Interpretation of the results on the cloud field photometry by applying the model of compound signal. Account for the unsteadiness of observation series

V.K. Oshlakov

Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk

Received May 7, 2004

We discuss the problems on the account of the unsteadiness of the time series of observations over the solar radiation brightness under cloudy conditions. Use of the model of signals with time compression to approximation of statistical characteristics of the brightness fluctuations observed at a cloud layer cross section is justified.

Introduction

One of the problems in the interpretation of observations of the brightness field of the downwelling solar radiation under cloudy conditions is the unsteadiness of the observation series. This paper is a continuation of the studies presented in Ref. 1 and I consider here the problem on the account of the unsteadiness in the series of observations of fluctuations of the solar radiation brightness. The validity of applying the model of signals with temporal compression to the interpretation problem stated is also analyzed.

The account of trends

In recording the sky brightness along some direction and within a narrow wavelength range, the linear, aperiodic, and quasiperiodic trends of the mean values and variances of the brightness essentially affect the series of observations. The main reasons of the brightness trends are the following: diurnal behavior of the Sun coordinates, spatiotemporal variability of the characteristics of a cloud layer, the variability of the optical thickness and the phase composition of the elements of a cloud layer. The view of a trend in a particular realization depends on the duration of the realization and the conditions of observations.[2,3]

The problem on isolating a trend is traditionally solved by means of smoothing procedures and numerical filtration of the temporal series.[2,4] At the same time, one can exclude the effect of, for example, the Sun position by performing observations of brightness within several wavelength ranges, varying the dimension of the data space and analyzing the variability of the color coordinates on the three-component diagram (TCD). In this case the processing of the results of observations becomes much simpler.[5–7] The matter is that the problem of choosing the information sign in the problems of classification, that is an independent problem, and this problem will be discussed in subsequent publications.

It is known that the modal radius of water droplets, in a liquid-droplet clouds is 4–5 μm.[8] Hence, if one compares the ratios of the read-outs of the brightness $a(\lambda_i)$ of, for example, cloudy and clear sky at zenith observed in different regions of the visible wavelength range (at its boundaries and in the center), then, according to Mie theory, the following relationship should be fulfilled for the cloudless sky $a(\lambda_1) > a(\lambda_2) > a(\lambda_3)$ (blue sky) and for cloudy sky it should be $a(\lambda_1) < a(\lambda_2) > a(\lambda_3)$ ("white" color), because such a relationship is characteristic of the source of radiation – the Sun. The differences manifest themselves qualitatively in the displacement of the maximum of the spectral distribution of energy from "green" region of the spectrum (for cloud particles) to shorter wavelengths (for the cloudless sky).

Variability of the positions of the color coordinates of the controlled information sign – spectral brightness – on TCD is explained by the variability of the optical characteristics of cloudiness and the cloudless sky (for example, the optical thickness) related to the variations of their composition (including the water phase composition of clouds) and, hence, the variability of the spectral distribution of radiation from the source with the constant spectral characteristics, Sun in the given case, scattered by atmospheric particles.[6,7] The conclusion on the possibility of making classification of the signals of spectral brightness by the sign "cloudiness" is based on the analysis of distributions of the brightness presented in Refs. 3 and 6.

One can present the results of statistical processing of one of the realizations as an example of isolation of the trend related to the change of the Sun position. Observations of the zenith brightness were carried out at three wavelengths: $\lambda_1 = 0.42$, $\lambda_2 = 0.53$, and $\lambda_3 = 0.69$ μm in winter at *Sc* cloud type and cloud fraction of 10. The time of

0235-6880/04/08 604-05 $02.00

observations was since 10 a.m. until 5.30 p.m. Precipitation was observed in the afternoon. The time step in data recording was 36 s. Preliminary data analysis was carried out taking into account the data of Hydrometeorological service and local station equipped in accordance with the requirements needed. The example was chosen as if only the layer thickness was changed during the cloud layer passage over the observation site, while the phase composition of cloudiness is being unchanged. In processing, the realization was divided into parts according to calculated data on the angular coordinates of the Sun for the given observation site: before noon, culmination, afternoon (zenith angles of the Sun less than 85°). It was supposed taking into account *a priori* data that the trends in the selected parts are close to linear.

The rank correlation coefficients $\hat{\tau}$ were estimated, as a criterion to the linear trend, for the sequences of the signal amplitudes at each wavelength $a(\lambda_1)$, $a(\lambda_2)$, $a(\lambda_3)$, the total signals S (the sum of the amplitudes) and color-divided signals m, n, l (the ratio of the amplitude at each wavelength to the total signal). Actually, these are the coefficients of correlation between the order of the signal amplitudes in the sequence of readouts and their order of magnitude. Then the amplitudes were calculated of the trend line and the value tanφ was determined – the tangent of the trend line slope. It was noted in Ref. 2 that the rank correlation coefficient τ is quite "powerful" criterion of the tendency to the linear trend: its asymptotic relative efficiency is equal to 0.98. The mathematical expectation τ is equal to 0, the variance is $\sigma^2(\tau) = \dfrac{2(2n+5)}{9n(n-1)}$, the distribution quickly tends to normal. To estimate the significance of the calculated value $|\hat{\tau}|$, it is enough to compare it with theoretical value $|3\sigma(\tau)|$ (testing of the zero hypothesis about the mean value $\tau = 0$ against the alternative: mean value τ is equal to the calculated value $\hat{\tau}$). In this case the value $3\sigma(\tau)$ is equal to 0.11. The values of the calculated rank correlation coefficients and the values tanφ of the trend lines are presented in the Table for the initial sequences at three wavelengths, the total signal, and three color-divided signals.

Signal sequence	Before noon		Afternoon	
	$\hat{\tau}$	tanφ	$\hat{\tau}$	tanφ
$a(\lambda_1)$	0.431	0.172	–0.854	–0.287
$a(\lambda_2)$	0.354	0.508	–0.819	–0.948
$a(\lambda_3)$	0.33	0.301	–0.787	–0.562
$S = \Sigma\, a(\lambda_i)$	0.354	0.979	–0.353	–0.158
$m = a(\lambda_1) / S$	0.304	0.000017	0.39	0.000017
$n = a(\lambda_2) / S$	0.166	0.000018	–0.684	–0.000047
$l = a(\lambda_3) / S$	–0.171	–0.000015	0.635	0.000032

The values of the slope tangent for the sequences of color-divided signals are practically equal to 0 at significant estimates of the tendency to a linear trend.

Testing the hypotheses on the variances and mean values of the spectral brightness observed

To test validity of the model chosen, the hypotheses were tested on the difference between the mean values of the brightness observed under cloudy and clear sky conditions. Observations were carried out at three wavelengths, namely at 0.42; 0.53, and 0.69 μm. Simultaneously recorded were the meteorological parameters; the height of the lower boundary was estimated, as well as the direction and velocity of the cloudiness motion, and the realizations were selected with coinciding parameters. In testing the hypotheses, 12 realizations were analyzed with duration from 3.5 to 7.5 hours (the observations were conducted along zenith direction): 7 for broken cloudiness and 5 for cloudless situations.

It was supposed in the analysis that the brightness of a break in a cloud layer tends to the brightness of the cloudless sky along the same angular direction and at the same Sun position; the sequence of readouts was represented by means of the model.[3,9,11] Let us remind that in this model the mean value $\overline{B_{\text{obs}}}$ and the mean square of the observed brightness are represented by linear dependences:

$$\bar{B}_{\text{obs}}(p) = p\bar{B}_1 + (1-p)\bar{B}_0 = p(\bar{B}_1 - \bar{B}_0) + \bar{B}_0, \quad (1)$$

$$\overline{B^2_{\text{obs}}(p)} = p\overline{B^2_1} + (1-p)\overline{B^2_0} = p(\overline{B^2_1} - \overline{B^2_0}) + \overline{B^2_0}, \quad (2)$$

and the variance σ^2_{obs} is presented by the square dependence on the probability of the presence of a cloud on the vision line[6] p:

$$\sigma^2_{\text{obs}}(p) = -p^2(\bar{B}_1 - \bar{B}_0)^2 + p[(\overline{B^2_1} - \overline{B^2_0}) - 2\bar{B}_0(\bar{B}_1 - \bar{B}_0)] + \sigma^2_0. \quad (3)$$

Indices "obs", 1, 0 in Eqs. (1)–(3) relate the respective values to the classes "observed value," "cloud," and "break in a cloud."

It is supposed that the sequences analyzed are free of trends related to the change of the Sun position.

Relationships between the sample variances

In testing the hypotheses on the equality of the variances of two samples, observed under cloudy conditions and in the absence of cloudiness, the statistic criterion (assuming that the distributions of the brightness readouts in two classes of values are normal) is the ratio of the sample variances[10]:

$$F^* = \sigma^2_{\text{obs}} / \sigma^2_0 > 1, \quad (4)$$

i.e., the larger of the sample variances is put in the nominator. Then F^* is compared with the critical value from the Fisher F-distributions $F_{\nu_{obs},\nu_0,1-\alpha/2}$, where α is the level of significance, $\nu_{obs} = N_{obs} - 1$ and $\nu_0 = N_0 - 1$ are the numbers of the degrees of freedom corresponding to the values σ^2_{obs} and σ^2_0, N_{obs} and N_0 are the sample sizes. The zero hypothesis is rejected if the ratio of two sample variances exceeded the critical one.

Taking into account Eq. (3), the ratio of the variances in the discussed representation is written in the form:

$$F^*(p) = \frac{\sigma^2_{obs}(p)}{\sigma^2_0} = 1 + \frac{-p^2(\bar{B}_1 - \bar{B}_0)^2 + p[(\overline{B^2_1} - \overline{B^2_0}) - 2\bar{B}_0(\bar{B}_1 - \bar{B}_0)]}{\overline{B^2_0} - (\overline{B_0})^2}. \quad (5)$$

As the brightness variance under cloudy conditions depends on the probability of the presence of a cloud on the vision line, then F^* also depends on p. If the number of the degrees of freedom has been much greater than 120, theoretical values of the Fisher F-distribution are close to 1. In this case the zero hypothesis, according the criterion (5), is accepted, if one of two conditions has been fulfilled: either $p = 0$ in the tested samples, or $p \neq 0$, but simultaneously $\overline{B_1} = \overline{B_0}$ and $\overline{B^2_1} = \overline{B^2_0}$. The model (1)–(3) loses meaning at fulfilling of the second conditions. Besides, it follows from the expression (1) that if $p \neq 0$, the values $\overline{B_1} \neq \overline{B_0}$ and the ratio of variances in Eq. (5) is greater than 1 for all $\overline{B_{obs}} \neq \overline{B_0}$. The restrictions to be imposed on the ratio of the mean squares are not determined. Hence, the criterion (5) should be sensitive at all values $p > 0$ and at fulfilling the condition of inequality of the mean brightness values of the cloud and the break in a cloud layer.

The ratios between the sample mean values

Let us consider the procedure of comparing the mean values of two populations: the reference "clear sky" and "observed under cloudy conditions."

The criterion for testing the mean values of two populations in the case when both variances are unknown and are not assumed to be equal is the following:

$$T = \overline{B_{obs}} - \overline{B_0} \Big/ \sqrt{\frac{\sigma^2_{obs}}{N_{obs}} + \frac{\sigma^2_0}{N_0}}, \quad (6)$$

its distribution is close to the Student t-distribution with the number of the degrees of freedom ν lying between the least value of $(N_{obs} - 1)$ and $(N_0 - 1)$ and their sum $(N_0 + N_{obs} - 2)$.[10]

Substituting the values $\overline{B_{obs}}(p)$ and $\sigma^2_{obs}(p)$ from Eqs. (1) and (3) into the Eq. (6), we obtain

$$T(p) = p(\overline{B_1} - \overline{B_0}) \Big/ \sqrt{\frac{\sigma^2_{obs}(p)}{N_{obs}} + \frac{\sigma^2_0}{N_0}}. \quad (7)$$

As both the mean value and variance of the observed brightness fluctuations depend on the probability of the presence of cloud on the vision line, the statistics of the criterion for testing the mean values depends on this value too. It follows from Eq. (7) that testing the hypothesis on the equality of the mean values is reduced to testing the condition of equality of the mean values of the brightness of a cloud and a break in the cloud layer. This condition is tested taking into account the value of the probability of the presence of a cloud on the vision line $p \neq 0$ and the estimates of the variances of the observed brightness $\sigma^2_{obs}(p)$ and clear sky σ^2_0, and that the estimates of the variances cannot be equal to zero simultaneously.

In practice the procedure of testing the hypothesis is reduced to testing the condition $T > t_{\nu,1-\alpha/2}$. If the condition has been fulfilled, the zero hypothesis on the equality of the mean values is rejected, hence, observations have been carried out under cloudy conditions. This condition for the set of the values α, B_0, σ^2_0, and observed $\overline{B_{obs}}$, σ^2_{obs} after small transformations taking into account Eq. (3) is written in the following form:

if

$$N_{obs}\left\{\left[\frac{(\overline{B_{obs}} - \overline{B_0})}{t_{\nu,1-\alpha/2}}\right]^2 - \frac{\sigma^2_0}{N_0}\right\} > \sigma^2_{obs}, \quad (8)$$

or, taking into account Eq. (1) for the set of p values

if

$$N_{obs}\left\{\left[\frac{p(\overline{B_1} - \overline{B_0})}{t_{\nu,1-\alpha/2}}\right]^2 - \frac{\sigma^2_0}{N_0}\right\} > \sigma^2_{obs}(p), \quad (9)$$

then for Eq. (8) the observed brightness values are not equal to the cloudless sky brightness values, and for Eq. (9) – at the set of the probability value of the presence of a cloud on the vision line the cloud brightness is not equal to the cloudless sky brightness. Hence, the sample of the observed values is related to observations under cloudy conditions.

Joint distributions of readouts relative to the mean values

It follows from the above stated formulations that the results of observations of brightness along a preset direction can be represented by a mixture of few classes of signals different in the mean values and, possibly, by the variances of the observed values. It is known that dividing the mixture of signals into a preset number of classes is performed on condition of inequality of the mean values in different classes of signals. The problem of detecting signals can be solved resulting from analysis of the empiric distributions of signals in the mixed sample. The sample of observed values can be described by bimodal distribution of probabilities of the mixture of signals from the cloud and in a break in the cloud layer:

$$f_p(x) = p_1 f_1(x) + p_0 f_0(x), \qquad (10)$$

where f_1 and f_0 are the distribution laws of the signals in the classes "cloud" and "break in the cloud layer," the probabilities p_1 and $p_0 = (1 - p_1)$ are *a priori* known.[12]

The following conclusions can be drawn from the analysis of the empirical distributions available. In the absence of trends in the mean values and the variances, the distribution of, for example, the results of processing the series of signals of the zenith color under conditions of broken cloudiness is subordinate to the distribution of the mixture of signals (10) (Fig. 1). Analogous conclusions have been drawn from processing the results of observations at a single wavelength (Fig. 2).

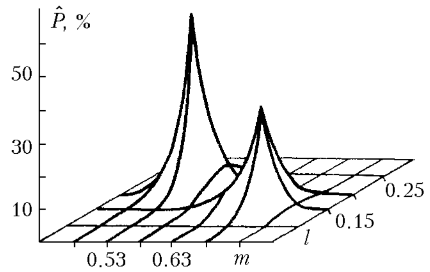
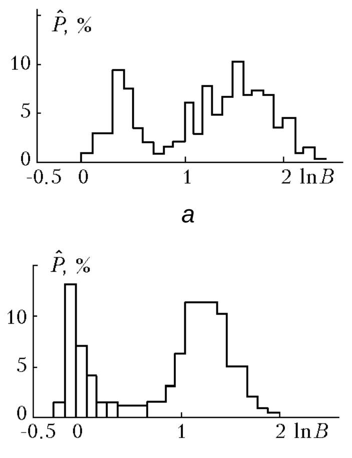


Fig. 1. Probability distribution of the zenith color coordinates under conditions of cumulus cloudiness.[6]



a

b

Fig. 2. Histograms of the distribution of the photometric readouts from cloudy zenith (λ = 0.69 μm): (*a*) before correction; (*b*) after correction for the height of the lower boundary of cloudiness and spectral sensitivity of the photometer.[3]

Let us remind that, in processing, the effect of trend related to the change of the Sun position was excluded, and, if necessary, corrections were introduced for the spectral sensitivity of the photometer and for the variations of the height of the lower boundary of cloudiness.[3,6] The hypotheses on the equality of the mean values of the signals and their variances in the presented examples (Figs. 1 and 2) were rejected at the level of significance of 0.01.

The variability of the position of the color coordinates on TCD (see Fig. 1) is explained by the difference in the optical characteristics of cloudiness and the cloudless sky related to their composition, and, hence, to the difference in the spectral distribution of the scattered radiation from the source with constant spectral characteristic (Sun).

As a rule, one usually removes the determined component (centers to the mean values) from the series of observations and then normalizes it to the rms deviation in order to reduce observations to the series with the zero mean value and unit variance. Further conclusions about the fluctuation components are drawn based on the analysis of the "residues." General form of the correlation function of fluctuations of brightness of solar (optical) radiation under conditions of broken cloudiness is presented in Ref. 1. It follows from analysis of the analytical form of the correlation function,[1] that if centering the initial series of the observed brightness to the mean value (1), the components will remain in the correlation function and then in the spectral density, which are related to the spatiotemporal inhomogeneity of both the elements of the cloudy field and the field itself. In order to exclude the effect of the "internal" inhomogeneities of the elements of the cloudy field, one should center to the mean value and normalize to the rms deviation in each class of signals separately. In this case, description of the observed processes can be reduced to combination of their mean values and rms deviations, what is confirmed by the form of empirical distributions of the observed brightness fluctuations (see Figs. 1 and 2).

Conclusions

Validity of using the model of signals with time compression in the problems of interpretation of the results of observations of the brightness fluctuations under cloudy conditions is tested using the results of processing the observations. The conclusions are drawn about significance of the difference in the mean values and the variances of the observed brightness of solar radiation scattered by the cloud field along a selected direction.

The presented relationships are used in the following cases:

– at classification of the samples with preliminarily removed trends;

– for testing the hypotheses about the relationship between the mean values and the variances of the samples obtained under the same conditions of observations;

– for formulation of the decision making rules on the presence of cloudiness on the vision line.

Two classes of signals are considered in the following combinations: cloudless sky and cumulus cloudiness (40 3.5-hour long realizations with time step of 15–20 s with preliminary classification according to the signs: season of observation, type of cloudiness, cloud fraction, Sun position), cloudless sky and cloudiness of arbitrary type – in all other cases. The possibility of classifying the signals into subsets is mentioned: clouds of liquid-droplet structure and clouds of crystal structure.[6]

Acknowledgments

The work was supported in part by Russian Foundation for Basic Research (grant No. 02–05–65169–à).

References

1. V.K. Oshlakov, Atmos. Oceanic Opt. **17**, No. 4, 305–309 (2004).
2. M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Griffin, London, 1958).
3. V.K. Oshlakov, Atmos. Oceanic Opt. **3**, No. 3, 255–259 (1990).
4. R.E. Bogner and A.G. Konstantinides, eds., *Introduction to Digital Filtering* (Wiley, New York, 1975).
5. R.E. Bykov and S.B. Gurevich, *Analysis and Processing of Color and Volume Images* (Radio i Svyaz', Moscow, 1984), 248 pp.
6. V.K. Oshlakov, Atmos. Oceanic Opt. **3**, No. 4, 392–396 (1990).
7. V.P. Galileiskii, A.M. Morozov and V.K. Oshlakov, in: *Regional Monitoring of the Atmosphere.* Part 2. *New Devices and Techniques for Measurements* (Spektr, Tomsk, 1997), 295 pp.
8. I.P. Mazin and A.K. Khrgian, eds., *Clouds and Cloudy Atmosphere*. Handbook (Gidrometeoizdat, Leningrad, 1989), 647 pp.
9. L. Franks, *Signal Theory* (Prentice-Hall, 1969).
10. N.L. Johnson and F.C. Leone, *Statistics and Experimental Design* (Wiley, New York, 1977).
11. A.S. Shalygin and Yu.I. Palagin, *Applied Methods for Statistical Modeling* (Mashinostroenie, Leningrad, 1986), 320 pp.
12. A.V. Milen'kii, *Classification of Signals under Conditions of Uncertainty. Statistical Methods for Self-Teaching in Image Recognizing* (Sov. Radio, Moscow, 1975), 328 pp.