Cloud cover estimation optical package: New facility, algorithms and techniques

Mikhail Krinitskiy

Articles you may be interested in

Cirrus heterogeneity effects on cloud optical properties retrieved with an optimal estimation method from MODIS VIS to TIR channels
AIP Conference Proceedings **1810**, 040002040002 (2017); 10.1063/1.4975504

Remote sensing of aerosol optical properties and solar heating rate by the combination of sky radiometer and lidar measurements
AIP Conference Proceedings **1810**, 060002060002 (2017); 10.1063/1.4975518

Seasonal clear-sky flux and cloud radiative effect anomalies in the Arctic atmospheric column associated with the Arctic Oscillation and Arctic Dipole
AIP Conference Proceedings **1810**, 070003070003 (2017); 10.1063/1.4975524

Does a relationship between Arctic low clouds and sea ice matter?
AIP Conference Proceedings **1810**, 070009070009 (2017); 10.1063/1.4975530

Verification of the ISO calibration method for field pyranometers under tropical sky conditions
AIP Conference Proceedings **1810**, 080004080004 (2017); 10.1063/1.4975535

Measuring fast optical depth variations in cloud edges with a CCD-array spectrometer
AIP Conference Proceedings **1810**, 080003080003 (2017); 10.1063/1.4975534

Cloud Cover Estimation Optical Package: New Facility, Algorithms And Techniques

Mikhail Krinitskiy[1, a)]

[1] *P.P.Shirshov Institute of Oceanology of the Russian Academy of Sciences*
36, Nakhimovski prospect, Moscow, Russia, 117997

[a)] Corresponding author: krinitsky@sail.msk.ru

Abstract. Short- and long-wave radiation is an important component of surface heat budget over sea and land. For estimating them accurate observations of the cloud cover are needed. While massively observed visually, for building accurate parameterizations cloud cover needs also to be quantified using precise instrumental measurements. Major disadvantages of the most of existing cloud-cameras are associated with their complicated design and inaccuracy of post-processing algorithms which typically result in the uncertainties of 20% to 30% in the camera-based estimates of cloud cover. The accuracy of these types of algorithm in terms of true scoring compared to human-observed values is typically less than 10%.
We developed new generation package for cloud cover estimating, which provides much more accurate results and also allows for measuring additional characteristics. New algorithm, namely SAIL GrIx, based on routine approach, also developed for this package. It uses the synthetic controlling index ("grayness rate index") which allows to suppress the background sunburn effect. This makes it possible to increase the reliability of the detection of the optically thin clouds. The accuracy of this algorithm in terms of true scoring became 30%.
One more approach, namely SAIL GrIx ML, we have used to increase the cloud cover estimating accuracy is the algorithm that uses machine learning technique along with some other signal processing techniques. Sun disk condition appears to be a strong feature in this kind of models. Artificial Neural Networks type of model demonstrates the best quality. This model accuracy in terms of true scoring increases up to 95,5%.
Application of a new algorithm lets us to modify the design of the optical sensing package and to avoid the use of the solar trackers. This made the design of the cloud camera much more compact. New cloud-camera has already been tested in several missions across Atlantic and Indian oceans on board of IORAS research vessels.

INTRODUCTION

Total cloud cover (TCC) is a key parameter in shortwave radiation parameterization [2]. In sea- and ground-based observations, TCC is commonly estimated visually following the technique [1], which results in large uncertainties due to observer subjectivity. This prevents the use of these observations for both shortwave radiation estimation and validation of cloud parameters in reanalysis [3]. To automate observations and increase their accuracy, state-of-the-art estimation techniques should be used, which are based on, e.g., data from wide-angle optical images of the visible sky hemisphere. Common sequentially produced packages for TCC estimation can be of two types: packages with spherical mirrors equipped with standard angle digital cameras and packages with fisheye lenses with viewing angles from 150° to 180° in the vertical plane and 360° in the horizontal plane. The peculiarities of digital shooting and algorithms used force one to use a sun-shading device (sun tracker) in the packages. Different techniques for TCC estimation are similar except for the way of calculating the control index. The algorithms described in [5, 8] calculate the index by the equation: $R_i = \frac{R}{B}$ (where R and B are the red and blue components of pixel color in the RGB model). Useful pixels of an image are classified by discrimination with an empirically selected threshold value of this index $R_{i_{thresh}} = 0.8$. The algorithm [10] uses the so-called *SkyIndex*, calculated as $SI = \frac{B-R}{B+R}$. The threshold for classifying

useful image pixels as CLEAR SKY and that CLOUD has been selected empirically as well: $SI_{thresh} = 0.23$. Again, TCC is defined as:

$$TCC = 10 * \text{int}\left[\frac{N_{cloud}}{N_{total}}\right] \tag{1}$$

where N_{cloud} is the number of pixels classified as CLOUD and N_{total} is the total number of useful image pixels.

More complex schemes similar to SkyIndex have been published; however, their use allows just an insignificant increase in TCC estimation accuracy. According to our studies, though the accuracy of these algorithms is acceptable in some special cases, in general they result in a statistically biased distribution as compared to human-observed values (Fig.1(a)).
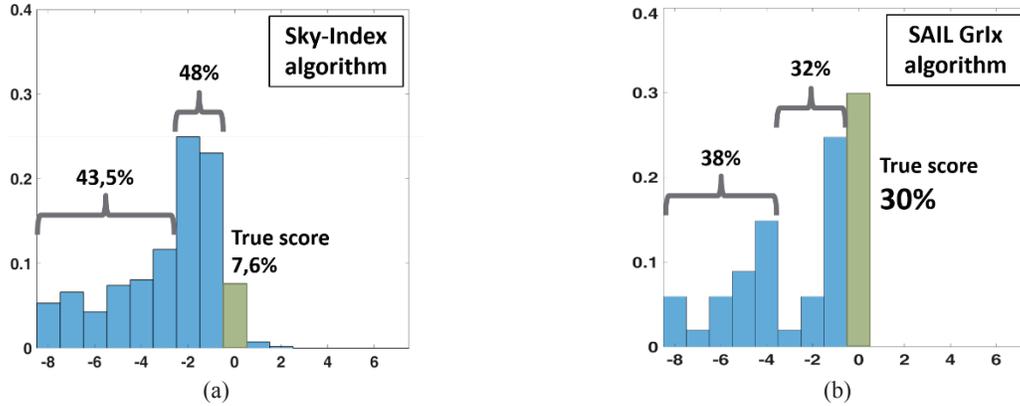

FIGURE 1. Frequency of deviations of TCC estimates compared to human-observed values. With the use of (a) SkyIndex algorithm and (b) SAIL GrIx algorithm. Based on field data collected in cruise of R/V "Akademik Nikolaj Strakhov" over Indian ocean, Mediterranean sea and Atlantic ocean (12/15/2015 - 01/21/2016).

SAIL GRIX TCC ESTIMATION ALGORITHM

SAIL GrIx algorithm basics

To exclude the disadvantages of routine algorithms, we have developed a new technique for TCC estimation based on a new synthetic index, described in details in [6], which we call the "grayness rate index" (**GrIx**). This index meaning is opposite to saturation in the HSV color model [4] and can be calculated for each pixel of the image using the formula:

$$GrIx = 1 - \frac{StdDev(R, G, B)}{Y} \tag{2}$$

Comparison of the indices R_i, **SI**, and **GrIx** has shown that the last index is the most sensitive to saturation gradients [6]. Thus, this index allows a significant increase in the reliability of discrimination of thin clouds.

The step allowed to estimate and, if necessary, filter the effects of sunlight scattering in the atmosphere in the part described by the Mie scattering is an individual branch of SAIL GrIx algorithm and has been described in details in [6]. With the use of this sunburn background filtering the SAIL GrIx algorithm demonstrates higher accuracy compared to routine TCC estimation algorithms. The quality of algorithm in terms of true score increased up to 30% when compared to concurrently human-observed TCC values (Fig.1(b)).

Sun disk condition estimation

Additional branch of SAIL GrIx algorithm allows to estimate the need of sunburn background suppressing. The Sun disk condition (SDC) classification is used as shown at Fig.2. In case of Sun[2] and Sun[1] the sunburn suppression scheme has to be applied. We have developed the machine learning approach to estimate SDC.
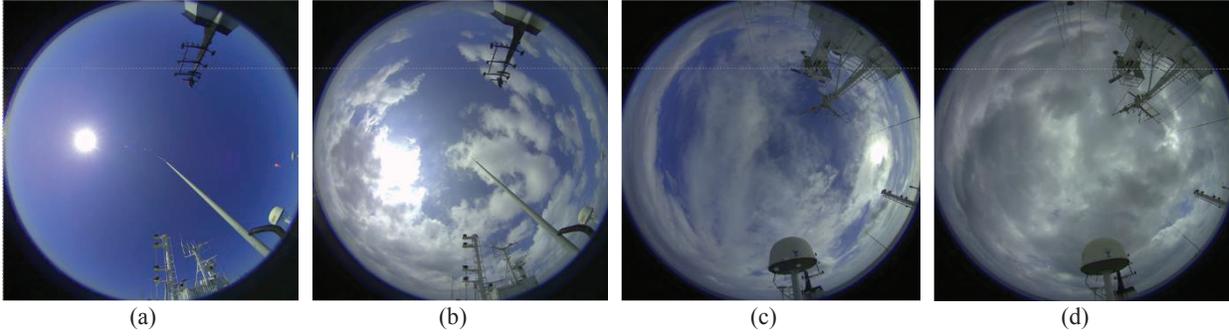
FIGURE 2. Sun disk condition classification: (a) Sun^2 (b) Sun^1 (c) Sun^0 (d) Cloudy

Among a few most widely used state-of-the-art machine learning methods the Artificial Neural Networks (ANN) demonstrated the best accuracy being applied to this problem. Using the multilayer ANN design we reached the 96,4% quality in terms of true score (Fig.3(a)). With this SDC prediction the SAIL GrIx algorithm accuracy increased up to 38% (Fig.3(b)). The effect of suppressing of the hard lower tail of the TCC deviations distribution is also achieved (Fig.3(b)).
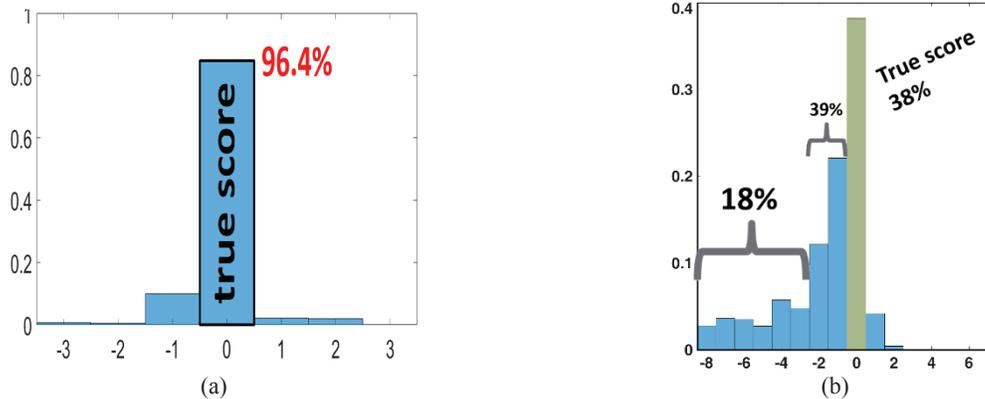

FIGURE 3. (a) SDC prediction deviations frequency with the encoding of the SDC degree: Sun^2=2; Sun^1=1; Sun^0=0; Cloudy=-1. (b) TCC estimates deviations frequency of SAIL GrIx algorithm with the use of SDC predicted values.

SAIL GrIx ML total cloud cover estimation algorithm

While the lower tail of the TCC errors distribution of SAIL GrIx algorithm remains hard (Fig.3(b)), there is machine learning (ML) approach that we have applied to the problem of TCC estimation directly. As well as in case of SDC prediction the following features set was used: ***R,G,B,Y*** and ***GrIx*** fields statistics (min, max, mean, variance and standard deviation, central moments, RMS and percentiles from 5 to 95 with the step of 5), and also SDC classes prediction probabilities and Sun elevation. These features are calculated over each sky image individually. Several state-of-the-art ML models were used and multilayer ANN demonstrated the best accuracy. ANN training was performed using the multi-start approach with randomly generated training subsets, with cross-validation accuracy control. The resulting best model configuration was evaluated using the hold-out subset of 20% of the whole dataset. Total objects count of the dataset is 12800 samples. The dataset was collected in the cruise of R.V. "Akademik Nikolaj Strakhov" over the Indian ocean, Mediterranean sea and Atlantic ocean started on 12/15/2015 from Colombo (Sri Lanka) ended on 01/21/2016 in Kaliningrad (Russia). All samples are supplied with concurrent observer records, for example, with the TCC value, which presents "target value" in terms of ML approach. The scheme of ANN is presented at the Fig.4(a). The following quality has been reached with the ML approach: accuracy = 95.5%; precision mean = 96%; recall mean = 96% (Fig.4(b)).

The ML approach used in SAIL GrIx ML algorithm demonstrates the best quality compared to routine TCC estimation algorithms and to SAIL GrIx algorithm.

Features influence estimation has been also performed using the "optimal brain damage" [7] approach. The most influential features appears to be the following: ***GrIx*** percentiles p5, p10, p15, p20, p25, p45, p50, p55, p75, p80, p85;

SDC prediction probabilities and **R** channel skewness. In case of ANN model type there is no more features with comparable influence among **B,G** or **Y** sky-image fields statistics.
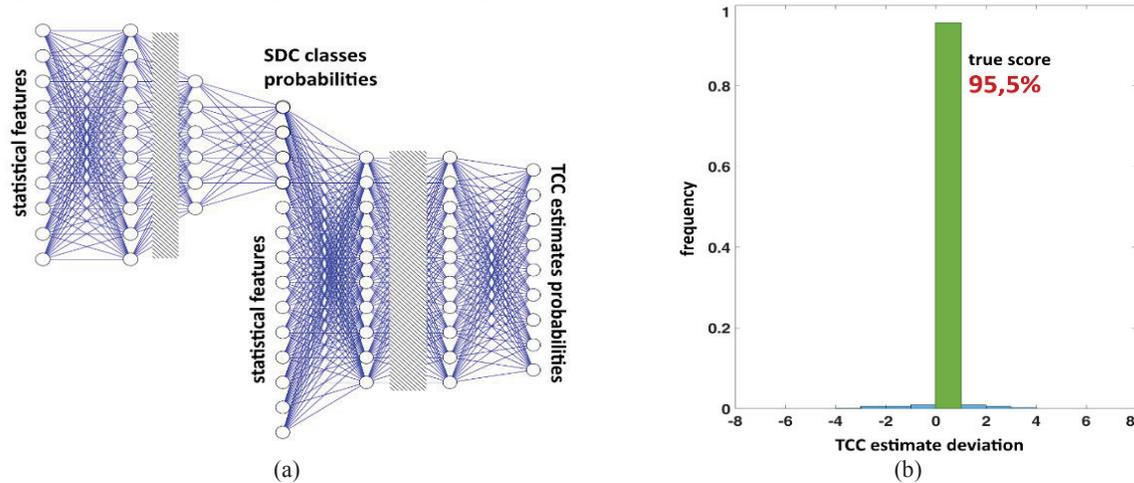

| (a) | (b) |

FIGURE 4. (a) ANN model for TCC estimation scheme. (b) TCC estimates deviations frequency of SAIL GrIx ML algorithm.

Detailed analysis of the ML approach is still has to be performed. The ML approach is the one used to predict the target value "like others close to the object" in the features space. So in terms of TCC prediction problem it predicts the value close to observer evidence. There always remains a question whether the observed values always have to be treated as true ones.

ACKNOWLEDGMENTS

REFERENCES

1. Recommendations for Hydrometeorological Stations and Offices, No. 3, Part 1, Chap. 16: Observation of the Clouds (Gidrometeoizdat, Leningrad, 1985) [in Russian]
2. M. P. Aleksandrova, S. K. Gulev, A. V. Sinitsyn, "An improvement of parametrization of short-wave radiation at the sea surface on the basis of direct measurements in the Atlantic" in *Russ. Meteorol. Hydrol.* 32 (4), p.p.245–251 (2007)
3. E. Bedacht, S. K. Gulev, A. Macke, "Intercomparison of global cloud cover fields over oceans from the VOS observations and NCEP/NCAR reanalysis," in *Int. J. Clim.* 27 (13), p.p.1707–1719 (2007)
4. H. J. George, D. Greenberg, "Color spaces for computer graphics," in Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'78) (Association for Computing Machinery, New York, NY, 1978), p.p.20–25
5. J. Kalisch, A. Macke, "Estimation of the total cloud cover with high temporal resolution and parameterization of short-term fluctuations of sea surface isolation," in *Meteorol. Z.* 17 (5), p.p.603–611 (2008)
6. M.A. Krinitskiy, A.V. Sinitsyn, "Adaptive algorithm for cloud cover estimation from all-sky images over the sea" in *Oceanology* (Engl. Transl.) 56 (3), p.p.315–319 (2016)
7. Yann LeCun, J. S. Denker, S. A. Solla, "Optimal Brain Damage," in Advances in Neural Information Processing Systems 2 (NIPS*89), edited by D. S. Touretzky (Morgan Kaufman, Denver, CO, 1990), p.p.598-605
8. C. N. Long and J. J. Deluisi, "Development of an automated hemispheric sky imager for cloud fraction retrievals," in Proceedings of the 10th Symposium on Meteorological Observations and Instrumentation, January 11–16, 1998 (Phoenix, AZ, 1998), p.p.171–174
9. A. V. Sinitsyn, S. K. Gulev, A. Macke, et al., "2006: MORE cruises launched," in *Flux News* No.1, p.p.11–13 (2006)
10. M. Yamashita, M. Yoshimura, and T. Nakashizuka, "Cloud cover estimation using multi-temporal hemisphere imageries," in Int. Arch. Photogr. Remote Sens. Spatial Inform. Sci. 35 (7), p.p.826–829 (2004)