NWCSAF/MSG PGE13 PHYSICAL RETRIEVAL PRODUCT VERSION 2010

**Miguel A. Martinez[1], Jun Li[2], Ramiro Romero[1]**

(1) Agencia Estatal de Meteorología (AEMET), Spain
(2) CIMSS/University of Wisconsin-Madison

Abstract

PGE13 SEVIRI Physical Retrieval (SPhR) is available as a new product of NWCSAF/MSG 2010 package. The NWCSAF/MSG physical retrieval algorithm and the improvements on the NWCSAF 2010 version are presented. The selected algorithm has been optimized for the purpose of an operational implementation; the algorithm targets the region where atmosphere is more unstable and will potentially cause convective storm development. The use of PGE13 in one case study and statistical validation of PGE13 are presented. Ideas for future versions of physical retrieval for MSG and MTG are also discussed.

INTRODUCTION

The main objective of NWC SAF is to produce Software packages for MSG and Polar satellite. In the case of NCWSAF/MSG package, the purpose of the NWC SAF algorithms is to derive each product every 15 minutes over a region selected by user with MSG data. The NWCSAF Steering Group requested to develop a physical retrieval product during CDOP phase. After Jun Li's Visiting Scientist Activities, the product was developed. PGE13 SPhR was approved and declared as pre-operational in the review that took place on April 2010 at Madrid. Thus, PGE13 SEVIRI Physical Retrieval (SPhR) software is included as a new product since version 2010. All documentation is available in the new NWC SAF Help Desk; click on *description* on the NWC SAF web page http://www.nwcsaf.org.
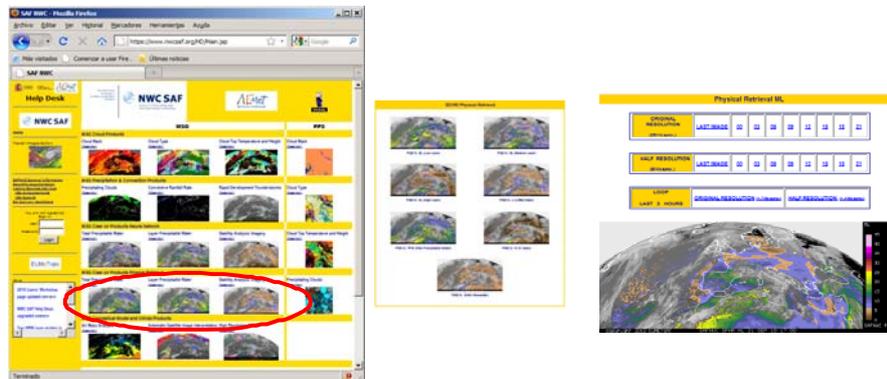
Figure 1: **NWC SAF web page showing the PGE13 SPhR main fields. Click on description for documentation. Click on thumb images to access to real time PGE13 SPhR last images and loops.**

Images and loops of the main outputs (precipitable water and instability indices parameters) are available on near real time in the NWCSAF Help Desk open area (click on the thumb images). The users can compare their outputs with the ones of the NWCSAF Reference system.
http://www.nwcsaf.org/products/SAFNWC_SPHR_ALL_Europe.html

The objective of this paper is to present to the users the main features of the new PGE13 SPhR. The paper is divided in two parts. In the first part, the algorithm proposed by Jun Li is briefly summarized together with some examples from the Convection Working Group (CWG) case study. In the second part, a summary of the big effort carried out in validation is made.

SUMMARY OF THE PHYSICAL RETRIEVAL ALGORITHM

Based on Jun Li' algorithm (Li 1999 and 2000), PGE13 SEVIRI Physical Retrieval (SPhR) is an optimal estimation algorithm with some differences over the classical approach. These ones are**:**

- Use of a non-linear regression to build the First Guess. Instead to use directly as First-Guess the background NWP profiles, the First-Guess is the result of non-linear regressions with the background NWP profiles and the bias corrected SEVIRI brightness temperatures.

- Use of empirical orthogonal functions (EOFs) to represent the temperature and moisture profiles in order to reduce the dimension of matrices to invert. In version 2010 of PGE13, three EOFs are used for humidity, two EOFs are used for temperature and one for Skin Temperature.

- Use of a regularization parameter (also called smoothing factor) introduced for convergence and solution stability

The algorithm is only applied on clear air pixels. RTTOV-9.3 is used for synthetic BT calculations (forward model) and Jacobian calculations. After First Guess (FG) regression calculation step, an estimation of the distance (BT_RMS) between SEVIRI BTs and synthetic BTs using as input to RTTOV the FG profile is made. BT_RMS is the distance on the non-window channel (WV6.2, WV7.3 and IR13.4) between SEVIRI BTs and RTTOV BTs. If the distance is greater than a configurable threshold (*BT_RMS_THRESHOLD*), the iterative physical retrieval process is applied.
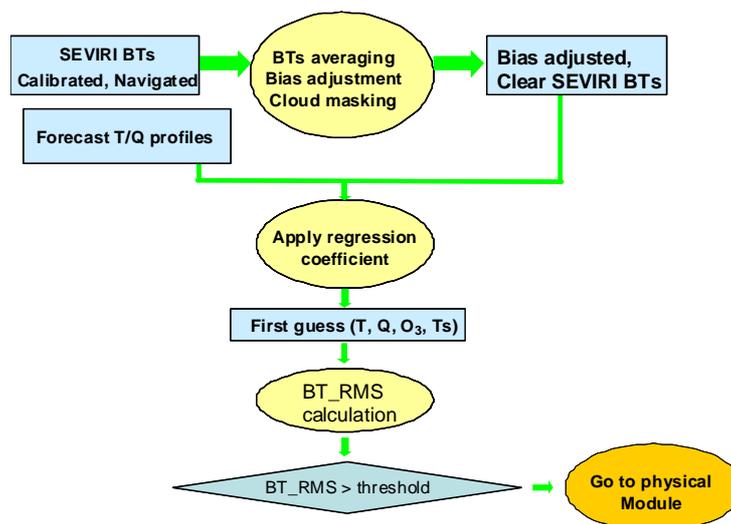

Figure 2: **PGE13 SPhR flowchart.**

For whole mathematical description and details, see the PGE13 Algorithm Theoretical Basis Document (ATBD) that is available at the NWC SAF web page.

MAIN FEATURES OF PGE13 SPHR

Since the physical retrieval can be applied only over the clear pixels, PGE01 Cloud Mask must be executed before the PGE13. In operational mode, NWCSAF/MSG Task Manager synchronises the execution of the products so the first product that is generated upon the arrival of a new image is the PGE01. Once the Cloud Mask is available, the Task Manager can execute PGE13 using as inputs SEVIRI images on HRIT format, background NWP at fixed pressure levels and the Cloud Mask.

Background NWP GRIB files from t+00 to t+24 hours forecast are needed. In real time operational mode, NWCSAF/MSG package has predefined tools (coordinated by Task Manager daemon) to make automatically the spatial remapping to the predetermined regions once new NWP GRIB files are receipted. This avoids repeating the process for every slot. Due to PGE13 needs to get NWP data collocated with SEVIRI data at the 43 RTTOV pressure levels, temporal and vertical interpolation is

made within PGE13 only at clear air window of MxM pixels. Users can use their own model and choice for the background NWP supply. Thus, any user that is running NWCSAF/MSG package could be able to run PGE13. As example, in the NWC SAF reference system, the NWP background model is ECMWF with 0.5º x 0.5º spatial resolution and 15 pressure levels, increasing to 0.125º x 0.125º spatial resolution on AEMET Intranet.

The main output is a file on HDF-5 format that, among others, includes four parameters related to the precipitable water, three instability indices and the differences between the parameters calculated from the retrieved and the background NWP profiles. See more details on the next section.
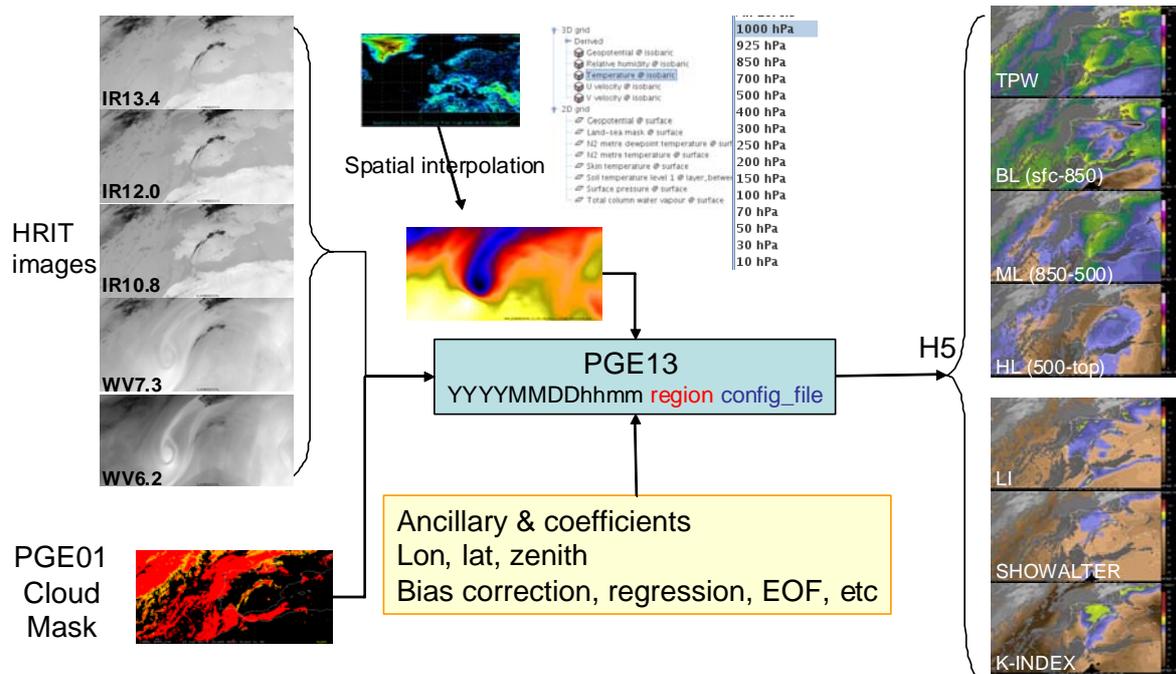

Figure 3: **PGE13 inputs and outputs scheme.**

The parameters of PGE13 program are the region where it will be processed and one ASCII configuration file.

Main PGE13 configurable options and parameters (PGE13 SPhR Model Configuration File)

The user can edit the ASCII configuration file that contains the coefficients file names and some parameters required by PGE13 software to run. See the PGE13 Product User Manual for the details available at the NWC SAF web page.

The user must establish a compromise between the machine characteristics (CPU speed and memory) where the NWCSAF/MSG package will be operated, the region (or regions) to process and some configurable keywords. The main ones are: the window size for processing in windows of M x M pixels (default 3 x 3), the number of iterations and the *BT_RMS_THRESHOLD* and *MAX_RESIDUAL* keywords. As example, the time execution multiplies by nine when PGE13 is run using a window of 1x1 instead of the default window of 3x3. The most CPU expensive part is the physical retrieval module due to the call to the *rttov_k* routine (needed to calculate the Jacobian). Thus, all keywords related to the physical retrieval part affect the time execution. Those keywords are: the number of iterations (maximum number three), the threshold to call the physical module (*BT_RMS_THRESHOLD*, see Figure 2) and the error threshold to stop the iterative process (*MAX_RESIDUAL* keyword).

The user management of the background NWP parameters is also made in the configuration file through the keywords that establish the fields, number of vertical levels and the spatial interpolation method. Another important configurable keyword is the method to calculate the BTs of the MxM

window (the mean of all clear pixels or the SEVIRI BTs of the warmest clear pixel at the IR10.8 channel).

PGE13 SPhR HDF-5 MAIN OUTPUTS

The main output of PGE13 is one HDF-5 file that contains several fields. The fields are key ingredients in convection triggering. There are four fields related with the content of the precipitable water calculated from the specific humidity retrieved profiles: Total Precipitable Water (**SPhR_TPW**) and the Layer Precipitable Water (LPW) in three layers of the atmosphere(**SPhR_BL** [Pressure$_{Surface}$, 850 hPa], **SPhR_ML** [850, 500 hPa] and **SPhR_HL** [500, 0.1 hPa]) and three fields related with the instability indexes calculated from the specific humidity and temperature retrieved profiles: Lifted Index (**SPhR_LI**), Showalter Index (**SPhR_SHOWALTER**) and K-Index (**SPhR_KI**).

Together with the parameters calculated directly from the retrieved profile, it was considered as adequate at 2007 Madrid Workshop to provide as other outputs the differences between the retrieved profiles and background NWP model. Following this idea, the differences between the parameters obtained with the retrieved profile and the same parameters obtained with the NWP model profile have been included as additional fields (see the PGE13 User Manual Document for details). Thus, seven fields have also been included: **SPhR_DIFFTPW**, **SPhR_DIFFBL**, **SPhR_DIFFML**, **SPhR_DIFFHL, SPhR_DIFFLI**, **SPhR_DIFFKI**, **SPhR_DIFFSHW**.

Finally two quality Flags fields are included: **SPhR_QUALITY** and **SPhR_SFLAG**.

OPTIONAL PGE13 SPhR 3D BINARY FILES OUTPUT. CONVERSION OF PGE13 BINARY FILES TO NETCDF FORMAT.

As an optional output, the 3D intermediate retrieved temperature and humidity profiles resulting from the physical retrieval module and the profiles from the background NWP model interpolated at 43 RTTOV pressure levels may be written on 3D binary files. The users can easily activate the generation by editing an option in the ASCII configuration file. Then, the binary files will be written in the $SAFNWC/tmp directory. This allows users to debug their local implementation, to get access to the 3D retrieved temperature and humidity profiles and to compare them with the background NWP profiles.
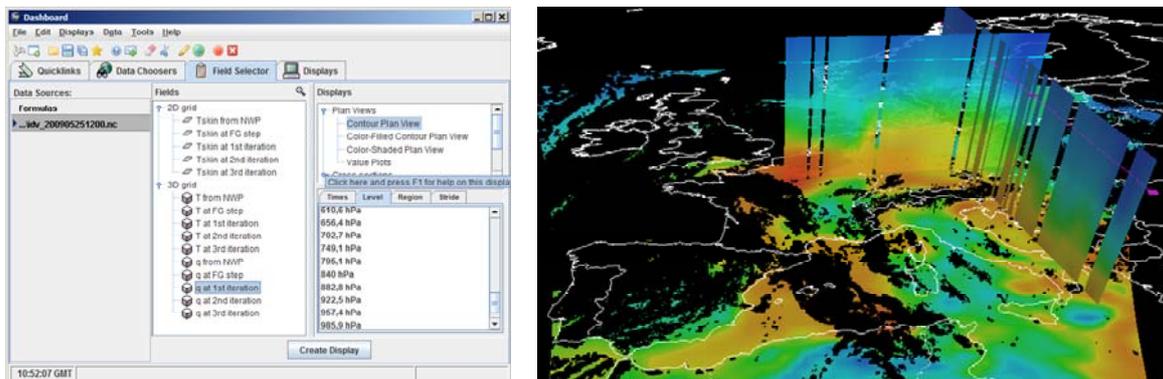

Figure 4. IDV visualization of retrieved q profiles with netCDF for 25th May 2009 at 12 UTC.

Since netCDF format can be managed by several tools, one IDL prototype to convert the PGE13 binary files to netCDF format has been codified. Among them, McIDAS-V and IDV are freely available on their respective web pages(http://www.ssec.wisc.edu/mcidas/software/v/, http://www.unidata.ucar.edu/software/idv/). In the Figure 4 an example is shown. Currently the IDL prototype will be supported as best effort basis and it is available on request to mmartinezr@aemet.es. It can be migrated to C or Fortran or added the possibility of writing directly in netCDF in future versions.

In the next figures, examples from the case study selected by the Convection Working Group (CWG) have been selected. The presentation with the loops is available at the NWCSAF web page.
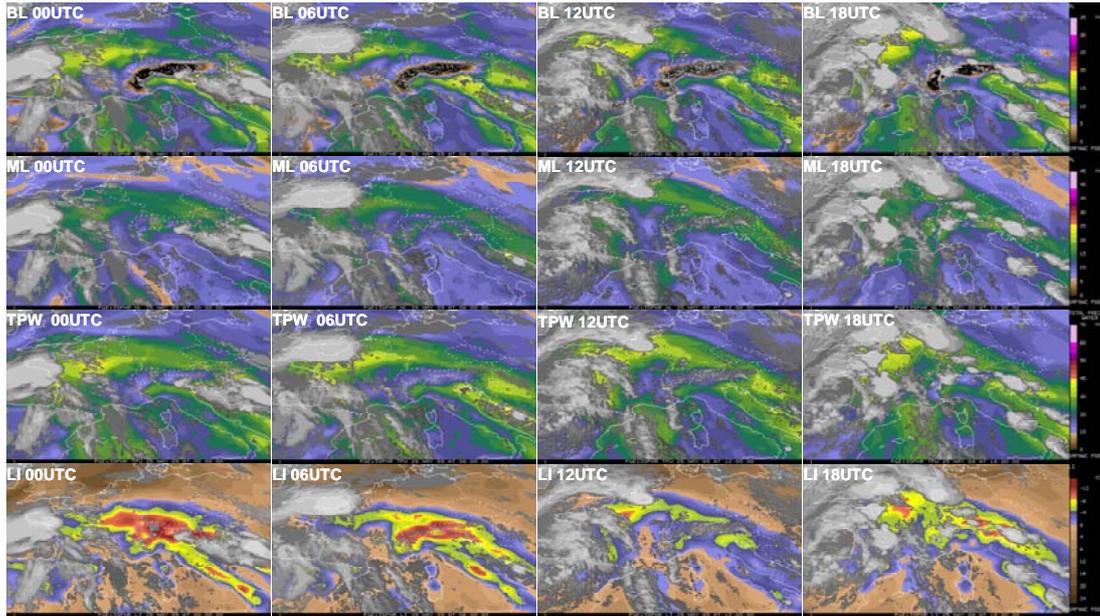
Figure 5: Strong convection over Europe: *1ˢᵗ CWG case study 25ᵗʰ May 2009* reprocessed at pixel by pixel scale. **From top to bottom, sequence of four images every six hours 00 to 18 UTC of PGE13 SPhR parameters BL (precipitable water surface-850 hPa); ML (precipitable water 850 hPa – 500 hPa); TPW and Lifted Index (LI)**.
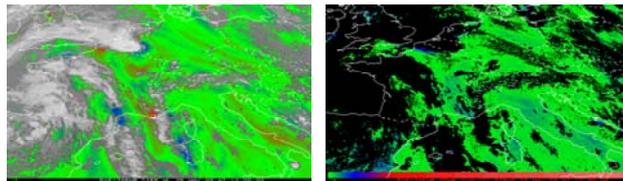
Figure 6: Strong convection over Europe: *1ˢᵗ CWG case study 25ᵗʰ May 2009* reprocessed at pixel by pixel scale. *On the left,* DIFFML (ML$_{retrieval}$-ML$_{ECMWF}$)*; on the right: the residual quality field.*

In the Figure 6, an example of the difference between the ML parameter calculated from the q retrieved profile and ML calculated from the background ECMWF q profile is provided; blue (red) regions indicate background NWP overestimation (underestimation) and they are the regions where the background NWP profiles have been modified by the physical retrieval with the satellite data. This can be used to investigate the value add of physical retrieval in relation with the background NWP.

PGE13 VALIDATION AND TRAINING DATASET

Building a good collocated SEVIRI BTs, NWP and radio sounding profiles dataset is a very important task for us. The use of this dataset has two main aims. The first one is to serve for the validation and tuning of the current version of the algorithm. The second one is the training, testing and validation of new versions of SPhR. Later, the use of the profiles of this dataset (containing several million profiles by year) with suitable RTTOV coefficients files will be an essential piece for the design of MTG era algorithms (both MTG-FCI and MTG-IRS). The procedure followed can be extended to other instruments such as the ABI instrument on GOES-R. In order to build the current PGE13 SPhR dataset, the PGE13 optional saving in binary files is used to get collocated data of MSG and background NWP model, in our case ECMWF model. Radio sounding profiles (hereafter RAOB) have been kindly provided by Larry D. Oolman from Wyoming University.

For the generation of PGE13 dataset, the 0 UTC and 12 UTC MSG images from January 2008 until today have been reprocessed. In order to reduce the execution time of each image, instead of processing pixel by pixel the full SEVIRI disk, a network of 13001 validation points has been designed. The validation network is composed for the pixels that inside the full disk SEVIRI image are nearest to the 1º x1º grid and the pixels closest to the RAOB stations provided by the University of Wyoming. For each image the following process is repeated. First, the program PGE01, using as background NWP the ECMWF analysis, is executed to calculate the optimal full disk cloud mask. Second, the full disk

cloud mask is filtered to retain as clear pixels only those ones that belong to the validation network of 13001 points while the rest are flagged as not clear. The filtered cloud mask is written as the PGE01 cloud mask to use hereafter. Then, the PGE13 program is executed twice using the filtered cloud mask (thus, instead of having several million of clear pixels by image, only several thousand of clear pixels at selected points are retained and the execution of PGE13 can be done in a few minutes for the entire disc).

The ECMWF analysis (t+00) is used as background NWP in the first run of PGE13, while in the second run of PGE13, the background NWP used is the ECMWF t+12 hours forecast from the previous 12 hours ECMWF model (i.e. for reprocessing the image of January 2009 at 0 UTC is used the ECMWF t+12 forecast from 31st December 2008 at 12 UTC ECMWF run). Since the two PGE13 executions were run with the option of saving in binary files the intermediate T and q profiles at clear air pixels (see PGE13 User Manual at NWC SAF website), it is possible to get spatially, temporally and vertically collocated profiles from the background NWP, the profiles after the FG step and the profiles after the physical retrieval step. After reading the PGE13 binary files in the directory $SAFNWC/tmp, a record is obtained for each clear air pixel. Each record contains: the SEVIRI BTs, all ancillary parameters (longitude, latitude, zenith angle, topographic height, land-sea mask, emissivity, date, time, etc), the temperature and q profiles from the ECMWF analysis (t+00) and ECMWF t+12 hours forecast of the previous ECMWF run, the T and q profiles after FG step, the T and q profiles after every iteration and the surface temperature at the same steps. Finally, the synthetic RTTOV BTs are calculated for each one of the steps (background NWP, after FG step and each iteration step) and added to the record. This process has been used to reprocess the 00 UTC and 12 UTC images from January 2008 to present. In order to make the management of the records easier, they have been joined in files by months. All the management of the records has been implemented using IDL. The PGE13 validation dataset has been used in the generation of BT bias correction coefficients, in the generation of the PGE13 FG regression coefficients and for the validation of the PGE13 version 2010. Full details are available in the PGE13 Validation Report (available at NWC SAF Website). Another collocated dataset contains the RAOB observation from Wyoming University.

BT bias correction

In order to calculate the BT bias correction, first the evolution of the BT bias correction between BT_SEVIRI and synthetic BT_RTTOV was studied using the PGE13 validation dataset. After the inspection of the evolution of the biases and after the analysis of the spatial distribution of the error, only sea clear pixels for period January 2009 to December 2009 have been used to calculate the bias correction. The calculation was made using robust regression between the collocated SEVIRI BTs and synthetic SEVIRI RTTOV BTs calculated from ECMWF analysis. The bias correction coefficients have been written as the default values in "*safnwc_pge13_msg2.cfm*".

FG training dataset FG regression coefficient file for 2010 generation Background NWP model ECMWF at 15 fixed pressure levels

Since the current version of NWC SAF library only allows reading GRIB files on fixed pressure levels and due to the fact that only 15 fixed levels pressure are available in the ECMWF, in the executions of PGE13 to build the first version of the PGE13 validation dataset, the temperature and specific humidity profile at the 43 RTTOV pressure levels were the result of interpolating them from the 15 fixed pressure levels available in the ECMWF (hereafter NWP15 profiles). Due to this fact, some vertical levels are linear combinations and it was needed to improve the vertical information available to each record. It was made by adding to every record another structure with the spatially and vertically interpolated profiles of T and q from interpolating the 91 hybrid pressure levels ECMWF profiles at the 43 RTTOV pressure levels (hereafter NWP91 profiles).

With this improved record version of the PGE13 validation dataset, the nonlinear regressions coefficients of the First-Guess step were calculated. Due to the PC memory constraints, only 1 out 6 records for period from January 2009 to December 2009 have been used to train the FG regressions. The FG regressions coefficient file contains 76 regression coefficients for every parameter; each regression corresponds to one local zenith angle ranging from 0 to 75 degrees.

Once the FG regressions were trained, they were validated using the PGE13 validation dataset. The profiles from the analysis (t+00) were considered as the truth and the 5 out of 6 observations of 2009 not used to build the training dataset were used to make the validation. Separated validations for sea, land and with RAOB observations are shown in the PGE13 Validation Report; it is available at the NWC SAF web page.

TEST OF A NEW FG REGRESSION WITH 91 HYBRID LEVEL PROFILES

At the 2010 PGE13 Validation Report, the temperature and q profiles used as input to PGE13 were the NWP15 profiles but some irregularities appeared in the inspection of the figures with the validation of the vertical profiles. To assess that these irregularities are caused by the use of the NWP15 profiles, a test has been made to train and validate a FG regression using as background NWP model ECMWF the NWP91 profiles. In the case of NWP91 FG training dataset the 43 levels are independent and it is possible to calculate the regression directly. The 43 levels have also been used (in the 2010 version the upper pressure levels were not used in the FG regressions). This new NWP91 FG regression file can not be implemented in the PGE13 SPhR without modifications in the NWCSAF library package and due to this fact, only the FG step can be validated. A full validation report is being written and the Figures 7 and 8 are shown as preliminary results of this NWP91 assessment study.
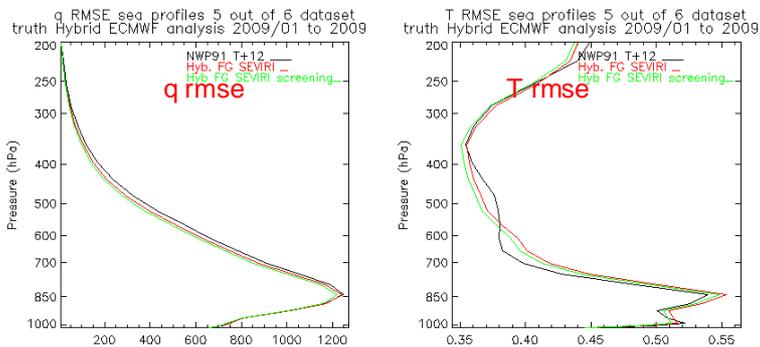

Figure 7. RMSE of q (ppmv) and of T (K) profiles compared with NWP91 ECMWF analysis in period January 2009 to December 2009 for sea 5 out 6 pixels (not used in FG regression calculation).
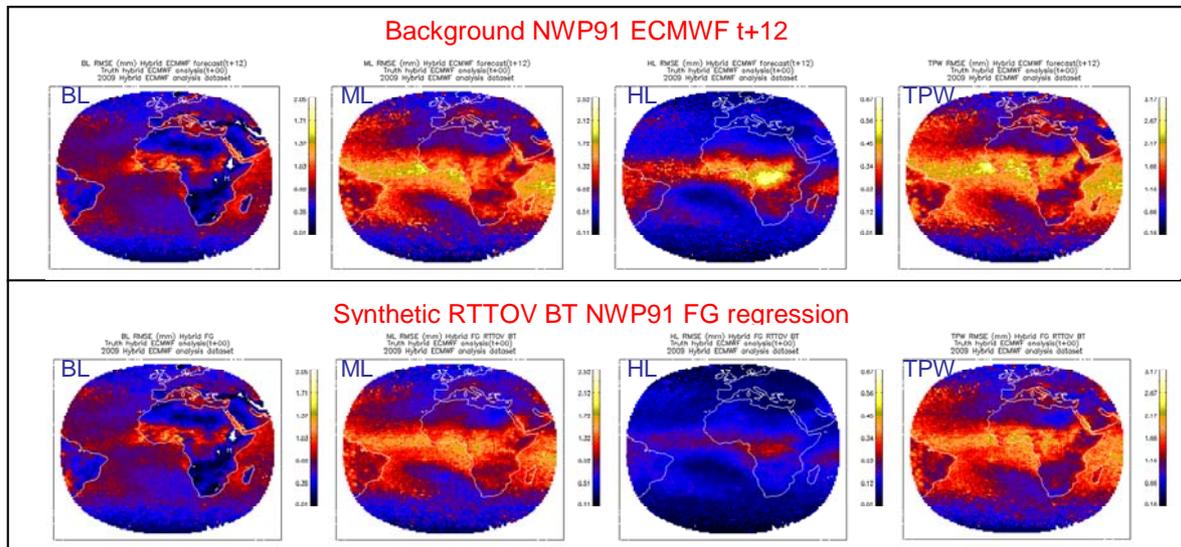

Figure 8. Spatial distribution of the LPW and TPW rmse over validation points. January 2009 to December 2009 for 5 out 6 pixels using ECMWF analysis as truth. Top: comparison with ECMWF t+12, bottom: NWP91 FG with BT RTTOV

In the Figure 7, it is represented: the temperature and q RMSE profile of the ECMWF t+12 forecast (black line) and after NWP91 FG regression using as inputs ECMWF t+12 and SEVIRI bias corrected BTs (red lines no screening is made, and green line after screening). The irregularities of the NWP15 validation (see the 2010 PGE13 Validation Report) have disappeared in the RMSE vertical profiles

and the improvement in the q profiles takes place in the middle layer. In the Figure 8, the spatial distribution of the RMSE is shown. The greatest values of NWP91 RMSE between ECMWF t+12 forecast and the ECMWF analysis appear near to the equatorial bell. When the NWP91 FG regressions are applied to the synthetic RTTOV BTs, there is a uniform reduction on the RMSE that is greater on the ML and HL layer. The synthetic RTTOV BTs were used here in order to assess that the error is not caused by the contamination by clouds.

CONCLUSIONS

A new PGE, PGE13 SPhR (SEVIRI Physical Retrieval), has been developed inside the NWCSAF/MSG package and it is available since the 2010 version. It has been extended validated with one year period (2009), for the full SEVIRI disk. Best results are obtained for humidity in medium layers due to the contribution of the two water vapor channels. SEVIRI has limited information to improve the vertical information beyond the forecast, but does provide useful spatial information. This limitation is clearer for the vertical information of temperature. Performance results present variations along the complete MSG disk but good results for all the parameters are obtained for European interest regions.

FUTURE WORKS

Next future works are related to increase the vertical information in the background NWP but previously, some modifications on the NWC SAF library and the approval of the Steering Group will be necessary. The training and validation dataset will be used to make a validation of the neural network products (PGE06, PGE07 and PGE08) and to make an assessment of their quality.

Concerning to the CDOP-2 phase, it is foreseen to prepare for MTG era and the validation dataset will be used to develop the MTG-FCI algorithm. The extension of the physical retrieval to other satellites is also foreseen in cooperation with Dr. Jun Li (CIMMS-Wisconsin). The IASI RTTOV coefficients and the conversion of IASI radiances to MTG-IRS BTs could be used to explore the development of MTG-IRS algorithms using the PGE13 validation dataset and the development experience.

Another issue is to explore the spatial and temporal resolution (gradients and trends between consecutive images) to obtain humidity flows/winds at different layers based on "optical flow" algorithm with collaboration of F.J Tapiador (Castilla-La Mancha University).

Last but not least, the collaboration with other SAFs and CAF will be an important task. It is foreseen the comparison of the physical retrieval skin temperature with OSI SAF's SST and Land SAF's LST. The continuity of the support to CM SAF to calculate the error estimation is also a priority.

ACKNOWLEDGEMENTS

REFERENCES

Li, J. and Huang, H.-L., 1999: *Retrieval of atmospheric profiles from satellite sounder measurements by use of the discrepancy principle.* Appl. Optics, Vol. 38, No. 6, 916-923.

Li, J.; et al., 2000. *Global soundings of the atmosphere from ATOVS measurements: The algorithm and validation.* J. Appl. Meteorol., Vol 39, pp 1248 – 1268.

Martínez M.A.; Li, J. (2009). *First operational results of the NWCSAF/MSG PGE13 SEVIRI Physical Retrieval algorithm.* Proc. The 2009 EUMETSAT Meteorological Satellite Conference, Bath, United Kingdom.