The EUMETSAT Network of Satellite Application Facilities

NWC SAF

Support to Nowcasting and Very Short Range Forecasting

Validation Report for "SEVIRI Physical Retrieval Product" (SPhR-PGE13) v1.2

SAF/NWC/CDOP/INM/SCI/VR/11, Issue 1, Rev.0

15 February 2012

Applicable to SAFNWC/MSG version 2012

Prepared by AEMET

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	Miguel A. Martinez (AEMET) Ramiro Romero (AEMET)		*15 February 2012*
Reviewed by	Marcelino Manso Rejón		*15 February 2012*
Authorised by	Pilar Fernández Álvarez SAFNWC Project Manager		*15 February 2012*

DOCUMENT CHANGE RECORD

Version	Date	Pages	CHANGE(S)
1.0	*15 February 2012*	28	Initial version for validation of "SEVIRI Physical Retrieval Product " (SPhR-PGE13) v1.2.

TABLE OF CONTENTS

1. INTRODUCTION....7

1.1 SCOPE OF THE DOCUMENT....7
1.2 SOFTWARE VERSION IDENTIFICATION....7
1.3 GLOSSARY....7
1.4 REFERENCES....7
1.4.1 NWC SAF Applicable Documents....7
1.4.2 External Reference Documents....8

2. DATA FILES USED....9

3. SEVIRI BRIGHTNESS TEMPERATURES BIAS CORRECTION....10

3.1 IR SEVIRI BRIGHTNESS TEMPERATURES BIAS ESTIMATION DATASETS DESIGN....10
3.2 USE OF THE BRIGHTNESS TEMPERATURES BIAS ESTIMATION DATASET....11
3.2.1 Spatial distribution of the error between SEVIRI_BT and RTTOV_BT....11
3.2.2 Evolution with the time of the bias in different channels....11
3.3 SEVIRI BT BIAS CORRECTION COEFFICIENTS FOR PGE13 SPHR VERSION 2012 FROM NWP15 BIAS ESTIMATION DATASETS....13

4. VALIDATION OF PGE13 SPHR OUTPUTS....15

4.1 ANALYSIS OF THE ALGORITHM CONVERGENCE BETWEEN THE SEVIRI BTS AND THE SYNTHETIC RTTOV BTS AT THE DIFFERENT STEPS OF THE ALGORITHM....17
4.2 ANALYSIS OF THE VERTICAL PERFORMANCE OF THE ALGORITHM AT THE FULL DISK DATASET....20
4.3 2D DIMENSIONAL HISTOGRAMS OF PGE13 SPHR PARAMETERS....21
4.4 SPATIAL ANALYSIS OF PGE13 SPHR PARAMETERS....24
4.5 VALIDATION AGAINST ECMWF: STATISTICAL SUMMARY....26

5. CONCLUSIONS....27

6. REFERENCES....28

List of Tables and Figures

Table 1: List of Applicable Documents. 8

Table 2: List of Referenced Documents.. 8

Table 3: Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in period 2009 for odd pixels dataset. Left) sea pixels, right) land pixels. 26

Table 4: Summary of the PGE13 SPhR statistical parameters in period January 2009 to December 2009 for odd pixels synthetic dataset. 27

Figure 1: Spatial distribution of the number of observation for each 25x25 FOR included in the PGE13 BT bias correction in the period November 2007 to September 2011. 10

Figure 2. Left: mean error between BT_SEVIRI and BT_RTTOV. Right: standard deviation for each 25x25 FOR in the period 2007/11 to 2011/09.. 11

Figure 3: Evolution of the BT bias correction between BT_SEVIRI and synthetic BT_RTTOV. The differences between a mean value before and after the BT bias correction calculated for a "moving" window of one month for five SEVIRI channels... 12

Figure 4: EUMETSAT GSICS Meteosat IR Inter-calibration with IASI snapshot: BT bias evolution from inter-comparison of equivalent infrared channels of geostationary SEVIRI and the polar-orbiting IASI sounder from collocated data.. 13

Figure 5: Scatter plots for IR13.4 SEVIRI BT at IR10.8 warmest clear pixel in 25 x 25 box SEVIRI BT versus ECMWF+RTTOV-9.3 synthetic BT over full disk region. On the left before BT bias correction, on the right after robust BT bias correction.. 14

Figure 6: Predefined set of points used in the FG regression dataset. Grid network of 1º x 1º plus Radiosonde Stations (red crosses). 15

Figure 7: Creation of PGE13 validation dataset for one image.. 16

Figure 8: Histogram of Error BT_RMS (distance in absorption channels) at different steps of the PGE13 SPhR over all 1 out of 2 pixels in period 2009/01 to 2009/12. Left) RTTOV reference case, right) PGE13 algorithm using as input actual SEVIRI case.. 18

Figure 9: Spatial distribution of mean distance between synthetic RTTOV from ground truth (ECMWF analysis from NWP91H dataset) and BTs at different step of PGE13 algorithm when using as input bias corrected SEVIRI BTs. Top) BT_distance is the distance using all SEVIRI channels, bottom) BT_RMS is the distance using WV6.2, WV7.3 and IR13.4 channels. Left) error calculated vs RTTOV BT calculated at NWP15 (T+12) background NWP profiles, middle) error calculated vs RTTOV BT after FG step profiles and right) error calculated vs RTTOV BT after 1st iteration step profiles. Odd pixels in period 2009/01 to 2009/12. 19

Figure 10: Same that Figure 9 but synthetic RTTOV BTs from NWP15(T+00) are used as input to the PGE13 algorithm. 19

Figure 11: RMSE of q (ppmv) of q profiles at different steps compared with ECMWF analysis hybrid profiles in period January 2009 to December 2009 for odd pixels. Left) sea pixels, right) land pixels. 20

Figure 12: Bias of q(ppmv) of q profiles at different steps compared with ECMWF analysis hybrid profiles in period January 2009 to December 2009 for odd pixels. Left) sea pixels, right) land pixels. 20

Figure 13: LPW and TPW 2D histograms over sea validation points in period January 2009 to December 2009 for odd pixels. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW parameters calculated after FG step profile, right) BL, ML, HL and TPW parameters calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles. 22

Figure 14: LPW and TPW 2D histograms over sea validation points in period January 2009 to December 2009 for odd pixels. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW parameters calculated after FG step profile, right) BL, ML, HL and TPW parameters calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles. 23

Figure 15: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in period January 2009 to December 2009 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles. 24

Figure 16: Same that Figure 15 but relative RMSE instead of RMSE. 25

1. INTRODUCTION

1.1 SCOPE OF THE DOCUMENT

The Eumetsat "Satellite Application Facilities" (SAF) are dedicated centres of excellence for processing satellite data, and form an integral part of the distributed EUMETSAT Application Ground Segment (http://www.eumetsat.int). This documentation is provided by the SAF on Support to Nowcasting and Very Short Range Forecasting, hereafter NWC SAF. The main objective of NWC SAF is to provide, further develop and maintain software packages to be used for Nowcasting applications of operational meteorological satellite data by National Meteorological Services. More information can be found at the NWC SAF webpage, http://www.nwcsaf.org. This document is applicable to the NWC SAF processing package for METEOSAT satellites meteorological satellites, SAFNWC/MSG.

The purpose of this document is to present the Scientific Validation Results for the PGE13 SEVIRI Physical Retrieval (SPhR) product and to show the compliance assessment of the PGE13 SPhR product against the requirement fixed in the Product Requirement Document **[AD.9]**.

The process to calculate and validate the NWC SAF SEVIRI brightness temperature bias correction is shown in Section 3. It has been calculated from the PGE13 SPhR validation dataset by using time and spatial collocated actual SEVIRI and synthetic RTTOV brightness temperatures.

In order to make more readable the report and to avoid some issues (as cloud contaminated pixels, emissivity issue, etc.), the scientific validation for 2012 version outputs shown in Section 4 has been mainly based on the validation of the PGE13 SPhR parameters using as input synthetic RTTOV SEVIRI brightness temperature. In order to avoid repeating the number of figures and tables, only the figures and tables considering as ground truth the parameters calculated directly from ECMWF analysis are shown in this report.

1.2 SOFTWARE VERSION IDENTIFICATION

The validation results presented in this document apply to the PGE13 SPhR SEVIRI Physical Retrieval Product (SPhR) v1.2 product included in the NWCSAF/MSG v2012 software package.

1.3 GLOSSARY

Please refer to the "Nowcasting SAF Glossary" [AD.8] for a glossary and a complete list of acronyms for the NWC SAF project.

1.4 REFERENCES

1.4.1 NWC SAF Applicable Documents

Ref.	Title	Code	Vers	Date
[AD.1]	Software User Manual for the SAFNWC/MSG Application: Software Part	SAF/NWC/CDOP/INM/SW/SUM/2	6.0	15/12/11
[AD.2]	Product User Manual for "PGE13 SEVIRI Physical Retrieval" (SPhR– PGE13 v1.2)	SAF/NWC/CDOP/INM/SCI/PUM/13	1.2	15/02/12
[AD.3]	Algorithm Theoretical Basis Document for PGE13 "SEVIRI Physical Retrieval Product" (SPhR) v1.2	SAF/NWC/CDOP/INM/SCI/ATBD/13	1.2	15/02/12
[AD.4]	Interface Control Document for the External and Internal Interfaces of the SAF NWC/MSG	SAF/NWC/CDOP/INM/SW/ICD/1	6.0	15/12/11
[AD.5]	SAFNWC/MSG Output Products format definition	SAF/NWC/CDOP/INM/SW/ICD/3	6.0	15/12/11
[AD.6]	Architectural Design Document for the INM related PGEs of the SAFNWC/MSG	SAF/NWC/CDOP/INM/SW/AD/4	6.0	15/12/11
[AD.7]	Software Version Description Document for the SAFNWC/MSG Application	SAF/NWC/CDOP/INM/SW/SVD/5	6.0	15/12/11
[AD.8]	The Nowcasting SAF Glossary	SAF/NWC/IOP/INM/MGT/GLO	1.3	10/11/09
[AD.9]	NWCSAF Product Requirements Document	SAF/NWC/CDOP/INM/MGT/PRD	1.0	28/07/09

Table 1: List of Applicable Documents.

1.4.2 External Reference Documents

Ref.	Title
[RD.1]	
[RD.2]	

Table 2: List of Referenced Documents

2. DATA FILES USED

In order to build the PGE13 SPhR datasets for validation purposes, data from MSG SEVIRI images, ECMWF GRIB files and radiosonde data have been used. At the time of writing this report the input data files used in the tuning and validation activities are:

From ECMWF outputs:

- 00 Z and 12 Z runs
- analysis (T+00 hours) and forecasts (T+12 hours)
- region: NW corner at (65º N, 65º W) and SE corner at (65º S, 65º E)
- time period: from 31 December 2007 12 Z to 30 September 2011 12 Z
- horizontal resolution: 0.5º by 0.5º
- vertical resolution: two different vertical resolutions are used
  - 15 fixed pressure levels at 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150 and 100 hPa (hereafter denoted as NWP15), and
  - 91 hybrid levels (hereafter denoted as NWP91H)
- parameters: temperature (T), humidity (relative humidity [RH] for NWP15 and specific humidity [q] for NWP91H).

Note: the NWP15 GRIB files are the ones used as input to the MSG NWCSAF software package.

Note: when the collocated records are written, the NWP(T+12) from previous 12 hour ECMWF run are collocated with the NWP(T+00). As example, the 01 January at 00UTC NPW(T+00) analysis profile is collocated with the NWP(T+12) from 31 December 12 UTC.

From MSG-2 SEVIRI Observations:

- 00 Z and 12 Z slots
- region: frame of 3400 x 3400 IR pixels centred at subsatellite position (only pixels with satellite zenith angle lower than 70º)
- time period: from 1 January 2008 00 Z to 30 September 2011 12 Z
- horizontal resolution: SEVIRI full resolution and MSG projection
- SEVIRI channels: All SEVIRI channels but HRVIS

These are the dynamic information datasets used for the tuning and validation activities. Specific datasets used for different objectives are in part generated from them and descriptions are provided in the respective sections.

3. SEVIRI BRIGHTNESS TEMPERATURES BIAS CORRECTION

3.1 IR SEVIRI BRIGHTNESS TEMPERATURES BIAS ESTIMATION DATASETS DESIGN

To get the infrared (IR) brightness temperatures (BTs) bias correction between SEVIRI BTs and synthetic RTTOV BTs, time and spatially collocated records are used. The PGE13 SPhR validation datasets have been generated since November 2007 and for near MSG full disk. Every record contains actual SEVIRI BTs and synthetic SEVIRI RTTOV BTs from ECMWF analysis (T+00).

Variables in every record: For every record added to the bias estimation dataset there is a structure with the following variables:

- *ancillary*: longitude, latitude, emissivities, etc.
- *date*: day, month, year, hour, minute.
- ***NWP from analysis (T+00):*** ECMWF temperature and humidity profile interpolated to the 43 RTTOV, T_{skin}, pressure at surface, etc.
- ***BT_SEVIRI(6)***: uncorrected bias BT from HRIT file. BT_SEVIRI[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$]
- ***BT_RTTOV(6)***: synthetic BT calculated using the RTTOV9.3 (see note bellow). BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$]

Note: for synthetic BT_RTTOV calculation a FORTRAN interface to RTTOV 9.3 outside of PGE13 SPhR is used.

Two kinds of datasets have been created: one built from NWP15 GRIB files and the other one built from NWP91H GRIB files. The reason for working in parallel with both is that results obtained just with NWP15 showed bigger biases than the ones published by GSICS and also found by Climate SAF. In order to make more readable, in this 2012 validation report only the results from NWP15 BT bias correction dataset are shown.

In order to avoid the inclusion of cloud contaminated pixels, the process to generate the collocated records uses the Cloud Mask (CMa-PGE01) and the PGE13-SPhR software configured for obtaining the required fields at the IR10.8 warmest clear pixel within a box (also denoted as Field of Regard [FOR]) from every 25 x 25 window inside the 3400 x 3400 region.

For the generation of the NWP15 derived dataset, an extensive use of the optional PGE13 SPhR capabilities to store interpolated (horizontal and vertical) intermediate profiles (keyword `SAVE_PROF` = 2) has been made (for details, see in Program User Manual for PGE13 SPhR [AD.2] document about how to activate the writing of intermediate temperature and humidity profiles on PGE13 SPhR binary files).

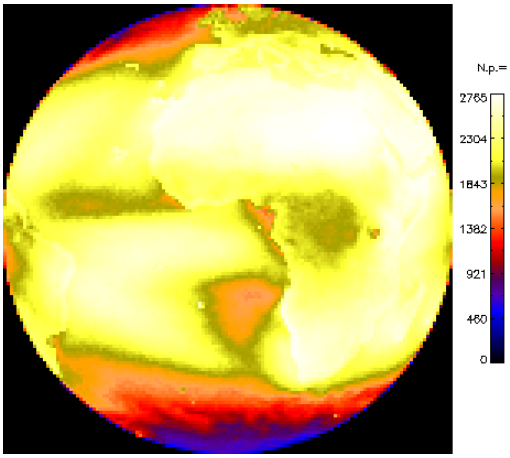

Figure 1: Spatial distribution of the number of observation for each 25x25 FOR included in the PGE13 BT bias correction in the period November 2007 to September 2011.

To facilitate the dataset management, information is grouped in files with a monthly basis. Later the monthly dataset can be joined for a desired wider period.

3.2 USE OF THE BRIGHTNESS TEMPERATURES BIAS ESTIMATION DATASET

3.2.1 Spatial distribution of the error between SEVIRI_BT and RTTOV_BT

The mean and standard deviation of the error between the vector of BT_SEVIRI and BT_RTTOV have been analysed. It can be seen in Figure 2 that the error over sea is lower. This is likely due to differences with actual skin temperatures in the ECMWF model skin temperature and to the emissivity issue over land pixels. Because of these reasons, land observations will not be used for bias correction estimation and NWC SAF SEVIRI **brightness temperature bias correction regression coefficients will be computed using only sea pixels**.

Note: to calculate error, the formula sqrt(total((bt_seviri-bt_rttov)^2)) has been used.

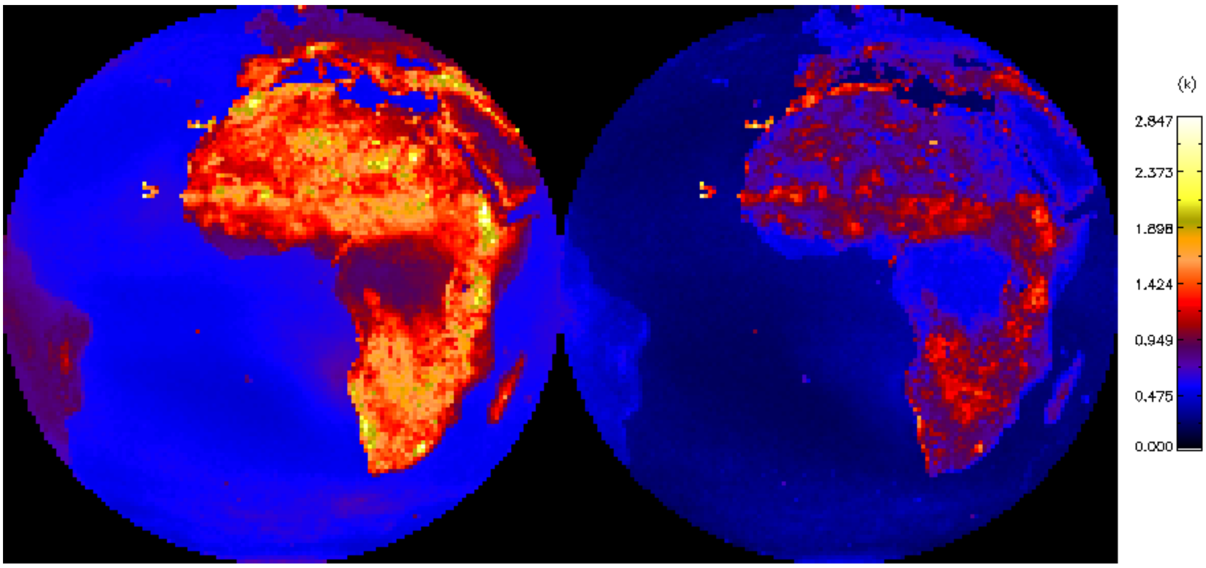

Figure 2. Left: mean error between BT_SEVIRI and BT_RTTOV. Right: standard deviation for each 25x25 FOR in the period 2007/11 to 2011/09.

3.2.2 Evolution with the time of the bias in different channels

Bellow, the process followed to determine the optimal period to calculate the default 2012 version PGE13 BT bias correction coefficients is described. The evolution of the bias between SEVIRI BT and synthetic RTTOV BT for several channels during the period from November 2007 to September 2011 has been calculated. The analysis of the evolution of the SEVIRI BT biases has been used to determine that the period April 2011 to September 2011 is the most adequate period to calculate the PGE13 BT bias correction coefficients.

For each SEVIRI IR channel and every day in the period from 1 November 2007 to 30 September 2011, all data from sea pixels for a "moving" window of 30 days in the 25x25 FOR full disk bias estimation dataset are used to calculate the "robust regression" for the BT bias correction over this 30 days window. In order to analyse the evolution of the BT bias on the different channels, the differences between a standard value (the mean of the all RTTOV BT for each channel in the period) with the result of applying the bias correction at this prefixed standard value have been represented in the Figure 3.

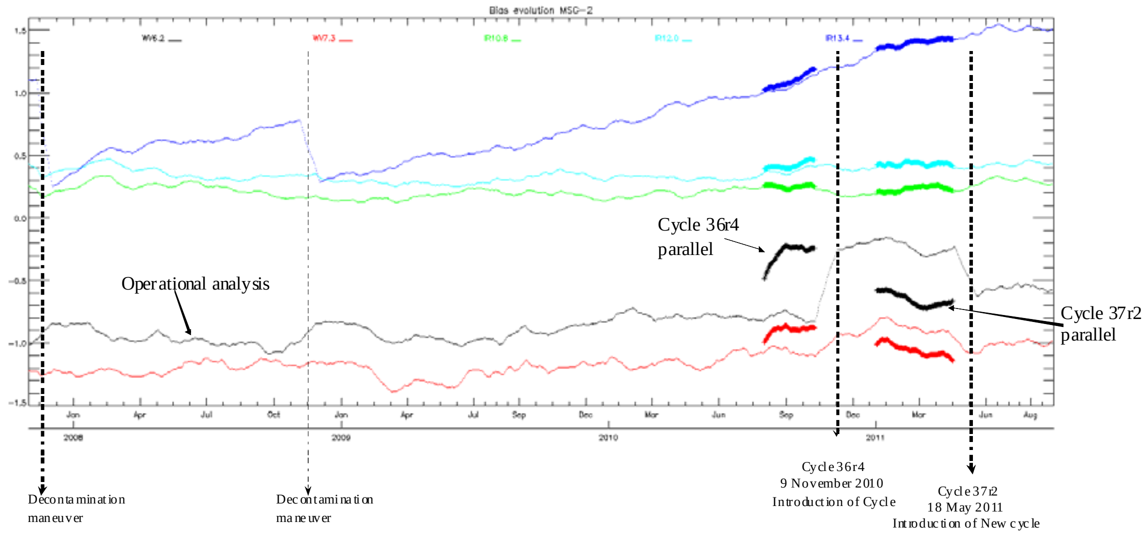

Figure 3: Evolution of the BT bias correction between BT_SEVIRI and synthetic BT_RTTOV. The differences between a mean value before and after the BT bias correction calculated for a "moving" window of one month for five SEVIRI channels.

It can be seen in the evolution of the BT bias of the IR13.4 channel (dark blue line) the deep steps in Dec 2008 and in Dec 2007. These are dates when decontamination of MSG-2 took place. To explain the continuous increase on the IR13.4 BT bias (dark blue line) EUMETSAT suggests that this behaviour is compatible with the accumulation of a light ice layer over the IR13.4 detector.

The period from November 2010 to May 2011 shows a different behaviour on the biases between SEVIRI BTs and ECMWF RTTOV BTs. Since November 2010, the main jumps in the evolution of the BT bias of WV channels (black and red lines) are due to the introduction of new cycles in the ECMWF. In order to assess this, the SEVIRI BTs biases calculated from GRIB analysis from the parallel runs have been also plotted on the Figure 3 (thick lines). These ECMWF parallel runs are made six months before the new cycles are introduced in operations. It can be seen in Figure 3 (thick lines) that parallel SEVIRI BTs biases anticipate the behaviour of the BT biases after introduction of the new cycle on the ECMWF.

As it can be seen in Figure 4, the evolution in the BT biases is similar to other BT bias estimation evolutions made by other authors as CM-SAF or EUMETSAT GSICS web page where an inter-calibration between collocated data of MSG and IASI is made. The URL with the EUMETSAT GSICS calibration is:

http://www.eumetsat.int/Home/Main/DataProducts/Calibration/Inter-calibration/GSICSBiasMeteosatIRInter-calibration/index.htm?l=en

By these reasons, a special PGE13 BT bias correction dataset has been created with the union of profiles from 1st April 2011 to 17th May from the parallel ECMWF analysis (MARS code v53) and from 18th May 2011 to 30th September 2011 from the operational ECMWF analysis. See (Martinez, 2011).

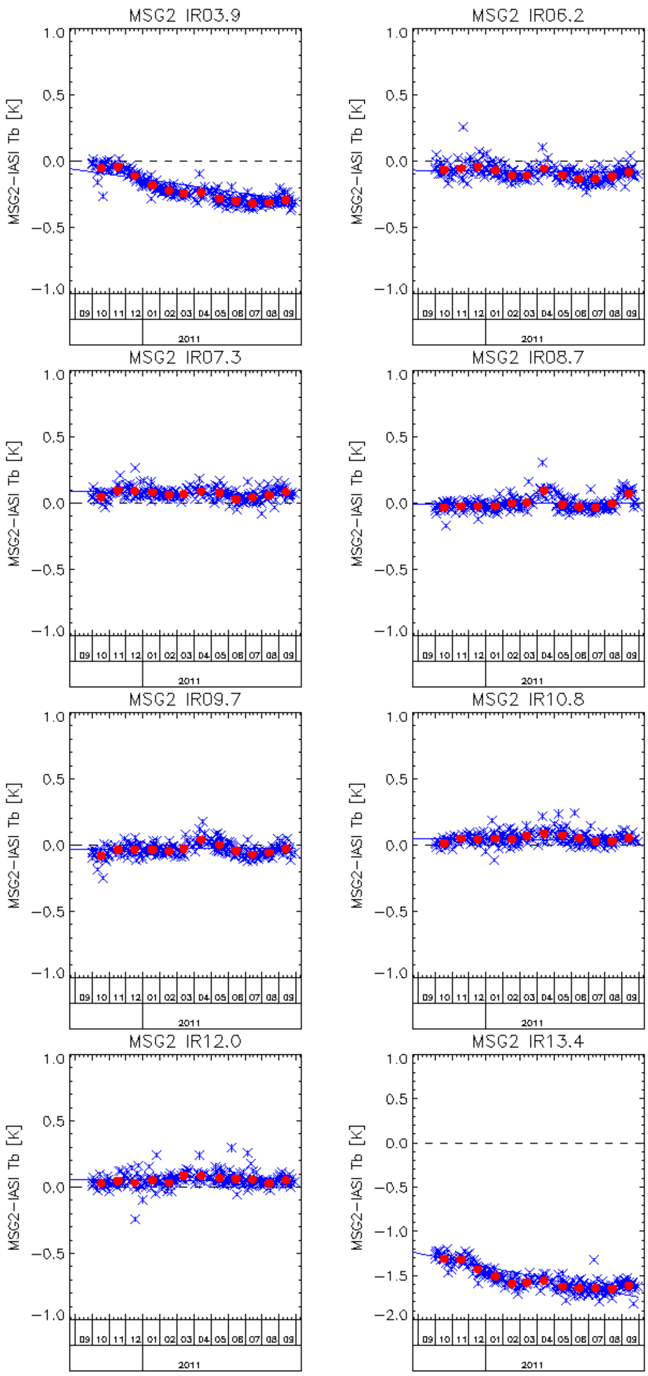

Figure 4: EUMETSAT GSICS Meteosat IR Inter-calibration with IASI snapshot: BT bias evolution from inter-comparison of equivalent infrared channels of geostationary SEVIRI and the polar-orbiting IASI sounder from collocated data.

3.3 SEVIRI BT BIAS CORRECTION COEFFICIENTS FOR PGE13 SPHR VERSION 2012 FROM NWP15 BIAS ESTIMATION DATASETS

Following the conclusions from the previous paragraphs and in order to calculate the default SEVIRI BTs bias correction coefficients for PGE13 SPhR version 2012, the period April 2011 to September 2011 has been chosen. A special PGE13 BT bias correction dataset has been created with the union of profiles from 1st April 2011 to 17th May from the parallel ECMWF analysis (MARS code v53) and from 18th May 2011 to 30th September 2011 from the operational ECMWF analysis on sea pixels.

All the observations that meet the following requirements on the PGE13 SPhR period 2011/04 to 2011/09 NWP15 dataset are used. These conditions are:

a) differences between actual SEVIRI BTs and synthetic RTTOV BTs, calculated as sqrt(BT_RTTOV -BT_SEVIRI), less than (mean+stdev),

b) sea pixels.

The number of observations used to calculate the BT bias corrections is larger that 2 million of pixels. The process followed here with the combination of the PGE01 Cloud Mask and the PGE13 SPhR programs is an example of the possibilities of the NWCSAF/MSG package to generate huge training and validation datasets.

Although PGE13 SPhR has been executed with the option 25x25 box processing and the warmest clear pixel in the IR10.8, because these options have the lowest probability to get contaminated by clouds pixels, there are some cloud contaminated pixels. By this reason the IDL command ***ladfit*** has been used to calculate the regression; ***ladfit*** uses "robust regression" (see references on the IDL ***ladfit*** help or in the Numerical Recipes book). In the Figure 5, the BT bias adjustment for IR13.4 SEVIRI channel is shown as example.

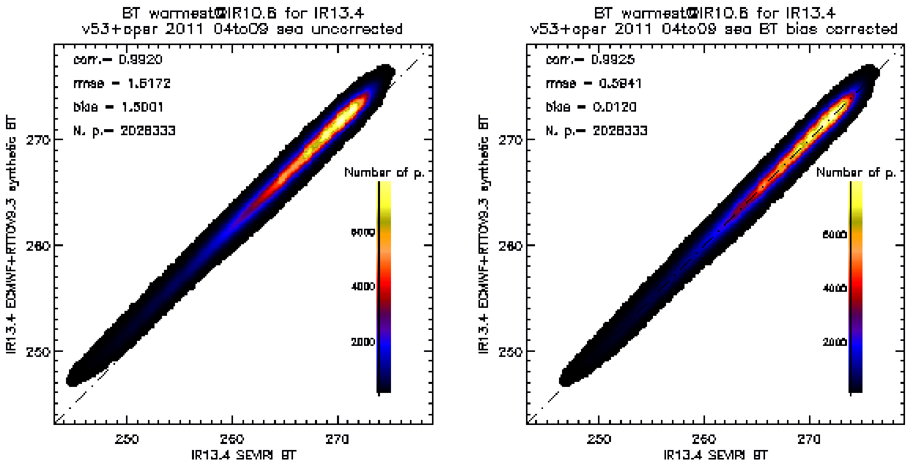

Figure 5: Scatter plots for IR13.4 SEVIRI BT at IR10.8 warmest clear pixel in 25 x 25 box SEVIRI BT versus ECMWF+RTTOV-9.3 synthetic BT over full disk region. On the left before BT bias correction, on the right after robust BT bias correction.

The SEVIRI BT bias correction regressions coefficients **have been written in the 2012 default PGE13 SPhR configuration file** "`$SAFNWC/config/`*`safnwc_pge13_msg2.cfm`*". For MSG-1 satellite, SEVIRI BT bias corrections regressions have also been calculated and written the 2012 version MSG-1 default configuration file ("`$SAFNWC/config/`*`safnwc_pge13_msg1.cfm`*").

The lines related to the SEVIRI BT bias correction are the ones with `BT_GLOBAL_SCALE_??` and `BT_GLOBAL_OFFSET_??` keywords (see the Program User Manual [AD.2] for one description of the configuration file).

In the case that the user doesn't want to apply any SEVIRI BT bias correction, the ASCII configuration file can be edited and fixing to 1 all the `BT_GLOBAL_SCALE_??` keywords and fixing to 0 all the `BT_GLOBAL_OFFSET_??` keywords.

4. VALIDATION OF PGE13 SPhR OUTPUTS

The main idea is to generate the whole dataset with the PGE13 SPhR software. Thus, all the profiles used in the validation have similar characteristics to the profiles used and retrieved with the PGE13 SPhR executed in the operational mode.

Spatial resolution of PGE13 SPhR dataset: In the case of PGE13 SPhR parameters validation, it has been designed a set of predefined positions of a 1º x 1º grid plus the RAOB stations positions are used. The set contains 13001 points in the actual mask. The positions where validation is made can be seen in Figure 6.

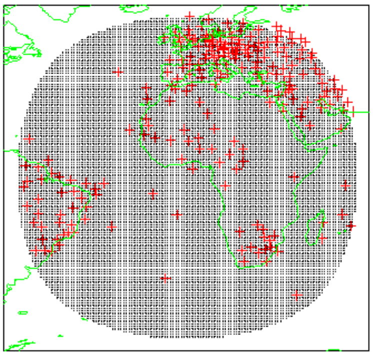

Figure 6: Predefined set of points used in the FG regression dataset. Grid network of 1º x 1º plus Radiosonde Stations (red crosses).

In order to speed-up the process the NWC SAF PGE01 cloud mask program is first executed and after reading the cloud mask, it is convolved with the mask of the 13001 positions (1 for validation points and 0 rest of pixels). The use of the convolved cloud mask speeds the process up because instead to execute the physical retrieval over several million of clear air pixels, this is only executed over the clear air pixels among the 13001 predefined positions.

Then, the PGE13 SPhR with option box size equal to 1x1 is executed twice and `BT_RMS_THRESHOLD` and `MAX_RESIDUAL` fixed to a very low value (0.001) in order to force always the three iterations in case of convergence. In the first PGE13 execution: the convolved cloud mask, the SEVIRI image and as background NWP the ECMW T+12 forecast from previous 12 hour ECMWF run (hereafter denoted as NWP15_12) are used as inputs. In the second PGE13 execution: the convolved cloud mask, the SEVIRI image and as background NWP the ECMW T+00 analysis (hereafter denoted as NWP15_00) are used as inputs. With the first execution it is possible to read the (T, q) profiles, from the *`$SAFNWC/tmp`* binary files, at the clear air predefined positions from T+12 hours ECMWF forecast from previous ECMWF run to the image and the (T, q) profiles at the different steps of the PGE13 algorithm (after FG regression step and after physical retrieval iterations). In the second execution it is possible to read the (T, q) profiles at the clear air predefined positions the (T, q) profiles from ECMWF T+00 analysis. The process can be seen in the Figure 7.

Variables in every record: Thus, for every observation added to dataset there is an IDL structure with some fields:

- *ancillary*: longitude, latitude, emissivities, etc.
- *date*: day, year, hour, etc.
- ***NWP from ECMWF forecast (T+12):*** ECMWF temperature and humidity profiles interpolated to the 43 RTTOV, T_{skin}, pressure at surface, etc. data from the previous run to the image ECMWF T+12 forecast (as example for image 20090101 at 00Z the T+12 forecast from 20081231 at 12 UTC ECMWF run is used).
- ***FG STEP:*** temperature and humidity profiles at the 43 levels and T_{skin}.

- ***1st Iteration STEP:*** temperature and humidity profiles at the 43 levels and T_{skin}.
- ***2nd Iteration STEP:*** temperature and humidity profiles at the 43 levels and T_{skin}.
- ***3rd Iteration STEP:*** temperature and humidity profiles at the 43 levels and T_{skin}.
- ***BT_SEVIRI(6)***: uncorrected BT from HRIT file. BT_SEVIRI[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].
- ***BT_RTTOV(6) from (T+12)***: synthetic BT calculated using the RTTOV9.3. BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].
- ***BT_RTTOV(6) at FG***: synthetic BT calculated using the RTTOV9.3; BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].
- ***BT_RTTOV(6) at 1st iteration***: synthetic BT calculated using the RTTOV9.3. BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].
- ***BT_RTTOV(6) at 2nd iteration***: synthetic BT calculated using the RTTOV9.3. BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].
- ***BT_RTTOV(6) at 3rd iteration***: synthetic BT calculated using the RTTOV9.3. BT_RTTOV[$_{WV6.2}$, $_{WV7.3}$, $_{IR8.7}$, $_{IR10.8}$, $_{IR12.0}$, $_{IR13.4}$].

To facilitate the dataset management, information is grouped in files on a monthly basis that can be joined easily for a wider period.

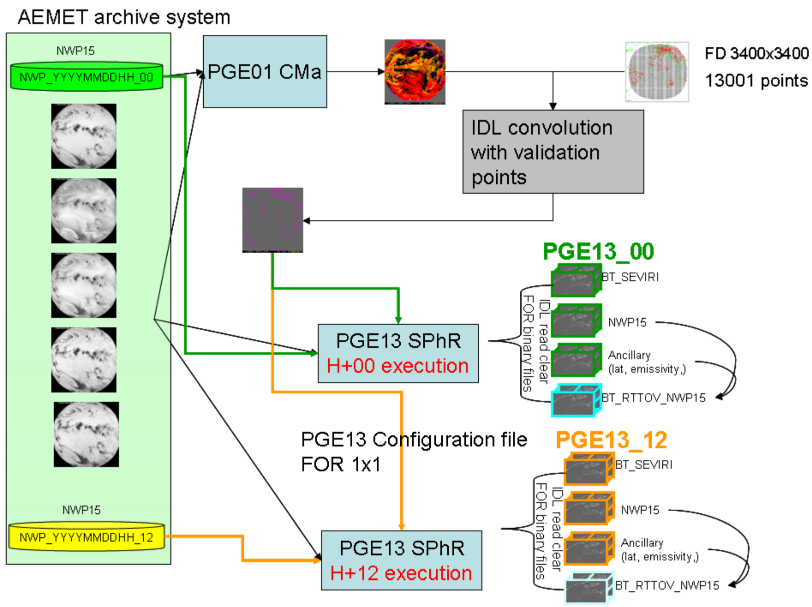

Figure 7: Creation of PGE13 validation dataset for one image.

In order to get the ground truth from the ECMWF analysis, IDL and FORTRAN programs have been developed to get the collocated profiles from the 91 hybrid levels (hereafter NWP91H) with every NWP15 profile. These programs read the ECMWF 91 hybrid levels GRIB profiles, make the interpolation to the 43 RTTOV pressure levels from the 91 hybrid levels and finally with the spatial interpolation to the NWP15 profile position.

In (Martinez, 2011), the main improvements of PGE13 2012 version and a brief comparison with the performance of PGE13 2011 version) are described. For the PGE13 2012 version the changes in the coefficients files are:

- ***New FG non-linear regression coefficients***: new noise figures have been added (greater than the 2011 ones) to the synthetic RTTOV BTs to calculate new FG regressions. In order to make less dependent of SEVIRI BTs the FG regression coefficients, the noise added in the calculation of new FG regression coefficients has been increased. With this increased noise synthetic SEVIRI RTTOV BTs dataset, the nonlinear regressions coefficients of the First-Guess step were calculated. Due to the PC memory constraints, only 1 out 2 records for period (even pixels) 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.

- ***New temperature and log(q) EOF files***: profiles after FG step have been used in calculation of new EOFs for temperature and log(q) profiles. Thus, these new EOFs files are based on 2009 ECMWF profiles instead on the original based on CIMSS datasets.
- ***New inverse of covariance matrix of the error*** $\mathbf{B}^{-1}$: the covariance matrix of the error between profiles after FG regression and ECMWF analysis profiles has been calculated from the 2009 PGE13 training dataset. Thus, the B^{-1} matrix is based on ECMWF profiles instead on the original based on CIMSS-Wisconsin datasets.

In order to a quick application of the new 2012 coefficients and to run physical retrieval over actual SEVIRI BTs or simulated RTTOV SEVIRI BTs, a environment has been developed. Thus, it is possible to test the new FG regression, EOFs and B^{-1} matrix avoiding the huge task to reprocess from the HRIT SEVIRI files and ECMWF GRIB files. This also allows to choose as input to the PGE13 algorithm SEVIRI BTs (bias corrected or uncorrected) or synthetic RTTOV SEVIRI BTs.

With the exception of Section 4.1, to avoid some issues (as SEVIRI BT biases, emissivity, contamination by clouds, need of screening in the selection of the records, etc.) and to make the document more readable, the assessment of the performances for these new PGE13 2012 version coefficients files is made here with synthetic RTTOV brightness temperature. It could be repeated with actual SEVIRI BTs in other report.

Since the 2009 year seems homogenous and it does not present the sharp changes in the profiles showed in the evolution of the BT bias after introduction of November 2010 cycle on the ECMWF (see Figure 3), the 2009 year has been chosen as the reference period for the PGE13 2012 version validation. The validation results obtained using as ground truth the ECMWF NWP91H analysis profiles are presented.

In order not to use for validation the same records used for calculation of the 2012 version coefficients, the records with odd position in the monthly datasets have been used.

Results presented are organised as follows. In Section 4.1, the analysis of the error between the SEVIRI BTs and the synthetic RTTOV BTs at the different steps of the algorithm is discussed. In section 4.2, an analysis of the vertical performance of the algorithm is made. In sections 4.3, the spatial validation results are presented.

4.1 ANALYSIS OF THE ALGORITHM CONVERGENCE BETWEEN THE SEVIRI BTS AND THE SYNTHETIC RTTOV BTS AT THE DIFFERENT STEPS OF THE ALGORITHM

In order to check the value add of the successive steps of the PGE13 SPhR algorithm, an inspection of the error between the bias corrected SEVIRI BTs and the synthetic RTTOV BTs at the different steps of the 2012 PGE13 SPhR algorithm is made first. This inspection has been divided in two parts. First, it is made the analysis of the histogram of the BT error and later it is made a spatial analysis of the BT error.

In order to check the error just over non window channels, the error between SEVIRI BTs and RTTOV BTs calculated from WV6.2, WV7.3 and IR13.4 (hereafter BT_RMS) has been also checked. The histograms with BT_RMS for synthetic RTTOV BTs case and actual SEVIRI BTs at the different steps of the PGE13 SPhR are shown in Figure 8.

It can be seen that the physical retrieval after the 1st iteration reduces significantly the BT_RMS (the peaks in the histograms present the lowest BT_RMS). Also, the FG step reduces slightly the BT_RMS. The BT_RMS has been checked because in the PGE13 code if the BT_RMS in the pixel is greater than a threshold, then the physical retrieval module is applied. The analysis of the histograms of error and BT_RMS can be used to select optimal values for the configurable parameters `BT_RMS_THRESHOLD` and `MAX_RESIDUAL` (these parameters are read from the ASCII PGE13 SPhR configuration file and an explanation of the impact in the selection in these parameters can be found in [AD.2]).

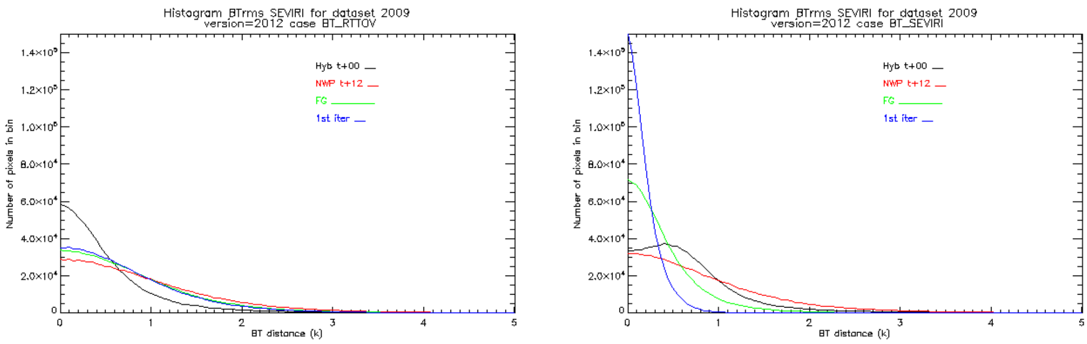

Figure 8: *Histogram of Error BT_RMS (distance in absorption channels) at different steps of the PGE13 SPhR over all 1 out of 2 pixels in period 2009/01 to 2009/12. Left) RTTOV reference case, right) PGE13 algorithm using as input actual SEVIRI case.*

In Figure 9, the spatial distribution of the mean distance between all SEVIRI channels and mean BT_RMS are shown for the case where actual BT bias corrected SEVIRI BTs are used as input to the PGE13 algorithm. It can be seen that the physical retrieval significantly reduces both. The regions with largest errors and residuals are located over land areas and especially over Sahara desert.

Figure 10 is similar to Figure 9 but in Figure 10 synthetic RTTOV BTs from NWP15(T+00) are used as input to the PGE13 algorithm. It can be seen that the 2012 PGE13 algorithm in Figure 10 has similar behaviour that in Figure 9 but without the regions with large errors that appear on the SEVIRI case.

The large residuals on the Figure 9 are associated to several issues such as emissivity fields and contamination with clouds. The errors between SEVIRI BT at window channels and the ones calculated from ECMWF analysis are due to skin temperature on ECMWF analysis does not represent the actual skin temperature over desert regions, mountains, etc. As conclusion of the analysis of the Figures 8 and 9, the physical retrieval algorithm implemented in the PGE13 SPhR works fine and the retrieved (T, q) profiles reduces significantly the error between the bias corrected SEVIRI BTs and the synthetic RTTOV BTs.

On the other hand, the analysis of PGE13 algorithm steps with synthetic RTTOV as input to the PGE13 can be representative of the performance of the PGE13 algorithm avoiding the issues on the previous paragraph. By these reasons and to avoid the need of screening filters, especially hard to make over land, the validation on the next sections is made only with the synthetic RTTOV BTs case.

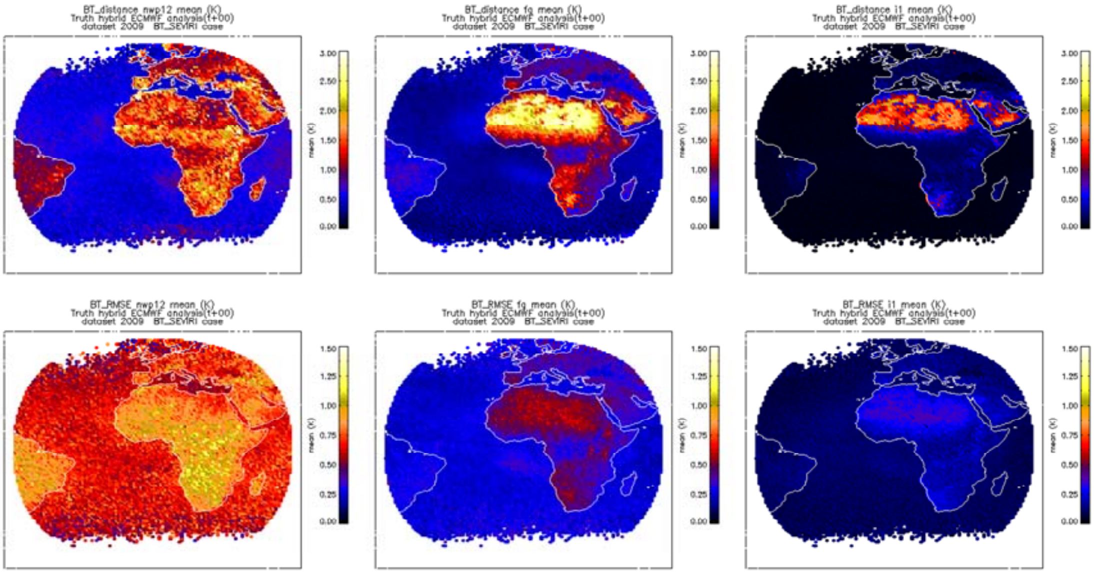

Figure 9: Spatial distribution of mean distance between synthetic RTTOV from ground truth (ECMWF analysis from NWP91H dataset) and BTs at different step of PGE13 algorithm when using as input bias corrected SEVIRI BTs. Top) BT_distance is the distance using all SEVIRI channels, bottom) BT_RMS is the distance using WV6.2, WV7.3 and IR13.4 channels. Left) error calculated vs RTTOV BT calculated at NWP15 (T+12) background NWP profiles, middle) error calculated vs RTTOV BT after FG step profiles and right) error calculated vs RTTOV BT after 1st iteration step profiles. Odd pixels in period 2009/01 to 2009/12.

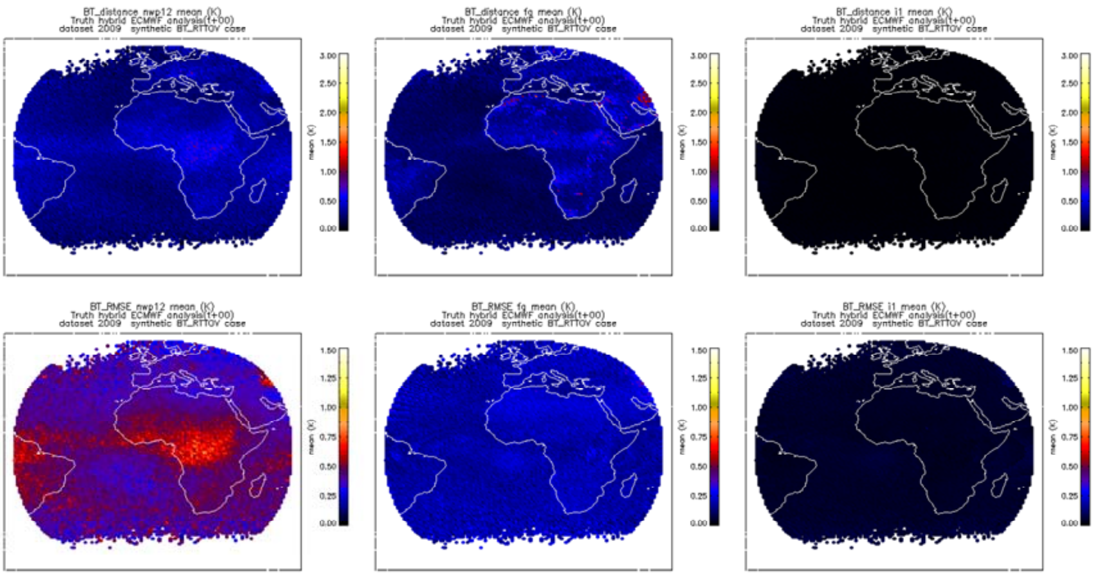

Figure 10: Same that Figure 9 but synthetic RTTOV BTs from NWP15(T+00) are used as input to the PGE13 algorithm.

4.2 ANALYSIS OF THE VERTICAL PERFORMANCE OF THE ALGORITHM AT THE FULL DISK DATASET

In the Figures 11 and 12, the RMSE and bias between the q profiles after several steps in the PGE13 SPhR algorithm for the Full Disk dataset at the 43 RTTOV levels have been represented.

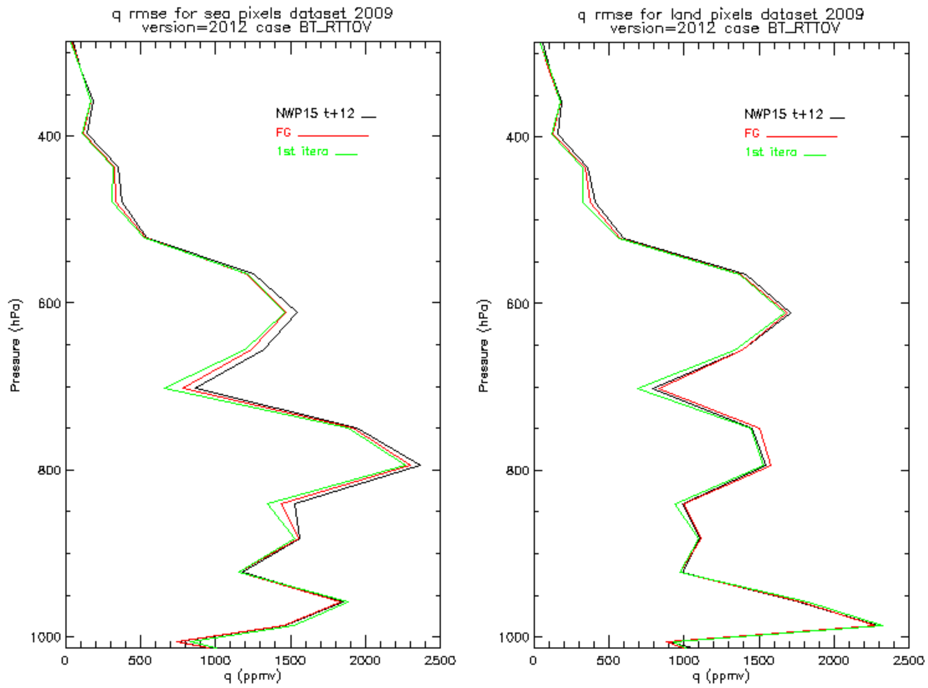

Figure 11: RMSE of q (ppmv) of q profiles at different steps compared with ECMWF analysis hybrid profiles in period January 2009 to December 2009 for odd pixels. Left) sea pixels, right) land pixels.

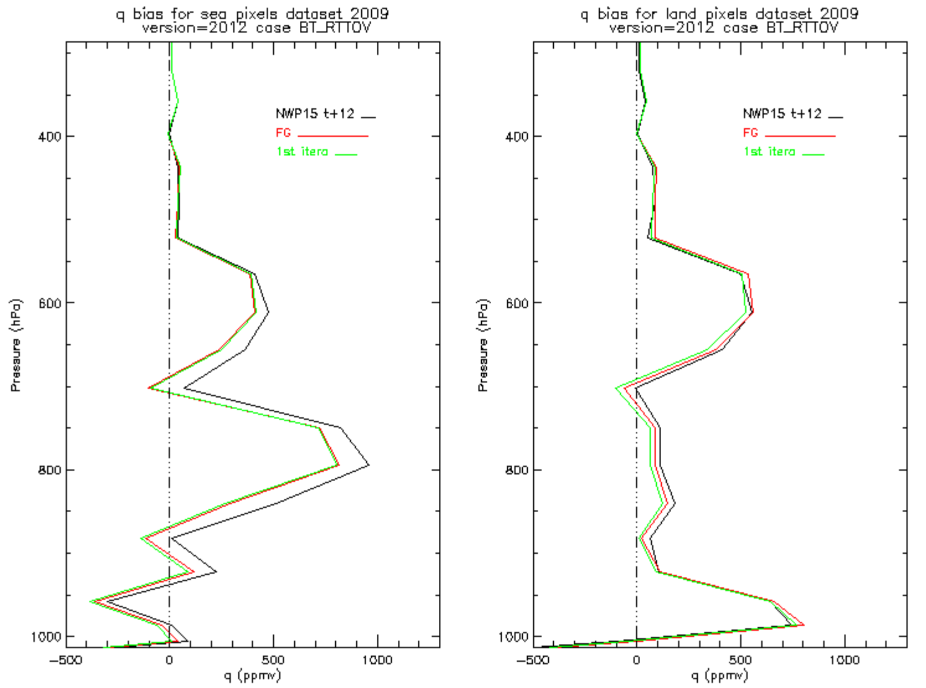

Figure 12: Bias of q(ppmv) of q profiles at different steps compared with ECMWF analysis hybrid profiles in period January 2009 to December 2009 for odd pixels. Left) sea pixels, right) land pixels.

The profiles of the ECMWF analysis (T+00) from NWP91H datasets (hybrid profiles interpolated to the 43 RTTOV level) have been considered as truth.

Only odd pixels have been used in vertical profile analysis here in order not to use the same ones used to calculate the 2012 PGE13 coefficients.

The statistical values for the specific humidity at mid levels show better performance for the FG regression and the physical retrieval than the background NWP model. Likely due to the value add of the WV SEVIRI channels, the reduction in the RMSE at these levels indicates that the PGE13 SPhR slightly improves the q background NWP profile.

The performance is better over sea pixels. The worse performance over land can be due to the poor vertical information of the background NWP used (only few pressure levels at low levels). The use of NWP15 profiles as input to the PGE13 SPhR creates several "peaks" and irregularities in the RMSE and bias vertical distribution centred at the 15 fixed pressure level. It can be seen in (Martinez, 2010) that when in a experiment similar 91 hybrid levels profiles (NWP91H) from T+12 are used as input to the PGE13 physical retrieval decrease the difference between land and sea pixels performance. 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, the 2012 NWC SAF/MSG package version will continue with the use of the 15 fixed pressure levels as background NWP input to PGE13. Due to this fact, in the validation dataset for the 2012 version, the temperature and specific humidity profiles at the 43 RTTOV pressure levels were the result of interpolating them from the 15 fixed pressure levels available in the ECMWF (NWP15 profiles).

> *Note: In the CDOP-2 phase it is foreseen the change in PGE13 to allow hybrid levels GRIB files as background NWP. But the change in the use of fixed pressure level GRIB files to hybrid level GRIB files in operational mode will need major changes in the NWC SAF library, in the PGE13 SPhR code and in the supply of the background NWP by the users.*

4.3 2D DIMENSIONAL HISTOGRAMS OF PGE13 SPHR PARAMETERS

To avoid multiplying the number of figures, only the two dimensional histograms for each one of the LPW and TPW parameters calculated from the profiles at different steps using as truth the calculated using the NWP91H (T+00) profiles are presented in Figure 13 for sea pixels and in Figure 14 for land pixels. The statistical values (RMSE, bias and correlation) that appear in the 2D histograms are also written in the tables of the Section 4.5 for a more comfortable read and comparison.

The PGE13 BL is the precipitable water in a layer between P_{sfc} to 850 hPa. The PGE13 ML is the precipitable water in a layer between 850 hPa to 500 hPa. The PGE13 HL is the precipitable water in a layer between 500 hPa to 0.1 hPa. The PGE13 TPW is the total precipitable water i.e the precipitable water in a layer between P_{sfc} to 0.1 hPa.

It can be seen in Figures 13 and 14 that statistical values of the PGE13 parameters reproduce the performance suggested by the vertical analysis from the figures 11 and 12. The parameters with the largest value added are ML and HL parameters; this fact is due to the WV channels.

Other important result is that the 2D histograms of the PGE13 SPhR parameters show no significant bias and it is not needed any correction in post processing.

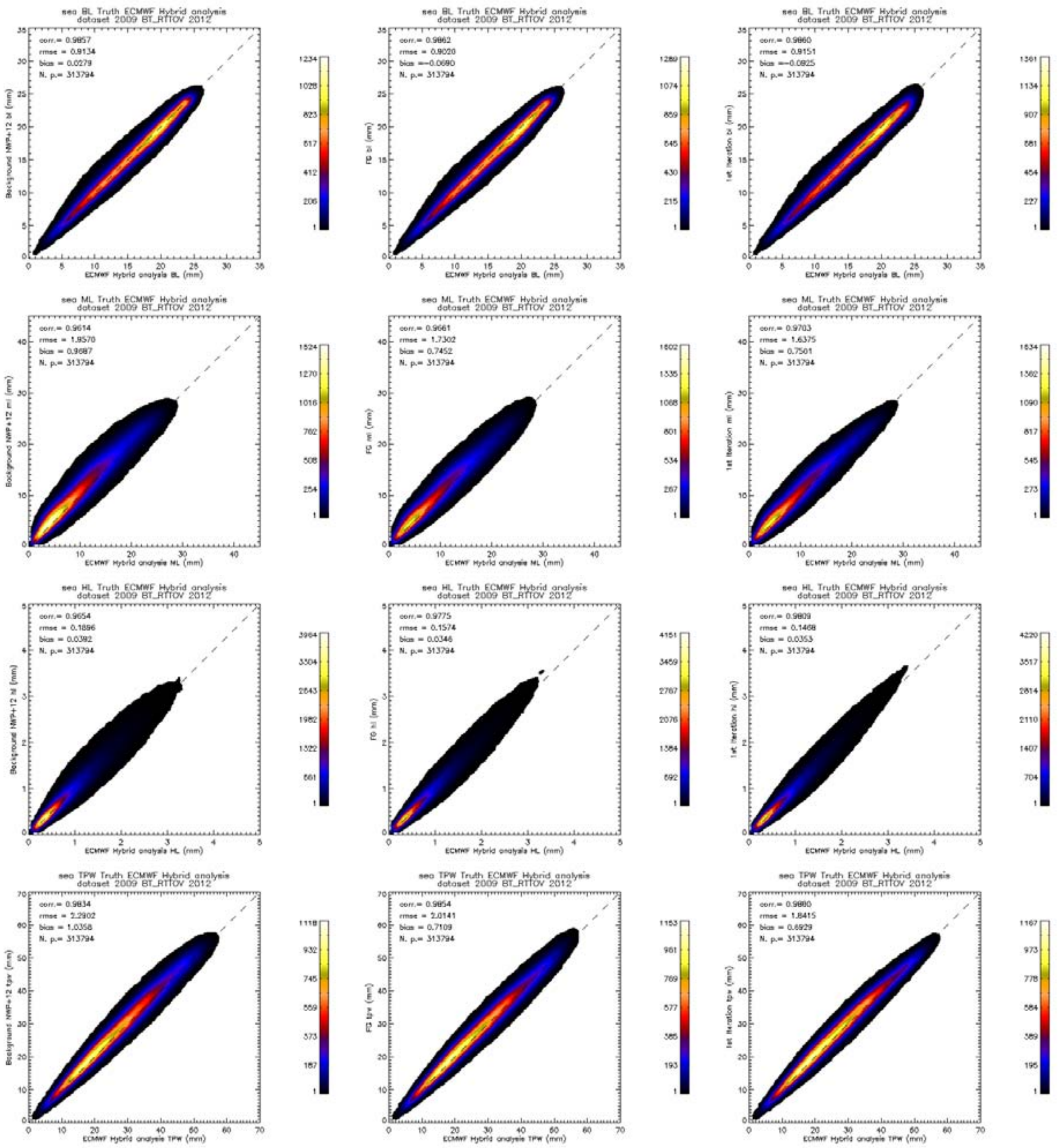

Figure 13: LPW and TPW 2D histograms over sea validation points in period January 2009 to December 2009 for odd pixels. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW parameters calculated after FG step profile, right) BL, ML, HL and TPW parameters calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles.

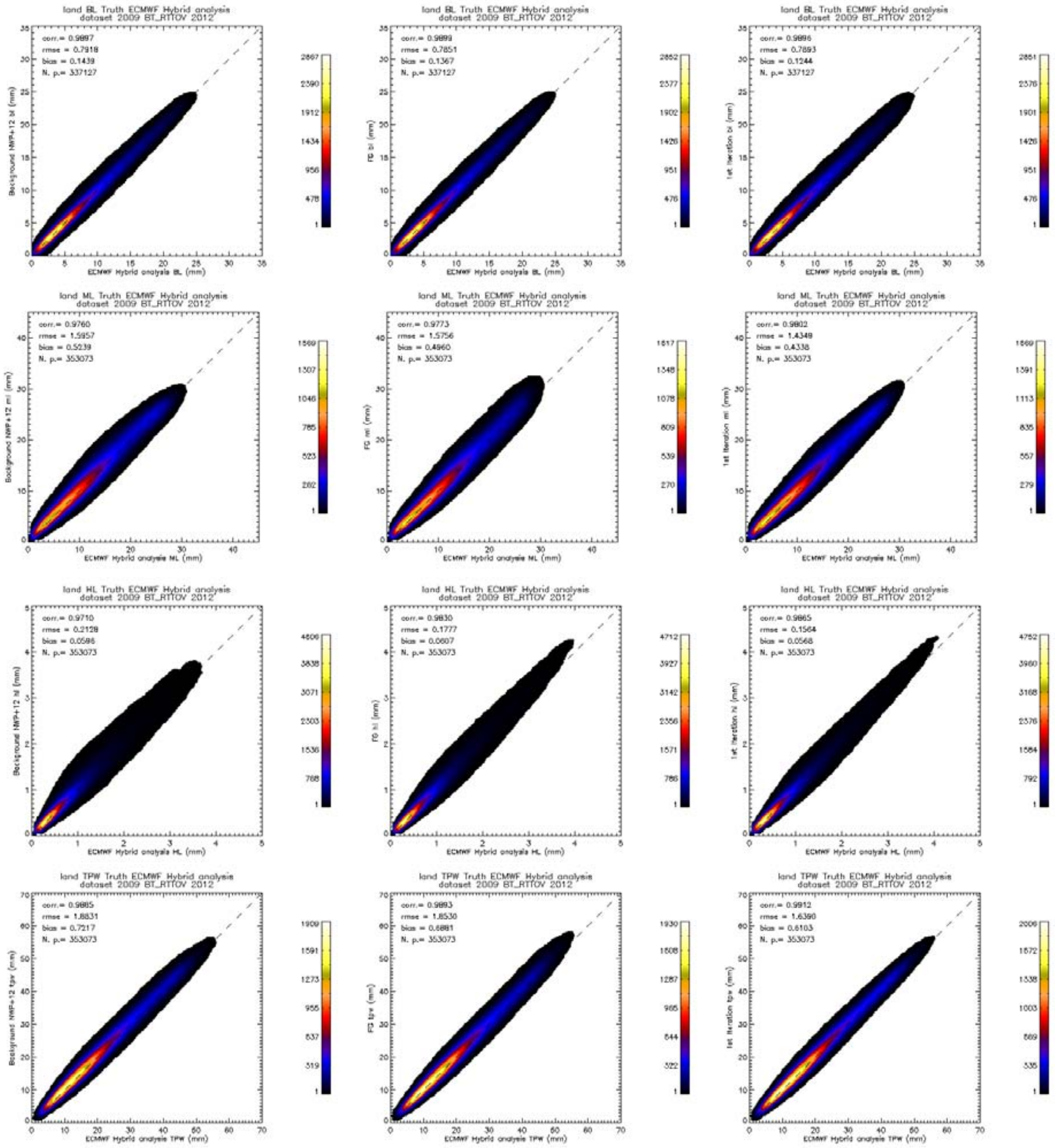

Figure 14: LPW and TPW 2D histograms over sea validation points in period January 2009 to December 2009 for odd pixels. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW parameters calculated after FG step profile, right) BL, ML, HL and TPW parameters calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles.

4.4 SPATIAL ANALYSIS OF PGE13 SPHR PARAMETERS

It can be seen in Figure 15 the spatial performance of the LPW and TPW parameters.

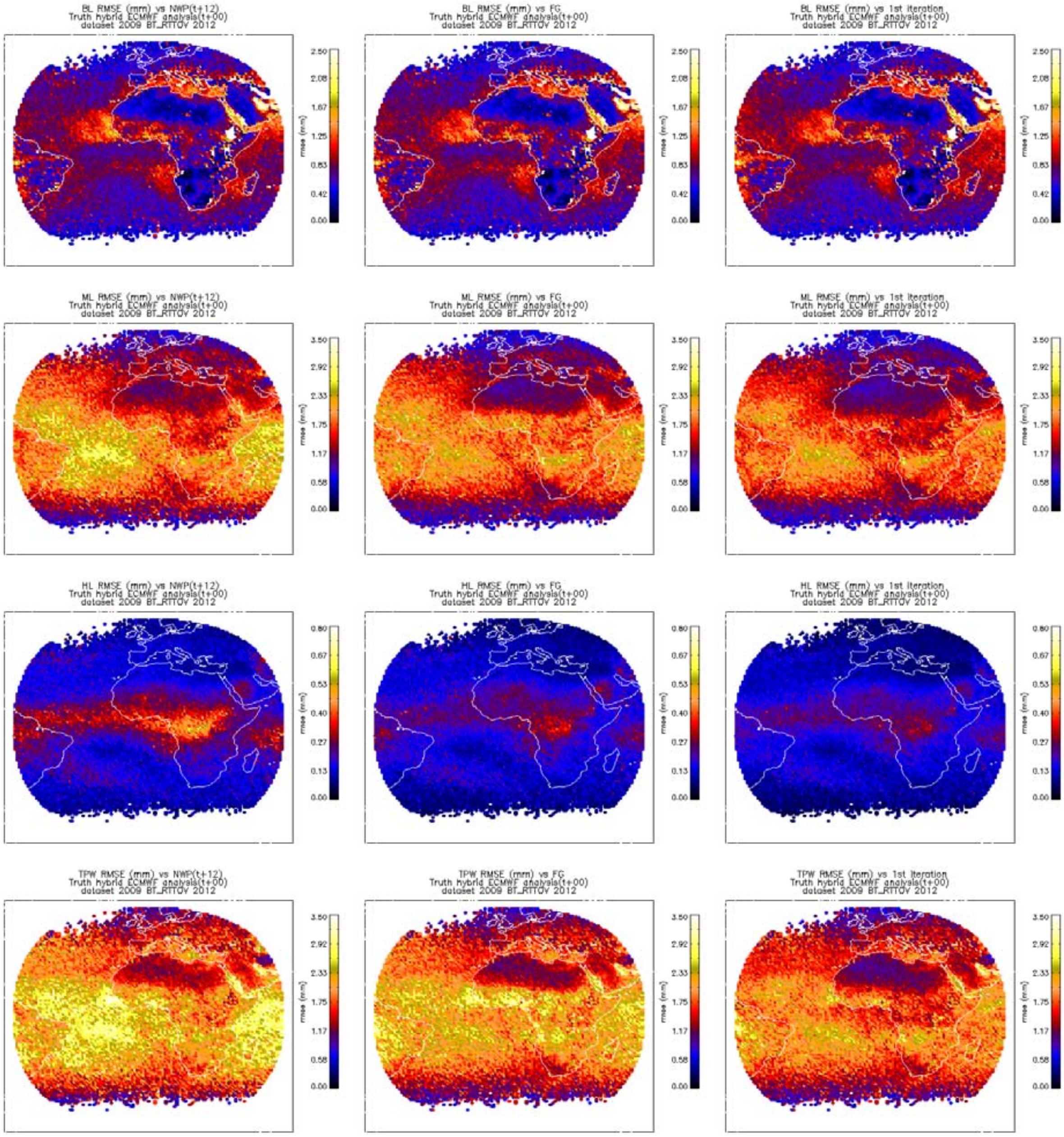

Figure 15: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in period January 2009 to December 2009 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background NWP15(T+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after 1st iteration step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP91H ECMWF analysis profiles.

The greatest values of ML and HL RMSE appear near to the equatorial belt. But, when the ML relative RMSE are calculated this effect disappears due to it had been caused by the high amount of precipitable water close to the equatorial belt. It can be seen this effect in the Figure 16.

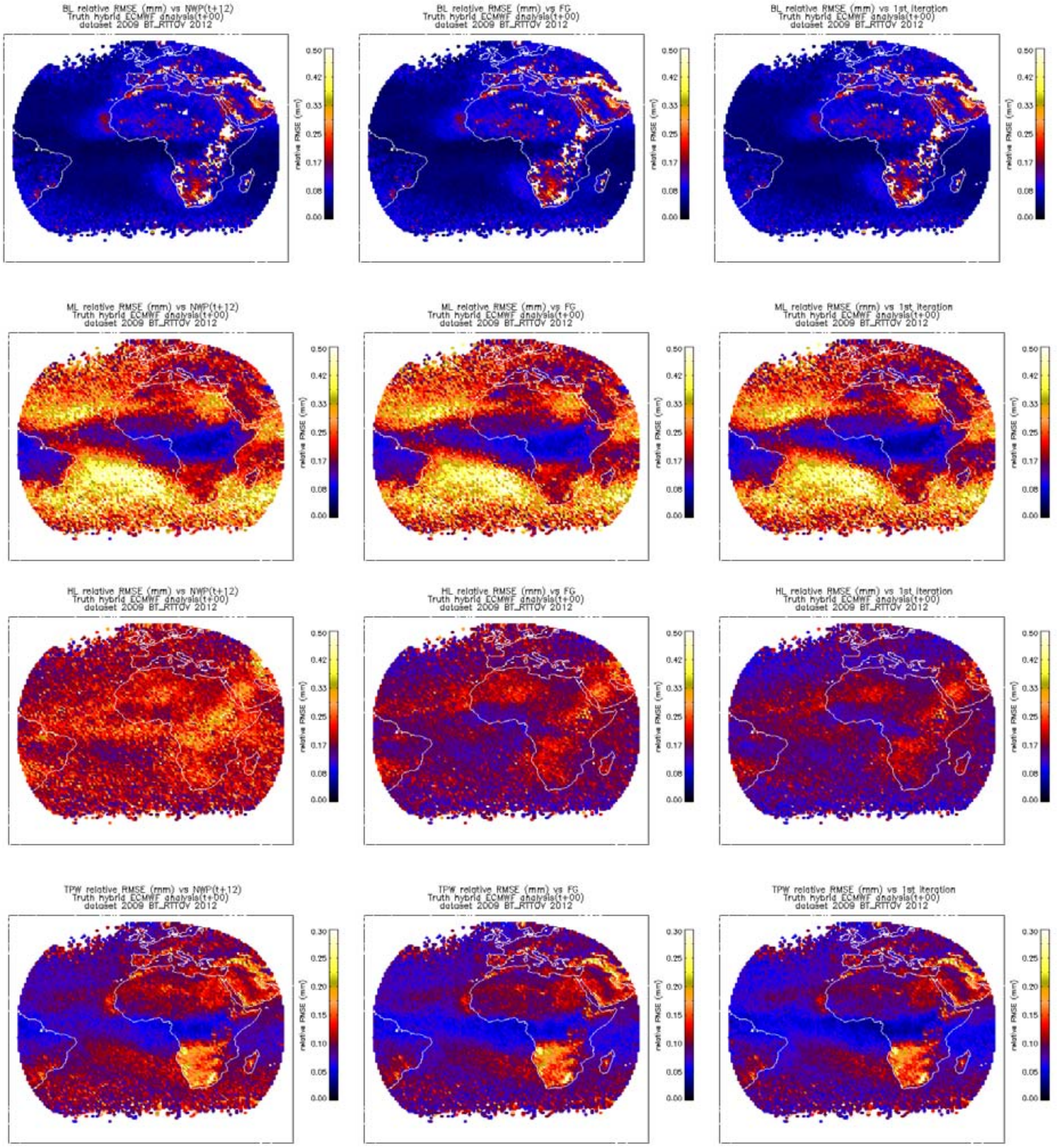

Figure 16: Same that Figure 15 but relative RMSE instead of RMSE.

4.5 VALIDATION AGAINST ECMWF: STATISTICAL SUMMARY

In order to allow a better comparison, the statistical values that appear inside the 2D histograms have been collected bellow in Table 3.

BL sea	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m^2)	0,913	0,902	0,915
BIAS (kg/m^2)	0,028	-0,069	-0,092
ML sea	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m2)	1,957	1,730	1,638
BIAS (kg/m2)	0,969	0,745	0,750
HL sea	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m2)	0,190	0,157	0,147
BIAS (kg/m2)	0,039	0,035	0,035
TPW sea	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m2)	2,290	2,014	1,842
BIAS (kg/m2)	1,036	0,711	0,693

BL land	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m^2)	0,792	0,785	0,789
BIAS (kg/m^2)	0,144	0,137	0,124
ML land	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m^2)	1,596	1,576	1,435
BIAS (kg/m2)	0,524	0,496	0,434
HL land	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m^2)	0,213	0,178	0,156
BIAS (kg/m2)	0,060	0,061	0,057
TPW land	NWP15(T+12)	FG	Phy. Retrieval
RMSE (kg/m^2)	1,883	1,853	1,639
BIAS (kg/m2)	0,722	0,688	0,610

Table 3: Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in period 2009 for odd pixels dataset. Left) sea pixels, right) land pixels.

ML parameter shows a significant reduction in RMSE. From the figures of Table 3, a reduction in ML RMSE of 12% for FG step and 16% after physical retrieval in sea pixels can be expected over sea pixels. The reduction of ML RMSE over land pixels is lower but it represents a 10% of reduction in the ML RMSE.

In the case of HL parameter the percentage in the reduction are even higher. In case HL RMSE reduction is around 21% over sea pixels and 26% over land pixels.

The former confirms the results of the vertical analysis of the performance made in Section 4.2 that showed a reduction in the RMSE and an improvement over the background NWP in the q profile at middle levels. The reduction of RMSE in the middle levels of the q profile is also likely the reason of reduction in the TPW RMSE.

In the case of PGE13 validation with actual SEVIRI BTs the performance has the same behaviour but with a higher figures and irregular distribution of the RMSE over the land pixels. These irregularities in the figures of the statistical value over land are due to the issues explained in Section 4.1. It is important to take into account that this is not always a negative aspect because it reflects the fact that actual SEVIRI BTs from the actual world are not the same that the synthetic and ideal RTTOV BTs.

In the case of the instability indices, not shown here, there is not a clear reduction in the RMSE. Likely, this is due to the fact that SEVIRI has limited information to improve the temperature vertical information beyond the forecast. But although the statistical validation is not much better, the PGE13 instability indexes have a great value added because SEVIRI provides useful spatial and temporal resolution. Thus, the PGE13 instability parameters are able to delimitate the region where instability is growing before convection triggering as it can be seen on the study case loops or in the near real time loops in http://www.nwcsaf.org.

5. CONCLUSIONS

After this validation, some conclusions can be obtained:

- Building a good training and validation dataset is a very important task for us. This continuous task has allowed the generation of a huge dataset of collocated SEVIRI BTs, ECMWF profiles from analysis and t+12 forecast, RAOB profiles, together with synthetic RTTOV BTs at several steps on the PGE13 algorithm. 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. In the 2012 version all the coefficients have been calculated from the PGE13 validation and training datasets (ECMWF profiles).
- The performance of PGE13 version 2012 is slightly better than the one for the PGE13 version 2011.
- Validation has been performed for the complete SEVIRI disk.
- Best results are obtained for humidity in medium layers due to the contribution of the two water vapor channel. In this layer the PGE13 SPhR improves the information beyond of the background NWP on the humidity profile.
- 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 and instability indexes.
- The RMSE of the PGE13 SPhR parameters are excellent and in all the parameters are better than the requested in the Product Requirement Document [AD.9] (see Table bellow)
- The validation should be repeated using as input to the PGE13 SPhR background NWP data with more vertical levels.
- The Figure 3 shows that a mechanism to calculate and distribute frequently an updated SEVIRI BT bias correction should be implemented for NWC SAF CDOP-2 phase. Thus, a web page set should be allocated, better inside the NWC SAF web, to provide frequent and rapid updates of the SEVIRI BT bias correction. A good time will be the launch of MSG-3.
- Validation has been performed for an extended period of a complete year – 2009. But there are available more years. During CDOP-2, a web page set should be created in order to maintain updated and detailed validation documentation, examples, repository of case studies, etc.

Next table summarizes the objective validation results in terms of RMSE; it summarizes the statistical values reported along the different sections of the document but the most important is that this figures represent a reduction in the RMSE from the background NWP greater than 20% for HL layer and 10% for ML layer.

PGE13 SPhR V1.2 summary of validation Results	Precipitable Water Low Layer – BL RMSE	Precipitable Water Medium Layer ML RMSE	Precipitable Water High Layer HL RMSE	Precipitable Water Total TPW RMSE	Lifted Index LI RMSE
Against ECMWF Analysis – Over Sea Full Disk validation	0.91 (kg/m^2)	1.64 (kg/m^2)	0.15 (kg/m^2)	1.84 (kg/m^2)	1.00 (K)
Against ECMWF Analysis – Over Land Full Disk validation	0.79 (kg/m^2)	1.43 (kg/m^2)	0.16 (kg/m^2)	1.64 (kg/m^2)	1.09 (K)

Table 4: Summary of the PGE13 SPhR statistical parameters in period January 2009 to December 2009 for odd pixels synthetic dataset.

6. REFERENCES

Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc., 43, 528–529.*

George, J. J., 1960: Weather Forecasting for Aeronautics. *Academic Press, 673 pp.*

Martínez M.A.; Li, J.; et al. 2008. Physical retrieval algorithm development for operational SEVIRI clear sky nowcasting products. *Proc. The 2008 EUMETSAT Meteorological Satellite Data User's Conference, Darmstadt, Germany.*

Martinez M.A.; Li, J., 2009. First operational results of the NWCSAF/MSG PGE13 SEVIRI Physical Retrieval algorithm. *Proc. The 2009EUMETSAT Meteorological Satellite Data User's Conference, Bath, United Kingdom.*

Martinez M.A.; Li, J.; Romero, R., 2010. NWCSAF/MSG PGE13 Physical Retrieval product version 2010. *Proc. The 2010 EUMETSAT Meteorological Satellite Data User's Conference, Córdoba,Spain.*

Martinez M.A.; Romero, R; Li, J., 2011. Status of NWCSAF/MSG PGE13 Physical Retrieval product. *Proc. The 2011 EUMETSAT Meteorological Satellite Data User's Conference, Oslo, Norway.*

Miller, R. C., 1972: Notes on Analysis of Severe-Storm Forecasting Procedures of the Air Force Global Weather Central. *AFGWC Tech. Rep. 200 (Rev.), Air Wea. Serv., Scott AFB, IL, 181 pp.*

NOAA, 1984: Convective stability indices. *Nat. Wea. Serv. West. Reg. Tech. Attach. No. 84-14, Salt Lake City, UT, 8 pp.*

Showalter, A. K., 1947: A stability index for forecasting thunderstorms. *Bull. Amer. Meteor. Soc., 34, 250–252.*