VICIRS
Development of vicarious calibration tools for MWI and ICI using radiosoundings (VICIRS)

D17: Final technical report

Deliverable #	D17
Doc type	Technical report
Authors	Domenico Cimini Francesco Di Paola Sabrina Gentile Salvatore Larosa Fabio Madonna Mario Montopoli Elisabetta Ricciardelli Filomena Romano Mariassunta Viggiano

Date: 03 July 2024

TABLE OF CONTENTS

1. Introduction

Calibration of satellite observations is a crucial step to ensure the level of data quality that is essential for the reliability of atmospheric products in different meteorological and climate applications. The methods and instrumentation involved in radiometric calibration can be grouped into three approaches (Dinguirard and Slater, 1999): on the ground prior to launch (preflight), on board the spacecraft post-launch, and vicarious or indirect. Although the pre-flight on ground calibration characterizes the sensor, frequent checks shall be performed to monitor sensor calibration in flight using on-board instruments (if present) or via vicarious calibration approach. Vicarious calibration methods are also needed to check the status of the on-board instrumentation and to monitor any post-launch degradation. Vicarious calibration methods are external to the satellite and depend on the accurate characterization of the reference scenes.

The VICIRS (VIcarious Calibration for MWI and ICI using RadioSoundings) study implemented a methodology and developed a vicarious calibration tool for the two conical-scanning radiometers planned to fly from 2025 onwards aboard the METOP-SG (Second Generation) satellites as part of the EUMETSAT Polar System (EPS) program. These radiometers are the Microwave Imager (MWI) and the Ice Cloud Imager (ICI). ICI will be the first operational sensor covering the mm/sub-mm wavelengths from 183 to 664 GHz. Its main objective is to provide data on humidity and especially on ice hydrometeors. MWI operates at 18 frequencies between 18 and 183 GHz. All the MWI channels up to 89 GHz and the ICI quasi-window channels at 243 and 664 GHz will observe in dual polarization, while only vertical polarization will be provided for the other frequencies. The combined use of MWI and ICI radiometers will provide an unprecedented set of microwave passive measurements, from 18.7 GHz up to 664 GHz.

The VICIRS tool compares clear-sky MWI/ICI brightness temperature (BT) observations with BTs simulated from radiosoundings (RS) profiles using the GRUAN processor (Carminati et al. 2019), modified to expand its capabilities to MWI/ICI and new RS types (currently at version 6.3.b.0.1) and to be interfaced with RTTOV v13.2. The validation process is particularly relevant for the ICI mission, because no similar instruments currently exist in orbit that deploy the same ICI frequencies within the sub-millimetre spectrum. Therefore, the radiometric validation of the ICI data can be achieved only by means of radiative transfer simulations, based either on Numerical Weather Prediction (NWP) forecasts or on in-situ measurements like RS.

The tool is designed to search match-ups between MWI/ICI and RS from the high-quality low-density Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) archive, as well as from a lower-quality but higher-density archive, the Radiosounding HARMonization (RHARM) data set. It also handles NWP profiles in order to fill the gap of surface parameters and of data over the RS top levels. For each match-up, different types of Target Areas (TAs) are considered depending on the availability of sonde drift. A statistical analysis of the difference between observed and simulated BTs, considering the overall uncertainty emerging from the various sources (e.g., instrumental uncertainties, forward model uncertainties, spatial and temporal mismatches) is determined for each MWI/ICI channel. Moreover, in order to characterize, simultaneously, the error structure of three collocated (in space and time) MWI/ICI, RS and NWP measuring systems, the Multi-source Correlative Methodology (MCM) analysis has also been implemented.

The VICIRS tool has been tested on MWI/ICI Level 1B simulated observations, provided by EUMETSAT, and corresponding RHARM RS, and on a dataset consisting of match-ups between observations from the NASA Global Precipitation Measurement Microwave Imager (GMI) and GRUAN and RHARM RS during 6 months of 2023 and 3 months of 2019, respectively.

The report is organized as follows, after a short review of the methods existing in literature for calibrating sensor operating at microwave frequencies (Section 2), Section 3 describes the radiosonde archives from which the RS used for calibration are collected, while the methodology analysis developed in VICIRS study is described in Section 4. The VICIRS-tool scheme is described in Section 5 and finally Section 6 deals with the verification and validation of the tool when applied to the MWI/ICI L1B simulated observations against RHARM profiles and to GMI real observations against RHARM and GRUAN profiles.

1.1 Applicable documents

[AD-1] Statement of Work, Study on the development of vicarious calibration tools for MWI and ICI using radiosoundings. EUM/RSP/SOW/22/1290081, Issue v1B, 22 June 2022.
[AD-2] ICI Calibration and Validation Plan, EUM/LEO-EPSSG/PLN/17/776069 v1D.
[AD-3] MWI Calibration and Validation Plan, EUM/LEO-EPSSG/PLN/14/776068 v1F.
[AD-4] MWI L1B Product Format Specification, EUM/LEO-EPSSG/SPE/14/767115 v4
[AD-5] ICI L1B Product Format Specification, EUM/LEO-EPSSG/SPE/14/771723 v4
[AD-6] EPS-SG Programme Overall Calibration and Validation Plan, EUM/LEO-EPSSG/PLN/14/758341
[AD-7] D05 Critical review of the literature methods
[AD-8] D07 Report on methodology analysis
[AD-9] D15 Algorithm Theoretical Basis Document (ATBD)

1.2 Acronyms

Abbreviations specific to this document are listed in the following table.

Acronyms	Definition
AMD	Air Motion Displacement
BT	Brightness Temperature
BT_O	Observed BT (by satellite-based sensor)
BT_S	Simulated BT
BT_TA	average of the FOVs-BT included in TA
DL	Dedicated Launch
FOV	Field Of View
GRUAN	Global Climate Observing System Reference Upper-Air Network
ICI	Ice Cloud Imager
LF	Land Fraction
MCM	Multi-source Correlative Methodology
MWI	Microwave Imager
NWP	Numerical Weather Prediction
nlev	Number of RS pressure levels
numTA	Number of TA types
P	Pressure
Pmin	minimum pressure level
RH	Relative Humidity
RHARM	Radiosounding HARMonization
RS	Radiosounding
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVS
SD	Standard Deviation
T	Temperature
TA	Target Area
TC	Triple Collocation
uBT	BT uncertainty
SD_TA	SD corresponding to BT_TA
SAT	Generic for MWI and ICI
VICIRS	**VI**carious **C**alibration for MWI and **I**CI using **R**adio**S**oundings

2. Review of vicarious calibration methods

Various methods have been proposed to calibrate a satellite sensor after launch (Slater et al., 1987, Santer et al., 1992). In general, a critical point to evaluate the radiometric accuracy of microwave measurements after launch is the lack of reference data. Generally, the vicarious calibration methods can use three types of reference-data sources: (i) radiative transfer (RT) simulations based on atmospheric profiles from radiosonde sensors or reanalysis data (e.g., Kerola, 2006; Moradi et al., 2013a; Saunders et al., 2013); (ii) similar instruments aboard aerial platforms (e.g. Wilheit, 2013); and (iii) similar instruments aboard other spaceborne platforms (Mo, 2007; Moradi et al., 2015;

Sapiano et al., 2013). Because of the unprecedented MWI and ICI channels that cannot be compared with reference observed data from space, the vicarious calibration in VICIRS is performed by using radiance-based method (i) that is characterized by the attempt to predict the radiance at the top-of-atmosphere (TOA), i.e. at sensor level, over selected “pseudo invariant” calibration sites at the time of satellite overpass and in a similar viewing geometry.

The first step of a calibration tool is the collocation process that consists in finding match-ups between satellite and reference data points in space and time. In comparing radiosonde and remote satellite measurements it is important to consider the different nature of the measurements and their different scale (Bobrishev et al., 2018). In fact, radiosondes provide in-situ measurement of the troposphere while the satellite observes the whole vertical extent of the atmosphere at a viewing angle (53° for MWI and ICI) covering an area corresponding to the satellite Field of View (FOV). While the radiosonde vertical profile is often treated as a point measurement related to a certain location, the sonde drifts horizontally depending on dynamical conditions, bursting at distances that may be tens of km away from the launch site. From the study of Seidel et al. (2011) that collected two years of RSs from 419 stations to perform a comprehensive global climatology of balloon drift distance, resulted that typical drift distances range from few km in the low troposphere to about 50 km in the lower stratosphere depending on some parameters such as wind speed and direction, height above the surface, latitude, and season. A very important aspect in defining temporal distance between satellite overpass and radiosonde launch, is the difference in the satellite and radiosonde acquisition time, in fact the radiosonde ascent time varies from a minimum of 1.43 hours to a maximum of 1.76 hours (Seidel et al., 2011) while the FOV information is acquired in less than 1 second. The satellite-radiosonde temporal distance is determined as the difference between the satellite overpass and the radiosonde target time that does not always coincide with the launch time of the sonde, but it is estimated by adding to the launch time a time interval depending on the radiosonde ascent time. Because of the non-perfect matching of the spatial and temporal sampling of radiosonde and satellite measurements, a particular attention is needed in defining collocation criteria to minimize the representativeness errors on the validation/calibration process. To this aim, many studies further analyze the collocated samples in terms of homogeneity, cloud presence and land surface emissivity effect on satellite measurements so as to avoid heterogeneous match-ups. The collocation criteria are usually based on spatial and temporal proximity of the two measurements. Suitable spatial and temporal criteria for establishing the maximum spatial/temporal distance depend mainly on the satellite FOV dimension and on the type of ground-based instruments. Furthermore, the analysis of the atmospheric conditions, which often dominate the uncertainty of the comparisons related to the match-up, is essential to establish whether the two measurement systems look at the same air mass, thus deciding the reliability of the match-up in the comparison. In general, the collocation criteria present in literature match a RS with the Single Closest (SC) satellite observation or with a Target Area (TA) built around the radiosonde launch-site. In most studies, once the match-ups have been identified, homogeneity tests, cloud detection, and emissivity analysis are applied to decide whether to consider the match-ups in the calibration process. Table 2.1 summarizes for each reference the collocation criteria and, when present, the method used for cloud detection and emissivity analysis.

Table 2.1: Collocation approaches in literature

REFERENCES	SATELLITE/ SENSOR	COLLOCATION	TARGET TIME	COLLOCATION WINDOW	CLOUD DETECTION APPROACH	EMISSIVITY ANALYSIS
Buehler et al., 2004	NOAA/ AMSU-B	TA (radius=50km)	Radiosonde launch time	3 hours, 50 km	Threshold test on 183.31±7.00 GHz	Emissivity fix at 0.95 in Lindenberg area
John et al., 2005	NOAA/ AMSU-B	TA (radius=50km)	Synoptic Time -30 minutes	3 hours, 50 km	-	-
Cherny et al., 2010	METEOR-M/MTVZA-GY	TA; test on WV content and Wind speed		10x10 pixel and more	-	-
Moradi et al. 2010	NOAA/ AMSU-B, MHS	TA (radius=50 km)	Radiosonde launch time + 45 minutes	2 hours,50 km	Threshold test on 183.31 GHz-band BT (Buehler et al. 2007 Burns et al. 1997)	-
Moradi et al., 2013a	NOAA/ AMSU-B, MHS	TA (radius=50 km)	Radiosonde launch time + 45 minutes	2 hours,50 km	Threshold test on 183.31 GHz-region BT (Buehler et al. 2007 Burns et al. 1997)	Emissivity screening: only profiles with TPW>5kg/m^2,TPW>10kg/m^2 and TPW>30 kg/m^2 were used to simulate BT at 183.31 ± 1.00 GHz, ± 3.00 GHz and ± 7.00 GHz, respectively.
Moradi et al., 2013b	NOAA/ AMSU-B, MHS	TA (radius=50 km)	Radiosonde launch time + 30 minutes	2 hours, 50 km	Threshold test on 183.31 GHz -region BT (Buehler et al. 2007, Burns et al. 1997)	Emissivity screening: only profiles with TPW>5kg/m^2 were used to simulate BT in 183.31 GHz band.
Clain et al., 2015	MT/ SAPHIR	Single closest pixel	Radiosonde launch time	45 minutes, 50 km	Hong et al. (2005) cloud screening method; analysis of radiosonde observations.	Emissivity taken from Prigent et al. (2006, 2008) or calculated internally to RTTOV by Fastem 4.
		TA		45 minutes, 5x5 pixel		
Bobryshev et al., 2018	NOAA/MHS, ATMS MT/SAPHIR	TA (radius=50 km)	Radiosonde launch time	3 hours, 50 km	Cloud Mask product from MSG/SEVIRI (Finkensieper et al. 2016)	
Sun et al., 2010	COSMIC/RO	SC		7 hours, 250 km	-	-
Reale et al., 2012	NPROVS 12 satellite sounding product	SC	Radiosonde launch time plus 45 minutes	6 hours, 250 km	Ancillary data analysis	-
Moradi et al., 2015	SNPP/ ATMS and MT/SAPHIR	SC in TA (radius=50 km)	Radiosonde launch time	1 hour, 50 km	Cloud screening from water vapor channels and IWP information	Emissivity screening: lower tropospheric and window ATMS channels not used for bias statistics over land.
Calbet 2016, Calbet et al. , 2010	METOP/ IASI	SC	Radiosonde launch time	5 minutes before the satellite overpass, IASI IFOV	-	-
Calbet et al., 2017	METOP/ IASI	SC	Radiosonde launch time	30 minutes, 25 km	IASI l1c cloud mask + AVHRR cloud mask	-
Sun et al., 2017	SNPP/CrIS, ATMS	SC	-	6 hours, 150 km	-	-
Di Paola et al., 2018	SNPP/ATMS	SC	-	10 minutes	-	-
Zhao et al., 2022	AQUA, TERRA/ MODIS	SC	-	4 hour	-	-

The SC approach compares the satellite measurements with the simulated observations obtained by applying the RTM to the closest RS and assuming a quasi-vertical ascent of the radiosonde without considering the horizontal drift of the radiosonde.
On the other hand, TA collocation strategy is used to account for balloon drift whose typical values are around 50 km depending on several factors. Some studies such as Buehler et al. (2004), John et al. 2005, Moradi et al. (2010), Moradi et al. (2013a,b), Bobryshev et al. (2018), compared radiosounding simulated BT with the average satellite BT of a TA centered at the radiosonde launch site with a radius of 50 km. Buehler et al. (2004), Moradi et al. (2010), Moradi et al. (2013a,b) also estimated the displacement of

the air mass during the time interval between target time and satellite overpass multiplying it by the average wind speed between 700 hPa and 300 hPa. The match-ups with displacement larger than 50 km were discarded. Moreover, in order to estimate the atmospheric inhomogeneity of the TA, Buehler et al. (2004) compared the Standard Deviation (SD) of the BTs of the FOVs included in the TA with the noise equivalent temperature (NEΔT) and filtered out the TA with SD>NEΔT. Moradi et al. (2010) used the inverse of the SD (so that a small weight equals a high SD) to weight each match-up for calculating statistics.
Many of the works listed in Table 2.1 applied cloud screening methods to filter out cloudy match-ups. Furthermore, in order to avoid the satellite measurements affected by the land-surface emissivity, in many cases emissivity analysis was also proposed.

Calibration coefficients are provided by the comparisons of the radiance measured by the sensor with the RT simulations from the radiosonde data. However, the methodology of directly comparing the observed BT from the target sensor with simultaneous measurements from similar channels on radiosonde sensors is prone to different sources of uncertainties, even though it is the most direct and, potentially, most accurate in-flight calibration method (Slater et al., 1996). A crucial point of this methodology is the evaluation of the uncertainties associated with all the contributing sources.
The availability of the satellite and radiosonde measurement uncertainties is fundamental to establish if the measurements and simulations agree within the uncertainty limits (Immler et al., 2010). So, a detailed analysis of the sources of uncertainty is the first, and often most important, step to improve the accuracy of satellite observations. However, once the sources have been identified, the quantification of uncertainties is probably the most difficult task. As stated by Calbet et al. (2017), it is unlikely that all the sources can be fully characterized, which somehow prevents a full metrological closure. Several papers are available in the literature addressing different aspects of the problem, providing an estimate of the uncertainty, sometimes through a deep investigation, more often with just a crude guess. The uncertainties issue will be fully treated in subsection 4.6.

3. Data description

VICIRS tool compares MWI/ICI BT with BTs simulated from radiosonde profile by the GRUAN processor (Carminati et al., 2019) adapted to the purposes of VICIRS (v6.3.b.1.0.0) (further detail in subsection 5.2.3). RSs are collected from two archives, GRUAN and RHARM, which are described in the following subsections.

3.1 GRUAN archive description

Homogenous upper-air data records with quantified uncertainties are the ideal candidate for the MWI/ICI calibration plan. As such, the primary dataset for the MWI and ICI calibration is the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN), which provides homogeneous and fully traceable upper-air measurements with quantified uncertainties. GRUAN was established in 2006 (Bodeker et al, 2015) and it provides long-term, high-quality radiosounding data at several sites worldwide. Although the GRUAN network includes 31 sites, data from 18 sites are routinely archived, of which 14 sites are certified to date. GRUAN measurements include uncertainties and are traceable to the SI international or other accepted standards, providing extensive metadata and comprehensive documentation of measurements and algorithms (Dirksen et al., 2014, Von Rhoden et al., 2021). Moreover, GRUAN includes the balloon position at all pressure/height levels using GNSS positioning. GRUAN is currently providing three radiosonde data products for four different types of radiosondes (Vaisala RS92 and RS41, Meisei iMS-100 and RS11-G). Although GRUAN represents the highest quality radiosounding product available at global scale, its spatial coverage may be insufficient for the purpose of ensuring calibration in various climate regimes and orography conditions. Therefore, an additional dataset recently provided within the Copernicus Climate Change Service (C3S), named Radiosounding HARMonization (RHARM, Madonna et al., 2022), was considered.

3.2 RHARM archive description

Building on the GRUAN expertise and WMO radiosonde intercomparison data, RHARM provides adjusted radiosounding observations of temperature, humidity and wind with estimated uncertainties at 700 stations, plus launches from a number of ships. The RHARM algorithm mimics the GRUAN procedure to process RS92 sonde types (other types are under implementation). RHARM also uses the 2010 WMO/CIMO radiosonde intercomparison data set to adjust the bias and estimate measurement uncertainties for several radiosonde types not covered at present by the GDPs.
The RHARM dataset includes twice daily (0000 and 1200 UTC) bias-adjusted radiosonde data at pressure levels in the range 1,000–10 hPa, from 1978 to present, using as input data source the Integrated Global Radiosonde Archive (IGRA), provided and maintained by the NOAA-NCEI. The applied adjustments are interpolated to all reported levels, when these are provided in IGRA and in the high-resolution BUFR files made available at a larger number of stations since 2014, the latter used an additional data and metadata source provided directly from ECMWF. RHARM is the first data set to provide homogenized time series with an estimation of the observational uncertainty at each sounding pressure level. By construction, RHARM adjusted fields are not affected by cross-contamination of biases across stations and are fully independent of reanalysis

data. In the upcoming new version of RHARM, also the exact balloon position at all pressure/height levels will be derived from its latitude and longitude estimations, obtained from the GNSS signal, or from the wind data.

Currently using IGRA version 2 dataset as input, the RHARM data set inherits the IGRA quality assurance procedures (Durré et al., 2008). Nevertheless, RHARM applies additional quality checks on: the metadata availability; physical plausibility; data completeness check; accuracy of the bias adjustment; removal of outliers; vertical correlation between structural breaks at the same station; coherency check for the adjustments applied at the significant levels.

The RHARM dataset is currently available via the Copernicus Climate Data Store (CDS) (https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-igra-baseline-network?tab=overview). The dataset is updated annually. In the near future, there is a plan to increase the monthly updates, depending only on the constraints applied by the CDS team, as the software generating the RHARM data can be operated in an operational fashion. This means that data can possibly be provided in NRT with a typical 1 day delay, the data update frequency for IGRA.

In general, the weaknesses of RHARM RS are (i) the coarse vertical resolution, (ii) the high RHARM top-pressure value and (iii) the lack of surface information. Section 5 in [AD-8] provides a quantification of these weaknesses on the simulated BT by comparing BT simulated from RHARM RS with BT simulated from GRUAN RS. The new version of RHARM (v2) is expected to improve the vertical resolution and top pressure, i.e. (i) and (ii).

3.3 Numerical Weather Prediction retrieve

The NWP data can be used in the GRUAN processor to simulate the BT and the associated uncertainty from RS profiles as a complement for missing fields in the RS datasets (such as surface parameters or data over the RS top level). NWP is mandatory in the framework of the multi-source correlative analysis.

The NWP used for the study is the ECMWF Integrated Forecasting System at highest spatial resolution (called HRES, with horizontal grid spacing of about 0.125°). ECMWF files are downloaded from the Meteorological Archival and Retrieval System (MARS) through the ECS system, which is accessible exclusively to registered users from Member and Co-operating States.

The characteristics of NWP files are set in a batch script: *date* and *time* select the simulation temporally closest before the RS launch time available in the archive, *step* indicates the hours of forecasts, the first 15 hours of forecast are downloaded with a step of 3 hours. A 2°x2° square (about 16*16 grid points) around the launch site is chosen as *area* and the profile variables are downloaded on the 137 model levels (*lev*). The single level atmospheric fields selected are sea ice area fraction (131), geopotential (129), surface pressure (134), 10-meter U wind component (165), 10-meter V wind component

(166), 2-meter temperature (167), 2-meter dew point temperature (168), land-sea mask (172), and skin temperature (235), and total cloud cover (164). The model level atmospheric profiles downloaded are temperature (130), U component of wind (131), V component of wind (132), specific humidity (133), logarithm of surface pressure (152), fraction of cloud cover (248), and ozone mass mixing ratio (203).
MARS archive contains the forecast runs out to 10 days based on the 00/12 UTC analysis forecast, while the 06 and 18 forecast runs are not archived.
The size of the single NWP file containing all the 15 hours of forecast is about 6-8 Mb and it is usually downloaded in 5-10 min depending on the crowding condition of the MARS system. The procedure to download the NWP files is completely automatic in the VICIRS tool.

4. VICIRS methodology: collocation criteria and match-up analysis

This section deals with the methodology as basis of VICIRS tool. In detail, subsection 4.1 describes the approach adopted for spatial and temporal collocation, while subsections 4.2 and 4.3 show the cloud screening tests applied to MWI/ICI observations and to RS, respectively. In subsection 4.3 the quality test applied to RS and the AMD test are also described. Subsection 4.4 analyzes the effect of land surface emissivity (LSE) on MWI/ICI channels observations. The brightness temperature simulation from RS and NWP are described in subsection 4.5. Subsection 4.6 deals with the source of uncertainties and analyzes them for VICIRS purposes. Finally subsection 4.7 describes the multisource correlative methodology (MCM) analysis for RS, NWP and MWI/ICI that is based on triple collocation methodology and allows for the characterization of the error structure of the three spatially and temporally collocated measuring systems.

4.1 Data spatial and temporal collocation

The spatial collocation criteria adopted in VICIRS tool is based on the TA approach (section 2). This approach is preferred to the single closest one because it takes in account the radiosonde drift and it allows to minimize the representativeness error due to the spatial and temporal collocation by applying homogeneity test and AMD test.
Generally, TA is a circle with a radius of 50 km centered at the radiosonde launch site (Buehler et al. 2004, Moradi et al. 2010, Bobryshev et al. 2018) with the related BT (hereinafter referred to BT_{TA}) defined as the average of the FOV-BTs included in it.

$$BT_{TA} = \sum_{i=1}^{N} BT_i \cdot \lambda_{i,j}$$

where N is the number of the MWI/ICI FOVs included in the TA and $\lambda_{i,j}$ is the weight for the i^{th} FOV:

$$\lambda_{i,j} = \frac{d_{0i}^{-j}}{\sum_{i=1}^{N} d_{0i}^{-j}}.$$

5 types of TA are obtained by modifying TA definition and by varying j $\lambda_{i,j}$ (the larger *j* is, the larger the weight of the FOVs closest to the launch site):

- **TA type 1: circular TA** where BT_{TA} is determined as the average of the BT of the FOVs included in TA (**j=0**);
- **TA type 2: circular TA** where BT_{TA} is determined for **j=1** (inverse distance weight);
- **TA type 3: circular TA** where BT_{TA} is determined for **j=2** (inverse squared distance weight);
- **TA type 4: RS-driven exact TA** as the set of FOVs closest to each pressure level of the radiosonde path (nearest neighbour approach), BT_{TA} is determined for **j=1;**
- **TA type 5: 3x3 RS-driven TA** as the set of the 9 FOVs closest to each RS pressure level, consisting of the FOV closest to the RS pressure level + the 8 FOVs closest to it.

The *TA-radius* is set equal to the maximum *sonde-drift* if sonde-drift<=50 km, otherwise *TA-radius*=50 km.

The **RS-driven TA** (types 4 and 5) can be used only when the latitude and longitude are available for each pressure level (GRUAN RS). The number of FOVs included in the RS-driven TA is lower than that in the circular TA. Note that, as such, different TA types related to the same match-up may be classified differently, because they may contain a different number of cloudy FOVs. As a result, the number of match-ups may be higher when RS-driven TA is chosen for statistics. Further details are given in subsection 5.2.2.

The temporal collocation will consider three options for the temporal distance between the sonde launch time and the satellite overpass (ΔT):

1. -15'≤ ΔT ≤45'
2. -1 hour ≤ ΔT ≤1 hour
3. -3 hours ≤ ΔT ≤3 hours.

The 3 temporal options are all configurable in the VICIRS tool (*config.ini* for collecting match-ups and *query.ini* for match-up and bias/uncertainty analysis).

Examples of TA examples for Potenza (POT) GRUAN site and spatially collocated MWI L1B observations at (183.31+/-4.9) GHz are shown in Figure 4.1.1.

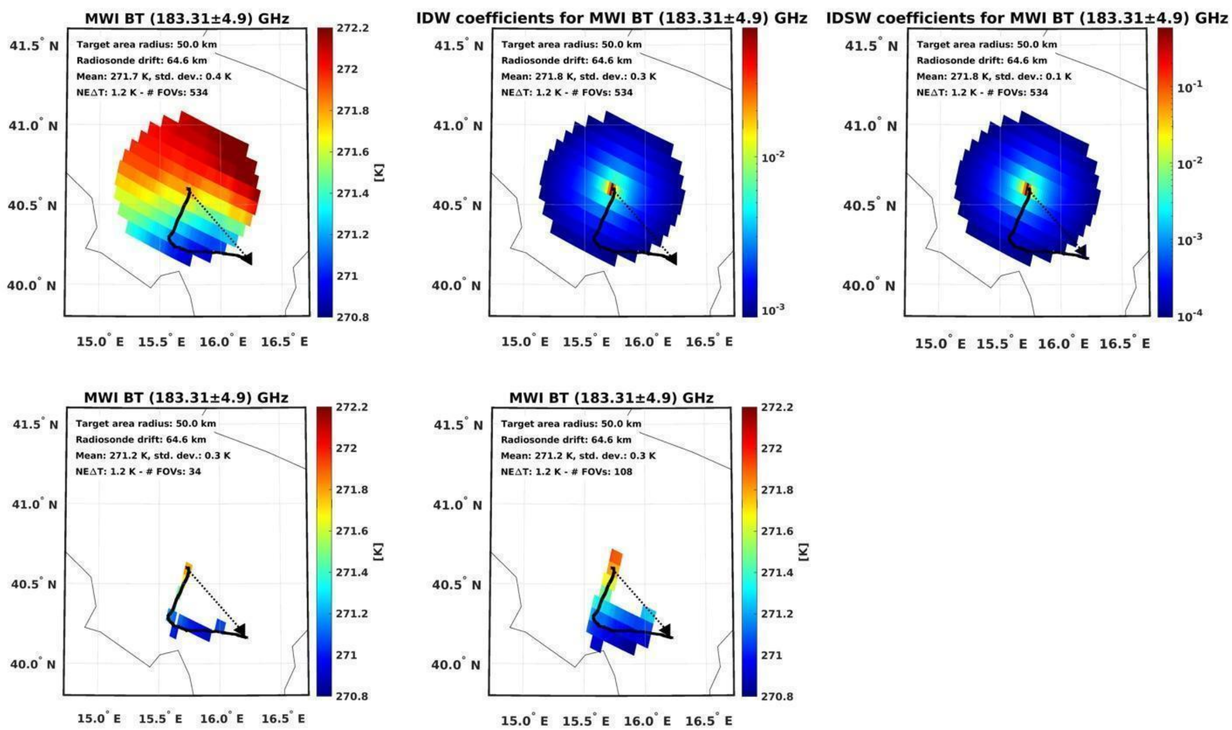


Figure 4.1.1 Potenza (POT) GRUAN site/MWI spatial match-up, TA examples: from top to bottom and from left to right example of **classical TA** (TA type 1.a, 1.b and 1.c), **RS-driven TA** (TA type 4 and 5), for sounding channel at (183.31+/-4.9) GHz. Note that the color-bar in top-left and bottom panels indicates BT (K), while it indicates distance weights in the top-center and top-right panels.

4.2 Cloud screening methodology for MWI and ICI

The MW cloud mask threshold tests available in literature (some of them are listed in Table 2.1) have been adapted and applied to the simulated MWI and ICI level 1B observations. The threshold tests used for MWI and ICI are listed in Table 4.2.1 and 4.2.2, respectively.

In detail, ICI and MWI Test 1 is based on the threshold test proposed for 183.31.31$\pm$1 GHz and 183.31.31$\pm$7 GHz AMSU-B band from Buehler at al. (2007). The AMSU-B channel at 183.31$\pm$1 GHz is not present in ICI/MWI and thus it is replaced with MWI/ICI 183.31$\pm$2 GHz in the tests listed in Tables 4.2.1 and 4.2.2. The threshold for ICI/MWI Test 1 (a and b) was chosen equal to the AMSU-B threshold at 44.55° viewing angle (Table 1 in Buehler et al (2007)), which is the closest to the MWI/ICI offset angle (44.82°, refer to pdf_science_epssg_mwi_ici_plan.pdf).

MWI-Test 1 and MWI-Test 2 related to 89 GHz and 165 GHz frequencies, respectively, are based on threshold tests proposed by Yaping et al. (2018). Yaping et al. (2018) tested the proposed-criteria to detect deep convective clouds on BT at 89 GHz (v-polarization) and 165 GHz (instead of AMSU-B 150 GHz frequency) and selected the 89 GHz and 150

GHz threshold values from a BT 3-year-dataset of AMSU-B observations acquired in summer.

MWI Test 2 on 89 GHz frequency (vertical (v), horizontal (h) polarization) and ICI Test 1 on 664 GHz (v and h polarization), are based on the study of Gong and Wu (2017). By examining the ice cloud scenes identified by the test on 183.31 GHz band (identified by the "3σ method" proposed by Gong and Wu (2017)), they found that the scattering by frozen particles was highly polarized, with v–h polarimetric differences (PD) being positive. In particular, the PD amplitude for 166 GHz and 89 GHz peaks at about 10 K in the tropics and it increases slightly with latitude, both over sea and land (Figure 4 in Gong and Wu, (2017)). They observed small values for the difference between vertical and horizontal polarized frequency along deep convective lines and higher values in the anvil and stratiform precipitation region, with more evidence at 166 than 89 GHz frequency, because of the increasing contribution of ice-particle scattering at the higher MW frequencies. The higher values of PD in the anvil and stratiform precipitation areas are due to the low multiple scattering process under relatively low turbulent conditions that saturates the polarization signatures.

Table 4.2.1 Cloud tests for MWI

MWI	183 GHz frequency	89 Ghz frequency	165 Ghz frequency
Test 1(a)	$(a) BT_{183.31\pm 2GHz} < 235.2\ K\ and\ (BT_{183.31\pm 7.0GHz} - BT_{183.31\pm 2GHz}) < 0$ (Buehler et al., 2007)	$BT_{89GHz,v} < 240\ K$ (Yaping et al., 2008, over land)	$BT_{165GHz} < 220\ K$ (Yaping et al., 2008)
Test 1(b)	$(b)\ BT_{183.31\pm 2GHz} < 235.2\ K\ and\ (BT_{183.31\pm 3.4GHz} - BT_{183.31\pm 2GHz}) < 0$ *(Buehler et al. 2007)*		
Test 2	$(BT_{183.31\pm 2.0GHz} - BT_{183.31\pm 7GHz}) \geq 0\ and\ (BT_{183.31\pm 2.0GHz} - BT_{183.31\pm 3.4GHz}) \geq 0\ and\ (BT_{183.31\pm 3.4GHz} - BT_{183.31\pm 7GHz}) \geq 0$ *(Hong et al. 2005, to detect deep convective clouds)*	$1 < BT_{89GHz,v} - BT_{89GHz,h} < 5\ K$ $and\ BT_{89GHz,v} < 265\ K$ *(over land);* $BT_{89GHz,v} - BT_{89GHz,h} \leq 20$ *(over sea)* *(based on Gong and Wu, 2017)*		
Test 3	$(BT_{183.31\pm 2.0GHz} - BT_{183.31\pm 7GHz}) \geq (BT_{183.31\pm 2.0GHz} - BT_{183.31\pm 3.4GHz}) \geq (BT_{183.31\pm 3.4GHz} - BT_{183.31\pm 7GHz}) > 0$ (Hong et al. 2005, to detect convective overshooting)		
Test 4	$BT_{183.31\pm 2.0GHz} > BT_{183.31\pm 3.4GHz} > BT_{183.31\pm 4.9GHz} > BT_{183.31\pm 6.1GHz} > BT_{183.31\pm 7GHz}$ (based on Clain et al. 2005)		

Since the scattering between ice hydrometeors and radiation induces a remarkable polarization signature strongly dependent on the size, shape, and orientation of non-spherical ice hydrometeors, a more comprehensive discussion of this topic is needed and can be found in Barlakas et al. (2021). Moreover, at 89 GHz the difference between v and h polarization is more sensitive to signals from the underlying surface, especially from sea. To investigate the cirrus clouds (ice clouds with small ice-crystals), they considered the higher frequency at 640 GHz, acquired by the NASA airborne Compact Scanning Submillimeter-wave Imaging Radiometer (CoSSIR) (Evans et al., 2005). Approximately, they found a value of 10 K for the peak of difference between BT at v and h polarization occurring at 220 K for 89 GHz and at 200 K for 640 GHz.

Table 4.2.2 Cloud tests for ICI

ICI	183 GHz frequency	664v GHz frequency
Test 1(a)	$BT_{183.31\pm2GHz} < 235.2\ K\ and\ (BT_{183.31\pm7.0GHz}$-$BT_{183.31\pm1GHz}) < 0$ *(Buehler et al., 2007)*	$BT_{664v\ GHz} < 220\ K$ *(based on Gong and Wu, 2017)*
Test 1(b)	$BT_{183.31\pm2GHz} < 235.2\ K\ and\ (BT_{183.31\pm3.4GHz}$-$BT_{183.31\pm1GHz}) < 0$ *(Buehler et al., 2007)*	
Test 2	$(BT_{183.31\pm2.0GHz}$-$BT_{183.31\pm7GHz}) \geq 0\ and\ (BT_{183.31\pm2.0GHz} - BT_{183.31\pm3.4GHz}) \geq 0\ and\ (BT_{183.31\pm3.4GHz} - BT_{183.31\pm7GHz}) \geq 0$ *(Hong et al., 2005, to detect deep convective clouds)*	$BT_{664v\ GHz} < 225\ K\ and\ 0 \leq\ BT_{664v\ GHz} - BT_{664h\ GHz} < 15K$ *(based on Gong and Wu, 2017)*
Test 3	$(BT_{183.31\pm2.0GHz}$-$BT_{183.31\pm7GHz}) \geq\ (BT_{183.31\pm2.0GHz} - BT_{183.31\pm3.4GHz}) \geq\ (BT_{183.31\pm3.4GHz} - BT_{183.31\pm7GHz}) > 0$ *(Hong et al., 2005, to detect convective overshooting)*	
Test 4	$BT_{183.31\pm2.0GHz} > BT_{183.31\pm3.4GHz} > BT_{183.31\pm7GHz}$ *(based on Clain et al., 2015)*	

The cloud detection results for a mid-latitude sub-region of the second summer orbit (n. 4656 from 20070912102225 to 20070912120114) are shown in Figure 4.2.1 and 4.2.2 for MWI and ICI, respectively. The MSG-SEVIRI natural RGB composition, spatially and temporally corresponding to the considered sub-region, is shown in order to reveal the cloudy-areas distribution.

In comparing the cloud test results with MSG-SEVIRI RGB, it is important to note that:

- the ERA-5 data used for the test data simulation have a horizontal resolution of about 30 km and this poses some limitations in interpreting the ICI/MWI observations whose spatial resolution is 16 km for ICI and ranges from 50 km to 10 km for MWI;
- for the simulation of the considered sub-region, the 2007-09-12 10:00 UTC ERA-5 data have been used, while the MSG-SEVIRI observations have been acquired on 2007-09-12 10:15 UTC.

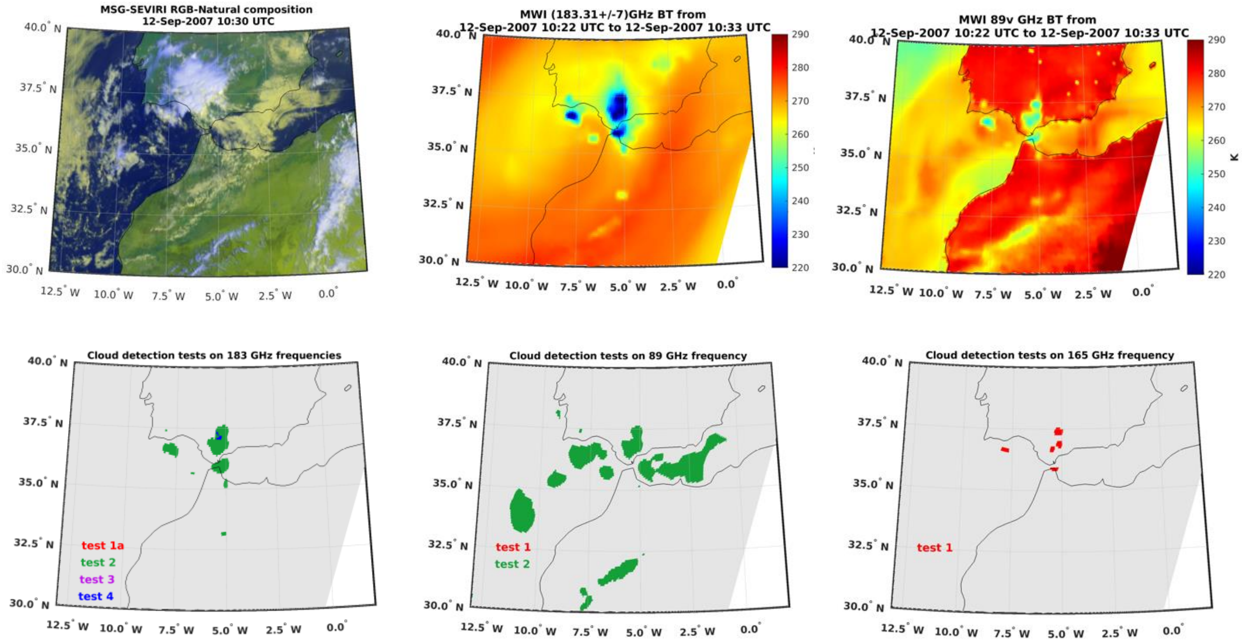


Figure 4.2.1 Example of cloud detection test result applied to the MWI level 1B simulated observations (from second summer orbit n. 4656 from 20070912102225 to 20070912120114). From left to right, in the first row: MSG-SEVIRI natural RGB composition, MWI 183±7.0GHz BT image, MWI 89 GHz BT image; in the second row: cloud-mask results obtained by tests on 183 GHz, at 89 GHz and at 165 GHz frequency, respectively.

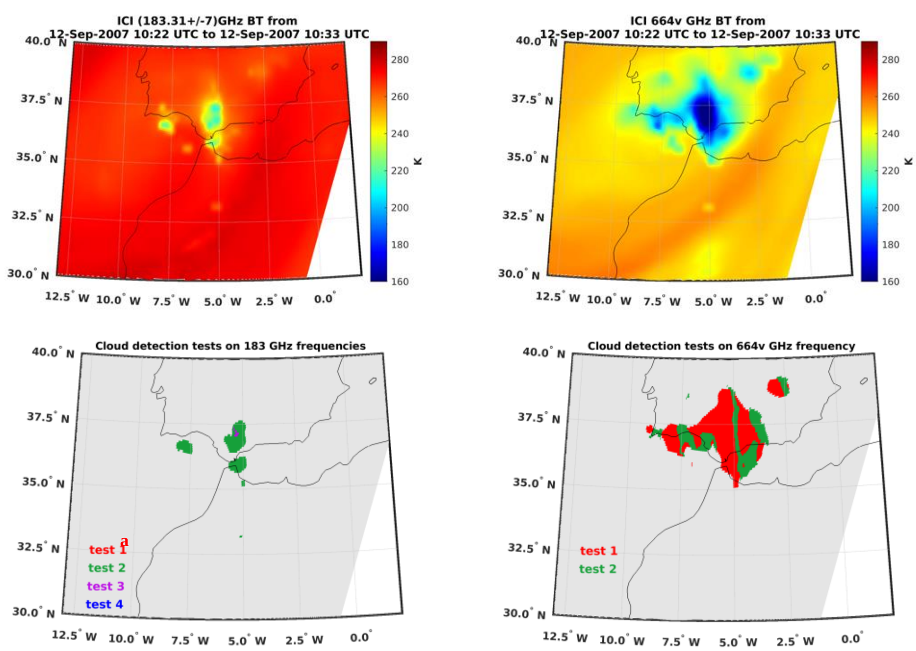


Figure 4.2.2 Example of cloud detection test result applied to the ICI level 1B simulated observations (from second summer orbit n. 4656 from 20070912102225 to 20070912120114). From left to right, in the first row the ICI 183±7.0GHz BT image and the ICI 664v GHz BT image; in the second row the results of the test applied to the 183 GHz frequency BTs and their combinations and to the 664v,h GHz BT, respectively.

The VICIRS tool applies all the tests listed above for MWI and ICI to all the FOVs included in the TA and it determines the percentage of cloudy FOVs. The percentage of cloudy FOVs detected by each test, as well as the maximum value among them, are stored in the match-up output file (for more details see subsection 5.5.2 and Table A.2). Note that in querying the match-up dataset, the user can select the maximum allowed value for TA cloudiness, which refers to the maximum cloudy percentage.

4.3 Radiosounding analysis and quality check

The RS analysis is mandatory to check whether the related match-up is usable for calibration purposes. The quality check of RS is done in terms of numbers of levels (*nlev*) and minimum pressure (*Pmin*) value, air mass displacement (AMD) and cloud contamination.
In detail, RS is considered for the calibration process when:

- *nlev*≥ 40 for P/T/RH profiles and the related uncertainties are available for each profile and for all the pressure levels;
- *Pmin*≤10hPa;
- RS is in clear sky. The presence of cloudy layers is verified by comparing the RH values with the reference values for clear sky as determined by Zhang et al. (2010). The RH-test outputs the number of levels contaminated by low, middle and high clouds. Note that this method was developed for RS92 sonde, but its performances with RS41 has been assessed by validating the RH-test results against the RS-*synopclouds* information for the RS measured in year 2022 from LAU, GVN, LIN, PAY and SNG GRUAN sites. The RS data files for these GRUAN sites are provided with cloud presence information (non-empty *SynopClouds* attribute). The quantitative evaluation of the tests for cloud detection was performed through a dichotomous statistical assessment in terms of Probability of Detection (POD), False Alarm Ratio (FAR), bias and accuracy, the bias being the deviation RH-test results minus the value of RS-SynopClouds attribute. The statistical scores are listed for the GRUAN sites in Table 4.3.1.
- AMD $\leq TA\,radius$. AMD is determined by multiplying the temporal distance between the satellite overpass and sonde launch, Δt, by the wind speed average between 700 hPa and 300 hPa, $\underline{w}$ (following Buehler et al. 2004, Brobyshev et al. 2017, Moradi et al. 2010). AMD test is used to reduce the variability caused by the horizontal inhomogeneity of the atmosphere.

Table 4.3.1 Dichotomous statistical scores between cloud detection methods (RH-test vs RS-SynopClouds).

GRUAN site	POD (%)	FAR (%)	Bias	Accuracy	Number of RS used for statistics
LAU	94	8.62	1.03	0.87	826
GVN	99	10.4	1.10	0.89	363
LIN	95	7.98	1.03	0.88	1601
PAY	88	2.16	0.89	0.87	745
SNG	100	0.00	1.00	1.00	17

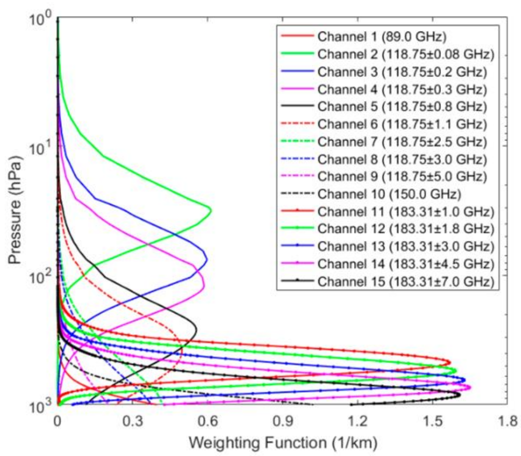
When the RS is classified as clear sky and the NWP option (NWP_opt>0) is activated by the user, the NWP profile is checked for cloud contamination, otherwise it is not considered for the calibration process. When NWP_opt>0 , the NWP profile spatially and temporally closest to the radiosonde launch is checked for the presence of low, medium or high clouds by examining the NWP field *lcc*, *mcc* and *hcc*, indicating respectively the fraction of low/medium/high cloud cover in the NWP profile. When the NWP profile is cloudy, it is not considered in the process, which will continue considering only RS (NWP_opt=1), unless the user has decided to discard the related match-up (NWP_opt=2, as detailed in subsection 5.2.1).

4.4 Emissivity screening considerations

Among the geophysical inputs required by the radiative transfer model to simulate MW BTs from the radiosonde profiles, the land surface emissivity needs a particular attention because of the complexity to model it.

Surface emissivity models, such as the TELSEM2 (Wang et al., 2017) and TESSEM2 (Prigent et al., 2017), are distributed with the current version of RTTOV. In detail, TESSEM2 provides parameterized sea surface emissivity, TELSEM2 provides parameterized land, snow and sea-ice surface emissivity. Both models have been extended with respect to their previous versions to cover the range up to 700 GHz and they are suitable for MWI and ICI simulations. Since RTTOV v13.2 (released in December 2022) a new option is available, namely the SURface Fast Emissivity Model for Ocean (SURFEM-Ocean). SURFEM-Ocean (Kilic et al., 2023) is a fast neural network parameterization of the PARMIO physical reference emissivity model simulating all Stokes components for channels in the range 0.5 – 700 GHz (Dinnat et al., 2023). To avoid uncertainties due to incorrect parameterization of LSE for the frequencies affected by LSE, the emissivity analysis will be applied to the clear sky match-ups considered in the calibration process.

The required performances for MWI and ICI are listed in Tables 4.4.1 and 4.4.2. In general, the observations at ICI frequencies (Table 4.4.2) are not affected by LSE because of the high atmospheric opacity. However, in very dry conditions (e.g. subarctic-winter conditions), few channels are sensitive to the surface contribution. This is the case of (i) the three outermost 183 GHz channels, (ii) the channels at 243 GHz, and (iii) the outermost 325 GHz channel (Buehler et al. 2012).
The RS from GRUAN sites on island (i.e., Tenerife (TEN), Graciosa (GRA) and Minatorishima (MIT)) will be used for the MWI channels affected by LSE. Most likely, these correspond to the first four channels (18.7, 23.8, 31.4, and 50.5 GHz), as well as 52.61, 53.24, 53.75, 89 GHz and 165.5 GHz in high latitude, which could be differently affected by LSE as it figures out from the weighting function in Figure 4.4.1. Similarly, particular attention must be paid also to MWI less opaque frequencies in the 118.75 GHz absorption band (e.g. 118.75±3.2 GHz).



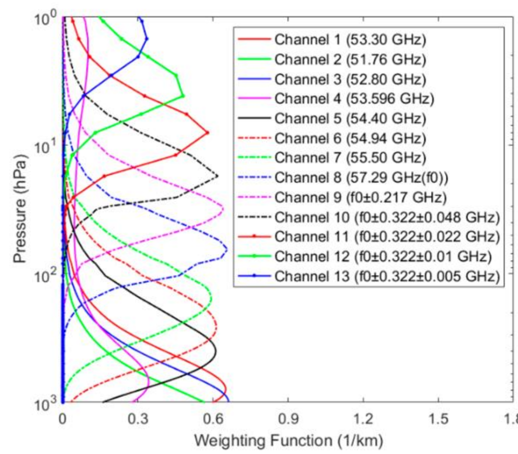


Figure 4.4.1. Weighting functions for 89-183 GHz channels in Microwave Humidity and Temperature Sounder (MWHTS) (a) and 51-58 GHz channels in Microwave Temperature Sounder II (MWTS II) (b) calculated from the U.S. standard atmospheric profile (He et al. 2022).

A useful test for screening MWI/ICI observations affected by LSE is the homogeneous test, consisting in comparing the SD_TA (Standard Deviation of the BTs included in TA) with the radiometric noise NEΔT and classifying as inhomogeneous the TAs with SD_TA > NEΔT (Buehler et al. 2004). The homogeneous-flag is an information provided by the TA analysis that does not cause the inhomogeneous TA to be removed. The NEΔT associated to the sample ($NE\Delta T_{sample}$) must be used, while $NE\Delta T$ in Table 4.4.1 for MWI and Table 4.4.2 for ICI is associated to the footprint ($NE\Delta T_{footprint}$). In fact, along-scan several samples within the footprint of each channel are taken (oversampling). As such, convolution of samples reduces the associated radiometric noise. $NE\Delta T_{sample}$ is simply derived from $NE\Delta T_{footprint}$ by considering the oversampling factor $1/K_{NE\Delta T}$, which is channel dependent (j indicates the channel index):

$$NE\Delta T_{sample} = NE\Delta T_{footprint}/K_{NE\Delta T}(j)$$

where:

$K_{NE\Delta T}(j) = \sqrt{\frac{T_{int}}{T_{int3dB}(j)}}$, T_{int} is the sample integration time and T_{int3dB} is the channel-dependent integration time over the 3dB antenna footprint.

T_{int} and T_{int3dB} values have been provided by EUMETSAT for MWI:
T_{int} = 0.394 ms

T_{int3dB} = [8.47 8.47 8.189 8.189 5.225 5.225 4.165 4.165 4.165 4.165 4.165 4.165 4.165 4.165 1.778 1.778 1.713 1.713 1.713 1.713 1.284 1.16 1.16 1.16 1.16 1.16] ms

and for ICI:
T_{int} = 0.663161278 ms

T_{int3dB} = [2.457 2.445 2.444 2.610 2.651 2.137 2.134 2.142 1.979 1.945 1.963 2.955 2.915] ms

These values are applicable to the test data used in the study, but they might be updated following the MWI and ICI evolution.

Table 4.4.1 MWI channels characteristics

Channel	*Frequency (GHz)*	*Bandwidth (MHz)*	*NEDT (K)*	*Radiometric Bias (K)*	*Polarization*	*Footprint Size at 3dB (km)*
MWI-1	18.7	200	0.8	1.0	V, H	50
MWI-2	23.8	400	0.7	1.0	V, H	50
MWI-3	31.4	200	0.9	1.0	V, H	30
MWI-4	50.3	180	1.1	1.0	V, H	30
MWI-5	52.7	180	1.1	1.0	V, H	30
MWI-6	53.24	400	1.1	1.0	V, H	30
MWI-7	53.750	400	1.1	1.0	V, H	30
MWI-8	89.0	4000	1.1	1.0	V, H	10
MWI-9	118.7503±3.20	2x500	1.3	1.0	V	10
MWI-10	118.7503±2.10	2x400	1.3	1.0	V	10
MWI-11	118.7503±1.40	2x400	1.3	1.0	V	10
MWI-12	118.7503±1.20	2x400	1.3	1.0	V	10
MWI-13	165.5±0.75	2x1350	1.2	1.0	V	10
MWI-14	183.31±7.0	2x2000	1.3	1.0	V	10
MWI-15	183.31±6.1	2x1500	1.2	1.0	V	10
MWI-16	183.31±4.9	2x1500	1.2	1.0	V	10
MWI-17	183.31±3.4	2x1500	1.2	1.0	V	10
MWI-18	183.31±2.0	2x1500	1.3	1.0	V	10

Table 4.4.2 ICI channels characteristics

Channel	***Frequency (GHz)***	***Bandwidth (MHz)***	***NEDT (K)***	***Radiometric Bias (K)***	***Polarization***	***Footprint Size at 3dB (km)***
ICI-1	183.31±7.0	2x2000	0.8	1	V	16
ICI-2	183.31±3.4	2x1500	0.8	1	V	16
ICI-3	183.31±2.0	2x1500	0.8	1	V	16
ICI-4	243.2±2.5	2x3000	0.7	1.5	V, H	16
ICI-5	325.15±9.5	2x3000	1.2	1.5	V	16
ICI-6	325.15±3.5	2x2400	1.3	1.5	V	16
ICI-7	325.15±1.5	2x1600	1.5	1.5	V	16
ICI-8	448±7.2	2x3000	1.4	1.5	V	16
ICI-9	448±3.0	2x2000	1.6	1.5	V	16
ICI-10	448±1.4	2x1200	2.0	1.5	V	16
ICI-11	664±4.2	2x5000	1.6	1.5	V, H	16

4.5 Brightness temperature simulation

The BT and its uncertainty simulated from RS profiles are computed using the GRUAN processor (Carminati et al., 2019) that has been developed to collocate GRUAN radiosonde profiles and NWP model fields, to simulate top-of-atmosphere BT at frequencies used by space-borne instruments, and to propagate GRUAN uncertainties in simulated BT. GRUAN processor has been modified for the purposes of VICIRS, starting a new branch, currently at version 6.3.b.0.1. This version uses the latest version of RTTOV (v13.2) and simulates BT from GRUAN/RHARM RS without using information from spatially and temporally collocated NWP profiles or using it. In detail, when the GRUAN/RHARM RS profiles are processed in combination with NWP data, the NWP information is used for:

- filling the gap of RS surface parameters. Among these, the skin temperature (T_{skin}) can be determined in two ways :
    - $T_{skin} = T_{2m} + (T_{skin}(NWP) - T_{2m}(NWP))$ where T_{2m} is determined from RS and $T_{skin}(NWP), T_{2m}(NWP)$ are T_{2m} *and* T_{skin} of the model;
    - $T_{skin} = T_{skin}(NWP)$;
- filling the gap of data over the RS top level;
- providing the ozone profile.

When NWP option is not activated

- surface parameter are determined directly from RS, considering the values closest to z=2m or to z=10m for wind components, where z is the *altitude field* in GRUAN RS and *geopotential_height field* in RHARM RS;
- the RTTOV reference ozone profile is used.

In the GRUAN processor an internal loop has been introduced on the different satellite angles to take into account that the zenith and azimuth angles vary for each ICI/MWI channel.

Regarding the emissivity model used in RTTOV, SURFEM-Ocean (Kilic et al, 2022) is chosen for sea. SURFEM-Ocean is a new microwave sea surface emissivity model available in RTTOV v13.2 valid across 0.5-700 GHz frequencies that should replace all FASTEM and TESSEM2 versions (https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs_rttov13/users_guide_rttov13_v1.2.pdf, Hocking et al. (2022b)). The emissivity model for land/ice is the TELSEM2 (Wang et al, 2016). *RTTOV v13 Users Guide* (2022) recommends TELSEM2 emissivity atlas instead of FASTEM land/sea-ice parameterization that will be deprecated in the future RTTOV versions.

4.6 Uncertainty analysis

Although the terms "error" and "uncertainty" are often treated as synonymous, in metrology they correspond to different definitions. According to the Vocabulaire International de Métrologie (VIM), published by the Joint Committee for Guides in Metrology (JCGM, 2012) of the International Bureau of Weights and Measures (BIPM), an error is defined as the measured value of a quantity minus the reference value of the same quantity. This error contains several components such as the instruments' error and a co-location error, and it can be either positive or negative. Conversely, the uncertainty is defined as a non-negative parameter characterizing the dispersion of the quantity values attributed to a measurand. Hence, the uncertainty quantifies the statistical properties of an ensemble of errors. In short, the term error is used for the deviation between a single value and the corresponding reference (unknown), while the term uncertainty indicates the statistical properties of these errors. Thus, the lack of exact knowledge of the value of the measurand forces the error to be characterized by a random variable, the uncertainty U. Note that the error is viewed as having two components, a random and a systematic one. Systematic errors may be fixed in time, or they may change slowly and can be dependent upon some operating conditions. The deviation of the measurement result from truth due to systematic errors defines the measurement bias (Immler et al., 2010). Although a correction may be applied to compensate for the systematic effect, there will still be a residual uncertainty associated with the correction. Following the "Guide to the expression of uncertainty in measurement" (JCGM/WG 1, 2008, GUM hereafter), it is important not only to correct for systematic effects but also to robustly ascertain and document the uncertainty of this correction. This level of knowledge of the systematic effects requires a detailed understanding of all aspects of the measurement. Assuming that proper corrections are made for all systematic effects, the expectation value of uncertainty U is zero. In this case, the uncertainty of the measurement result can therefore be expressed by one single value, the standard uncertainty u, which is the estimated standard deviation of the random variable U. All sources of uncertainty should be summarized to an uncertainty budget. The overall resulting uncertainty is calculated from independent sources of uncertainties according to the rule of uncertainty propagation:

$$u_y = \sqrt{\sum_{n=1}^{N} \left(\frac{\delta y}{\delta x_n}\right)^2 u_n^2 + 2\sum_{m=1}^{N} \left(\sum_{n=1(n\neq m)}^{N} \frac{\delta y}{\delta x_n} \frac{\delta y}{\delta x_m} u_{m,n}\right)} \qquad (4.6.1)$$

where $u_n^2 = u_{n,n}$ and $u_{n,m}$ $(n \neq m)$ indicate respectively the variance and covariance of input variables x_n. Thus, to derive the uncertainty budget rigorously, correlations between the different sources must be considered. In the simplest case of uncorrelated sources, the different uncertainty contributions can be summed quadratically.

Once the uncertainties have been evaluated and determined, two independent measurements can be cross-checked for consistency, which is achieved when the independent measurements agree within their individual uncertainties.
The vicarious calibration aims to cross-check two independent measurements m_1 and m_2 of the same measurand (e.g., the brightness temperature BT) with standard uncertainties u_1, and u_2, respectively. However, a compromise must be made between abundance of comparison pairs on the one hand, and on the other hand the spurious uncertainty due to several sources, such as non-perfect co-location in space and time between satellite and radiosonde measurements. Such uncertainty sources cannot be eliminated, so that the uncertainty budget of the vicarious calibration must account not only for the measurement uncertainties themselves but also for the other sources related to differences in sampling and smoothing of the inhomogeneous and variable atmospheric field. Calling σ the intrinsic uncertainties of the comparison (e.g. the colocation uncertainty), and assuming true the hypothesis that $m_1 = m_2$ with normally distributed uncertainties, the probability:

$$|m_1 - m_2| < k\sqrt{\sigma^2 + u_1^2 + u_2^2} \qquad (4.6.2)$$

depends on the coverage factor *k*, which determines an interval about the mean value as a multiple of standard uncertainty (Immler et al., 2010). If the results agree within *k*=1 the data are “consistent”, while within *k*=2 they are “in (statistical) agreement”. Conversely, if the results do not agree within *k*=2, the data are “significantly different”, while

"inconsistent" if the data do not agree within *k*=3. In such cases, it is very likely that a bias is present, i.e. an unaccounted systematic effect needs to be removed.
One way to visualize the uncertainty contributions is to draw a metrological uncertainty model chain, as recommended by the GAIA-CLIM project[1]. The metrological uncertainty model chain describes the flow diagram of the measurement process, including references to calibration, uncertainty sources, and linkages to reference standards. Figure 4.6.1 shows the uncertainty model diagram developed for the vicarious calibration in this study, using the module convention suggested by the GAIA-CLIM project. The next paragraphs introduce the uncertainty sources reviewed within the VICIRS study.

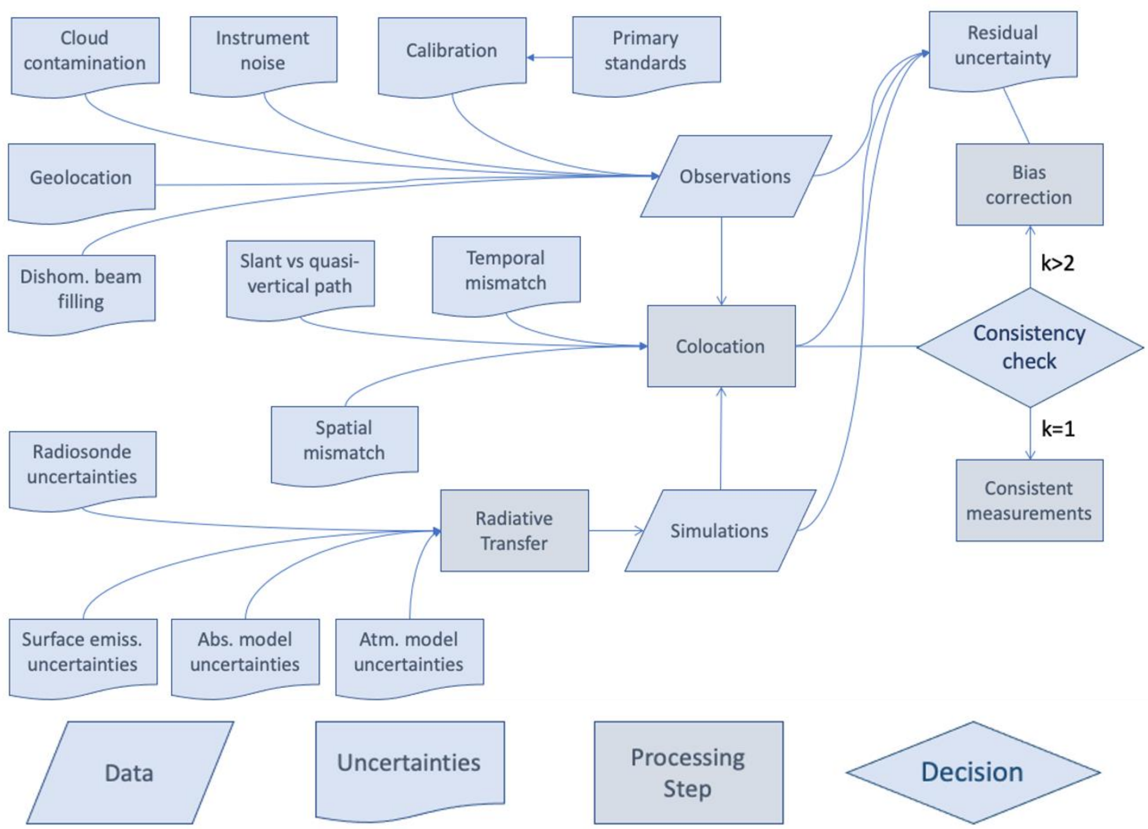


Figure 4.6.1 Top: Uncertainty model diagram for the vicarious calibration/validation using radiosoundings. Bottom: Legend for uncertainty model diagram blocks.

4.6.1 Review of uncertainty sources

A detailed analysis of the sources of uncertainty is the first and likely most important step to improve the accuracy of satellite observations. However, once the sources have been identified as above, the quantification of uncertainties is probably the most difficult task. Several papers are available in the literature addressing different aspects of the problem,

1 http://www.gaia-clim.eu/page/traceability-model-diagrams

providing an estimate of the uncertainty, sometimes through a deep investigation, more often with just a crude guess. Thus, the uncertainty sources have been reviewed within the VICIRS study, providing the estimated value when available, and attempting the estimate of the identified contribution that are not available.

The instrument's radiometric accuracy is one source of uncertainty. The instruments of interest, i.e. ICI and MWI, have been designed and built in response to the requirements set by the meteorological satellite user community, including the radiometric accuracy (low bias) and precision (high repeatability). The accuracy is obtained through the pre-flight and in-flight calibrations, which is somehow linked to primary or secondary metrological standards. Specifications on on-ground and in-flight instrument calibration are reported on EUMETSAT documents (such as EPS-SG Programme Overall Calibration and Validation Plan, ICI and MWI Calibration and Validation Plan) and references therein. In-flight deep space calibration data from roll maneuvers are part of Cal/Val activities, used for determining important parameters for the antenna pattern correction, such as spillover radiation and near-sidelobe radiation corrections. The instrument precision is characterized by the noise equivalent delta T (NE∆T). As reported in Table 4.4.1 and 4.4.2, Footprint-NE∆T depends on channel and ranges for MWI from 0.7 to 1.3 K while for ICI from 0.8 to 2.0 K (from [AD-2], [AD-3]).

Another source of uncertainty is from the **radiosonde measurements**, which are obtained by disposable temperature and humidity sensors, whose uncertainty is characterized by the manufacturer. However, common radiosonde measurements are not provided with the associated uncertainty, this being the case for example of the comprehensive Integrated Global Radiosonde Archive (IGRA). The radiosonde uncertainty is also independently characterized by GRUAN through laboratory and inflight tests, as described for RS92-SGP (Dirksen et al., 2014) and RS41 (Dirksen et al., 2020; von Rohden et al., 2022). The resulting uncertainty is provided within the operational GRUAN standard radiosonde products. Apart from GRUAN, also the RHARM data set provides an estimation of the radiosonde uncertainty at each sounding level for the adjusted twice daily (0000 and 1200 UTC) radiosonde data (700 radiosounding stations world-wide from 1978 to present). The radiosonde uncertainty profiles, either from GRUAN or RHARM dataset, are propagated through the GRUAN processor (Carminati et al. 2019) to compute the uncertainty associated with simulated radiosonde BT. Note that the GRUAN processor propagates GRUAN uncertainties in radiance space via perturbation of the temperature, humidity, and pressure profiles by plus and minus their uncertainty, thus assuming complete correlation of the uncertainties at all levels. This is a conservative assumption and the resulting uncertainty obtained in radiance space is therefore representative of a maximum uncertainty of the GRUAN component. The true GRUAN uncertainty in radiance space must be smaller than that calculated as such, as only a fraction of GRUAN uncertainty is really correlated over the entire profile. Comparing infrared observations with radiosonde simulations, Calbet et al. (2017) assumed no uncertainty correlation between GRUAN levels, although recognizing that such an assumption prevents full metrological closure, and that work is needed on estimating the full GRUAN uncertainty covariance matrices. During the 15th GRUAN implementation and coordination meeting (ICM-15, Bern, March 2024) a small working group has been established to address this knowledge gap.

At the core of the GRUAN processor is the fast radiative transfer code RTTOV. RTTOV is parameterized, in the sense that the atmospheric optical depths are computed from the thermodynamics of each layer through regression. The regression is trained and tested against channel-integrated spectrally resolved line-by-line (LBL) reference calculations, and thus the regression uncertainty contributes to the overall uncertainty. This **RT parameterization** uncertainty contribution is evaluated by the EUMETSAT NWPSAF as part of their continuous development of RTTOV, and thus was performed also for MWI and ICI channels in preparation for EPS-SG. The uncertainty contribution of RT parameterization is evaluated as the std of the differences between RTTOV and LBL BT computed for a set of diverse 83 profile set and six zenith angles, assuming constant unit surface emissivity, top-hat pass bands, and MW v13 predictor coefficient files[2].

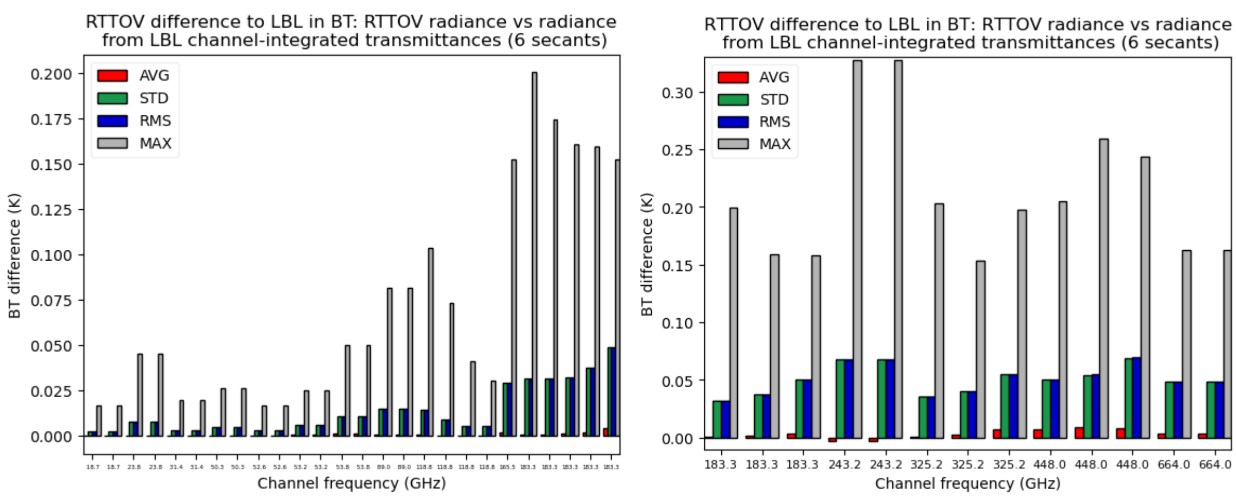


Figure 4.6.2 Uncertainty of simulated BT for MWI (left) and ICI (right) channels due to approximation of the parameterized radiative transfer code RTTOV with respect to the reference line-by-line code (source: NWP-SAF[3]).

[2] https://nwp-saf.eumetsat.int/site/software/rttov/download/coefficients/coefficient-download/#MW_optical_depth_coefs_and_RTTOV-SCATT_optical_properties
[3] https://nwp-saf.eumetsat.int/site/software/rttov/download/coefficients/comparison-with-lbl-simulations/#mw

Statistics are reported in Figure 4.6.2, showing that std are below 0.1 K for all MWI and ICI channels.

However, even the reference line-by-line **absorption model** is affected by uncertainty, due to the computational or experimental uncertainty underlying the adopted values of spectroscopic parameters. The evaluation of the absorption model uncertainty was first evaluated for down-welling radiation (Cimini et al. 2018; 2019) and recently extended to upwelling radiation in a much larger frequency range (16 to 700 GHz). The study is reported in Gallucci et al. (2024), while selected results are shown in Figure 4.6.3 for uncertainties at MWI and ICI channels due to 135 dominant water vapour and oxygen spectroscopic parameters. While also ozone contributes to line absorption in this range, the uncertainty was found negligible for ICI/MWI channels (<0.1 K). The simulated observation geometry mimics the observations from MWI and ICI, i.e., down-looking from top-of-the-atmosphere with 53° incident angle. The emissivity of a sea background (covering 72% of the globe) is considered, assuming typical conditions (8 m/s wind speed; 290 K sea surface temperature; 35 PSU salinity). The uncertainty has been evaluated for six typical climatology conditions (tropical, midlatitude summer, midlatitude winter, sub-arctic summer, sub-arctic winter, U.S. standard). The uncertainty on simulated BT for MWI and ICI has been calculated by convolving the uncertainty spectra at 50 MHz spectral resolution within the instrument first-order approximation bandpass filters (i.e., a rectangular box-average). Modeling the spectral response of the MW channels adds uncertainty due to lack of accurate knowledge of the filter **spectral response function** (SRF) and consequently the need to approximate the filter with idealized profiles, such as rectangular band pass. Buelher et al. (2004) evaluated this contribution, reporting BT differences well below 0.1 K for either rectangular or Gaussian band pass shapes.

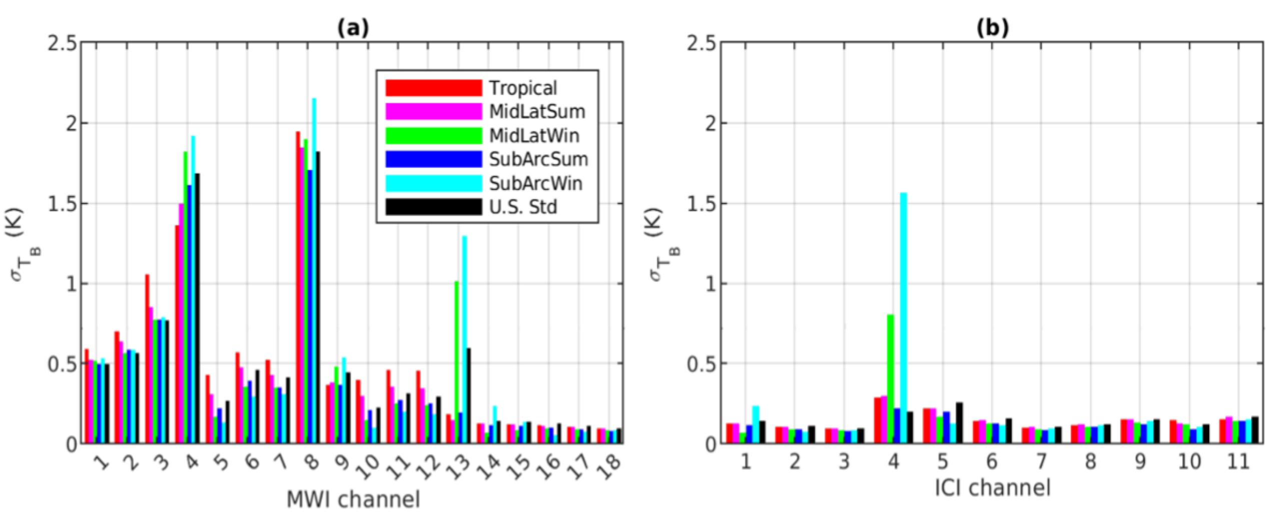


Figure 4.6.3 Estimated brightness temperature uncertainty due to absorption model convolved on MWI (a) and ICI (b) channel bandwidths for 6 typical climatology conditions (from Gallucci et al, 2024). Down-looking

view from top-of-the-atmosphere with 53° incident angle. Sea surface emissivity at typical conditions (8 m/s wind speed; 290 K sea surface temperature; 35 PSU salinity).

Vertical discretization also generates uncertainty as the naturally continuous atmospheric profiles are represented by discrete levels. As such, the uncertainty decreases as the profile vertical resolution increases. GRUAN radiosonde data are provided at original high resolution, i.e., thousands of levels from surface to below 30 hPa pressure. Conversely, RHARM data are available at much less pressure levels between 1000 and 10 hPa. As one can imagine, the low-resolution data provide far less detail than the high resolution. However, Buelher et al. (2004) reported that low-vertical-resolution data, as found in operational archives, are sufficient to accurately calculate satellite radiances, provided that low-resolution data are interpolated to a fine grid before calculating column quantities. They also report that the interpolation scheme is crucial: the best results are obtained when humidity is interpolated in relative humidity, whereas interpolation of humidity in volume mixing ratio leads to a large discrepancy between the two data sets. The 54 pressure levels used within RTTOV are deemed sufficient to accurately calculate satellite radiances and other path-integrated quantities. The vertical interpolation uncertainty was evaluated for channels of the Advanced Technology Microwave Sounder (ATMS) at the EUMETSAT NWPSAF (Hocking, 2014). These results were extended within the VICIRS study to MWI and ICI channels, most of which match very closely one ATMS channel. For the remaining MWI and ICI channels, that do not match closely one ATMS channel, the analysis extrapolated the vertical discretization uncertainty assuming a linear relationship with atmospheric opacity.

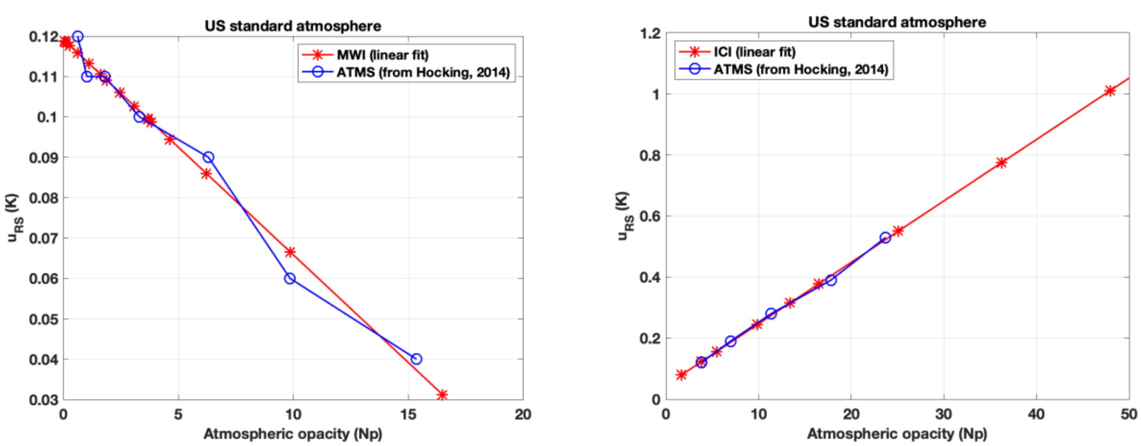


Figure 4.6.4 Estimated relationship between vertical discretization uncertainty and atmospheric opacity. Left: relationship estimated for MWI channels from ATMS oxygen channels. Right: relationship estimated for ICI channels from ATMS water vapour channels.

The relationship was estimated separately for MWI and ICI channels, computing atmospheric opacity from US standard atmosphere profiles, and considering both the cases in which the number of user levels is higher/lower than the number of RTTOV coefficient levels (Cases 1 and 2 in Hocking, 2014, respectively). Results are reported in Figure 4.6.4 for Case 2 and in Tables 4.6.1 and 4.6.2 for MWI and ICI, respectively.

Table 4.6.1 Vertical discretization uncertainty for MWI channels estimated extending results for ATMS oxygen channels. Case 1 corresponds to the number of user levels (Nu) being larger than the number of RTTOV coefficient levels (Nc), while Case 2 corresponds to the opposite. Values in black result from the ATMS channel nearest in frequency. Values in red result from the interpolation/extrapolation based on atmospheric opacity in Figure 4.6.4.

MWI Chan #	Freq	Case 1 (Nu>Nc)	Case 2 (Nu<Nc)
1v	18v	0.12	0.12
1h	18h	0.12	0.12
2v	23v	0.13	0.13
2h	23h	0.13	0.13
3v	31v	0.14	0.14
3h	31h	0.14	0.14
4v	50_3v	0.12	0.12
4h	50_3h	0.12	0.12
5v	52_8v	0.10	0.11
5h	52_8h	0.10	0.11
6v	53_24v	0.08	0.10
6h	53_24h	0.08	0.10
7v	53_75v	0.08	0.10
7h	53_75h	0.08	0.10
8v	89v	0.01	0.01
8h	89h	0.01	0.01
9v	118_32v	0.11	0.11
10v	118_21v	0.11	0.11
11v	118_14v	0.10	0.10
12v	118_12v	0.10	0.10
13v	165v	0.04	0.06
14v	183_7v	0.07	0.12
15v	183_6v	0.09	0.09
16v	183_4v	0.10	0.19
17v	183_3v	0.15	0.28
18v	183_2v	0.20	0.39

Table 4.6.2 Vertical discretization uncertainty for ICI channels estimated extending results for ATMS water vapour channels. Case 1 corresponds to the number of user levels (Nu) being larger than the number of RTTOV coefficient levels (Nc), while Case 2 corresponds to the opposite. Values in black result from the ATMS channel nearest in frequency. Values in red result from the interpolation/extrapolation based on atmospheric opacity in Figure 4.6.4. Values at channels 9v and 10v (in bold) have been extrapolated well outside the available opacity range and thus are doubtful (likely overestimated).

ICI Chan #	Freq	Case 1 (Nu>Nc)	Case 2 (Nu<Nc)
1v	183_7	0.07	0.12
2v	183_3	0.15	0.28
3v	183_2	0.20	0.39
4v	243v	0.05	0.08
4h	243h	0.05	0.08
5v	325_9	0.09	0.16
6v	325_3	0.17	0.32
7v	325_1	0.29	0.55
8v	448_7	0.52	1.01
9v	448_3	**1.30**	**2.53**
10v	448_1	**2.45**	**4.78**
11v	664v	0.40	0.78
11h	664h	0.40	0.78

4.6.2 Estimate of contribution from surface emissivity uncertainty

Surface emissivity affects the outgoing radiation from the Earth surface and thus modulates the background radiation traveling through the atmosphere and reaching the space-borne radiometers. Although this applies in general at all frequencies, some channels may be unaffected by the surface because of the corresponding high opacity of the atmosphere. To our knowledge, the contribution of surface emissivity uncertainty to the uncertainty of brightness temperature simulations has not been quantified before. Quantification of the uncertainty affecting sea surface emissivity modeling is available at some channels and in certain conditions, while the uncertainty propagation to simulations is currently lacking. Therefore, a dedicated analysis has been performed within the VICIRS study. Surface emissivity models, such as TELSEM2 (Wang et al., 2017) and SURFEM (Kilic et al., 2023), are distributed with the current version of RTTOV and are thus used to estimate surface emissivity in this study. While SURFEM provides parameterized sea surface emissivity, TELSEM2 provides parameterized surface emissivity for land, snow and sea-ice. Albeit with some limitations (e.g., no frequency dependence of sea-ice emissivity above 183 GHz due to the lack of available information), these models are suitable for MWI and ICI simulations. In terms of accuracy, TELSEM2 emissivity up to 325 GHz has been validated against airborne observations from the International Submillimeter Airborne Radiometer (ISMAR) and the Microwave Airborne Radiometer Scanning System (MARSS), reporting consistent estimates in spatially homogeneous regions, especially at 89 and 157 GHz (Wang et al., 2017). Wang et al. (2017) shows histograms of retrieved minus TELSEM2 emissivity differences at 89, 118, 157, 183, 243, and 325 GHz channels, reporting biases and standard deviation of the order of 0.01 and 0.04, respectively. Thus, a conservative value of 0.05 has been assumed for land surface emissivity. The uncertainty of land surface emissivity has been mapped on simulated BT in the six typical climatology conditions introduced above to quantify the BT uncertainty due to the surface emissivity. Results are reported in Figure 4.6.5 for MWI and ICI, showing that the surface emissivity uncertainty suggested by Wang et al. (2017) lead to large BT uncertainty, especially at lower frequency and most transparent channels. The same analysis was performed for sea surface emissivity, considering an uncertainty value of 0.018 derived from the data reported by Kilic et al. (2023), resulting in Figure 4.6.6 for MWI and ICI.

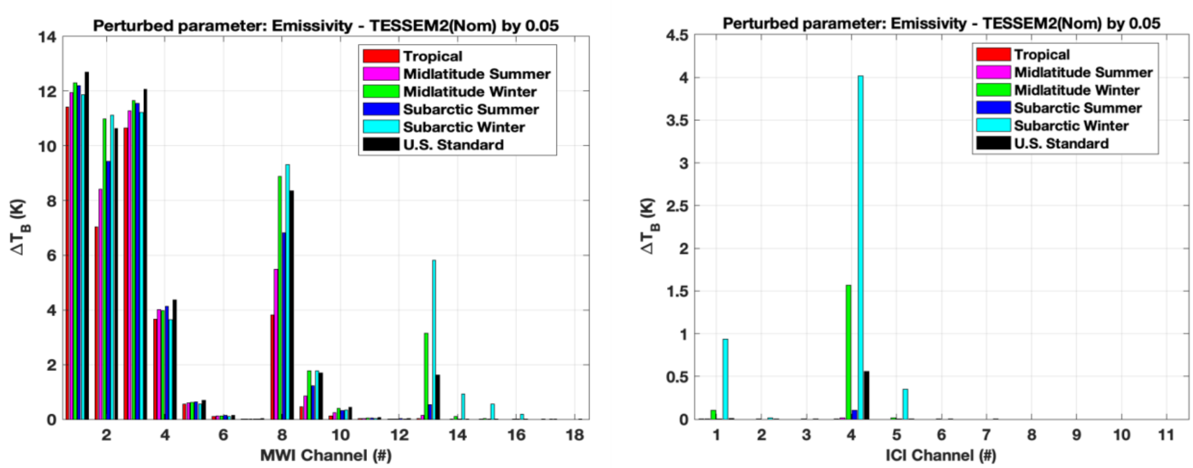


Figure 4.6.5 Uncertainty of simulated BT for MWI (left) and ICI (right) channels due to uncertainties in land surface emissivity estimated from Wang et al. (2017). Down-looking view from top-of-the-atmosphere with 53° incident angle. Color bars indicate six typical climatology conditions (tropical, midlatitude summer, midlatitude winter, sub-arctic summer, sub-arctic winter, U.S. standard).

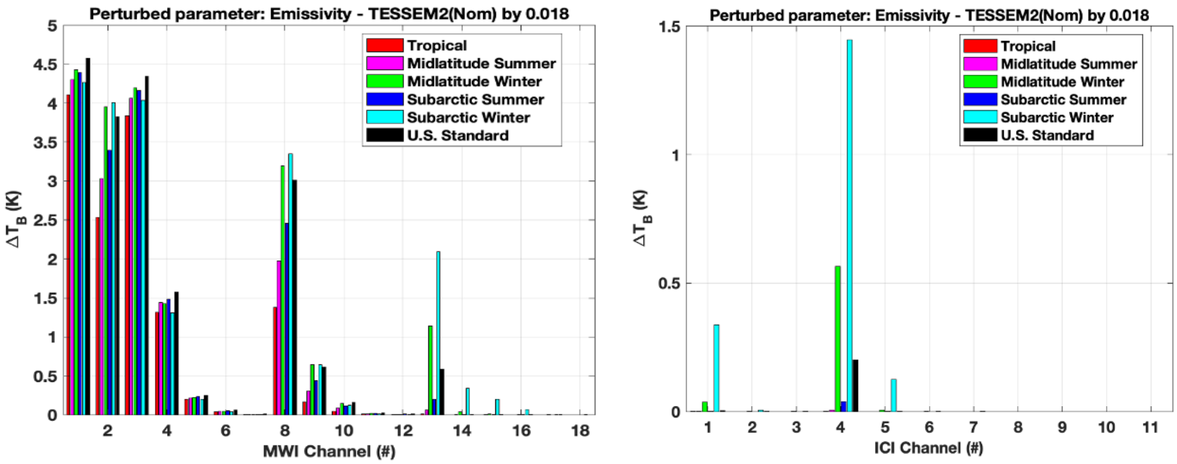


Figure 4.6.6 As Figure 4.6.2 but for uncertainties in sea surface emissivity estimated from Kilic et al. (2023).

Note that the impact of surface emissivity does depend on atmospheric conditions affecting atmospheric opacity. The most evident case is at 243 GHz (ICI channel 4), for which the contribution of surface emissivity uncertainty is negligible (<0.1 K) in warm and humid conditions (tropical and midlat summer) but becomes substantial (0.5-1.5 K) in cold and dry conditions (e.g., midlatitude and subarctic winters). This suggests a parametric approach to derive the most appropriate values, e.g. depending upon latitude, month, column-integrated water vapour (IWV), and/or surface temperature. If conditions depart substantially from the six typical climatology introduced above, the effect of atmospheric opacity may be accounted for by multiplying the given uncertainty by a “surface efficiency” (0-1), e.g., a factor proportional to the normalized land surface contribution (LSC). LSC is a measure of surface contribution to the radiance received by each channel, calculated as the difference between the simulated BT when the surface emissivity varies from 0 to 1 (Moradi et al., 2013). Putting in relation the LSC with TPW, Moradi et al. (2013b) suggest that channels 183.31±1, ±3, and ±7 GHz are not affected by surface emissivity if TPW is larger than 5, 10, and 30 kg/m^2, respectively. This method applies as is to MWI and ICI channels around 183 GHz and could be extended to higher frequency channels (e.g., 243 and 325 GHz).

4.6.3 Estimate of contribution from geolocation uncertainty

The uncertainty in geolocating the instantaneous field-of-view also contributes to the uncertainty of the observed BT. This contribution was evaluated by assuming an average geolocation uncertainty of 6 km, based on the analysis of Papa et al. (2021) who report average geolocation error between 5.4 and 6.2 km estimated from boundaries of ice shelves, mountainous lakes, and sea bays. The corresponding uncertainty was evaluated for each MWI and ICI channel using simulated test data (provided by EUMETSAT), as the BT variability (standard deviation) over a 3-by-5 box (3 along-track, 5 across-track). Since along-track and cross-track distance are ~9 and ~2 km, such a box corresponds to an area of (9+9)*(4+4)=144 km2, larger than the circle corresponding to 6 km geolocation uncertainty (~113 km2). The BT uncertainty was evaluated for each MWI and ICI channel as the average of the standard deviation over 3-by-5 boxes extracted from one entire orbit of test data (from 2007/09/12 08:43:21 to 2007/09/12 10:22:24, processed on 2022/06/13 10:44:00), as pictured in Figure 4.6.7.

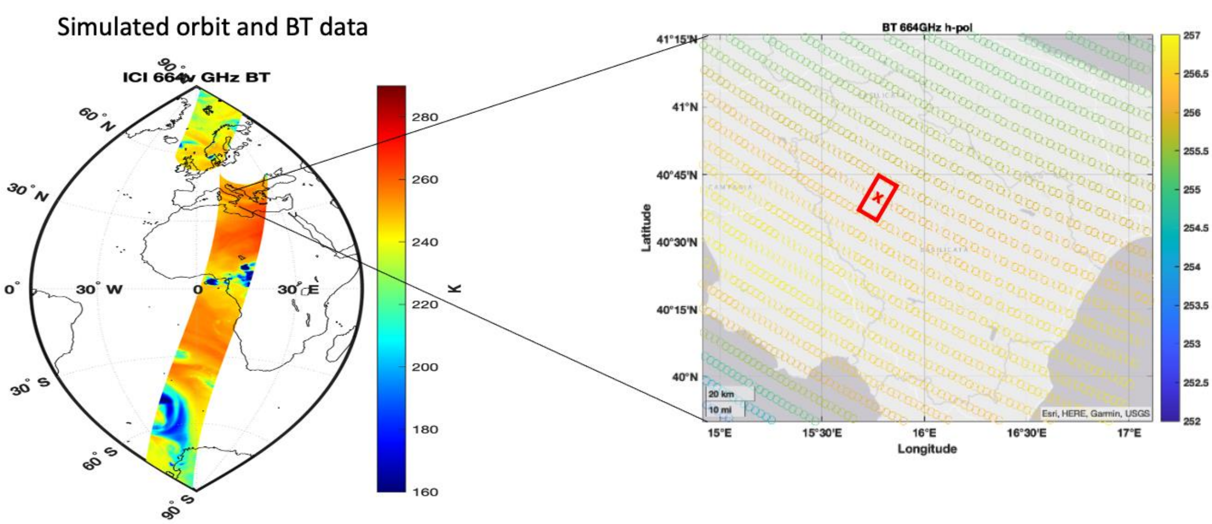


Figure 4.6.7 One full orbit of simulated ICI test data at 664 GHz (from 2007/09/12 08:43:21 to 2007/09/12 10:22:24) and zoom on one 3-by-5 box used to estimate the contribution of geolocation uncertainty.

4.6.4 Knowledge gap analysis

The review of uncertainty sources indicated knowledge gaps in their evaluation. Some of these sources have been evaluated within the VICIRS study as outlined above. As stated by Calbet et al. (2017), it is unlikely that all the sources can be characterized fully, which somehow prevents a full metrological closure. For example, to our knowledge the correlation of radiosonde uncertainty between levels has not been studied yet, and just that makes an enormous difference in the estimation of overall uncertainty. Nevertheless, the knowledge gap analysis performed within VICIRS aimed to advance the awareness and characterization of the contributing uncertainties, identifying the dominant contributions, to verify the consistency of independent measurements, making conclusions consistent to the extent possible. The dominant uncertainty contribution likely comes from spatial/temporal colocation, as already suggested by Buehler et al. (2004), although they probably underestimate the contribution of other uncertainties being evaluated here. In the following, paths for future analysis are suggested to refine the uncertainty characterization of the vicarious calibration procedure.

One uncertainty source that has not being characterized comes from the **spatial and temporal representativeness** of radiosonde data with respect to the spatial and temporal colocation criteria adopted for the match-ups, i.e. how representative are the radiosonde profiles of the atmospheric spatial and temporal variability within the selected target area and time window. The radiosonde temporal representativeness could be characterized by analyzing the typical temporal variability within different time windows mapped into clear-sky simulated BT, i.e. computing simulated clear-sky BT from a dataset of realistic atmospheric profiles (e.g., ERA5 reanalysis) for the same site but within time windows of different amplitudes (e.g., +/-1h to +/-3h), and then evaluating how the difference changes with temporal distance. This could also be linked to the meteorological conditions by using proxies, such as total column water vapor, instability indices, convective available potential energy, or wind speed/direction. The radiosonde spatial representativeness could be evaluated following the analysis outlined in Calbet et al. (2018, 2022), who derived the spatial structure functions (closely related to auto-correlation) of atmospheric water vapor and temperature from sequential radiosonde launches. Considering that water vapor and temperature are the main drivers of MWI/ICI simulated observations, their structure functions could be used as a proxy to compute the BT auto-covariance function. Other approaches may be considered as well, e.g., exploiting (*i*) available NWP data to evaluate the variability of BT for the set of NWP profiles falling within the considered target area (*ii*) available airborne observations to compute the BT auto-correlation function for each channel.

As anticipated, the **vertical correlation of radiosonde uncertainty** between levels has not been studied yet, although it is recognized to have a large impact on estimated BT uncertainty from radiosondes (Calbet et al. 2017). To this aim, a working group has been established within GRUAN to cover the needed expertise and deliver a first estimate of vertical correlation at the GRUAN ICM-16 (fall 2025).

Another unaccounted uncertainty contribution is the contamination of **undetected clouds** within the field of view, i.e. relative thin clouds with small water amounts that are not detected and screened out by the applied cloud tests. These clouds cause a residual signal that is not modeled by clear-sky RT calculations. Brogniez et al. (2016) evaluated the uncertainty related to undetected clouds for 183 GHz channels by comparing observations detected as clear-sky with simulations from ECMWF profiles in either clear-sky or all-sky computations, showing that the all-sky calculations lead to smaller biases in the lower peaking channels (e.g. by 0.4 K in the 183±7 GHz channel). A similar approach may be extended to other MWI and ICI channels.

Table 4.6.3 List of identified uncertainty sources with relevant references and status of understanding.

UNCERTAINTY SOURCE	STATUS OF UNDERSTANDING	REFERENCES
Uncertainty propagation	Theory is well understood.	JCGM, 2008; 2012 Immler et al., 2010 Carminati et al. 2019
Radiometric sensors	Quantification is available by instrument characterization and calibration.	AD-4 AD-5
Radiosonde sensors	Quantification is available at point measurement level. Lack of information on vertical correlation (estimate expected by 2025 by a dedicated GRUAN working group).	Madonna et al., 2022 Von Rhoden et al., 2022 Dirksen et al., 2020 Carminati et al., 2019 Dirksen et al., 2014
Geolocation	Quantification is available in terms of geolocation uncertainty. Mapped into BT using MWI and ICI test data (one orbit) within the VICIRS study.	Papa et al., 2021
Colocation (temporal, spatial)	Recommendations are available to reduce the uncertainty and provide a rough estimate. A dedicated study is recommended.	Calbet et al., 2022 Calbet et al., 2018 Bobryshev et al., 2018 Brogniez et al., 2016 Verhoelst et al., 2015 Ignaccolo et al., 2015 Fassò et al., 2014 Moradi et al., 2013 Calbet et al., 2011 Buehler et al., 2004
Discretization	Recommendations are available to reduce the uncertainty and provide a rough estimate. A study focusing on ATMS has been extended to MWI and ICI channels within the VICIRS study, with limitations for high-opacity channels.	Buehler et al., 2004 Hocking, 2014
Absorption model	Quantification is available for 16-700 GHz range and for MWI, ICI, MWS, and ATMS channels.	Gallucci et al., 2024
Surface emissivity	Quantification is available in terms of surface emissivity uncertainty at some channels and in certain conditions. Suggested surface emissivity uncertainty has been propagated to BT within the VICIRS study. A dedicated study to link expected uncertainty to atmospheric conditions is suggested.	Wang et al., 2017 Pringet et al., 2017
Cloud screening	Quantification is available only for a few channels (e.g., 183±7 GHz). A dedicated study to extend results to other MWI and ICI channels is suggested.	Brogniez et al., 2016 Moradi et al., 2013a John et al., 2012 Buehler et al., 2004

Table 4.6.3 presents a summary of the identified uncertainty sources with the status of understanding at the end of the VICIRS study, relevant references, and suggestions for further studies.

4.7 Multi-source correlative methodology for RS, NWP and SAT

The Multi-source Correlative Methodology (MCM), also known as Triple Colocation (TC) allows for the characterization of the error structure of three collocated (in space and time) measuring systems. In the following sections the mathematical formalism underpinning MCM is described with emphasis to the assumptions required to properly implement MCM.

4.7.1 MCM assumptions

We assume to have three measuring systems (x_i) with the index $i = 1, 2, 3$ (Stoffelen et al., 1998). Note here that the term "measuring systems" is not restricted to actual measured data only, but it can be extended to a numerical system that simulates a measured quantity, as well. Each of the terms x_i can be thought as the result of a measuring process that introduces some amplification (a_i), biases (b_i) and noise (ε_i), to a true, but unknown, geophysical quantity (t):

$$x_i = b_i + a_i\, t + \varepsilon_i \tag{4.7.1}$$

where:

a_i: calibration scaling of measuring system $i - th$

b_i: calibration bias of measuring system $i - th$

t: unobserved truth which is common to all the measuring systems

ε_i: measurement random error of system $i - th$

Some assumptions on the different terms of (4.7.1) are required to greatly simplify the mathematics and quantify the error variance ($\sigma^2_{\varepsilon_i}$) which is the ultimate goal of MCM. Some other additional assumptions are required to estimate the calibration parameters of two measuring systems out of three. The main MCM assumptions are:

Assumption1 (A1): eq. (4.7.1) holds, that implies linearity between t and x_i holds as well.

Assumption2 (A2): the error ε_i is zero average random error that means $\langle \varepsilon_i \rangle = 0$ where $\langle \cdot \rangle$ is the average operator.

Assumption3 (A3): the error ε_i is independent by the truth, t, which means $\langle \varepsilon_i\, t \rangle = 0$.

Assumption4 (A4): the errors of the various measuring systems are independent by each other, i.e.: $e_{ij} = \langle \varepsilon_i\, \varepsilon_j \rangle = 0$ with $i \neq j$ and (i, j) describing all the combinations in the interval from 1 to 3.

Assumption5 (A5): both x_i and ε_i are stationary processes, i.e. they should have constant mean and standard deviation in the analyzed domain.
Assumption6 (A6): the three measuring systems, x_i, $i = 1, 2, 3$ must observe the same quantity t.
The additional assumptions required to find a_i and b_i will be discussed in the next section.

4.7.2 MCM random error estimation

Under the assumptions A1-A6, and assuming that the error variances are referred to the observation scale of the third measuring systems, it can be demonstrated (see Appendix B) that the error variance of the three measuring systems in (4.7.1) can be estimated ($\hat{\cdot}$) as follows:

$$\hat{\sigma}^2_{\varepsilon_1} = \sigma^2_{x_1} - \frac{C_{13}}{C_{23}}(C_{12} - e_{12}) \tag{4.7.2a}$$

$$\hat{\sigma}^2_{\varepsilon_2} = \sigma^2_{x_2} - \frac{C_{23}}{C_{13}}(C_{12} - e_{12}) \tag{4.7.2b}$$

$$\hat{\sigma}^2_{\varepsilon_3} = \sigma^2_{x_3} - C_{13}\, C_{23} \left(\frac{1}{C_{12} - e_{12}}\right) \tag{4.7.2c}$$

In eq.s (4.7.2), $\hat{\sigma}^2_{x_i}$ is the estimated variances of the measured quantities from the $i-th$ measuring system whereas C_{ij} are the covariances terms defined as usual:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{4.7.3}$$

More elaborated arguments need to be spent for the term e_{12}. The latter is defined as the error covariance of errors of systems 1 and 2:

$$e_{12} = \langle \varepsilon_1\, \varepsilon_2 \rangle = \rho_{12}\, \sigma_{\varepsilon_1}\, \sigma_{\varepsilon_2}. \tag{4.7.4}$$

Where ρ_{12} is the correlation coefficient between ε_1 and ε_2. Then $e_{12} \neq 0$ seems to violate the MCM A4. However, as will be discussed in a later section, the MCM theory can accept some measurement errors to be correlated to describe some representativeness errors in the measurements. In eq.s (4.7.2), it is implicitly assumed that the system number 3 has the lowest spatial resolution compared to the other two systems, and it is taken as reference for the scale of analysis (i.e. the error variances, $\hat{\sigma}^2_{\varepsilon_i}$, will be referred to the poorer spatial scale of the system 3). Consequently, measurements from systems 1 and 2, thanks to their higher variability that is caused by their higher resolution than system 3, will pay for an additional error. Such an extra error depends on the way the systems 1 and 2 observe (i.e. represent) the scene and for this reason it is referred to as the representativeness error in the MCM framework. The representativeness error is then included in the term e_{12} eq. (4.7.2) since it contributes to an increase of the error variance $\hat{\sigma}^2_{\varepsilon_1}$ and $\hat{\sigma}^2_{\varepsilon_2}$. In other words, the term $(C_{12} - e_{12})$ is lowering as e_{12} is increasing, thus making $\hat{\sigma}_{\varepsilon_1}^2$ closer to $\hat{\sigma}_{x_1}^2$ and lowering $\sigma_{\varepsilon_3}^2$. A similar reasoning applies to the system 2). Estimation of e_{12} can be critical and requires spatial spectral analysis of the data for systems 1 and 2 (see later sections). Obviously, a different choice of the reference scale of analysis will lead to a different formulation of (4.7.2). However, in our case the third system could be associated with MWI/ICI whereas the other two could be RS and NWP.

4.7.3 MCM correlation coefficient estimation

A second output quantity provided by the MCM is the correlation coefficient ($\rho_{t,i}$) between the actual unobserved value t and each input time series x_i (McColl et al., in 2014):

$$\hat{\rho}_{t,1} = \frac{1}{\hat{\sigma}_{x1}}\sqrt{\frac{(C_{12}-e_{12})\,C_{13}}{C_{23}}} \tag{4.7.5a}$$

$$\hat{\rho}_{t,2} = \frac{1}{\hat{\sigma}_{x2}}\sqrt{\frac{(C_{12}-e_{12})\,C_{23}}{C_{13}}} \tag{4.7.5b}$$

$$\hat{\rho}_{t,3} = \frac{1}{\hat{\sigma}_{x3}}\sqrt{\frac{C_{23}C_{13}}{(C_{12}-e_{12})}} \tag{4.7.5c}$$

These quantities are additional metrics to characterize the errors in the three measuring systems, providing important new information about their performance. Another quantity that can be derived by the correlation coefficients is the signal to noise ratios of the three systems defined as the ratio between the signal and the error averaged squared value:

$$SNR_i = \frac{\langle (x_i')^2 \rangle}{\langle (\varepsilon_i)^2 \rangle} = \frac{\rho_{t,i}^2}{1+\rho_{t,i}^2} \tag{4.7.6}$$

Where x_i' is the signal part of x_i (i.e. $x_i' = b_i + a_i\, t$).

4.7.4 MCM calibration parameter estimation

Another important achievement of MCM is the possibility to calculate the calibration parameters a_i and b_i. This is done by assuming one of the three systems as a reference. This implies that for the reference system we assume to know its calibration parameters. Without loss of generality, hereafter, we consider system 1 to act as the reference systems so that, for system 1 the calibration parameters a_1 and b_1 are known inputs. Under such assumption the calibration parameters of the other two systems can be estimated as follows:

$$\hat{a}_2 = \frac{C_{23}}{C_{13}} a_1 \tag{4.7.7a}$$

$$\hat{a}_3 = \frac{C_{23}}{(C_{12}-e_{12})} a_1 \tag{4.7.7b}$$

and

$$\hat{b}_2 = \langle x_2 \rangle - \frac{C_{23}}{C_{13}} \langle x_1 \rangle + \frac{C_{23}}{C_{13}} b_1 \tag{4.7.8a}$$

$$\hat{b}_3 = \langle x_3 \rangle - \frac{C_{23}}{(C_{12}-e_{12})} \langle x_1 \rangle + \frac{C_{23}}{(C_{12}-e_{12})} b_1 \tag{4.7.8b}$$

It is worth noting that assuming the knowledge of calibration parameters $a_1$and b_1 is equivalent to assume that the measuring system 1 is perfectly calibrated, because it is always possible to calibrate system 1 from the knowledge of $a_1$and b_1 (i.e. produce a calibrated system $\underline{x}_1 = \frac{x_1 - b_1}{a_1}$). If $\underline{x}_1$ is considered instead of x_1 then, in eq.s (4.7.7)-(4.7.8) $a_1 = 1$ and $b_1 = 0$ should be considered.
However, leaving explicit the parameters $a_1$and b_1 in (4.7.7)-(4.7.8) it easily gives the opportunity to quantify the error standard deviation ($\sigma_{\hat{a}_2}$), ($\sigma_{\hat{a}_3}$), ($\sigma_{\hat{b}_2}$) and ($\sigma_{\hat{b}_3}$) in $\hat{a}_2$, $\hat{a}_3$, $\hat{b}_2$ and $\hat{b}_3$ estimates, respectively, as a function of the error standard deviations (σ_{a_1}) and (σ_{b_1}) with which reference calibration parameters a_1 and b_1 are known. In this respect, error propagation theory finds an easy application in our case, yielding the following:

$$\sigma_{\hat{a}_2} = \frac{C_{23}}{C_{13}} \sigma_{a_1} \tag{4.7.9}$$

$$\sigma_{\hat{a}_3} = \frac{C_{23}}{(C_{12}-e_{12})} \sigma_{a_1} \tag{4.7.10}$$

$$\sigma_{\hat{b}_2} = \frac{C_{23}}{C_{13}} \sigma_{b_1} \tag{4.7.11}$$

$$\sigma_{\hat{b}_3} = \frac{C_{23}}{(C_{12}-e_{12})} \sigma_{b_1} \tag{4.7.12}$$

4.7.5 Synthetic experiment setup

Simulations of BT (K) are used to synthetically reproduce the three measuring systems x_1, x_2, x_3 and implement the MCM equations. **Figure 4.7.1** shows the flow chart followed to generate and evaluate the outcomes of the synthetic experiment performed. BT considered are those from MWI simulations at 183GHz $\pm$2GHz V, as provided by EUMETSAT ([AD-4], [AD-5]). They include four reference Metop-A orbits:
● orbits 4655 and 4656: from 08:00 to 13:00 UTC from ERA5 forecast@2007-09-12T06:00:00
● orbit 6985: from 08:00 to 11:00 UTC from ERA5 forecast @2008-02-23T06:00:00
● orbit 9744: from 13:00 to 16:00 UTC from ERA5 forecast @2008-09-04T06:00:00
with a horizontal resolution (0.25°, i.e. about 30 km), vertical resolution variable up to an altitude of 80 km, and time sampling of 1 hour. Values of BT corresponding to cloud free

conditions and for the orbits 4655 and 4656, are selected. However, the specific channel selected does not have a significant impact on the final result because, in our experiment, the error structure assigned to each measuring system is a user choice and then the MCM outcomes are not affected by the specific choice of the *BT* values. **Figure 4.7.2** (left) shows the time series of the input *BT* considered. However, as explained later, to investigate the sensitivity of MCM outcomes to the number of measurements available, a random selection of *BT* values (reshuffling) from those available from simulations, is performed (**Figure 4.7.2**, right). The setup of the experiment foresees the simulations of the three measuring systems x_1, x_2, x_3 as follows:

$$x_1 = b_1 + a_1\, BT + \varepsilon_1 \qquad (4.7.13a)$$
$$x_2 = b_2 + a_2\, BT + \varepsilon_2 \qquad (4.7.13b)$$
$$x_3 = b_3 + a_3\, BT + \varepsilon_3 \qquad (4.7.13c)$$

where the calibration coefficients a_1, a_2, a_3, b_1, b_2, b_3 are fixed through the various experiments presented later, whereas the zero-mean Gaussian error terms, $\varepsilon_1, \varepsilon_2$, and ε_3 are generated assigning the error variances $\sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}$ which are varied among fixed values: [0.1^2, 0.5^2, 1.0^2, 1.5^2, 2^2] K (Table 4.7.1).

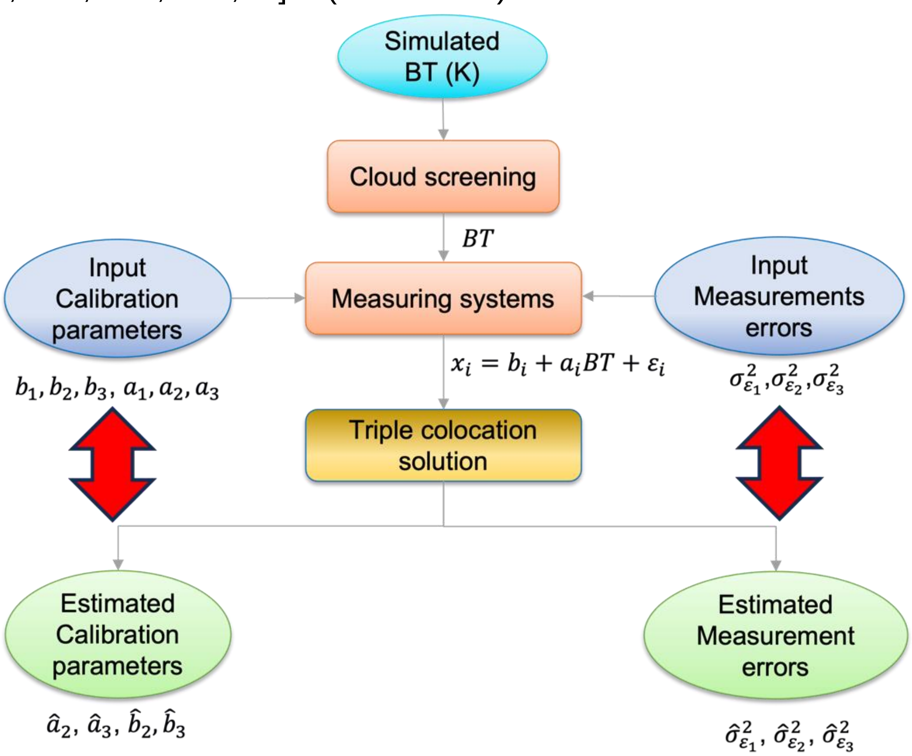


Figure 4.7.1: Flowchart followed for the synthetic experiment to test the triple colocation.

With this choice N_t=5 x 5 x 5=125 combinations of the error triplets are generated. For each of the 125 error triplets, N_s randomly selected samples of *BT* are generated (see **Figure 4.7.2**, right, with N_s in [10^2, 10^6]). Then, $\hat{\sigma}_{\varepsilon_i}$ is estimated using eq. (4.7.2) and the error difference, $\hat{\sigma}_{\varepsilon_i,} - \sigma_{\varepsilon_i}$, is calculated. Consequently, the root mean square error ($RMSE_i$) is defined as:

$$RMSE_i = \sqrt{\langle\left(\hat{\sigma}_{\varepsilon_i,} - \sigma_{\varepsilon_i}\right)^2\rangle_{N_t}} \quad \text{with } i = 1, 2, 3. \tag{4.7.14}$$

where the average operator "$\langle\cdot\rangle_{N_s}$" extends across the N_t=125 error cases tested.

Note that the reshuffling operation (**Figure 4.7.2**, right) tends to level any trend (i.e. non-stationarity) which may be present in the initial dataset (**Figure 4.7.2**, left) and this makes the input data more in agreement with the stationarity assumption required by MCM.

Some test cases, obtained varying the MCM parameters, are applied to illustrate the expected performance of MCM. Their set up is listed in **Table 4.7.2** in which the error covariance term, e_{12}, is factorized into $e_{12} = \rho_{12}\, \sigma_{\varepsilon_1}\, \sigma_{\varepsilon_2}$ with the term ρ_{12} representing the correlation coefficients between the error terms ε_1 and ε_2.

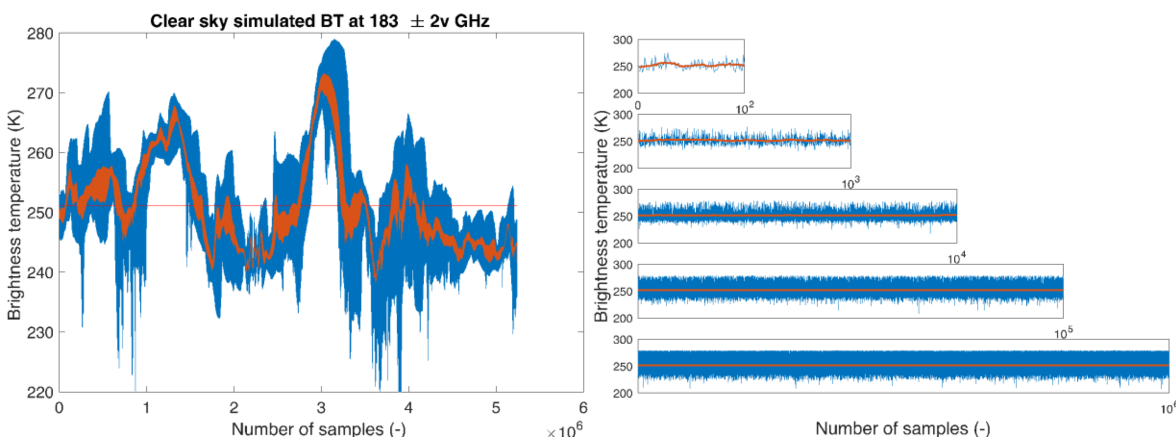


Figure 4.7.2: simulated time series BT (K) with cloud filtered for the orbits 4655 and 4656 on 2007-09-12 for the MWI channel at 183± 2 v GHz (left). Randomly selected samples from the time series on the left for (from top to down) increasing number of selected samples, N_s. Red curves are the running average for each time series.

Table 4.7.1. Set of error variances used to generate the error terms $\varepsilon_1, \varepsilon_2, \varepsilon_3$.

Error variances $\sigma^2_{\varepsilon_i}$ (K^2)	0.1^2	0.5^2	1.0^2	1.5^2	2^2

Table 4.7.2. Setup of the test experiments implemented.

		Parameters used to generate input triples (x_1, x_2, x_3)						
Cases		a_1	a_2	a_3	b_1	b_2	b_3	ρ_{12}
1) Ideal case	(ID)	1	1	1	0	0	0	0
2) Intermediate case	(IN)	1	1.25	0.75	0	1	1	0
3) Worst case 1	(W1)	0.75 – 1.25	1.25	0.75	0	1	1	0
4) Worst case 2	(W2)	0.75 – 1.25	1.25	0.75	0	1	1	0.1
5) Worst case 3	(W3)	0.75 – 1.25	1.25	0.75	0	1	1	0.3

4.7.6 MCM algorithm implementation

The MCM algorithm implemented descends directly from eq.s (4.7.2), (4.7.7), (4.7.8). We verified that $\hat{\sigma}_{\varepsilon_i,}$ may strongly depend by the scaling parameters a_2 and a_3. This is shown in **Figure 4.7.3** where two identical triplets are considered but in one case (left) we implemented the ordinary MCM in eq. (4.7.2) whereas in a second case (right) we firstly estimated the calibration parameters $\hat{a}_i$ $\hat{b}_i$ (*i*=2, 3) using eq. (4.7.7) and (4.7.8), and then we implemented eq. (4.7.2) with calibrated triplets ($\underline{x}_i$) instead of x_i:

$$\underline{x}_i = \frac{x_i - \hat{b}_i}{\hat{a}_i} = BT + \frac{\varepsilon_i}{\hat{a}_i} \text{ with } i = 2, 3 \qquad (4.7.15)$$

Such a two-step approach with calibrated triplets is used in the presentation of the experiment results in the next section and it is the approach suggested for the final implementation of the MCM method. The effect of the calibration is clearly visible in Figure 4.7.3 (right) as a spread reduction in the estimated $\hat{\sigma}_{\varepsilon_i,}$ compared to the uncalibrated case (left). It is worth underlying that the biases $\hat{b}_i$ do not play a role in the accuracy of $\hat{\sigma}_{\varepsilon_i,}$ since the covariances from which they are derived are not affected by b_i terms. Note that in the case of calibrated systems the error component, ε_i, is scaled by $\hat{a}_i$ (see the third term of eq. (4.7.15)) and consequently the estimated error STD must be multiplied by $\hat{a}_i$ too (i.e. $\hat{\sigma}_{\varepsilon_i,}$ in the eq. (4.7.14) must be replaced by $\hat{a}_i$ $\hat{\sigma}_{\varepsilon_i,}$). A detailed and self-explicative flow chart of the MCM implementation is shown in figure 4.7.4. A last important consideration concerns the decreasing trend of retrieval errors in figure (4.7.3) as a function of the number of triplets considered. To explain this behavior, we set up a synthetic experiment generating two zero-mean, unitary-variance, 0.5-correlated, random noises with progressive increment of the number of samples in each time series. This allows us to check how fast the covariance and the average of the generated time series converge to the expected values, i.e., in our case, 0.5 and 0, respectively. The experiment result is shown in Figure (4.7.5), which clearly shows that a number of samples of the order of 10^5 or more is needed to stabilize the statistical moments. When applying MCM this aspect can be particularly important because averages and covariance terms combine together as in eq.s (4.7.2), (4.7.7), (4.7.8), and small inaccuracies in these terms can produce large uncertainties in the MCM output quantities.

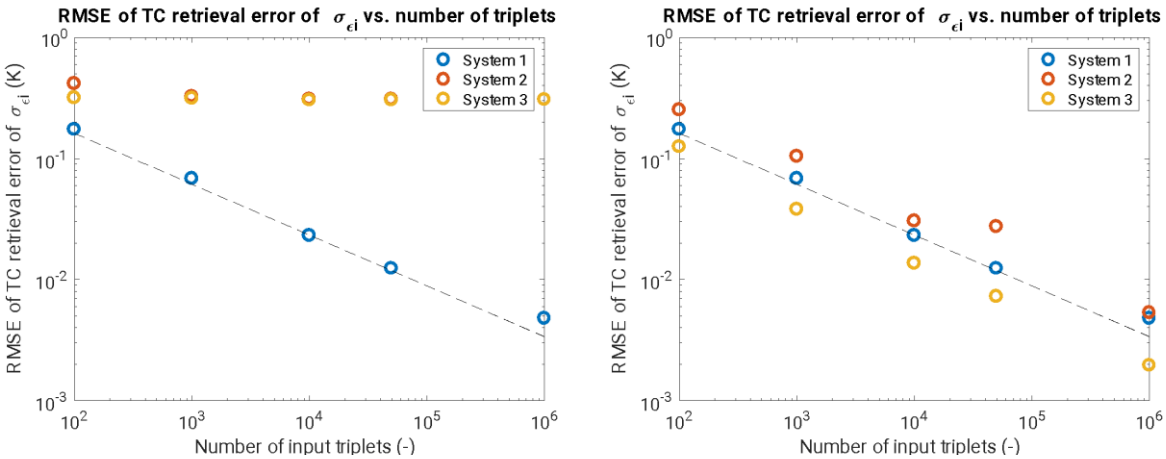


Figure 4.7.3 RMSE of $\hat{\sigma}_{\varepsilon_i}$ in eq. (4.7.14) for the intermediate case listed in table 4.7.2 without (left panel) and with (right panel) the calibration procedure in eq. (5.3).

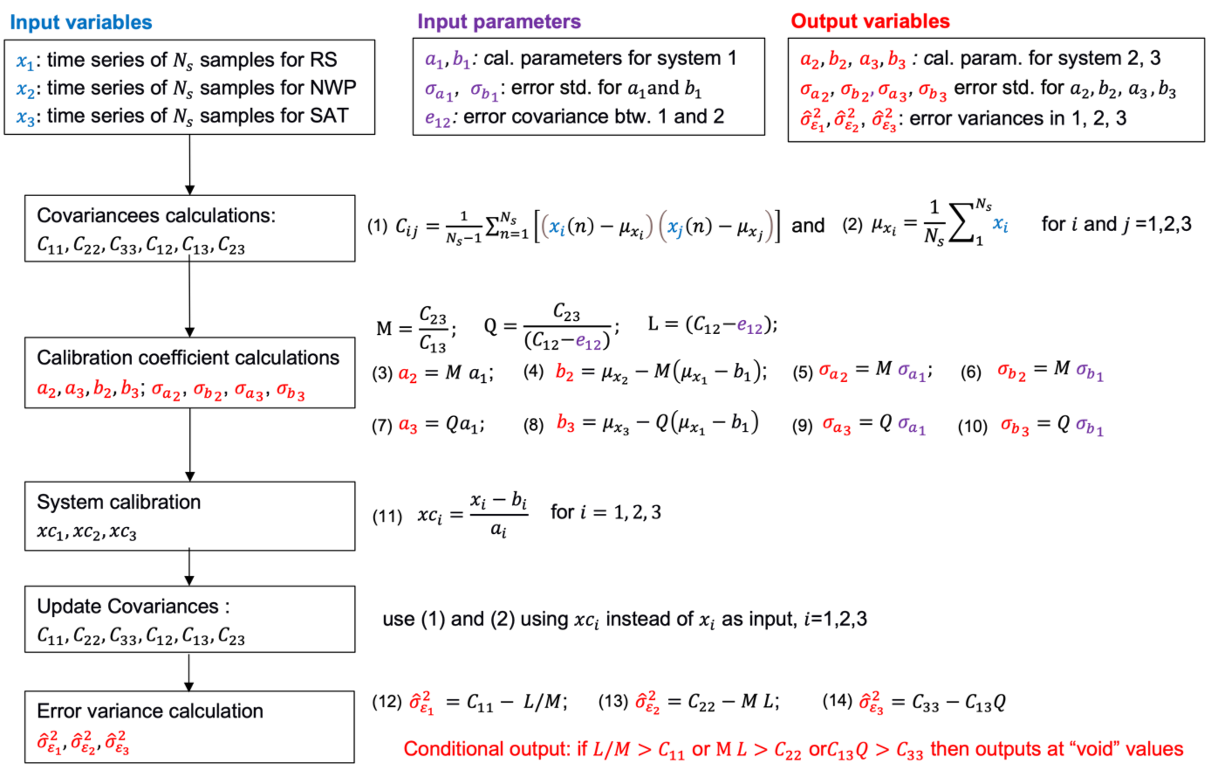


Figure 4.7.4: Flow chart of the MCM implementation. Input/output variables as well as input parameters are color highlighted. Measuring system 1 is assumed to be the reference one with known calibration parameters a_1, b_1.

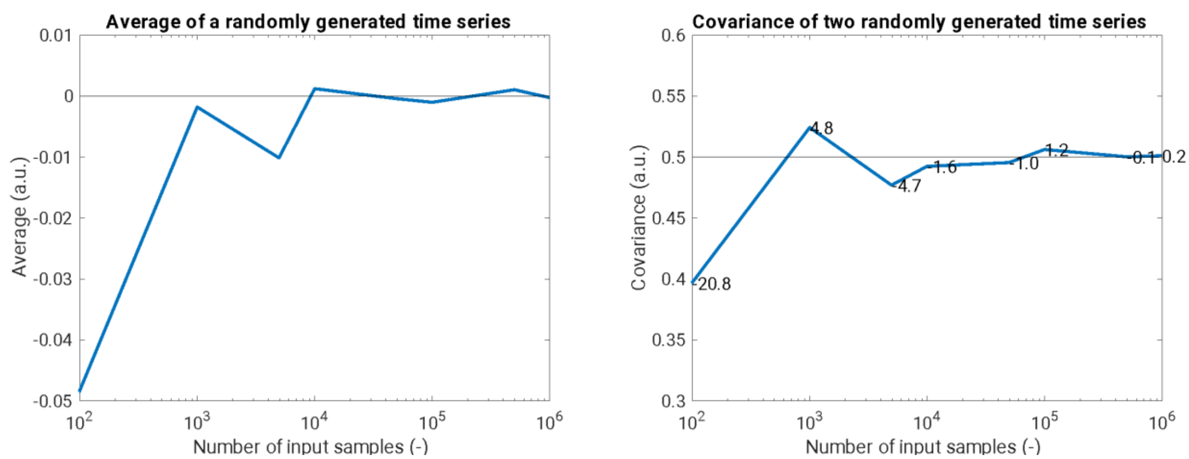


Figure 4.7.5: average of one member zero-mean, unitary-variance random noise time series (left) and covariance of two members zero-mean unitary-variance and 0.5-correlated random noise time series (right). The numbers in the plot of the right panel represent the relative errors (%).

4.7.7 Results of MCM performances

To assess the dependence of MCM performance, the MCM equations, as summarised in Figure 4.7.4, are implemented for various test cases listed in Table 4.7.2 and used to generate input triples (x_1, x_2, x_3)

Figure 4.7.6 (left) shows, for the ideal case of Table 4.7.2, the MCM error difference statistic as a function of the number of samples, *N*s, whereas the RMSE is shown in the right panel. Each error bar in the left panel for a fixed value of *N*s, includes 125x*N*s samples. From this figure it is clear as MCM yields more robust error standard deviations,, σ_{ε_i}, with an increasing *N*s. This is probably because the covariance terms, C_{ij}, that are involved in the MCM formulation, acquires statistical robustness for a large number of samples as demonstrated in Figure 4.7.6. A level of 1000 triples could seem enough to maintain the MCM RMSE error retrieval below 0.1 K (right panel), although other factors need to be considered for the evaluation of the minimum number of triplets necessary to achieve a reasonable MCM estimation error (see later). Another aspect to note is the RMSE difference among three systems that seems to vary a bit as *N*s increases. This is likely caused by the large dynamic that the imposed error variance can assume across the three systems. For a fixed *N*s and a given experiment case, the error level added to the three systems can differ considerably (see table 4.7.1). This unbalance causes some outliers in the estimation of the error variance (e.g. figure 4.7.6 left, red dots for *N*s=5000) thus making the RMSE higher (red circle in the right panel for *N*s=5000).

Similarly, to Figure **4.7.6**, **Figures 4.7.8**- **4.7.10** show the other cases of Table 4.7.2 with a progressive relaxing of some of the MCM assumptions. For example, in the intermediate case (**Figure 4.7.7**) system 2 and 3 are uncalibrated (i.e. the scaling parameter are of the order of $\pm 25\%$ with respect to ideal conditions) and the overall effect is an increment of variation of RMSE with respect to the ideal case. In particular, increasing a_2 by a factor of $+25\%$ makes the RMSE of system 2 larger than the ideal case, whereas decreasing a_2 by the same factor, makes the RMSE lower. Such error variation involves system 1 too when assuming a variation of $\pm 25\%$ for the scaling parameter a_1 system 1 (**Figure 4.7.8,** worst case 1). Eventually, when assuming a small correlation between errors of systems 1 and 2 (i.e. $\rho_{12} = 0.1$), large increments of the number of samples do not bring any benefit as in the previous test cases discussed (**Figure 4.7.9**) and a plateau is reached after N_S =1000. From this latter experiment it can be argued that, at worst, for N_S=100, $\hat{\sigma}_{\varepsilon_i}$ from MCM can be estimated with an accuracy lower than 0.3 K RMSE.

Things are getting worse and worse as ρ_{12} increases (e.g. in the worst case with $\rho_{12} = 0.3$, **Figure 4.7.10**). In this case, the error dependence with respect to N_S tends to vanish with the RMSE of $\hat{\sigma}_{\varepsilon_i}$ of the order, at worst, of 0.4 K for N_S=100. Thus, the detrimental effect of an increased correlation between measurement errors of system 1 and 2 is to increase the uncertainty in the estimates of $\hat{\sigma}_{\varepsilon_i}$. However, in our case $\rho_{12} \neq 0$ could be due to the fact that system 1 and 2, namely RS and NWP, share the same RTM to produce the output BT. Consequently, in this circumstance, we could assume ρ_{12}=1. Thus, recalling that, in general, the error covariance of measuring system 1 and 2 can be expressed as $e_{12} = \rho_{12}\, \sigma_{\varepsilon_1}\, \sigma_{\varepsilon_2}$, we have $e_{12} = \sigma_{\varepsilon_1}\, \sigma_{\varepsilon_2}$. The latter can be further simplified into $e_{12} = \sigma^2_{\varepsilon_{RTM}}$ assuming that the RTM error is the only one which equally interest system 1 and 2 (i.e. producing $\sigma_{\varepsilon_1} = \sigma_{\varepsilon_2} = \sigma_{\varepsilon_{RTM}}$).
If $\sigma_{\varepsilon_{RTM}}$ term was known, it could be ingested in the MCM procedure allowing to take into account the correlation term ρ_{12}. In this case, the results from the intermediate test case (i.e. that with $\rho_{12} = 0$), would give a reasonable error frame reference.

Figures 4.7.11-4.7.15 show the RMSE retrieval errors of $\hat{b}_i$ (left panels) and $\hat{a}_i$ (right panels) for $i = 2, 3$ through all the experiments in Table 4.7.2. For system 1 the key assumption is that we know its calibration parameters so that we can take them into a proper consideration. Later we will relax this assumption by assuming to know system 1 calibration parameters with some degree of uncertainty. Two important things are noteworthy. The first thing is that the bias has a very strong dependence from the number of triplets considered and only a very high number of them guarantees acceptable estimation accuracy (e.g. 10^4 triplets are needed to achieve an accuracy less than 0.9K for $\hat{b}_i$ in the "worst case 3" experiment). The second consideration is that RMSE of $\hat{a}_i$ are not much sensitive to the number of input triplets in the presence of miscalibrations (i.e. when worst cases are considered). In those cases, the RMSE is always lower than 0.5. Just for reference, an accuracy of 0.5 correspond to a variation of $\pm$13° with respect to the 1:1 line in the scatterplot of the measurement x_i vs. the truth t.

In the previous experiments presented, the calibration parameters of the reference measuring system number 1, that is a_1 and b_1, are assumed to be perfectly known. If we remove this assumption, eqs. (4.7.9)-(4.7.12) can be used to propagate the uncertainty of a_1 and b_1, which is σ_{a_1} and σ_{b_1}, respectively, into those of the calibration parameters of the other two measuring systems: σ_{a_2}, σ_{b_2}, σ_{a_3}, σ_{b_3}. **Figure 4.7.16** shows the result of the uncertainty propagation for the various test cases implemented in Table 4.7.2. In the figure we assumed σ_{a_1} and σ_{b_1} to be known and both varying from 0.1 to 0.4 at step of 0.1, respectively. The different test cases produce different values of the covariance terms $\frac{C_{23}}{C_{13}}$ and $\frac{C_{23}}{(C_{12}-e_{12})}$ in eqs. (4.7.9)-(4.7.12), so that σ_{a_2}, σ_{b_2} seem to be progressively amplified with respect to the reference values of σ_{a_1} or σ_{b_1} (**Figure 4.7.16, left**) whereas and opposite behavior is registered for σ_{a_3}, σ_{b_3} (right). This figure suggests that if we assume that the reference system 1 is calibrated with a good degree of uncertainty of the order of 0.1 K for both a_1 and b_1, respectively, the resulting uncertainty in the estimated calibration parameters (a_2, a_3) and (b_2, b_3) will be reasonably well below 0.2 K

IDEAL CASE: ERRORS OF $\hat{\sigma}_{\varepsilon_i}$

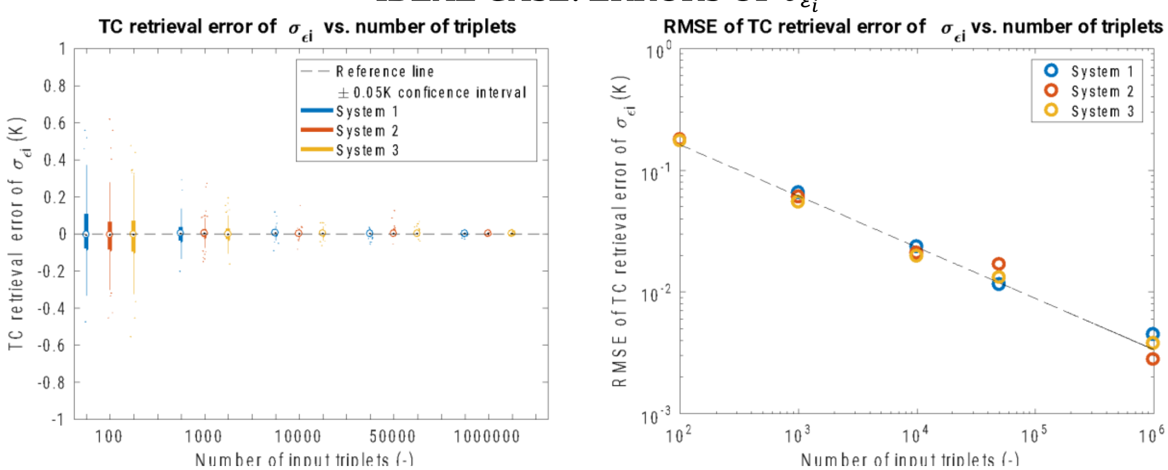


Figure 4.7.6. Box plot MCM retrieval error (left) of $\hat{\sigma}_{\varepsilon_1}$, $\hat{\sigma}_{\varepsilon_2}$ and $\hat{\sigma}_{\varepsilon_3}$ as a function of the number (N_S) of considered triplets (x_1, x_2, x_3) and corresponding RMSE (right). In the box plot, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The thin vertical lines extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '.' symbol. Selected MCM parameters are those for the ideal case in table .25. Horizontal dashed lines in the left panel are thresholds at ± 0.05 K whereas in the right panel the linear fit *y=mx+q* is shown (*m*= -0.42, *q*= 0.11).

INTERMEDIATE CASE: ERRORS OF $\hat{\sigma}_{\varepsilon_i}$

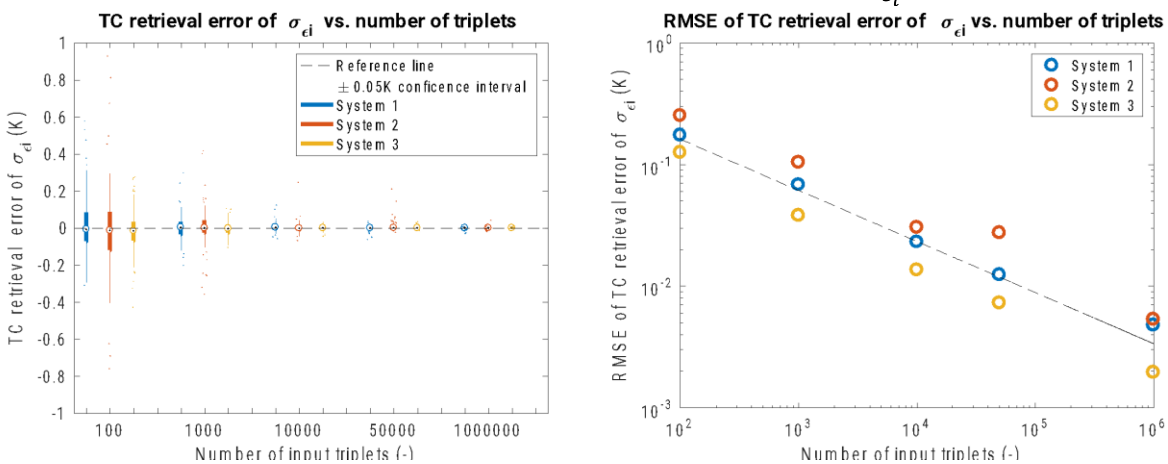


Figure 4.7.7. As in figure 4.7.6 but for the intermediate case in table 4.7.2.

WORST CASE 1: ERRORS OF $\hat{\sigma}_{\varepsilon_i}$

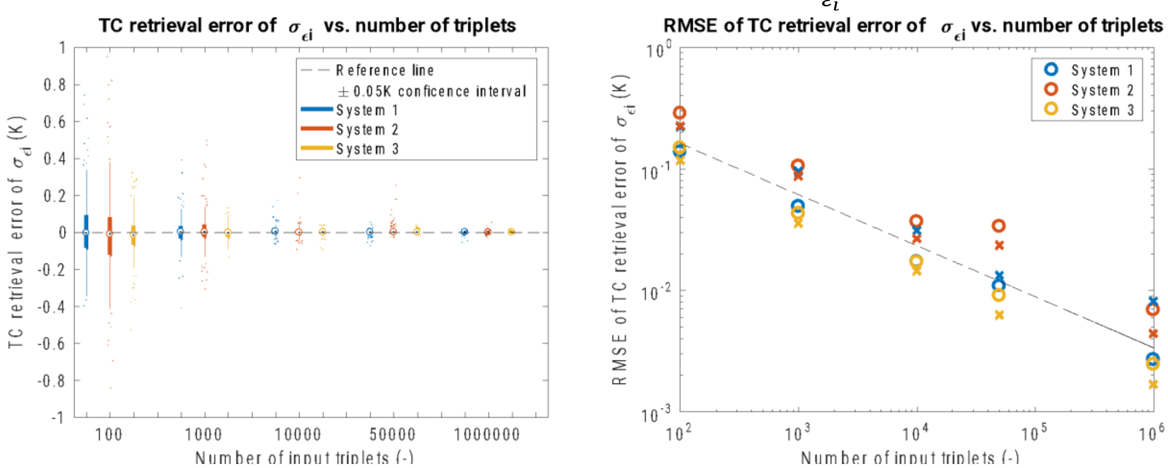


Figure 4.7.8. As in figure 4.7.6 but for the worst case 1 in table 4.7.2. In the right panel, circles refer to $a_1 = 0.75$ whereas cross markers to $a_1 = 1.25$.

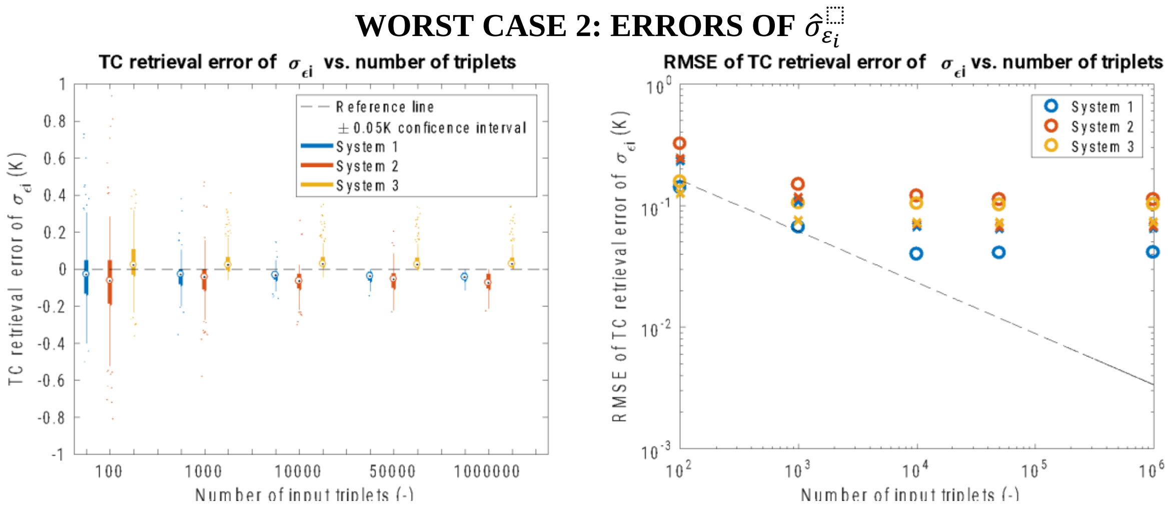


Figure 4.7.9. As in Figure 4.7.6 but for the worst case 2 in table 4.7.2.

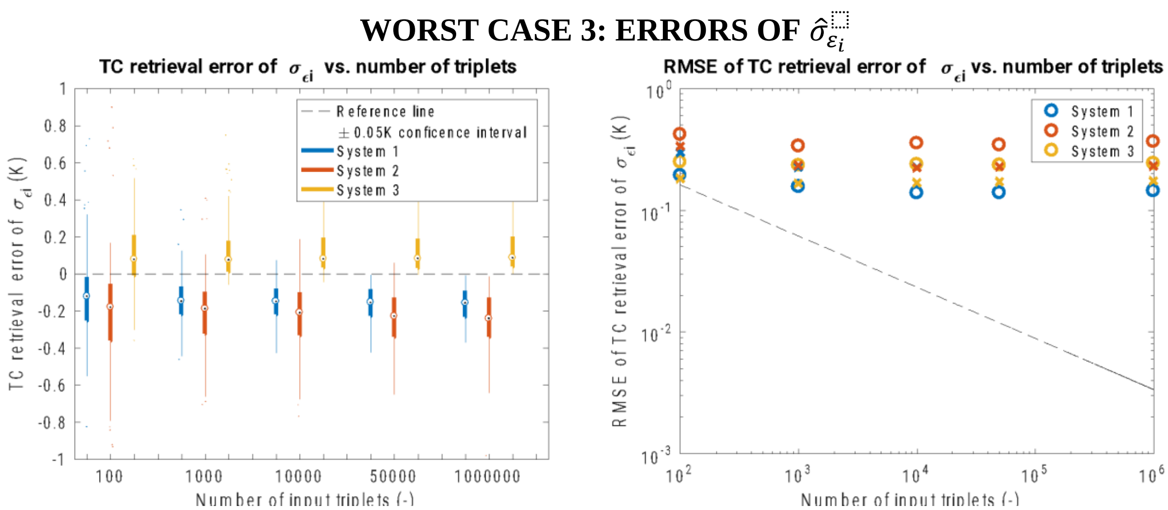


Figure 4.7.10. As in Figure 4.7.6 but for the worst case 3 in Table 4.7.2.

IDEAL CASE: ERRORS OF $\hat{b}_i, \hat{a}_i$

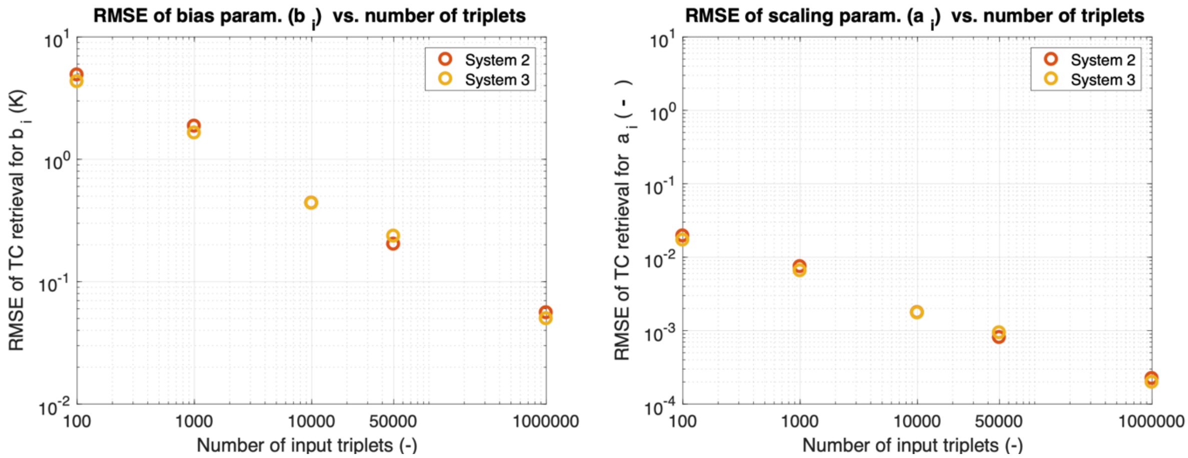


Figure 4.7.11. RMSE of calibration bias $\hat{b}_i$ (left) and calibration scaling $\hat{a}_i$ (right) for $i = 2{,}3$ as a function of the number (N_S) of triplets in the ideal case of Table 4.7.2.

INTERMEDIATE CASE: ERRORS OF $\hat{b}_i, \hat{a}_i$

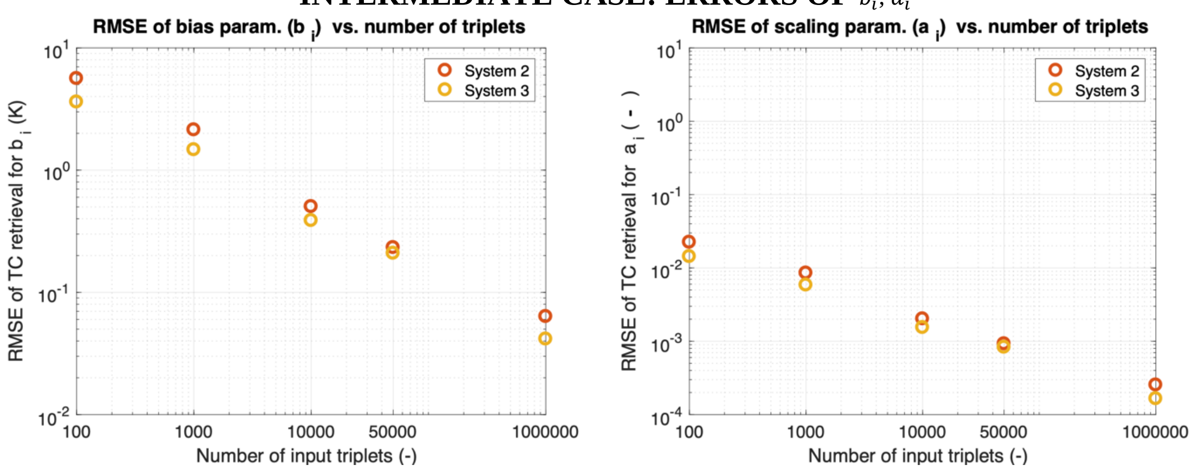


Figure 4.7.12. As in Figure 4.7.11 but for the intermediate case in table table 4.7.2.

WORST CASE 1: ERRORS OF $\hat{b}_i$, $\hat{a}_i$

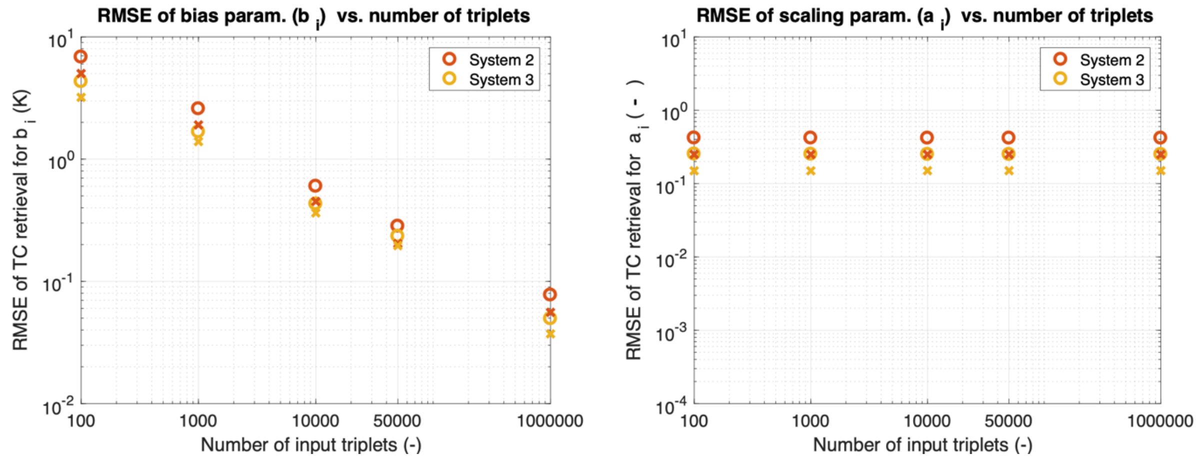


Figure 4.7.13. As in Figure 4.7.11 but for the worst case 1 in table Table 4.7.2.

WORST CASE 2: ERRORS OF $\hat{b}_i$, $\hat{a}_i$

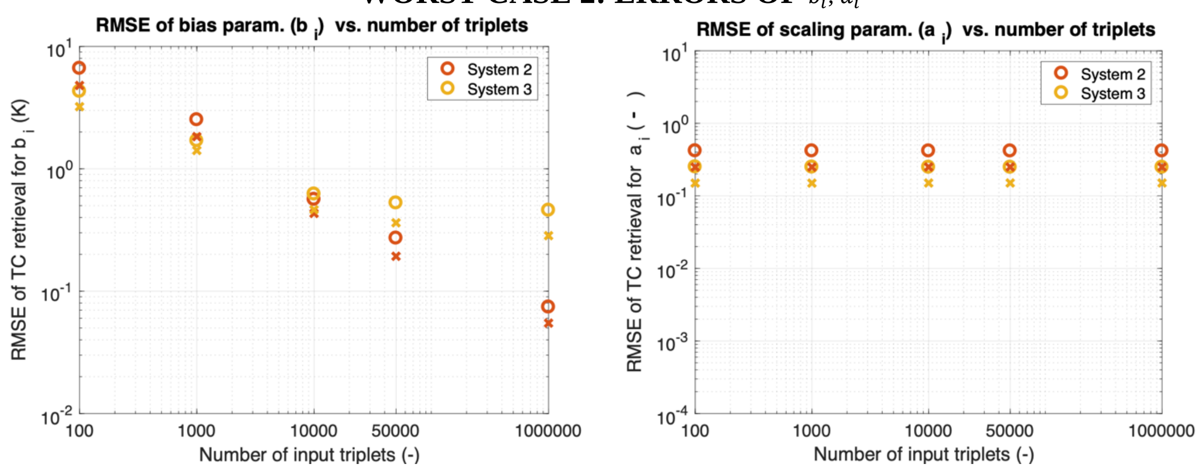


Figure 4.7.14. As in Figure 4.7.11 but for the worst case 2 in table Table 4.7.2.

WORST CASE 3: ERRORS OF $\hat{b}_i$, $\hat{a}_i$

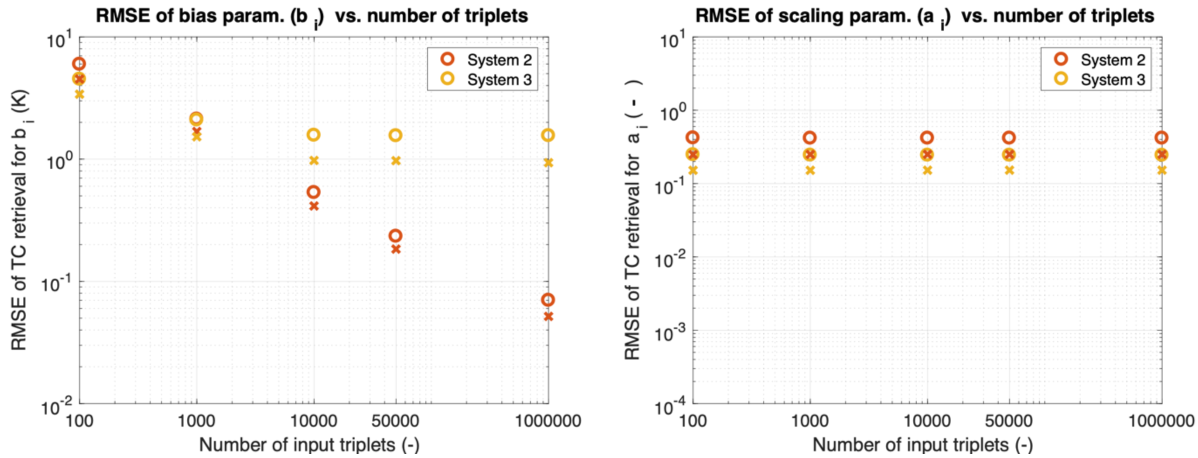


Figure 4.7.15. As in Figure 4.7.11 but for the worst case 3 in table table 4.7.2.

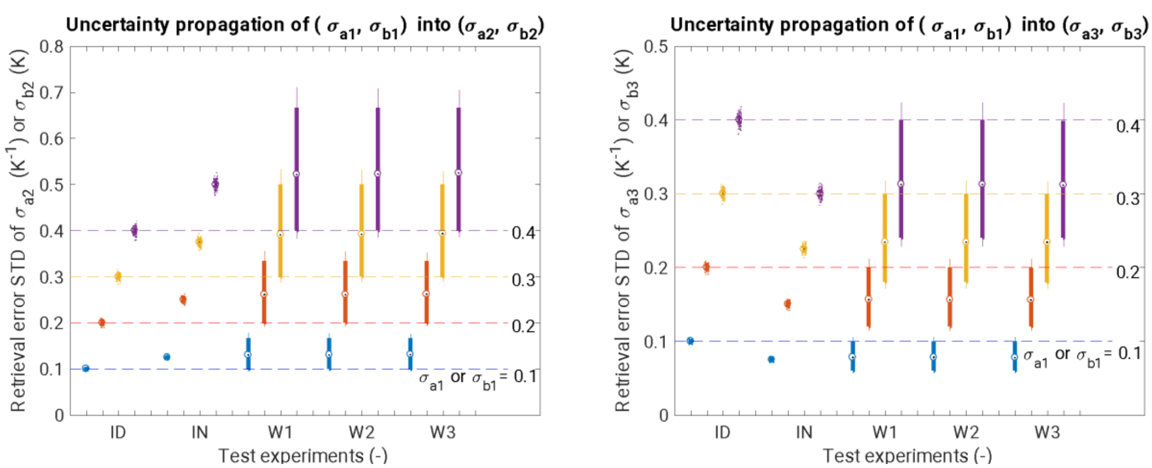


Figure 4.7.16. Uncertainty propagation of $(\sigma_{a_1}, \sigma_{b_1})$ into $(\sigma_{a_2}, \sigma_{b_2})$ (left) using eq.s (4.7.9), (4.7.11) and into $(\sigma_{a_3}, \sigma_{b_3})$ (right) using eq.s (4.7.10), (4.7.12), as a function of the test experiments in Table 4.7.2.

5. VICIRS tool description

This section describes the design of the VICIRS tool, analyzing each part with reference to the methodology analysis discussed in the previous sections. VICIRS tool input/output is indicated for each task, together with the processing blocks. The tool is intended for the vicarious calibration of both ICI and MWI. It has been designed for validating ICI and MWI observations against RSs collected from GRUAN and RHARM data archives in terms of BT. Moreover, VICIRS tool handles Numerical Weather Prediction (NWP) profiles and RS profiles as indicated in subsection 4.5. BT simulations are performed by the GRUAN processor. For simplicity, MWI and ICI data hereinafter will be referred to as SAT data.
The VICIRS-tool flowchart (Figure 5.2) can be divided into 7 Steps (Figure 5.1), grouped in the following 3 Blocks:

1. **Block 1** is the front-end of the VICIRS tool; it takes in input the search parameters set by the user in *config.ini*, finds the match-ups between SAT and RS and downloads the RS/SAT/NWP data related to the match-ups;
2. **Block 2** is the core of VICIRS tool; it consists in the loop on match-ups finalized to collect the data useful for the statistical analysis;
3. **Block 3** depends on the previous blocks, it performs (i) statistical analysis of differences between BT observed and simulated and related uncertainties for each SAT channel and for each match-up; (ii) BIAS and uncertainty analysis for all the match-ups; (iii) Multisource Correlative Methodology (MCM) analysis**.**

The VICIRS-tool Steps can be in more than one block. In particular, **Step *** is present in all the Blocks, calling all the other Steps and handling their input/output. In **Step *, main.py** reads the user settings in *config.ini* and transmits them to **Step I,** which searches the SAT/RS match-ups, downloads the corresponding data and optionally the collocated NWP (if activated in *config.ini*). These operations are included in **Block 1**. Once the list of match-ups is available, *main.py* begins the loop on the match-ups (included in **Block 2**) and passes the information about RS and SAT orbit to **Step II,** which analyses the RS to decide if it can be included in the calibration process. If the RS passes all the tests in **Step II**, **main.py** calls the **Step I** module for extracting the circular Target Area (TA) and transmits TA and RS information to **Step III. Step III** extracts two other TA types from circular TA and analyzes all the three TAs in terms of cloud screening and emissivity screening. The results of the TA-analysis as well as information about RS are written in a *netcdf* file. **Step IV** writes the input files for the GRUAN-processor and launches the GRUAN-processor to compute simulated BT. The output is then read by the **Step-V** module that computes the observed (BT_O) minus simulated (BT_S) BT difference and the related uncertainty, updating the *netcdf* file from **Step III**. The *main.py* moves on to the next match-up in the list. For each match-up a *netcdf* file is generated that can be queried in **Block 3** to build the sub-sample used for statistical analysis of bias and uncertainty and in, **Step VI,** for the MCM analysis.

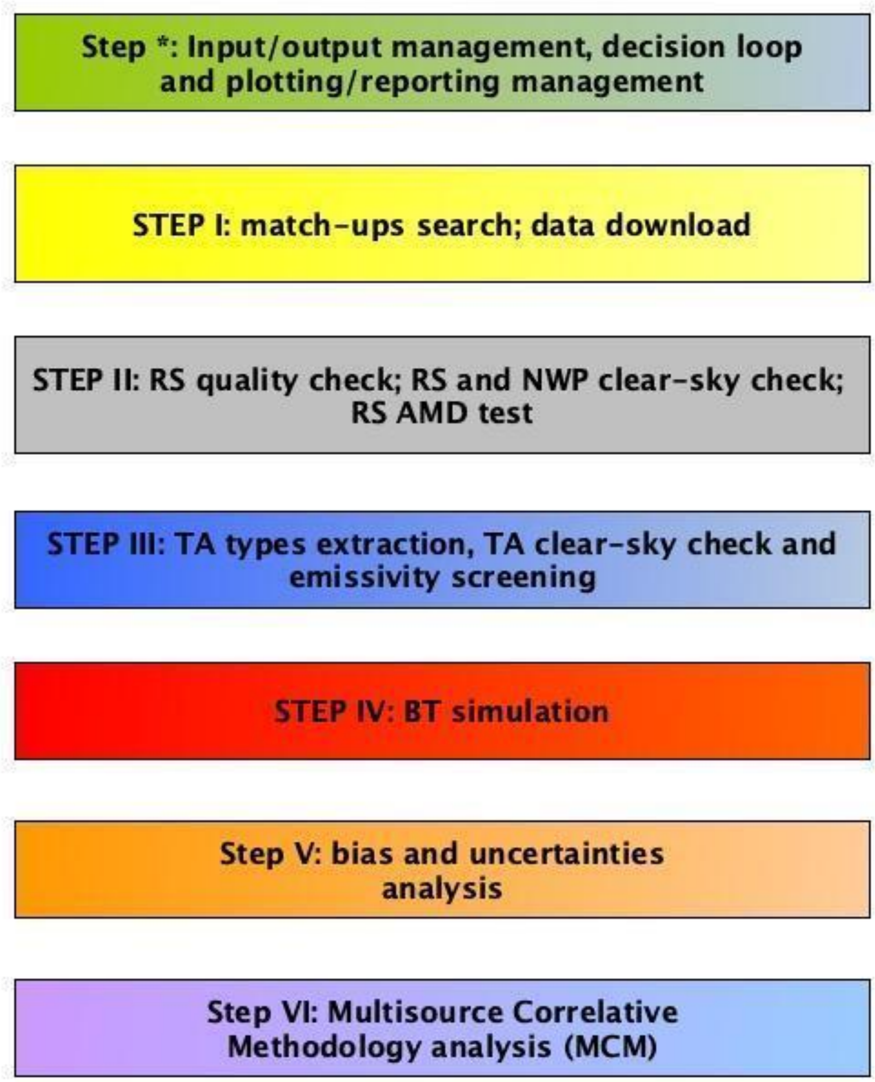


Figure 5.1 VICIRS tool Steps description. Each step is colored in a different manner to identify it in the flowchart.

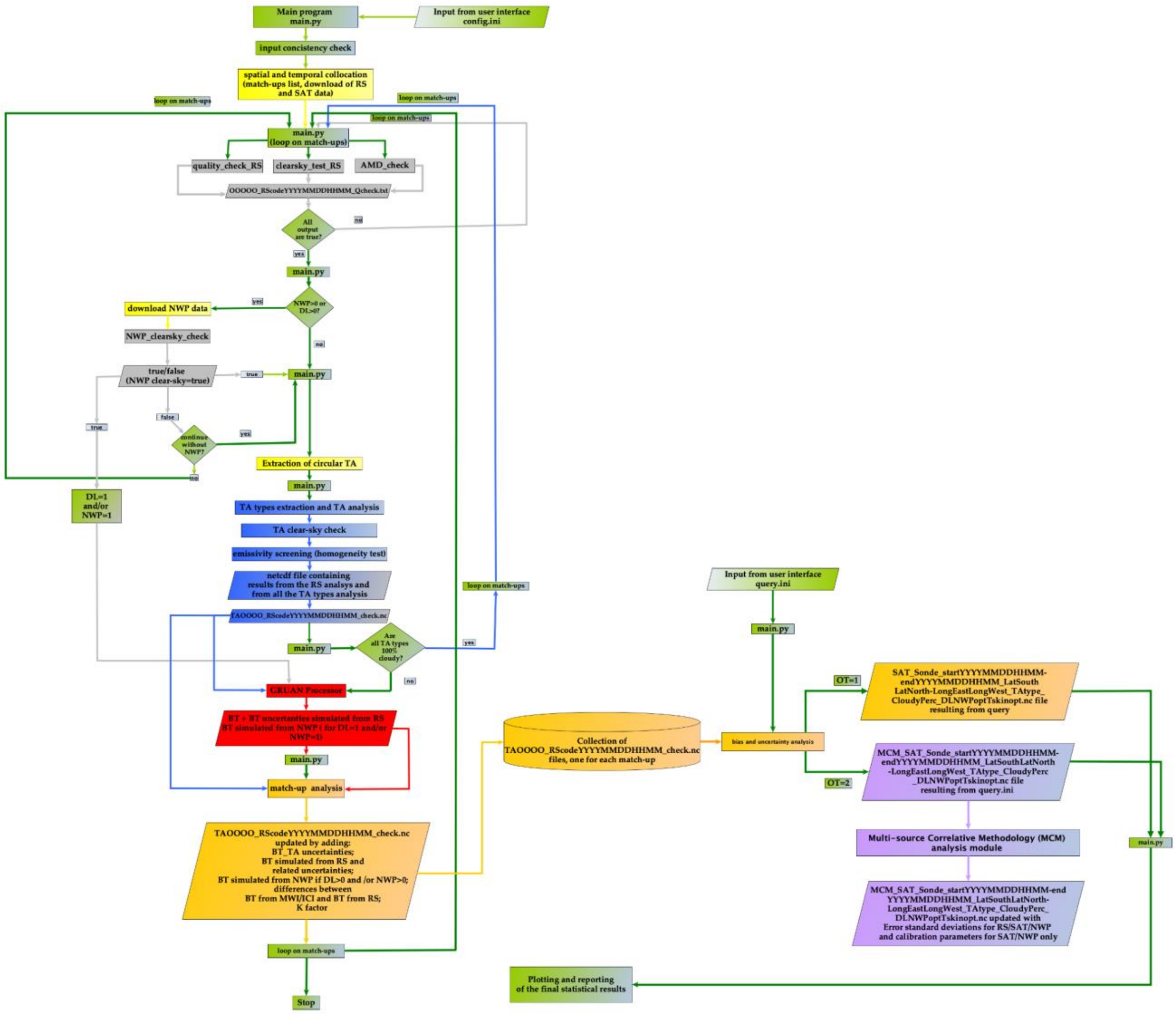


Figure 5.2 VICIRS-tool flow chart

5.1 Block 1: match-up search and data download

The match-up search is done on the basis of the input parameters set by the user in *config.ini* (Table 5.1.1). VICIRS Blocks-1 and -2 are designed to collect a dataset of match-ups characterized by temporal distance (between radiosonde launch and satellite overpass) ranging from (-15m/+45 min) to (-3h/+3h) and globally distributed according to the radiosonde sites. The first two Blocks are preparatory to the Block-3 statistical results analysis. To optimize the procedure, MWI and ICI observations are collected and processed simultaneously in Blocks 1 and 2, whereas, in Block 3, the user can choose only one instrument for carrying out the statistics. In *config.ini* the user can set several parameters. In detail, the *config.ini* parameters consist in the choice of:

1. the radiosonde archive (RHARM or GRUAN);
2. the temporal range;

3. the temporal distance. The default value is option 3 (-3h/+3h), which is preferable for collecting all the spatial match-ups available (also for the other options without repeating the search); the statistical analysis can then be performed on shorter time distances;
4. the spatial range. The default value is the whole Earth (90/-180/-90/180);
5. the use of Dedicated Launches (DL=1) or operative launches (DL=0) only. The option for using NWP information (Table 5.1.1 option 6) also in operational launches can be activated by setting NWP_opt=1, as explained in the next point;
6. the use of NWP information to supply NWP surface parameters to RTTOV (GRUAN processor) for BT simulation when surface information is not available from RS (NWP_opt=1);
7. source of skin temperature, T_{skin} (activated in case NWP_opt=1). Default choice: deriving T_{skin} from the difference between the T_{skin} (NWP) and the NWP 2-meter temperature (T_{2m}(NWP)) added to the T_{2m} calculated from the RS profile (Carminati et al., 2019); optional choice: T_{skin} equal to the model T_{skin} (T_{skin} (NWP));
8. RLF option for selecting only the FOVs with LF≤RLF to be included in TA (when LF≤100 all the FOVs are selected).

Before transmitting the information set in *config.ini* to the modules for searching match-ups, *main.py* checks the consistency of the parameters set in 1), 5), 6) and 7). The DL option is considered only when the option 1) about the radiosonde archive is 1 (GRUAN). The Tskin_opt (7) is considered only when the NWP_opt has been activated with NWP_opt=1 and/or when DL=1 (option 5). In case of inconsistency between the input parameters, *main.py* stops the VICIRS flow and requires consistent inputs.

Table 5.1.1 Input from user: ***config.ini***.

Parameter	Value
1) radiosonde archive	RS=1(GRUAN),RS=2(RHARM)
2) temporal range start	*aaaa-mm-dd hh:mm*
stop	*aaaa-mm-dd hh:mm*
3) temporal distance (default: 3)	**1** (-15m/+45min); **2** (-1h/+1h); **3** (-3h/+3h)
4) spatial range (default: 90/-180/-90/180)	N/W/S/E
5) dedicated launches	**DL=0** operative launches; **DL=1** dedicated launches, if NWP cloudy: use RS only; **DL=2** dedicated launches; if NWP cloudy: move to next match-up
6) NWP_opt (for supplying or not NWP surface information for simulating BT_RS)	**NWP_opt=0** no NWP; **NWP_opt=1** use NWP; if NWP cloudy use RS only; **NWP_opt=2** if NWP cloudy: move to the next match-up
7) Tskin_opt (only if NWP=1)	**Tskin_opt=0** T_{skin} determined from RS **Tskin_opt=1** $T_{skin} = T_{2m}(RS) + (T_{skin}(NWP) - T_{2m}(NWP))$; **Tskin_opt=2** $T_{skin} = T_{skin}(NWP)$
8) Maximum Land Fraction LF (%) of the SAT FOV	**RLF=LF** (0<=LF<=100) for selecting MWI /ICI FOVs to be included in TA

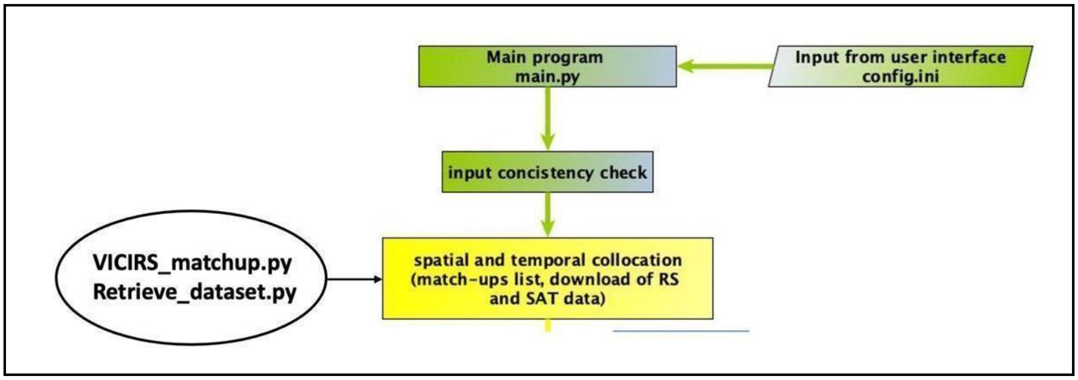


Figure 5.1.1 VICIRS tool flowcharts: Block 1, Step 1.

After consistency check, main.py passes config.ini information to the **pyvicirs.matchup** and **pyvicirs.retrieve_dataset** modules. The **pyvicirs.matchup** module has two main function, *find_matchup_from_obs* to get all match-up occurrences from observations dataset (dataset already available or retrieved from the *retrieve_dataset* module), and *find_matchup_from_tle*, to find overpasses by reconstruction of the satellite orbit and FOVs using orbital element (TLE) and geometry characteristic of the sensors.
The **pyvicirs.retrieve_dataset** module provides four main functions to download public dataset by date and time: *GMIDownloader*, *RHARMDownloader* and *RS41Downloader, NWPDownloader* to retrieve respectively GMI granule from NASA, RHARM radiosondes dataset from Climate Data Store, RS41 radiosondes dataset from GRUAN and NWP dataset from ECMWF data archive.

The Block-1 output is the file **list_out_I.csv** listing the information of match-ups useful for the processing of RS and SAT data (see Table 5.1.2). Furthermore, a module for TA extraction from SAT observations is also present in Step I. The **pyvicirs.ta_creator** module has two main functions, *MetopTACircular* and *GPMTACircular* to create circular TA which are then passed to Step III to analyse it (see subsection 5.2.2). There is also a **pyvicirs.utils** module containing several functions shared with the other modules and a **pyvicirs.read_config** module to read options from *config.ini* file.

Table 5.1.2 - Example of output from Step I.

rs_code	sat_orbit	scan_time_matchup	sat_filename	rs_filename
POT	54641	2023-10-10 22:31:55.000085	1B.GPM.GMI.TB2021.20231010-S215820-E233052.054641.V07A.HDF5	POT-RS-02_2_RS41-GDP_001_20231010T220000_1-000-001.nc

5.2 Block 2: loop on match-ups

Block 2 is the core of the VICIRS tool, including Steps from I to V. In particular, it includes Step II for the analysis of the RS, Step I for extracting the circular TA and Step III for the analysis of the TA, Step IV for the BT simulation by using GRUAN processor modified for the VICIRS purposes, Step V for collecting $BT_{observed}$-$BT_{simulated}$ and the related uncertainty for each match-up.

5.2.1 Step II: RS and NWP analysis

Step II deals with RS analysis in terms of numbers of levels (**nlev**) and pressure minimum value (**Pmin**), Air Mass Displacement (AMD), cloud contamination, and NWP analysis in terms of cloud contamination. Step-II flow chart is shown in Figure 5.2.1.

The RS analysis is mandatory to decide if the related match-up is useful for calibration purposes. main.py passes the *RS filename* and the *satellite overpass time* to the executable of the Fortran90 modules :

- *VICIRS_qualitycheck_GRUAN.f90 (VICIRS_qualitycheck_RHARM.f90)* checks the availability of the Temperature (T), Pressure (P), Relative Humidity (RH) profiles and related uncertainties, wind speed and wind direction (or the meridional and zonal wind component) at 10 m to be used as surface parameters in BT simulation. It gives in output *FlagRS=True (1)* when **nlev**≥ *40* for P/T/RH profiles and the **Pmin≤10hPa** ([AD-8], Section 5). *VICIRS_qualitycheck_GRUAN.f90* outputs also the *flagSP* about the availability of surface parameters information that is usually stored in the *global attributes* of GRUAN RS data file. FlagSP=True (1) when the surface-parameter information is available. *FlagSP* is considered only in case of GRUAN RS and NWP_opt>0 for deciding to use NWP or not for supplying the surface information for simulating BT: (i) NWP_opt>0 and FlagSP=1, surface parameters from GRUAN-RS *global attributes* are used*; (ii)* NWP_opt>0 and FlagSP=0, NWP surface information is used;
- *VICIRS_clearcheck_RS.f90* checks the presence of cloudy layers by comparing the RH values with the reference values for clear sky as determined by Zhang et al. (2010). Note that this method was developed for RS92 sonde, but its performances with RS41 have been assessed (see [AD-8], Section 3.2). It outputs the number of levels contaminated by low, middle and high clouds. The RS is defined clear-sky when the number of contaminated layers is 0, and *VICIRS_clearcheck_RS.f90* outputs Flag=1 (True);
- *VICIRS_AMD_GRUAN.f90 (VICIRS_AMD_RHARM.f90)* determines AMD ([AD-8], Section 3.2) by multiplying the temporal distance between the satellite overpass and sonde launch, Δt, by the wind speed average between 700 hPa and 300 hPa, $\underline{w}$. The output is Flag=1 (True) when $AMD \leq TA\, radius$.

The match-up is considered for the validation only if all the three outputs are *True* (flag=1), otherwise the match-up is discarded and not considered for the calibration process. After, *main.py* moves on to the next match-up. The *VICIRS_qualitycheck_GRUAN.f90 (VICIRS_qualitycheck_RHARM.f90)* executable collects the output of the three tests in the intermediate file ***RSfilename_OOOOO_check.nc*** (whose fields are listed in Table A.1 in appendix A) where **RSfilename** is the RS name and **OOOOO** is the satellite orbit number. ***RSfilename_OOOOO_check.nc*** files are queried from Step-V modules for generating the statistics related to RS used and discarded from the calibration process.

If **RS quality check flag=1** and the option 5) or option 6) in config.ini have been activated (**NWP_opt>0** and/or **DL>0)**, *main.py* calls the module for downloading NWP profiles, i.e. the *NWPDownloader* function from the **pyvicirs.retrieve_dataset** module. NWP file is checked by the *VICIRS_clearcheck_NWP.f90* module to detect any low, medium or high cloud layers in the NWP profile spatially and temporally closest to the radiosonde launch. *VICIRS_clearcheck_NWP.f90* returns values ranging between 0 and 1 of the four cloud fractions *tcc* (total cloud cover), *lcc* (low cloud cover), *mcc* (medium cloud cover) and *hcc* (high cloud cover). The module also provides information about the cloud presence in the temporally next three NWP profiles, at t_0+3h, t_0+6h and t_0+9h. This further information can be useful for determining whether the scenario is clear or cloudy. In case of cloudy NWP, *main.py*

- moves on the next match-up if options 5) and/or 6) are **DL=2** and/or **NWP_opt=2**;
- moves to Step III for TA analysis without considering NWP, if **DL=1** and/or **NWP_opt=1**.

When **RS quality check flag=1** and **DL=0** and **NWP=0**, *main.py* calls the Step I module ***pyvicirs.ta_creator{****MetopTACircular, GPMTACircular****}*** and passes the circular TA to Step III for analyzing it.

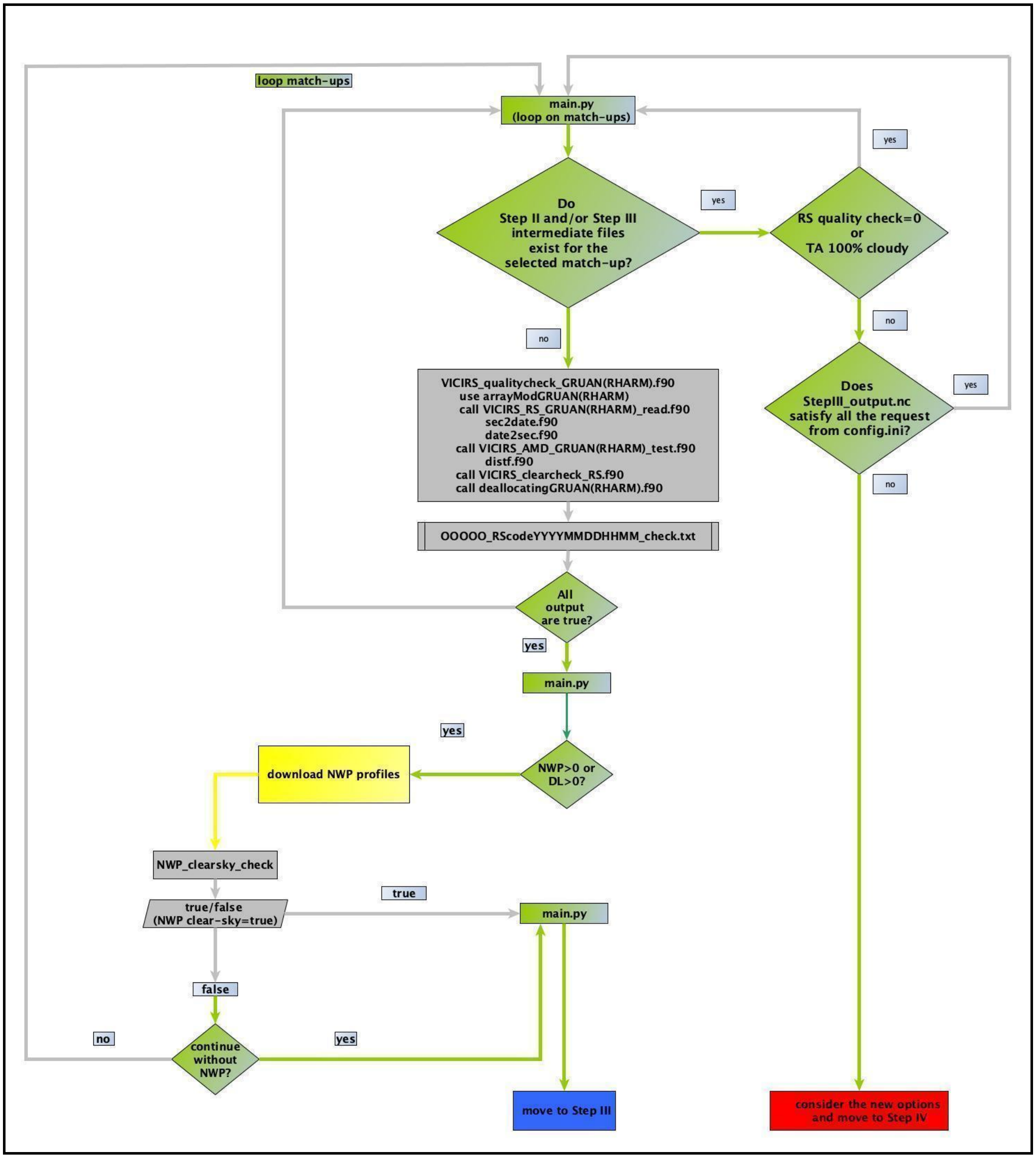


Figure 5.2.1 VICIRS tool flowcharts, Block 2 Step II: RS and NWP analysis.

5.2.2 Step III: TA types extraction and analysis

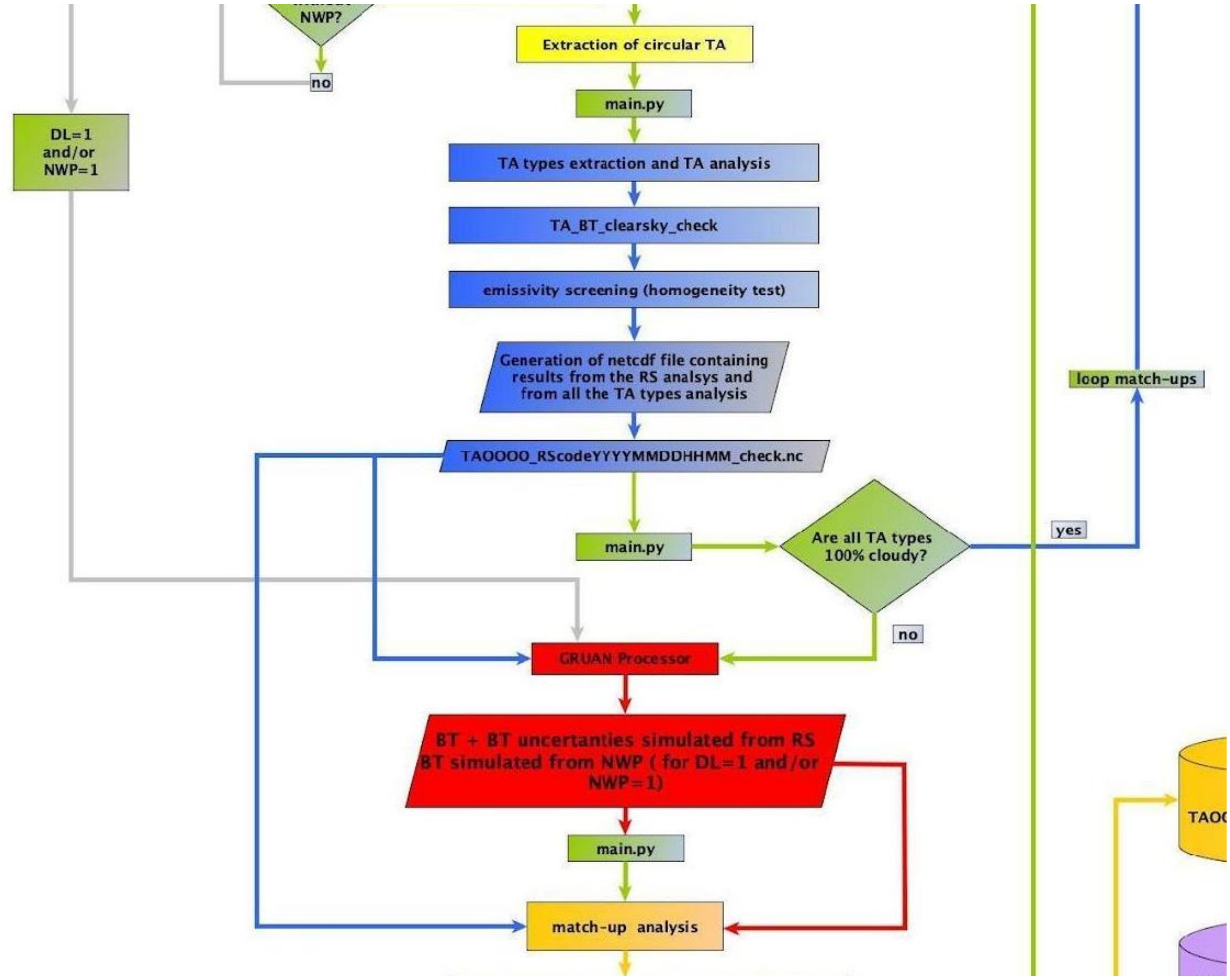


Figure 5.2.2 VICIRS tool flowcharts, Block 2: Step III (TA analysis), Step IV (GRUAN processor) and Step V (bias and uncertainty analysis).

On the basis of the spatial collocation proposed and analyzed in subsection 4.1, Step III defines three types of TA that are implemented for both ICI and MWI starting from the classical circular TA extracted by ***pyvicirs.ta_creator*** from the MWI/ICI file indicated in **list_out_l.csv***.*

The procedure for processing the TA types is different for GRUAN RS and RHARM RS, depending on the availability of the latitude and longitude values for each pressure level that are always available in GRUAN but not in the RHARM current version. For this reason, the number of TA types related to GRUAN RS and RHARM RS is 5 and 3.

In detail, when latitude and longitude are not available for each RS pressure level (RHARM RS), it is possible to define the first three types of TA as described in subsection 4.1. Moreover, for GRUAN RS, the TA types 4 and 5 are also considered.

The lack of latitude/longitude information for RHARM-RS pressure levels influences not only the number of TA types to be investigated but also the definition of the TA radius (**TA$_R$**) and fraction of RS-sonde path over land (**land_frac_RS**). Consequently, Step III has been designed differently depending on the availability of latitude and longitude values for the RS pressure levels. In case of GRUAN RSs, the procedure for defining **TA$_R$** and **land_frac_RS** is based on the compared analysis of the circular TA data and the matched RS profile, in detail:

- TA$_R$ default value is 50 km (Buehler et al. 2004, Moradi et al. 2010, Bobryshev et al. 2018), but TA$_R$= *RS-sonde path length* when *RS-sonde path length* $<$ 50 km;
- **land_frac_RS** is determined for each SAT channel by associating to each RS pressure level the land fraction of the SAT FOV closest to it and averaging on all the pressure levels. RS land fraction is memorized in the Step-III output file as *land_frac_RS*, a 26-elements array for MWI and 13-elements array for ICI. In detail:

$$land\_frac\_RS(i) = \frac{\sum_{l=1}^{nlev} lf\_sat(i,l)}{nlev}$$

where *i* is the *ith SAT*channel, lf_sat(i,l) is the land fraction of the FOV closest to the lth sonde-path pressure level for the *ith SAT* channel and $nlev$ is the number of the sonde-path pressure levels .

Current RHARM archive does not provide any information about the pressure level position, and consequently, when **RHARM RS** are considered, **TA$_R$** and ***land_frac_RS*** are defined as follows:

- **TA$_R$**=50 Km;
- ***land_frac_RS*** for each MWI/ICI channel is the average of the land fraction of all the FOVs included in the circular TA. RS land fraction, memorized in the Step-III output file as *land_frac_RS* is determined as follows:

$$land\_frac\_RS(i) = \frac{\sum_{j=1}^{nFOVS\_TA} lf\_sat(i,j)}{nFOVs\_TA}$$

where *i* is the *ith SAT channel*, lf_sat(i,j) is the land fraction (lf) of the jth FOV included in TA for the *ith SAT channel* and $nFOVs\_TA$ is the number of the FOVs included in TA.

Land_frac_RS is essential for defining the RS surface type (**surf_type**) that will be passed to Step IV in order to simulate BT from RS (and NWP). **surf_type** is defined by averaging ***land_frac_RS*** for each channel over the number of channels (*nf*) and it will be:

1) ***surf_type=0***, sea surface, when $\frac{\sum_{i=1}^{nf} land\_frac\_RS(i)}{nf} = 0$;
2) ***surf_type=1***, land surface, when $\frac{\sum_{i=1}^{nf} land\_frac\_RS(i)}{nf} = 1$;

3) ***surf_type=2***, mixed, when $0 < \frac{\sum_{i=1}^{nf} land\_frac\_RS(i)}{nf} < 1$.

In 1) and 2) the surf_type is uniquely defined. In case 3) it is necessary to introduce a land-fraction threshold below which RS surf_type is sea surface, but this could not be true for all the channels. To overcome this question, when ***surf_type=2,*** two sets of BT are simulated from RS (NWP) by setting surf_type=0 (BT_RS(NWP)_sea) and surf_type=1 (BT_RS(NWP)_land); the final BT (BT_RS) is computed from BT_RS_sea and BT_RS_land weighted according to the land fraction for each channel (i) and TA type (j). In detail:

$$BT\_RS(i,j) = LAND\_FRAC\_TA(i,j) \cdot BT\_RS\_land(i) + (1 - LAND\_FRAC\_TA(i,j)) \cdot BT\_RS\_sea(i)$$

The field $LAND\_FRAC\_TA$ is the average of the land fraction of the FOVs included in the TA, for each TA and for each channel.

After the extraction of TA types and the definition of RS surf_type, the executable of *VICIRS_clearcheck_TA.f90* applies the tests listed in Table 4.2.1 and Table 4.2.2 (subsection 4.2) to the MWI and ICI observations, respectively. The match-up is avoided from the calibration process only when all the MWI/ICI tests detect as cloudy the 100% of the FOVs included in all the TA types. The percentage of cloudy FOVs included in the 3 (5) TA types as well as the (BT_TA, SD_TA) for all the TA types and MWI/ICI channels, are written in the netcdf file **TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc**. **TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc** contains all the information to be supplied for initializing the GRUAN processor (Step IV).

After BT simulation from RS profile and, in case of DL=1, from NWP profile, **TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc** will be updated by adding the BT simulated (BT_RS, uBT_RS) and, if present, (BT_NWP).

Figure 5.2.2 shows the scheme of Step III and how it is linked to Step IV and Step V. *main.py*, when the match-up is not rejected, calls the executable of the *VICIRS_write_gprocnl.f90* that writes the two *namelist* (for ICI and MWI) for initializing the GRUAN processor.

5.2.3 Step IV: GRUAN processor, BT simulation from GRUAN/RHARM RS and NWP profile

In Block 2, MWI and ICI BT are simulated separately for the same match-up. *main.py* calls the GRUAN processor (Carminati et al. 2019) and provides the following parameters through the file *namelist.nl* (the parameters **in bold** are mandatory):

- ***inst***: “mwi”/ “ici”;
- ***rs_opt:*** "gruan"/"rharm"

- ***surf_type***= 0 for sea,1 for land, 2 for mixed surf_type as defined in Step III (subsection 5.2.2);
- ***satellite zenith angle***: from Step III, the values corresponding to the MWI/ICI FOV closest to the sonde launch site; it is a 26-elements array when *inst*=”mwi” and a 13-elements array when *inst*=”ici”;
- ***satellite azimuth angle***: from Step III, the values corresponding to the MWI/ICI FOV closest to the sonde launch site; it is 26-elements array when *inst*=”mwi” and a 13-elements array when *inst*=”ici”;
- ***sonde_datafile***= path + RS profile name to be processed for simulating the corresponding BT and its uncertainty;
- ***T_{skin_opt}* =1 / 2** option 7 in *config.ini* is considered only when the field “*model_datafile*” is not empty, otherwise surface parameters are determined directly from RS (values closest to z=2m or to z=10m for wind components, where z is the *altitude field* in GRUAN RS and *geopotential_height field* in RHARM RS). Table 4.3 shows how the surface parameters are defined on the basis of the RS archive and of the user options set in *config.ini*;
- ***model_datafile***=path + NWP profile name to be processed for simulating BT from NWP and/or for supplying the estimate of the surface parameters in the BT_RS.

Other information concerns the platform (*metopsg)*, the model data type, rttov_rmse and rttov_bias (different for ICI and MWI), and indication about the output file as well as the output destination.
Regarding the emissivity model used in RTTOV, SURFEM-Ocean (Kilic et al, 2022) is chosen for sea. SURFEM-Ocean is a new microwave sea surface emissivity model available in RTTOV v13.2 valid across 0.5-700 GHz frequencies that should replace all FASTEM and TESSEM2 versions (https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs_rttov13/users_guide_rttov13_v1.2.pdf, Hocking et al. 2022b). The emissivity model for land/ice is the TELSEM2 (Wang et al, 2016). *RTTOV v13 Users Guide* (2022) recommends TELSEM2 emissivity atlas instead of FASTEM land/sea-ice parameterization that will be deprecated in the future RTTOV versions.
The GRUAN processor version 6.3 was made available at the kick-off of VICIRS. This has been modified for the purposes of VICIRS, starting a new branch, currently at version 6.3.b.0.1. In the following, for GRUAN processor, we refer to version 6.3.b.0.1, if not otherwise specified. The new version of GRUAN processor works with GRUAN/RHARM RS (operational launches) without using information from spatially and temporally collocated NWP profiles. In addition, the GRUAN/RHARM RS profiles can be processed in combination with NWP data in case of DL=1 (option 5 in *config.ini*) (GRUAN DL) or if NWP_opt=1 (option 6 in *config.ini*)**.**

It is important to underline that NWP profile can be used also in case of operational GRUAN/RHARM RS (in *config.ini*, option 5 DL=0) by setting NWP_opt=1 (option 6) and, consequently, Tskin_opt=1/2 (option 7).

In the GRUAN processor an internal loop has been introduced on the different satellite angles to take into account that the zenith and azimuth angles vary for each ICI/MWI channel. In addition, also the effect of the vertical ozone profile has been inserted. In detail, in the case of DL=1 (option 5 in *config.ini*) (GRUAN DL) or if NWP_opt=1 (option 6 in *config.ini*) the ozone profile is provided from the NWP data otherwise the RTTOV reference ozone profile is used.

In case of operational GRUAN/RHARM RS and GRUAN DL, the GRUAN processor takes in input:

1. GRUAN/RHARM P, RH and T profiles with related uncertainties;
2. GRUAN/RHARM wind speed profiles to determine V_{10m} and U_{10m} (Table 5.2.3);
3. RS surface parameters (defined on the basis of *config.ini* options, Table 5.2.3);
4. NWP surface parameters and NWP P, T, Q (water vapor mixing ratio) profile over the RS top level when *option 6* NWP_opt=1 in *config.ini* for both GRUAN and RHARM RS, and option 5 DL=1 only for GRUAN RS.

Table 5.2.3 Surface parameters definition (z indicates the altitude, that is the *alt_asml* field in GRUAN RS and *geopotential_height* field in RHARM RS).

RS type	P_{2m}	T_{2m}	T_{skin}	V_{10m}	U_{10m}
GRUAN *g.SurfaceObs*: available **In config.ini: NWP_opt= 0**	P_{2m} *(g.SurfaceObs)*	T_{2m} *g.SurfaceObs*	$T_{skin}=T_{2m}$ *g.SurfaceObs*	**V_{10m}** =*V(z) value* closest to z_{10m}=10m (*if* $\|z-10m\|$<=*10m)* **V_{10m} =V_{2m}** (*g.SurfaceObs)* *(if* $\|z-10m\|$>*10m) and* V_{2m} *(g.SurfaceObs) if available*	**U_{10m} =*U(z)*** *value* closest to z_{10m}=10m (*if* $\|z-10m\|$<=*10m)* **U_{10m} =U_{2m}** *(g.SurfaceObs)* *(if* $\|z-10m\|$>*10m) and* U_{2m} *(g.SurfaceObs) if available*
GRUAN *g.SurfaceObs*: no available **In config.ini: NWP_opt= 0**	P_{2m}=P(z) closest to z=2m	T_{2m}=T(z) closest to z=2m	$T_{skin}=T_{first\ level}$	**V_{10m} =*V(z)*** closest to z_{10m}=10m	**U_{10m} =*U(z)*** closest to z_{10m}=10m
GRUAN *g.SurfaceObs*: available Config.ini-option 6: NWP= 1 option 7: Tskin_opt=1 or 2	P_{2m} *g.SurfaceObs*	T_{2m} *g.SurfaceObs*	$T_{skin}=T_{2m}+(T_{skin}(NWP)-T_{2m}(NWP))$ (Tskin_opt=1) or $T_{skin}=T_{skin}(NWP)$ (Tskin_opt=2)	**V_{10m} =*V(z)*** *value* closest to z=10m (*if* $\|z-10m\|$<=*10m)* **V_{10m} =V_{2m}** *(g.SurfaceObs)* *(if* $\|z-10m\|$>*10m) and* V_{2m} *(g.SurfaceObs) available*	**U_{10m} =*U(z)*** *value* closest to z=10m (*if* $\|z-10m\|$<=*10m)* **U_{10m} =U_{2m}** *(g.SurfaceObs)* *(if* $\|z-10m\|$>*10m) and* U_{2m} *(g.SurfaceObs) available*
GRUAN *g.SurfaceObs*: no available Config.ini-option 6: NWP= 1 option 7: Tskin_opt=1 or 2	P_{2m}(NWP)	T_{2m}(NWP)	$T_{skin}=T_{2m}(GRUAN)+(T_{skin}(NWP)-T_{2m}(NWP))$ (if Tskin_opt=1) or $T_{skin}=T_{skin}(NWP)$ (if tskin_opt=2)	**V_{10m}** *(NWP)*	**U_{10m}** (NWP)

RHARM Config.ini-option 6: NWP= 0	P_{2m}=P(z) closest to z=2m	T_{2m} =T(z) closest to z=2m	T_{skin}=T_{2m}	**V_{10m} =*V(z)*** closest to z=10m	**U_{10m} =*U(z)*** closest to z=10m
RHARM Config.ini-option 6: NWP= 1 option 7: Tskin_opt=1 or 2	P_{2m}(NWP)	T_{2m}(NWP)	$T_{skin} = T_{2m}(RHARM) + (T_{skin}(NWP) - T_{2m}(NWP))$ (Tskin_opt=1) or $T_{skin} = T_{skin}(NWP)$ (Tskin_opt=1)	*v10(NWP)*	u10(NWP)

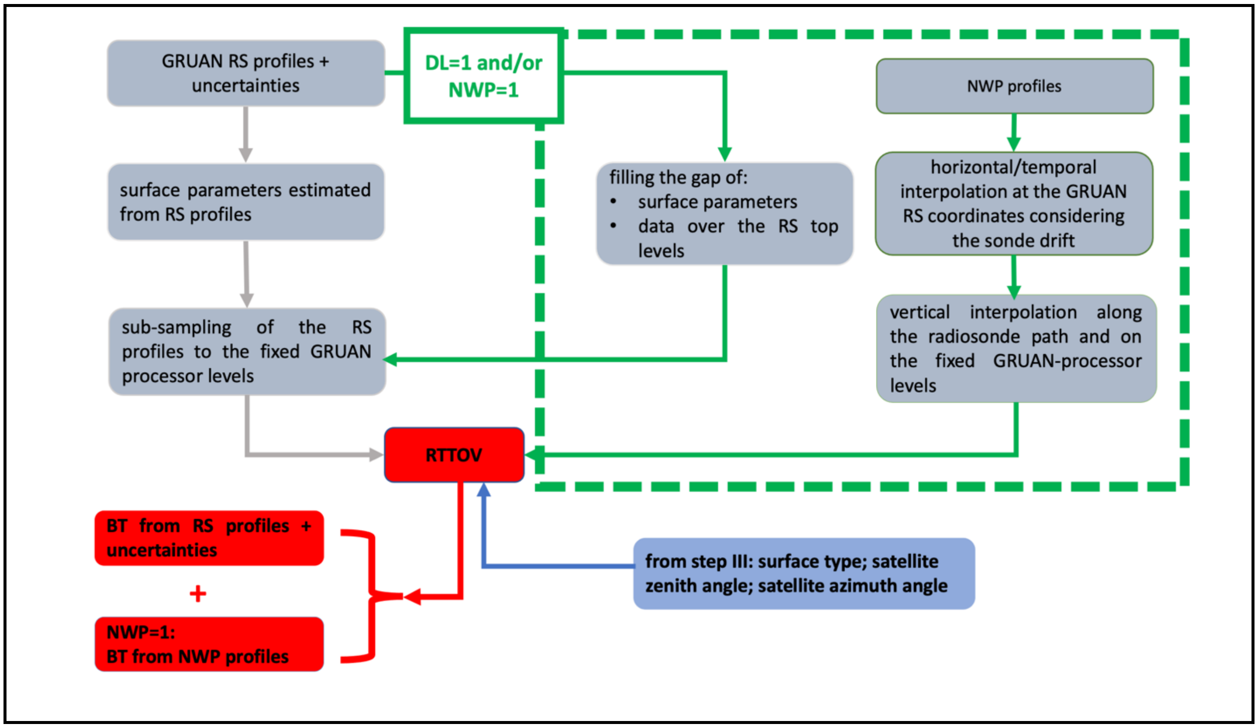


Figure 5.2.3a GRUAN-processor flow-chart scheme with GRUAN RS.

GRUAN-processor scheme is different for RHARM-RS (Figure 5.2.3a) and GRUAN RS (Figure 5.2.3b) depending on the different number of levels characterizing GRUAN RS (vertically dense profile) and RHARM RS (less dense profile) and on the different surface parameters often available in GRUAN RS but never available in RHARM RS (Table 5.2.3).

In detail, the GRUAN-RS profiles usually count several thousand levels, and thus are subsampled to the nearest level of the 310 selected fixed GRUAN-processor pressure levels, imposing that the ratio between the GRUAN profile pressure and the processor pressure is lower than 0.1% (Carminati et al. 2019). This subsampling is performed to homogenize the profiles that can be characterized by different vertical resolutions. On the other hand, RHARM-RS profiles are passed, as they are, to RTTOV for the BTs simulation without an intermediate interpolation to the fixed GRUAN-processor levels.

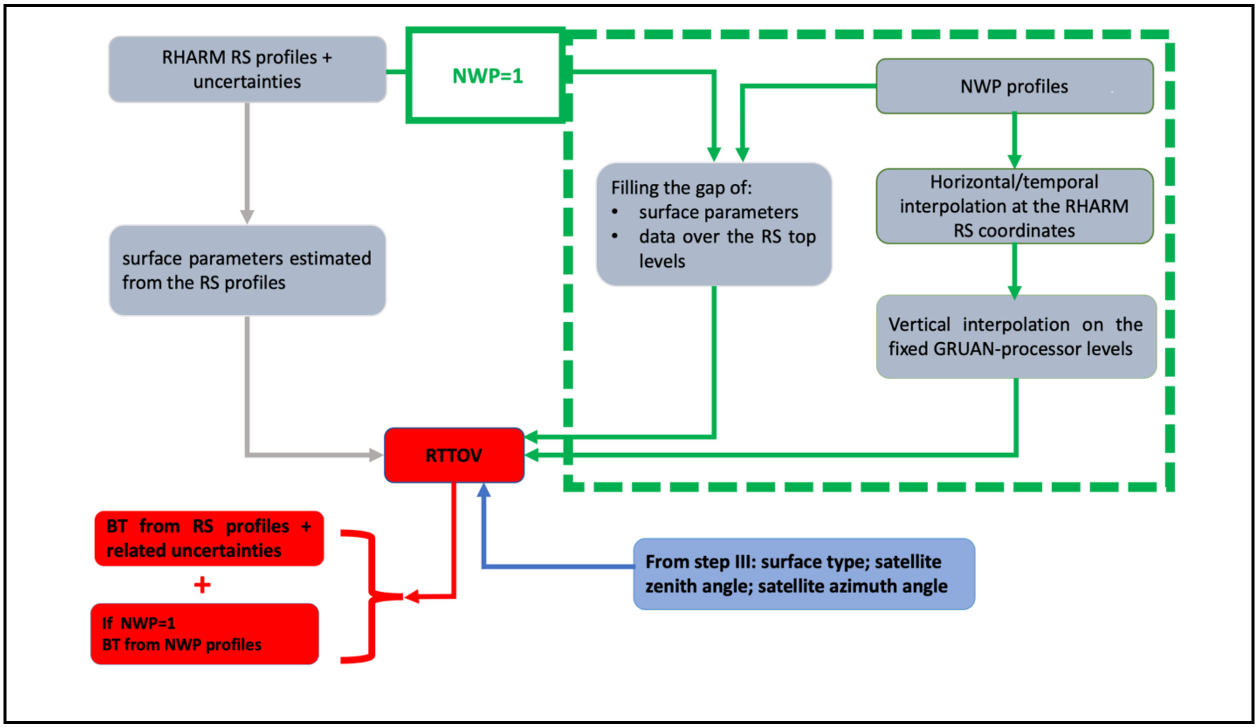


Figure 5.2.3b GRUAN-processor flow-chart scheme with RHARM RS. The RHARM-RS profile is not sub-sampled as for GRUAN-RS

The choice of not interpolating RHARM-RS comes out from the results obtained from the statistical analysis of MWI/ICI BT simulated from RHARM-RS compared with the BT simulated from corresponding GRUAN-RS (further details in [AD-8], Section 5). The use of the RHARM-RS interpolated profiles minimizes the BT-differences against the corresponding GRUAN-RS when nlev<40. The improvement in the range of (BT; SD) difference induced by the use of interpolated RHARM-RS profile with nlev≥ 40 and Pmin$\leq 11\ hPa$ is about **(3%; 5%)** for MWI simulation and **(-4%;-5%)** for ICI simulation. These results have been obtained for a sub-dataset D13 consisting of only 3 profiles. The statistical analysis will be updated when the RHARM v2 dataset will be available and until then, RHARM-RS will be passed to RTTOV without interpolation.

In case of surf_type=0/1, the GRUAN processor outputs consists of three types of (BT and related uncertainties) simulated both from GRUAN and RHARM RS, that will be indicated as:

1. (BT_RS, u_BT_RS) when NWP=0 (*config.ini* option 6 NWP_opt=0);
2. (BT_RS_1, u_BT_RS_1) when NWP_opt=1 and Tskin_opt=1 (*config.ini* option 6 NWP_opt=1 and option 7 Tskin_opt=1);
3. (BT_RS_2, u_BT_RS_2) when NWP_opt=1 and Tskin_opt=2 (*config.ini* option 6 NWP=1 and option 7 Tskin_opt=2);

4. (BT_NWP) BT simulated from NWP when DL=1 and/or NWP_opt=1.

When surf_type=2, GRUAN processor gives as output two sets of BT, the related uncertainties and BT_NWP for land surface (BT_RS_(*)_land, u_BT_RS_(*)_land, BT_NWP_land) and for sea surface (BT_RS_(*)_sea, u_BT_RS_(*)_sea, BT_NWP_sea), where '*' indicates the BT type. The final BT_RS, u_BT_RS and BT_NWP, will be calculated as follows:

1) $BT\_RS(i,j) = LAND\_FRAC\_TA(i,j) \cdot BT\_RS\_land(i) + (1 - LAND\_FRAC\_TA(i,j) \cdot BT\_RS\_sea(i)$
2) $u\_BT\_RS(i,j) = LAND\_FRAC\_TA(i,j) \cdot u\_BT\_RS\_land(i) + (1 - LAND\_FRAC\_TA(i,j) \cdot U\_BT\_RS\_sea(\mathrm{i})$
3) $BT\_NWP(i,j) = LAND\_FRAC\_TA(i,j) \cdot BT\_NWP\_land(i) + (1 - LAND\_FRAC\_TA(i,j) \cdot BT\_NWP\_sea(i)$

GRUAN-processor reports the results in the *Gproc-GRUAN_metopsg_2_mwi_RScode-YYYYMMDDHHMM(_land/_sea).nc* and *Gproc-GRUAN_metopsg_2_ici_RScode-YYYYMMDDHHMM(_land/_sea).nc* files for RS GRUAN and in the *Gproc-RHARM_metopsg_2_mwi_RScode-YYYYMMDDHHMM(_land/_sea).nc* and *Gproc-RHARM_metopsg_2_ici_RScode-YYYYMMDDHHMM(_land/_sea).nc* files for RS GRUAN. Moreover, when DL=0 and/or NWP=1, BTs are simulated from NWP profiles and the results are written in *Gproc-EC_metopsg_2_mwi_RScode-YYYYMMDDHHMM(_land/_sea).nc* and *Gproc-EC_metopsg_2_ici_RScode-YYYYMMDDHHMM((_land/_sea).nc* files. *main.py* passes the two GRUAN-processor output files to the *VICIRS_matchup_analysis.f90* that updates the Step III output *TAOOOOO_RScodeYYYYMMDDHHMM_check.nc* by writing the BT simulated by GRUAN-processor and the related uncertainties. In Table 4.4, the fields that are added in *TAOOOOO_RScodeYYYYMMDDHHMM_check.nc* by *VICIRS_matchup_analysis.f90* executable are marked in bold. For each match-up, only one type of the above-listed BT and BT uncertainties simulated from RS (1, 2 or 3) can be added to the *TAOOOOO_RScodeYYYYMMDDHHMM_check.nc,* and in case of DL=1 and/or NWP_opt=1 also BT_NWP is written.

5.2.4 Step V: match-ups statistical analysis

Step V concludes Block 2 by determining the residuals (TA_RS_SAT) between BT observed for each TA type (BT_TA) and the BT simulated from RS (BT_RS), the overall uncertainties (u_all) and the coverage factor k (subsection 4.6).

In detail, the value of the coverage factor k is determined considering the relation from Immler et al. 2010:

$$|m_1 - m_2| < k\sqrt{\sigma^2 + u_1^2 + u_2^2} \quad . \tag{5.2.1}$$

By replacing the terms in eq. (5.2.1) with the corresponding terms in TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc, the eq. (5.2.1) becomes:

$$TA\_RS\_SAT < K\_FACTOR \cdot u\_all \tag{5.2.2}$$

where $TA\_RS\_SAT = |BT\_TA - BT\_RS|$ is the difference between BT observed and simulated for each SAT frequencies and the related uncertainty u_all includes all the independent sources of uncertainties identified in (subsection 4.6), e.g.:

$$u\_all = \sqrt{u\_col^2 + u\_obs^2 + u\_sim^2} \tag{5.2.3}$$

where for all the MWI/ICI frequencies (j-th index) and for all the TA types (i-th index):

- ***u_obs***: uncertainty related to observation
  u_obs= $\sqrt{(NE\Delta T(i)/\sqrt{nFOVs(i,j)})^{2} + u_{geol}{}^2 + \ldots}$;
- ***u_col***: uncertainty related to collocation
  u_col= $\sqrt{SD\_TA(i,j)^2}$*;*
- u_sim is the uncertainty of simulated BT
  $u\_sim(i,j) = \sqrt{uBT\_RS(i,j)^2 + uBT\_ABS(j)^2 + uBT\_EMIS(j)^2 + uRTMlbl(j)^2 + uRTMlev(j)^2}$
  where
  - **uBT_RS** is the uncertainty related to the BT simulated from RS that accounts for T, RH and P profiles uncertainties (from Step IV). These quantities can be determined for NWP_opt=0 (**BT_RS**, **uBT_RS**), NWP_opt=1 and TSkin_opt=1(**BT_RS1**, **uBT_RS1**), NWP_opt=1 and TSkin_opt=2(**BT_RS2**, **uBT_RS2**);
  - *uBT_ABS* is the absorption model uncertainty of simulated BT due to uncertainties in H_2O and O_2 spectroscopic parameters;
  - *uBT_EMIS* is the uncertainty in surface emissivity (Wang et al, 2017);
  - uRTMlbl is the radiative transfer model (RTM) uncertainty obtained by comparing RTTOV radiance with those obtained by line-by-line calculation;
  - $uRTMlev$ is the RTM uncertainty due to the discrete levels versus dense level.
- ***K_FACTOR***: that assumes the value satisfying 5.2.2. In particular, when *k=1* the two measurements are consistent, when *k=2* they are in statistical agreement, while when *k=3* the measurements are inconsistent.

In Step V, the VICIRS_matchup_analysis.f90 executable *updates TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc* by adding the parameters described above. (Table A.2 in appendix A).

5.3 Block 3: statistical analysis

Block 3 (Figure 5.3.1) is the final part of the VICIRS tool. Here, the *TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc* files are used for implementing the bias and statistical analysis with related reporting and plotting (Step V described in subsection 5.3.1) and MCM analysis (Step VI described in subsection 5.3.2).

5.3.1 Step V: BIAS and uncertainty analysis

Step V uses the executable of *VICIRS_matchup_query.F90* to collect the match-ups satisfying the search-parameters set in *query.ini* (Table 5.3.1) to perform the statistical analysis and plot the results. *query.ini* is used for setting the parameters and for collecting data for both Steps V and VI.

Table 5.3.1 *query.ini* parameters

Parameter	Value
1. radiometer	MWI=1, ICI=2
2. radiosonde archive=1	RS=1(GRUAN)
3. temporal range start	*aaaa-mm-dd hh:mm*
stop	*aaaa-mm-dd hh:mm*
4. temporal distance (TD) **(default: 3)**	**1** (-15m/+45min); **2** (-1h/+1h); **3** (-3h/+3h)
5. spatial range (default: 90/-180/-90/180)	N/W/S/E
6. dedicated launches	**DL=0** (operative launches), (**DL=1**) dedicated launches only
7. NWP_opt	**NWP_opt=0 (**not using NWP); **NWP_opt=1**
8. Tskin_opt (only if NWP=1)	**Tskin_opt=0** T_{skin} determined from RS **Tskin_opt=1** $T_{skin} = T_{2m}(RS) + (T_{skin}(NWP) - T_{2m}(NWP))$; **Tskin_opt=2** $T_{skin} = T_{skin}(NWP)$
9. TA type	**TA type: 1, 2, 3, 4,5**
10. TA cloudy percentage maximum value	maximum value of the percentage of cloudy FOVs included in TA
11. Land fraction minimum/maximum value	minimum/maximum value of TA land fraction
12. Maximum LF=RLF	Maximum value of LF for selecting TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc files to be queried
13. Output type OT	OT=1 for residuals analysis and for Bias and uncertainty analysis OT=2 for MCM analysis

Depending on the set *Output type (OT)* set by the user in query.ini, *VICIRS_matchup_query* outputs (the file name is a combination of the *query.ini* parameters set by the user):
(H_)(MCM_)SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorth-LonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt.nc
The prefix *H_* is added to the StepV-output file name when the user sets the option Homogeneous TA=1 in *query.ini.* In this case, only homogeneous match-ups are selected for the statistical analysis. The prefix MCM is added when the user sets OT=2 in *query.ini* for performing MCM analysis, in this case only the *TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc* files corresponding to the MWI/ICI - GRUAN RS match-ups with DL=1 and/or NWP_opt=1 are considered.

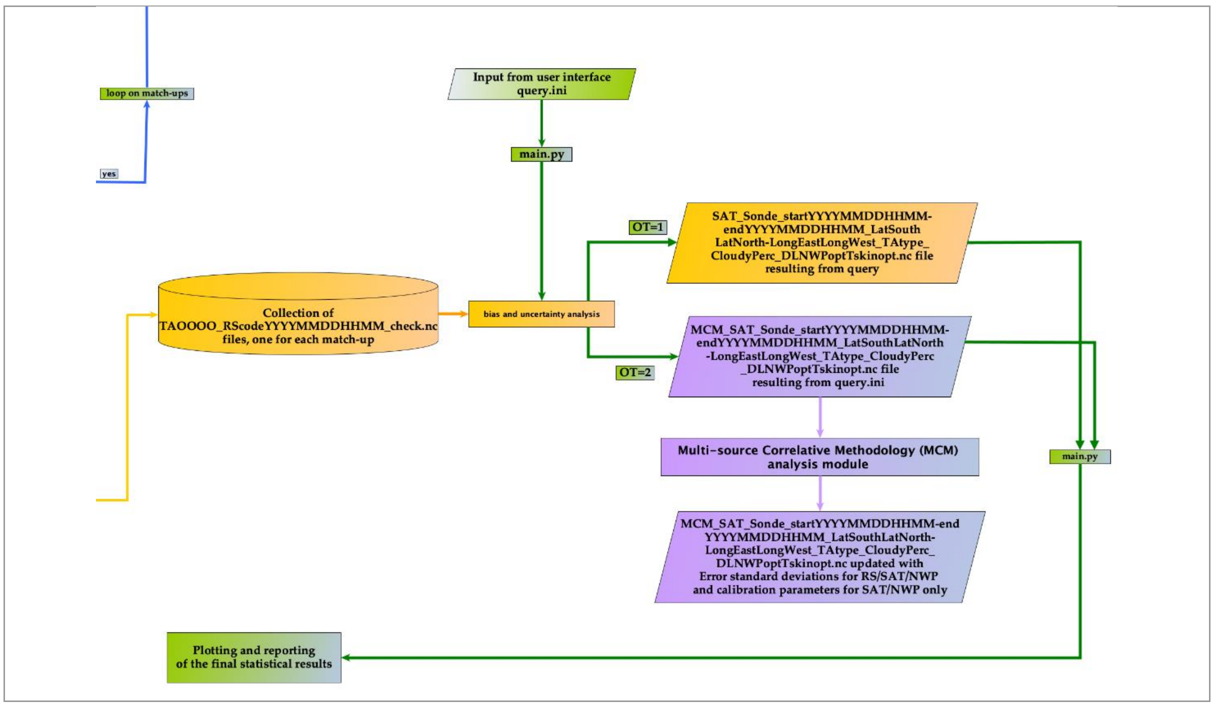


Figure 5.3.1 VICIRS tool-Block 3 scheme.

The *VICIRS_matchup_query* outputs variables are (further details in Table A.3 of Appendix A):

- **RSlatitude**, **RSlongitude**: RS site latitude and longitude for each match-up;
- **Pmin**: minimum value of RS pressure for each match-up;
- **RS_Lev**: number of RS level for each match-up;
- **NSAMPLE:** number of samples useful for statistics for each frequency;
- **BT_TA**, **u_col_SAT, u_obs_SAT**: the observed BT and the related uncertainties for all the frequencies, for all the match-ups and for the selected TA type;

- **BT_RS**, **u_sim_SAT**: the BT simulated from RS and the related uncertainties for all the frequencies and for all the match-ups;
- **BT_NWP:** the BT simulated from NWP (when OT=2) for all the frequencies and for all the match-ups;
- **TA_RS**: (BT_TA -BT_RS) for all the frequencies and for all the match-ups (**TA_RS_SAT** in TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc);
- **u_all:** uncertainty related to **TA_RS** for each channel and for each match-up, corresponding to **u_all_SAT** in TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc;
- **K_FACTOR:** coverage factor related to **TA_RS** and **u_all** for each match-up and for each channel;
- **BIAS_TA_RS:** the mean value of **TA_RS** for all the SAT channels**;**
- **SD_TA_RS**: Standard Deviation of the **TA_RS** corresponding to the search criteria;
- **u_BIAS**: uncertainty in the bias defined for each channel *j* (from *Managing error and uncertainties Lab. Manual*):

$$u\_BIAS(j) = SD\_TA\_RS(j)/\sqrt{nsample}$$

  where *nsample* is the number of match-ups used for statistics
- **wBIAS:** normalized BIAS of **TA_RS** (Moradi et al., 2010) accounting for the overall uncertainty **u_all**. **wBIAS(j)** is weighted by $w_{i,j} = \frac{1}{u\_all(i,j)^2}$

$$wBIAS(j) = \frac{\sum_{i=1}^{nsample} w_{i,j} \cdot TA\_RS(i,j)}{\sum_{i=1}^{nsample} w_{i,j}}$$

- **u_wBIAS:** uncertainty in w**BIAS**:

$$u\_wBIAS(j) = \sqrt{\frac{1}{\sum_{i=1}^{nsample} w_{i,j}}}$$

- **SDw_TA_RS**: Standard Deviation of the **TA_RS** weighted on the inverse of squared uncertainties $w_{i,j}$:

$$SDw\_TA\_RS(j) = \sqrt{\frac{\sum_{i=1}^{nsample} w_{i,j} \cdot (wBIAS(j) - TA\_RS(i,j))^2}{\sum_{i=1}^{nsample} w_{i,j} - (\sum_{i=1}^{nsample} w_{i,j}^2 / \sum_{i=1}^{nsample} w_{i,j})}}$$

The following outputs are generated:

- SKEW_KURTOSIS_**file_name**.png: shows Skew/Kurtosis for each SAT channel;
- SPATIAL_RS_PMIN_NLEV_**file_name**.png: shows the spatial distribution of match-ups used for statistics and the histogram of RS Pmin and RS number of levels;

- UNC_**file_name**.png: shows the TA_RS, u_all indicating also K-FACTOR for each match-up and for each SAT channel;
- BIAS_SD_**file_name**.png: BIAS/SD_BIAS plot for each SAT-channel;
- WBIAS_SD_**file_name**.png: BIASn/u_BIASn plot for each SAT-channel;
- scatter plot of BT observed versus BT simulated, colored differently to distinguish GRUAN sites (plot available only for GRUAN-SAT match-ups);
- RS statistics:
  - for all latitudes;
  - for polar latitudes;
  - for mid-latitudes;
  - for subtropical latitudes;
  - for tropical latitudes.

(**file_name**=(H_)(MCM)*SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorth LonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt* and *SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_TemporaleDistance.nc*).

5.3.2 Step VI: MCM analysis

The Multi-source Correlative Methodology (MCM) allows for the characterization of the error structure of three collocated (in space and time) measuring systems (subsection 4.7). It is implemented in Block 3, Step VI where the executable of *VICIRS_MCM_analysis.F90* takes as input:

1. BT_RS, BT_NWP and BT_TA from MCM_*SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorth LonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt*.nc
2. error covariance elements for RS system and NWP system (**e12**), the scaling (**a1**) and bias (**b1**) calibration parameters for RS reference system with their related error standard deviations (**s_a1**) and (**s_b1**).

The parameters in 2. can be initialized by the user in *MCMconfig.ini* (Figure 5.3.2).

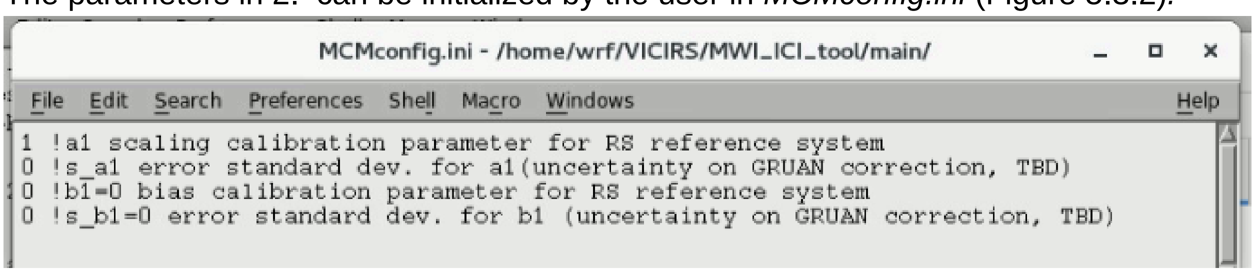


Figure 5.3.2 *MCMconfig.ini example.*

VICIRS_MCM_analysis outputs the error variances for each channel and for each measuring system and calibration parameters for SAT/NWP only. The fields added by MCM analysis in MCM_*SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorth LonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt*.nc are listed in Appendix A, Table A.4.
Note that MCM can provide unrealistic results when a limited number of input BT triplets is considered. Tests performed indicate that 100 or more triplets are necessary to have reasonable values of the output quantities.

6. Verification and validation of VICIRS tool

This section describes the activities focusing on testing the VICIRS tool. The testing has been performed by using two datasets:

- MWI and ICI L1B simulated dataset, provided by EUMETSAT at kick-off and described in [AD-4] and [AD-5], and spatially and temporally collocated RSs from RHARM archive and NWP profiles;
- Global Precipitation Measurement (GPM) Microwave Imager (GMI) observations, and spatially and temporally collocated RSs from GRUAN archive and from RHARM archive and NWP profiles.

Subsection 6.2 describes the VICIRS-tool configuration and the statistical results obtained by analyzing the match-ups between MWI/ICI simulated dataset and RHARM radiosondes. Subsection 6.3 describes (i) the modules added to the VICIRS tool for handling GMI observations, and (ii) the GMI-GRUAN-NWP and GMI-RHARM-NWP datasets and corresponding statistics results.

6.1 Test with MWI/ICI L1B – RHARM

The VICIRS tool has been tested on the MWI/ICI Level-1B simulated dataset, covering the three Metop-A reference orbits available (listed in Table 6.1.1), and the spatially/temporally collocated RHARM RSs.

Table 6.1.1 Reference Metop-A orbits provided by EUMETSAT with simulated MWI/ICI Level-1 B test data.

Orbit number	Time interval (UTC, Format:YYYY-MM-DD HH:MM:SS)	Comment
4655	2007-09-12 08:43:03 to 2007-09-12 10:22:03	First summer orbit
4656	2007-09-12 10:22:03 to 2007-09-12 12:04:03	Second summer orbit
6985	2008-02-23 08:46:03 to 2008-02-23 10:28:03	First winter orbit

6.1.1 Config.ini settings

The match-ups have been collected by running the VICIRS tool with the parameters set in *config.ini*. Figure 6.1.1 shows the settings used for the 4655 and 4656 Metop-A orbits.

config.ini - /home/vicirs/vicirs-tool/ (su Poirot)

File Edit Search Preferences Shell Macro Windows

```
[USER_DEF]
#radiometer MWI(ICI) or GMI
radiometer = MWI

# Temporal range start/stop
temporal_range_strt = 2008-02-23 00:00
temporal_range_stop = 2008-02-23 23:30

# Spatial range window (N/W/S/E)
spatial_range = 90.00/-180.00/-90.00/180.00

# radiosonde GRUAN or RHARM
radiosonde = RHARM

# Temporal colocation criteria
# 1 for -15m/+45m
# 2 for 2 for -1h/+1h
# 3 for 3 for -3h/+3h
temporal_distance = 3

# NWP
# 0 not using NWP
# 1 using NWP only if NWP is in clear sky
# 2 pass to the next match-up when NWP is cloudy
nwp = 0

# Dedicated launches
# 0 all launches
# 1 dedicated launches only if NWP is in clear sky
# 2 pass to the next match-up when NWP is cloudy
dedicated_launches = 0

# Tskin_opt
# 0 Tskin determined from RS
# 1 Tskin from RS and NWP
# 2 Tskin(NWP)
Tskin_opt = 0

# maximum LF (%) for selecting MWI/ICI FOVs to be included in TA
RLF = 100

[GLOBAL_VAR]
ICI_path=data_in/ICI/ICI_L1B_TDP_in_granules/6985
MWI_path=data_in/MWI/MWI_L1B_TDP_in_granules/6985
GMI_path=data_in/GMI/HDD
GRUAN_path=data_in/GRUAN
RHARM_path=data_in/RHARM
TA_path=data_in/TA_data
GP_path=./GRUAN_Processor_v6.3.b.0.1/bin/
data_out=data_out
```

Figure 6.1.1 *config.ini* used for searching and analyzing match-ups between simulated ICI/MWI data and RHARM RS (orbits 4655/4656).

For the Metop-A orbit 6985, *config.ini* was modified by simply replacing:

1) temporal range *temporal_range_start* = ***2007-09-12 08:00***
   temporal_range_stop = ***2007-09-12 12:30***
2) ICI_path=data_in/ICI/ICI_L1B_TDP_in_granules/**4655_4656**
3) MWI_path=data_in/MWI/MWI_L1B_TDP_in_granules/**4655_4656**

with:

1) temporal range *temporal_range_start* = ***2008-02-23 08:00***

temporal_range_stop = ***2008-02-23 11:00***

2) ICI_path=data_in/ICI/ICI_L1B_TDP_in_granules/**6985**
3) MWI_path=data_in/MWI/MWI_L1B_TDP_in_granules/**6985.**

The land fraction (LF) within a target area (TA) is tunable through the RLF option in *config.ini*. To test this capability, three types of Step-III output have been generated (*TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc*) for each match-up by setting RLF=0, RLF=30 and RLF=100. In the first case, only the MWI/ICI FOVs with LF[4]=0% are included in TA, while in the second and third cases only FOVs with LF<=30% and LF<=100% are considered, respectively. The examples of TA (for MWI 89 GHz) including only FOVs with LF=0 , LF≤30% , LF≤100%, are shown in Figure 6.1.2 where a match-up between MWI/ICI (orbit n. 4656) and RHARM (lat/lon=[18.1°N 15.9°W]) is analyzed for the three RLF obtaining three different outputs:

- TA4656_MRM000061442-200709121200_lf0_check.nc
- TA4656_MRM000061442-200709121200_lf30_check.nc
- TA4656_MRM000061442-200709121200_lf100_check.nc

which are characterized by a different number of FOVs for each TA and for each MWI/ICI channel.

In general, varying the maximum LF allowed within each FOVs (RLF parameter in *config.ini*) while keeping the other parameters in *config.ini* unchanged is useful because it allows to investigate the trade-off between number of land-contaminated FOVs and total number of FOVs for TAs centered near coastlines. For example, referring to Figure 6.1.2, the user can:

1. include FOVs over land and over sea for all the frequencies (LF<=100%). BT_TA is the average of all the FOVs included in circular TA. BT simulated from RS (BT_RS) is the linear combination of BT simulated over sea (BT_RS_s) and over land (BT_RS_l):
   $$BT\_RS = LAND\_FRAC\_TA \cdot BT\_RS\_l + (1 - LAND\_FRAC\_TA) \cdot BT\_RS\_s)$$
   where $LAND\_FRAC\_TA$ is LF determined for each MWI/ICI frequency and for each TA. The right panel in Figure 6.1.2 shows MWI-89GHz TA with LF≤100%. The resulting BT are BT_TA=281.76K, BT_RS=281.03K, with high variability (SD_TA=10.46K) due to the contrast between land and sea surface emissivity;
2. exclude some land-contaminated FOVs (LF<=30%) to reduce land influence and deal with more homogeneous TA, as shown in the middle panel in Figure 6.1.2 for which BT_TA=267.09K, SD_TA=0.86K, BT_RS=272.04K.

[4] Note that an apparent error was found in the test data provided by EUMETSAT. For the MWI data groups 1,2,5,6,7,8 (see [AD-11] for further details) the elevation and land fraction at coarser resolution are erroneously associated with the high frequency channels, while the high resolution ones are associated with the low frequency channels. The land fraction and elevation terrains for the middle frequency channels (data groups 3 and 4) are unaffected.

3. exclude all land-contaminated FOVs for all the frequencies (LF=0%) so to completely avoid land influence, as in the left panel in Figure 6.1.2. Here, only FOVs with LF=0 are considered for determining BT_TA (267.15K) and SD_TA (0.01K). In such a case, BT_RS=270.33K results from BT simulated over sea only.

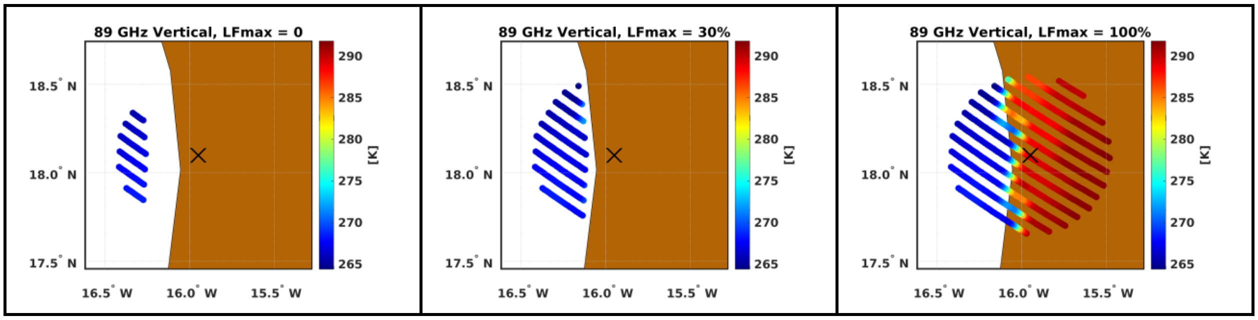


Figure 6.1.2 Example of TA for MWI 89 GHz vertical polarization built around NOUAKCHOTT RHARM-site including only FOVS with LF=0 (left panel), LF≤30% (middle), LF≤100% (right).

The statistics related to three types of TA (RLF=0, RLF=30 and RLF=100) are given in Table 6.1.6, to give an idea of the LF impact on statistics.
When a RS site is near the coastline (for example GRUAN sites such as TEN, GRA, MTS) and RLF=100 is set in *config.ini*, the resulting TAs are characterized by LF higher than 0 for all the channels (e.g., see LAND_FRAC_TA field in *TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc*). This is because LAND_FRAC_TA is the average of all the LF included in the TA, and LF>0 for the FOVs closest to the launch site for all the MWI/ICI frequencies. Conversely, RLF=0 reduces the number of FOVs in TA to only those completely over sea. This aspect is very important for collecting match-ups over sea with no land contamination in performing the bias and uncertainty analysis.

6.1.2 Analysis of radiosoundings used for MWI/ICI(L1B)-RHARM match-up analysis

The RS analysis in Step-II is mandatory to decide if the related match-up is useful for calibration (subsection 4.3). In summary, the match-up is considered for calibration if:

1. number of pressure levels (*nlev*) ≥ 40;
2. pressure minimum value, *Pmin* ≤ 10 hPa;
3. RS in clear sky;
4. Air Mass Displacement (AMD) ≤ Target-Area (TA) radius.

In particular, points 1 and 2 have been derived from the analysis of BT simulated from a set of RHARM and GRUAN clear sky RSs, measured from the same site and at the same time. 111 clear-sky RSs profiles were extracted from a larger number of RSs (1446) filtered with the method described in subsection 4.3. The remaining RHARM RSs have nlev ranging from 15 to 51 and Pmin ranging from 10 to 250 hPa. The set of 111 clear-sky cases was divided in 13 sub-datasets, according to binned Pmin and nlev values. The best match between RHARM and GRUAN were obtained from the sub-dataset corresponding to the Pmin and nlev values in point 1 and 2 above. Further details are in Section 5 of [AD-8].

Among the 92 initial MWI/ICI-RHARM match-ups for 6985, 4655 and 4656 orbits, only one satisfies all the Step-II tests. To test the VICIRS tool on a larger number of ICI/MWI-RHARM match-ups, the test 1. and 2. have been relaxed as follows:

1. *nlev*≥15;
2. no filter applied to Pmin;

Applying these filters, the number of match-ups lowered from 92 to 29. In detail, 9.8% of the initial RS is removed because nlev<15 and 59.8% is removed because the clear-sky test failed.

Table 6.1.2 Statistical scores related to RHARM-MWI/ICI match-ups

	Polar latitude	**Mid-latitude**	**Subtropical latitude**	**Tropical latitude**	**All latitude**
RS total (#)	27	43	9	13	92
RS discarded (#)	20	29	5	9	63
RS useful (#)	7	14	4	4	29
QC fails (%)	7.4	9.3	11.1	15.4	9.8
Cloudy fails (%)	66.7	60.5	44.4	53.9	59.8
AMD fails (%)	0.0	0.0	0.0	0.0	0.0
total fails (%)	74.1	67.4	55.6	69.2	68.5

Table 6.1.2 shows the statistics of the analyzed RS grouped according to their latitude:

- Polar latitude (from 60° to 90° South and from 60° to 90° North);
- Mid-latitude (from 37° to 60° North and from 35° to 60° South);
- Subtropical latitude (from 23°26’ to 37° North and from 23°26’ to 35° South);
- Tropical latitude (from 23°26’ South to 23°26’ North).

The percentage of failure in all the latitude bands mainly depends on cloud cover, being higher than 44% in all latitude bands.

6.1.3 Query.ini settings and discussion of query output

Once the match-ups have been collected according to the parameters set in *config.ini*, it is possible to query them according to the radiometer (MWI or ICI).

```
query.ini - /home/vicirs/vicirs-tool/ (su Poirot)
File  Edit  Search  Preferences  Shell  Macro  Windows
#radiometer MWI or ICI
MWI

# Temporal range start/stop #2007-09-12 2008-02-23
"2007-09-12 0:00"
"2008-02-23 23:30"

#spatial_range = "lat North/long West/lat South/long East"
"90.00/-180.00/-90.00/180.00"

#radiosonde GRUAN or RHARM
RHARM

# Temporal collocation criteria
# 1 for -15m/+45m
# 2 for 2 for -1h/+1h
# 3 for 3 for -3h/+3h
3

# NWP
# 0 not NWP
# 1 with NWP
0

# Dedicated launches
# 0 all launches
# 1 dedicated launches only
0

# Tskin_opt
# 0 Tskin determined from RS
# 1 Tskin from RS and NWP
# 2 Tskin(NWP)
0

# Target area type
# 1 circular with BT averaged without weighting on distance from RS launch site
# 2 circular with BT averaged by weighting on distance from RS launch site
# 3 circular with BT averaged by weighting on squared-distance from RS launch site
# 4 RS-driven TA
# 5 3x3 RS-driven TA
1

# Target area cloudy percentage maximum value (%)
0

# Target area land fraction "minimum/maximum value" (%) example "0.00/100.00"
"0/100"

#maximum LF (%) (RLF) for selecting TAOOOOO_RSCODE-aaaammddhhmm_lfRLF_check.nc to be queried
"100"

# Output type Ot=1 (BT_diff, Bias and uncertainty analysis); Ot=2 (MCM analysis)
1

# Homogeneous TA (1), not Homogeneous TA (0)
0

#matchup_file directory for reading TA
"../data_out/StepIII"
#Step-V output directory for reading TA
"../data_out/StepV"
```

Figure 6.1.3 *query.ini* for querying the StepIII-output related to MWI/RHARM match-ups and for reporting and plotting of the statistical results (in Block 3, Step V)

In fact, MWI and ICI are processed simultaneously in the collection of match-ups, whereas they are queried and analyzed separately. The information required for querying the output files (TAOOOOO_RSCODE-aaaammddhhmm_lfRLF_check.nc), are set in *query.ini* (Figure 6.1.3) for selecting match-ups, computing the statistics and plotting the results (subsection 5.2.5). *query.ini* parameters are similar to the ones initialized in *config.ini*, but they can be varied in order to analyze the collected dataset with different combinations of

While the data collection set with *config.ini* is relatively slow, the data analysis set with *query.ini* is relatively fast. In particular, the user can choose:

1. radiometer;
2. temporal range;
3. spatial range;
4. radiosonde archive;
5. temporal collocation criteria (maximum lag between satellite overpass and radiosonde launch time);
6. use of NWP information for filling the RS gaps (i.e., surface parameters, profiles above the maximum RS altitude, and NWP Ozone profile instead of fixed RTTOV climatological Ozone profile);
7. Tskin_opt that indicates how to determine the skin temperature (T_{skin}) for simulating BT using surface information from NWP (further details in subsection 5.2.3):
    1. $T_{skin} = T_{2m} + (T_{skin}(NWP) - T_{2m}(NWP))$ where T_{2m} is the 2-meter temperature determined from RS and $T_{skin}(NWP), T_{2m}(NWP)$ are T_{2m} and T_{skin} of the model;
    2. $T_{skin} = T_{skin}(NWP)$.
8. TA type;
9. TA cloudy percentage corresponding to the percentage of the cloudy FOVs included in TA, according to the results of all the cloudy tests applied in Step III;
10. minimum and maximum land fraction (%) (**LFmin/LFmax**) for selecting observations with LF in [LFmin : LFmax] for each frequency in the selected TA;
11. maximum LF (%) (**RLF)** for selecting output files to be queried (e.g. if **RLF=0** in query.ini, only TAOOOOO_RSCODE-aaaammddhhmm_**lf0**_check.nc files will be queried. Note that these files exist if already created in Block-2 by setting RLF=0 in *config.ini*.

Note that the above points 10. and 11. may seem identical and could create confusion. For the sake of clarity, they differ in:

- **10-query.ini** (**LFmin/LFmax**) **used for defining the land type of the match-ups to be investigated**, for example: **0/0** selects match-ups with TA completely over sea; **0/30** selects match-ups over coast with TA-LF<=30%; **100/100** selects match-ups completely over land with TA-LF=100%.
- **11-query.ini (RLF) used for selecting the output files to be queried** , for example: when **RLF=0**, the tool queries TAOOOOO_RSCODE-aaaammddhhmm_lf**0**_check.nc files; **RLF=100**, the tool queries TAOOOOO_RSCODE-aaaammddhhmm_lf**100**_check.nc files. Here **RLF** indicates that in TAOOOOO_RSCODE-aaaammddhhmm_lf**RLF**_check.nc the BT related to TA has been determined considering only the FOVs with LF<=RLF.

It is important to bear in mind that LFmax/LFmin filter (10 query.ini) is applied in *query.ini* to the mean LF within the TA, whereas RLF filter (11) in *config.ini* is applied to LF within each FOV in building TA, while in *query.ini* is applied for selecting only the TA including FOVs with LF<=RLF.

A number of cases have been investigated as described in Table 6.1.3, presenting the results obtained for MWI and ICI. Note that the number of match-ups depends on the selected TA cloudy percentage (e.g., 0 or 30%), because it downselects the available RS. Note that the following caveats affect the presented results:

- no filter is applied to *Pmin* and the minimum value of *nlev* has been lowered to 15. This is to test the VICIRS tool and the plotting output on a larger number of ICI/MWI-RS match-ups;
- LF of some simulated MWI L1B observations is not correct, as mentioned earlier[4].

Table 6.1.3 List of *query.ini* parameters set for the statistical analysis of the RHARM-ICI/MWI test dataset. Temporal range is in UTC (format: YYYY-MM-DD HH:MM:SS). Case number (CN) and number of available match-ups (#MU) are also reported. The number of *query.ini* parameters is indicated within parentheses following the list described above.

Temporal range (UTC)(2)	Temporal collocation criteria(5)/ TA type(8)	TA cloudy %(9)	RLF (11)	[LF min: LF max] (10)	NWP_opt(6)/ Tskin_opt(7)	CN(#MU)	Figure	Tables
2007-09-12 08:00 to 2008-02-23 11:00	3/1	0	100	[0:100]	0/0	**1.a (10)**	**6.1.5, 6.1.6 (MWI) 6.1.7, 6.1.8 (ICI)**	**6.1.4(MWI) 6.1.5(ICI)**
	3/1	30	100	[0:100]	0/0	**1.b(19)**	**6.1.5 (MWI) 6.1.7 (ICI)**	
	3/1	100	100	[0:100]	0/0	**1.c(29)**		
	3/1	0	30	[0:30]	0/0	**2(8 for MWI and 10 for ICI)**		**6.1.6(MWI) 6.1.7(ICI)**
	3/1	0	0	[0:0]	0/0	**3(14 for MWI and 18 for ICI)**	**6.1.9**	
	3/1	0	100	[0:100]	1/1	**4(5) 4.a(5 match-ups without NWP)**	**6.1.10, 6.1.11,6.1.12 (MWI)**	**6.1.8(MWI) 6.1.9(ICI)**
	3/1	0	100	[0:100]	1/2	**5(5)**	**6.1.10, 6.1.13,6.1.14 (ICI)**	

Among the 29 ICI/MWI-RHARM match-ups with TA characterized by RLF=100 (stored in output file TAOOOOO_RSCODE-aaaammddhhmm_lf**100**_check.nc where TAs have been built considering FOVs with LF<=100%), only 10 match-ups with TA in clear sky have been identified (**Case 1.a**). The geographical distribution of **Case 1.a** match-ups is shown in Figure 6.1.4 where histograms of *nlev* and *Pmin* are also shown: *Pmin* ranges from 10 to 300 hPa and the number of RS pressure-levels is higher than 40 for 2 match-ups.

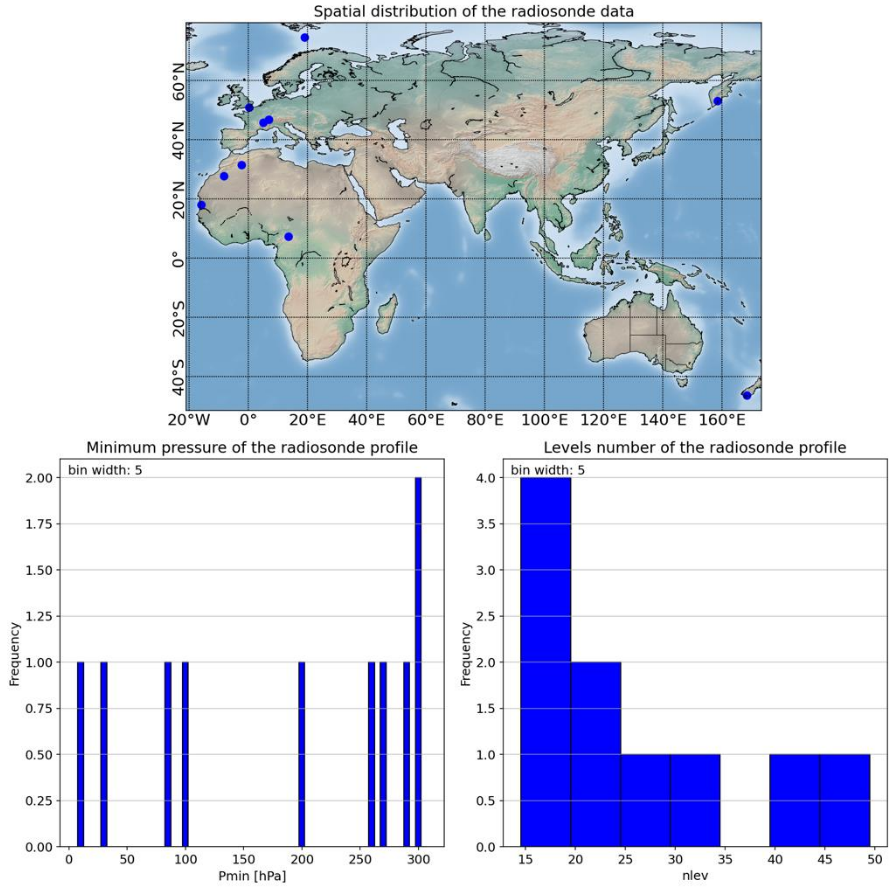


Figure 6.1.4: **Case 1.a** (TA with **LF<=100**, corresponding to TAOOOOO_RSCODE-aaaammddhhmm_**lf100**_check.nc). Top: spatial distribution of the 10 MWI-ICI/RHARM match-ups (cloud screening applied to both RS and TA). Bottom: histograms of RS minimum pressure (*Pmin,* left) and number of levels (nlev, right).

The statistical results are presented in term of:

- *BIA*S: mean value of the residuals TA_RS=(BT_TA-BT_RS)
- SD: standard deviation of TA_RS
- *u_BIAS*: BIAS uncertainty (standard error of the mean)
- the same quantities weighted by the inverse of the squared overall uncertainties.

In detail, in Figures 6.1.5, 6.1.7, 6.1.11, 6.1.13:

- the top panels show $BIAS(j) \pm SD\_TA\_RS(j)$, $BIAS(j) \pm u\_BIAS(j)$, where:

$$BIAS(j) = \frac{\sum_{i=1}^{nsample} TA\_RS(i,j)}{nsample} \text{ and}$$

$$SD\_TA\_RS(j) = \sqrt{\frac{\sum_{i=1}^{nsample}(BIAS(j) - TA\_RS(i,j))^2}{nsample - 1}},$$

and

$$u\_BIAS = SD\_TA\_RS/\sqrt{nsample}$$

nsample is the number of match-ups, *j* indicates the MWI(ICI) frequency, *i* indicates the match-up number and $TA\_RS(i,j) = BT\_TA(i,j) - BT\_RS(i,j)$ is the difference between observed and simulated BT. These data correspond to BIAS_TA_RS, SD_TA_RS and u_BIAS fields in Step-V output (*SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorthLonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage_LF_DL/NWPopt/Tskinopt.nc*;

- the bottom panels show the plot of $wBIAS(j) \pm SD\_wTA\_RS(j)$, $wBIAS(j) \pm u\_wBIAS(j)$ where $wBIAS, u\_wBIAS,$ and SD_wTA_RS are respectively the weighted bias, its uncertainty, and the standard deviation of the weighted residuals, where the weights are defined as the inverse of the squared overall uncertainty (Buehler et al. 2004; Moradi et al., 2010) . In detail:

$$wBIAS(j) = \frac{\sum_{i=1}^{nsample} w_{i,j} \cdot TA\_RS(i,j)}{\sum_{i=1}^{nsample} w_{i,j}}$$

$$u\_wBIAS(j) = \sqrt{\frac{1}{\sum_{i=1}^{nsample} w_{i,j}}}$$

$$SDw\_TA\_RS(j) = \sqrt{\frac{\sum_{i=1}^{nsample} w_{i,j} \cdot (wBIAS(j) - TA\_RS(i,j))^2}{\sum_{i=1}^{nsample} w_{i,j} - (\sum_{i=1}^{nsample} w_{i,j}^2 / \sum_{i=1}^{nsample} w_{i,j})}}$$

where $w_{i,j} = 1/(u\_all(i,j))^2$ and $u\_all(i,j) = \sqrt{u\_col^2 + u\_obs^2 + u\_sim^2}$ (further details in subsections 5.2.4 and 5.2.5).

The corresponding statistics are compiled in Tables 6.1.4, 6.1.5, 6.1.6, 6.1.7, 6.1.8, 6.1.9. Figures 6.1.6, 6.1.8, 6.1.12 and 6.1.14 show the plot of the residuals TA_RS with the associated overall uncertainty, u_all as error bar for each MWI/ICI frequency and match-up. The points are colored differently depending on the resulting coverage factor k, which determines an interval about the mean value as a multiple of standard uncertainty assuming that the uncertainty is normally distributed (subsection 4.6). Each residual TA_RS is checked against the overall uncertainty as:

$$TA\_RS(i,j) \le k(i,j) \cdot u\_all(i,j)$$

where *i* and *j* indicate the match-up and frequency, respectively. The coverage factor k is color-coded as follows:

- green when the results agree within **k=1** (data are *consistent*);
- orange when the results agree within **k=2** (data are *in statistical agreement*);
- red when the results agree within **k=3** (data are *significantly different*);.
- black when the results do not agree within **k>3**.

If the last two are dominant, they likely indicate that either a bias is present, or the overall uncertainty has been underestimated. The overall uncertainty data and coverage factor are stored as *u_all* and *K_FACTOR*, respectively, in output files such as *TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc* and *SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorthLonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage_LF_DL/NWPopt/Tskinopt.nc.*

Figure 6.1.5 shows the statistical results for the 10 match-ups of Case 1.a (Table 6.1.3) on the left and for the 19 match-ups of Case 1.b on the right. Case 1.b/1.c match-ups consist of Case 1.a match-ups plus match-ups with TA cloudy for a maximum of 30%/100% of the FOVs included in it. In both cases, the highest value for BIAS (and wBIAS) is related to 18.7 GHz (horizontal polarization, H): BIAS≃10.0 K and SD=13.1 K. The high BIAS derives from the high residuals at 18.7 GHz (H) for about 4 match-ups as shown in Figure 6.1.6 where the residuals are plotted for each frequency and for each match-up for Case 1.a. The behavior is similar for Case 1.b and 1.c (not shown in order to avoid redundant information), where BIAS and SD are higher in relatively transparent channels while lower in most opaque channels, except for 89 GHz (V) that shows the lowest BIAS (≃0.20K). The opaque channels at 183 GHz show a BIAS ranging from 2.1 to 3.4 K with a SD ranging from 5.6 to 9.8 K, due to the presence of high residuals for a match-up located near coastlines (RSM00032540; lat=53.08, lon=158.58; PETROPAVLOVSK-KAMCHATSKIJ). In Case 1.b (1.c), the effect of this residual in the 183-GHz region is reduced by the higher number of match-ups, as testifies also by the uncertainty in BIAS (u_BIAS) that is about 1 K lower than the correspondent in Case 1.a for all the 183.31 GHz frequencies. Table 6.1.4 shows the statistics related to Case.1.a (left) and Case 1.b (middle) and Case 1.c (right).
The ICI statistical results for Case 1.a and 1.b are shown in Figures 6.1.7, 6.1.8 and Table 6.1.5 where also the statistics for Case 1.c is shown. The number of match-ups examined is the same for both MWI and ICI, due to the fact that the two instruments are processed simultaneously. The results obtained for the MWI and ICI at similar channels (183.31 GHz frequency band) are colored similarly. For ICI channels at 183.31±7.0, 183.31±3.4 and 183.31±2.0 GHz, BIAS values are ~0.6 K higher and SD are ~0.3 K lower than for the same MWI channels. BIAS ranges from 2.9 to 3.9 K for ICI and from 2.1 to 3.4 K for MWI; SD from 5.2 to 9.6 K for ICI and from 5.6 to 9.8 K for MWI. This is due to the differences between BT observed by MWI and by ICI. In fact, for the same match-up, ICI TA is wider than the MWI TA, as the two instruments have different spatial resolutions (10 km for MWI and 16 km for ICI). As for MWI, the statistical scores related to 183.31 GHz band decrease when cloudy TA are also considered (Case 1.b and 1.c in middle and right panel, respectively) but they are worse for the higher frequencies due to the impact of clouds. Comparing residuals in Figures 6.1.6 and 6.1.8, it is evident that also for ICI, the

statistical results in 183.31 GHz band are strongly influenced by the high residual in RSM00032540 RHARM site. In clear-sky case, the lowest ICI BIAS (-0.38K) is related to 243.2±2.5GHz(V) and the worst to 183.31±3.4GHz (3.9K). In Figure 6.1.6 the residuals at 52-55 and 118 GHz channels often show k>=3, meaning that the relation $TA\_RS(i,j) \leq k(i,j) \cdot u\_all(i,j)$ is not satisfied for lower k values. This may indicate that either the overall uncertainty is underestimated or there is an unaccounted systematic difference due, e.g., to temperature mismatch at the surface and/or in the lower stratosphere. As shown in Figure 6.1.12, the inclusion of NWP data (both at surface and lower stratosphere) reduces the systematic difference, thus pointing to the second hypothesis. This feature will be further investigated.

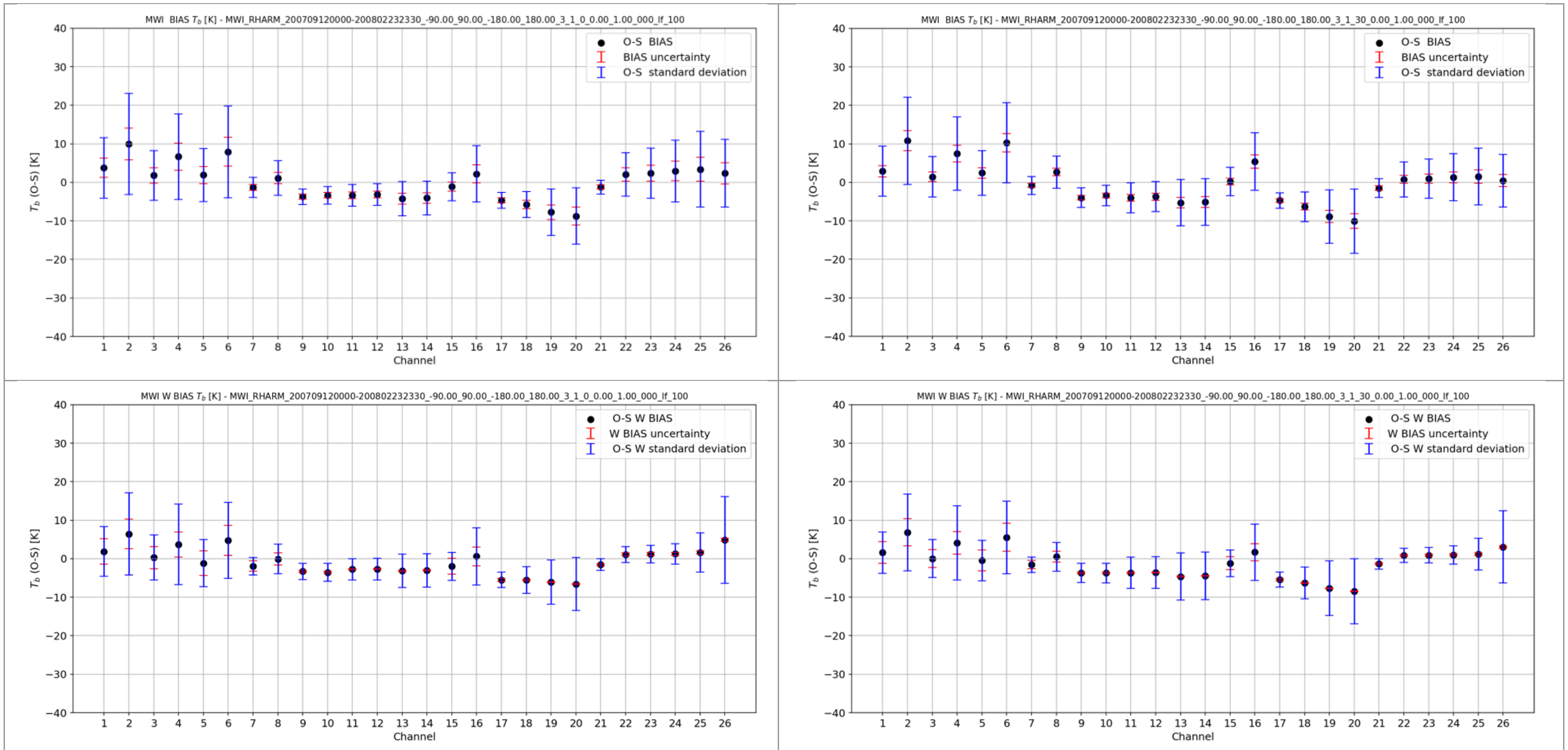


Figure 6.1.5 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom w_BIAS ± SD_wTA_RS ± u_wBIAS. **(Left) Case 1.a** MWI statistics from 10 match-ups with LF<=100% in clear sky. (**Right**) **Case 1.b** MWI statistics from 19 match-ups with LF<=100% and percentage of cloudy FOVs<=30%.

Figure 6.1.6. Case 1.a (as introduced in Table 6.1.3) for MWI: (BT(obs)-BT(sim)) $\pm$ u_all (TA-LF<=100%,10 match-ups related to TA and RS in clear sky).

Table 6.1.4 BIAS, SD_TA_RS and u_BIAS for Case 1.a and Case 1.b. The highlighted rows correspond to MWI channels similar to ICI channels.

	Case 1.a (10 match-ups in clear sky, with LF<=100%)			**Case 1.b (19 match-ups with LF<=100% and TA cloudy FOVs>=30%)**			**Case 1.c (29 match-ups with LF<=100% and TA cloudy FOVs>=100%)**		
MWI frequency (GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
18.7V)	3,74	7,89	2,50	2,88	6,49	1,49	2,34	8,46	1,57
18.7(H)	9,96	13,13	4,15	10,81	11,32	2,60	11,73	17,21	3,20
23.8(V)	1,79	6,46	2,04	1,44	5,29	1,21	0,61	7,08	1,31
23.8(H)	6,67	11,10	3,51	7,46	9,52	2,18	8,11	15,27	2,84
31.4(V)	1,90	6,91	2,18	2,43	5,77	1,32	1,53	7,70	1,43
31.4(H)	7,93	11,91	3,77	10,28	10,38	2,38	10,74	15,69	2,91
50.3(V)	-1,32	2,55	0,81	-0,81	2,33	0,54	-0,76	2,43	0,45
50.3(H)	1,11	4,53	1,43	2,67	4,19	0,96	2,82	5,53	1,03
52.610(V)	-3,72	2,02	0,64	-3,97	2,55	0,59	-3,43	2,67	0,50
52.610(H)	-3,35	2,25	0,71	-3,40	2,64	0,60	-2,90	2,68	0,50
53.24(V)	-3,37	2,79	0,88	-4,00	3,92	0,90	-3,38	3,94	0,73
53.24(H)	-3,13	2,83	0,89	-3,72	3,93	0,90	-3,12	3,92	0,73
53.750(V)	-4,24	4,41	1,39	-5,27	6,04	1,39	-4,41	5,98	1,11
53.750(H)	-4,07	4,41	1,40	-5,10	6,04	1,39	-4,24	5,97	1,11
89.9(V)	-1,11	3,64	1,15	0,18	3,68	0,84	-0,14	3,80	0,71
89.9(H)	2,20	7,29	2,31	5,40	7,50	1,72	5,40	10,91	2,03
118.7503±3.20	-4,66	2,03	0,64	-4,70	2,03	0,47	-4,23	2,55	0,47
118.7503±2.10	-5,79	3,37	1,06	-6,33	3,83	0,88	-5,48	4,28	0,79
118.7503±1.40	-7,75	6,00	1,90	-8,86	6,92	1,59	-7,49	7,34	1,36
118.7503±1.20	-8,74	7,31	2,31	-10,06	8,35	1,91	-8,45	8,80	1,64
165.5±0.75	-1,22	1,77	0,56	-1,48	2,44	0,56	-1,93	2,67	0,50
183.31±7.0	2,06	5,61	1,77	0,77	4,54	1,04	0,06	3,94	0,73
183.31±6.1	2,37	6,50	2,05	0,94	5,10	1,17	0,37	4,32	0,80
183.31±4.9	2,95	8,03	2,54	1,30	6,11	1,40	0,95	5,14	0,95
183.31±3.4	3,37	9,81	3,10	1,51	7,35	1,69	1,47	6,43	1,19
183.31±2.0	2,36	8,77	2,77	0,43	6,84	1,57	0,84	6,65	1,24

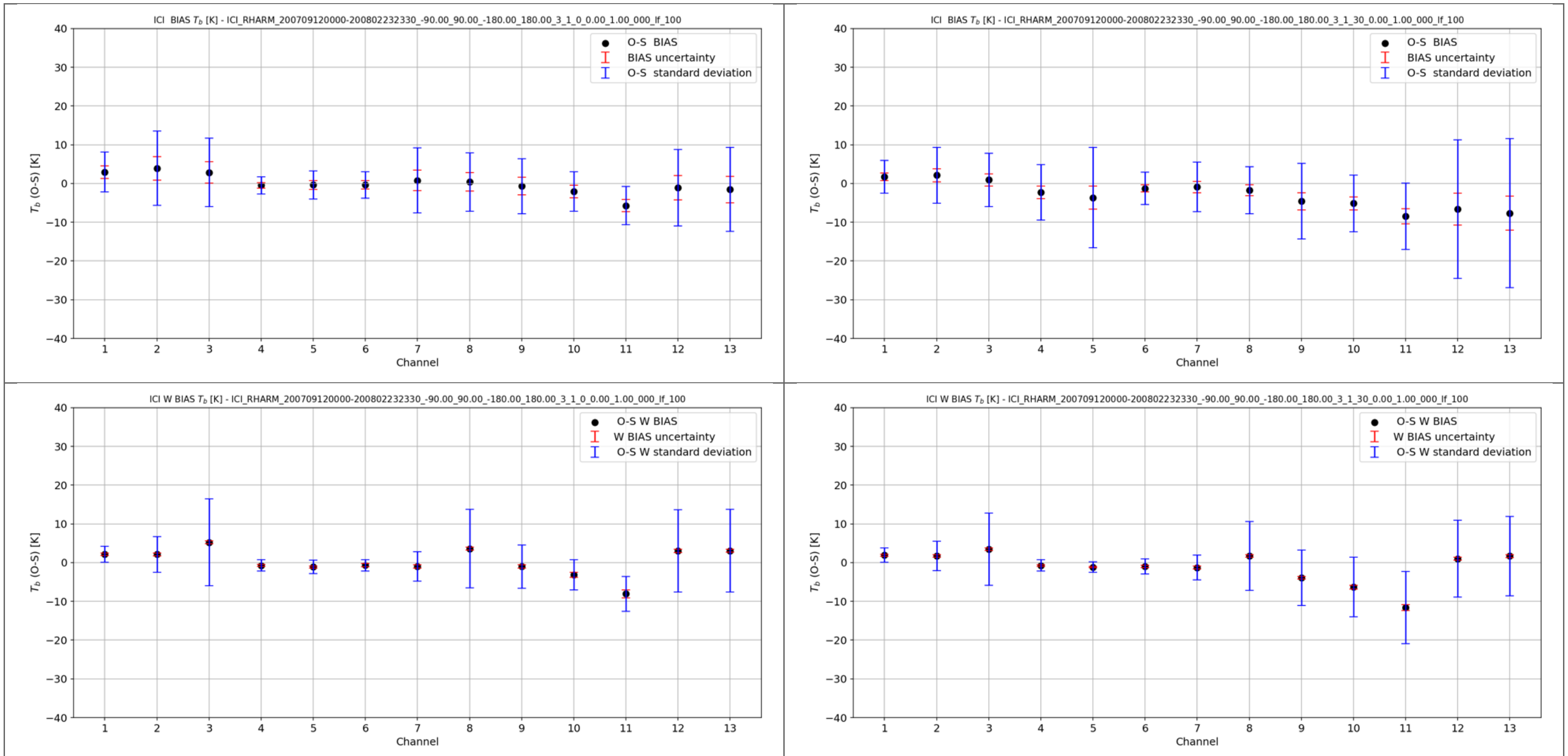


Figure 6.1.7 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom w_BIASn ± SD_wTA_RS ± u_wBIAS. **(Left) Case 1.a** ICI statistics from 10 match-ups with LF<=100% in clear sky. (**Right**) **Case 1.a** ICI statistics for 19 match-ups with LF <=100% and percentage of cloudy FOVs<=30%.

Figure 6.1.8. Case 1.a (as introduced in Table 6.1.3) for ICI: (BT(obs)-BT(sim)) ± u_all (LF<=100%,10 match-ups in clear sky).

Table 6.1.5 BIAS, SD_TA_RS and u_BIAS for Case 1.a and Case 1.b, The highlighted rows correspond to the ICI channels similar to MWI channels.

	Case 1.a (10 match-ups in clear sky, with LF<=100%)			Case 1.b (19 match-ups with LF<=100% and TA cloudy FOVs>=30%)			Case 1.b (29 match-ups with LF<=100% and TA cloudy FOVs>=100%)		
ICI frequency (GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
183.31±7.0	2,94	5,16	1,63	1,75	4,23	0,97	1,05	3,71	0,69
183.31±3.4	3,92	9,59	3,03	2,10	7,19	1,65	2,06	6,28	1,17
183.31±2.0	2,85	8,85	2,80	0,93	6,88	1,58	1,33	6,66	1,24
243.2±2.5(V)	-0,48	2,20	0,69	-2,29	7,20	1,65	-2,46	6,40	1,19
243.2±2.5(H)	-0,38	3,66	1,16	-3,64	12,92	2,96	-2,75	11,02	2,05
325.15±9.5	-0,37	3,42	1,08	-1,27	4,17	0,96	-1,83	4,14	0,77
325.15±3.5	0,79	8,36	2,64	-0,89	6,42	1,47	-0,95	5,55	1,03
325.15±1.5	0,38	7,51	2,38	-1,75	6,10	1,40	-1,34	6,00	1,11
448±7.2	-0,69	7,11	2,25	-4,59	9,76	2,24	-3,81	8,58	1,59
448±3.0	-2,05	5,07	1,60	-5,13	7,33	1,68	-4,68	6,96	1,29
448±1.4	-5,71	4,92	1,55	-8,49	8,56	1,96	-7,67	8,94	1,66
664±4.2(V)	-1,09	9,84	3,11	-6,60	17,85	4,10	-5,03	15,09	2,80
664±4.2 (H)	-1,55	10,84	3,43	-7,67	19,21	4,41	-5,95	16,36	3,04

In order to compare the statistics related to the different choices of *LFmax/LFmin* and *RLF* parameters (above-mentioned points 10 and 11 of *query.ini*) and to evaluate the impact of LF on the statistics, **Case-1.a** (**LF<=100%** and TAOOOOO_RSCODE-aaaammddhhmm_lf**100**_check.nc files queried), **Case-2** (**LF<=30%** and TAOOOOO_RSCODE-aaaammddhhmm_lf**30**_check.nc files queried) and **Case-3** (**LF=0** and TAOOOOO_RSCODE-aaaammddhhmm_lf**0**_check.nc files queried) statistical scores are summarized in Table 6.1.6 for MWI and Table 6.1.7 for ICI. Unfortunately, due to the low number of match-ups in clear sky, there are no common match-ups for the three cases, thus preventing from isolating the effect of LF and therefore of land surface emissivity.

The number of match-ups useful for statistics is higher for LF=0 than for LF=100 because the statistics consider only TA in clear-sky: TA with LF=100 includes a higher number of FOVs than LF=0 with higher probability of finding cloudy FOVs causing the related match-up to be removed.

The Case 1.a statistical results obtained for MWI are better than the other two cases except for the 183.31 GHz frequencies statistics in Case 3. In fact, in Case 3 BIAS ranges from 0.5 K for 183.31±4.9 GHz to 1.4 K for 183.31±3.4 GHz (from 2.1 K for 183.31±7.0 GHz to 3.4 K for 183.31±3.4 GHz in Case 1.a) and SD_TA_RS ranges from 2.7 K for 183.31±7.0 GHz to 5.9 K for 183.31±2.0 GHz (from 5.6 K for 183.31±7.0 GHz to 9.8 K for 183.31±3.4 GHz in Case 1.a). The improvement is due to the removal of the high residuals related to the RSM00032540 RHARM site. The same behavior can be noted in Case-3 ICI results but the difference between BIAS is not as high as in MWI case. In MWI cases 2 and 3 the BIAS/SD for less opaque channels are very high and this may depend on the wrong LF reported in MWI L1B data([4]) that impacts both observed and simulated BT.

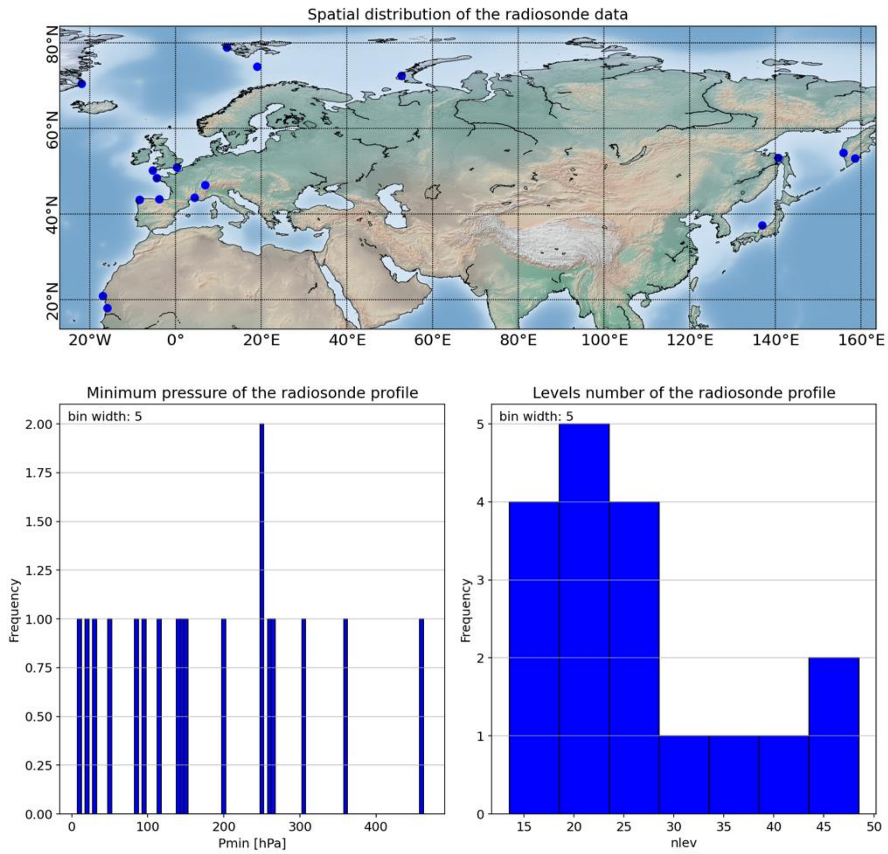


Figure 6.1.9 Case 3: spatial distribution of the match-ups with LF=0 used for ICI statistics; histograms of RS-Pmin (bottom left) and RS-levels number (bottom right)

Table 6.1.6 BIAS, SD_TA_RS and u_BIAS for **MWI Case 1.a, 2 and 3.** The highlighted rows correspond to MWI channels similar to ICI channels.

	Case 1.a (10 match-ups with LF<=100%)			**Case 2 (8 match-ups with LF<=30%, RLF=30%)**			**Case 3 (14 match-ups with LF=0, and RLF=0)**		
MWI frequency (GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD(TA_RS) (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
18.7(V)	3,74	7,89	2,50	11,67	12,44	4,40	9,41	9,38	2,51
18.7(H)	9,96	13,13	4,15	25,91	17,15	6,06	19,33	14,19	3,79
23.8V)	1,79	6,46	2,04	5,34	11,82	4,18	4,34	8,95	2,39
23.8(H)	6,67	11,10	3,51	17,98	15,16	5,36	13,11	13,22	3,53
31.4(V)	1,90	6,91	2,18	-2,07	12,13	4,29	-1,59	9,89	2,64
31.4(H)	7,93	11,91	3,77	5,57	13,92	4,92	1,63	13,97	3,73
50.3(V)	-1,32	2,55	0,81	-0,61	2,34	0,83	-0,80	1,97	0,53
50.3(H)	1,11	4,53	1,43	2,30	4,91	1,74	1,45	5,44	1,46
52.610(V)	-3,72	2,02	0,64	-1,56	1,69	0,60	-2,27	2,14	0,57
52.610(H)	-3,35	2,25	0,71	-1,14	1,65	0,58	-1,80	1,91	0,51
53.24(V)	-3,37	2,79	0,88	-1,21	2,04	0,72	-2,05	2,90	0,78
53.24(H)	-3,13	2,83	0,89	-1,01	1,99	0,71	-1,82	2,82	0,75
53.750(V)	-4,24	4,41	1,39	-1,45	3,08	1,09	-2,64	4,46	1,19
53.750(H)	-4,07	4,41	1,40	-1,31	3,07	1,09	-2,49	4,45	1,19
89.9(V)	-1,11	3,64	1,15	-2,97	2,61	0,92	-0,88	3,76	1,01
89.9(H)	2,20	7,29	2,31	-5,66	8,26	2,92	0,12	13,69	3,66
118.7503±3.20	-4,66	2,03	0,64	-2,16	2,86	1,01	-2,95	2,51	0,72
118.7503±2.10	-5,79	3,37	1,06	-2,73	3,17	1,12	-4,19	3,66	1,06
118.7503±1.40	-7,75	6,00	1,90	-3,54	4,77	1,69	-5,96	6,12	1,77
118.7503±1.20	-8,74	7,31	2,31	-3,82	5,72	2,02	-6,71	7,41	2,14
165.5±0.75	-1,22	1,77	0,56	0,63	2,35	0,83	-0,14	1,53	0,41
183.31±7.0	2,06	5,61	1,77	3,76	6,36	2,25	0,81	2,65	0,73
183.31±6.1	2,37	6,50	2,05	4,36	7,35	2,60	0,91	3,06	0,85
183.31±4.9	2,95	8,03	2,54	5,46	9,05	3,20	0,49	2,94	0,85
183.31±3.4	3,37	9,81	3,10	6,81	11,08	3,92	1,37	4,82	1,34
183.31±2.0	2,36	8,77	2,77	6,74	10,30	3,64	0,98	5,86	1,63

Table 6.1.7 BIAS, SD_TA_RS and u_BIAS for **ICI Case 1.a, 2 and 3.** The highlighted rows correspond to ICI channels similar to MWI channels.

	Case 1.a (10 match-ups, with LF<=100% RLF=100%)			**Case 2 (10 match-ups with LF<=30%, RLF=30%)**			**Case 3 (18 match-ups with LF=0, RLF=0%)**		
ICI frequency(GHZ)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
183.31±7.0	2,94	5,16	1,63	3,78	5,16	1,63	2,70	4,16	0,98
183.31±3.4	3,92	9,59	3,03	6,20	9,82	3,11	3,47	8,06	1,90
183.31±2.0	2,85	8,85	2,80	5,92	9,63	3,05	2,60	8,42	1,98
243.2±2.5(V)	-0,48	2,20	0,69	0,90	3,32	1,05	0,70	2,66	0,63
243.2±2.5(H)	-0,38	3,66	1,16	2,26	3,64	1,15	2,08	3,20	0,76
325.15±9.5	-0,37	3,42	1,08	-2,39	5,54	1,85	-1,44	4,17	1,01
325.15±3.5	0,79	8,36	2,64	1,79	9,10	3,03	0,06	6,85	1,66
325.15±1.5	0,38	7,51	2,38	2,64	8,59	2,86	-0,35	7,52	1,82
448±7.2	-0,69	7,11	2,25	1,14	7,31	2,31	-1,33	6,98	1,64
448±3.0	-2,05	5,07	1,60	-0,84	5,05	1,60	-3,28	7,01	1,65
448±1.4	-5,71	4,92	1,55	-4,08	5,37	1,70	-6,56	8,89	2,10
664±4.2(V)	-1,09	9,84	3,11	0,34	11,80	3,73	-0,62	7,86	1,85
664±4.2 (H)	-1,55	10,84	3,43	0,01	12,94	4,09	-1,35	8,88	2,09

Generally, from the comparison of Case 1.a, 2, and 3 for ICI match-ups, the better results are obtained in Case 1.a for all the frequencies except for 183.31 GHz frequencies and for 325.15±3.5 GHz that show lower BIAS values in Case 3. The BIAS uncertainty (u_BIAS) is lower for Case 3 than for the other case due to the Case-3 higher number of samples.

Case 4 and 5 are related to the statistics on match-ups when NWP information is used for filling the RS data gap (surface parameters, data above the RS maximum altitude, and NWP ozone profile instead of RTTOV fixed climatological profile) for simulating BT from RS. Figure 6.1.10 shows the spatial distribution of the 5 match-ups (with LF<=100%) used for MWI/ICI Case 4 and 5. The number of match-ups is lower than for Case 1.a due to the removal of the match-ups related to cloudy NWP profiles.

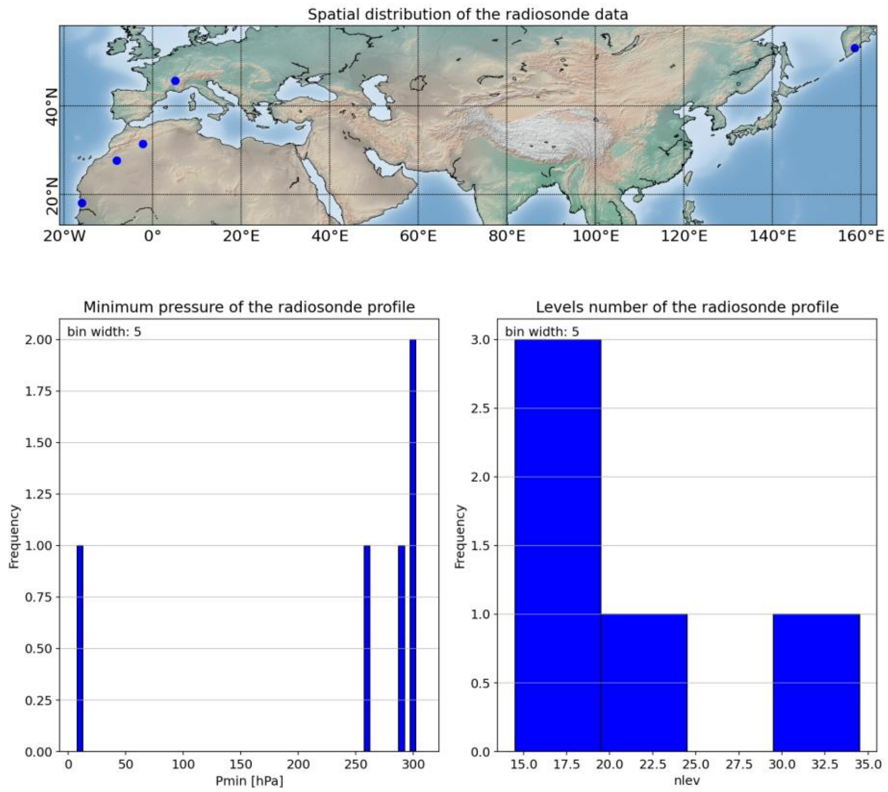


Figure 6.1.10 Case 4/5 (as introduced in Table 2.3.1): spatial distribution of the match-ups with LF<=100% used for ICI/MWI statistics when NWP information is used for simulating BT from RHARM; histograms of RS-Pmin (bottom left) and RS-levels number (bottom right)

By comparing the plot of BIAS and SD on the top left (no NWP) and on the top right panel (NWP used) of Figure 6.1.11, it is evident that the use of NWP information (both for Tskin_opt=1 and Tskin_opt=2) impacts positively the statistics for 53.24 GHz (V, H) and 53.75 GHz (V,H) (for example for 53.75 GHz(V) BIAS passes from -6.6 to -0.4 K and SD from 3.4 to 0.3 K) and for frequencies at 118.7503 $\pm$ 1.2 GHz. In Figure 6.1.12, the residuals for the above-mentioned frequencies are more confident when NWP is used (right panel) and it generally depends on the lowering of BT(obs)-BT(sim) and on the higher value of uncertainty related to BT simulation. For more opaque frequencies, NWP impact on BIAS statistics is not very evident, especially for the 183.31-GHz band. The weighted BIAS is lower for MWI more transparent frequencies (bottom right panel of Figure 6.1.11 and right panel of Figure 6.1.12), due to the higher uncertainties related to the simulated BT.
Figures 6.1.13 and 6.1.14 show the statistical scores obtained by using NWP information in BT simulated from RS without NWP information (left panel) and using NWP information (right panel). There seems to be no improvement when NWP information is used, except for 448±1.4 GHz frequency for which BIAS decreases (in module) from 8.4 to 1.4 K. Moreover, as for MWI, the uncertainty related to simulated BT is higher causing the data related to match-ups to be more confident when NWP is used (Figure 6.1.14, right panel (with NWP) versus left panel (no NWP)).

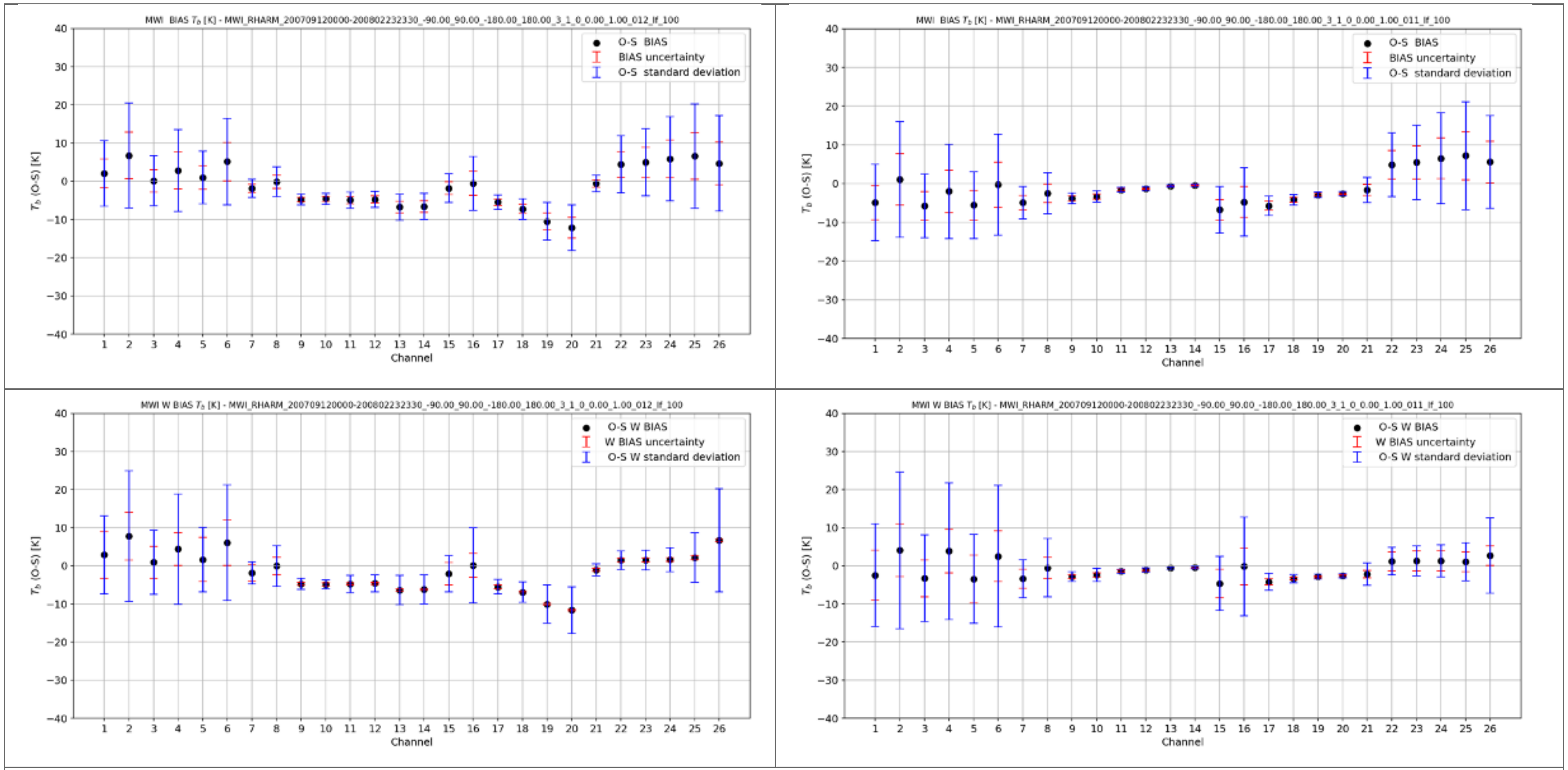


Figure 6.1.11 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom w_BIAS ± SD_wTA_RS ± u_wBIAS. **(Left) Case 4.a** MWI statistics from 5 match-ups with LF<=100 in clear sky without NWP information (**Right**) **Case 4** MWI statistics from 5 match-ups with LF<=100%, BT simulated from RS using NWP information, TSkin_opt=1.

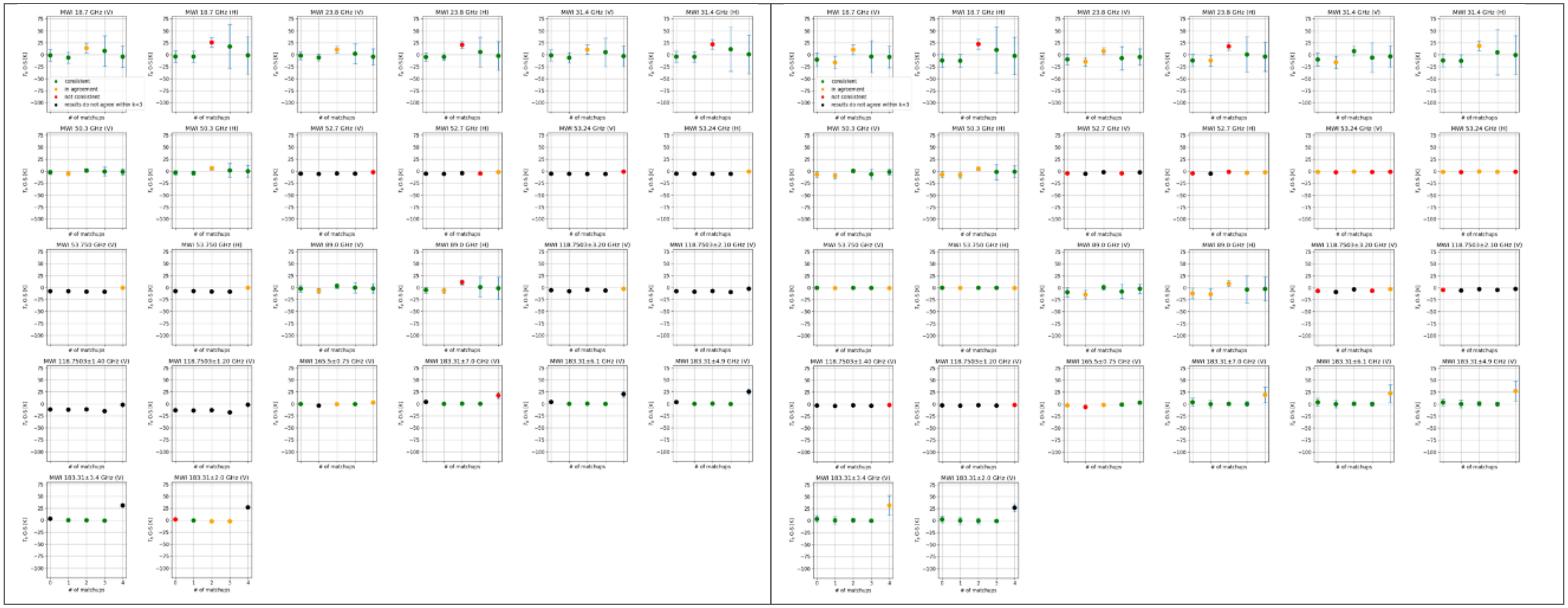


Figure 6.1.12 Case 4 MWI: TA_RS u_all (5 match-ups, LF<=100%), BT from RS simulated without NWP information (Case 4.a on the left) using NWP information, Tskin_opt=1 (Case 5 on the right).

Table 6.1.8 BIAS, SD_TA_RS and u_BIAS for **MWI Case 4,5** and and the **same case without NWP information**

	Case 4.a (5 match-ups with LF<=100%, no NWP)			Case 4 (5 match-ups with LF<=100%, Tskin_opt=1)			Case 5 (5 match-ups with LF<=100, Tskin_opt=2)		
MWI frequency (GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
18.7(V)	2,05	8,56	3,83	-4,87	9,88	4,42	-7,79	10,12	4,53
18.7(H)	6,74	13,69	6,12	1,11	14,97	6,69	-1,29	14,51	6,49
23.8V)	0,11	6,55	2,93	-5,73	8,21	3,67	-8,21	8,60	3,85
23.8(H)	2,82	10,76	4,81	-2,00	12,16	5,44	-4,06	11,91	5,33
31.4(V)	1,01	6,84	3,06	-5,58	8,65	3,87	-8,43	9,26	4,14
31.4(H)	5,14	11,32	5,06	-0,28	13,03	5,83	-2,65	12,87	5,76
50.3(V)	-1,87	2,36	1,06	-4,91	4,16	1,86	-6,64	4,94	2,21
50.3(H)	-0,14	3,90	1,74	-2,51	5,34	2,39	-3,97	5,54	2,48
52.610(V)	-4,82	1,42	0,63	-3,81	1,41	0,63	-4,44	1,75	0,78
52.610(H)	-4,60	1,45	0,65	-3,34	1,44	0,64	-3,87	1,61	0,72
53.24(V)	-4,94	2,10	0,94	-1,58	0,55	0,24	-1,84	0,65	0,29
53.24(H)	-4,74	2,08	0,93	-1,29	0,54	0,24	-1,51	0,58	0,26
53.750(V)	-6,77	3,42	1,53	-0,61	0,29	0,13	-0,68	0,27	0,12
53.750(H)	-6,61	3,42	1,53	-0,42	0,30	0,14	-0,48	0,28	0,12
89.9(V)	-1,83	3,73	1,67	-6,78	5,99	2,68	-8,96	6,88	3,08
89.9(H)	-0,51	7,03	3,14	-4,75	8,83	3,95	-6,65	9,04	4,04
118.7503±3.20	-5,48	1,85	0,83	-5,71	2,50	1,12	-6,66	3,00	1,34
118.7503±2.10	-7,27	2,73	1,22	-4,18	1,31	0,59	-4,66	1,58	0,71
118.7503±1.40	-10,49	4,95	2,21	-2,92	0,71	0,32	-3,09	0,82	0,37
118.7503±1.20	-12,14	6,01	2,69	-2,60	0,56	0,25	-2,70	0,62	0,28
165.5±0.75	-0,53	2,18	0,97	-1,63	3,25	1,45	-2,25	3,71	1,66
183.31±7.0	4,41	7,46	3,34	4,88	8,24	3,68	4,85	8,26	3,69
183.31±6.1	4,93	8,77	3,92	5,49	9,56	4,28	5,47	9,57	4,28
183.31±4.9	5,88	11,01	4,92	6,49	11,75	5,26	6,49	11,76	5,26
183.31±3.4	6,61	13,67	6,12	7,19	13,95	6,24	7,19	13,95	6,24
183.31±2.0	4,70	12,49	5,59	5,59	11,99	5,36	5,59	11,99	5,36

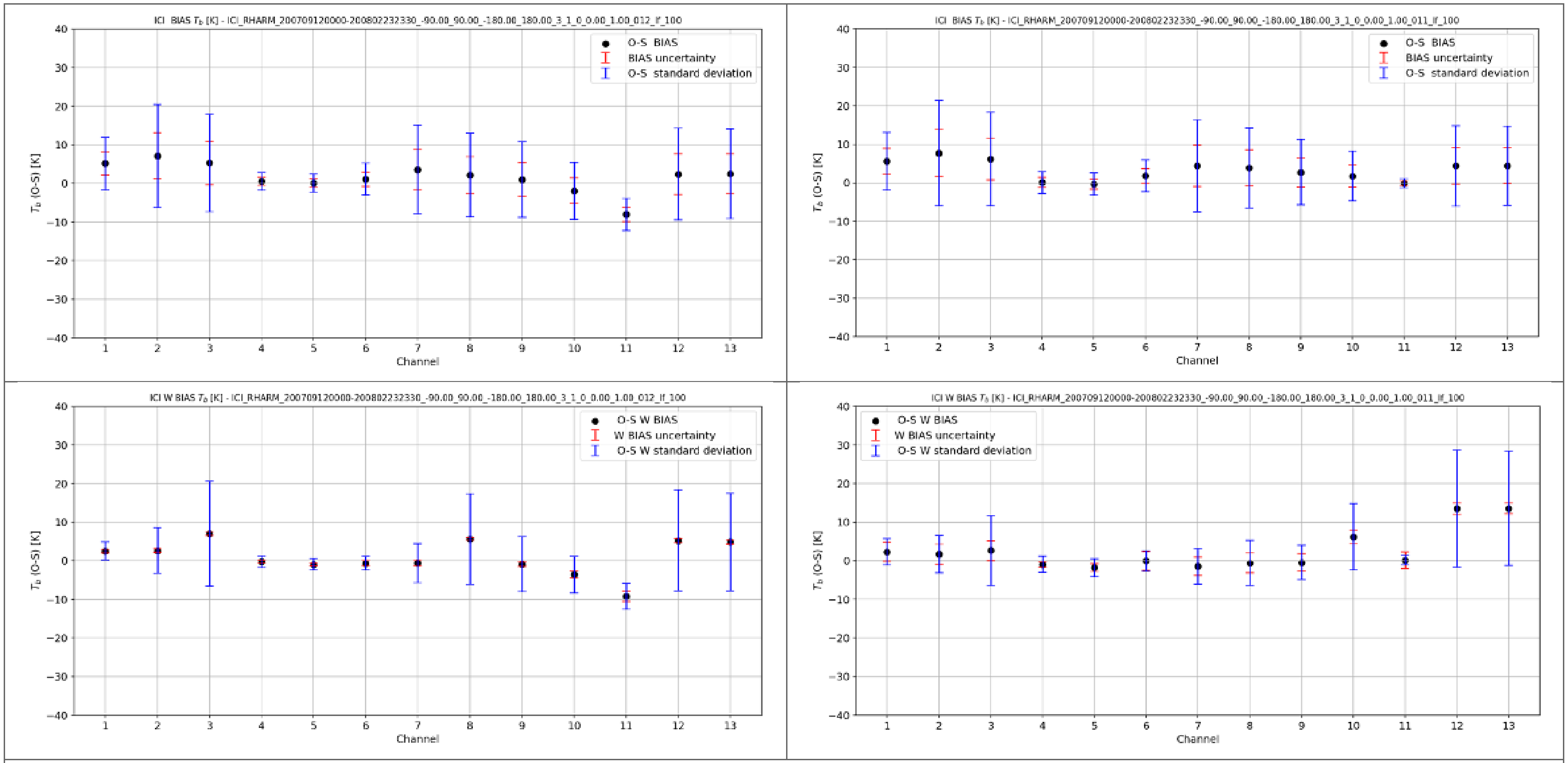


Figure 6.1.13 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom w_BIAS ± SD_wTA_RS ± u_wBIAS. **(Left) Case 4.** ICI statistics from 5 match-ups with LF<=100% in clear sky without NWP information (**Right**) **Case 4** ICI statistics from 5 match-ups with LF<=100%, BT simulated from RS using NWP information, TSkin_opt=1.

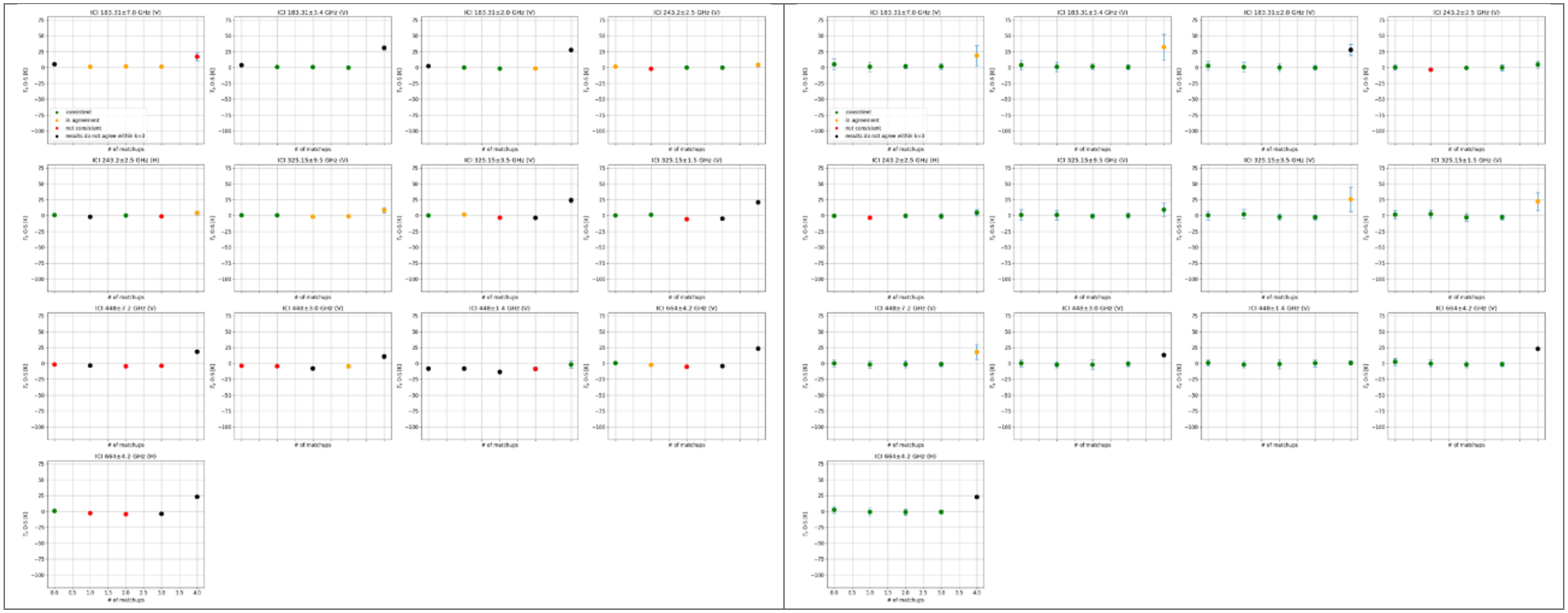


Figure 6.1.14 Case 4 ICI: TA_RS u_all (5 match-ups, LF<=100%), BT from RS simulated using NWP information, Tskin_opt=1 (on the right), without NWP information (on the left).

Table 6.1.9 BIAS, SD_TA_RS and u_BIAS for **ICI Case 4, 5** and analysis of the same match-ups without NWP information (**Case 4.a**)

	Case 4.a (5 match-ups, with LF<=100% no NWP information)			**Case 4 (5 match-ups with LF<=100%, Tskin_opt=1)**			**Case 5 (5 match-ups with LF<=100%, Tskin_opt=2)**		
ICI frequency(GHZ)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
183.31±7.0	5,17	6,78	3,03	5,64	7,49	3,35	5,61	7,51	3,36
183.31±3.4	7,13	13,34	5,96	7,73	13,68	6,12	7,73	13,68	6,12
183.31±2.0	5,26	12,59	5,63	6,16	12,14	5,43	6,16	12,14	5,43
243.2±2.5(V)	0,57	2,28	1,02	0,07	2,88	1,29	-0,27	3,16	1,41
243.2±2.5(H)	0,14	2,39	1,07	-0,31	2,92	1,31	-0,63	3,16	1,41
325.15±9.5	1,09	4,09	1,83	1,82	4,10	1,83	1,81	4,10	1,84
325.15±3.5	3,61	11,57	5,17	4,40	11,97	5,35	4,40	11,97	5,35
325.15±1.5	2,15	10,80	4,83	3,92	10,39	4,64	3,92	10,39	4,64
448±7.2	1,00	9,79	4,38	2,72	8,51	3,81	2,72	8,51	3,81
448±3.0	-1,94	7,39	3,31	1,77	6,44	2,88	1,77	6,44	2,88
448±1.4	-8,05	4,15	1,86	-0,13	1,24	0,55	-0,13	1,24	0,55
664±4.2(V)	2,41	11,86	5,30	4,38	10,50	4,69	4,38	10,50	4,69
664±4.2 (H)	2,48	11,61	5,19	4,46	10,27	4,59	4,46	10,27	4,59

6.2 Test with GMI L1B – GRUAN/RHARM

The VICIRS tool has been tested also on real radiometric observations and corresponding GRUAN and RHARM RS profiles separately. Among the currently operating radiometers, the GMI has been chosen due to the similarity in scanning strategy and the partial overlap in channel frequencies with MWI and ICI (only at 183 GHz).
Subsection 6.2.1 describes the main characteristics of GMI, while subsection 6.2.2 describes the code added to adapt the VICIRS tool to GMI. The dataset used for testing the VICIRS tool on GMI-GRUAN and GMI-RHARM match-ups and the analysis of the related statistics are shown in subsection 6.2.3 and 6.2.4, respectively.

6.2.1 GMI Overview

The GMI instrument is a multi-channel, conical-scanning, microwave radiometer serving an essential role in the near-global-coverage and frequent-revisit-time requirements of GPM. GMI observations are used principally for the retrieval of solid and liquid precipitation and of the near-surface wind speed. GMI flies at an altitude of 407 km and scans with an off-nadir angle of 48.58 degrees (Hou et al. 2014). It has two swaths, S1 and S2, about 885 km wide, and 13 channels (as in Table 3.1). Channels at 10.6, 18.7, 36.5, 89, and 166 GHz have both horizontal and vertical polarization while all other channels have only vertical (Table 6.2.1). S1 swath has nine channels with frequencies below 166 GHz while S2 swath has four channels with frequencies at and above 166 GHz. For the same GMI-RS match-up, the S1-circular TA and S2-circular TA are not coincident and, consequently, neither the 5 (3 for RHARM RS) TA types. The different S1/S2 circular TA dimensions affect the results of the cloudy tests, which may lead to an underestimation of the TA cloudy percentages. In fact, the GMI-RS match-ups are found considering S1 swath, covering the TA with S1 FOVs fully, while only partially with S2 FOVs. Thus, S2-based cloud tests may miss the cloudiness in the S1-S2 remaining area (i.e., 183.31 and 166.0 GHz clear-sky test in Table 6.2.3). This issue will be subject of future investigation.

Table 6.2.1 List of GMI channels and their characteristics (from space.oscar.wmo.int)

Central frequency (GHz)	Bandwidth (MHz)	Polarizations	NEΔT	IFOV	Pixel
10.65	100	V, H	0.96 K	30x50 km	24x12.8 km
18.7	200	V, H	0.84 K	17x29 km	12x12.8 km
23.8	400	V	1.05 K	14x24 km	12x12.8 km
36.5	1000	V, H	0.65 K	13x22 km	6.0x12.8 km
89.0	6000	V, H	0.87 K	6.9x11 km	3.0x12.8 km
166.0	3000	V, H	1.5 K	6.9x11 km	3.0x12.8 km
183.31 ± 3	3500	V	1.5 K	6.9x11 km	3.0x12.8 km
183.31 ± 7	4500	V	1.5 K	6.9x11 km	3.0x12.8 km

6.2.2 Including GMI into VICIRS-tool

The VICIRS-tool code has been updated for working with GMI observations. The flowchart is the same described in Figure 5.2 for MWI and ICI, but further code has been implemented for:

- searching GMI/RS match-ups;
- creating circular TA from GMI observations;
- extracting and analyzing TA types in terms of cloud-screening and emissivity screening;
- adapting the *bias and uncertainty analysis* (Step-V for Block 2 and 3) to BT observed and simulated for GMI by using the GRUAN processor.

The cloud test for GMI consists of threshold tests for the observations acquired in the water vapor band centered at 183.31 GHz, as well as at 89.0 GHz (V,H) and 166.0 GHz (V). The tests are listed in Table 6.2.3

Table 6.2.3 Cloud tests applied to GMI channels

MWI	183 GHz frequency	89 GHz frequency	166 GHz frequency
Test 1	$(BT_{183.31\pm3.4GHz} - BT_{183.31\pm7GHz}) > 0$ (Hong et al. 2005, to detect convective overshooting)	$BT_{89GHz,v} < 240\ K$ (by Yaping et al. (2008), over land)	$BT_{166GHz,v} < 220\ K$ by Yaping et al. (2008)
Test 2		$1 < BT_{89GHz,v} - BT_{89GHz,h} < 5\ K$ $and\ BT_{89GHz,v} < 265\ K$ (over land); $BT_{89GHz,v} - BT_{89GHz,h} \leq 20$ (over sea) (based on Gong and Wu, 2017)	

The Step-III output, *TAOOOOO_RScodeYYYYMMDDHHMM_check.nc*, is a NetCDF-4 file organized as the Step-III output for MWI/ICI. The RLF parameter in *config.ini* is not considered in the extraction of TA type because the LF information is not available for GMI FOVs as for MWI/ICI. The surface type (1 for land and 0 for sea surface) is defined for each GMI FOV by the VICIRS_GMI_LF.f90 module that defines the surface type of each GMI FOV included in TA on the basis of the 1 km spatial resolution land surface map used for AVHRR in ITPP-5 (Smith et al. 1993). The TA LF is defined for each TA type and for each GMI frequency as the percentage of land-surface GMI FOVs included in TA.

In Step IV, *main.py* calls GRUAN processor v6.3.b.0.1 executable to simulate GMI BTs from RS, using or not NWP profiles based on the NWP_opt/Tskin_opt option set by the user.

Step-V main programs have been implemented for handling GMI/RS match-ups in Block 2 and Block 3. The fields of the *TAOOOOO_RScodeYYYYMMDDHHMM_check.nc* are listed in Table A.5 of the Appendix A.

6.2.3 GMI – GRUAN test dataset description and results discussion

An example of *config.ini* for initializing the VICIRS tool to run with GMI data is shown in Figure 6.2.1. The *radiometer* and *radiosonde* options are set to “GMI” and ”GRUAN”, respectively, for the main.py to call the python code for searching match-ups between GMI and GRUAN RS, and the F90 code for processing GMI observations.

The RLF option in *config.ini* is not considered for initializing Block-2 loop, due to the unavailability of LF for GMI FOVs.

A total of 1139 GMI/GRUAN match-ups have been collected for 6 months of 2023 (January to April and September to November).

config.ini - /home/vicirs/mwi_ici_cvt_bias_radiosonde-main/ (su Poirot)

File Edit Search Preferences Shell Macro Windows

```
[USER_DEF]
#radiometer MWI(ICI) or GMI
radiometer = GMI

# Temporal range start/stop
temporal_range_strt = 2023-01-01 00:00
temporal_range_stop = 2023-04-30 23:30

# Spatial range window (N/W/S/E)
spatial_range = 90.00/-180.00/-90.00/180.00

# radiosonde GRUAN or RHARM
radiosonde = GRUAN

# Temporal colocation criteria
# 1 for -15m/+45m
# 2 for 2 for -1h/+1h
# 3 for 3 for -3h/+3h
temporal_distance = 3

# NWP
# 0 not using NWP
# 1 using NWP only if NWP is in clear sky
# 2 pass to the next match-up when NWP is cloudy
nwp = 0

# Dedicated launches
# 0 all launches
# 1 dedicated launches only if NWP is in clear sky
# 2 pass to the next match-up when NWP is cloudy
dedicated_launches = 0

# Tskin_opt
# 0 Tskin determined from RS
# 1 Tskin from RS and NWP
# 2 Tskin(NWP)
Tskin_opt = 0

# maximum LF (%) for selecting MWI/ICI FOVs to be included in TA
RLF = 100

[GLOBAL_VAR]
ICI_path=data_in/ICI/ICI_L1B_TDP_in_granules/4655_4656
MWI_path=data_in/MWI/MWI_L1B_TDP_in_granules/4655_4656
GMI_path=data_in/GMI/HDD
GRUAN_path=data_in/GRUAN
RHARM_path=data_in/RHARM
TA_path=data_in/TA_data
GP_path=./GRUAN_Processor_v6.3.b.0.1/bin/
data_out=data_out

[MWI_ICI_DOWNLOAD]

[GRUAN_DOWNLOAD]

[RHARM_DOWNLOAD]
```

Figure 6.2.1 *config.ini* for searching and analyzing match-ups between GMI observations and GRUAN RS.

As for ICI/MWI-RHARM match-ups (subsection 6.1), the RS analysis in Step-II is mandatory to decide if the related match-up is useful for calibration (as described in subsection 6.1.2). Among the 1139 initial GMI-GRUAN match-ups only 89 satisfy all the Step-II tests (QC test, the clear-sky test and the AMD test). In detail, 92.2 % of the initial 1139 RS was removed. Among these, 40.6% of the initial RS failed QC test, 76.6% failed the clear-sky test and 43.2% failed AMD test. Table 6.2.4 shows the statistics of the analyzed GRUAN RS grouped according to their latitude and it is evident that the percentage of failure in all the latitude bands depends above all on cloud cover, being higher than 51.5% in all latitude bands.

Table 6.2.4 Statistics of the RS related to GRUAN-GMI match-ups (from 01/01/2023 to 30/04/2023 and from 01/09/2023 to 30/11/2023)

	Polar latitude	**Mid-latitude**	**Subtropical latitude**	**Tropical latitude**	**All latitude**
RS total (#)	193	660	196	90	1139
RS discarded (#)	177	604	179	89	1050
RS useful (#)	16	56	16	1	89
QC fails(%)	70.5	18.6	67.4	80.0	40.6
Cloudy fails(%)	80.8	81.7	51.5	84.4	76.6
AMD fails(%)	35.2	47.3	34.7	50.0	43.2
total fails(%)	91.7	91.5	91.8	98.9	92.2

The cases described in Table 6.2.5 have been investigated to query the GMI-GRUAN dataset. The Case 1 for GMI-GRUAN match-ups includes 68 samples (corresponding to clear-sky TA only) related to the GRUAN sites spatially distributed as shown in Figure 6.2.2.

Table 6.2.5 List of query.ini parameters set for the statistical analysis of the GRUAN-GMI test dataset. Temporal range is in UTC (format: YYYY-MM-DD HH:MM:SS). Case number (CN) and number of available match-ups (#MU) are also reported.

Temporal range (UTC)	**Temporal collocation criteria/TA type**	**RLF (%) LF min: LF: max**	**TA cloudy %**	**NWP_opt/ Tskin_opt**	**CN (#MU)**	**Figures**	**Tables**
from 2023-01-01 0:00 to 2023-04-30 23:30 and from 2023-09-01 0:00 to 2023-10-30 23:30	3/1	[0:100]	0	0/0	**1 (68)**	6.2.2, 6.2.3, 6.2.4, 6.2.5	6.2.7
	1/1	[0:100]	0	0/0	**2 (19)**	6.2.3, 6.2.4, 6.2.5	
	3/1	[100:100]	0	0/0	**3 (45)**	6.2.6	6.2.8
	3/1	[0:80]	0	0/0	**4(15)**		
	3/1	[0:100]	0	1/1	**5(22)**	6.2.7, 6.2.8,6.2.9,6.2.10	6.2.9 6.2.9
	3/1	[100:100]	0	1/2	**6(22)**		
	3/1	[100:100]	0	Case 5(6) without NWP information	**7(22)**		

The GRUAN sites used for the GMI-GRUAN statistics are TEN (island), CAB, HKO (near coastline), PMO, LAU, PAY, LIN, SOD and POT. Five dedicated launches (DLs) (listed in Table 6.2.6) were performed in POT and have been used for the VICIRS-tool test on GMI.

Although ten DLs were originally planned, so far only five have been launched because of cloudy sky conditions in correspondence of GPM overpasses and the temporary unavailability of GMI data due to GPM Core Observatory satellite orbit boost maneuvers. As a result of these maneuvers, the GMI observing parameters, the footprint sizes and Earth incidence angle changed and thus the data processing needed to be updated. This caused a lack of GMI data approximately from November 2023 until March/April 2024.

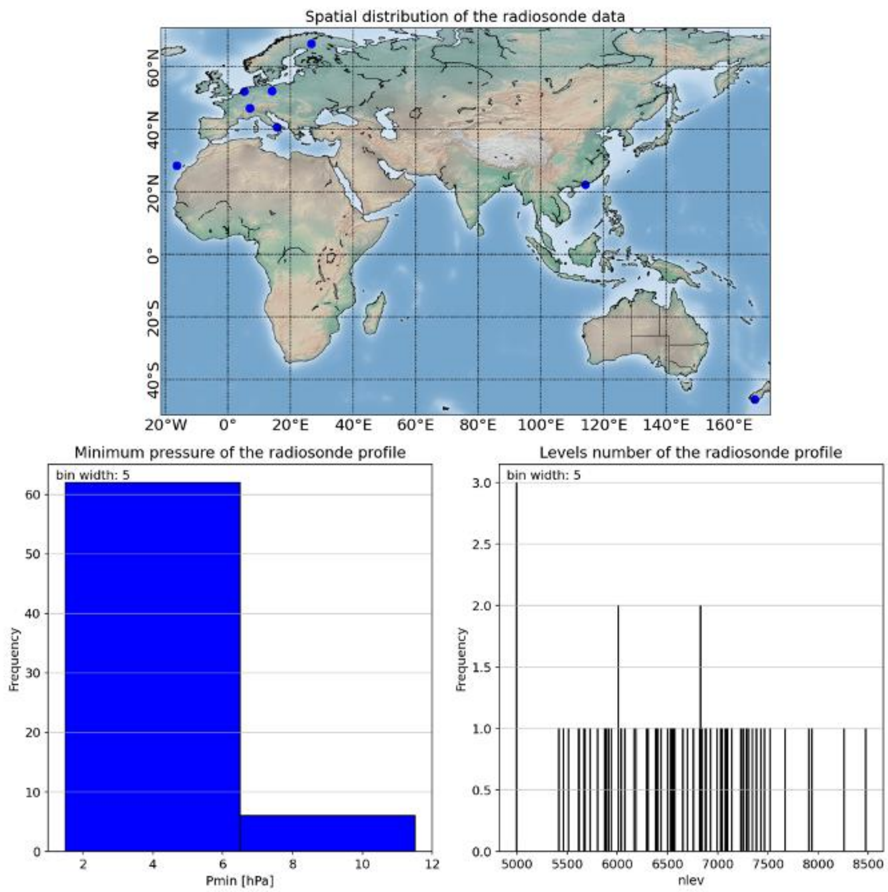


Figure 6.2.2 Spatial distribution of RS GRUAN sites used for **Case-1** GMI statistics.

Table 6.2.6 List of DL from POT GRUAN site

Step-III output	GRUAN-DL/GPM-overpass time difference	TA radius
TA54641_POT-202310102200_check.nc	1898 seconds= 31 minutes	50,0 Km
TA54662_POT-202310120657_check.nc	1248 seconds=20 minutes	36,7 Km
TA54841_POT-202310231830_check.nc	1607 seconds=27 minutes	50,0 Km
TA54887_POT-202310261718_check.nc	2082 seconds=35 minutes	50,0 Km
TA54933_POT-202310291615_check.nc	2161 seconds=36 minutes	50,0 Km

Case-1 statistics are shown in Figure 6.2.3 (left panel), Figure 6.2.4 (left panel) and Table 6.2.7. Right panel of Figures 6.2.3 and 6.2.4 and Table 6.2.7 show Case-2 statistics. Case-2 match-ups differ from Case-1 ones for the choice of temporal distance reduced to -15'<=Dt<=45' to evaluate the impact of the time lag between sonde launch and satellite overpass on the statistics.
The top panel of Figure 6.2.3 shows the plot of BIAS$\pm$ SD_TA_RS (u_BIAS) while the bottom panel shows the plot of the wBIASn$\pm$ SD_wTA_RS (u_wBIAS), similarly to ICI/MWI-RHARM (subsection 6.1).
BIAS$\pm$SD ranges from (-10.0$\pm$16.9)K for 10.65GHz to (0.2$\pm$1.4)K for 183.31$\pm$3GHz for Case 1. Case-2 BIAS and SD decrease and BIAS$\pm$SD ranges from (-6.1$\pm$13.0)K for 10.65GHz to (0.2$\pm$1.1)K for 183.313 GHz. Large biases are found for less opaque frequencies, mainly at GRUAN sites located near coastlines (TEN, HKO and LAU) where the approximated LF affects negatively the BT simulated from RS (combination of BT simulated over land and BT simulated over sea). In fact, TA-LF is determined as the average of the surface type (0/1 for sea/land) estimated for each GMI FOVs included in TA using an indirect method, i.e., ITPP-5 land surface map. This is likely much less accurate than the direct estimate that will be available from MWI/ICI FOVs. This can be seen from the statistics related to the match-ups with LF<=80% (Case 4), shown in Table 6.2.8 and in the right panel of Figure 6.2.6.

Table 6.2.7 BIAS, SD_TA_RS and u_BIAS for GMI Case 1 and 2

	Case 1 (68 match-ups, with LF<=100%,-3h<=Δt<=3h)			**Case 2 (19 match-ups with LF<=100%, -15'<=Δt<=45')**		
GMI frequency(GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
10.65(V)	-9,99	16,90	2,05	-6,05	12,96	2,97
10.65(H)	-8,66	18,72	2,27	-5,01	15,28	3,50
18.7(V)	-6,43	12,09	1,47	-3,41	9,57	2,19
18.7(H)	-4,84	15,19	1,84	-1,87	12,57	2,88
23.8(V)	-4,31	8,65	1,05	-2,05	7,11	1,63
36.5(V)	-1,73	8,61	1,04	-0,07	6,20	1,42
36.5(H)	-1,82	12,55	1,52	0,30	9,45	2,17
89.0(V)	0,90	5,38	0,65	0,95	2,75	0,63
89.0(H)	0,83	7,27	0,88	1,65	4,56	1,05
166.0(V)	-0,56	1,76	0,24	-0,95	1,52	0,38
166.0(H)	-0,07	3,22	0,45	-0,94	2,66	0,66
183.31±3	0,16	1,43	0,20	-0,15	1,13	0,28
183.31±7	-0,79	1,01	0,14	-0,80	0,65	0,16

Case-2 statistical scores for window channels are slightly lower than in Case-1 because some match-ups characterized by higher residuals (1 HKO-GMI match-up, 5 LAU-GMI match-ups and 10 TEN-GMI match-ups) have been removed. This can be observed by comparing the left panel (Case-1) and the right panel (Case-2) of Figure 6.2.5. In figure 6.2.5 the observed BT are plotted against the simulated ones and the markers are colored differently according to the GRUAN site. Note that when comparing the left and right panels the site color legend changes (in particular, TEN is in blue on the left and in green on the right).

Table 6.2.8 BIAS, SD_TA_RS and u_BIAS for GMI Case 3, Case 4 and Case 3 without outliers (SOD, TEN and HKO match-ups)

	Case 3 (45 match-ups, with LF=100%)			**Case 4 (15 match-ups with LF<=80%)**			**Case 3 without SOD, TEN and HKO outliers (39 match-ups)**		
GMI frequency(GHZ)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
10.65(V)	-6,09	11,90	1,77	-21,46	21,11	5,45	-3,94	3,94	0,63
10.65(H)	-6,23	12,52	1,87	-17,87	26,75	6,91	-4,61	5,25	0,84
18.7(V)	-4,36	8,55	1,28	-13,51	16,14	4,17	-3,05	3,32	0,53
18.7(H)	-3,86	10,42	1,55	-10,70	22,42	5,79	-3,10	4,82	0,77
23.8(V)	-2,69	5,96	0,89	-9,70	11,78	3,04	-2,07	2,77	0,44
36.5(V)	0,23	7,24	1,08	-7,15	9,82	2,54	0,12	2,86	0,46
36.5(H)	-0,12	10,10	1,51	-7,31	16,70	4,31	-0,26	4,00	0,64
89.0(V)	1,65	6,21	0,93	-0,18	2,75	0,71	0,27	2,55	0,41
89.0(H)	1,05	7,31	1,07	1,23	7,65	1,97	-0,21	3,71	0,59
166.0(V)	-0,41	1,85	0,33	-0,31	1,46	0,39	-0,63	1,42	0,25
166.0(H)	-0,25	1,82	0,32	1,48	4,09	1,09	-0,45	1,42	0,26
183.31±3	0,39	1,53	0,27	-0,29	1,21	0,32	0,44	1,56	0,28
183.31±7	-0,71	1,09	0,19	-1,03	0,80	0,21	-0,71	1,11	0,20

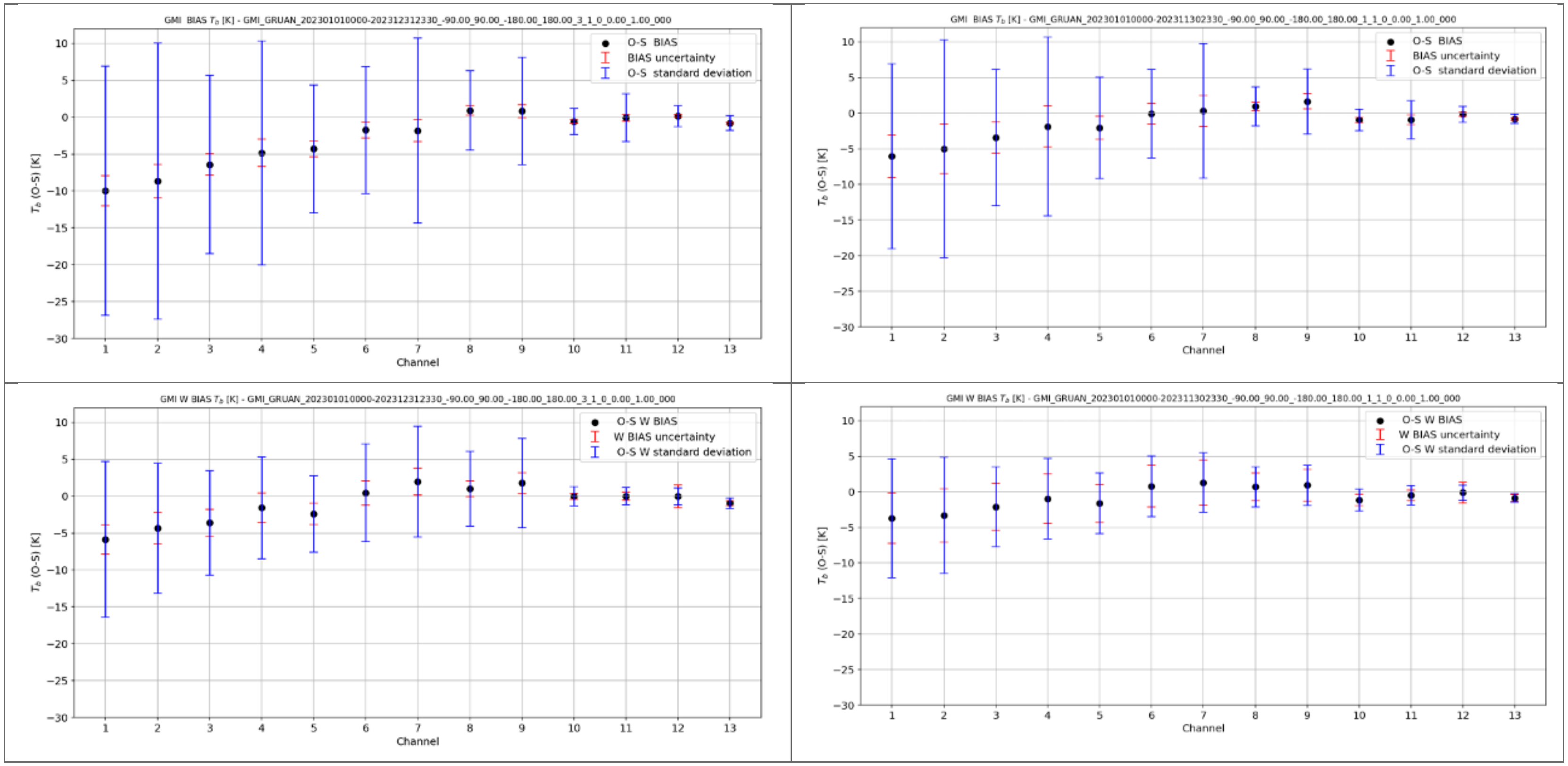


Figure 6.2.3 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom wBIAS ± SD_wTA_RS ± u_wBIAS for TA type=1. **Case 1 (left), case 2(right).**

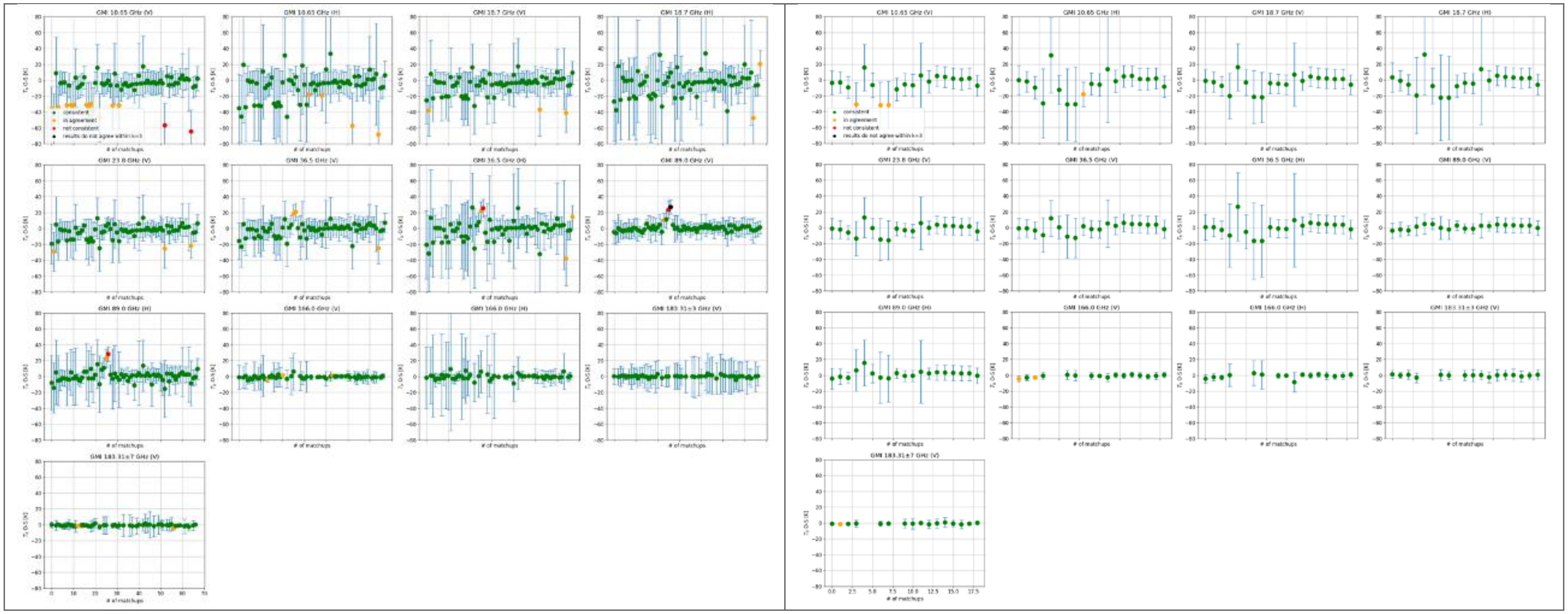


Figure 6.2.4. TA_RS u_all for **Case 1** (left), 68 match-ups in clear-sky, LF<=100%,-3h<=Δt<=3h; for **Case 2** (right), 19 match-ups in clear-sky, LF<=100%,-15'<=Δt<=45'.

.

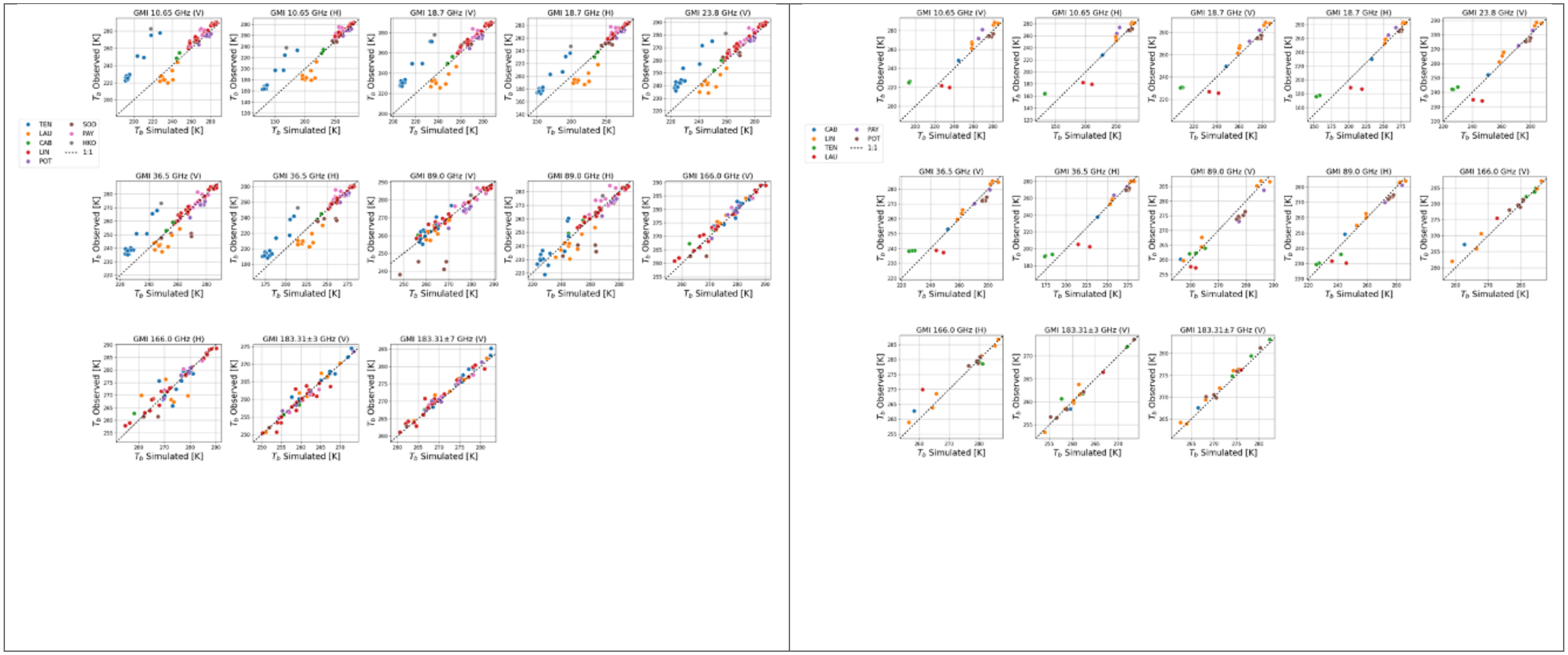


Figure 6.2.5. Scatterplot of BT(Obs) vs BT(sim) for **Case-1 (left)** and **Case-2 (right).**

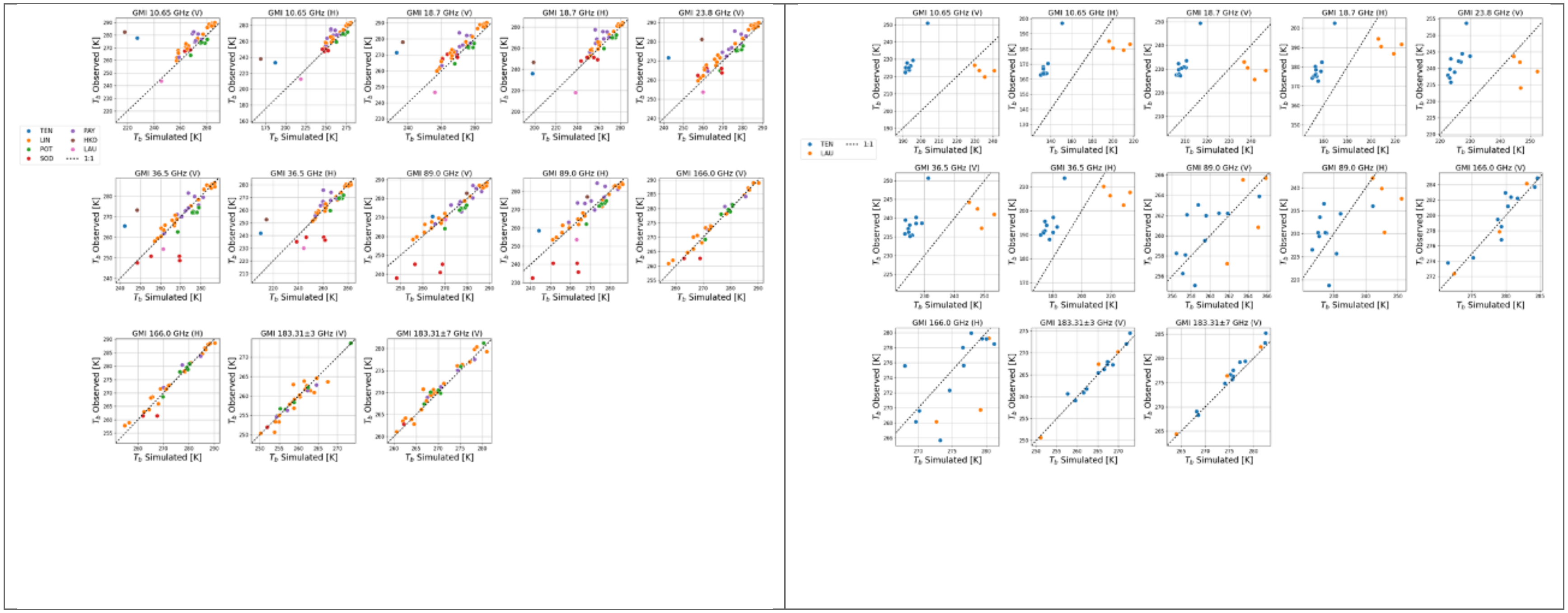


Figure 6.2.6. Scatterplot of BT(Obs) vs BT(sim) for **Case-3 (left)** and **Case-4 (right).**

In order to better understand the influence of LF on GMI-GRUAN overall statistics, Case 3 (LF=100%) and Case 4 (LF<=80%) have been examined. Table 6.2.8 shows the results for Case 3 and Case 4. BIAS values for match-ups completely over land (Case 3) lowers by about 50% when compared with Case-1 statistics. Case-3 statistics are influenced by outliers (few match-ups with residuals higher than 20K) related to SOD, TEN, and HKO sites. Despite TEN and HKO being island and coastline sites, Case-3 correspondent match-ups are classified as completely over land because the radiosonde path is completely over land and the absence of LF-FOV information causes inaccurate estimates of the LF associated to each GMI frequency. When TEN and HKO outliers are removed from Case 3, the bias for less opaque channels decreases by 30%, as shown in Table 6.2.8.

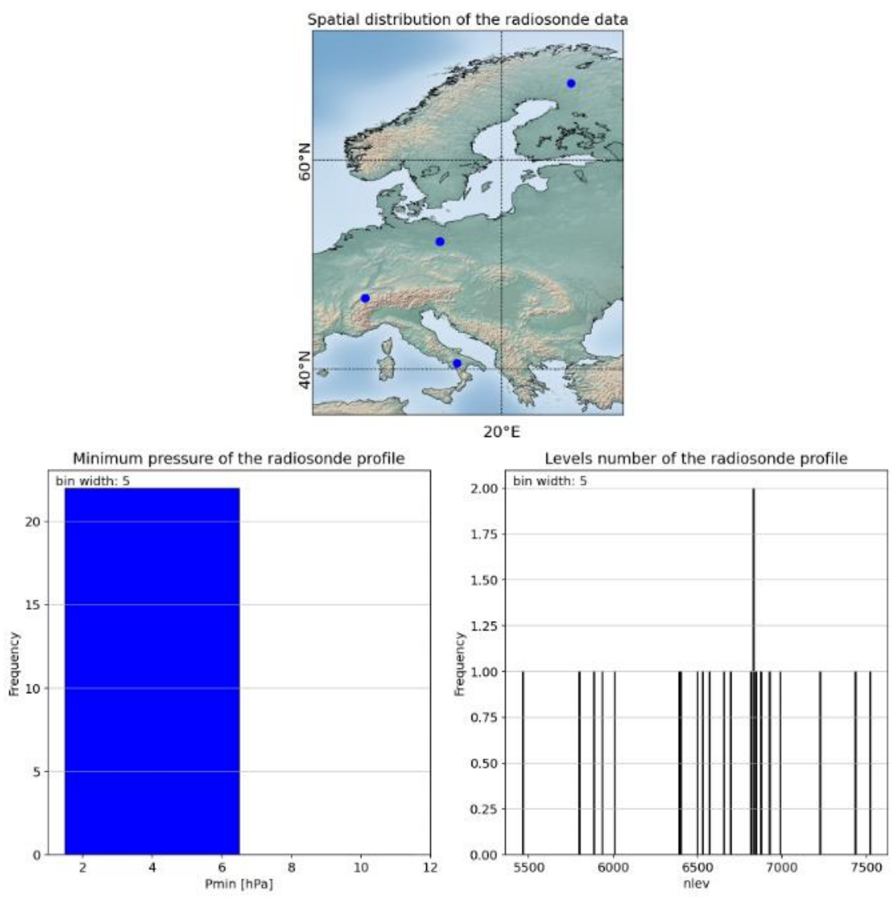


Figure 6.2.7 Spatial distribution of RS GRUAN sites used for GMI statistics (Case 5, 6 and 7 in Table 6.2.5).

Cases 5 and 6 indicate match-ups for which the NWP information has been added to RS for filling the gap of surface parameters and of data above the RS pressure top level. For these cases the match-ups number lowers to 22 because only match-ups related to NWP profile in clear sky are considered. Figure 6.2.8 shows the statistics for original (top) and weighted (bottom) residuals. The left panel shows the statistics for Case 7 that analyzes the Case 5 (6) match-ups without considering NWP information, while the right panel for Case 5 (with NWP information). NWP information decreases BIAS slightly, in fact the difference in BIAS between Case 7 (without NWP) and Case 5/Case 6 (with NWP) is about 0.1/0.4 K for 10.65 GHz and about 0.5 K for 166.0 GHz. On the contrary, SD_TA_RS values are higher when NWP information is used: SD_TA_RS for Case-5/Case-6 10.65 GHz (V) is 7.2/6.9 K whereas it is 6.1 K when NWP information is not supplied (Case 7). Statistical scores for Case 7 are lower than for Case-1, resulting from the removal of some high residuals according to the clear-sky test applied to the collocated NWP profiles. The improvement of Case-5/6 with respect to Case 7 is due to the useful information brought by NWP. All three cases (5/6/7) show better statistics with respect to Case 1 because of the removal of faulty match-ups from TEN, HKO, and SOD.

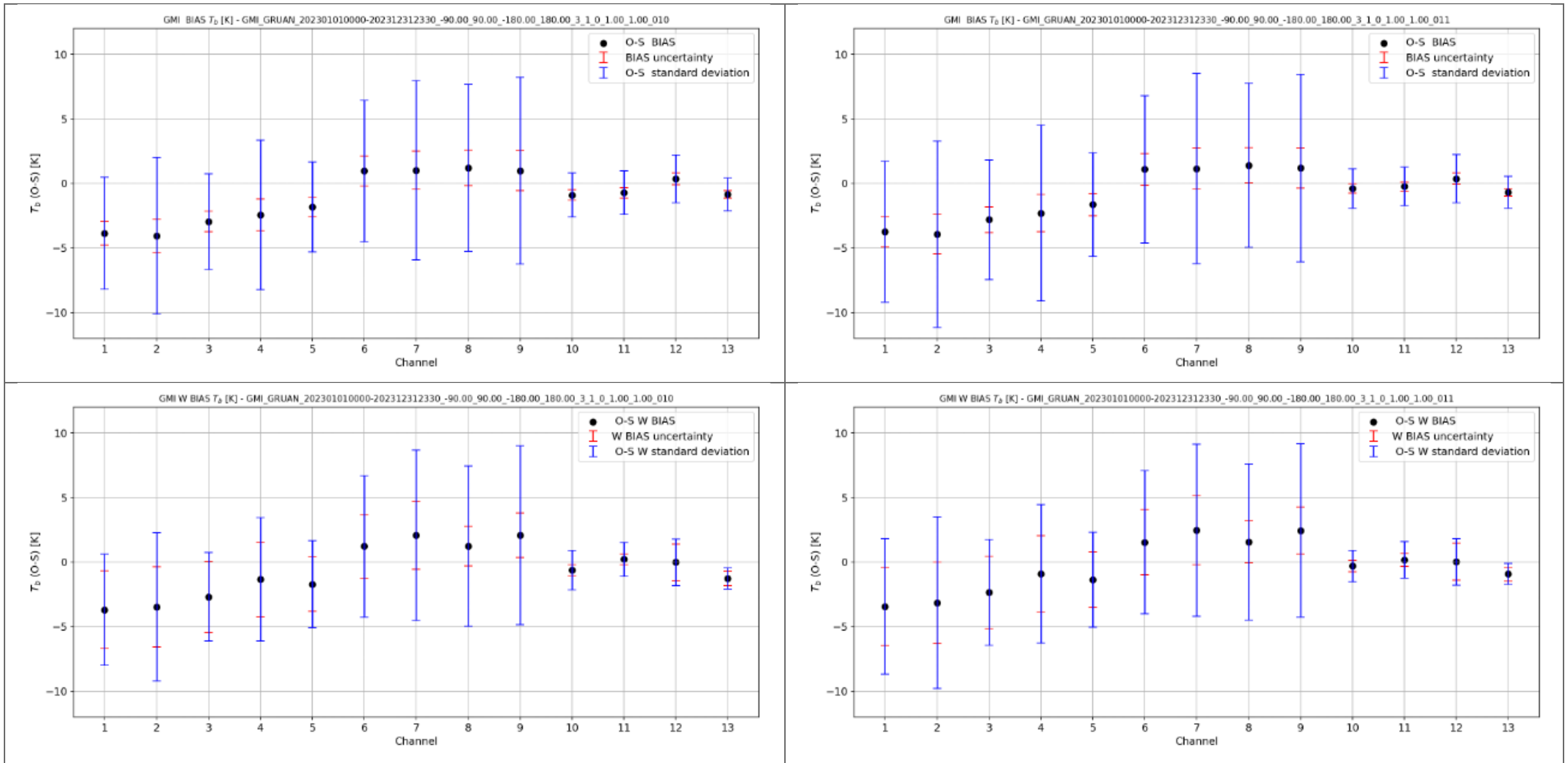


Figure 6.2.8 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom wBIAS ± SD_wTA_RS ± u_wBIAS (22 match-ups) for TA type=1, **Case 5** (on the right), with NWP information, TSkin_opt=1; **Case 7** without NWP information (on the left).

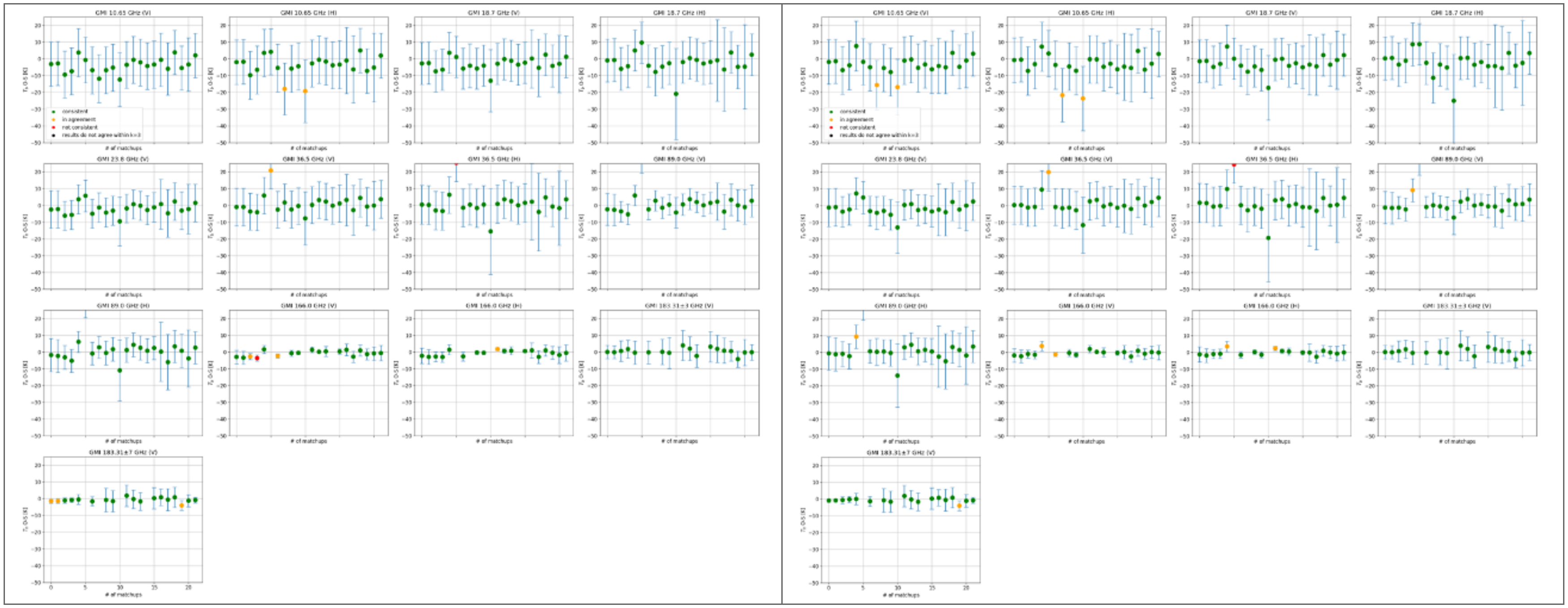


Figure 6.2.9 TA_RS u_all (22 match-ups, LF=100%), **Case-5** BT from RS simulated using NWP information, Tskin_opt=1 (on the right), **Case 7** without NWP information (on the left).

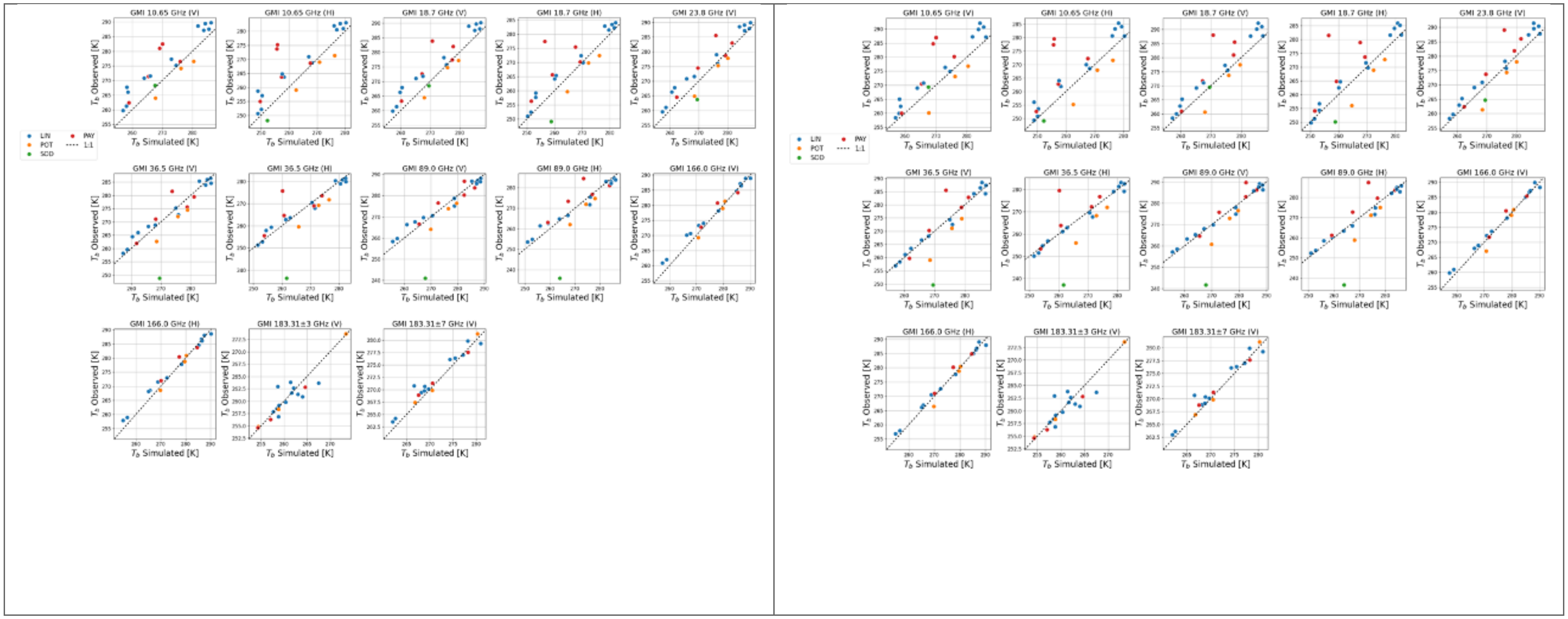


Figure 6.2.10. Scatterplot of BT(Obs) vs BT(sim) for **Case 7,** without NWP (left) and **Case 5,** with NWP information and TSkin_opt=1(right**).**

Table 6.2.9 BIAS, SD_TA_RS and u_BIAS for GMI Case 5, 6 and 7

	Case 7 (22 match-ups, no NWP information)			**Case 5 (22 match-ups NWP information, TSkin_opt=1)**			**Case 6 (22 match-ups NWP information, TSkin_opt=2)**		
ICI frequency(GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
10.65(V)	-3,86	4,34	0,93	-3,74	5,46	1,16	-3,45	4,94	1,05
10.65(H)	-4,05	6,05	1,29	-3,93	7,20	1,54	-3,66	6,90	1,47
18.7(V)	-2,95	3,71	0,79	-2,81	4,63	0,99	-2,54	4,30	0,92
18.7(H)	-2,44	5,80	1,24	-2,30	6,81	1,45	-2,04	6,71	1,43
23.8(V)	-1,82	3,49	0,74	-1,64	4,01	0,86	-1,40	3,87	0,83
36.5(V)	0,96	5,46	1,16	1,10	5,70	1,22	1,35	5,75	1,23
36.5(H)	1,01	6,94	1,48	1,15	7,37	1,57	1,39	7,46	1,59
89.0(V)	1,20	6,47	1,38	1,41	6,36	1,36	1,61	6,52	1,39
89.0(H)	0,99	7,23	1,54	1,19	7,27	1,55	1,38	7,48	1,59
166.0(V)	-0,90	1,72	0,41	-0,40	1,53	0,36	-0,36	1,70	0,40
166.0(H)	-0,73	1,69	0,40	-0,24	1,50	0,35	-0,21	1,68	0,40
183.31±3	0,35	1,86	0,44	0,37	1,86	0,44	0,36	1,87	0,44
183.31±7	-0,86	1,26	0,30	-0,69	1,23	0,29	-0,71	1,31	0,31

6.3 GMI – RHARM test dataset description and results discussion

A total of 3156 GMI/RHARM match-ups have been collected for 3 months of 2019 (September to November) from 20° to 65° latitude North and from 15° to 40° longitude East. As for GMI-GRUAN analysis, *Pmin<=10* hPa and *nlev*>=40 filters have been applied.

Among the 3156 initial match-ups only 82 satisfy all the Step-II tests. Table 6.3.1 shows the statistics of the RHARM RS. In detail, 97.4% of the initial 3156 RS was removed. Among these, 90.7% of the initial RS failed QC test, 59.2% failed the clear-sky test and 8.5% failed AMD test. In Table 6.3.1 it is evident that the percentage of failure in all the latitude bands depends above all on QC tests, being higher than 78.0% in all latitude bands.

Table 6.3.1 Statistics of the RS related to RHARM-GMI match-ups (from 09/01/2019 to 30/11/2019)

	Polar latitude	Mid-latitude	Subtropical latitude	Tropical latitude	All latitude
RS total (#)	513	2346	252	45	3156
RS discarded (#)	493	2290	246	45	3074
RS useful (#)	20	56	6	0	82
QC fails(%)	78.0	92.6	96.8	100.0	90.7
Cloudy fails(%)	82.3	58.6	26.6	11.1	59.2
AMD fails(%)	10.1	7.8	10.3	13.3	8.5
total fails(%)	96.1	97.6	97.6	100.0	97.4

Table 6.3.2 List of query.ini parameters set for the statistical analysis of the GRUAN-GMI test dataset. Temporal range is in UTC (format: YYYY-MM-DD HH:MM:SS). Case number (CN) and number of available match-ups (#MU) are also reported.

Temporal range (UTC)	Temporal collocation criteria/TA type	RLF (%) LF min: LF: max	TA cloudy %	NWP_opt/ Tskin_opt	CN (#MU)	Figures	Tables
2019-09-01 0:00 to 2019-11-30 23:00	3/1	[0:100]	0	0/0	**1 (63)**	6.3.1, 6.3.2, 6.3.3	6. 3.3
	1/1	[100:100]	0	0/0	**2 (41)**	6.3.2,6.3.3	

Figure 6.3.1 shows the geographical distribution of RHARM sites used for statistics and with the histograms of *Pmin* and of *nlev* (from 40 to 60, with one RS with nlev=99). Statistical scores for both Case 1 and Case 2 are shown in Figures 6.3.2 and 6.3.3 left and right panel, respectively, and in Table 6.3.3.

As for GMI-GRUAN analysis, the statistics lower when only match-ups over land are considered, e.g. BIAS/SD for 10.65GHz (V) decreases from (-3.3±14.9)K to (0.4±10.2)K. The largest values for both cases, (8.4±11.9)K and (10.1±12.6)K, are related to 183.31±3.4 GHz and are due to the high residuals with k>=3 in Figure 6.3.3.

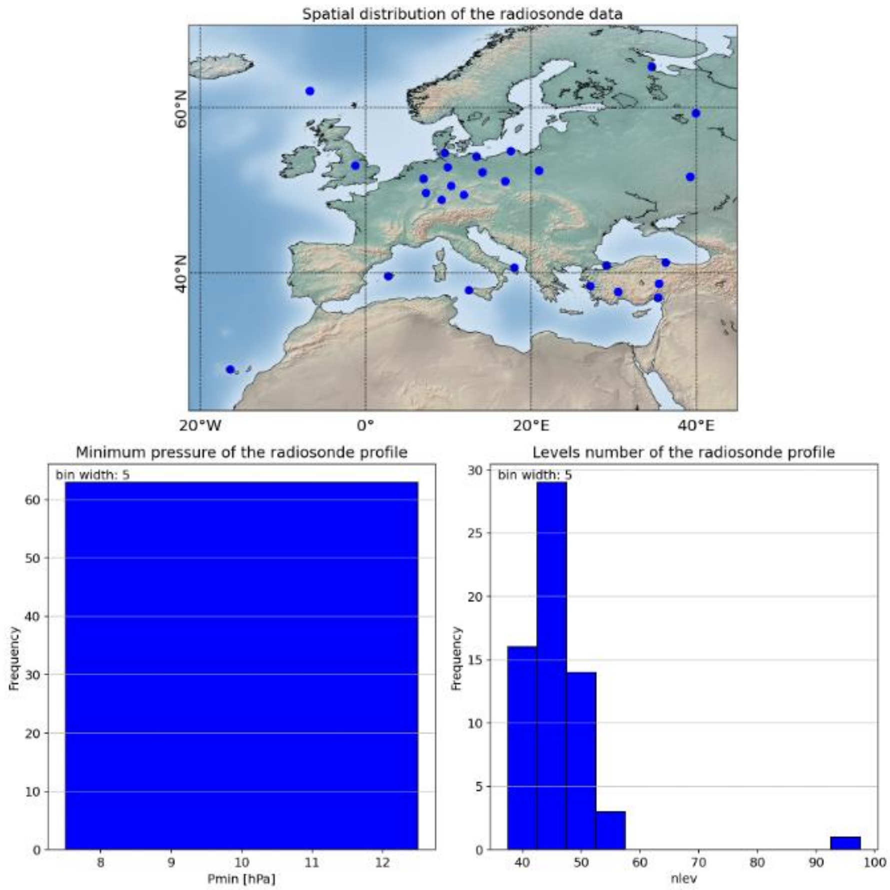


Figure 6.3.1: Top: spatial distribution of the 63 GMI/RHARM match-ups (cloud screening applied to both RS and TA). Bottom: histograms of RS minimum pressure (*Pmin,* left) and number of levels (nlev, right).

The higher values of some residuals in the 183.31 GHz region are due to very low simulated BT with respect to the observed one. From the analysis of the RHARM profiles it resulted that the low 183.31±3.4 GHz BT values are associated with high values of RH at the lowest pressure levels (~10 hPa). In particular, residuals lower than 2 K are related to profiles with RH (at 10 hPa) in the range 3.6±3.2 % while residuals higher than 20 K are related to RH in the range 15.2±5.8 %. This may be related to undetected cases of the well-known RS issues with RH measurements at low temperatures, which are corrected in the GRUAN dataset and should also be corrected in the RHARM dataset (Madonna et al., 2022).

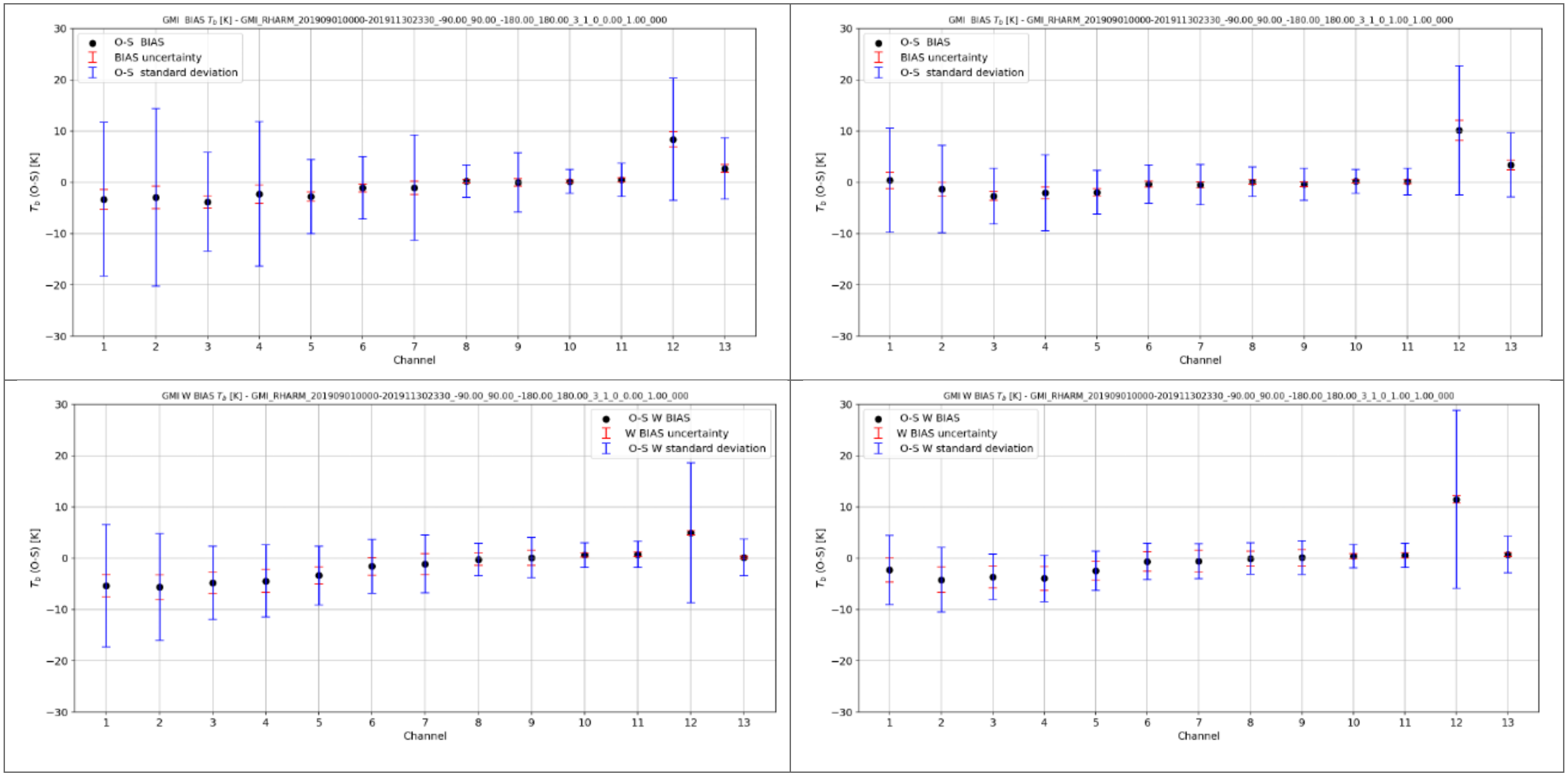


Figure 6.3.2 Top: BIAS_TA_RS ± SD_TA_RS ± u_BIAS; Bottom wBIAS ± SD_wTA_RS ± u_wBIAS. **Case 1** (63 match-ups) (left), **Case 2** (41) match-ups) (right).

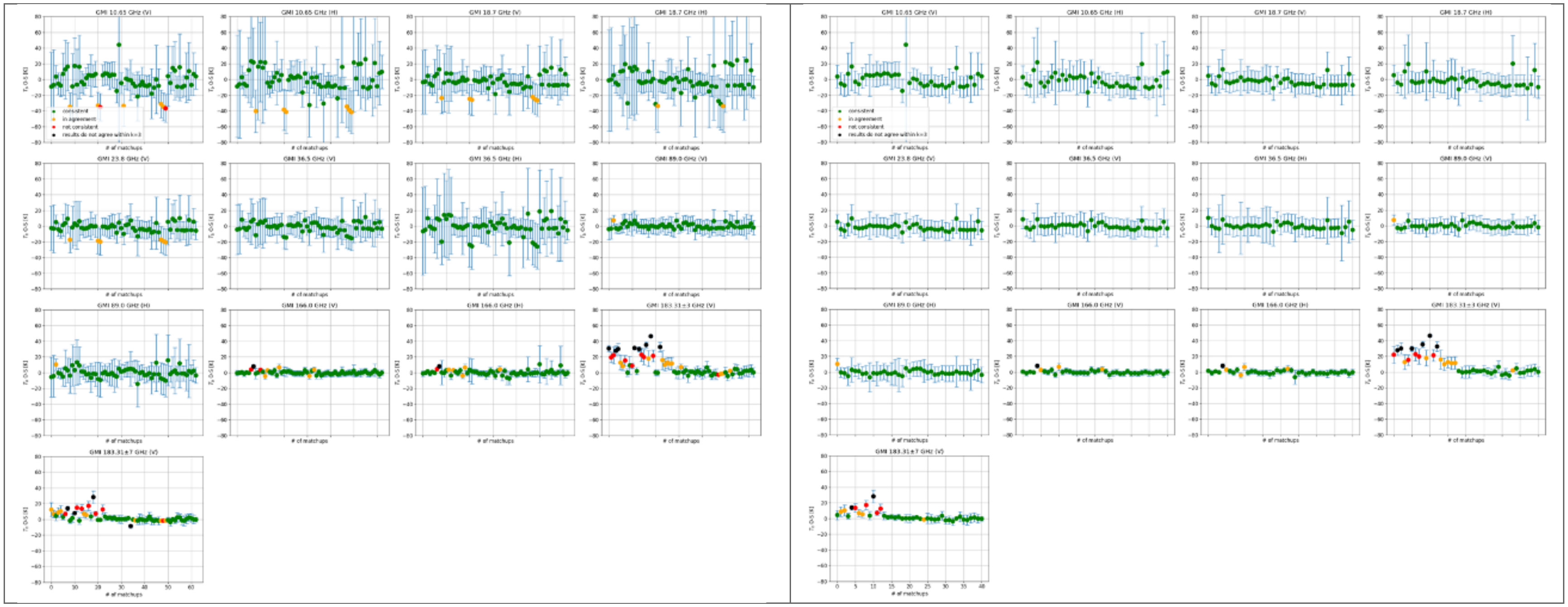


Figure 6.3.3 TA_RS u_all for **Case 1** (left), 63 match-ups in clear-sky, LF<=100%; for **Case 2** (right), 41 match-ups in clear-sky, LF=100%.

Table 6.3.3 BIAS, SD_TA_RS and u_BIAS for GMI Case 1 and 2

	Case 1 (63 match-ups, with LF<=100%)			Case 2 (41 match-ups with LF=100%)		
ICI frequency(GHz)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)	BIAS (K)	SD_TA_RS (K)	u_BIAS (K)
10.65(V)	-3,32	14,99	1,89	0,37	10,15	1,59
10.65(H)	-2,95	17,31	2,18	-1,33	8,52	1,33
18.7(V)	-3,81	9,62	1,21	-2,67	5,44	0,85
18.7(H)	-2,26	14,11	1,78	-2,04	7,38	1,15
23.8(V)	-2,79	7,24	0,91	-1,93	4,24	0,66
36.5(V)	-1,08	6,11	0,77	-0,35	3,74	0,58
36.5(H)	-1,05	10,27	1,29	-0,46	3,93	0,61
89.0(V)	0,21	3,17	0,40	0,09	2,87	0,45
89.0(H)	0,00	5,77	0,73	-0,36	3,13	0,49
166.0(V)	0,18	2,33	0,29	0,21	2,32	0,36
166.0(H)	0,45	3,26	0,41	0,14	2,60	0,41
183.31±3	8,40	11,92	1,50	10,14	12,62	1,97
183.31±7	2,70	5,99	0,75	3,40	6,28	0,98

7. Conclusions

The VICIRS tool has been designed and developed to assist the Cal/Val activities of MWI/ICI by collecting and comparing satellite observations with RT simulations from RS, considering the overall uncertainty arising from the two sources and their collocation. The VICIRS tool is currently developed to search match-ups between MWI/ICI and RS from the GRUAN and RHARM archives. It also handles ancillary NWP profiles to fill the data gap of surface parameters and levels above the RS burst altitude. For each match-up, different types of TAs are considered depending on the availability of sonde drift. A statistical analysis of the difference between observed and simulated BTs is determined for each MWI/ICI channel, considering the overall uncertainty emerging from the various sources (e.g., sensors, absorption model, surface emissivity, geolocation, among others).

In addition, to characterize the error structure of the three spatially and temporally collocated measuring systems (i.e MWI or ICI, RS and NWP) a Multi-source Correlative Methodology (MCM) analysis has been implemented. It provides the estimate of the (i) error variance of the three sources of BT (RS, NWP and SAT) and the (ii) calibration parameters (bias and scaling) of two out of three sources (e.g. SAT and NWP if RS is considered as calibrated reference). Note that the estimation accuracy of the MCM tool strongly depends on the number of collocated triplets, so that only a high number of them can guarantee acceptable and accurate results.

The VICIRS tool has been tested on different settings, collecting simulated and real satellite observations collocated with either GRUAN or RHARM radiosondes. In particular, the tool was tested on the following datasets:

- simulated MWI/ICI Level 1B data and spatially/temporally collocated RHARM RS;
- real GMI observations and spatially/temporally collocated GRUAN profiles (for a 6-month period during 2023);
- real GMI observations and spatially/temporally collocated RHARM profiles (for a 3-month period during 2019 (September to November), over the area extending from 20° to 65° latitude North and from 15° to 40° longitude East).

Information is provided in output about the percentage of RS discarded from the calibration process. The RSs are grouped according to latitude range (polar, mid-latitude, sub-tropical, tropical) and, for each group the percentage of discarded is provided in output, as well as the percentage of data discarded due to clear-sky test, AMD test, and QC test. Table 7.1 summarizes the percentage of match-ups removed from all the aforementioned tests because of the failure of RS or TA tests. The largest percentage corresponds to cloud contamination for the examined GRUAN RS and on low nlev for RHARM RS.

Different combinations of query parameters have been tested to demonstrate the flexibility of the tool and give a preliminary idea of how the statistics vary by changing LF and NWP settings.

Thus, the proposed combinations are examples of how to handle the tool for investigating the impact of LF and NWP on statistics, but it should be noted that the results presented here are based on a small number of match-ups which is deemed not statistically representative enough to indicate the best combination.

Table 7.1 Statistics of the match-ups used in the MWI/ICI and GMI statistics

SAT-RS dataset	**#MU**	**#MU removed by RS tests(% of #MU removed)**	**#MU removed by clear-sky test on TA (% of #MU removed)**	**#MU useful(% of #MU useful)**	**Notes**
MWI/ICI-RHARM filtered on nlev=>15	92	63 (68.5%)	19(20.7%)	10(10.8%)	from 2007-09-12 08:00 to 2007-09-12 12:00 and from 2008-02-23 08:00 to 2008-02-23 10:00
GMI-RHARM	3156	3074(97.4%)	21(0,7%)	61(1.9%)	3 months of 2019 (Sep, Oct, Nov)
GMI-GRUAN	1139	1050 (92.2%)	21(1.8%)	68(6.0%)	6 months of 2023 (Jan, Feb, Mar, Apr, Sep, Oct)

Generally, MWI statistics show high BIAS and SD values for less opaque channels at 18.7 GHz(H) and 23.8 GHz(H), for which smaller values are related to TA built with no LF filter applied. For some more opaque MWI channels (e.g., 183.31 GHz band and 165.050.75 GHz), the statistics decrease when considering only TA with LF=0. High values of BIAS and SD for ICI are found for 183.31 and 448.01 GHz channels, which lowers when NWP information is supplied for simulating BT from RS.
The highest biases for GMI are found for less opaque channels (e.g., 10.65 and 18.7 GHz). These are likely affected by the unavailability of LF information for GMI FOV, which is supplied with the surface type (sea or land) and possibly causing a crude approximation of LF percentage for each GMI channel and for each TA, impacting negatively the statistics of window channels.
Although the tool has been demonstrated end-to-end, the resulting statistics should be taken with caution, as it is important to consider that:

- in MWI/ICI L1B comparison with RHARM RS:
    - no filter has been applied to RHARM-RS top pressure value (*Pmin*) and the minimum number levels (*nlev*) has been lowered to 15, in order to collect a higher number of match-ups. Without these assumptions, only one match-up would have been available for the MWI/ICI L1B-RHARM statistics;
    - LF of simulated MWI L1B observations is incorrect for MWI groups 1, 2, 5, 6, 7, 8 ;
- in the GMI comparison with GRUAN and RHARM RS
    - LF information is not available for GMI-FOV, causing inaccurate estimates of the simulated BT from RS especially for sites near coastline;
    - different TAs correspond to GMI S1 and S2, which may lead to inaccurate cloud detection as the cloud tests assume the same scene is observed.

In summary, it has been shown that VICIRS is a flexible and user-friendly tool that will be valuable for the calibration/validation of MWI/ICI observations against radiosoundings, characterizing the uncertainty propagation from the data sources (radiometric observations, GRUAN and RHARM RS datasets) throughout the collocation and comparison chain. In the future, the VICIRS tool may be adapted to operate with other radiometers and radiosonde archives.

References

Barlakas, V., A.J. Geer and P. Eriksson: Introducing hydrometeor orientation into all-sky microwave and submillimeter assimilation. Atmos. Meas. Tech., 14, 3427–3447, https://doi.org/10.5194/amt-14-3427-2021, 2021.

Bodeker, G.E., S. Bojinski, D. Cimini, R.J. Dirksen, M. Haeffelin, J.W. Hannigan, D. Hurst, F. Madonna, M. Maturilli, A.C. Mikalsen, R. Philipona, T. Reale, D.J. Seidel, D.G.H. Tan, P.W. Thorne, H. Vömel, and J. Wang: Reference upper-air observations for climate: From concept to reality. Bull. Amer. Meteor. Soc., doi: 10.1175/BAMS-D-14-00072.1, 2015.

Bobryshev, O., S.A. Buehler, V.O. John, M. Brath, and H. Brogniez: Is There Really a Closure Gap Between 183.31-GHz Satellite Passive Microwave and In Situ Radiosonde Water Vapor Measurements?. IEEE Transactions on Geoscience and Remote Sensing,vol. 56, no. 5, pp. 2904-2910, doi: 10.1109/TGRS.2017.2786548, 2018.

Buehler, S. A., M. Kuvatov, V.O. John, U. Leiterer and H. Dier: Comparison of Microwave Satellite Humidity Data and Radiosonde Profiles: A Case Study. J. Geophys. Res., 109, D13103, https://doi.org/10.1029/2004JD004605, 2004.

Buehler, S. A., M. Kuvatov, T.R. Sreerekha, V.O. John, B. Rydberg, P. Eriksson, and J. Notholt: A cloud filtering method for microwave upper tropospheric humidity measurements, Atmos. Chem. Phys., 7, 5531–5542, https://doi.org/10.5194/acp-7-5531-2007, 2007.

Buehler, S. A., E. Defer, F. Evans, S. Eliasson, J. Mendrok, P. Eriksson, C. Lee, C. Jiménez, C. Prigent, S. Crewell, Y. Kasai, R. Bennartz, and A. J. Gasiewski: Observing ice clouds in the submillimeter spectral range: the CloudIce mission proposal for ESA's Earth Explorer 8, Atmos. Meas. Tech., 5, 1529–1549, https://doi.org/10.5194/amt-5-1529-2012, 2012.

Calbet, X., R. Kivi, S. Tjemkes, F. Montagner, and R. Stuhlmann: Matching radiative transfer models and radiosonde data from the EPS/Metop Sodankylä campaign to IASI measurements. Atmos. Meas. Tech., 4, 1177–1189, https://doi.org/10.5194/amt-4-1177-2011, 2011.

Calbet, X.: Assessment of adequate quality and collocation of reference measurements with space-borne hyperspectral infrared instruments to validate retrievals of temperature and water vapour. Atmos. Meas. Tech., 9, 1–8, https://doi.org/10.5194/amt-9-1-2016, 2016.

Calbet, X., N. Peinado-Galan, P. Rípodas, T. Trent, R. Dirksen, and M. Sommer: Consistency between GRUAN sondes, LBLRTM and IASI. Atmos. Meas. Tech., 10, 2323–2335, https://doi.org/10.5194/amt-10-2323-2017, 2017.

Clain, G., H. Brogniez, V.H. Payne, V.O. John, and M. Luo: An Assessment of SAPHIR Calibration Using Quality Tropical Soundings, Journal of Atmospheric and Oceanic Technology, 32(1), 61-78. Retrieved Jan 18, 2023, from https://journals.ametsoc.org/view/journals/atot/32/1/jtech-d-14-00054_1.xml, 2015.

Carminati, F., S. Migliorini, B. Ingleby, W. Bell, H. Lawrence, S. Newman, J. Hocking,, and A. Smith: Using reference radiosondes to characterize NWP model uncertainty for improved satellite calibration and validation. Atmos. Meas. Tech., 12, 83–106, https://doi.org/10.5194/amt-12-83-2019, 2019.

Cherny, I.V., Mitnik, L. M., Mitnik, M. L., Uspensky A. B. and Streltsov A. M., "On-orbit calibration of the “Meteor-M” microwave imager/sounder," 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 2010, pp. 558-561, doi: 10.1109/IGARSS.2010.5651139.

Dinguirard, M., P. Slater: Calibration of Space-Multispectral Imaging Sensors: A Review. Remote Sens. Environ., 68(3), 194-205, https://doi.org/10.1016/S0034-4257(98)00111-4, 1999.

Dinnat, E., and Coauthors, 2023: PARMIO: A Reference Quality Model for Ocean Surface Emissivity and Backscatter from the Microwave to the Infrared. Bull. Amer. Meteor. Soc., 104, E742–E748, https://doi.org/10.1175/BAMS-D-23-0023.1.

Dirksen, R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, and H. Vömel: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. Atmos. Meas. Tech., 7, 4463-4490, doi:10.5194/amt-7-4463-2014, 2014.

Dirksen, R.J., G.E. Bodeker, P.W. Thorne, A. Merlone, T. Reale, J. Wang, D.F. Hurst, B.B. Demoz, T.D. Gardiner, B. Ingleby, M. Sommer, C. von Rohden, and T. Leblanc: Managing the transition from Vaisala RS92 to RS41 radiosondes within the Global Climate Observing System Reference Upper-Air Network (GRUAN): a progress report, Geosci. Instrum. Method. Data Syst., 9, 337–355, https://doi.org/10.5194/gi-9-337-2020, 2020.

Di Paola, F., E. Ricciardelli, D. Cimini, A. Cersosimo, A. Di Paola, D. Gallucci, S. Gentile, E. Geraldi, S. Larosa, S.T. Nilo, E. Ripepi, F. Romano, P. Sanò, M. Viggiano: MiRTaW: An Algorithm for Atmospheric Temperature and Water Vapor Profile Estimation from ATMS Measurements Using a Random Forests Technique. Remote Sens., 10, 1398, doi:10.3390/rs10091398, 2018.

Durre, Imke; Yin, Xungang; Vose, Russell S.; Applequist, Scott; Arnfield, Jeff; Korzeniewski, Bryant; Hundermark, Bruce. (2016) Integrated Global Radiosonde Archive (IGRA), Version 2. [indicate subset used]. NOAA National Centers for Environmental Information. DOI:10.7289/V5X63K0Q

Evans, K. F., Wang, J. R., Racette, P. E., Heymsfield, G., and Li, L.: Ice Cloud Retrievals and Analysis with Data from the Conical Scanning Submillimeter Imaging Radiometer and the Cloud Radar System during CRYSTAL-FACE, J. Appl. Meteorol., 44, 839–859, 2005.

Gallucci, D., Cimini, D., Turner, E., Fox, S., Rosenkranz, P. W., Tretyakov, M. Y., Mattioli, V., Larosa, S., and Romano, F.: Uncertainty of simulated brightness temperature due to sensitivity to atmospheric gas spectroscopic parameters, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-3160, 2024

Gong, J. and Wu, D. L.: Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements, Atmos. Chem. Phys., 17, 2741–2757, https://doi.org/10.5194/acp-17-2741-2017, 2017.

Hocking, J., R. Saunders, A. Geer, and J. Vidot: RTTOV v13 Users Guide. Tech. Rep. NWPSAF-MO-UD-046, EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWPSAF), 2022b.

Hong, G., Heygster, G., Miao, J., and Kunzi, K. (2005), Detection of tropical deep convective clouds from AMSU-B water vapor channels measurements, J. Geophys. Res., 110, D05205, doi:10.1029/2004JD004949.

He, Q., Li, J., Wang, Z., & Zhang, L. (2022). Comparative study of the 60 GHz and 118 GHz oxygen absorption bands for sounding sea surface barometric pressure. Remote Sensing, 14(9), 2260.

Immler, F.J. J. Dykema, T. Gardiner, D.N. Whiteman, P.W. Thorne, and H. Vömel: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products. Atmospheric Measurement Techniques, 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

JCGM: Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), Tech. Rep. JCGM 100: 2008, International Bureau of Weights and Measures (BIPM), https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf, 2008.

JCGM: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM3), Tech. Rep. JCGM 200: 2012, International Bureau of Weights and Measures (BIPM), https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf, 2012.

John V. O. and S. A. Buehler, “The impact of ozone lines on AMSU-B radiances,” Geophys. Res. Lett., vol. 31, p. L21 108, Nov. 2004.

John, V.O. and S.A. Buehler: Comparison of microwave satellite humidity data and radiosonde profiles: A survey of European stations, Atmos. Chem. Phys., 5, 1843–1853, https://doi.org/10.5194/acp-5-1843-2005, 2005.

Kerola, D.X.: Calibration of Special Sensor Microwave Imager/Sounder (SSMIS) upper air brightness temperature measurements using a comprehensive radiative transfer model, Radio Sci., 41, RS4001, https://doi.org/10.1029/2005RS003329, 2006.

Kilic L., C. Prigent1,2, C. Jimenez, E. Turner, J. Hocking, S. English, T. Meissner, E. Dinnat, Development of the SURface Fast Emissivity Model for Ocean (SURFEM-Ocean) based on the PARMIO2 radiative transfer mode, submitted to JGR Oceans, 2023.

Liu Q, Weng F, English SJ. 2011. An improved fast microwave water emissivity model. IEEE Trans. Geosci. Remote Sens. 49: 1238–1250, doi: 10.1109/TGRS.2010.2064779.

Madonna, F., R. Kivi, J.-C. Dupont, B. Ingleby, M. Fujiwara, G. Romanens, M. Hernandez, X. Calbet, M. Rosoldi, A. Giunta, T. Karppinen, M. Iwabuchi, S. Hoshino, C. von Rohden, and P.W. Thorne: Use of automatic radiosonde launchers to measure temperature and humidity profiles from the GRUAN perspective, Atmos. Meas. Tech., 13, 3621–3649, https://doi.org/10.5194/amt-13-3621-2020, 2020.

Madonna, F., E. Tramutola, S. SY, F. Serva, M. Proto, M. Rosoldi et al.: The new Radiosounding HARMonization (RHARM) data set of homogenized radiosounding temperature, humidity, and wind profiles with uncertainties. J. Geoph. Res.: Atmosph., 127, e2021JD035220. https://doi.org/10.1029/2021JD035220, 2022.

McColl, K.A., J. Vogelzang, A.G. Konings, D. Entekhabi, M. Piles, A. Stoffelen: Extended triple collocation: Estimating Errors and Correlation Coefficients with respect to an Unknown Target. Geophys. Res. Lett. 41, 6229–6236, 2014.

Mo, T.: Postlaunch Calibration of the NOAA-18 Advanced Microwave Sounding Unit-A, IEEE T. Geosci. Remote, 45, 1928–1937, doi:10.1109/TGRS.2007.897451, 2007.

Moradi, I., S.A. Buehler, V.O. John, and S. Eliasson: Comparing upper tropospheric humidity data from microwave satellite instruments and tropical radiosondes, J. Geophys. Res., 115, D24310, doi:10.1029/2010JD013962, 2010.

Moradi, I., B. Soden, R. Ferraro, P. Arkin and H. Vömel: Assessing the quality of humidity measurements from global operational radiosonde sensors, J. Geophys. Res.-Atmos., 118, 8040– 8053, https://doi.org/10.1002/jgrd.50589, 2013a.

Moradi, I, S. A. Buehler, V. O. John, A. Reale and R. R. Ferraro: Evaluating Instrumental Inhomogeneities in Global Radiosonde Upper Tropospheric Humidity Data Using Microwave Satellite Data, IEEE Trans. Geosci. Rem. Sens., vol. 51, no. 6, pp. 3615-3624, June 2013, doi: 10.1109/TGRS.2012.2220551, 2013b.

Moradi, I., R. Ferraro, P. Eriksson, and F. Weng.: Intercalibration and Validation of Observations From ATMS and SAPHIR Microwave Sounders, IEEE T. Geosci. Rem. Sens., 53, 5915–5925, https://doi.org/10.1109/TGRS.2015.2427165, 2015.

Prigent, C., F. Aires, D. Wang, S. Fox, and C. Harlow: Sea-surface emissivity parametrization from microwaves to millimetre waves. Q.J.R. Meteorol. Soc., 143: 596-605. https://doi.org/10.1002/qj.2953, 2017.

Reale, T., B. Sun, F. H. Tilley, and M. Pettey, 2012: The NOAA Products Validation System (NPROVS). J. Atmos. Oceanic Technol., 29, 629–645, https://doi.org/10.1175/JTECH-D-11-00072.1.

Santer, R., J.L. Deuze, C. Devaux, E. Vermote: SPOT calibration at the La Crau test site (France). Remote Sens. Environ., 41(2-3), 227-237, https://doi.org/10.1016/0034-4257(92)90080-4, 1992.

Sapiano, M., W. Berg, D. McKague, and C. Kummerow.: Toward an Intercalibrated Fundamental Climate Data Record of the SSM/I Sensors, IEEE T. Geosci. Remote, 51, 1492–1503, https://doi.org/10.1109/TGRS.2012.2206601, 2013.

Saunders, R., T. Blackmore, B. Candy, P. Francis, and T. Hewison: Monitoring Satellite Radiance Biases Using NWP Models, IEEE T. Geosci. Rem. Sens., 51, 1124–1138, https://doi.org/10.1109/TGRS.2012.2229283, 2013.

Seidel, D. J., B. Sun, M. Pettey, and A. Reale: Global radiosonde balloon drift statistics, J. Geophys. Res., 116, D07102, doi:10.1029/2010JD014891, 2011.

Slater, P.N., S.F. Biggar, R.G. Holm, R.D. Jackson, Y. Mao, M.S. Moran, J.M. Palmer, B. Yuan: Reflectance and radiance based methods for the in-flight absolute calibration of multispectral sensors. Remote Sens. Environ., 22, 11–37, 1987.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res., 103, 7755–7766, doi:10.1029/97JC03180, 1998.

Sun, B., A. Reale, D.J. Seidel, and D.C. Hunt: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, J. Geophys. Res., 115, D23104, doi:10.1029/2010JD014457, 2010.

Von Rohden, C., M. Sommer, T. Naebert, V. Motuz, and R.J. Dirksen: Laboratory characterisation of the radiation temperature error of radiosondes and its application to the GRUAN data processing for the Vaisala RS41, Atmos. Meas. Tech., 15, 383–405, https://doi.org/10.5194/amt-15-383-2022, 2022.

Von Rohden, C., M. Sommer, T. Naebert, V. Motuz, and R.J. Dirksen: Asset package related to AMT article “Laboratory characterisation of the radiation temperature error of radiosondes and its application to the GRUAN data processing for the Vaisala RS41”, GRUAN Lead Centre [data set], https://doi.org/10.5676/GRUAN/dpkg-2021-1, 2021.

Wang, D., C. Prigent, L. Kilic, S. Fox, C. Harlow, C. Jimenez, F. Aires, C. Grassotti, and F. Karbou: Surface Emissivity at Microwaves to Millimeter Waves over Polar Regions: Parameterization and Evaluation with Aircraft Experiments, Journal of Atmospheric and Oceanic Technology, 34(5), 1039-1059, 2017.

Wilheit, T.: Comparing Calibrations of Similar Conically Scanning Window-Channel Microwave Radiometers, IEEE T. Geosci. Remote, 51, 1453–1464, https://doi.org/10.1109/TGRS.2012.2207122, 00006, 2013.

Wu, Z., J. Li, Z. Qin: Development and Evaluation of a New Method for AMSU-A Cloud Detection over Land. Remote Sens. 13, 3646. https://doi.org/10.3390/rs13183646, 2021.

Yaping, Z., L. Jianwen, C. Zhoujie: Detection of Deep Convective Clouds Using AMSU-B and GOES-9 Data, China-Japan Joint Microwave Conference, pp. 278-281, doi: 10.1109/CJMW.2008.4772425, 2008.

Zhang, J., Chen, H., Li, Z., Fan, X., Peng, L., Yu, Y., and Cribb, M. (2010), Analysis of cloud layer structure in Shouxian, China using RS92 radiosonde aided by 95 GHz cloud radar, J. Geophys. Res., 115, D00K30, doi:10.1029/2010JD014030.

Zhao, T., Yu, P. Shi, J. et al. Global spatiotemporally continuous MODIS land surface temperature dataset. Sci Data 9, 143 (2022). https://doi.org/10.1038/s41597-022-01214-8

Appendix A

Table A.1 RSfilename_OOOOO_check.nc: list of variables

variable name	definition	type	dim	unit
Sonde_type	Sonde type: 1=GRUAN, 2=RHARM	NC_SHORT	1	
RS_filename	name of RS-file stored in /data_in/GRUAN (Sonde_type=1) or /data_in/RHARM (Sonde_type=2)	NC_STRING	1	
RS_launchtime	sonde launch time: seconds since 2001-01-01 00:00:00	NC_SHORT	1	seconds
SAT_overpass_date	date of SAT overpass “yyyy-mm-dd”	NC_STRING	1	
SAT_overpass_hhmmss	time of SAT overpass “hh:mm:ss.ddd”	NC_STRING	1	
RS_launch_latitude	latitude North of RS launch site	NC_DOUBLE	1	degree (°)
RS_launch_longitude	longitude East of RS launch site	NC_DOUBLE	1	degree (°)
Sat-RS_time_difference	difference between SAT-overpass time and sonde launch time	NC_SHORT	1	seconds
TA_radius	dimension of TA	NC_DOUBLE	1	m
AMD	Air Mass Displacement	NC_DOUBLE	1	m
RS_lev	number of RS pressure levels	NC_SHORT	1	
Pmin	top pressure value in hPa	NC_DOUBLE	1	hPa
Psurf	surface pressure value in hPa	NC_DOUBLE	1	hPa
t2m	temperature at 2m, in K	NC_DOUBLE	1	hPa
rh2m	RH at 2 m	NC_DOUBLE	1	%
zsurf	height of station above sea level	NC_DOUBLE	1	m
v10m	10m eastward wind component in \"m s-1\""	NC_DOUBLE	1	m/s
u10m	10m northward wind component in \"m s-1\""	NC_DOUBLE	1	m/s
low_cloud	number of RS levels low-cloud contaminated	NC_DOUBLE	1	
middle_cloud	number of RS levels middle-cloud contaminated	NC_DOUBLE	1	
high_cloud	number of RS levels high-cloud contaminated	NC_DOUBLE	1	
flagall	flag=1 RS useful for calibration, flag=0 RS no useful for calibration	NC_SHORT	1	
flagRS	flag=1 RS of good quality, flag=0 RS of no good quality	NC_SHORT	1	
flagc	flag=1 RS in clear sky, flag=0 RS in cloudy sky	NC_SHORT	1	
flagw	flag=1 AMD test passed, flag=0 AMD test not passed	NC_SHORT	1	

Table A.2 ***TAOOOOO_RScodeYYYYMMDDHHMM_lfRLF_check.nc***: list of variables

ROOT				
variable name	**definition**	**type**	**dim**	**unit**
TA_filename	name of circular-TA file extracted from SAT orbit file from **pyvicirs.ta_creator** and stored in /data_in/TA_data	NC_STRING	1	
Sonde_type	RS archive: 1 for GRUAN 2 for RHARM	NC_SHORT	1	
RS_filename	name of RS-file stored in /data_in/GRUAN (Sonde_type=1) or /data_in/RHARM (Sonde_type=2)	NC_STRING	1	
SAT_overpass_date	date of SAT overpass “yyyy-mm-dd”	NC_STRING	1	
SAT_overpass_hhmmss	time of SAT overpass “hh:mm:ss.ddd”	NC_STRING	1	
RS_launch_latitude	latitude North of RS launch site	NC_DOUBLE	1	degree (°)

RS_launch_longitude	longitude East of RS launch site	NC_DOUBLE	1	degree (°)
Sat-RS_time_difference	difference between SAT-overpass time and sonde launch time	NC_SHORT	1	seconds
TA_radius	dimension of TA	NC_DOUBLE	1	km
MWI GROUP				
land_frac_RS	LF percentage determined for RS for each MWI channel	NC_DOUBLE	26	
MWI_azimuth_angle	azimuth satellite angles corresponding to the MWI FOV closest to the sonde launch site	NC_DOUBLE	26	degree (°)
MWI_zenith_angle	zenith satellite angles corresponding to the MWI FOV closest to the sonde launch site	NC_DOUBLE	26	degree (°)
nFOVs_TA	number of MWI FOVs include in each TA	NC_SHORT	(26, numTA)	
LAND_FRAC_TA	LF percentage corresponding to each MWI channel for each TA	NC_DOUBLE	(26, numTA)	
teston89	percentage of FOVs declared cloudy by the two 89-GHz cloudy tests for each TA	NC_DOUBLE	(2, numTA)	
teston165	percentage of FOVs declared cloudy by the 165-GHz cloudy test for each TA	NC_DOUBLE	(1, numTA)	
teston183	percentage of FOVs declared cloudy by the 4 183.31-GHz cloudy tests for each TA	NC_DOUBLE	(4, numTA)	
max_cld	maximum percentage of cloudy FOVs among the results from the 3 sets of cloudy tests	NC_DOUBLE	numTA	
NEDT_MWI	NEDT for the 26 MWI channels	NC_DOUBLE	26	K
BT_TA	TA BT determined for each TA and for each MWI channel ([AD-10] subsection 4.2)	NC_DOUBLE	(26, numTA)	K
SD_TA	SD determined for each TA and for each MWI channel	NC_DOUBLE	(26, numTA)	K
HOMOGENEOUS_MWI	Index of homogeneity obtained for each MWI channel and for each TA by comparing SD and NEDT (Buehler et al. 2004)	NC_SHORT	(26, numTA)	
u_obs_MWI	uncertainty related to observations	NC_DOUBLE	(26, numTA)	K
u_col_MWI	uncertainty related to collocation	NC_DOUBLE	26, numTA)	K
BT_RS	BT simulated from RS for NWP_opt=0	NC_DOUBLE	(26, numTA)	K
uBT_RS	uncertainty related to BT_RS that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(26, numTA)	K
u_sim_MWI	uncertainty BT_RS that accounts for uBT_RS, absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(26, numTA)	K
TA_RS_MWI	difference between BT_TA and BT_RS determined for NWP_opt=0	NC_DOUBLE	(26, numTA)	K
u_all_MWI	uncertainty related to TA_RS_MWI that includes all the independent sources of uncertainties	NC_DOUBLE	(26, numTA)	K
K_FACTOR	coverage factor determined from the relation TA_RS_MWI<K_FACTOR*u_all_MWI for NWP_opt=0	NC_DOUBLE	(26, numTA)	
BT_RS1	BT simulated from RS for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(26, numTA)	K
uBT_RS1	uncertainty related to BT_RS1 that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(26, numTA)	K

u_sim_MWI1	uncertainty BT_RS1 that accounts for uBT_RS1 , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(26, numTA)	K
TA_RS_MWI1	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(26, numTA)	K
u_all_MWI1	uncertainty related to TA_RS_MWI1 that includes all the independent sources of uncertainties	NC_DOUBLE	(26, numTA)	K
K_FACTOR1	coverage factor determined from the relation TA_RS_MWI1<K_FACTOR1*u_all_MWI1 for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(26, numTA)	
BT_RS2	BT simulated from RS for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(26, numTA)	K
uBT_RS2	uncertainty related to BT_RS2 that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(26, numTA)	K
u_sim_MWI2	uncertainty BT_RS2 that accounts for uBT_RS2, absorption-model uncertainties(uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(26, numTA)	K
TA_RS_MWI2	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(26, numTA)	K
u_all_MWI2	uncertainty related to TA_RS_MWI2 that includes all the independent sources of uncertainties	NC_DOUBLE	(26, numTA)	K
K_FACTOR2	coverage factor determined from the relation TA_RS_MWI2<K_FACTOR2*u_all_MWI2 for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(26, numTA)	
ICI GROUP				
land_frac_RS	LF percentage determined for RS for each ICI channel	NC_DOUBLE	13	
ICI_azimuth_angle	azimuth satellite angles corresponding to the ICI FOV closest to the sonde launch site	NC_DOUBLE	13	degree (°)
ICI_zenith_angle	zenith satellite angles corresponding to the ICI FOV closest to the sonde launch site	NC_DOUBLE	13	degree (°)
nFOVs_TA	number of ICI FOVs include in each TA	NC_SHORT	(13, numTA)	
LAND_FRAC_TA	LF percentage corresponding to each ICI channel for each TA	NC_DOUBLE	(13, numTA)	
teston183	percentage of FOVs declared cloudy by the 4 183.31-GHz cloudy tests for each TA	NC_DOUBLE	(4, numTA)	
teston664	percentage of FOVs declared cloudy by the 2 664-GHz cloudy tests for each TA	NC_DOUBLE	(4, numTA)	
max_cld	maximum percentage of cloudy FOVs among the results obtained from the 2 sets of cloudy tests	NC_DOUBLE	numTA	
NEDT_ICI	NEDT for the 13 ICI channels	NC_DOUBLE	13	K
BT_TA	BT determined for each TA and for each ICI frequency ([AD-10] subsection 4.2)	NC_DOUBLE	(13, numTA)	K
SD_TA	SD determined for each TA and for each ICI channel	NC_DOUBLE	(13, numTA)	K
HOMOGENEOUS_ICI	Index of homogeneity obtained for each ICI channel and for each TA by comparing SD and NEDT (Buehler et al. 2004, subsection 3.3)	NC_SHORT	(13, numTA)	
u_obs_ICI	uncertainty related to observations	NC_DOUBLE	(13, numTA)	K
u_col_ICI	uncertainty related to collocation	NC_DOUBLE	(13, numTA)	K
BT_RS	BT simulated from RS for NWP_opt=0	NC_DOUBLE	(13, numTA)	K

uBT_RS	uncertainty related to BT_RS that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)	K
u_sim_ICI	uncertainty BT_RS that accounts for uBT_RS , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)	K
TA_RS_ICI	difference between BT_TA and BT_RS determined for NWP_opt=0	NC_DOUBLE	(13, numTA)	K
u_all_ICI	uncertainty related to TA_RS_ICI that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)	K
K_FACTOR	coverage factor determined from the relation TA_RS_ICI<K_FACTOR*u_all_ICI for NWP_opt=0	NC_DOUBLE	(13, numTA)	K
BT_RS1	BT simulated from RS for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)	K
uBT_RS1	uncertainty related to BT_RS1 that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)	K
u_sim_ICI1	uncertainty BT_RS1 that accounts for uBT_RS1 , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)	K
TA_RS_ICI1	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)	K
u_all_ICI1	uncertainty related to TA_RS_ICI1 that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)	K
K_FACTOR1	coverage factor determined from the relation TA_RS_ICI1<K_FACTOR1*u_all_ICI1 for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)	
BT_RS2	BT simulated from RS for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)	K
uBT_RS2	uncertainty related to BT_RS2 that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)	K
u_sim_ICI2	uncertainty BT_RS2 that accounts for uBT_RS2 , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)	K
TA_RS_ICI2	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)	K
u_all_ICI2	uncertainty related to TA_RS_ICI2 that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)	K
K_FACTOR2	coverage factor determined from the relation TA_RS_ICI2<K_FACTOR2*u_all_ICI2 for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)	K

Table A.3 VICIRS_query_matchup output file: (H_)(MCM)*SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorthLonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt.nc*: list of variables

variable name	definition	type	dimension	unit
dim_MCM	Dimension, number of measuring system (when OT=2)	NC_SHORT	1	
dim_row	Dimension, number of match-ups found	NC_SHORT	1	
dim_chan	Dimension, number of SAT channels	NC_SHORT	1	
RSlatitude	RS site latitude	NC_DOUBLE	dim_row	degree(°)
RSlongitude	RS site longitude	NC_DOUBLE	dim_row	degree(°)
Pmin	minimum value of RS pressure for each match-up	NC_DOUBLE	dim_row	hPa
RS_Lev	number of RS level for each match-up	NC_SHORT	dim_row	
NSAMPLE	number of samples useful for statistics for each channel	NC_DOUBLE	dim_row	
TA_RS	BT_TA -BT_RS	NC_DOUBLE	(dim_chan, dim_row)	K
u_all	overall uncertainty related to TA_RS	NC_DOUBLE	(dim_chan, dim_row)	K
K_FACTOR	coverage factor related to TA_RS and u_all	NC_SHORT	(dim_chan, dim_row)	
BIAS_TA_RS	the mean value of TA_RS	NC_DOUBLE	dim_chan	K
SD_TA_RS	SD of the TA_RS	NC_DOUBLE	dim_chan	K
u_BIAS	uncertainty of BIAS	NC_DOUBLE	dim_chan	K
wBIAS	weighted BIAS of TA_RS	NC_DOUBLE	dim_chan	K
SDw_TA_RS	SD of TA_RS weighted on the inverse of squared overall uncertainty	NC_DOUBLE	dim_chan	K
u_wBIAS	uncertainty of BIASn	NC_DOUBLE	dim_chan	K
BT_TA	observed BT	NC_DOUBLE	(dim_chan,dim_row)	K
BT_RS	BT simulated from RS	NC_DOUBLE	(dim_chan, dim_row)	K
BT_NWP	BT simulated from NWP (when OT=2 in *query.ini*)	NC_DOUBLE	(dim_chan, dim_row)	K

Table A.4 *MCM_SAT_SondeArchive_startYYYYMMDDHHMM-endYYYYMMDDHHMM_LatSouthLatNorthLonEastLonWest_TemporaleDistance_TAtype_CloudyPercentage__LF_DL/NWPopt/Tskinopt.nc updated after MCM analysis*: list of variables add to Table A.4 when OT=2 in *query.ini*

variable name	definition	type	dimension
dim_MCM	Dimension, number of measuring system (when OT=2)	NC_SHORT	1
calibration_coeffients_a	calibration-coefficient **a** for SAT/NWP	NC_DOUBLE	(dim_chan, dimMCM)
error_std_calibration_coefficients_a	error SD for calibration-coefficient **a**	NC_DOUBLE	(dim_chan, dimMCM)
calibration_coeffients_b	calibration-coefficient **b** for SAT/NWP	NC_DOUBLE	(dim_chan, dimMCM)
error_std_calibration_coefficients_b	error SD for calibration-coefficient **b**	NC_DOUBLE	(dim_chan, dimMCM)
error_std	error in standard deviation for RS/SAT/NWP	NC_DOUBLE	(dim_chan, dimMCM)

Table A.5 ***TAOOOOO_RScodeYYYYMMDDHHMM_check.nc*** list of variables for all the steps of GMI/RS match-up processing

ROOT			
variable name	**definition**	**type**	**dimension**
TA_filename	name of circular-TA file extracted from GMI orbit file from **pyvicirs.ta_creator** and stored in /data_in/TA_data	NC_STRING	1
Sonde_type	RS archive: 1 for GRUAN 2 for RHARM	NC_SHORT	1
RS_filename	name of RS-file stored in /data_in/GRUAN (Sonde_type=1) or /data_in/RHARM (Sonde_type=2)	NC_STRING	1
SAT_overpass_date	date of SAT overpass “yyyy-mm-dd”	NC_STRING	1
SAT_overpass_hhmmss	time of SAT overpass “hh:mm:ss.ddd”	NC_STRING	1
RS_launch_latitude	Latitude North of RS launch site	NC_DOUBLE	1
RS_launch_longitude	Longitude East of RS launch site	NC_DOUBLE	1
Sat-RS_time_difference	Difference between SAT-overpass time and sonde launch time (in seconds)	NC_SHORT	1
TA_radius	Dimension of TA (in km)	NC_DOUBLE	1
GMI GROUP			
GMI_azimuth_angle	azimuth satellite angles corresponding to the GMI FOV closest to the sonde launch site	NC_DOUBLE	13
GMI_zenith_angle	zenith satellite angles corresponding to the GMI FOV closest to the sonde launch site	NC_DOUBLE	13
nFOVs_TA	number of GMI FOVs include in each TA	NC_SHORT	(13, numTA)
LAND_FRAC_TA	LF percentage corresponding to each GMI channel for each TA	NC_DOUBLE	(13, numTA)
teston89	percentage of FOVs declared cloudy by the two 89-GHz cloudy tests for each TA	NC_DOUBLE	(2, numTA)
teston165	percentage of FOVs declared cloudy by the 165-GHz cloudy test for each TA	NC_DOUBLE	(1, numTA)
teston183	percentage of FOVs declared cloudy by the 2 183.31-GHz cloudy tests for each TA	NC_DOUBLE	(2, numTA)
max_cld	maximum percentage of cloudy FOVs obtained from the 3 cloudy tests	NC_DOUBLE	numTA
NEDT_GMI	NEDT for the 26 MWI channels (in K)	NC_DOUBLE	13
BT_TA	TA BT determined for each TA and for each GMI channel ([AD-10] subsection 4.2) (in K)	NC_DOUBLE	(13, numTA)
SD_TA	SD determined for each TA and for each GMI channel	NC_DOUBLE	(13, numTA)

HOMOGENEOUS_GMI	Index of homogeneity obtained for each GMI channel and for each TA by comparing SD and NEDT (Buehler et al. 2004, subsection 3.3)	NC_SHORT	(13, numTA)
u_obs_GMI	uncertainty related to observations	NC_DOUBLE	(13, numTA)
u_col_GMI	uncertainty related to collocation	NC_DOUBLE	(13, numTA)
BT_RS	BT simulated from RS for NWP_opt=0	NC_DOUBLE	(13, numTA)
uBT_RS	uncertainty related to BT_RS that accounts for RS T, RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)
BT_NWP	BT simulated from NWP (NWP_opt>0)		
u_sim_GMI	uncertainty BT_RS that accounts for uBT_RS , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)
TA_RS_GMI	difference between BT_TA and BT_RS determined for NWP_opt=0	NC_DOUBLE	(13, numTA)
u_all_GMI	uncertainty related to TA_RS_MWI that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)
K_FACTOR	coverage factor determined from the relation for NWP_opt=0	NC_DOUBLE	(13, numTA)
BT_RS1	BT simulated from RS for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)
uBT_RS1	uncertainty related to BT_RS1 that accounts for RS T , RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)
u_sim_GMI1	uncertainty BT_RS1 that accounts for uBT_RS1 , absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)
TA_RS_GMI1	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)
u_all_GMI1	uncertainty related to TA_RS_MWI1 that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)
K_FACTOR1	coverage factor determined from the relation f.for NWP_opt=1, TSkin_opt=1	NC_DOUBLE	(13, numTA)
BT_RS2	BT simulated from RS for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)
uBT_RS2	uncertainty related to BT_RS2 that accounts for RS T, RH and P profiles uncertainties	NC_DOUBLE	(13, numTA)
u_sim_GMI2	uncertainty BT_RS2 that accounts for uBT_RS2, absorption-model uncertainties (uABS) and surface-emissivity (uEMIS) uncertainties	NC_DOUBLE	(13, numTA)
TA_RS_GMI2	difference between BT_TA and BT_RS1 determined for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)
u_all_GMI2	uncertainty related to TA_RS_GMI2 that includes all the independent sources of uncertainties	NC_DOUBLE	(13, numTA)
K_FACTOR2	coverage factor determined from the relation for NWP_opt=1, TSkin_opt=2	NC_DOUBLE	(13, numTA)

Appendix B

Derivation of error variances

We start from eq. (4.7.1) that describe three collocated measurements of a same quantity t, that we rewrite here for ease of understanding.

$$x_i = b_i + a_i\, t + \varepsilon_i \tag{B.1}$$

Where a_i is the calibration scaling, b_i is the calibration bias, t is unobserved truth which is common to all the measuring systems and ε_i is the measurement random error of system $i - th$ with the index $i = 1, 2, 3$. The derivation of eq.s (4.7.2), (4.7.3) and (4.7.4) stats from the calculation of the covariance between to measurements x_i and x_j:

$$C_{ij} = \langle \left(x_i - \mu_{x_i}\right)\left(x_j - \mu_{x_j}\right)\rangle \tag{B.2}$$

In which, $\mu_x = \langle x \rangle$, $\langle \cdot \rangle$ is the average operator, and using assumption 2 in the main text (i.e. $\langle \epsilon_i \rangle = 0$) we have:

$$\mu_{x_i} = \langle x_i \rangle = b_i + a_i\, \mu_t \tag{B.3}$$

Then using (B.1) and (B.3) into (B.2) we obtain:

$$C_{ij} = a_i a_j \langle (t - \mu_t)^2 \rangle + a_i \langle (t - \mu_t)\varepsilon_j \rangle + a_j \langle (t - \mu_t)\varepsilon_i \rangle + \langle \varepsilon_i \varepsilon_j \rangle \tag{B.4}$$

In (B.4) the terms $\sigma_t^2 = \langle (t - \mu_t)^2 \rangle$, $\langle (t - \mu_t)\varepsilon_j \rangle = \langle (t - \mu_t)\varepsilon_i \rangle = 0$ due to assumptions 2 and 3, while the term $e_{ij} = \langle \varepsilon_i \varepsilon_j \rangle$ is left, for the moment, explicitly different from zero. Thus, (B.4) reduces to:

$$C_{ij} = a_i a_j \sigma_t^2 + e_{ij} \tag{B.5}$$

Then, two equations can be obtained from (B.5): case $i = j$ and $i \neq j$. In the first case we have:

$$C_{ii} = \sigma_{x_i}^2 = a_i^2\, \sigma_t^2 + \sigma_{\varepsilon_i}^2 \tag{B.6}$$

Whereas for $i \neq j$, from (A.5) we can write:

$$\frac{(C_{ij}-e_{ij})(C_{ik}-e_{ik})}{(C_{jk}-e_{jk})} = \frac{(a_i a_j\ \sigma_t^2)(a_i a_k\ \sigma_t^2)}{(a_j a_k\ \sigma_t^2)} = a_i^2\ \sigma_t^2 \tag{B.7}$$

Where k expresses the third reference system so in practice for example $i = 1$; $j = 2$; and $k = 3$. Thus, using (B.7) into (B.6), (B.7) can be written as:

$$\sigma_{\varepsilon_1}^2 = \sigma_{x_1}^2 - \frac{(C_{13}-e_{13})}{(C_{23}-e_{23})}(C_{12} - e_{12}) \tag{B.8a}$$

$$\sigma_{\varepsilon_2}^2 = \sigma_{x_2}^2 - \frac{(C_{23}-e_{23})}{(C_{13}-e_{13})}(C_{12} - e_{12}) \tag{B.8b}$$

$$\sigma_{\varepsilon_3}^2 = \sigma_{x_3}^2 - (C_{13} - e_{13})\ (C_{23} - e_{23}) \left(\frac{1}{C_{12}-e_{12}}\right) \tag{B.8c}$$

If we considering $e_{13} = e_{23} = 0$ and $e_{12} \neq 0$ under the assumption that the spatial scale of the third measuring system is taken as reference scale of analysis (see the section of MCM representativeness error for more details), we obtain the equation system of MCM system which is typically used:

$$\sigma_{\varepsilon_1}^2 = \sigma_{x_1}^2 - \frac{C_{13}}{C_{23}}(C_{12} - e_{12}) \tag{B.9a}$$

$$\sigma_{\varepsilon_2}^2 = \sigma_{x_2}^2 - \frac{C_{23}}{C_{13}}(C_{12} - e_{12}) \tag{B.9b}$$

$$\sigma_{\varepsilon_3}^2 = \sigma_{x_3}^2 - C_{13}\ C_{23} \left(\frac{1}{C_{12}-e_{12}}\right) \tag{B.9c}$$

which coincides with eq.s: (4.7.2a)- (4.7.2c) in the main text.

Derivation of correlation coefficients

The derivation of correlation coefficients in eq.s (4.7.4) follow the typical definition of the correlation coefficient

$$\rho_{t,i} = \frac{\langle (x_i - \mu_{x_i})(t - \mu_t) \rangle}{\sqrt{\sigma_{xi}^2}\sqrt{\sigma_t^2}} \tag{B.10}$$

Now, using (B.3) and (B.1) into (B.10) it can be easily demonstrated that the term $\langle x_i - \mu_{x_i} \rangle = a_i \sigma_t$ producing:

$$\rho_{t,i} = \frac{a_i \sigma_t}{\sqrt{\sigma_{xi}^2}} = \frac{a_i \sigma_t}{\sqrt{a_i^2\ \sigma_t^2 + \sigma_{\varepsilon_i}^2}} \tag{B.11}$$

Then, applying the result of eq. (B.7), the final result is obtained:

$$\rho_{t,1} = \frac{1}{\sigma_{x1}}\sqrt{\frac{(C_{12}-e_{12})\,C_{13}}{C_{23}}} \tag{B.12a}$$

$$\rho_{t,2} = \frac{1}{\sigma_{x2}}\sqrt{\frac{(C_{12}-e_{12})\,C_{23}}{C_{13}}} \tag{B.12b}$$

$$\rho_{t,3} = \frac{1}{\sigma_{x3}}\sqrt{\frac{C_{23}C_{13}}{(C_{12}-e_{12})}} \tag{B.12c}$$

Where it has been assumed $e_{13} = e_{23} = 0$ and $e_{12} \neq 0$ as done previously.
From (B.12) we can obtain a relation between $\rho_{t,i}$ and the signal to noise ratio SNR_i of *i*-th measuring system:

$$SNR_i = \frac{\langle (x_i')^2 \rangle}{\langle (\varepsilon_i)^2 \rangle} = \frac{\sigma_{x_i}^2 - \sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_i}^2} = \frac{a_i^2 \sigma_t^2}{\sigma_{\varepsilon_i}^2} \tag{B.13}$$

And substituting $a_i^2 \sigma_t^2 = SNR_i\, \sigma_{\varepsilon_i}^2$ into (B.11) we obtain:

$$SNR_i = \frac{\rho_{t,i}^2}{1+\rho_{t,i}^2} \tag{B.14}$$

Derivation of calibration parameters

For the calibration scaling parameters, they can be derived from (B.5) assuming the knowledge of system 1 calibration parameters (a_1 and b_1). Under this condition we have:

$$C_{12} - e_{12} = a_1 a_2\, \sigma_t^2 \tag{B.15a}$$

$$C_{13} = a_1 a_3\, \sigma_t^2 \tag{B.15b}$$

$$C_{23} = a_2\, a_3\, \sigma_t^2 \tag{B.15c}$$

Equating (B.15b) and (B.15c) we have $\sigma_t^2 = \frac{C_{13}}{(a_1\, a_3)} = \frac{C_{23}}{(a_2\, a_3)}$ from which a_2 estimate is obtained:

$$\hat{a}_2 = \frac{C_{23}}{C_{13}} a_1 \tag{B.16}$$

while using (B.15a) into (B.15c) we obtain $a_3 = \frac{C_{23}}{(a_2\, \sigma_t^2)} = \frac{(C_{23}\, a_1)}{(C_{12}-e_{12})}$ and finally:

$$\hat{a}_3 = \frac{C_{23}}{(C_{12}-e_{12})} a_1 \tag{B.17}$$

On the other hand, the derivation of the calibrating bias can be obtained by eq. (B.3). From this equation, and bearing in mind that we assumed to know a_1 and b_1 (i.e. $\mu_t = \left(\frac{\mu_{x_1}-b_1}{a_1}\right)$), we can write:

$$\hat{b}_2 = \mu_{x_2} - \hat{a}_2\, \mu_t = \mu_{x_2} - \hat{a}_2 \left(\frac{\mu_{x_1}-b_1}{a_1}\right) = \mu_{x_2} - \frac{C_{23}}{C_{13}}\mu_{x_1} + \frac{C_{23}}{C_{13}} b_1 \qquad \text{(B.18)}$$

$$\hat{b}_3 = \mu_{x_3} - \hat{a}_3\mu_t = \mu_{x_3} - \hat{a}_3 \left(\frac{\mu_{x_1}-b_1}{a_1}\right) = \mu_{x_3} - \frac{C_{23}}{(C_{12}-e_{12})}\mu_{x_1} + \frac{C_{23}}{(C_{12}-e_{12})} b_1 \qquad \text{(B.19)}$$

Where in the last terms of (B.18)-(B.19) we used (B.16)-(B.17), respectively.
Obviously, when system 1 is perfectly calibrated, that is $a_1 = 0$ and b_1=0, eq.s (B.16)-(B.19) simplify accordingly.

[END OF D17:FINAL REPORT]