Cloud detection using SEVIRI IR channels for the GERB processing

A. Ipe, L. Gonzalez Sotelino, E. Baudrez, N. Clerbaux, I. Decoster, S. Dewitte, S. Nevens, A. Velazquez Blazquez

Royal Meteorological Institute of Belgium, Department of Observations, Section Remote Sensing from Space, Avenue Circulaire 3, B-1180 Brussels, Belgium.

gerb@oma.be – http://gerb.oma.be

1 Motivations

The Geostationary Earth Radiation Budget (GERB) processing currently relies on a scene identification applied to the Spinning Enhanced Visible and InfraRed Imagers (SEVIRI) for the estimation of the top–of–the–atmosphere (TOA) solar fluxes on a near realtime basis from the GERB broadband radiometers. More specifically, this scheme delivers a cloud mask based on the visible SEVIRI channels to properly select the angular dependency models (ADMs) for the radiance–to–flux conversion. It results that such product is unavailable during nighttime. Therefore, we have decided to develop a cloud detection algorithm solely based on thermal SEVIRI measurements. Its associated thermal cloud mask will be available under all observing conditions including sunglint and will allow to perform studies on clouds and aerosols radiative forcing over any time period.

2 Strategy

Major cloud detection algorithms based on the multispectral threshold technique in the thermal wavelengths rely on ancillary data. For example, the Nowcasting Satellite Application Facility (NWCSAF) [1] as well as the EUMETSAT CLoud Mask (CLM) [3] need numerical weather prediction (NWP) model fields to correctly discriminate clearsky from cloudy scenes. These fields include surface temperatures, air temperatures at various altitudes as well as the total water vapor content of the atmosphere. While these data are not mandatory for the NWCSAF software, its accuracy is significantly decreased when there are missing since climatologies are then used.

Such multispectral threshold approach for the GERB processing is discarded for several reasons:

1. GERB products would not anymore allow a fully independent validation of GCMs,
2. the latter would rely on an external data source which could be unavailable, thus resulting in delayed processing (near-realtime constraint),
3. the quality of these ancillary data could not be guaranteed over long time period due to NWP software updates, and therefore bias could be introduced in the thermal cloud mask (release of consistent product datasets),
4. the overhead in terms of implementation complexity would be prohibitive.

Instead, we decided to develop a simple but robust method which can be independently applied to the 3 thermal SEVIRI channels: 8.7, 10.8 and 12 μm. The suggested method is considering time–series of 60 days measurements at the same time of day and at the pixel–level. Therefore we are maintaining the native spatial (3 km at nadir) and temporal (15 minutes) samplings of SEVIRI. This selected approach implicitly considers that clearsky conditions will at least occur once over this 60 days time period.

Several factors are known to affect these channels' associated brightness temperatures (BTs) (equivalent black body temperatures): surface emissivity, atmospheric profiles and cloudiness. Changes of surface emissivity due to precipitation or vegetation and atmospheric states result in a significant variability of the BT over cloud–free pixels. Therefore, it is obvious that the sensitivity of the developed cloud detection scheme will be impacted. While thick clouds characterized by low cloud top temperatures can be detected due to the drastic drop of their BT compared to clearsky conditions, thin (cirrus) as well as low (stratocumulus) clouds are not significantly modifying the clearsky BT signal and therefore could be misidentified as clearsky.

3 Algorithm

As mentioned above, we are considering a 60 days time–series of BTs for each pixel and infrared channel at a given time of day. These 60 values can usually be classified into 3 groups with decreasing BT:

- clearsky,
- thin or warm (low) clouds,
- thick cold clouds.

This classification is achieved by means of a modified k–means unsupervised clustering algorithm [2] as described below:

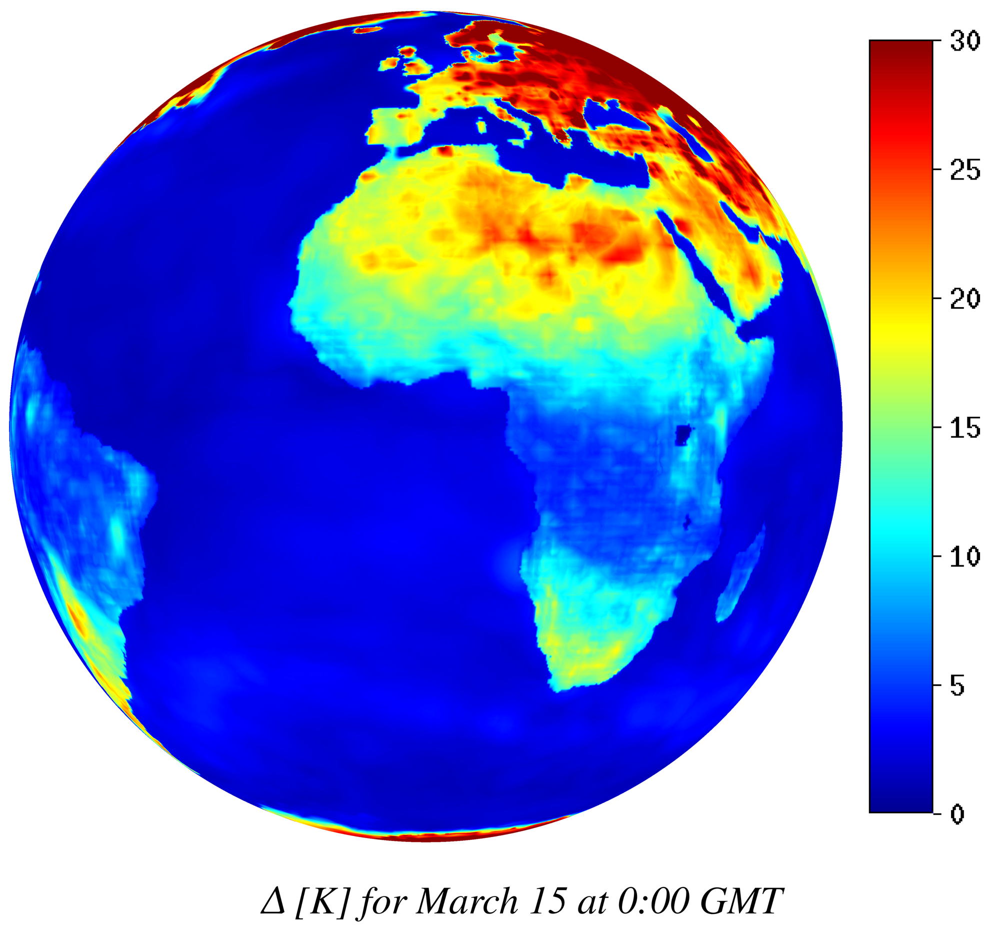
1. Let $i = 0$ and initialize the 3 clusters' centers $C_k^{(i)}$ $(k = 0, 1, 2)$,
2. Classify all 60 BTs according to their "nearest" cluster's center $C_k^{(i)}$,
3. Recompute the 3 clusters' centers $C_k^{(i+1)}$,
4. Let $i = i + 1$ and repeat from step 2 until $C_k^{(i)}$ and $C_k^{(i+1)}$ does not change more than 0.01 K.

The modification lies in the distance d used for the assignment of each sample $\mathrm{BT}^{(n)}$ $(n = 1, \ldots, 60)$ to a specific class C_k. We make the assumption that within each class, all samples are distributed according to a normal distribution $N(\mu_k, \sigma_k)$. Thus, the distance is given by the Bayesian discriminant function

$$d(\mathrm{BT}^{(n)}, C_k) = \frac{(\mathrm{BT}^{(n)} - \mu_k)^2}{2\sigma_k^2} + \log \sigma_k^2$$

if we assume that the prior probability to belong to each class is identical.

However, the k–means clustering method suffers from a major drawback since its results depend on the clusters' initialization $C_k^{(0)}$ [2]. To circumvent such issue, we are attempting to find the best first guess estimate of the clearsky cluster's width Δ allowing to initialize $C_k^{(0)}$. This is achieved by considering 10 years (from 1991 to 2001) of skin surface temperatures T_s from the ECMWF ReAnalysis project (ERA–40) [4] which are available every 6 hours on a 0.25° × 0.25° grid. We are assuming that the surface temperature is the major factor of variation of the BTs measured by the satellite and thus neglecting the contribution of the atmosphere. These data once reprojected into the SEVIRI field–of–view are used to compute for every pixel on a 3–hourly basis t $\Delta(d, t) = T_s^{(\max)} - T_s^{(\min)}$ at a given date d from the previous 60–days T_s time–series. Then, a climatology of Δ for every pixel can be estimated by taking the monthly percentile at 95 % of these 10 years of instantaneous $\Delta(d, t)$ values. The initialization of the clusters $C_k^{(0)}$ then follows:



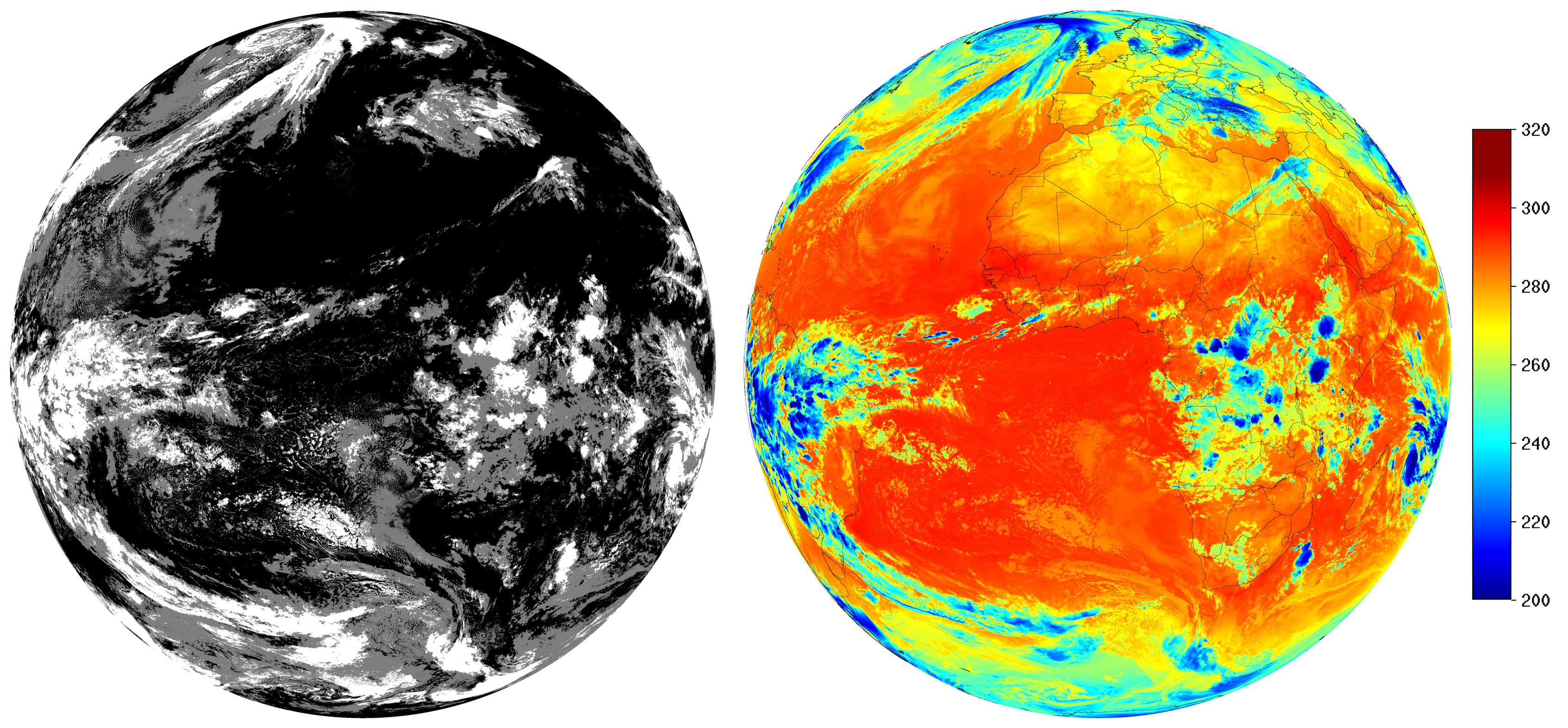
Δ [K] for March 15 at 0:00 GMT

1. For $C_2^{(0)}$ (clearsky): $\mu_2^{(0)} = \mathrm{BT}^{(\max)} - \frac{1}{2}\Delta$ and $\sigma_2^{(0)} = \Delta/3.25$,
2. For $C_1^{(0)}$ (thin or warm clouds): $\mu_1^{(0)} = \mu_2^{(0)} - \Delta$ and $\sigma_2^{(0)} = \Delta/3.25$,
3. For $C_0^{(0)}$ (thick cold clouds): $\mu_0^{(0)} = \mathrm{BT}^{(\min)}$ and $\sigma_2^{(0)} = \Delta/3.25$.

However, if $C_0^{(0)}$ and $C_1^{(0)}$ are not at least separated by Δ, then the clustering is only performed on the 2 upper clusters. If the initialization fails with 2 clusters, then the times–series is assumed to be entirely clearsky.

4 Preliminary results

In the following figures, we have plotted the resulting cloud mask when our method is applied to the SEVIRI 8.7 μm channel. One may note that our modified k–means clustering is successfully detecting low contrasted clouds in terms of brightness temperatures over the Atlantic Ocean. Moreover, convective cloud fields near the Equator are also correctly identified by our scheme. However, further quantitative comparisons with another cloud mask are required.



Cloud mask based on the 8.7 μm band (left) where black is for clearsky, gray is for low contrasted clouds, white is for high contrasted clouds and associated BTs [K] (right) for March 11 2007 at 0:00 GMT

5 Preliminary comparisons

As mentioned in section 2, two cloud masks are routinely derived from SEVIRI imagery. While the product from NWCSAF can be considered as one of the most accurate cloud detection scheme, the EUMETSAT CLM was primarily designed for robustness. Nevertheless, both are using NWP fields in their numerous threshold tests as well as visible and NIR channels during daytime. Thus, in the following, we are considering the NWCSAF cloud mask as the truth and comparing both the CLM and GERB products to it. Since the CLM is not delivering results at viewing zenith angles above 75°, we are restricting GERB comparisons to pixels below this limit.

Depending on the surface type and its associated emissivity, it is obvious that the classification is improved for SEVIRI IR channels exhibiting the highest contrast between clearsky and cloud objects. This is illustrated in the following table.

	ocean	vegetation	desert
8.7 μm	86.12 ± 0.45	87.50 ± 0.90	93.13 ± 1.66
10.8 μm	85.95 ± 0.46	88.47 ± 0.72	94.30 ± 1.23
12.0 μm	84.46 ± 0.53	89.12 ± 0.61	94.88 ± 0.96

Averages of the weighted daily means according to the number of night pixels of the hourly pixels' agreement (in percent) between the NWCSAF CMa and the GERB IR cloud masks for August 11–17 2007 for the 3 IR channels and geotypes. The uncertainties are given at 2σ.

Thus, simply by combining the classification from the SEVIRI 8.7 and 12.0 μm channels respectively for ocean and land surfaces, we can improve the single channel results as demonstrated in the following table.

Cloud mask	Geotype			
	ocean	vegetation	desert	all
MPEF CLM	85.73	88.70	91.19	87.20
GERB IR	86.12	89.12	94.88	88.02
GERB IR+	89.97	90.67	96.02	90.94

Cloud mask	Geotype			
	ocean	vegetation	desert	all
MPEF CLM	84.93	88.97	90.73	86.70
GERB IR	84.51	86.20	94.00	86.17
GERB IR+	89.01	88.33	94.87	89.60

Weighted means according to the number of night (left) and all (right) pixels of the hourly pixels' agreement (in percent) of the MPEF CLM, GERB IR cloud masks with respect to the NWCSAF CMa cloud mask for August 11–17 2007. GERB IR+ designates the GERB IR cloud mask supplemented with the NWCSAF CMa spatial texture filter.

Since no spatial textural filtering is performed within the GERB IR cloud detection algorithm compared to the 2 other schemes, we also have calculated the agreement with an "augmented" GERB IR method by performing the NWCSAF CMa spatial filtering as post–processing (denoted in table by "GERB IR+"). As expected, the "augmented" GERB IR method always performs better than without the spatial filter. It can be noted that the MPEF CLM systematically exhibits a lower agreement than the standard GERB IR cloud detection over all geotypes for nighttime conditions.

Nevertheless, low water clouds (stratocumulus) are usually misidentified due to low BT contrast (≈ 1 K). The use of the 3.7 μm channel through BT differences with the 10.8 and 12.0 μm channels (NWCSAF–like) should be investigated.

References

[1] M. Derrien and H. Le Gléau. MSG/SEVIRI cloud mask and type for SAFNWC. *Int. J. Remote Sensing*, 26 (21):4707–4732, 2005. doi: 10.1080/01431160500166128.

[2] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification — Second Edition*. Wiley Interscience, New York, 2001.

[3] H.J. Lutz. Cloud processing for Meteosat Second Generation. Technical Report EUMETSAT Technical Memorandum No. 4, EUMETSAT, Darmstadt, Germany, 1999.

[4] S.M. Uppala, P.W. Kallberg, A.J. Simmons, U. Andrae, V. Da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J.F. Mahfouf, J.J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen. The ERA–40 re–analysis. *Quart. J. R. Meteorol. Soc.*, 131(612):2961–3012, 2005.

Acknowledgment

This study was supported by EUMETSAT and the Belgian Science Policy Office under the contract EUMETGERB.