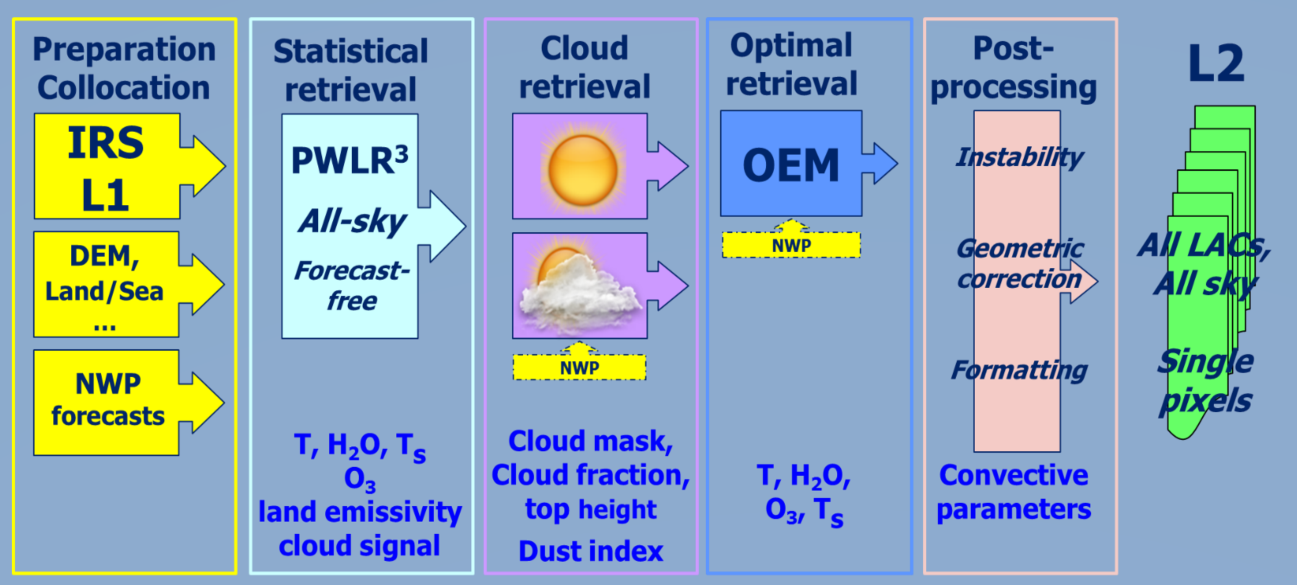
EUMETSAT L2 Processor Prototype for MTG-IRS

C. Goukenleuque[1], T. Hultberg[2], T. August[2], P. Dussarrat[2], O. Samain[3]
[1]Hamtec Consulting Ltd, UK, [2]*EUMETSAT*, [3]*Exostaff GmbH*

Summary

The Meteosat Third Generation (MTG) program is a cooperation between EUMETSAT and the European Space Agency (ESA)in which ESA develops and procures the satellites on behalf of EUMETSAT. EUMETSAT develops and operates the ground segment used to control the satellites, acquires and processes the data, and delivers the products to users worldwide.
MTG is the next generation of the European geostationary Meteosat satellite system for numerical weather prediction and nowcasting. The MTG space segment consists of 4 satellites for the imaging applications (MTG-I) and 2 for the sounding applications (MTG-S). The 6 satellites in the MTG satellite fleet will provide the European national meteorological services and international users with improved imaging and infrared sounding capabilities for meteorological and climate applications in the next 20-year timeframe.
The hyperspectral Infrared Sounder IRS, deployed on MTG-S, is primarily designed to support numerical weather predictions at regional and global scales, including nowcasting. It will provide 4-D hyper-spectral soundings of, inter alia, temperature, water vapor, ozone vertical profiles and trace gases within two spectral bands, 680-1210 and 1600-2250 cm^{-1}, with a spectral sampling of ~0.625 cm^{-1}, and a footprint of 4x4 Km at nadir. The scan of the full Earth's disk can be achieved within an hour with a revisit over Europe every 30 minutes.
The level 2 products will also provide instability parameters. It is foreseen to supplement the products at a later stage with the Atmospheric Motion Vectors derived from the temperature, moisture and ozone vertical profiles.
We present here an overview of the architecture and algorithmic implementation of the L2 MTG-IRS processor prototype. In addition, geophysical parameters retrievals, based on IRS L1b simulated data, are shown in a selection of dwells over Europe.

Overview



The five main steps of the processing chain are:

- Pre-processing (PREP)
  - Collocation of elevation and land/sea mask
  - Collocation of NWP forecast (optional *a priori* information)
- First retrieval, statistical method (PWLR)
  - Fast all-sky statistical retrieval of temperature (T), water vapour (W), ozone (O), emissivity, surface skin temperature (T_s) and cloud signal.
  - Retrieval includes reliable quality indicators (error estimates) for each of the retrieved parameters.
- Cloud retrieval/detection (CLOUD)
  - 1D-var retrieval of effective cloud fraction and cloud top pressure of up to two cloud layers using the simple cloud parametrisation of RTTOV as forward operator and analytic expressions for optimal effective cloud fractions for given cloud top pressures.
  - Cloud phase by simple brightness temperature difference tests
  - Dust index by linear regression
- Second retrieval, optimal estimation (OEM)
  - Clear sky optimal estimation of T, W, O, CO_2 (as optional by-product of T), emissivity.
- Post-processing
  - Reconstruction of the profiles at the vertical of the target point on Earth.
  - Calculation of the instabilities

The PREP, PWLR, CLOUD and OEM are developed in Julia for fast performance (comparable to C) and experimentation flexibility and the POST module is developed in Java.

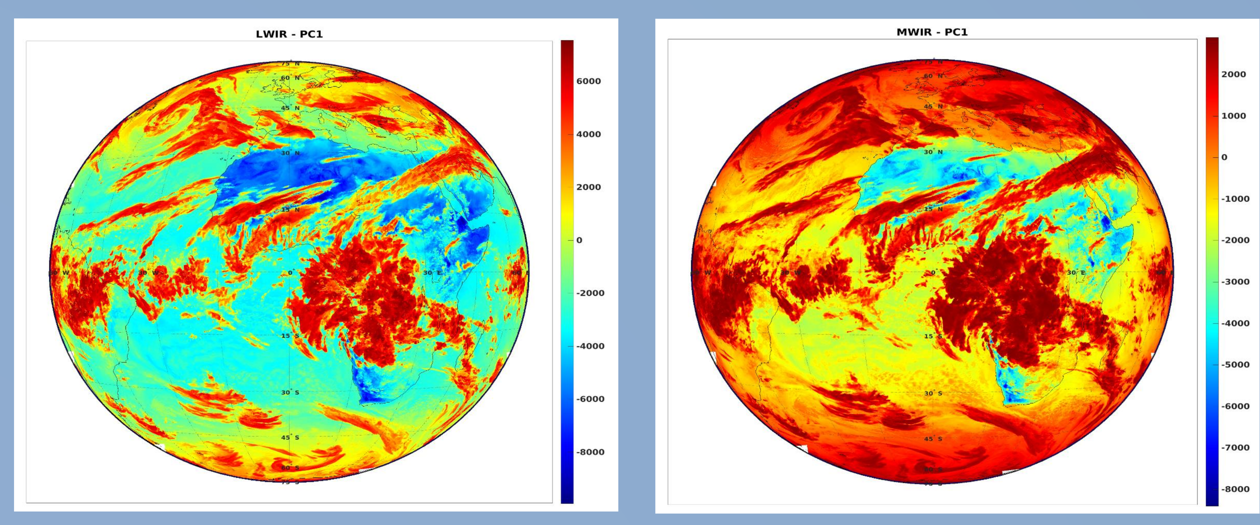
Prototype Testing

In order to preliminarily verify and test the implementation of the prototype, a L1b dataset has been generated for all the realistic dwells of each of the four MTG-IRS Local Area Coverage (LAC) of the Earth's disk. The cloudy-sky radiances have been calculated by RTTOV, with true 3D viewing geometry, including atmospheric profiles constructed from ECMWF forecast fields of March, 15 2016 at 1200UTC, along the instrument's line of sight.

The dataset consists of the 150 principal components scores of the synthetic radiance in each of the two MTG-IRS spectral bands, 680-1210 and 1600-2250 cm^{-1}, respectively. This dataset serves two purposes:

- L1D measurements as an input to the processor prototype
- Training database to compute the regression coefficients of the PWLR

The prototype has been tested on 6 continuous dwells centred around France, which allows the reconstruction of the vertical profiles from the slanted profiles performed in the post-processing stage.



Scores for the leading eigenvector in low-wavelength range band (left) and mid-wavelength range band (right).

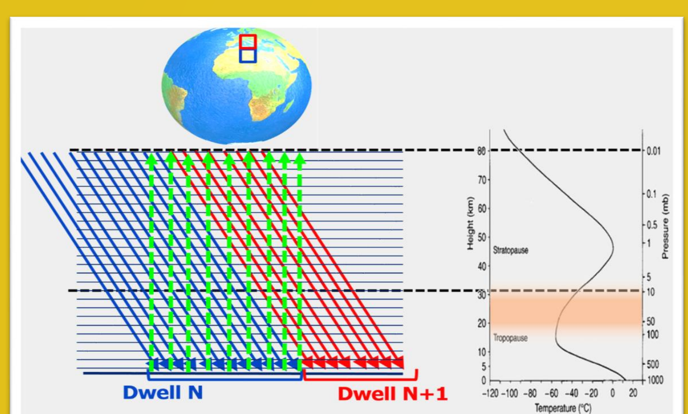
Pre-processing

The main input data to the IRS L2 processing are the IRS L1 radiances represented as principal components (PC) scores, which is the baseline for Near Real Time dissemination in the operational processor. It contains two types of radiance PC scores. The "global" scores, derived from static eigenvectors, and the "local" scores, derived from dynamic eigenvectors generated on the fly during the L1b product generation. For the IRS L2 product generation, only the global scores are used.

The pre-processing module extracts the global scores and geometric information from the input data chunks and performs data quality checks. It then calculates for each IRS pixel the instrument PSF-weighted land fraction and the surface elevation from a high resolution digital elevation model. Finally, it collocates the ECMWF fields along the satellite slant view. This implies the calculation of the Sub-Profile Point.

The collocated NWP data can optionally be replacing the PWLR profiles as *a priori* for the optimal estimation.

In the testing phase, the spatial sampling of the forecast fields is 0.25°.



The vertical atmospheric profiles are reconstructed from the collocated ECMWF profile variables on the slanted satellite's line of sight. Observed slanted profile (red) and vertical 1D profile (green) as stored in the forecast data.

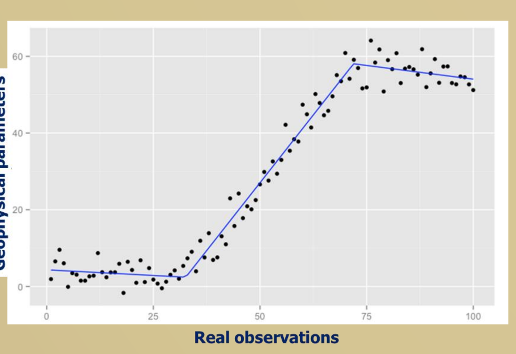
1st retrieval: PWLR

PWLR is a statistical retrieval algorithm relying on regression classes. These classes represents similar meteorological regimes or observation conditions. The PWLR concept is illustrated below.

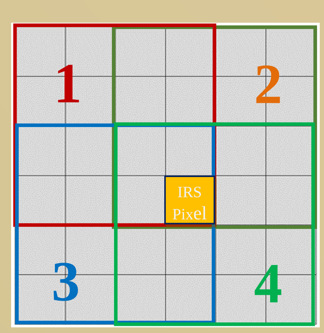
The classification itself is obtained with the k-mean clustering technique applied to a subset of the leading PC scores.

Different regression coefficients are also applied depending on surface type (land or sea), time (day and night) and the LAC being processed.

Because the full vector of input L1 PC scores for the 16 pixels would contain a big portion of co-linearity coming both from the spectral correlation (between the two bands) and the spatial correlation (between the 16 adjacent pixels), it is necessary to first compute PC scores of the input PC scores. This is done in two hierarchical steps as follows for improved computational efficiency and simpler handling of missing pixels (for example off-disc pixels or pixels with bad quality spectra). First the PC scores in both band 1 and band 2 of each group of 4 adjacent (2 by 2) pixels are combined to a single set of PC scores, then the final set of PC scores for the 4 by 4 pixels is obtained by combining the 4 sets of the



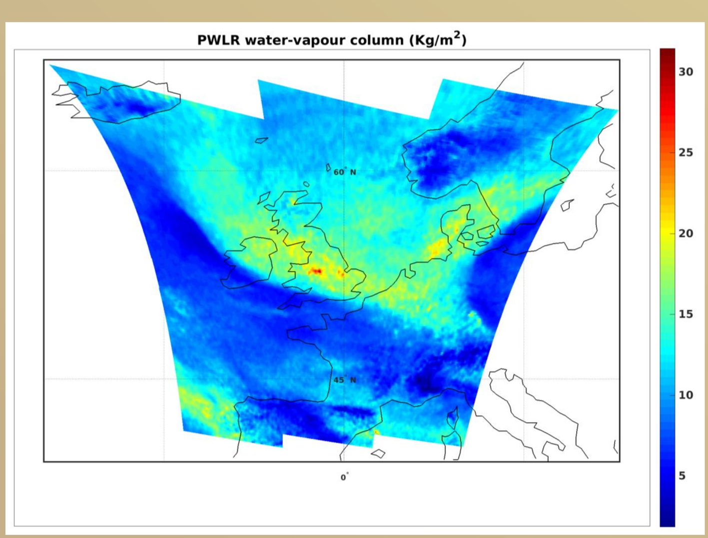
Piece-wise linear regression concept.



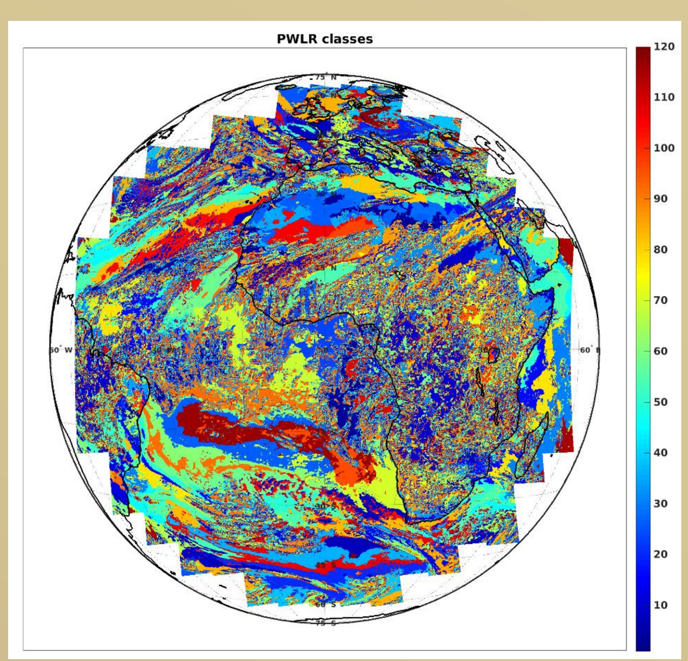
PWLR aggregation boxes.

PC scores for 2 by 2 pixels computed in the first step. These final PC scores are supplemented in the PWLR input vector by the secant of the average of the 16 satellite zenith angles and the surface elevation in meters of the 16 pixels in line by line order. The co-linearity introduced by the inclusion of 16 adjacent surface elevations is taken care of by using ridge regression for the computation of the regression coefficients.

The algorithm specified here also makes use of the 3D correlations within the observations by performing retrieval for groups of 4x4 IRS pixels. These retrieval boxes are shifted by 2 pixels in both horizontal and vertical directions, creating an overlap as shown above. This results in 79x79 PWLR retrievals per IRS dwell. The final outputs for each IRS pixel are obtained through a combination of the 4 separate retrievals that included this pixel (only 2 for the pixels at the edges of the dwell).



Water vapour total column amount from integration of PWLR retrieval.



Testing PWLR classes based on ECMWF forecast on March 15,2016. Only dwells with all valid pixels are shown.

The training for the regression classes is performed offline using re-analysis fields from ECMWF and the synthetic observations training set for the predictors. The classification itself is obtained with the K-means clustering technique applied to a subset of the leading L1 PC scores.

Cloud analysis

The purpose of this algorithm is to identify, characterise and geo-localize the clouds in the field of view. A simple but useful model in satellite products represents the clouds as black bodies, characterised with an emitting top located at a certain altitude (pressure level) and an effective fractional coverage.

With this simple cloud modelling, the observed radiances at the top of the atmosphere R_v at wavenumber v writes as a linear combination of the clear-sky and cloudy radiance terms:

$$R_v = (1 - \alpha_v) * R_v^{clear} + \alpha_v * R_v^{cloudy}(p)$$

where R_v^{clear} is the cloud-free radiance, $R_v^{cloudy}(p)$ the overcast radiance with cloud-top at pressure p and α_v is the effective cloud coverage at wavenumber v. The RTTOV radiative transfer model is used to simulate the cloud-free radiances as well as overcast radiances with a hypothetical cloud top height at different pressure levels.

The minimisation of a cost function based on the differences between simulated radiances and IRS reconstructed radiances allows to identify whether a pixel is cloudy, and what is the most probable cloud top height and cloud fraction. The algorithm applies two cost functions independently; one assuming a single layer cloud and one assuming a 2-layer cloud. However, the 2-layers cloud model is still experimental.

The cloud phase is determined by a simple brightness temperature difference tests on 3 channels between 8 and 12 micrometres. The difference is compared to a threshold that discriminates between ice, liquid water and mixed phase.

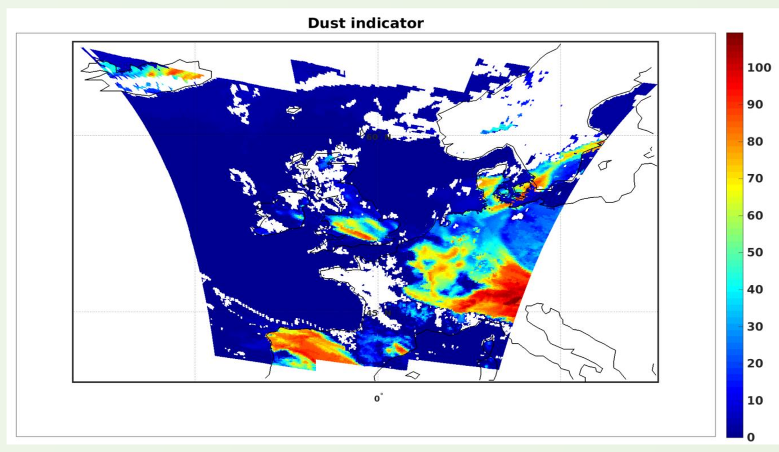
The unitless retrieved dust indicator is a weighted projection of the observed signal onto typical aerosol signature:

$$R = k^T S^{-1}(y - \mu_c)$$

where:

- k is the dust Jacobians
- S is the covariance matrix of dust-free spectra
- y is the observed spectrum
- μ_c is the mean dust-free spectrum.

k, S and μ_c can be defined empirically from a static training base of real observations, including dust-free and dust-contaminated pixels. Different static configuration of the gain matrix $G = k^T S^{-1}$ are pre-computed off-line for the different land/sea and day/night combination separately. The presence of dust can be suspected when the indicator R exceeds a configurable threshold, typically of 2 to 3.



The dust indicator is part of the scene characterization.

2nd retrieval: OEM

While the first retrieval is performed in both clear and cloudy situations, the second retrieval – the optimal estimation method – is only performed in pixels which are clear, or considered clear enough to apply cloud-free radiative transfer modelling. The retrieval of atmospheric and surface parameters by this variational method is formulated as the minimization of the cost function J, sum of the background (J_x) and the observation (J_y) terms:

$$\begin{aligned} J &= J_x + J_y \\ &= (x - x_a)^T S_x^{-1} (x - x_a) + (F(x) - y)^T S_y^{-1} (F(x) - y) \end{aligned}$$

where:

- $x = t_1,\dots,t_{nt},w_1,\dots,w_{nw},o_1,\dots,on_o,c_1,\dots,c_{nc},T_s$ is the state vectors to be retrieved in the form of PC scores. *T, w, o, c* are the temperature profile, water vapour profile, ozone profile and carbon dioxide, represented as *nt, nw, no, nc* PC scores, respectively. T_s is the surface skin temperature. OEM CO_2 retrieval is optional and is not yet active in the current version of the prototype. The CO_2 profile from PWLR is thus carried over as an input to the forward model.
- x_a is the *a priori* state vector (out of PWLR, default configuration).
- y is the observation vectors and it is a subset of reconstructed radiances which suppress instrument artefacts and preserves the full information of the signal.
- $F(x)$ is the simulated observations vectors using the forward model (RTTOV)
- S_x and S_y are the background error and observation error matrices, respectively. The full matrix S_y is the combined measurement and forward model error.

The cost function is minimized using Newton's method.

The PC scores used in the state vector are based on the deviations from the *a priori* profile given at the 101 fixed pressure levels RTTOV grid. Therefore, by construction, the *a priori* state vector in the PC space, is the null vector. If X_a is the *a priori* state vector in the profile space, the profiles X in the profile space can be calculated as:

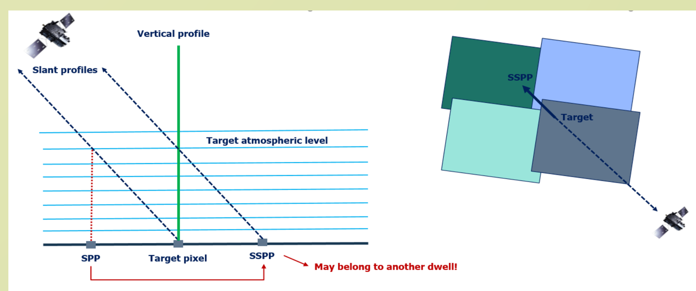
$$X = X_a + E_x.x$$

where E_x are the PC eigenvectors calculated offline.

At the outcome of this activity, the temperature and water vapour profiles are corrected for super adiabatic and super saturation, respectively.

Plus, error covariance matrices for each profiles are extracted.

Post-processing



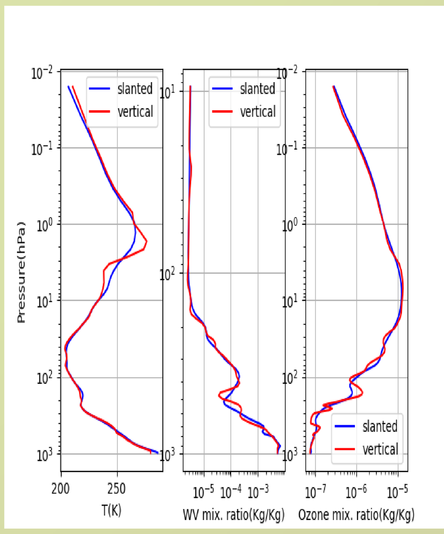
Observed slanted and vertical profile provided to the end-users.

The purpose of the Post-Processing is to reconstruct the information at the vertical of the target point from the slanted retrievals that are intercepting the vertical line, as the profiles at the vertical of the target point are intended to be delivered to the end-users.

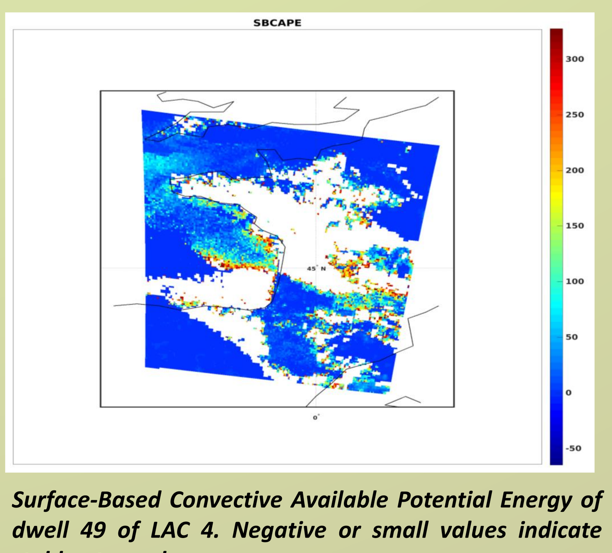
In addition to the profile conversion described above, the Post-Processing activity also includes the computation of the instability indices derived from the retrieved profiles, as well as the encoding of the retrieved geophysical quantities and associated processing information into final IRS L2 Products.

Short description	Content	Purpose
Atmospheric Temperature and WaterVapour profiles	'All-sky' first retrieval + quality indicators Clear-sky variational retrieval	NRT Dissemination Archive
Surface parameters	Skin surface temperature over ocean (SST) and continental surfaces (LST) Land surface emissivity in selected channels (of the order of 10), from which the entire spectrum can be reconstructed by application of PC transformations	NRT Dissemination Archive
Cloud products	Cloud detection Cloud top pressure and effective fractional coverage Cloud phase Dust indicator	NRT Dissemination Archive
Atmospheric ozone profiles	'All-sky' first retrieval + quality indicators Clear-sky variational retrieval	NRT Dissemination Archive
Convective parameters	Collection of instability indices and integrated quantities to support the assessment of potential convective initiation	NRT Dissemination Archive
Error covariance matrix	Is the theoretical posterior error estimate computed after Rodgers [Rodgers 2000] for the parameters retrieved with the OEM: temperature, humidity, ozone	Archive

List of IRS L2 products.



Comparison of slanted and reconstructed vertical OEM retrievals (temperature, water vapour and ozone, respectively).



Surface-Based Convective Available Potential Energy of dwell 49 of LAC 4. Negative or small values indicate stable atmosphere.

Instability indices are used to identify atmospheric instability conditions that can lead to convection processes. There exists a large diversity of stability indicators which have been conceived from atmospheric sounding profiles and past satellite missions.

The list of indices to be calculated within the IRS L2 processing facility aims at ensuring continuity with the MSG GII product and consistency with the MTG-FCI follow-up products. It is completed with a few more indices collected in the literature and from initial interactions with forecasters.

Please send comments and questions to: cedric.goukenleuque@external.eumetsat.int