Detection and identification of unresolved clouds in the measurements from the Infrared Atmospheric Sounding Interferometer (IASI)

A.Rublev[a,*], A.Uspensky[b], T.Udalova[a]

[a] Russian Research Center KURCHATOV INSTITUTE, Moscow 123182, Russia – (rublev,udalova @imp.kiae.ru)
[b] Scientific Research Center PLANETA, Moscow 123242, Russia – uspensky@planeta.litpp.ru

Abstract -The present cloud detection (CD) and cloud identification (CI) procedures have been developed for processing measurements of the new Infrared Atmospheric Sounding Interferometer (IASI). The CD/CI uses the computational modeling of the global sample of synthetic IASI data in dedicated channels calculated for diverse atmospheric models coupled with consistent cloudy scenes. In this line the auxiliary database named Cloud Data Set (CDS) has been generated using well-known NOAA 88/89 dataset of radiosonde measurements.

Testing of the developed procedures as well as verification experiments to estimate achievable cloud discrimination accuracies have been performed using synthetic IASI measurements in dedicated channels subset extracted from season - and region-cloud- labeled samples of CDS. Additionally, the experimental validation of developed CD/CI schemes has been carried out using National Aircraft Sounding Testbed-Infrared (NAST-I) measurements available as a result of the CAMEX-3 campaign performance on Bahamas during September 13, 1998 (http://ghrc.nsstc.nasa.gov/camex3/)

Keywords: IASI, NAST, clouds detection, Monte Carlo method, radiosonde observations

1. INTRODUCTION

The basic aim of realizing investigations is the development of methods and algorithms of cloudiness parameters identification, particularly, the cloud detection on IASI measurements without attraction of additional information from other satellite devices. Previously, it's possible to relate the common amount, average optical thickness, and altitude of top level of cloudiness to parameters that can be estimated from IASI measurements.

Stochastic behavior of optical-geometric cloudiness parameters supposes the probabilistic (statistical) approach to evaluation of algorithms quality. At the same time, it's necessary to provide sufficiently presentable statistic of IASI signals for the adjusted value of evaluating cloudy parameters in different atmospheric cases. At low statistic it's impossible to detect the kind of probability law, its characteristics, and as a consequence - the precision and reliability of CD/CI algorithm. On the other hand, it's necessary to provide the determinacy of situations, i.e. to know the pure values of all the basic factors determining the radiation transfer of infrared emission at every concrete case. Otherwise, the unknown potent factors will mask a bond between the evaluating cloudy parameter and IASI signals.

It's possible to perform these both conditions by getting the IASI signals statistic using the IASI mathematical modeling for different concrete cloudy cases. For statistic description we suggest to use the range of fully determined cloudy models obtained on the basis of real radiosonde measurements of atmospheric parameters. It's obvious that the practicability of such approach owing to necessity to perform several thousands calculations depends of universality and speed of the Radiative Transfer Model (RTM) that is intended to simulate IASI signals under different atmospheric conditions including broken and multi-layers clouds. At the same time for the same reason it's required to select the most informative for retrieval of cloudy parameters IASI channels and limit their amount by several tens. Bigger amount will have made the suggested statistic approach practically unrealizable. In that way, the development of methods for cloudiness parameters determination (detection) on the basis of mathematical modeling includes five conditionally interdependent tasks:

1) mathematical formalization of likelihood approach;
2) development of universal high-speed RTM;
3) compilation of the special subset of IASI channels that are acceptable for cloud parameters retrieval;
4) selection of indicators whose quantitative values would be bound with cloudy parameters and development of algorithms of their identifications from IASI measurements;
5) evaluation of precision and reliability of determination of cloudy parameters in different geographical regions.

The validation of developed CD/CI procedures has been carried out involving experimental high resolution IR radiance spectra measured by airborne interferometer NAST-I. The research objectives of validation tests include quantifying the capabilities and limitations of proposed CD/CI techniques.

2. METHODICAL APPROACHES OF CD/CI

2.1. Bayesian approach overview

According to our approach that has been developed in recent years, it was reasonable to realize CD/CI using the maximum likelihood method to estimate total cloud amount, and then using this estimate to derive parameters of cloud structure for the IASI pixel. Given the probability distributions for a priori specified m cloud classes ω, the most likely i-th class ω_i can be determined by means of the Bayesian classifier (decision rule). Its application minimizes the classification error (Fukunaga, 1972) and includes several stages.

At the first stage, six classes of scenes in the IASI IFOV have been separated using relevant values of fractional cloud cover α, namely:
ω_1 : $\alpha = 0$ – clear; ω_2 : $0 < \alpha < 0.25$ – light clouds; ω_3 : $0.25 < \alpha < 0.15$ – broken clouds; ω_4 : $0.65 < \alpha < 0.95$ – heavy clouds; ω_5 : $\alpha > 0.95$ – overcast; ω_6: $\alpha=1$ –thin cirrus layer.

In the measurement space, the $(n\times1)$ feature vector $\mathbf{y} = || y_1, \ldots y_n ||^T$ consisted of IASI radiances in *n* channels has been identified. Let us assume that the sample $\{\mathbf{y}_i\}$ of IASI data associated with i-th class ω_i is described by Probability Density Function (PDF) $f(\mathbf{y}|\omega_i)$ of known form. Also let $P(\omega_i)$ to be the *a priori* probability of occurrence for the i-th class. Then according to Bayesian formula, the *a posteriori* probability $f(\omega_i \mid \mathbf{y})$ of the membership of the observed vector **y** in the ω_i is estimated by the following expression

$$f(\omega_i \mid \mathbf{y}) = \frac{P(\omega_i) f(y|\omega_i)}{\Sigma P(\omega_j) f(y|\omega_j)}. \qquad (1)$$

* Corresponding author
** This work was funded by Eumetsat (contract EUM/CO/01/938/PS)

The expected class membership for particular IASI data array $\mathbf{y}_i$ can be estimated using classifier or Bayesian Discriminate Function BDF (see, e.g., Fukunaga, 1972):

$\mathbf{y} \in \omega_k$, where the class number k is specified according to $f(\omega_k | \mathbf{y}_j) = f(\omega_i | \mathbf{y}_j)$. (2)

The explicit calculations of BDF are possible if we know the form of underlying distribution $f(\omega_i | \mathbf{y}_j)$. The analysis of the IASI data samples for all ω_i classes under consideration (i = 1, …,m) demonstrates that relevant distributions can be approximated (with satisfactory level of accuracy) by multi-variate Gaussian, i.e.

$$f(\mathbf{y} | \omega_i) = (2\pi)^{-n/2} (\det \mathbf{S}_i)^{-1/2} \exp\{-\frac{1}{2} d^2(\mathbf{y})\}, \quad (3)$$

where n is a total number of the IASI channels selected for CI; m is the total number of ω_i classes,

$d^2(\mathbf{y}) = (\mathbf{y} - \overline{\mathbf{y}}_i)^T S_i^{-1} (\mathbf{y} - \overline{\mathbf{y}}_i)$; and $\overline{\mathbf{y}}_i$, S_i are ω_i class mean and covariance (n×n) matrix.

In our case, because of the lack of any *a priori* information about the cloud classes, it is reasonable to define $P(\omega_i)$ in (1) as $P(\omega_i) = m^{-1}$.

Hence solved in order to implement the Bayesian CD procedure the rest points listed above have to be performed.

2.2. 3D Monte-Carlo algorithm

The specially developed for CD/CI aims RTM is based on the simulation of photon trajectories by a Monte Carlo method (MC). The algorithm is intended for calculation of apodized values of the intensity of outgoing radiation at IASI channels. The trajectory of single photon is considered as sequence of collisions with particles of atmosphere and underlying surface. At every collision, the photon can change its movement direction or be absorbed by atmospheric gases or aerosol as well as by the surface.

The conjugated trajectory method that uses the optical reciprocity theorem was applied. In this case the radiative transfer simulations are made as if the source (the atmosphere or the surface) and the radiation receiver (light-receiving IASI device) trade places, i.e. each photon trajectory is simulated from the light-receiving IASI device till photon absorption in the atmosphere or at the surface. The energy contribution of a photon in estimated radiance is determined by Plank function at given temperature in the point of photon absorption. Gaseous absorption coefficients at different altitudes are calculated using line-by-line method for given IASI band with spectral resolution 0.001 cm^{-1}. The apodization is made using random choice of photon frequency by the normal law. The average of its distribution coincides with mean value for the channel and dispersion depends on the apodization function. Thus, the integration on space is made simultaneously with the integration on spectra. For cloudy conditions, the improvement in time of the calculations with the same precise achieved 50÷60 times against "standard" method that uses transmission function and effective absorption masses to take into account gaseous absorption along photon trajectory.

So far the 3D MC code allows carrying out calculations in presence of up to two broken cloud layers and single cirrus one. The model of broken cloud layer is specified by Gaussian random field. The field is horizontally isotropic and limited from the bottom at definite atmosphere level. Such model was developed by (Rublev, and Golomolzin, 1992) and provides reasonable agreement between simulated and real observations (Geogdzhaev et al, 1997). The main input parameter to the model is a cloud amount. Other geometrical parameters, namely mean cloud diameter and thickness are derived from statistical data. The cloud optical parameters include mean extinction coefficient, scattering phase function and single scattering albedo that are evaluated using known models of the cloud particle size distribution.

Optical models of cirrus clouds are based on four crystal size distributions as dependent on the cloud temperature. Ice crystals are represented in the form of hexagonal prisms with given relationships between bottom diameter and length and between density of ice particles and prism length. The calculations of basic optical characteristics are based directly on the Mie theory using different effective radius and give good fit the results obtained by accurate method (Fu et al., 1998).

The MC algorithm has been applied for simulation of the IASI measurements in conditions of diverse cloudy scenes. The calculated spectra have been compared with those calculated on the basis of "standard" LITMS_FRTM version 2.0 (Trotsenko et al., 2000) under horizontally inhomogeneous conditions. We find generally good agreement between these models. As a result, the MC based algorithm can be reputed as the appropriate radiative transfer model capable of sufficiently fast and reasonably accurate simulation of the IASI level 1c radiances correspondent to diverse cloud conditions (including broken clouds).

2.3. Selection of channel set in IASI spectral bands

For selection of the channel subsets suitable for CD and CI the IASI channels with wavenumbers less than 1300 cm^{-1} have been considered. This was done in order to totally exclude the impact of the reflected solar radiation. Basically, the spectral domain ranging from 1000 to 1200 cm^{-1} has been analyzed. Within the above spectral range two multi-factorial computational experiments have been conducted involving both the stratus and broken cloud conditions. For all IASI channels, the brightness temperature y_i have been calculated using standard atmospheric models and the surface cover emitted as blackbody. As a result a standard linear regression model in the following statement has been developed: $y_i = b0 + b_1 x_1 + b_2 x_2 + b_3 x_3$, where b_n (n= 1,2,3) correspondingly denote the regression coefficients for the type of atmospheric model, cloud amount or optical thickness and clouds altitude

From the analysis (Uspensky et al., 2001) of b_2&b_3 magnitudes, the subset of 69 IASI channels has been selected as the most perspective for the CD/CI procedure.

2.4. Generation of the database

The development of efficient CD/CI procedures should be based on the representative database which incorporates a global sample of diverse atmospheric models consistent with the cloud scenes. Such database was generated using the NOAA 88/89 radiosonde dataset and the approach reported in (Chernykh et al., 1996).

Table 1 illustrates how representative the database is regarding the availability of data correspondent to the mid-latitude (ML), high-latitude (SA) and tropical (TR) atmosphere conditions, over the land (L) or sea (S). The following *a priori* hypotheses has been adopted:

- the clouds can exist at three levels, namely low; medium, and high;
- cloud amount is derived using well-known bimodal distribution consistent with spatial resolution of IASI measurements (for the area of about 50×50 km^2 formed by 4 (2×2) IASI pixels);

- typical values of cloud parameters (such as CWL, etc.) have been extracted from (Mazin and Khargian, 1989).

Table 1. Number of soundings in the radiosonde dataset

Zone	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
ML/L	760	177	123	89	227	195	262	206	43	200	186	233
ML/S	206	53	41	36	90	93	82	80	10	66	79	74
SA/L	572	102	84	68	214	158	174	140	30	107	162	107
SA/S	49	17	14	11	35	36	25	37	9	10	12	32
TR/L	544	132	121	54	196	164	218	190	46	101	124	140
TR/S	134	31	27	21	53	60	52	44	7	18	54	68

After some refinement and tuning the cloud parameters derivation technique the statistical verification of updated cloud data set has been performed using available satellite-based experimental datasets (TOVS, GOES, ISCCP). Table 2 contains cloud amount statistics extracted from the listed data sources as well as respective statistics for CDS. The first row in the table is taken from (Mazin and Khargian, 1989).

Table. 2. Intercomparison of the cloud amount statistics averaged over the year

Data source	Europe		S-E Asia	North America	South America	Austr
	East	West				
Handbook	4.5	-	-	4.8	5.4	3.6
TOVS 1988	3.8	4.4	4.2	3.6	4.4	2.7
GOES-1 1988	6.4	4.6	5.2	5.0	6.4	3.8
ISCCP 1988	5.3	6.0	5.8	5.5	6.2	4.4
ISCCP 1989	6.6	6.2	6.6	6.4	6.7	5.6
Mean	5.3	5.2	5.4	5.1	5.8	4.2
IASI CDS	5.4	4.8	5.0	4.7	4.6	4.3

As it can be seen from comparison of the last two rows (table 2) the values of mean cloud amount (averaged for year) are in a good agreement.

2.5. Testing of the CD scheme

Testing of the developed CD procedure has been performed in the same manner, as in (Uspensky et al., 2001). General restrictions to carry out verification experiments (established empirically through performing a series of tests) are as follows: the sample size of each class must be much more greater than the number of the used channels (in fact, the maximum number of used channels was not above 15 for all the zones from Table 1).

The accuracy or skill of the CD/CI scheme performance is generally described by the contingency or the chance tables. Cloud amount chance tables have been established for different cloud samples. As example, the table 3 is concerned to tropic conditions sample. An entry n_{ij} (i, j=1,…,6) in both table panels (frames) represent the probability (in per cent) that the i-th predicted class falls in the j-th true class. In other words, the accuracy of CD scheme is characterized by matrices composed from probabilities of the correct classification and errors of the first and second kind. With regard to this, the diagonal elements are the probabilities of correct classification, while the others are the probabilities of the errors.

Table 3. BDF-based chance table for the tropics

Cloud class	Simulated (true)					
Predicted	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
ϖ_1	**88**	3	5	4	1	1
ϖ_2	4	**66**	27	17	6	9
ϖ_3	2	12	**43**	19	3	2
ϖ_4	3	9	15	**49**	4	4
ϖ_5	1	6	6	7	**67**	18
ϖ_6	0	4	4	4	17	**66**

2.6. Derivation of clouds parameters

The developed software is based on the CI scheme involving the evaluation cloud layers number, Cloud Top Height (CTH), and mean cloud thickness τ. The proposed approach is based on successive application of Bayesian classifier and least-squares fitting. To account for the needs of CTH estimating the original feature vector y (IASI measurements in dedicated 69-channel subset) has been complemented by ancillary information, i.e. a background temperature profile T(z) at sounding location. The required profiles T(z) should be assigned at several levels with 1-km vertical sampling (from ground level up to 8-10 km) and accuracies better than 2.0K. The AMSU/Metop-based T-retrievals as well as NWP output products can be considered as source candidates for supplying this information.

After performing the CD procedure, that distinguishes particular ω_i class and specifies (α), the well-known least-squares fitting technique is applied. The cost function, which measures how much extended feature vector differs from the tested model, could be the sum of squared residuals

$$\chi_i^2 = \chi_{TP}^2 + \chi_{BT}^2 .$$

Here index i relates to individual model realization (within preselected ω_s-class),

$$\chi_{TP}^2 = \frac{1}{N}\sum_{n=1}^{N}\frac{(T_n - T_n^m)^2}{\sigma_{Tn}^2} ,$$

where N is a total number of levels z_i, T(z) is assigned at N atmosphere levels (N = 2-10), σ_{Tn}^2 is the temperature variance at level z_n (also within ω_s class); and

$$\chi_{BT}^2 = \frac{1}{L}\sum_{l=1}^{L}\frac{(BT_l - BT_l^m)^2}{\sigma_{BTl}^2} ,$$

where L is a total number of channels in the dedicated subset (dimension of vector y), BT_l and BT_l^m are measured and simulated Brightness Temperatures (BT) in l-th channel respectively, σ_{BTl}^2 is a BT variance in l-th channel calculated for ω_s class.

The sought cloud parameters can be derived via minimization of the cost function χ_i^2. The number k of the suitable ("true") model is the solution of the following minimization problem

$$\chi_k^2 = \min_i(\chi_i^2) .$$

To solve this problem, the straightforward method, namely, the direct sorting algorithm is applied.

In order to clarify the developed CI technique, it should be reminded, that the identification of k-th model enables to specify cloud parameters of interest, i.e. the number of cloud layers, together with the CTH and τ estimates. The point is that within RTM MC calculations and establishment of training samples, the clouds with concrete properties and corresponding atmosphere models are inserted to simulate measured radiances. Thus, while minimizing the cost function, we can determine the most likely cloud parameter values for real radiances measured by IASI *in situ*. As it follows from sensitivity studies (regarding CTH, τ estimation) the developed CI method is the most efficient for one layer of thick clouds. The accuracy of identification degrades significantly for scenes with semi-transparent and multi-layers clouds.

3. Validation results against real measurements

Additionally, the experimental validation of developed CD/CI schemes has been carried out on the NAST-I measurements available as a result of the CAMEX-3 campaign performance. According to CAMEX-3 Programme schedule the day September, 13 was dedicated to Cal/Val activities (regarding all instruments used).

Testing of developed CD/CI techniques as well as validation of output products has been performed via intercomparison of NAST-I- based cloud parameter estimates and “true” ones that were obtained using MAMS (Multispectral Atmospheric Mapping Sensor) images and radiosonde data.

The point is that the NAST-I spectral and performance characteristics are rather close to IASI. The airborne MAMS imager is a multispectral scanner which measures reflected radiation from the Earth's surface and clouds in 8 visible/near infrared channels and thermal emission from the surface, clouds, and atmospheric constituents (primarily water vapor) in four infrared bands. The scanner views a 37 kilometer wide scene of the Earth from the ER2 aircraft altitude of about 20 kilometers. Each MAMS footprint (individual field of view-IFOV) has a horizontal resolution of 100 meters at nadir.

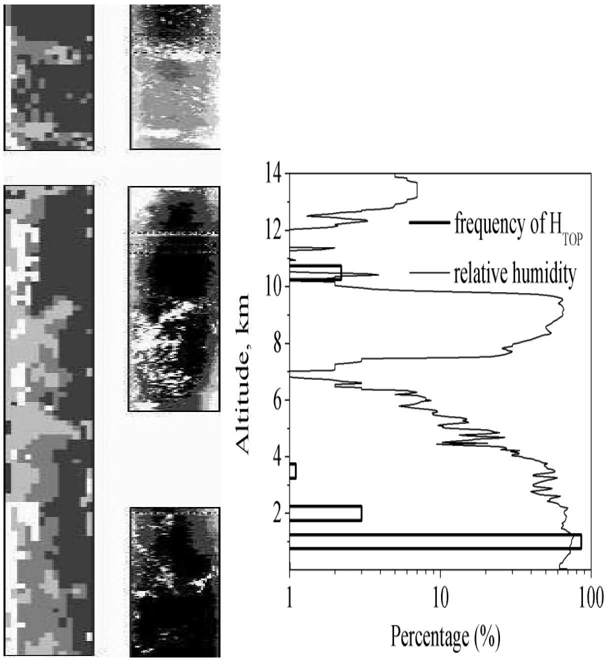
The high spatial resolution of MAMS is of great importance for CD validation, since it enables to perform cloud detection over the entire image (pixel-by-pixel) with good efficiency. As follows from the visual analysis of MAMS imagery with averaging over NAST-I pixel (2.5 x 2.5 km^2) the predominant clouds within cloud scene are Cumulus with mean cloud fraction $n_0 \approx 0.1 \div 0.2$.

Verification of CD procedure has been carried out using synchronous and merged NAST-I data and MAMS imagery. The area of interest or cloud scene covers the region about 385 x 37 km^2 and incorporates 2002 NAST-I pixels. The fragment of the cloud images from retrieved from NAST-I results and MAMS observations are demonstrated at Figure 1 (a).

Five gradations from black to white correspond to five magnitudes of cloud fraction from n_0=0 (clear sky) to n_0=1.0 (overcast). Visual comparison of both figure sides demonstrates efficiency of the developed CD technique that reproduces the spatial structure of cloud scene rather well.

In the main, the output products of the procedure reproduce correctly spatial structure of cloudy scenes).

As an important part of validation studies, the verification of CI procedure with respect to CTH evaluation has been performed. In order to derive “true” CTH values the radiosonde relative humidity profiles have been visually analyzed. As it is clearly seen from Fig1 (b), within the observed region, the cumulus (in the layer about 1÷2 km) as well as cirrus cloud layer at 8-12 km height are distinguished. The CTH estimates based on NAST-I spectra are similar. They have detected cumulus clouds with top heights closely to 1 km. The presence of cirrus clouds may be also detected on the Fig.1 (b) from NAST-I data within 10 ÷ 11 km.



a) b)

Figure 1. Comparison of cloud parameters retrievals

a) cloud fraction from NAST-I spectra (left) against MAMS observations (right)
b) frequency distribution of cloud top height H_{TOP} against vertical profile of relative humidity

In that way, the validation exercise shows perceptivity of the developed approaches for retrieval of cloud parameters from spectral measurements.

Acknowledgments. We would like to thank the members of the I3RC and IASI science working groups for useful discussions.

REFERENCES

Chernych I.V., R.E.Eskridge. 1996: Determination of cloud amounts and level from radiosonde soundings. J.Appl.Met., 35, 8, pp1362-1369

Fu, Q., et al., 1998: An accurate parametrization of the infrared radiative properties of cirrus clouds for climate models. Journal of Climate, 11, pp 2223-2237

Fukunaga K. 1972: Introduction to statistical pattern recognition. Ac. Press. N.- 4., London

Geodzhaev, I.V, et al., 1997: UV radiation transfer through broken cloud fields. Modeling and comparison with measurements. Izvestiya, Atmospheric and Oceanic Physics, No. 5, pp 630-635

Mazin and Khargian, 1989: Clouds and cloudy atmosphere. Handbook, Gidrometeoizdat, Leningrad, 646 p. (in Russian)

Rublev, A.N, and V.V. Golomolzin, 1992: Modeling of cumulus clouds. Preprint of RRC Kurchatov Institute, KIAE-5567/16, 12p. (in Russian)

Trotsenko, A., et al., 2000: LITMS-FRTM: Fast radiative transfer model to simulate IASI/MetOp measurements. In Proc. of the 2000 Eumetsat Meteor. Satellite Data Users’ Conference, Bologna, Italy, 29 May-2 June, 2000, EUM, pp 231-238

Uspensky et al, 2001: A Bayesian cloud detection scheme using high spectral resolution IR sounder observations And Experimental Cloud Data Set: A Simulation Study. The 2001 Eumetsat meteorological satellite data user's conf. Antalya. EUM-B33. 638-645