EUMETSAT NWC SAF

SUPPORT TO NOWCASTING AND VERY SHORT RANGE FORECASTING

Early adaptation of iSHAI v2016 to future MTG-I FCI using 2013 dataset

NWC/CDOP3/GEO/AEMET/SCI/RP/01, Issue 1, Rev.0

31 January 2020

Applicable to close backlog NWC-CDOP2-BL-06

Prepared by Agencia Estatal de Meteorología (AEMET)

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	Miguel Angel Martinez (AEMET)		*31 January 2020*
Reviewed by	Pilar Ripodas and Llorenç Lliso, AEMET		*31 January 2020*
Authorised by	Pilar Ripodas, AEMET (NWC SAF Project Manager)		*31 January 2020*

DOCUMENT CHANGE RECORD

Version	Date	Pages	Changes
1.0	*31 January 2020*	50	Report to close backlog NWC-CDOP2-BL-06.

Table of Contents

List of Tables and Figures

1. INTRODUCTION

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://nwc-saf.eumetsat.int. This document is applicable to the NWC SAF processing package for geostationary meteorological satellites, NWC/GEO.

1.1 Purpose and Scope of the Document

iSHAI (**i**maging **S**atellite **H**umidity **A**nd **I**nstability) is since version 2016 the name of the clear air product of the NWC/GEO software package.

GEO-iSHAI product is a heritage of former PGE13 SPhR (SEVIRI Physical Retrieval) product. The name of the product was changed to avoid the word SEVIRI because since release 2018 of the NWC/GEO software package it can be used with other GEO imager instruments different from MSG (AHI on board Himawari satellites and ABI on board GOES-R class satellites).

The name iSHAI was chosen to be coherent with names of NWC SAF products from MTG-FCI and MTG-IRS ("sSHAI) in CDOP-3 and future CDOP-4. Thus prefix "**i**" indicates generated with input data from one GEO **i**mager instrument and "**s**" prefix will indicate generated with input data from MTG-IRS **s**ounder instrument.

The main objective is to improve GEO-iSHAI algorithm and to be ready for the MTG-FCI era. The purpose of this document is to present a summary of the early activities made at the end of CDOP-2 phase and early CDOP-3 phase related to the adaptation of the iSHAI version 2016 software from MSG to other GEO satellites and to explore how to adapt to future MTG-I FCI data. This last activity was planned in the CDOP-2 working package WP4314E (NWC-033) "Studies and preparation of physical retrieval for MTG-I". As it is shown in this report the activity for this working package was made but due to lack of time the scientific report was not written till now.

Some of the results and Figures described in this Scientific Report were shown in April 2018 at the Convection Working Group Workshop 2018 that took place in Ljubljana; see Martinez presentation "*iSHAI and PGE00 as key tools in pre-convection*" with links (pdf) or (.pptx file including animations).

After the delivering of iSHAI software and documents for revision of the version 2018 patch 1 for the supporting of GOES-R class some available time has allowed to write this Scientific Report to close the backlog.

In version 2018.1 the supported GEO imager instruments are SEVIRI on board METEOSAT Second Generation satellites, AHI on board Himawari satellites and ABI instrument on board GOES-R class satellites (it will be distributed as a patch to the version 2018 at the end of 2019). Thus, GEO-iSHAI is updated to the state of art satellites and the GEO-iSHAI products could be compared with the products using the JMA and NOAA legacy algorithms.

In iSHAI version 2018.1 there are separated source directories with similar version of iSHAI code adapted for each imager instrument. Following this design, it was started the process to made a parallel copy of the process of generating and validating coefficients that could be used for MTG-FCI algorithm and to test iSHAI MTG-FCI version with test and synthetic data.

This Scientific Report with the 2013 dataset has been written to close the backlog NWC-CDOP2-BL-06 with the activities started at end of CDOP-2. As a continuation it will be repeated in next months the process using synthetic RTTOV MTG-FCI BTs from the 2017 iSHAI training dataset. The performance of MTG-FCI with the 2017 iSHAI training dataset will be publish in other Scientific Report.

1.2 DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Please refer to the "Nowcasting SAF Glossary" document in the NWC SAF web for a wider glossary and a complete list of acronyms for the NWC SAF project.

ABI	Advanced Baseline Imager
AEMET	Agencia Estatal de Meteorología Meteorology State Agency (Spain)
AHI	Advanced Himawari Imager
ASCII	American Standard Code for Information and Interchange
ATBD	Algorithm Theoretical Basis Document
BL	Precipitable water in low layer (P_{sfc} – 850 hPa)
BT	Brightness Temperature
CDOP-2	Continuous Development and Operations Phase 2
CDOP-3	Continuous Development and Operations Phase 3
CIMSS	Cooperative Institute for Meteorological Satellite Studies (USA)
CMa	Cloud Mask
ECMWF	European Centre for Medium-range Weather Forecasts
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCI	Flexible Combined Imager (MTG)
FG	First Guess
FOV	Field Of View
FOR	Field Of Regard
GEO	Geostationary Satellites
GEO-CMa	GEO Cloud Mask and Cloud Amount
GEO-iSHAI	GEO imaging Satellite Humidity And Instability
GRIB	Gridded Information in Binary Form
HL	Precipitable water in High Layer (500 – 0 hPa)
hPa	Hecto Pascal
HRIT	High Rate Image Transmission
IDL	Interactive Data Language
IR	InfraRed
IREMIS	InfraRed Emissivity

iSHAI	imaging Satellite Humidity And Instability
K	Kelvin
KI	K-Index
km	kilometre
LI	Lifted Index
LPW	Layer Precipitable Water
LST	Land Surface Temperature
MARS	ECMWF Meteorological Archive and Retrieval Facility
ML	Precipitable water in Medium Layer (850 – 500 hPa)
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
MTG-FCI	Meteosat Third Generation Flexible Combined Imager
MTG-IRS	Meteosat Third Generation Infrared Sounder
NRT	Near Real Time
NWC	Nowcasting
NWC/GEO	Geostationary part of the Nowcasting SAF
NWCLIB	Nowcasting Library
NWCSAF	Nowcasting SAF
NWP	Numerical Weather Prediction
NWP SAF	SAF for Numerical Weather Prediction
LPW	Layer Precipitable Water
PGE	Product Generation Element PGE01 Cloud Mask (GEO-CMa) Product Generator PGE13 SEVIRI Physical Retrieval (SPhR) Product Generator
PW	Precipitable Water
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVs
SAF	Satellite Application Facility
SEVIRI	Spinning Enhanced Visible Infrared Imager
SG	Steering Group
SHAI	Satellite Humidity And Instability

SHW	Showalter Index
SKT	Skin Temperature
SST	Sea Surface Temperature
TOZ	Total ozone
TPW	Total Precipitable Water
UM	User Manual
VR	Validation Report
WV	Water Vapour Channel

1.3 REFERENCES

1.3.1 NWC SAF Applicable Documents

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD.X]

For versioned references, subsequent amendments to, or revisions of, any of these publications do not apply. For unversioned references, the current edition of the document referred applies.

Current documentation can be found at the NWC SAF Helpdesk web: http://nwc-saf.eumetsat.int.

Ref.	Title	Code	Vers
[AD.1]	Proposal for the Second Continuous Development and Operations Phase (CDOP-2)	NWC/CDOP2/MGT/AEMET/PRO	1.0
[AD.2]	NWC SAF CDOP-2 Project Plan	NWC/CDOP2/SAF/AEMET/MGT/PP	1.0
[AD.3]	Request for Changes in the GEO Project Plan for the CDOP-2	NWC/CDOP2/GEO/AEMET/MGT/RP/02	1.0

Table 1: List of Applicable Documents

1.3.2 Reference Documents

The reference documents contain useful information related to the subject of the project. These reference documents complement the applicable ones, and can be looked up to enhance the information included in this document if it is desired. They are referenced in this document in the form [RD.X]

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies.

Ref.	Title	Code	Vers	Date
[RD.1]	Scientific and Validation Report for the iSHAI Processors of the NWC/GEO	NWC/CDOP3/GEO/AEMET/SCI/VR/iSHAI	1.0	21/01/19
[RD.2]	User Manual for iSHAI Product Processors of the NWC/GEO: Science Part	NWC/CDOP3/GEO/AEMET/SCI/UM/iSHAI	1.0	21/01/19
[RD.3]	Algorithm Theoretical Basis Document for iSHAI Product Processors of the NWC SAF/GEO	NWC/CDOP2/GEO/AEMET/SCI/ATBD/iSHAI	1.0	21/01/19
[RD.4]	Scientific Report: improvements in "PGE13 SEVIRI Physical Retrieval Product (SPhR)" using as input ECMWF GRIB files on hybrid levels	SAF/NWC/CDOP2/INM/SCI/RP/02	1.0	15/07/13
[RD.5]	Validation Report for "Clear Air Products"	SAF/NWC/CDOP2/INM/SCI/RP/02	1.0	15/10/16

Table 2: List of Referenced Documents

1 INTRODUCTION AND AIM OF THE STUDIES

In iSHAI some IR brightness temperatures measured by the supported GEO imager instruments on board meteorological satellites are used as the main input. The iSHAI algorithm is composed of a statistical retrieval followed by a physical retrieval of temperature and moisture profiles using infrared (IR) brightness temperatures (BTs) measured by GEO imager satellites and a background NWP forecast as main inputs. GEO-iSHAI algorithm remains based in Jun Li's algorithm for GOES Sounder physical retrieval algorithm (physical retrieval approach with non-linear regression as first guess) but adapted for RTTOV and other GEO imager instruments. Jun Li (CIMSS-Wisconsin) algorithm is also the one proposed as Day-1 algorithm for the GOES-R processing by NOAA (Li, 2010).

The iSHAI algorithm retrieves temperature and moisture profiles that are used to derive total precipitable water (TPW), layer precipitable water (LPW), Lifted Index (LI), Showalter Index (SHW), K-Index (KI), total ozone (TOZ) and the skin temperature (SKT) parameters from clear sky brightness temperatures within individual pixels or in a Field of Regard (FOR); a FOR is a region of M x M IR satellite pixels. Also the difference of these parameters with the calculated ones from the background NWP are output parameters.

In the iSHAI Algorithm Theoretical Basis Document (ATBD) the scientific algorithm, the needed input data and the resulting output are described in detail. It also provides basic information on the algorithm implementation; see [RD.3].

In CDOP-2 proposal [AD.1], the working package *WP4314E* was foreseen as one activity to do at the end of CDOP-2 phase (March 2012- March 2017).

In Table 3 it is shown the task proposed in this WP. It must be taken into account that due to the SAFs five year scheduling basis, at the moment when the proposal for CDOP-2 was written in 2011 the schedule was different and the launch of MTG-I has been delated several years and it has not been launched yet.

<table>
<tr><td rowspan="2">WP4314E (NWC-033)</td><td colspan="4">Title: GEO-TqPh-MTG-I renamed GEO-iSHAI-MTG-I</td></tr>
<tr><td colspan="4">Comments:</td></tr>
<tr><td>Start:
Mar-15</td><td>End:
Feb-17</td><td>Effort (mm):
2.00</td><td>Cost:
21.788 K€
(50%Eumetsat)</td><td>Responsible partner:
AEMET</td></tr>
<tr><td>WP Content</td><td colspan="4">Objective:
Studies and preparation of physical retrieval for MTG-I.

Tasks:
• Adaptation of physical retrieval software for MTG-I: calculations of coefficients with synthetic data.
• Checking with synthetic data and/or with proxy data of MTG-I software.</td></tr>
<tr><td>WP Input</td><td colspan="4">Pre-existing physical retrieval software and documentation.
NWC SAF library.
GEO physical retrieval training and validation dataset and tools.
MTG-I technical information.</td></tr>
<tr><td>WP Output</td><td colspan="4">Physical retrieval software and documentation prepared for MTG-I processing.
Tools ready for MTG-I tuning and validation by using the GEO physical retrieval datasets.</td></tr>
<tr><td>WP Interfaces</td><td colspan="4"></td></tr>
<tr><td colspan="2">Interactions with other SAFS and/or Federated Activities</td><td colspan="3">Coordination with CAF, NWP SAF</td></tr>
</table>

Table 3: Original formulation of WP4314E in CDOP-2 proposal in 2011.

When the CDOP-2 proposal was written it was foreseen to have an early GEO/NWCSAF software for testing just before the MTG-I launch on early CDOP-3 phase. Thus, the WP was introduced originally in order to create an iSHAI version for MTG-FCI in preparation of the launch of MTG-I.

At the end of the CDOP-2 phase, after delivering of iSHAI version 2016, it was initiated the early activities to explore how to adapt the iSHAI algorithm from SEVIRI instrument (on board of MSG class satellites) to the AHI instrument (on board of the satellites of the Japanese space agency HIMAWARI) and ABI instrument (on board the GOES-R class satellites from the NOAA) for iSHAI version 2018.1.

AHI and ABI instruments are two third generation instruments with 16 channels (visible and infrared channels), 2 km spatial resolution in the IR, 10 minutes temporal resolution and apparently with similar performance to the future MTG-FCI instrument.

But from the iSHAI product point of view, AHI and ABI instruments have a great improvement over the MTG-FCI instrument because they have ten channels in the infrared while SEVIRI and MTG-FCI only have eight channels in the infrared (wavelength greater than 4µm).

AHI and ABI instruments have three water vapour channels and three channels in the split-window region; instead of two water vapour channels and two channels in the split-window region in the SEVIRI and MTG-FCI instruments.

Therefore, the first to be taking into account when began to design the process of migration and generation of the iSHAI coefficients for 10 IR channels instruments is the differences in the number of channels. At the same time, since FCI will have similar channels to SEVIRI, a parallel version of the process but without the increase in the number of IR channels was started to repeat in parallel the developing for the future MTG-FCI.

It was made first the adaptation of the IDL program to calculate the First-Guess (FG) regression coefficients to support the AHI instrument on board HIMAWARI satellite. It was used as basis the IDL program to estimate the First-Guess (FG) regression coefficients using the SEVIRI instrument on board MSG satellites. This migration of the IDL code to AHI instrument included the change of the RTTOV call from SEVIRI to AHI, the increase in the size of arrays from 8 to 10, the generation of emissivity values for AHI instrument, etc.

Other initial task was the selection of the dataset for the calculation of the coefficients. There were two options:

- A) Several datasets for the full disk region of each imager instrument.
- B) Same common training dataset but using synthetic RTTOV BTs for calculation for each imager instrument

Since the main objective is to improve GEO-iSHAI algorithm and to be ready for the MTG-FCI era it was selected the option B. Thus, it was used as common training dataset the MSG 2013 iSHAI training dataset. This MSG 2013 iSHAI training dataset was built with selected pixels in the MSG full-disk and with synthetic RTTOV BTs for each imager instrument. To see the characteristic and how was built MSG 2013 iSHAI training dataset see Annex I. This option has as complementary advantage that as the NWP data are always the same and the radiative transfer model is the same with the only variant of the RTTOV coefficients depending on each one of the imager instrument. Then the performance differences are only due the differences in the satellites allowing a direct comparison and to assess the relative performance of iSHAI MTG-FCI with respect to the other GEO imager instruments.

Since dataset with real AHI data was not created, to calculate the AHI iSHAI coefficients it was used the 2013 iSHAI training dataset used for iSHAI MSG 2016 version but with array structures changed to 10 channels to allow store the synthetic RTTOV data BTs for AHI instrument. Once the migration of the program was finished, the coefficients for HIMAWARI were analysed and checked through the generation of the same figures used to write the version 2016 iSHAI Validation Report for MSG [RD.5].

An analysis of the differences between the two programs allowed to identify that the majority of the changes were due to the change in the name of the RTTOV coefficients, the change in the dimensions of the arrays size (eight to ten) and changes in the names of the input and output files to allow the identification of the files.

Therefore, once the AHI HIMAWARI version was available, the adaptation of the prototype IDL programs for GOES-R ABI was easily done repeating the changes made from MSG to HIMAWARI not related to modify the number of channels in a separate directory.

This pioneering experience of the generalization of the iSHAI algorithm from MSG to the rest of the imager-type geostationary instruments allowed to identify the main changes to be made in the algorithm. This identification of the differences allowed a rapid generalization of the process for generation of the FG regression coefficient files for the ABI instrument on board of GOES-R class satellites. The identification of the differences was made using the Linux *diff* and *kompare* commands; this allows to guarantee the needed changes traceability and that all from AHI to ABI was repeated correctly but from MSG to MTG-FCI.

Repeating the process of the adaption of the iSHAI IDL prototype from AHI to ABI but using as basis MSG, one early IDL prototype for iSHAI MTG-FCI was made. This iSHAI IDL prototype allowed to apply iSHAI algorithm to one array of pixels without the need to wait for the whole set of NWCSAF library, NWCSAF code and iSHAI code for one satellite.

The objective in this Scientific Report is to compare the validations with the 2013 iSHAI training dataset from the different imager instruments, which would ensure that all were performed correctly and the final objective to asses that the iSHAI algorithm from MTG-FCI will have the same or similar performances as the current one based on MSG-SEVIRI.

In this Scientific Report the Figures include in the title one identifier of the test for each one of GEO imager instrument:

- ✓ **FG_v2018** indicates that have been generated using as input MSG SEVIRI synthetic RTTOV BTs
- ✓ **FG_v2018_FCI** indicates that have been generated using as input MTG-I FCI synthetic RTTOV BTs
- ✓ **FG_v2018_ABI** indicates that have been generated using as input GOES-R ABI synthetic RTTOV BTs
- ✓ **FG_v2018_AHI** indicates that have been generated using as input Himawari AHI synthetic RTTOV BTs

After the experience with the migration of calculation of FG regression coefficients from MSG to other GEO imager instruments, instead to develop a single iSHAI software using an abstract satellite model with flow process routines depending on the satellite case, the iSHAI software has been designed to have different executable programs for each one of GEO imager instrument with adapted sources in different directories (with few and traceable modifications). Then, a UNIX script is written as the main user iSHAI command; this main iSHAI script execute the appropriate iSHAI program variant according to the active instrument read from the satellite configuration file (*sat_conf_file* in *$SAFNWC/config*).

This option will also allow in the future a separated evolution of the iSHAI MTG-FCI version of the ten-channel instruments AHI and ABI iSHAI versions. Thus, the AHI and ABI iSHAI versions could be used to explore the use of the additional IR channel and it will allow to incorporate and testing future developments made for USA of Japanese agencies or organisms. Also it will allow to simplify the stop of developing of the iSHAI MSG-SEVIRI version when MTG-I FCI be operational just simply retiring the directory with iSHAI MSG-SEVIRI version and there will no need the pruning of the iSHAI code.

2 ANALYSIS OF THE PERFORMANCE AT DIFFERENT VERTICAL LEVELS.

In Figures 1 to 4, the RMSEs profiles at the 54 RTTOV levels of the q profiles after FG regression step from several GEO imager instruments and the RMSEs profiles of the q profiles from ECMWF *t+12* forecast are shown. The profiles of the ECMWF analysis (*t+00*) from NWP-Hyb datasets (137 hybrid levels interpolated to the 54 RTTOV level) have been considered as the truth. It has been used the 2013 iSHAI training dataset. It have been split the data from sea and land pixels.

The statistical values for the specific humidity at mid-levels show better performance for the FG regression step than the background NWP model (ECMWF GRIB files on hybrid levels from *t+12* forecast). This is likely due to the added value of the imager WV channels, the reduction in the RMSE at these levels indicates that the GEO iSHAI slightly improves the q profile from background NWP.

Hereafter, in the figures the meaning of the labels is the following:

1. ***Hybrid t+12***: indicates that the value has been calculated using the comparison of ECMWF *t+12* forecast and the ECMWF analysis *t+00* as truth.
2. **FG**: indicates that the value has been calculated using comparison of the FG result (executing the non-linear regressions of FG step over the satellite synthetic BTs using ECMWF *t+12* forecast as background) and the ECMWF analysis *t+00* as truth.

It can be seen in Figures 1 to 4 that the performance in all GEO imager instruments is similar and that the performance of MTG-FCI (Figure 1) is not worse but similar to the MSG-SEVIRI one (Figure 2).

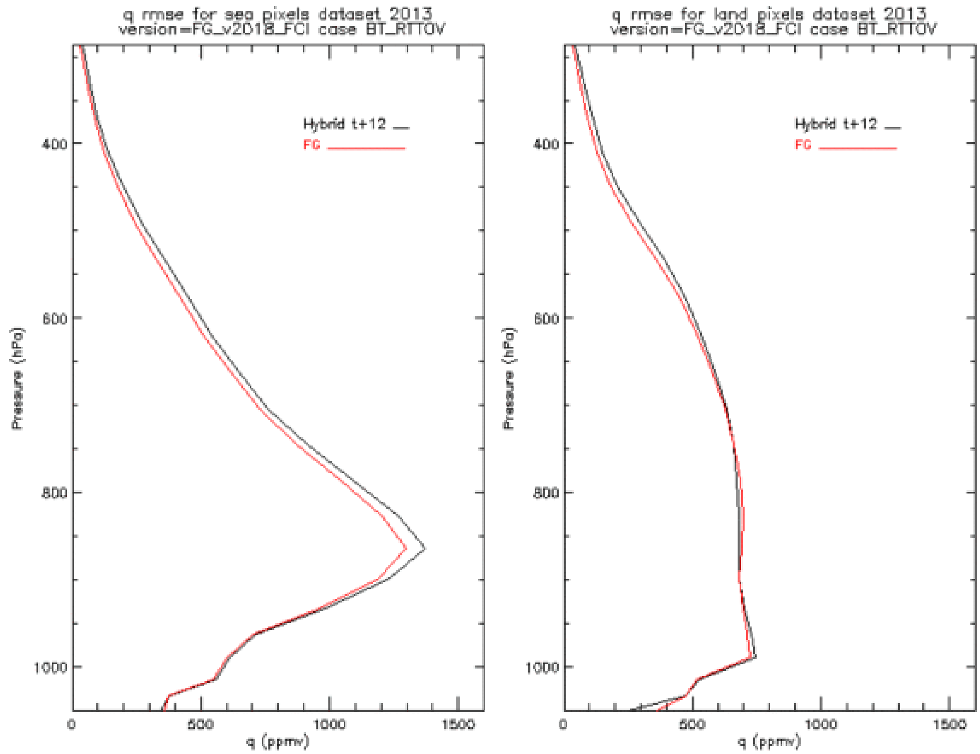

Figure 1: RMSE q profiles (ppmv) from ECMWF t+12 forecast and after FG linear regression step compared with ECMWF analysis (t+00) hybrid profiles ***for synthetic MTG-FCI BT_RTTOV****. Right) over sea pixels, left) over land pixels.*

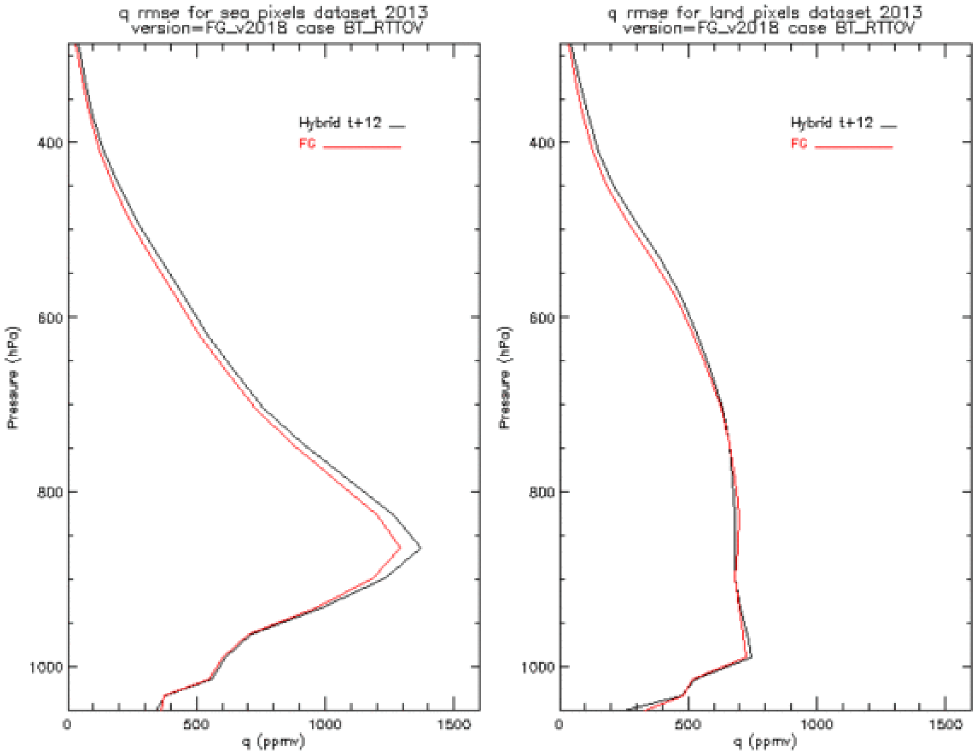

*Figure 2: RMSE q profiles (ppmv) from ECMWF t+12 forecast and after FG linear regression step compared with ECMWF analysis (t+00) hybrid profiles **for synthetic MSG-SEVIRI BT_RTTOV**. Right) over sea pixels, left) over land pixels.*

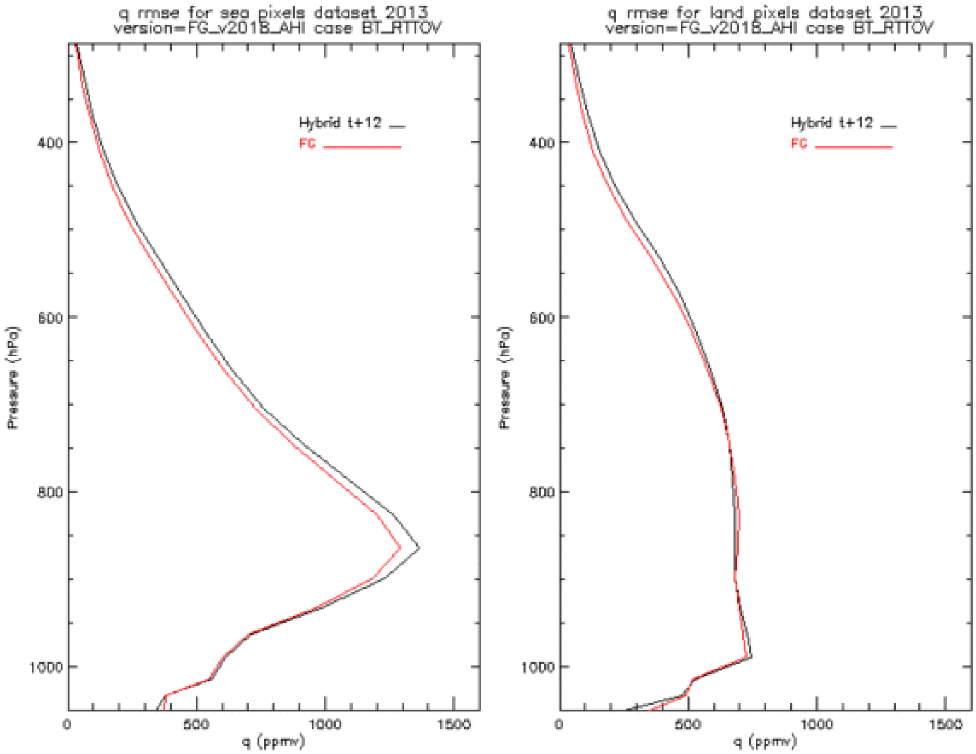

*Figure 3: RMSE q profiles (ppmv) from ECMWF t+12 forecast and after FG linear regression step compared with ECMWF analysis (t+00) hybrid profiles **for synthetic Himawari-AHI BT_RTTOV**. Right) over sea pixels, left) over land pixels.*

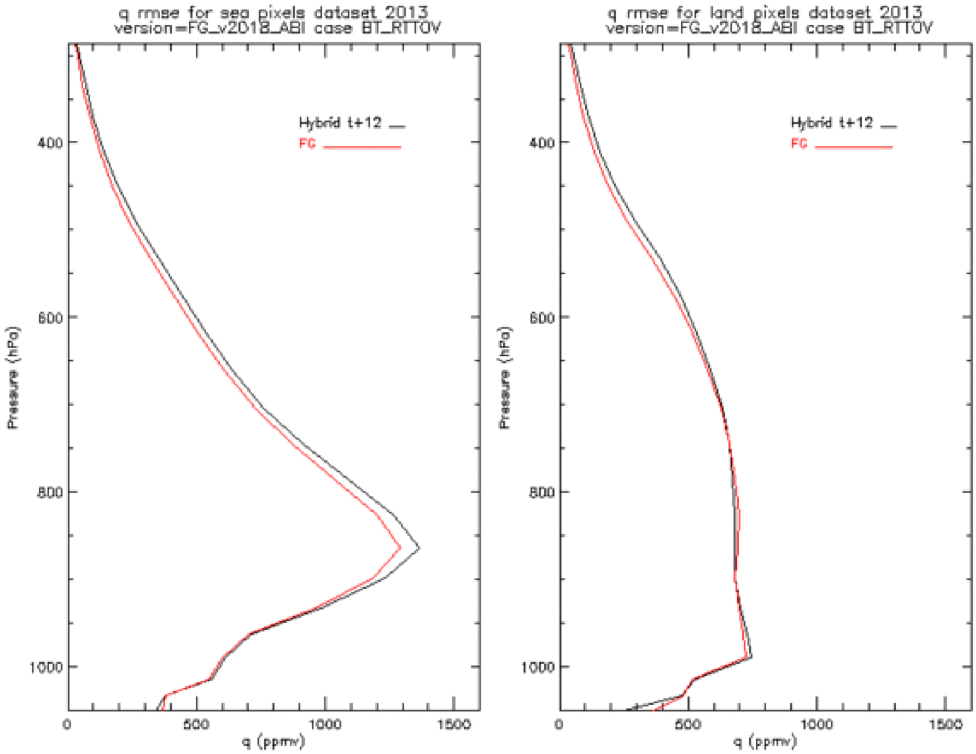

Figure 4: RMSE q profiles (ppmv) from ECMWF t+12 forecast and after FG linear regression step compared with ECMWF analysis (t+00) hybrid profiles ***for synthetic GOES-16-ABI BT_RTTOV****. Right) over sea pixels, left) over land pixels.*

The retrieved profiles are smoother than the analysis and forecasted ECMWF profiles. By this reason, when RMSE or other statistical parameters are calculated on every pressure level they could be slightly greater than the ones calculated from the t+12 forecast and analysis ECMWF profiles on some pressure levels; this artefact is more evident on pixels and levels with high spatial and vertical variability as in land pixels on low levels. This has been checked by comparison of the plots of retrieved profiles versus the analysis and t+12 forecast profiles. This artefact is the cause that in land pixels and in some levels (as near 800 hPa) the RMSE from the retrieved profiles is slightly greater to the t+12 forecast RMSE.

3 2D DIMENSIONAL HISTOGRAMS OF GEO ISHAI PARAMETERS

To avoid multiplying the number of Figures, only the two dimensional histograms for each one of the Layer Precipitable Water LPW parameters (BL: Bottom layer, ML: Mid Layer and HL: High Layer) and Total Precipitable Water TPW parameters calculated from the retrieved profiles from ECMWF t+12 forecast and after FG linear regression step are presented here.

Note that GEO-iSHAI BL is the precipitable water in a layer between P_{sfc} to 850 hPa. GEO-iSHAI ML is the precipitable water in a layer between 850 hPa to 500 hPa. GEO iSHAI HL is the precipitable water in a layer between 500 hPa to 0.1 hPa. GEO iSHAI TPW is the total precipitable water i.e the precipitable water in a layer between P_{sfc} to 0.1 hPa.

It has been used always as truth the ECMWF analysis denoted here as NWP-Hyb (*t+00*) profiles. In Figures 5 to 8 for sea pixels and in Figures 9 to 12 for land pixels for synthetic BT_RTTOV test for different GEO imager instruments.

The statistical values (RMSE, bias and correlation) that appear in the 2D histograms are also written at the end of this report in the tables of the Section 7 for a more comfortable read and comparison.

It can be seen in Figures 5 to 12 that statistical values of the GEO-iSHAI parameters reproduce the performance suggested by the vertical analysis from Figures 1 to 4. The parameters with the largest added value are ML and HL parameters; this fact is due to the contribution of the WV channels.

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

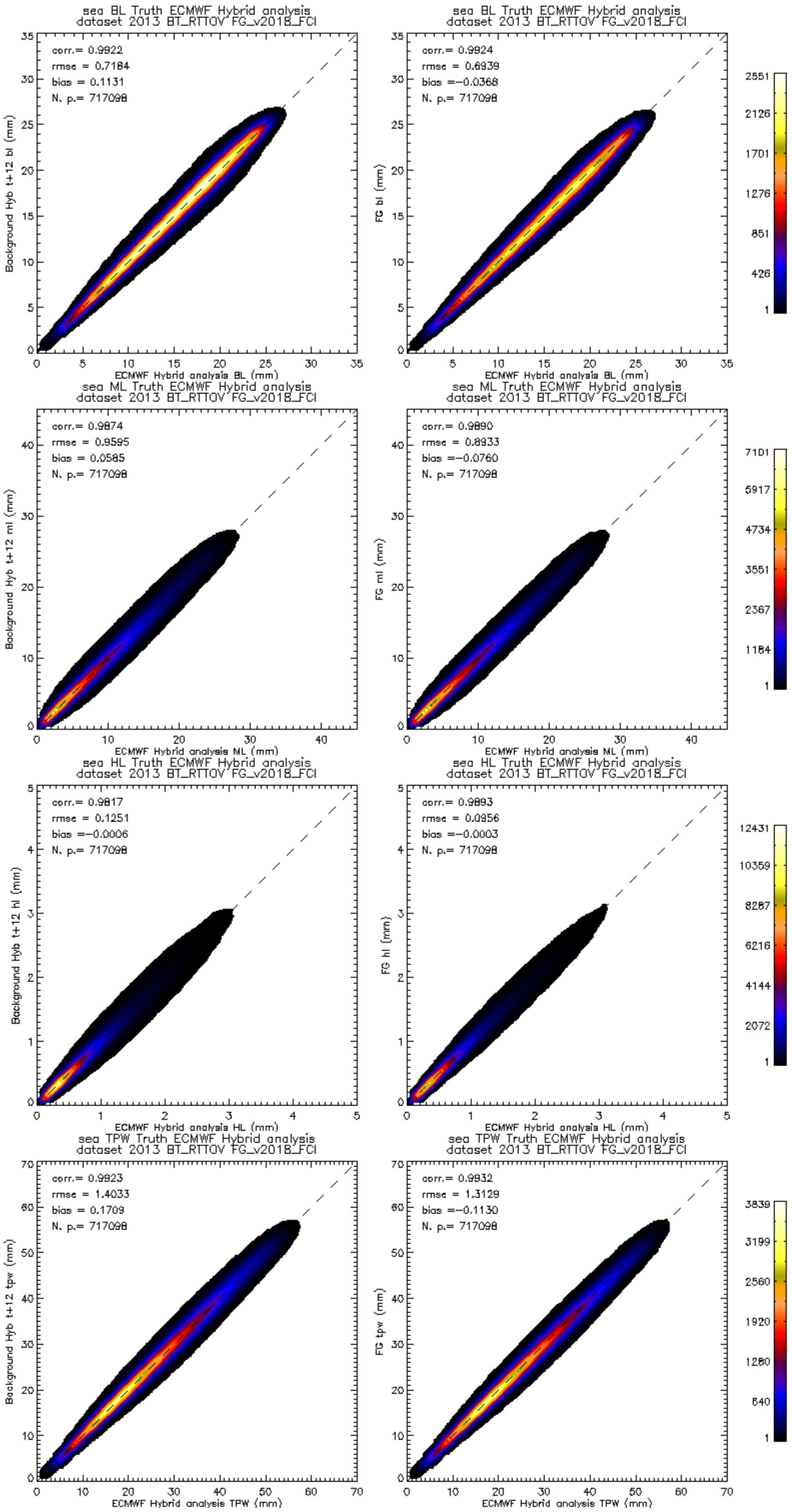

*Figure 5: **BT_RTTOV test** for synthetic MTG-FCI: LPW and TPW 2D histograms over **sea**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

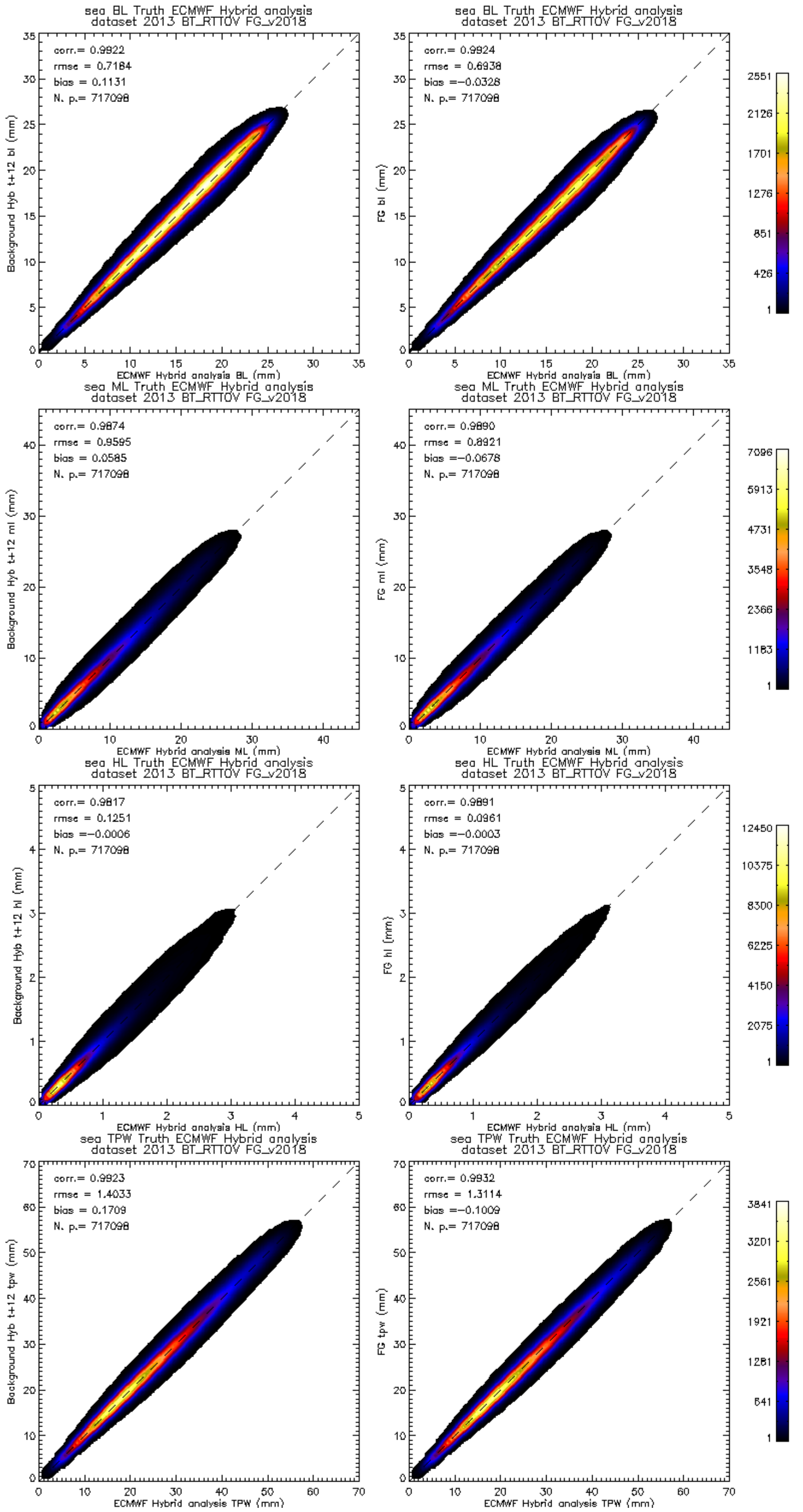

*Figure 6: **BT_RTTOV** test for synthetic MSG-SEVIRI: LPW and TPW 2D histograms over **sea**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

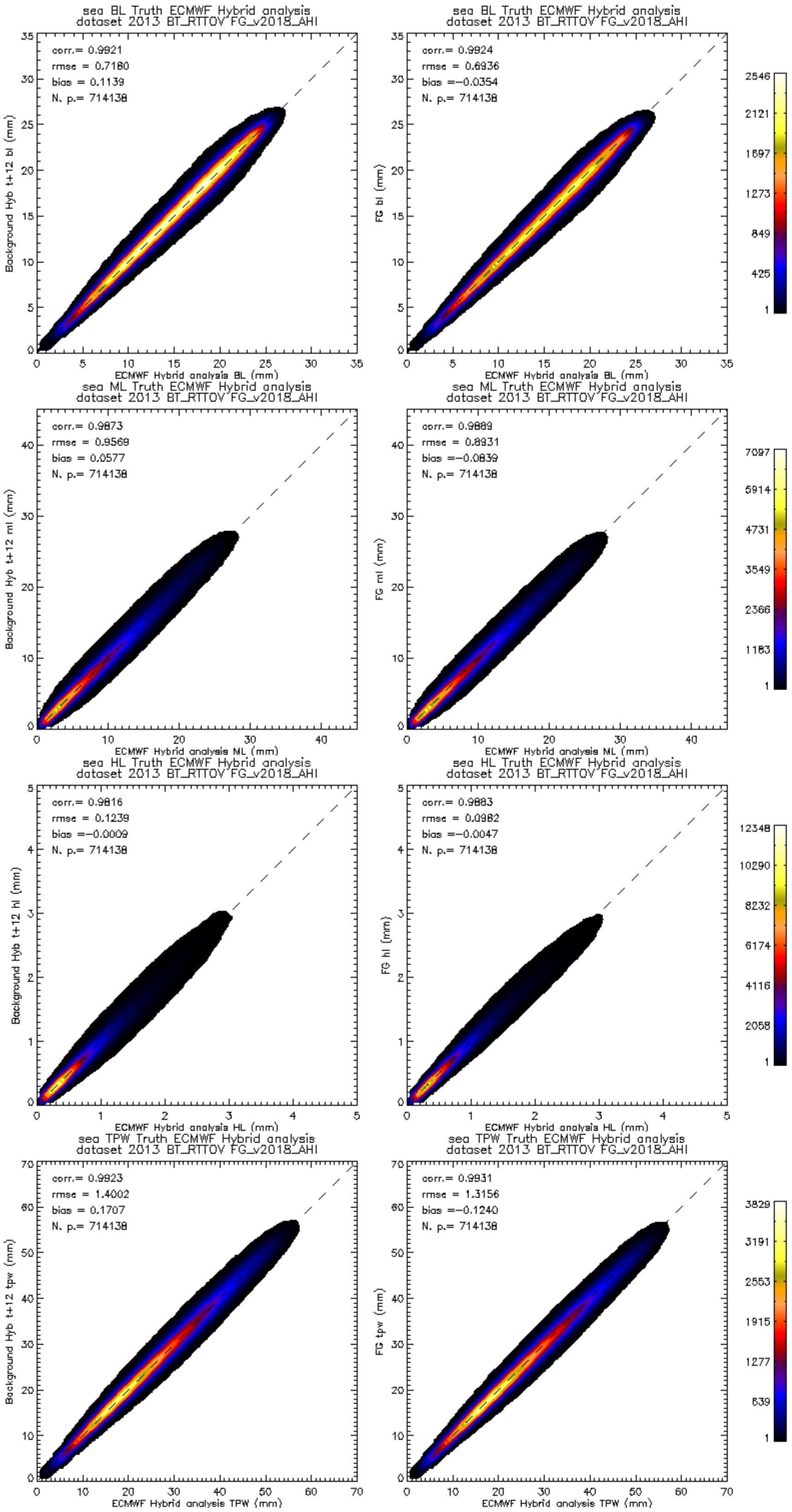

*Figure 7: **BT_RTTOV** test for synthetic Himawari-AHI: LPW and TPW 2D histograms over **sea**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

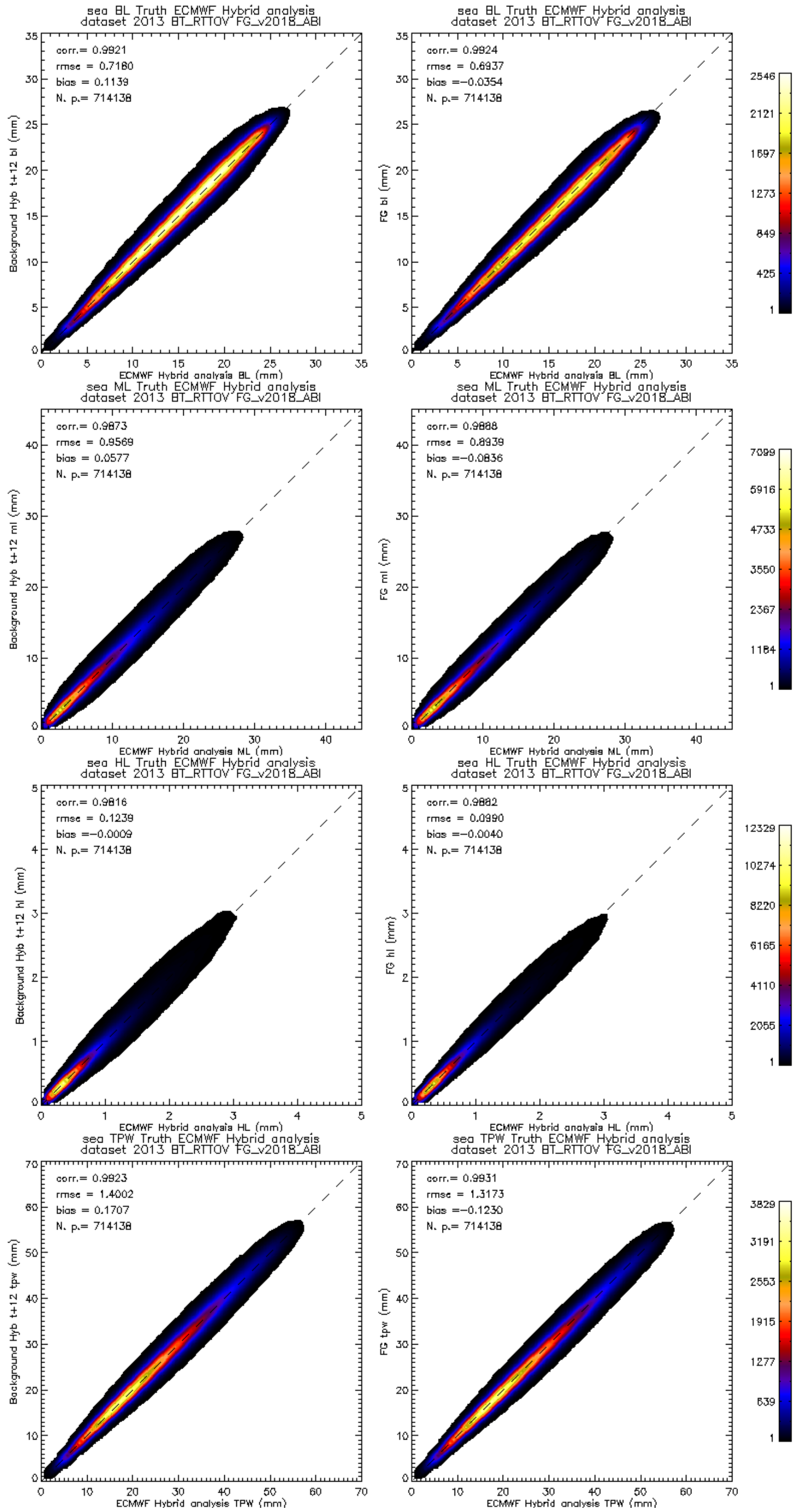

*Figure 8: **BT_RTTOV** test for synthetic GOES-16-ABI: LPW and TPW 2D histograms over **sea**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

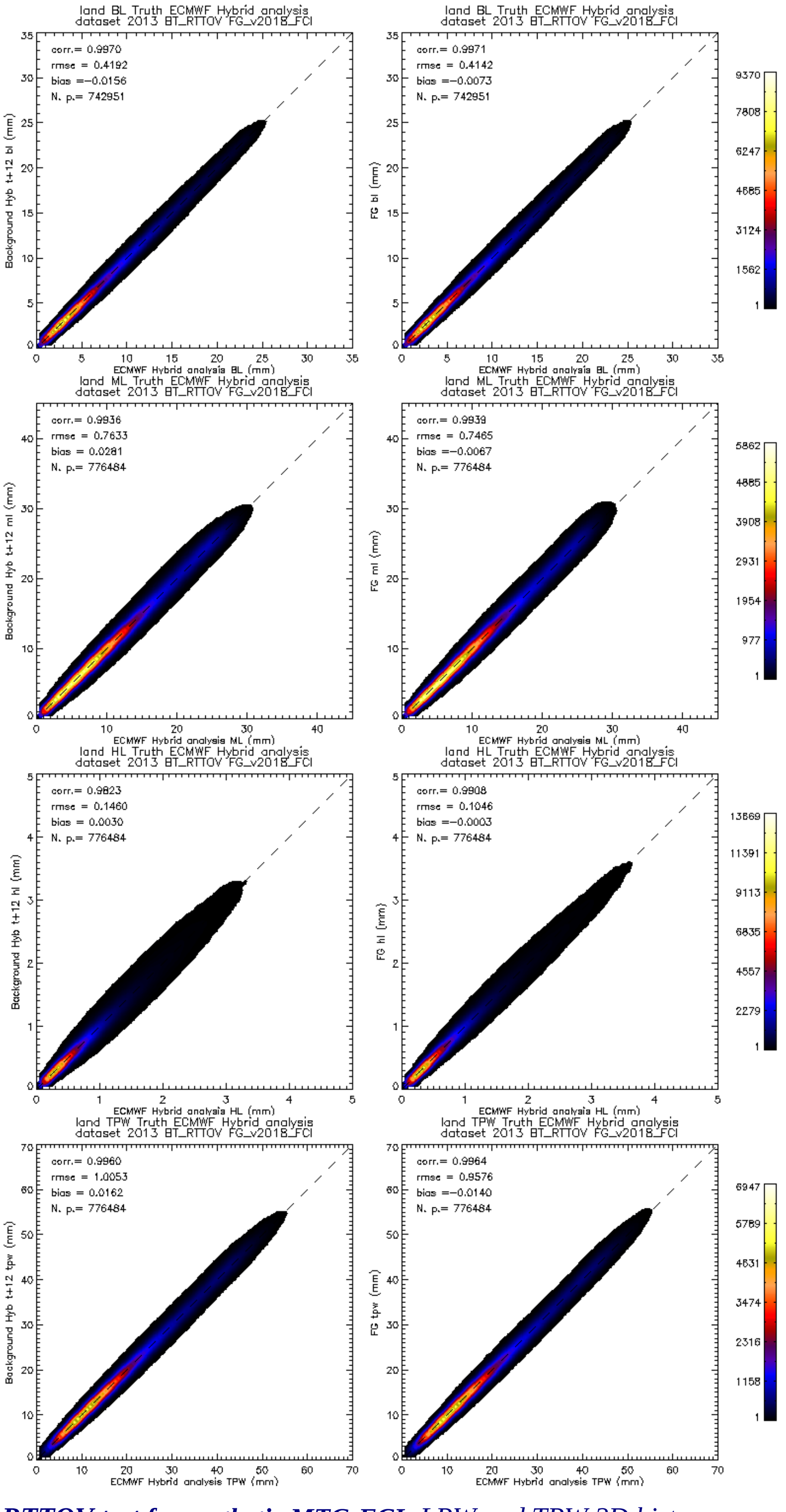

*Figure 9: **BT_RTTOV test for synthetic MTG-FCI**: LPW and TPW 2D histograms over **land**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

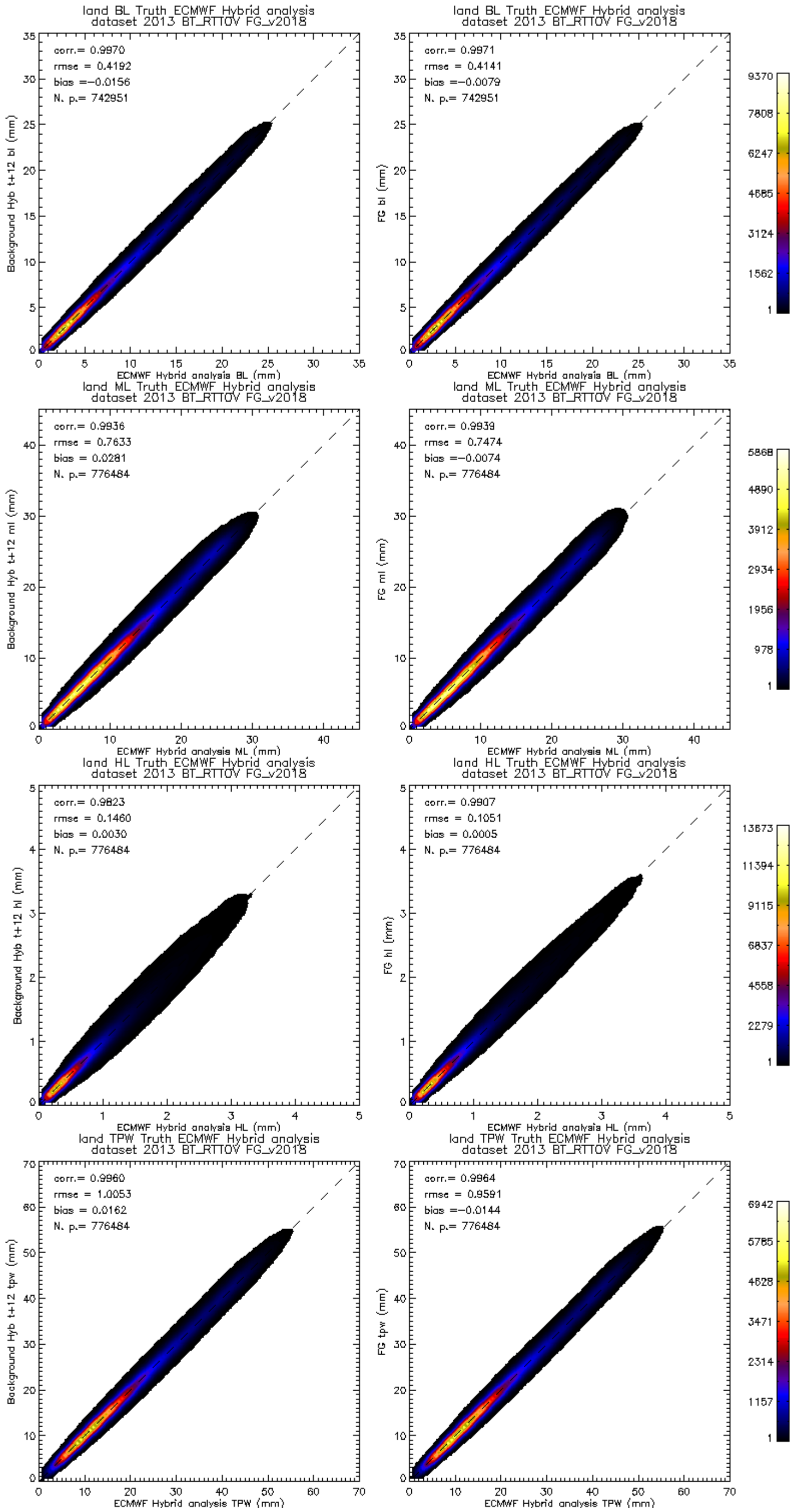

*Figure 10: **BT_RTTOV test for synthetic MSG-SEVIRI**: LPW and TPW 2D histograms over **land**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

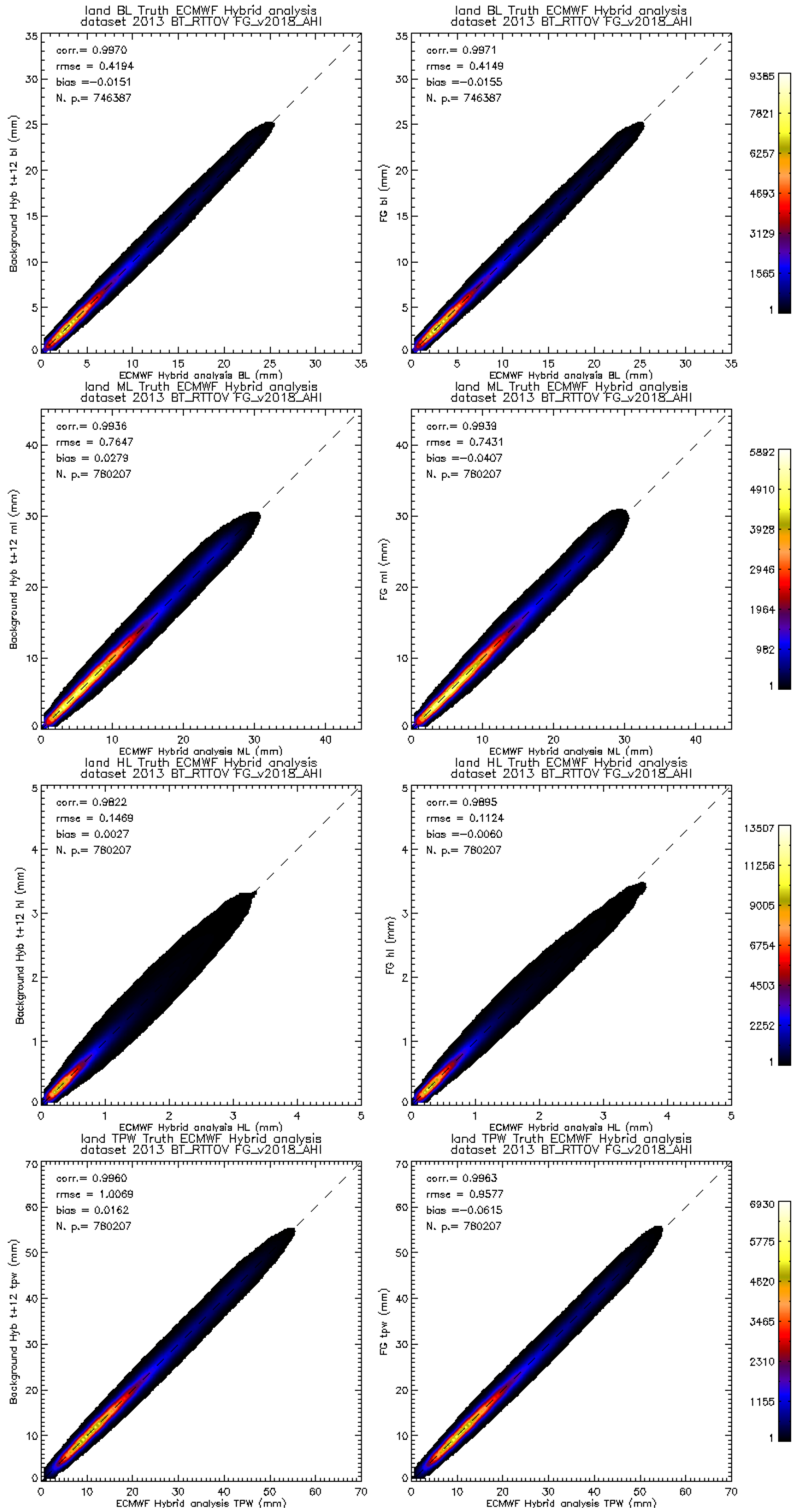

*Figure 11: **BT_RTTOV test for synthetic GOES-16-ABI**: LPW and TPW 2D histograms over **land**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

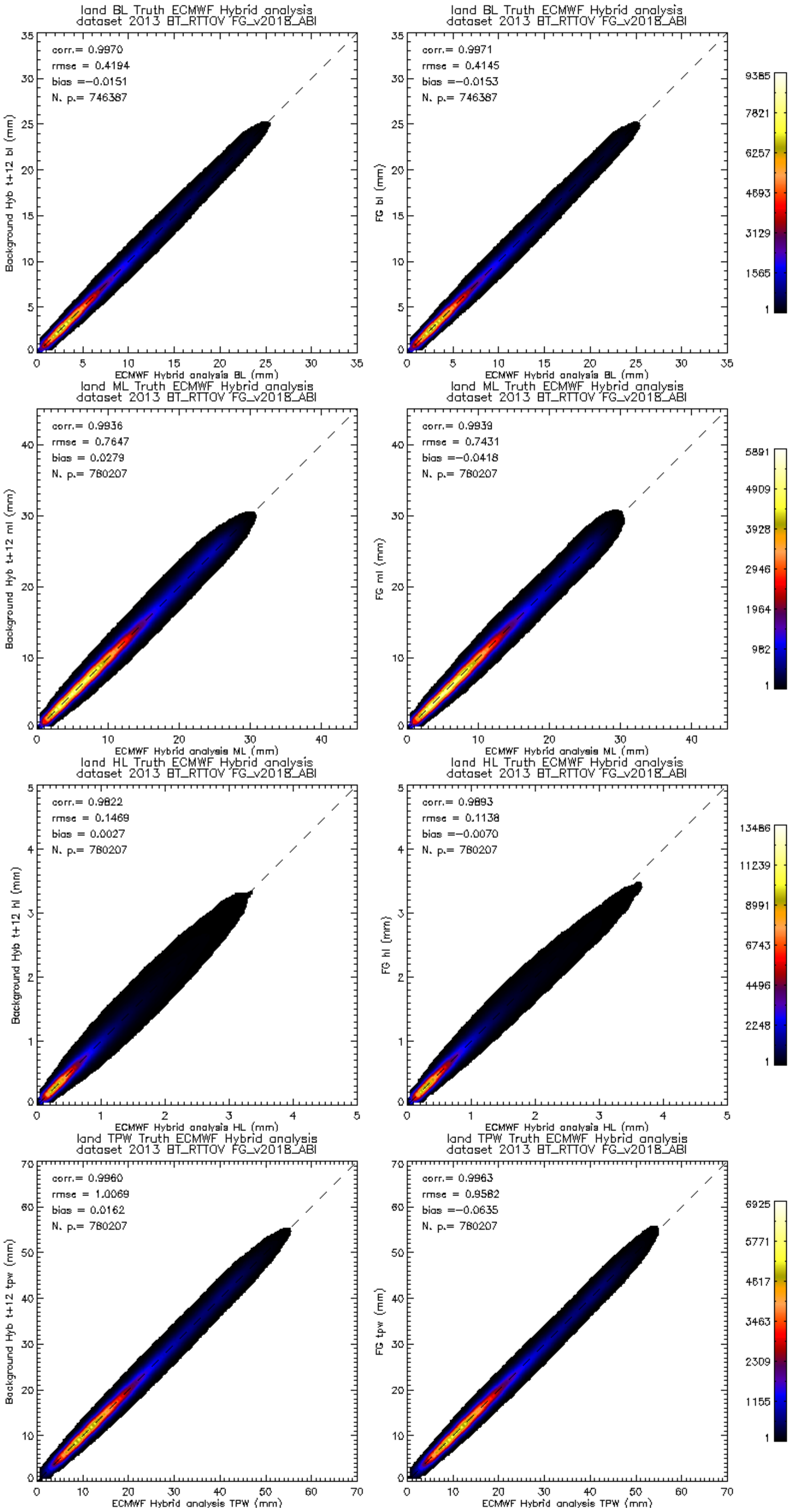

*Figure 12: **BT_RTTOV test for synthetic GOES-16-ABI**: LPW and TPW 2D histograms over **land**. From top to bottom BL, ML, HL and TPW parameters. Left) calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, right) calculated after FG step profile using as input BT_RTTOV (t+00). In all cases ground truth calculated from ECMWF analysis (t+00) profiles.*

4 SPATIAL ANALYSIS OF GEO ISHAI PARAMETERS

In Figures 13 to 16 the spatial performance of the LPW and TPW parameters RMSEs for synthetic BT_RTTOV test for GEO imager instruments respectively are shown.

The greatest values of ML and HL RMSE appear near the equatorial belt. But, when the relative ML RMSE are calculated, this effect disappears due to the high amount of precipitable water close to the equatorial belt, see as example [RD.1] for MSG-SEVIRI, AHI and ABI instruments. The relative ML RMSE has not been calculated in the case of MTG-FCI dataset 2013 and it is not shown here. You can see the comparison of ML and HL with relative ML and HL RMSE figures in the iSHAI Validation Report [RD.1] using synthetic and real data cases. Also in the iSHAI Validation Report [RD.1] the statistical performance only for Europe region are shown.

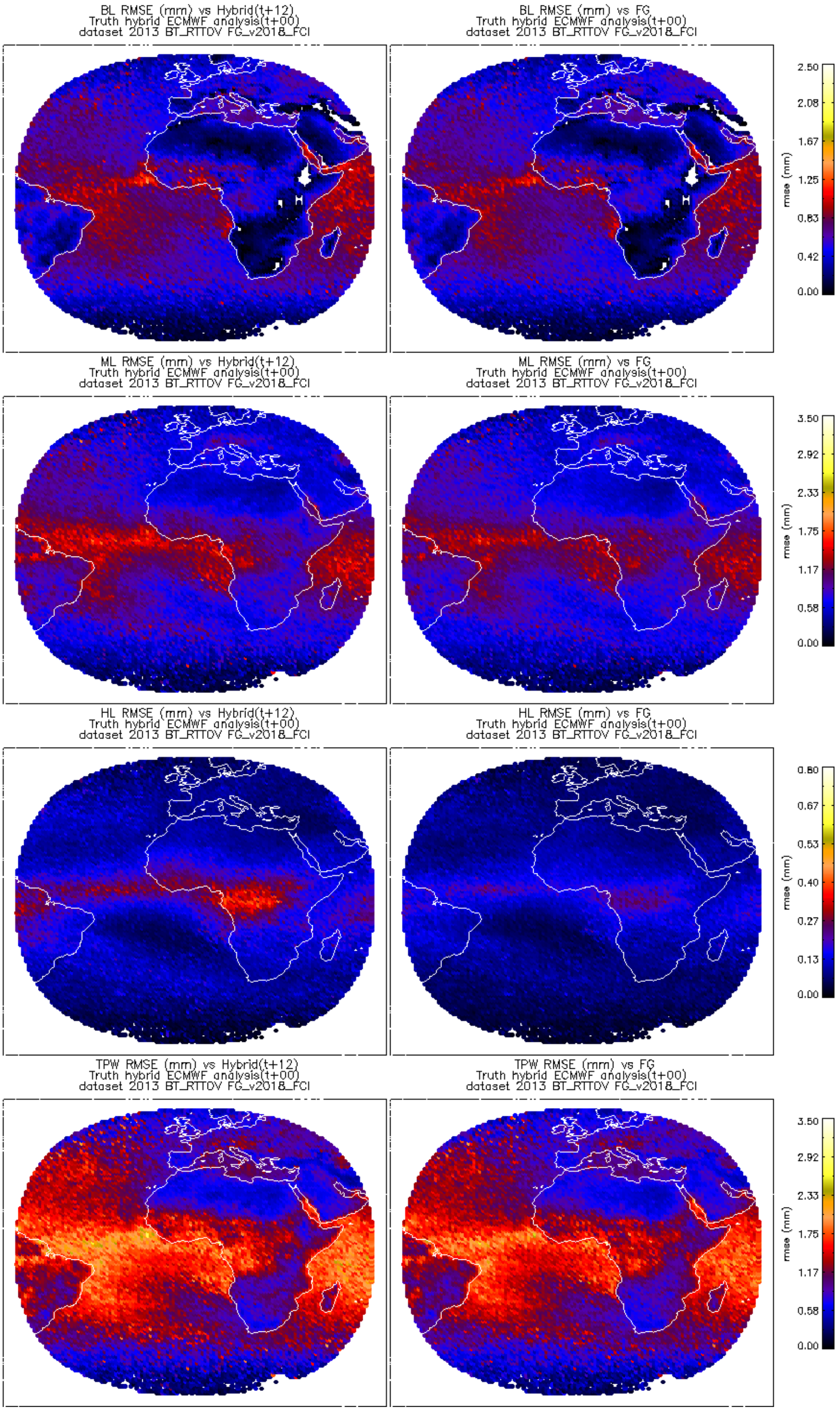

Figure 13: ***BT_RTTOV test for synthetic MTG-I FCI:*** *Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) BL, ML, HL and TPW RMSE calculated after FG step profile. In all cases the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

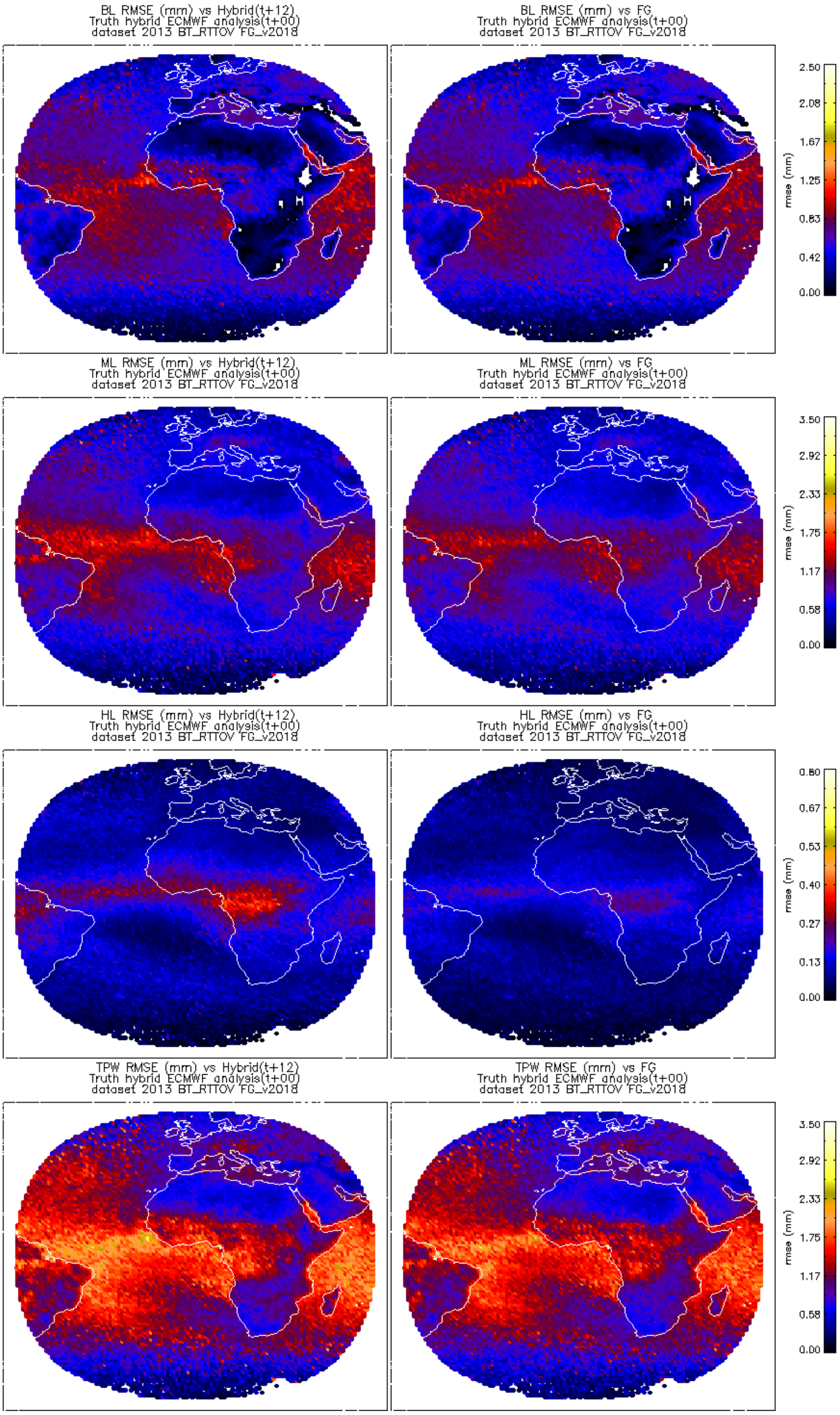

*Figure 14: **BT_RTTOV test for synthetic MSG-SEVIRI:** Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) BL, ML, HL and TPW RMSE calculated after FG step profile. In all cases the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

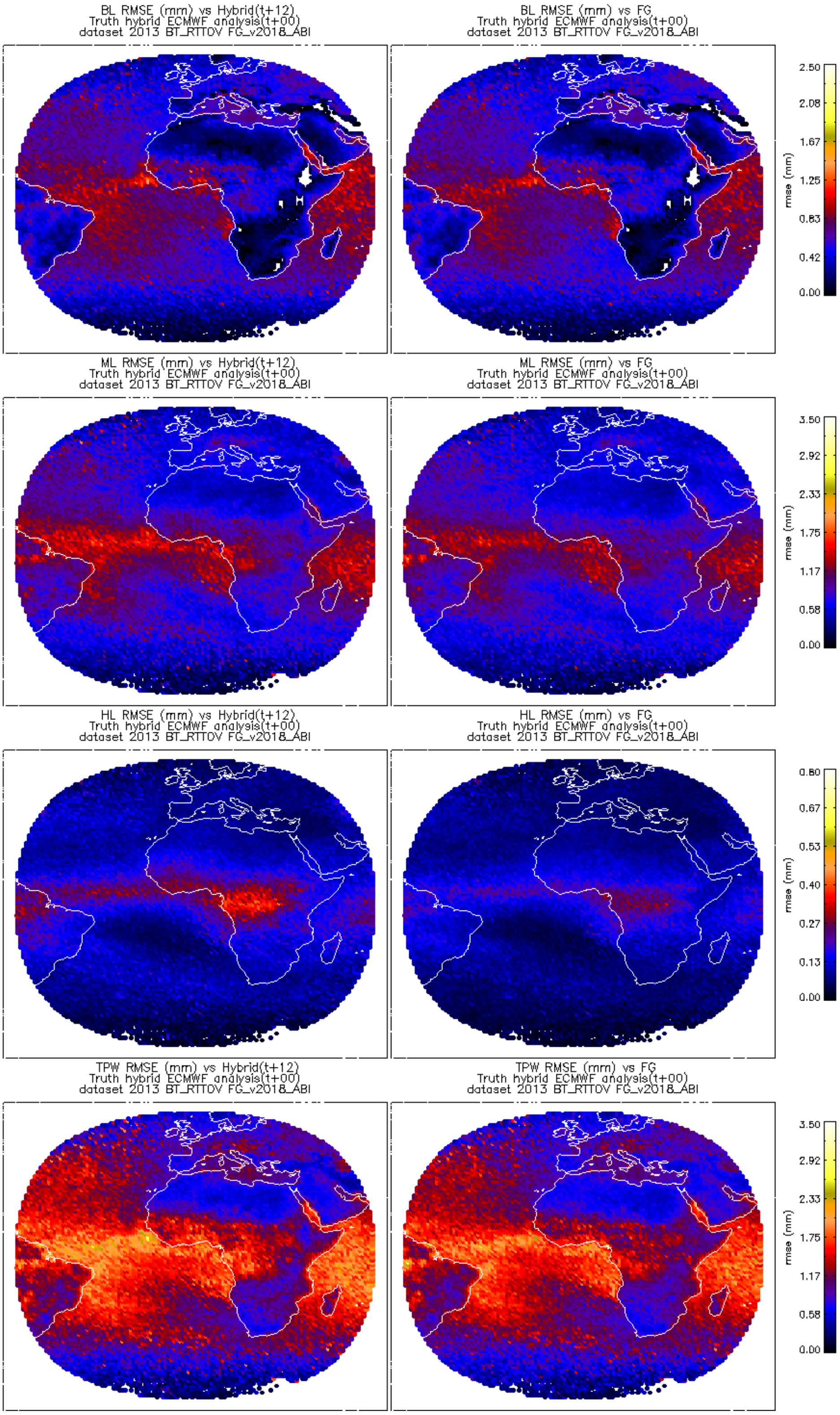

Figure 15: ***BT_RTTOV test for synthetic Himawari-AHI:*** *Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) BL, ML, HL and TPW RMSE calculated after FG step profile. In all cases the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

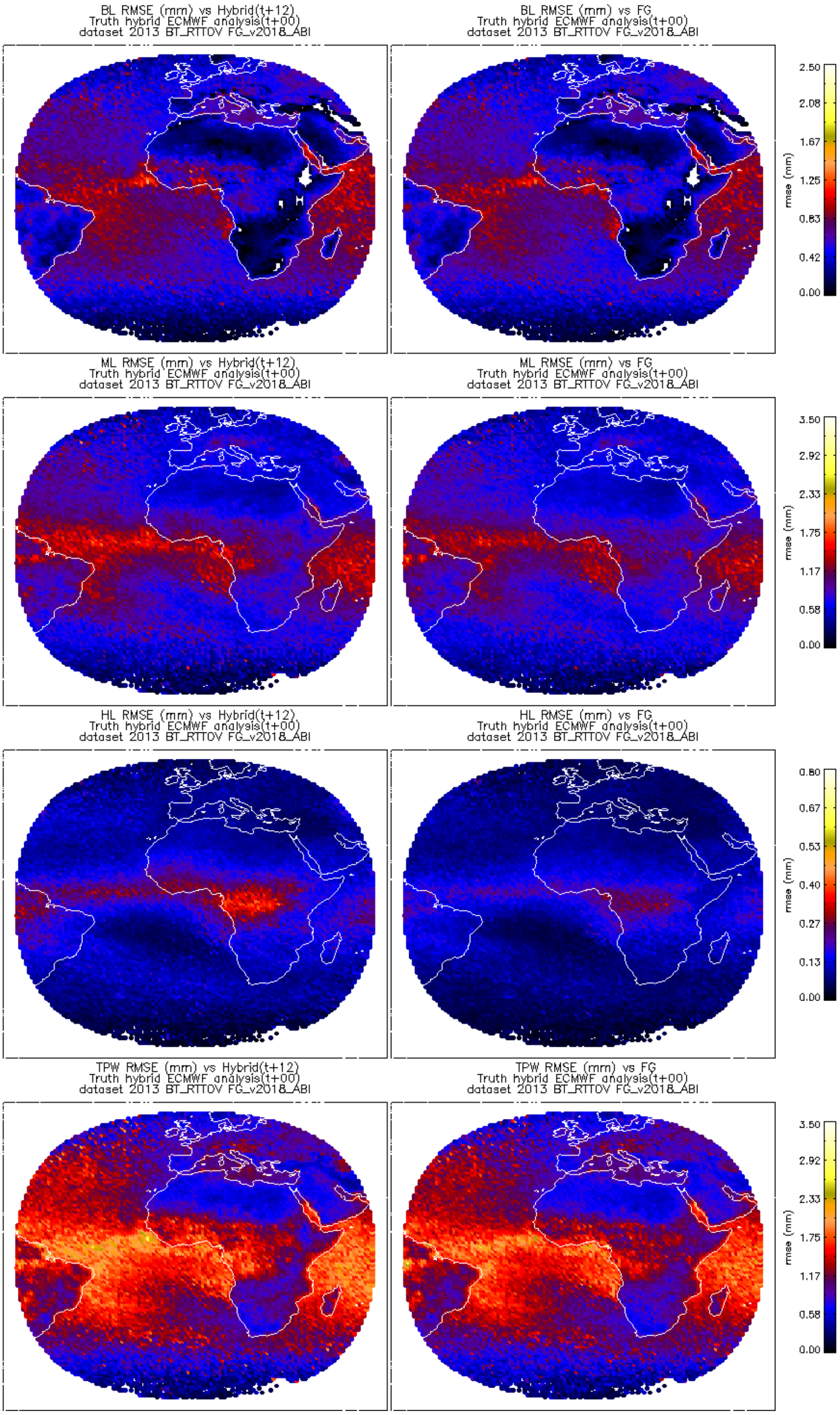

*Figure 16: **BT_RTTOV test for synthetic GOES-16-ABI:** Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) BL, ML, HL and TPW RMSE calculated after FG step profile. In all cases the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

5 VALIDATION OF GEO-ISHAI SKT: SKIN TEMPERATURE

In this Section the validation results of the Skin Temperature (SKT) are shown.

As SKT is an important internal variable used in the RTTOV calculations, SKT has been introduced as a new iSHAI output since release 2016. SKT is written just for nowcasting purposes and in order for the users to have access to this parameter. For example, the inspection of spatial gradients and temporal tendency of the SKT output could be used to detect the presence of non-adequately detected clouds or errors in the background NWP SKT. **<u>The SKT should be taken as an indicative output and it should not be considered as SST or LST products because more controls and spatial and temporal tests would be needed for this.</u>**

The SKT field of ECMWF has not a great quality especially over land pixels. The result of the Figures and Tables in this SKT Section indicates that GEO-iSHAI SKT could be used to inform users of the discrepancies between the background NWP SKT and one optimal retrieved SKT in clear pixels. Here all the figures has been generated from synthetic BTs; it must be taken into account that with real data the discrepancies could be due to physical reasons, due to undetected clouds or due to error in the emissivities, etc.

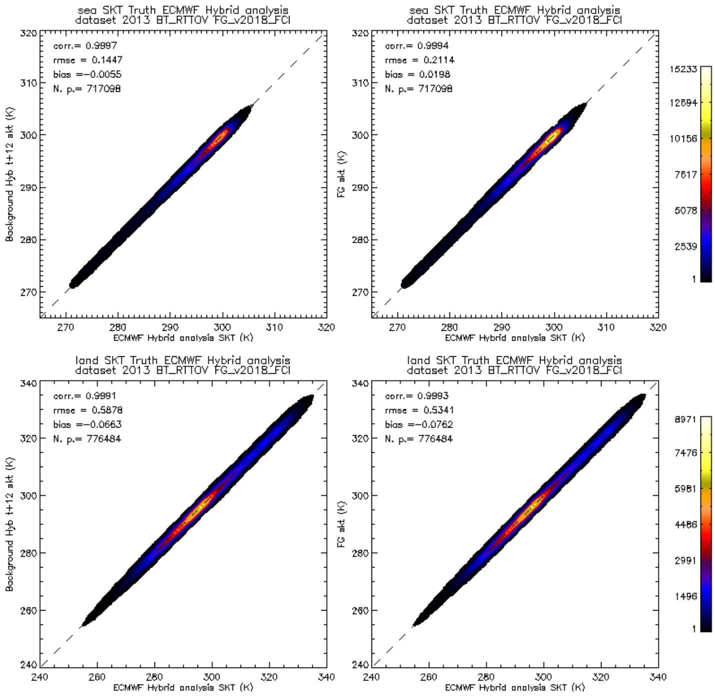

*Figure 17: **BT_RTTOV test for synthetic MTG-I FCI:** SKT 2D histograms. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) SKT RMSE calculated after FG step profile. In all cases the ground truth is SKT calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

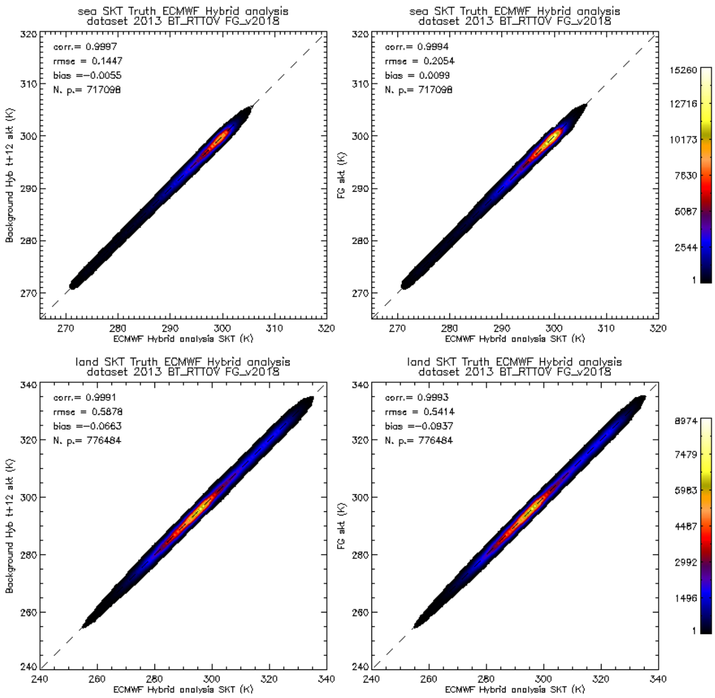

*Figure 18: **BT_RTTOV test for synthetic MSG-SEVIRI:** SKT 2D histograms. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) SKT RMSE calculated after FG step profile. In all cases the ground truth is SKT calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

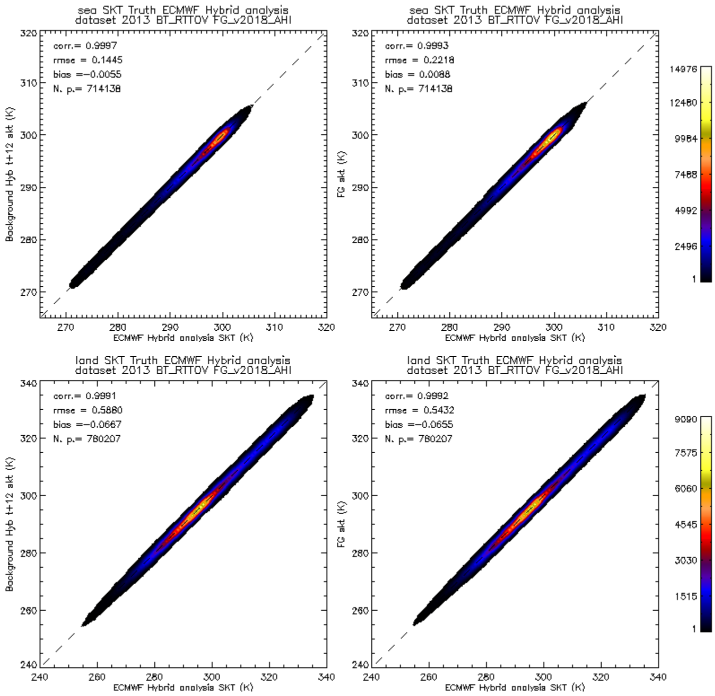

*Figure 19: **BT_RTTOV test for synthetic Himawari-AHI:** SKT 2D histograms. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) SKT RMSE calculated after FG step profile. In all cases the ground truth is SKT calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

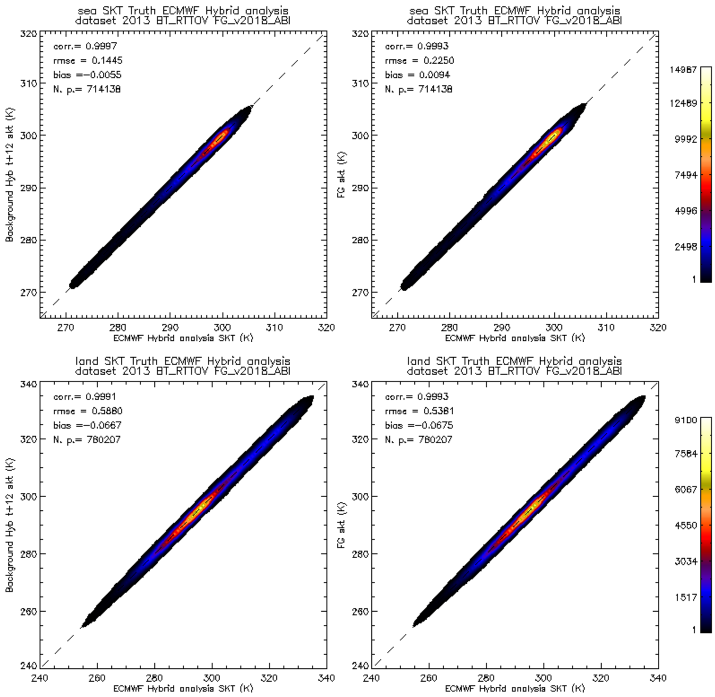

*Figure 20: **BT_RTTOV test for synthetic GOES-16-ABI:** SKT 2D histograms. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) SKT RMSE calculated after FG step profile. In all cases the ground truth is SKT calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

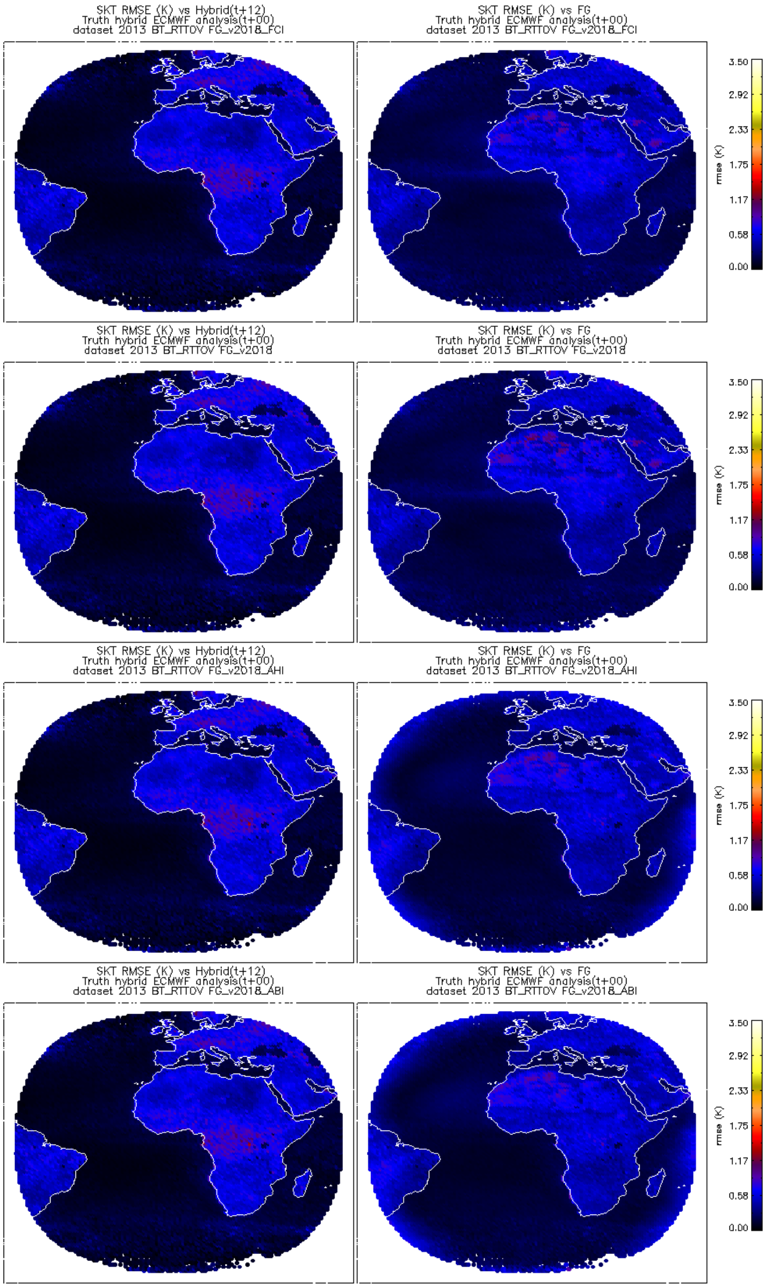

Figure 21: Spatial distribution of the SKT RMSE BT_RTTOV test. ***From (top) to (bottom) for synthetic MTG-I FCI, for synthetic MSG-SEVIRI, for synthetic Himawari-AHI and for synthetic GOES-16-ABI****. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) SKT RMSE calculated after FG non-linear regression step. In all cases the ground truth are SKT calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

6 VALIDATION OF GEO-ISHAI TOZ: TOTAL OZONE

In this Section the validation results of the Total Ozone (TOZ) are shown. Figures equivalent to those presented in previous sections for LPW, TPW and SKT are presented for the TOZ.

GEO-iSHAI TOZ output was introduced as a new output since release 2016 and TOZ is calculated from the ozone profile after applying only the non-linear regression step. In the ozone non-linear regressions, the collocated real bias corrected BTs and the temperature profile, the logarithm of ozone profile and the skin temperature from the iSHAI retrieved profile are used as inputs.

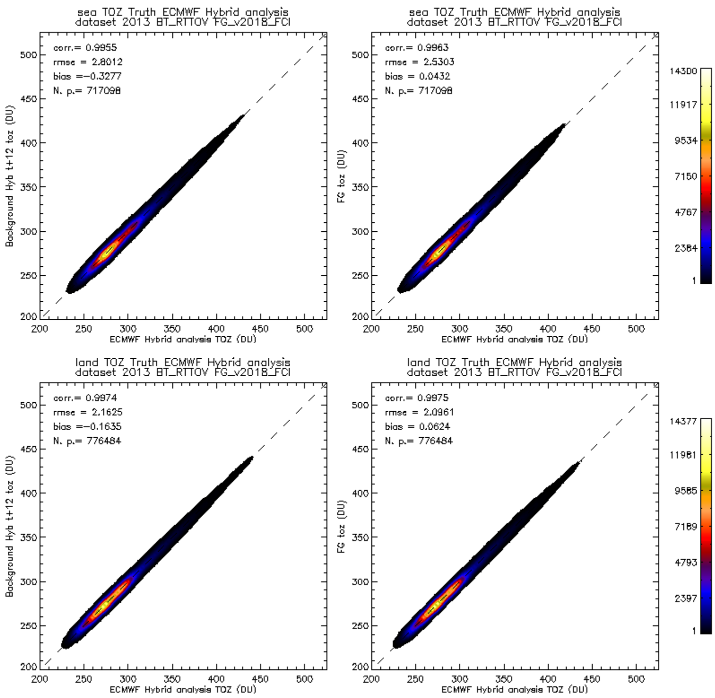

*Figure 22: **BT_RTTOV test for synthetic MTG-I FCI:** TOZ 2D histograms. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) TOZ RMSE calculated after FG step profile. In all cases the ground truth is TOZ calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

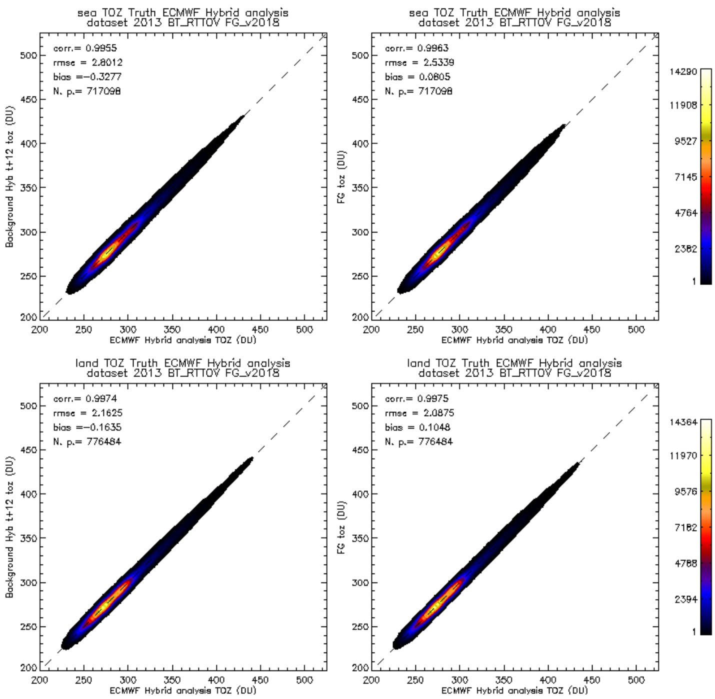

*Figure 23: **BT_RTTOV test for synthetic MSG-SEVIRI:** TOZ 2D histograms. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) TOZ RMSE calculated after FG step profile. In all cases the ground truth is TOZ calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

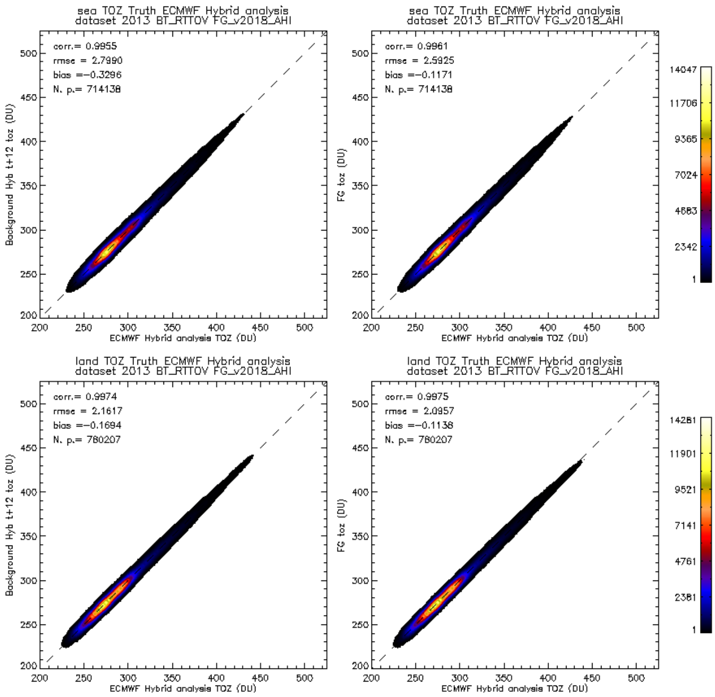

*Figure 24: **BT_RTTOV test for synthetic Himawari-AHI:** TOZ 2D histograms. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) TOZ RMSE calculated after FG step profile. In all cases the ground truth is TOZ calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

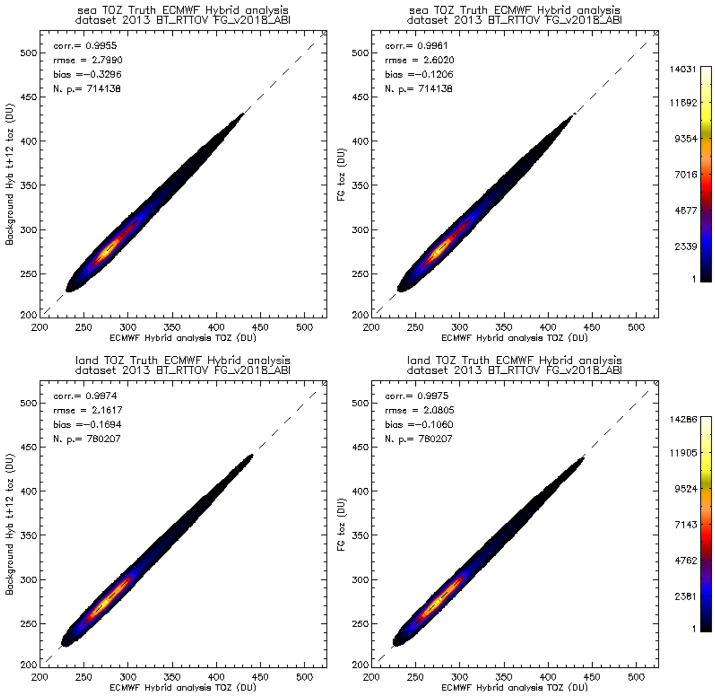

Figure 25: ***BT_RTTOV test for synthetic GOES-16-ABI:*** *TOZ 2D histograms. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) TOZ RMSE calculated after FG step profile. In all cases the ground truth is TOZ calculated from NWP-Hyb ECMWF analysis (t+00) profiles.*

7 STATISTICAL SUMMARY

In order to allow a better comparison, the statistical values that appear inside the 2D histograms have been summarized below in Tables 4 to 6.

It can be seen that in this synthetic BT_RTTOV test the SEVIRI and FCI instruments have similar figures and then iSHAI from MTG-I FCI and MSG-SEVIRI will have similar theoretical performance. The performance of iSHAI for AHI and ABI instruments is similar and slightly better than the performance of iSHAI for SEVIRI and FCI instruments

BL	NWPHyb(t+12) sea pixels RMSE (kg/m²)	FG sea pixels RMSE (kg/m2)	NWPHyb(t+12) land pixels RMSE (kg/m²)	FG land pixels RMSE (kg/m²)
MTG-I FCI	0,7184	0,6939	0,4192	0,4142
MSG-I SEVIRI	0,7184	0,6938	0,4192	0,4141
HIMAWARI-AHI	0,7180	0,6936	0,4194	0,4149
GOES-16 ABI	0,7180	0,6937	0,4194	0,4145
ML	NWPHyb(t+12) sea pixels RMSE (kg/m²)	FG sea pixels RMSE (kg/m2)	NWPHyb(t+12) land pixels RMSE (kg/m²)	FG land pixels RMSE (kg/m²)
MTG-I FCI	0,9595	0,8933	0,7633	0,7465
MSG-I SEVIRI	0,9595	0,8921	0,7633	0,7474
HIMAWARI-AHI	0,9569	0,8931	0,7647	0,7431
GOES-16 ABI	0,9569	0,8939	0,7647	0,7431
HL	NWPHyb(t+12) sea pixels RMSE (kg/m2)	FG sea pixels RMSE (kg/m2)	NWPHyb(t+12) land pixels RMSE (kg/m²)	FG land pixels RMSE (kg/m²)
MTG-I FCI	0,1251	0,0956	0,1460	0,1046
MSG-I SEVIRI	0,1251	0,0961	0,1460	0,1051
HIMAWARI-AHI	0,1239	0,0982	0,1469	0,1124
GOES-16 ABI	0,1239	0,0990	0,1469	0,1138
TPW	NWPHyb(t+12) sea pixels RMSE (kg/m2)	FG sea pixels RMSE (kg/m2)	NWPHyb(t+12) land pixels RMSE (kg/m²)	FG land pixels RMSE (kg/m²)
MTG-I FCI	1,4033	1,3129	1,0053	0,9576
MSG-I SEVIRI	1,4033	1,3114	1,0053	0,9591
HIMAWARI-AHI	1,4002	1,3156	1,0069	0,9577
GOES-16 ABI	1,4002	1,3173	1,0069	0,9582

Table 4: Synthetic ***BT_RTTOV test for GEO imager instruments:*** *Summary of statistical parameters for LPW and TPW parameters over the MSG Full Disk validation points in year 2013.*

SKT	NWPHyb(t+12) sea pixels RMSE (k)	FG sea pixels RMSE (k)	NWPHyb(t+12) land pixels RMSE (k)	FG land pixels RMSE (k)
MTG-I FCI	0,1447	0,2114	0,5878	0,5341
MSG-I SEVIRI	0,1447	0,2054	0,5878	0,5414
HIMAWARI-AHI	0,1445	0,2218	0,5880	0,5432
GOES-16 ABI	0,1445	0,2250	0,5880	0,5381

Table 5: Synthetic ***BT_RTTOV test for GEO imager instruments:*** *Summary of statistical parameters for Skin Temperature (SKT) parameter over the MSG Full Disk validation points in year 2013.*

TOZ	NWPHyb(t+12) sea pixels RMSE (DU)	FG sea pixels RMSE (DU)	NWPHyb(t+12) land pixels RMSE (DU)	FG land pixels RMSE (DU)
MTG-I FCI	2,8012	2,5303	2,1625	2,0961
MSG-I SEVIRI	2,8012	2,5339	2,1625	2,0875
HIMAWARI-AHI	2,7990	2,5925	2,1617	2,0957
GOES-16 ABI	2,7990	2,6020	2,1617	2,0806

Table 6: Synthetic ***BT_RTTOV test for GEO imager instruments:*** *Summary of statistical parameters for Total Ozone (TOZ) parameter over the MSG Full Disk validation points in year 2013.*

8 CONCLUSIONS.

In this document it is provided a description of an early attempt to adapt the algorithm for generation of FG regression coefficients from SEVIRI instrument (on board of MSG satellites) to the one from FCI instrument (on board of future MTG-I satellites). Following this design, it was started the process to made a parallel copy of the process of generate and validate coefficients that could be used for MTG-FCI algorithm and to test iSHAI MTG-FCI version with test and synthetic data.

It can be seen that the performance in all GEO imager instruments is similar and that the performance of MTG-FCI is not worse but similar to the MSG-SEVIRI one.

The works, studies and ideas described in this report has been used also to develop the iSHAI versions for processing AHI and ABI instruments in version 2018.1. It is also the basis to develop the future NWC SAF iSHAI MTG-FCI Day-1 version to be delivered around 2021-2022.

Taking into account that MTG-FCI will have better spatial and temporal resolution than MSG-SEVIRI, the MTG-FCI performance with real data will be better.

To have separated iSHAI version for each GEO imager satellite kind also will allow in the future a separated evolution of the iSHAI MTG-FCI version respect to the ten-channel instruments AHI and ABI ones. Thus, the AHI and ABI iSHAI versions could be used to explore the use of the additional IR channel and it will allow to incorporate and testing future developments made for USA of Japanese agencies or organisms in the AHI/ABI iSHAI version. Also it will allow to simplify the stop of developing of the iSHAI MSG-SEVIRI version when MTG-I FCI be fully operational just simply retiring the directory with iSHAI MSG-SEVIRI version and there will no need the elimination of dead MSG related lines in the iSHAI code.

The main objective is to improve GEO-iSHAI algorithm and to be ready for the MTG-FCI era.

8.1 FUTURE WORKS

These kind of activities has not been closed yet and will be revisited during CDOP-3 and finished in CDOP-4 when MTG-FCI be fully operational.

In the 2018 Convection Working Group (CWG) workshop at Ljubljana, it was also commented that these early results (based on the 2013 iSHAI training dataset) were provisional and that the effort to create an iSHAI training datasets with the 2017 year would be made. The 2017 iSHAI training dataset has been used to train the coefficients of the 2018.1 iSHAI version for MSG/SEVIRI and HIMAWARI/AHI as it GOES-R_class/AHI.

Once the generation of coefficients for GOES-R_class/AHI is finished and all the documentation is available there is time to repeat soon the training of the regression coefficients for MTG-FCI using the 2017 iSHAI training dataset and to complement the validation for synthetic RTTOV case with the four NWCSAF supported GEO imager instruments. The MTG-FCI results with 2017 iSHAI training dataset will not be added to [RD.1] and it will be publish in a Scientific Report.

9 REFERENCES

Li, J. et al., 2010. NOAA-NESDIS ATBD for ABI Legacy Atmospheric Moisture Profile / Legacy Atmospheric Temperature Profile / Total Precipitable Water / and Derived Atmospheric Stability Indices. *http://www.goes-r.gov/products/ATBDs/baseline/Sounding_LAP_v2.0_no_color.pdf*

Li, J. et al., 2010. NOAA-NESDIS ATBD for ABI Ozone Detection. *http://www.goes-r.gov/products/ATBDs/option2/AAA_Ozone_V2.0_no_color.pdf*

Martinez M. A., Calbet X, 2018. iSHAI and PGE00 as key tools in pre-convection. *2018 Convection Working Group Workshop in Ljubljana, Slovenia.*

Martinez M.A., Calbet X, 2016. NWCSAF/GEO clear air product version 2016. Evolution to the MTG era. Proc. *Proc. The 2016 EUMETSAT Meteorological Satellite Data User's Conference, Darmstadt, Germany*.

ANNEX I: GEO-ISHAI VALIDATION DATASET

DESCRIPTION OF FILES USED

In order to build the GEO-iSHAI datasets for training and validation purposes, data from MSG SEVIRI images and ECMWF GRIB files have been used. The input data files available for training, tuning and validation activities are:

From ECMWF fields:

- 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 2012 12 Z to 31 December 2013 of each day during this period.
- horizontal resolution: 0.5º by 0.5º
- vertical resolution: two different vertical resolutions are used
    - Hybrid levels (hereafter denoted as NWP-Hyb): The number is 91 or 137 levels depending on the ECMWF cycle.
    - Fixed pressure levels (hereafter denoted as NWP-P): These GRIB files are needed only for PGE01 CMa processing. The pressure levels available on MARS are typically 15 or 25 synoptic levels (as example: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100 .. hPa)

      Note: the NWP-P GRIB files are the ones used as input to the MSG NWCSAF software package.
- parameters: temperature (T), humidity (specific humidity [q] for NWP-Hyb and relative humidity [RH] for NWP-P files).

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

From MSG 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 2013 to 31 December 2013
- horizontal resolution: SEVIRI full resolution and MSG projection
- SEVIRI channels: All SEVIRI channels but HRVIS
- In this report the period selected is the year 2013 and for this reason MSG-3 is the satellite used when actual SEVIRI brightness temperatures (hereafter BTs) are used.

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.

DESCRIPTION OF PROCESS TO GENERATE

To build one good validation and training dataset is a high priority and standing task for us. The process to build the GEO-iSHAI v2016 validation dataset is a heritage of the process used to build the PGE13 SPhR validation dataset.

The main idea is to generate the whole validation dataset using PGE01 CMa and a modified version of PGE13 SPhR from NWSAF/MSG v2013 software called PGE00Hyb which uses RTTOV-11.2 instead of RTTOV-9.3.

PGE13 SPhR, PGE00Hyb and GEO-iSHAI allow activating an option in their configuration files to write in optional binary files structures at clear pixels with real SEVIRI BTs together with the T, q and O_3 profiles, surface and ancillary parameters collocated spatially, temporally and vertically interpolated to the position and time of the clear SEVIRI pixels.

The use of RTTOV-11.2 implies that the GEO-iSHAI validation dataset is based now on profiles with 54 pressure levels; instead of the 43 pressure levels from RTTOV-9.3. In the near future the process will be migrated to NWC/GEO v2016 GEO-CMA and GEO-iSHAI. Thus, all the profiles used in the validation have similar characteristics to the profiles used and retrieved within the GEO-iSHAI execution in operational mode.

The 2013 year has been chosen as the reference period for the GEO-iSHAI validation dataset. The validation results obtained using as ground truth the ECMWF NWPHyb analysis (*t+00*) profiles are presented here. Also the validation results using as input to GEO-iSHAI real SEVIRI BTs bias BT corrected are included in order to show the deviation from the theoretical ones and to advice users on the limitation of the algorithm.

To avoid using for validation the same records as the ones used for calculation of the GEO-iSHAI coefficients, the records with odd position in the complete dataset have been used for validation.

Positions of GEO iSHAI validation dataset: For GEO iSHAI parameters validation, a set of predefined positions of a 1º x 1º grid plus the RAOB stations positions have been chosen. The set contains 13001 points in the actual mask. The positions where validation is made can be seen in Figure 26.

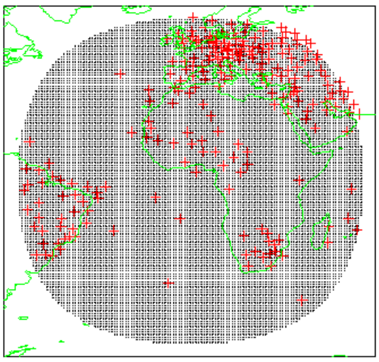

Figure 26: Predefined set of 13001 validation points used in validation datasets. Grid network of 1º x 1º plus Radiosonde Stations (red crosses).

Process to build the GEO iSHAI validation dataset: The actual process to build the validation and training dataset is the following:

a) <u>Calculate Cloud Mask (PGE01 CMa):</u> the cloud mask generation is the first step. The PGE01 program is first executed. The results of PGE01 v2013 program are HDF-5 files with the cloud mask located at *$SAFNWC/export/PGE01*.

Please, note that for PGE01CMa the NWP GRIB files used as input need to be on fixed pressure levels whilst for this Validation Report all calculations come from hybrid ECMWF GRIB files (91 hybrid levels till June 2013 and 137 hybrid levels in rest of 2013). In this report all background NWP profiles have been downloaded from ECMWF.

b) Selection of clear validation locations by screening the cloud mask: the NWC SAF PGE01 CMa cloud mask file is overwritten with the multiplication of PGE01CMa mask with the 13001 validation position mask (1 for selected pixel and 0 in the rest). The use of this screened cloud mask speeds up the process because instead of executing the physical retrieval over several millions of clear air pixels it is executed only over the clear air pixels among the 13001 predefined positions.

The process to get the screened cloud mask is:

- the cloud mask matrix is read from the PGE01 CMa file
- the cloud mask matrix is multiplied with the 13001 validation positions mask (matrix with same dimensions of CMa with values: 1 for validation points and 0 for the rest of pixels) and the result is the screened cloud mask.
- This screened cloud mask is used to overwrite the PGE01_CMa cloud mask HDF-5 file on *$SAFNWC/export/PGE01*.

c) To obtain the profiles from analysis (*t+00*) and forecast (*t+12*): the PGE00Hyb program is executed twice for each slot. The PGE00Hyb program calculates the profiles by interpolating the ECMWF fields from hybrid levels to 54 levels in the vertical and also in time and space. It also calls RTTOV-11.2 to calculate the synthetic BTs.

In the first PGE00Hyb execution: the screened cloud mask, the SEVIRI image and as background NWP-Hyb the *t+12* forecast ECMWF GRIB file from previous 12 hour ECMWF run (hereafter denoted as NWPH12) are used as inputs.

In the second PGE00Hyb execution: the screened cloud mask, the SEVIRI image and as background NWP the ECMW *t+00* analysis GRIB file (hereafter denoted as NWPH00) are used as inputs.

With the first execution it is read the (T, q, O_3) profiles and some surface parameters (P_{sfc}, T_{skin}, etc), from the *$SAFNWC/tmp* binary files, at the clear air predefined positions from *t+12* hours ECMWF forecast.

In the second execution it is read the (T, q, O_3) profiles and some surface parameters at the clear air predefined positions from ECMWF *t+00* analysis.

Together with the *t+00* and *t+12* profiles, ancillary data (as emissivities, longitude, latitude, zenith angle, etc) are also read from the *$SAFNWC/tmp* binary files. The process can be seen in the Figure 27. The result is one binary file by slot that can be easily read on IDL with the *restore* command. The mean number of retained clear pixels by slot is 4829.

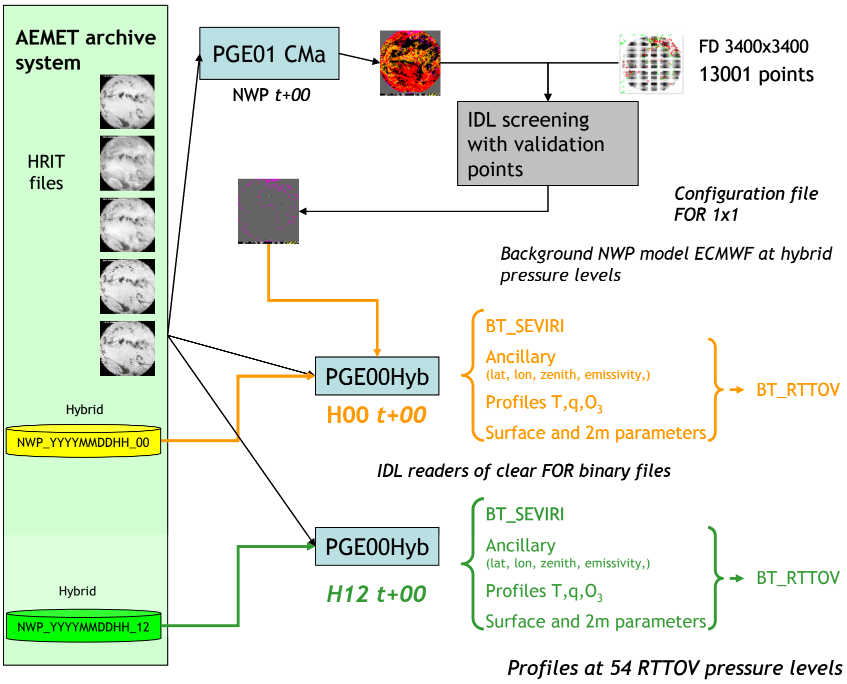

Figure 27: Generation of the records for adding to GEO iSHAI validation dataset from one image.

d) Joining the files for every slot on monthly files: In order to allow an easy management of the datasets, the slot binary files are joined in one file for every month. It is made with one IDL procedure. This monthly binary files are the base for the validation process since files on a monthly basis can be joined easily to build a wider period dataset.

e) Joining the monthly files on a period file: Once a period is selected for validation or training, one period binary file is generated joining the monthly files for the months in the period. It is made with one IDL procedure. In this validation report it has been used 1 out 2 (pixels with odd positions) clear pixels for the 2013 year. The number of pixels used here is 1,709,518 pixels. Other reason to use 1 out 2 of the 2013 clear pixels is that is needed 3.5 GB to store it in memory (the compressed file has a size of 2.5 GB).

f) Write binary file which can be used as input to the validation version of GEO-iSHAI: One array with selected parameters is written on a binary file in a format that will be used later with the validation version of GEO-iSHAI. This validation version processes the data record by record instead of processing a MSG image, as is done in the operational version.

g) Build the GEO iSHAI validation dataset: The outputs of the previous step are blended with the structures of the validation dataset (ancillary fields as emissivity values, land/sea mask, height, etc.) and one IDL binary file for restoring is generated.

Structure of the records in the GEO iSHAI validation dataset: After the execution of the previous steps the validation dataset for a period is one array (that could contains several millions of records) written on an IDL binary file. The array can be restored easily with the use of IDL command *restore*. Every record is one IDL structure with the following parameters or fields:

- ***Ancillary***: longitude, latitude, emissivity values, etc.
- ***Date***: day, year, hour, etc.
- ***NWP from ECMWF analysis (t+00):*** ECMWF temperature and humidity profiles interpolated to the 54 RTTOV pressure levels interpolated vertically from the 91 or 137 hybrid levels, T_{skin}, pressure at surface, etc. from the analyses (*t+00*) ECMWF GRIB files. It will be used as the validation truth.

- ***NWP from ECMWF forecast (t+12):*** ECMWF temperature and humidity profiles interpolated to the 54 RTTOV pressure levels interpolated vertically from the 91 or 137 hybrid levels, T_{skin}, pressure at surface, etc. 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).
- ***BT_SEVIRI(8)***: uncorrected BT from HRIT file. *BT_SEVIRI[IR3.9,WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]*.
- ***BT_RTTOV(8) from NWPHyb(t+00)***: Synthetic BTs calculated using the RTTOV-11.2 with the analysis (t+00). *H00.BT_RTTOV[IR3.9,WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]*.
  BT_RTTOV(8) from NWPHyb(t+12): Synthetic BTs calculated using the RTTOV-11.2 with the forecast (t+12). *H12.BT_RTTOV[IR3.9,WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]*.

These basic validation and training datasets have been used for several tasks. One of them consisting of splitting the records in the global dataset into records to generate the validation datasets (odd positions) and the training ones (even positions).

To see more examples of the use and the characteristic and how was built MSG 2013 iSHAI training dataset see as example [RD.4] or [RD.5].