed at the Geophysical Observatory at Belsk once a day within 10–12 GMT interval, and changes in AOD values within a day can exceed 100%, AODs registered within 10–12 GMT interval have been chosen from measurements made all over the day. Averaged AODs for that interval have been calculated, and these averaged values of AOD were correlated with values of surface visibility.

391 pairs (surface visibility, AOD) obtained from measurements which were made within 10–12 GMT interval have been analyzed. Time series of surface visibility and AOD are shown in Figs. 1 and 2, respectively.

Fig. 1. Time series of surface visibility.

Fig. 2. Time series of aerosol optical depth.
Correlation matrix of characteristics $X$ (surface visibility – discrete characteristic) and $Y$ (AOD – continuous characteristic) has been made on the basis of the data sets. The 0.1 spread of class interval has been taken in the case of AOD, so 16 right-hand closed class intervals have been obtained.

3. Results of analysis

Results of some calculations for both examined characteristics are presented in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Surface visibility</th>
<th>AOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value</td>
<td>4</td>
<td>0.044</td>
</tr>
<tr>
<td>Maximal value</td>
<td>9</td>
<td>1.549</td>
</tr>
<tr>
<td>Average value</td>
<td>7.4</td>
<td>0.420</td>
</tr>
<tr>
<td>Variance</td>
<td>0.94</td>
<td>0.092</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.97</td>
<td>0.303</td>
</tr>
<tr>
<td>Variability coefficient</td>
<td>0.1</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Empirical marginal distributions of characteristics are shown in Figs. 3 and 5. Their empirical cumulative distribution functions are shown in Figs. 4 and 6.

![Fig. 3. Empirical marginal distribution of surface visibility.](image-url)
The most numerous are AODs from the 0.2–0.3 interval. AODs greater than 0.8 make only 12% of all cases, but they shift the average AOD by nearly about 70% to the value of 0.420. Nearly 50% of values of AOD do not exceed 0.3.
On the other hand, in marginal distribution of surface visibility large values predominate. Visibilities smaller than 4 (below 1 km) have not been noticed in analyzed period of time. Analysis of variance of the distribution, standard deviation and variability coefficient indicates small diversification of surface visibility and large diversification of AOD.
After standardization of AOD it turned out that there are 6 outliers in empirical set of the \(Y\) characteristic. As it was difficult to establish whether a mistake occurred while taking measurements, those values have not been removed from the data set.

The joint distribution of both characteristics is shown in Fig. 7. The average values of surface visibility and aerosol optical depth have been marked by dashed lines.

The covariance in two-dimensional empirical distribution of surface visibility and AOD is negative. It means that for greater values of surface visibility, values of AOD are generally smaller.

The \(\chi^2\) test has been used to examine the hypothesis of independence (or dependence) of the analyzed variables:

\[
\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \bar{n}_{ij})^2}{\bar{n}_{ij}},
\]

where \(k\) is the number of classes of \(X\) characteristic, \(l\) is the number of classes of \(Y\) characteristic, \(n_{ij}\) is the \(ij\)-th element of correlation matrix \((i = 1, \ldots, k; j = 1, \ldots, l)\), \(\bar{n}_{ij}\) are the expected (hypothetical) cell frequencies in correlation matrix under an assumption that \(X\) and \(Y\) characteristics are independent \((H_0\) is true), and there are \(n\) data items in a sample.

The test is a measure of discrepancy between observed and hypothetical distributions. The \(X\) and \(Y\) values had been separated into groups before the \(\chi^2\) statistic was calculated, so that sufficiently large numbers of observations of hypothetical distribution could be found in every element of correlation matrix of hypothetic frequencies of \(X\) and \(Y\) characteristics. Making the assumption that significance level \(\alpha = 0.05\), the critical value \(\chi^2_{0.05;15} = 24.996\) has been got. As the calculated value \(\chi^2 = 65.4\) is contained in critical region determined by \(P(\chi^2 \geq \chi^2_{\alpha(k-1)(l-1)}) = \alpha\) relation, the null hypothesis about lack of dependence between surface visibility and AOD should be rejected for advantage of an alternative hypothesis. Frequencies in empirical distribution differ too much from those which should occur if variables were independent. The null hypothesis should be rejected also for the significance level \(\alpha = 0.0005\). Hence, with 0.0005 probability of wrong decision, we can state that there is a relationship between analyzed characteristics.

Cramer’s convergence coefficient was used to measure the strength of the relationship:

\[
V = \sqrt{\frac{\chi^2}{n(m-1)}},
\]

where \(m = \min(k, l)\). \(V = 0.24\) was obtained. It indicates rather moderate relationship between surface visibility and AOD.

Conditional distributions of variables specified in correlation matrix were analysed as well, in order to estimate additionally the correlation between variables (sur-
face visibility and AOD). Empirical regression curves have been drawn to state the shape of dependence of analyzed characteristics.

Figure 8 shows empirical curve of regression of surface visibility in relation to AOD, while Fig. 9 shows the second empirical curve of regression (AOD in relation to surface visibility).

Fig. 8. Empirical curve of regression of surface visibility in relation to AOD.

Fig. 9. Empirical curve of regression of aerosol optical depth in relation to surface visibility.
In particular, Fig. 9 indicates the non-linearity of the relationship. When the surface visibility increases, the averaged (conditional) AOD at first increases and then decreases.

Similarly, in Fig. 8, as AOD increases, the surface visibility decreases at first and then, after temporary fluctuations, begins to increase. That may be a result of small number of data with big values of AOD. Besides – as it was mentioned earlier – the six biggest AOD values are untypical.

Analysis of variance between analyzed characteristics has been carried out in order to make an additional determination of the strength of dependence between them. Each time, the other of variables was used as classification factor, and its values were used to divide the population into groups. The values of correlation ratios of one variable in relation to the other are results of that analysis.

We have got \( e_{xy}^2 = 0.1340 \), \( e_{yx}^2 = 0.0930 \), \( e_{xy} = 0.37 \), \( e_{yx} = 0.31 \). This means that 13.4\% of total variability of surface visibility \((X)\) can be assigned as an influence of AOD \((Y)\), and 9.3\% of total variability of AOD \((Y)\) can be assigned to variations of surface visibility \((X)\). So the dependence between surface visibility and AOD, and vice versa, is not too strong.

The statistics:

\[
F_{xy} = \frac{e_{xy}^2/(l-1)}{(1-e_{xy}^2)/(n-l)}
\]

and

\[
F_{yx} = \frac{e_{yx}^2/(k-1)}{(1-e_{yx}^2)/(n-k)}
\]

where \( n \) is the size of the sample, \( l \) is the number of conditional distributions of variable \( Y \), \( k \) is the number of conditional distributions of variable \( X \), \( e_{xy} \) and \( e_{yx} \) are the correlation ratios from the sample, have been used in order to verify the hypothesis of lack of correlation dependence of surface visibility on AOD and vice versa. The lack of correlation dependence of surface visibility on AOD or vice versa was the null hypothesis towards the alternative hypothesis that such a dependence exists. If null hypothesis would be true, statistics \( F_{xy} \) and \( F_{yx} \) would have the F-Snedecor’s distribution with \( v_1 = l - 1 \) \((v_1 = k - 1)\) and \( v_2 = n - l \) \((v_2 = n - k)\) degrees of freedom. The critical region of the test is specified by the value of \( F_\alpha \) that fulfils the relation: \( P(F \geq F_\alpha) = \alpha \), where \( \alpha \) is the significance level. Values of \( F_{xy} = 3.87 \) and \( F_{yx} = 7.90 \) have been obtained.

For the significance level \( \alpha = 0.05 \) the critical levels are:

\[
F_{0.05} = F_{0.05;15;375} = 1.67,
\]

\[
F_{0.05;5;385} = 2.21.
\]

As values of both statistics, \( F_{xy} \) and \( F_{yx} \), calculated from the sample exceeded the critical values, the null hypotheses about the lack of dependence of surface visibil-
ity on AOD and vice versa should be rejected for the advantage of an alternative hypotheses. This means that correlation ratios from the sample took values significantly different from zero.

The next parameter that was used for estimating the strength of relation between variables was the correlation coefficient that measures the linear correlation of variables:

$$r = \frac{c_{xy}}{s_x s_y},$$

where $c_{xy}$ is the covariance in two-dimensional empirical distribution of $X$ (surface visibility) and $Y$ (AOD) variables, $s_x$ and $s_y$ are the standard deviations in empirical marginal distributions of $X$ and $Y$ variables.

The result of $r = -0.28$ has been obtained for the study period. This result indicates rather moderate linear correlation between surface visibility and AOD with negative direction. However, the estimation of strength of this relation can be underrated if it would be – as empirical curves of regression indicate – nonlinear. Then correlation ratios would be more suitable as measures of the strength of correlation. Therefore, the differences between correlation ratios and correlation coefficient, which are measures of nonlinearity, have been calculated. We obtained:

$$m_{xy} = e_{xy}^2 - r^2 = 0.0554,$$

$$m_{yx} = e_{yx}^2 - r^2 = 0.0145.$$  

As the obtained values of measures of nonlinearity are not too big, it seems that relations between variables are linear.

Because relations between correlation ratios and correlation coefficient can be only the basis for general estimation whether the dependence between variables can be recognized as linear or nonlinear, the linear dependence between investigated characteristics has been assumed and classical models of linear regression have been built on the basis of experimental data, showing how conditional expected values of one of the variables change according to the values, the other variable adopts. Then the examination of their curvilinearity has been done. This allowed to draw out more precise conclusions concerning linearity of regression. The sum of square deviations that are not explained by regression has been separated into the sum of square deviations that represent pure error and square deviations resulting from the lack of fitting of straight line to empirical data. Errors of fitting were differences between the values of straight line function fitted by the least squares method and the values of empirical regression resulting from averaged values of $Y$ variable in subsets of the sample with the same values of $X$. Comparison of averaged square deviations for fitting error (of model) and pure error made it possible to verify the hypothesis whether regression function $Y$ in relation to $X$ is linear in population. The testing statistic had the form:
The numerator of the fraction represents the mean square of fitting error (error of model) while the denominator represents the mean square of pure error. Factor \((r - 2)\) is the number of degrees of freedom for the sum of squares of fitting error, while \((n - r)\) is the number of degrees of freedom for the sum of squares of pure error. Factor \(r\) is the number of classes of \(X\) variable, while \(n_i\) expresses the number of repeated observations for \(X_i\).

If regression \(Y\) in relation to \(X\) is linear then \(F\) statistic has \(F\)-Snedecor’s distribution with the number of degrees of freedom of numerator \(v_1 = r - 2\) and of denominator \(v_2 = n - r\). Critical region for this test – at \(\alpha\) significance level – is expressed by \(F_{\alpha, r-2, n-r}\) value that comply with the condition:

\[
P(F \geq F_{\alpha, r-2, n-r}) = \alpha.
\]

If the \(F\) value is not less than \(F_{\alpha, r-2, n-r}\) (critical), then we reject the hypothesis about linearity of regression.

Taking as \(X\) the surface visibility and AOD as \(Y\), we got: \(F = 1.5197, F_{\alpha, r-2, n-r} = F_{0.05,4,3.85} = 2.37\) and \(F < F_{0.05,4,3.85}\), so there is no reason for rejecting the null hypothesis about linearity of regression of AOD in relation to surface visibility.

Taking AOD as \(X\) and surface visibility as \(Y\), we got: \(F = 1.7883, F_{\alpha, r-2, n-r} = F_{0.05,14,3.75} = 1.71\) and \(F < F_{0.05,14,3.75}\), so – at the 0.05 significance level – the null hypothesis about linearity of regression of surface visibility in relation to AOD should be rejected.

At last, the hypothesis that correlation coefficient equals zero has been verified. If this hypothesis would be true, the statistic:

\[
t = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2},
\]

where \(r\) is an empirical correlation coefficient from the sample, would have \(t\)-Student’s distribution with \((n-2)\) degrees of freedom. This time the critical region is defined by the following relation:

\[
P(|t| \geq t_\alpha) = \alpha,
\]

where \(\alpha\) is the significance level, \(t_\alpha\) is the critical value taken from \(t\)-Student’s distribution table.

As the alternative hypothesis is the hypothesis saying that correlation coefficient is (significantly) less than zero, the one sided test is appropriate in this case. It is defined by the relation:

\[
P(t \leq -t_\alpha) = \alpha.
\]
Taking $\alpha = 0.05$ we obtained:

\[ t = -5.76 , \]
\[ t_{2\alpha} = t_{0.1;389} = 1.645 . \]

As $t = -5.76 < -1.65 = -t_{0.1;389}$, the value of statistic from the sample lies inside the critical region and the hypothesis about the lack of dependence between surface visibility and AOD should be rejected for advantage of an alternative hypothesis. This means that the value of correlation coefficient from the sample is significantly less than zero.

References


Li, F., and D. Lu, 1997, Features of aerosol optical depth with visibility grade over Beijing, Atmospheric Environment 31, 20, 3413-3419.


Received August 21, 2006
Accepted September 18, 2006

KORELACJA POMIĘDZYZ WIDZIALNOŚCIĄ METEOROLOGICZNĄ I GRUBOŚCIĄ OPTYCZNĄ ATMOSFERY NA PODSTAWIE POMIARÓW WYKONANYCH W OBSERWATORIUM GEOFIZYCZNYM W BELSKU

Streszczenie

Na podstawie materiału obserwacyjnego z okresu od marca 2001 do maja 2005 roku. zebranego w Obserwatorium Geoﬁzycznym w Belsku, zbadano zależności ko-
relacyjne pomiędzy widzialnością meteorologiczną mierzoną za pomocą reperów i grubością optyczną aerozolu mierzoną radiometrem. Określono rozkłady brzegowe i warunkowe obu cech oraz dystrybuanty tych rozkładów. Średnia widzialność meteorologiczna mierzona w dziesięciostopniowej skali w godzinach południowych wynosiła 7,4 (ok. 15 km), zaś średnia grubość optyczna aerozolu wyniosła 0,420. Ko-wariancja w dwuwymiarowym rozkładzie empirycznym obu cech ma wartość ujemną. Wartości stosunków korelacyjnych z próbą (e_yx = 0,37; e_xy = 0.31) wskazują na to, iż ok. 13,5% całkowitej zmienności widzialności meteorologicznej może być przypisana wpływowi grubości optycznej aerozolu, zaś ok. 9,5% całkowitej zmienności grubości optycznej aerozolu może być przypisana wpływowi widzialności meteorologicznej. Wartość współczynnika korelacji r = –0,28 wskazuje na umiarkowane liniowe skorelowanie widzialności meteorologicznej i grubości optycznej aerozolu. Empiryczne krzywe regresji cech względem siebie wskazują na nieliniowość związków, zaś wskaźniki krzywoliniowości zależności jednej cechy od drugiej sugerują liniowość związków. Badanie liniowości zbudowanych na podstawie danych empirycznych liniowych modeli regresji wskazuje na leżącą na granicy liniowości zależność obu cech. Na poziomie istotności 0,05 zależność grubości optycznej aerozolu od widzialności meteorologicznej jest liniowa, zaś zależność widzialności meteorologicznej od grubości optycznej aerozolu jest nieliniowa. Przeprowadzone na poziomie istotności 0,05 testy statystyczne wskazują na statystyczną istotność związku między analizowanymi cechami, istotnie różne od zera wartości stosunków korelacyjnych oraz istotnie mniejszą od zera wartość współczynnika korelacji.