EUMETSAT Satellite Application Facility on Climate Monitoring

Visiting Scientist Report

The EUMETSAT Network of Satellite Application Facilities

CM SAF Climate Monitoring

QUASAR

QUality Assessment of SAtellite and Radiosonde data

CDOP-2 AVS Study 13_03

By
Noëlle A. SCOTT
ARA/ABC(t)/LMD
On behalf of the ARA/ABC(t)/LMD Team
Ecole Polytechnique, RD36
91128 Palaiseau Cedex France

Date: 18.12.2015

Table of Contents

List of Figures

1 Purpose of the study

This report summarises the results of the activity **QUASAR** (**QU**ality **A**ssessment of **SA**tellite and **R**adiosonde data) which was dedicated to support CM SAF and the GEWEX water vapor assessment.

The major objective of the CM SAF is the exploitation of satellite observations to derive information on key climate variables of the Earth system. The CM SAF focuses on the atmospheric part of the Essential Climate Variables defined within the framework of the Global Climate Observing System (GCOS). The CM SAF produces (among others) data sets of humidity and temperature from various satellite instruments using homogenised and recalibrated radiance data sets of high temporal stability, e.g., upper tropospheric humidity (UTH) from homogenised AMSU-B data and free tropospheric humidity (FTH) from homogenised MVIRI and SEVIRI observations. The FTH product has been successfully transferred from LMD to CM SAF in a Research and Operations activity between both parties during CDOP1.
The CM SAF products are generated in a sustained and operational environment. Prior to release the products are fully documented, validated and reviewed in an external review process. Focus is on the validation of the products and the proven quality relative to the product and user requirements. The FTH product utilises clear sky radiances (CSR). In previous studies the CSR quality has been evaluated (Brogniez et al., 2006 and Brogniez et al., 2009) against the Analysed RadioSoundings Archive (ARSA, http://ara.abct.lmd.polytechnique.fr/index.php?page=arsa) data showing good quality (bias and stability) of the CSR data. On the contrary the work of John and Buehler, 2005; Moradi et al., 2010; Kottayil et al., 2012 identified suspicious quality of the Integrated Global Radiosonde Archive (IGRA) archive and operational radiosonde humidity data in general by comparison against AMSU-B radiances. This has previously been shown by Soden and Lanzante (1996) using HIRS observations.
The GEWEX Data and Assessments Panel (GDAP) has initiated a water vapor assessment project (G-VAP) intended to both quantify the state of the art in water vapor products as well as to eventually select a product for use by GDAP in its production of globally consistent water and energy cycle products. The assessment will analyse UTH/FTH, TCWV and profile products. The CM SAF co-chairs the assessment.

At the core of the assessment is GEWEX's need to gain insight into a number of water vapor products now being constructed for climate applications. Because each product can have slightly different users and objectives, it was thought important to clearly describe the data set objectives in the introduction to the assessment in order to place each product in the proper perspective relative to climate needs. The assessment focuses on overall characteristics of participating satellite data records and reanalysis as determined from inter-comparisons and comparisons against in situ observations and ground-based products.

Therefore, the validation data base is central to the assessment. The ARSA, IGRA and homogenised IGRA data records are candidates for the validation data base.

This AS supports CM SAF DRRs and the GEWEX water vapor assessment and is the consequence of CM SAF's response to the CDOP2 proposal review (A-CM-06), presented to CDOP2 SG1.

2 Executive summary

In this AS the radiosonde data records from ARSA and homogenised IGRA have been inter-compared. The focus has been on *stability and bias*, to lead to the identification of stations appropriate for the validation of long-time series data sets within CM SAF and the GEWEX water vapor assessment.

The ARSA database was produced at and provided (N.A. Scott, 2009, private communication) by ARA/ABC(t)/LMD, Paris, France. ARSA is mainly based on radiosonde observations that have successfully completed extensive qualitative and quantitative tests: the required minimal information being to have measured points from surface up to 30 hPa for temperature profiles and from surface to 300 hPa for water vapour profiles. Moreover, in order to give a continuous description of the atmospheric state from the surface to the top of the atmosphere (~0.002hPa), these radiosonde observations have been extended above their highest measured point with ERA-Interim data (temperature, water vapour and ozone up to 0.1 hPa) and then with SciSat ACE FTS level2 data (from 0.1hPa to the top the atmosphere: 0.0026 hPa).
At LMD, the validation of ARSA, currently relies upon the study of statistics (bias, standard deviation) between simulated and observed satellite radiances: TOVS and ATOVS, as well as, in the more recent years, the MetOp A&B IASI, HIRS4 and MHS observations. The simulated data are generated by the 4A/OP radiative transfer model, fed with the ARSA profile that is the closest (in space and time) from the satellite observation. Due to the excellent stability of the IASI radiances and the accuracy of the 4A/OP model, the quality of the ARSA profiles may then be assessed. Also, it has to be noticed that the spectral resolution (0.50 cm-1, apodized) and continuity (645 to 2760 cm-1) of IASI spectra helps doing these validations in a coherent way.
The comparison to satellite data records is also done for AMSU-B, MHS (microwave spectral domain) and HIRS (thermal infrared) radiance space.

The IGRA data base is described in (Durre et al., 2006; Durre and Yin, 2008) and has been homogenised as described in Dai et al. (2011).

During this AS the following specific tasks have been performed:

1. Gather required data (ARSA, IGRA, homogenised IGRA and AMSU-B, MHS, HIRS) and radiative transfer model (4A/OP).
2. Inter-compare the in-situ data records with focus on stability and bias in the UT region.
3. Simulate satellite radiances for sub-periods such as 2 years in each decade between 1980 and 2010. In recent years IASI can serve as reference. The HIRS and MHS instrument on board the same satellite will also be used in order to ease the interpretation.
4. Identify stations with unsuspicious quality (stability and bias) on basis of results from tasks 2 and 3 and recommend on utilisation for validation of regional (METEOSAT) and global data records at CM SAF and for G-VAP.

This report summarizes the work performed to answer each of these questions.

3 Organisation of the document

As stated above, several datasets are involved in this inter-comparison of radiosonde databases: ARSA and IGRA_Homogenized, as well as satellite data. An auxiliary dataset has been added: the Raw Radiosonde reports collocated with ARSA and the IGRA_Homogenized datasets.

This report is organized as follows:

We first describe the databases and models required to perform the four tasks mentioned above. Section 4 is for databases and models produced and maintained at LMD. Section 5 is dedicated to a brief description of the IGRA_Homogenized as delivered to LMD.
In Section 6 we describe methodology and the upstream activities to start the inter-comparison “ARSA versus IGRA” or “simulations of satellite observations (based on ARSA and the 4A/OP forward model) versus satellite observations”.
In Section 7 and 8, we describe all the datasets generated for this study and the deliverables (files, statistics, graphs).
In Section 9 we detail some points on the results.
Section 10 draws Conclusions to this study.
Section 11 gives some Recommendations.
Sections 12 gives overview of the Technical Memo, deliverables, …
Sections 13 and 14 give some references and milestones of this study.
Last two sections 15 and 16 give additional technical descriptions of the ARSA and IGRA_Homogenized databases.
Acknowledgements are in Section 17.

4 State of the art of the LMD databases and models involved in QUASAR

In this Section is given the information on databases and radiative transfer models developed and maintained at LMD which have been used as support to the current QUASAR Project.
Eventually, work within this contract has been based upon three datasets (see Sections 4.1 to 4.3) and a forward radiative transfer model (see Section 4.4).
In order to fully describe these datasets and situate them in the frame of this contract, we also give in Section 4.5, the method and results of the validation of the ARSA database.

4.1 ARSA

4.1.1 The origin of the ARSA database

Processing and validation of level1 and level2 satellite data require auxiliary datasets, a key one being radiosonde measurements. Eventually, radiosonde reports are critical for a wide range of applications as forward and inverse models validation, verification of satellite measurements, and any other application dealing with earth observations from vertical or limb sounders or imagers (radiometers, spectrometers, interferometers) on board polar or geostationary satellites, for operational or research purposes. In order to be fully useful for these applications, and mainly for the one involving forward radiative transfer simulations, the ARSA (Analyzed RadioSoundings Archive) database has been elaborated, starting from observations by worldwide distributed radiosonde stations and combining them with surface and other auxiliary observations.
The current ARSA database starts in January 1979, and is extended onwards, on a monthly basis. It is available upon request at LMD. Validation of the ARSA database is currently performed using numerous comparisons between simulated and observed brightness temperatures of satellite borne instruments, among many others is the IASI/MetOp hyperspectral infrared sounder.

Before starting a more detailed description of the ARSA data base, and to make clear the way to its elaboration and validation, it is important to recall that the main purpose of ARSA as defined by ARA/ABC(t)/LMD group is to serve as input to forward radiative transfer models to simulate satellite observations and study the sensitivity of individual sounding channels, or to be space-time collocated with them to help the validation of level2 products.

Following Sections 4.1.2 to 4.1.5 report about the work to collect and adjust radiosonde reports to make them suitable for the abovementioned applications.

Radiosonde reports are extracted from the ECMWF archive. They come from 1472 globally distributed stations. The ARSA data base covers the period 1979-onwards and is updated on a monthly basis.

The effort is comprehensive in the sense that radiosonde data, reanalysis products from the ECMWF ERA_Interim, and simulated and observed satellite data from IASI/Metop have been simultaneously quality controlled, merged, intercompared.

The work to produce and use ARSA started at LMD in the late 90's. The current 2.7 version was distributed in the 2005.

4.1.2 The elaboration of the ARSA database

In order to be fully useful for the abovementioned applications, and mainly for the one involving forward radiative transfer simulations, several tasks are required to answer the questions of quality and completeness of the ***radiosonde reports***. From the raw radiosonde archive of ECMWF to the end product, an entirely automatic multi step procedure has been applied to achieve the desired quality.

Our aim was to achieve a high level of completeness (spatial and vertical coverage) and quality. Obviously, the density and spatial representativeness of the data is related to the density of the RAOBS.

At the very beginning of the procedure, 22 millions of RS from 11,742 stations have been processed **(corresponding to** 480 million of measurements). These numbers are for the January 1979 to June 2014 period.

The ***first step*** is to develop and apply physically coherent quality control tests to detect/ eliminate gross errors: format problems, redundant RS and levels, unrealistic jumps, physically implausible values, temporal and vertical inconsistencies in temperature, dew point temperatures.

> After the step 1, we come to 5 million of RS accepted for subsequent steps from 10,583 stations (corresponding to 230 million of measurements).

In the ***second step***, our QC have to ensure that every RS report kept after the first step is also fully compatible with the forward radiative transfer simulations. Such requirements are that it has to lead to a relevant discretization in pressure. To achieve this, it is thus required that temperature measurements be available at least up to 30 hPa, that water vapor measurements be available at least up to 350 hPa, and that surface pressure be not smaller than a given value (currently: 850 hPa over land and 950 hPa over sea). Also, the TIGR climatological dataset (http://ara.abct.lmd.polytechnique.fr/) helps removing values that deviate by more than a given number of standard deviation from their respective air-mass (tropical, mid-lat, polar) mean values.

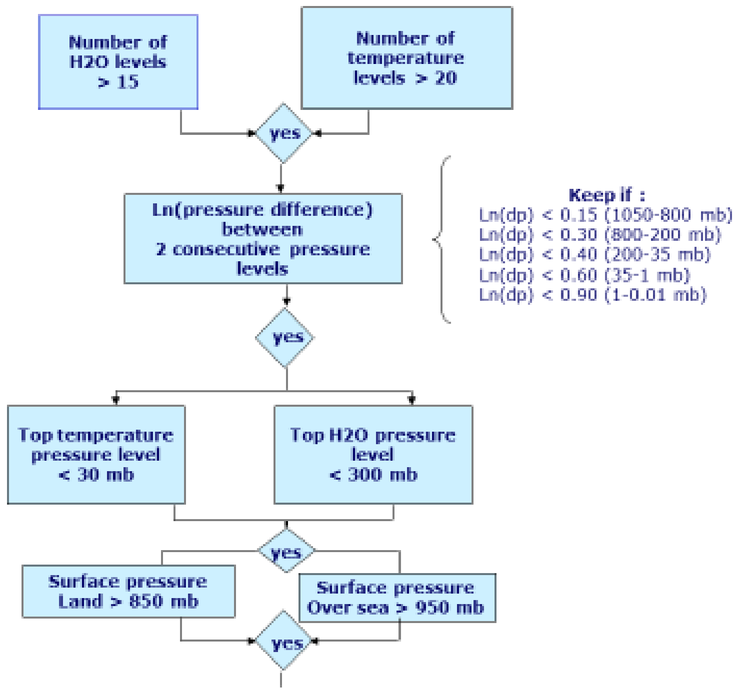


Figure 4-1: *ARSA Second step of QC: Constraints on the number and distribution of the RS pressure levels.*

In the ***third step***, whenever and wherever required information is missing, we combine existing radiosonde measurements with other reliable data sources in order to complete the description of the atmospheric state as high as the 0.0026 hPa pressure level.

Temperature and water vapor profiles are extrapolated by ERA_interim outputs between. 30 hPa and 0.1 hPa for temperature and between 300 hPa and 0.1 hPa for water vapor. Above 0.1 hPa, these same profiles are extrapolated up to 0.0026 hPa using a climatology of ACE/Scisat level2 products.

Ozone: since most of the radiosonde reports do not provide information on ozone, the current version of ARSA takes its ozone profiles from ERA_interim outputs, space and time collocated with the considered radiosonde station.

Surface temperature: when not available in the radiosonde report, the current version of ARSA takes its surface temperature value from the surface station archive of ECMWF. As for ozone, this is performed through a space and time collocation of the surface station archive of ECMWF with the considered radiosonde station.

In the ***fourth step***, temperature, water vapor, ozone profiles are interpolated on a multi-level pressure grid: a nominal 43-level pressure grid is used between sea level pressure and 0.0026hPa, or a “smaller than 43” pressure wherever necessary for radiosonde stations in altitude.

The resulting data set is the ARSA (Analyzed RadioSoundings Archive) database, regularly extended on a monthly basis. Validation of ARSA data is currently performed using numerous comparisons between 4A/OP simulated and observed brightness temperatures of IASI/Metop.

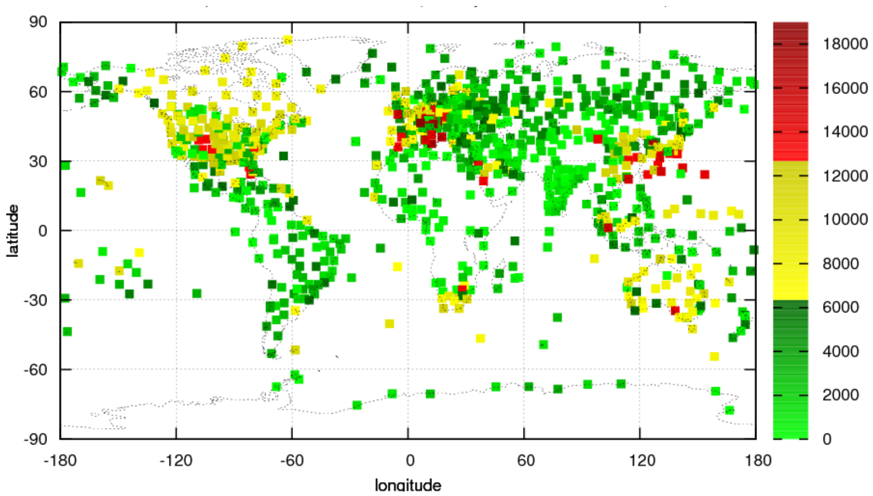


Figure 4-2*: Number and Location of the ARSA profiles: January 1979 to December 2013.*

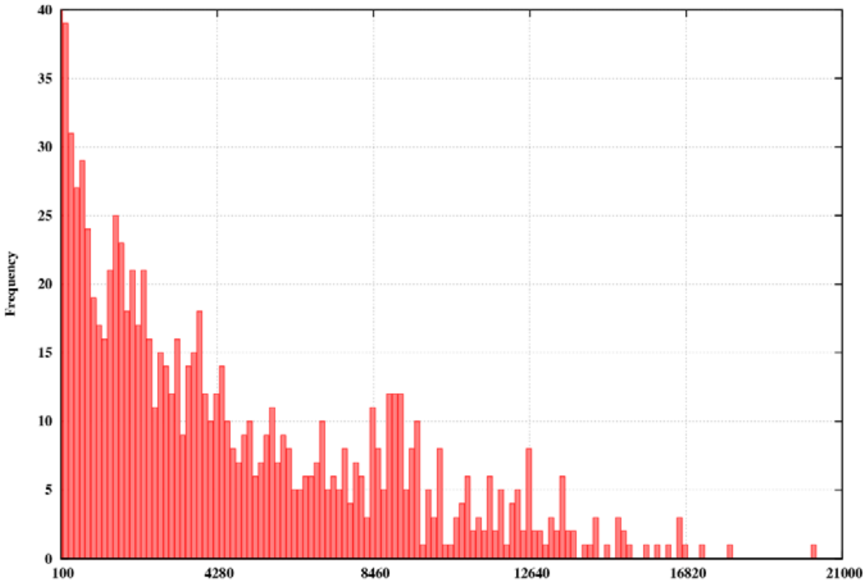


Figure 4-3: *Histogram of radiosonde reports kept during the January 1979 to December 2013 period of ARSA: x-axis: number of profiles kept in the ARSA data base; y-axis: number of stations associated to this number of profiles. Data from 1134 stations went through step 1 to 4 of the QC tests.*

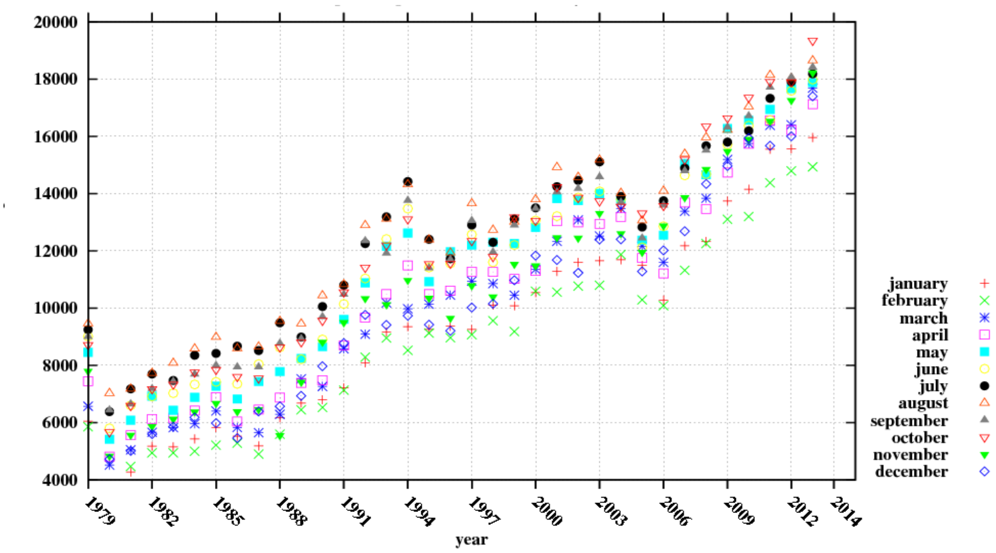


Figure 4-4. *Number of profiles per month kept in ARSA from January 1979 to December 2013.*

The overall increase of the number of measurements as shown on Figure 4.4 indirectly illustrates the fact that the quality, the vertical resolution and the vertical extent of radiosoundings improve significantly over time.

4.2 ARSA metadata

As already stated above, for our in-house LMD applications, the radiosonde reports we consider as “relevant” are the ones which describe "correctly" the whole atmospheric column. From the raw radiosonde measurements extracted from ECMWF up to the converged ARSA product, due to the “keep/reject“ quality control tests and interpolation, extrapolation processes, several steps are performed (cf Section 3.1) and intermediate files are generated

These files are of three types:

1. rejected raw radiosonde reports (selection based on the quality controls): they are archived on their nominal pressure grid (type 1);
2. kept (selection based on the quality controls) raw radiosonde reports: they are archived on their nominal pressure grid (type 2);
3. kept radiosonde reports that are then interpolated, completed, extrapolated: they are archived on the 43-level 4A/OP pressure grid (between surface and 0.0026hPa). Such reports constitute the ARSA database (type 3);

Besides these three files, we also generate and archive the following metadata:

- List of RS stations, their WMO code, …
- Monthly statistics on all the kept/rejected elements at the end of the full QC process Numbers are given in percentage of stations processed and measurements processed.

However, some comparisons are performed between ARSA and products of type 2 (ie “kept raw radiosonde reports”: see definition of type 2 above) to evaluate the impact of interpolation/extrapolations. This product of type 2 has also been used for the evaluation of IGRA.

4.3 The 4A/OP forward radiative transfer model

The 4A line-by-line model (Scott and Chédin, 1981) to calculate forward radiative transfer is an advanced version of the nominal line-by-line STRANSAC model (Scott, 1974). These forward models have become more and more accurate and efficient through the exploitation of new mathematical techniques, the availability of faster and faster computer systems, and, last but not least, the provision of better spectroscopic data (GEISA database). The 4A model has a long history of validation within the frame of the international radiative transfer community. Most of the validation results have been extensively discussed in a number of inter-comparison exercises and in particular during the ITRA (Inter-comparison of Transmittance and Radiance Algorithms) working groups - 1983, 1985, 1988, 1991 of the International Radiation Commission.(see e.g. Chédin et al., 1988) and during the ICRCCM (Inter-comparison of Radiation Codes in Climate Models) campaigns (see e.g. F. Luther et al. 1988). More recently, launch of hyperspectral sounders (AIRS on Aqua platform and IASI on

Metop platform) have led to more and more extensive validations, still within the frame of international campaigns or working groups.
4A/OP has been chosen by CNES as the official radiative transfer model for IASI level1 CAL/VAL and level1 operational processing.

The description of the analytic computation of the Jacobians has been given in the mid 90’s (F. Chéruy et al, J. Quant. Spectrosc. Radiat. Transfer, 1995).
So far, the computation of jacobians for temperature, absorbing gases, surface temperature and emissivity is optionally performed.

During these last years, the model underwent important transformations, in relation with our in house applications (aerosols retrievals from AIRS and IASI, processing of the ACE-Scisat instrument for the retrieval of CO_2 profiles, …), or in relation with applications within the frame of the CNES activities, more recently related to SWIR applications and in development for the UV/Vis domain.

4A is maintained at LMD which includes introducing the newly derived parameters for spectroscopy, for line-coupling or for continua, for aerosols and CFCs, ..., as soon as they have been validated. The current version is referred to as 4A-2012. It differs from previous versions (2006, 2009) as indicated in the table below:

Items	2012 Version	2009 Version	2006 Version
Spectroscopic database	Geisa 2011	Geisa 2009	Geisa 2003
H2O continuum	Clough et al., 1989	Clough et al., 1989	Clough et al., 1989
N2 continuum	Lafferty et al., 1996	Lafferty et al., 1996	Lafferty et al., 1996
O2 continuum	Thibault et al.,1997	Thibault et al.,1997	Thibault et al.,1997
Partition Function	Gamache et al., 2000	Gamache et al., 2000	Gamache et al., 2000
Line coupling	Niro et al., 2008 updated by Hartmann	Niro et al., 2004	Rodrigues et al., 1998
□correction of Voigt profiles	Perrin et al., 1989	Perrin et al., 1989	Perrin et al., 1989
pressure shift for H2O	Yes	yes	no
Spectral dependence of the surface emissivity	Yes	yes	no
Simulation of Scattering effects of aerosols	yes, based on DISORT K. Stamnes et al., 1988	yes, based on DISORT K. Stamnes et al., 1988	no
Implementation of the limb viewing geometry	yes	Yes	no
Number of nominal layers / Top of the atmosphere	43 / 0.0026 hPa	43 / 0.0026 hPa	40 0.05 hPa
Adjustment of gas concentration profiles in the Atlas computations	yes	Yes	N.R.
Spectral Domain	TIR+SWIR	TIR+SWIR	TIR
Spectral Domain	UV/VIS (in progress)		

4.4 Validation of the ARSA database based on the study of simulated versus observed IASI brightness temperatures: impact on water vapor and ozone profiles

4.4.1 Method for the validation of the ARSA database

The method, used for many years in the ARA group at LMD for the quality control of satellite level1 or level2 products relies on simulations of satellite observations based on the use the 4A/OP radiative transfer model and the ARSA data base. Such simulations have been performed for instruments operated by NASA and NOAA including TOVS, ATOVS, AIRS/Aqua,… as well as the instruments operated by EUMETSAT including IASI, HIRS4 and MHS observations in the more recent years.

The procedure goes through the study of statistics (bias, standard deviation) between simulated and observed radiances. We call such quantities: residuals. The multistep process is as follows:

- Identify spectral regions where an unexpected bias behaviour arises.
- Identify the absorbing gas source of this bias.
- Iteratively refine the stats by tuning the profile.

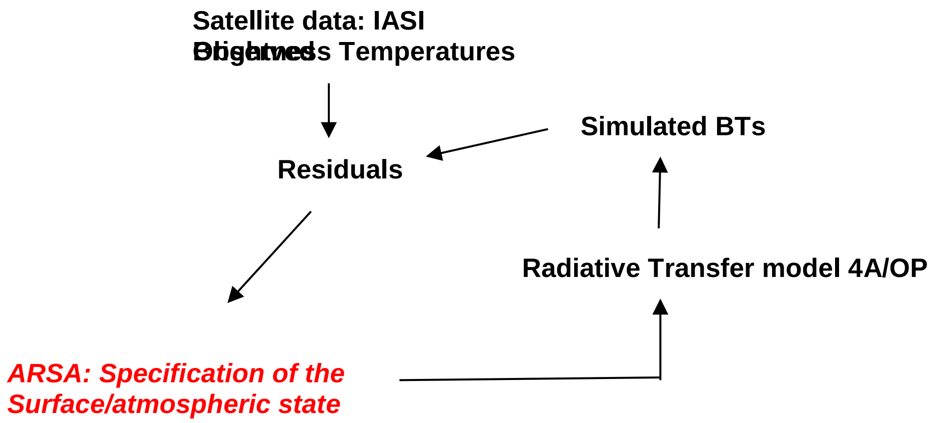


Figure 4-5. *The validation process at LMD combining IASI/MetOp data, ARSA and the Radiative transfer model 4A/OP.*

4.4.2 Satellite data used in the validation process

4.4.2.1 IASI on board MetOpA

IASI, the Infrared Atmospheric Sounding Interferometer is a key payload element of the METOP series of European meteorological polar-orbit satellites.
Developed by CNES in collaboration with EUMETSAT, the IASI instrument (Chalon et al., 2001; http://smsc.cnes.fr/IASI), onboard the MetOp-A polar platform, is a Fourier Transform Spectrometer that measures Earth-emitted infrared radiation. Launched in October 2006 and operational since July 2007, it provides 8461 spectral channels, between 15.5 µm (645 cm-1) and 3.63 µm (2755 cm-1) with

a spectral resolution of 0.50 cm−1 after apodisation, and a regular spectral sampling interval of 0.25 cm−1. MetOp-A crosses the Equator at 9:30 p.m., Local Time (LT), on its ascending node. IASI provides a near global coverage twice a day at a spatial resolution of 12 km at nadir.

The first flight model was launched in 2006 onboard the first European meteorological polar-orbiting satellites, METOP-A. The second instrument, mounted on the METOP-B satellite, was launched in September 2012. The third instrument will be mounted on the METOP-C satellite with launch scheduled for October-November 2016.

4.4.2.2 HIRS

4.4.2.2.1 HIRS3 on board the NOAA-series

The HIRS is a discrete stepping, line-scan instrument designed to measure scene radiance in 20 spectral bands to permit the calculation of the vertical temperature profile from the Earth's surface to about 40 km. Multispectral data from one visible channel (0.69 micrometers), seven shortwave channels (3.7 to 4.6 micrometers) and twelve longwave channels (6.5 to 15 micrometers) are obtained from a single telescope and a rotating filter wheel containing twenty individual filters.

An elliptical scan mirror provides cross-track scanning of 56 increments of 1.8 degrees. The mirror steps rapidly (<35 msec), then holds at each position while the 20 filter segments are sampled. This action takes place each 100 msec. The instantaneous FOV for each channel is approximately 1.4 degrees in the visible and shortwave IR and 1.3 degrees in the longwave IR band which, from an altitude of 833 kilometers, encompasses an area of 20.3 kilometers and 18.9 kilometers in diameter, respectively, at nadir on the Earth.

The above information is extracted from:
http://www.nsof.class.noaa.gov/data_available/tovs_atovs/index.htm

4.4.2.2.2 HIRS4 on board MetOpA

The HIRS/4 instrument measures the incident radiation primarily in the infrared region of the spectrum in 19 channels, including both longwave (6.5 to 15 µm) and shortwave (3.7 to 4.6 µm) regions, and it also has one channel in the visible (0.69 µm).
HIRS/4 is an across-track scanning system with a rotating mirror and a scan range of ±49.5° with respect to the nadir direction. The instantaneous field of view (IFOV) of each channel is approximately 0.69°, leading to a circular IFOV size close to 10.0 km at nadir for a nominal altitude of 833 ± 19 km. The major difference between HIRS/3 and HIRS/4 is that HIRS/3 has an IFOV size close to 20 km. Each scan line takes 6.4 s to complete. At the end of the scan line, the mirror rapidly returns to its home position (8 retrace steps of 100 ms each) and the scanning pattern is repeated.

There are 56 Earth view samples per scan for a swath width of ±1080.35 km (sampling time of 100.0 ms). The sampling angular interval is close to 31.42 milliradians (1.8°). The distance between two consecutive scans is approximately equal to 42.15 km.
IFOV type is: circular, 10.0 km at Nadir, 17.03 km (edge)-along track, 33.27 km (edge)-across track.
The above information is extracted from http://oiswww.eumetsat.org/WEBOPS/eps-pg/ATOVS-L1/ATOVSL1-PG-4ProdOverview.htm.

4.4.3 MHS on board MetOpA

The MHS is the follow-on instrument to the Advanced Microwave Sounding Unit-B (AMSU-B) which flew as a part of ATOVS on the NOAA-KLM satellite series. It is procured by EUMETSAT for the Metop and NOAA satellites.

MHS is a five-channel microwave radiometer, which complements the Advanced Microwave Sounding Unit-A (AMSU-A) channels.
(H1) channel 16: 89 GHz
(H2) channel 17: 157 GHz
(H3 and H4) channels 18 & 19: 183.311 +/- 1 and +/- 3 GHz
(H5) channel 20: 190.311 GHz

It is planned to derive from these frequencies humidity profiles and cloud liquid water content. Additionally, the instrument's sensitivity to large water droplets in precipitating clouds can provide a qualitative estimate of precipitation rates.

MHS is an across-track scanning system with a scan range of ±49.44° with respect to the nadir direction. The IFOV of each channel is approximately 19.2 milliradians (1.1°) leading to a circular instantaneous field of view size close to 15.88 km at nadir for a nominal altitude of 833 km. Each scan takes 2.667 seconds to complete.

The scan of the MHS instrument is synchronised with the AMSU-A scan, i.e. there are three scans of MHS for each scan of AMSU-A.

There are 90 Earth samples per scan and per channel for a swath width of ±1077.68 km (sampling time of 19.0 ms). The sampling angular interval is close to 19.39 milliradians (1.1111°), which is slightly larger than that of AMSU-B (1.1000°). The distance between two consecutive scans is approximately equal to 17.56 km.
IFOV type is: circular, 15.88 km at Nadir, 27.10 km (edge)-along track, 52.83 km (edge)-across track.
The above information is extracted from http://oiswww.eumetsat.org/WEBOPS/eps-pg/ATOVS-L1/ATOVSL1-PG-4ProdOverview.htm.

4.4.4 Validation of the ARSA database: impact on Water vapor profiles

With respect to the very first version of ARSA (hereafter referred to as v2.5), two modifications have been brought

- Extension of radiosonde water vapor profiles above 350 hPa using ERA_interim profiles (up to 0.1 hPa) instead of a previous extrapolation method has led to i) considerably reducing the standard deviation in the 6.3 micron spectral region of IASI MetOpA while ii) introducing a *negative bias*. This *negative bias* indicates, in the ERA_interim profiles, too high a quantity of water vapor in – roughly - the ~[160 to 350hPa] pressure range.
- Consequently, further iterative comparisons between simulated and observed IASI spectra have led to *empirically correct the ERA_Interim water vapor* profiles between ~[160 and 350 hPa.].

In summary, the modifications brought to the very first version of the ARSA profiles are as follows:

Water vapor:

- Above 350 hPa, ARSA is extended with ERA_Interim.
- In addition a ***linear correction*** on the ERA_Interim water vapor profile. After several iterative comparisons with observations, the double linear correction is as follows involving levels at 170 hPa, 270 hPa, 400 hPa.
  170 hPa : -   0 % ERA_Interim
  270 hPa : - 20 % ERA_Interim
  400 hPa : -   0 % ERA_Interim

Temperature:

- ARSA extended to 0.1 hPa by ERA_Interim starting at 37 hPa.

As shown on figure 4.6, the residuals (simulated-observed) IASI brightness temperatures obtained after improving the water vapour profile turn out to be improved both in bias (top) and standard deviation (bottom).

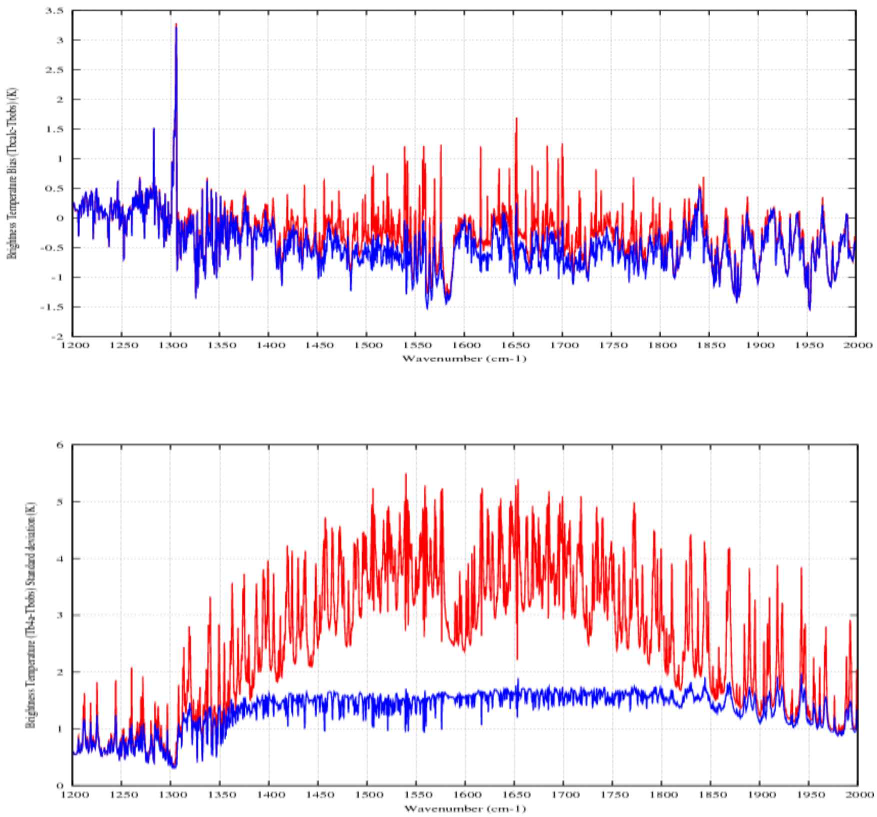
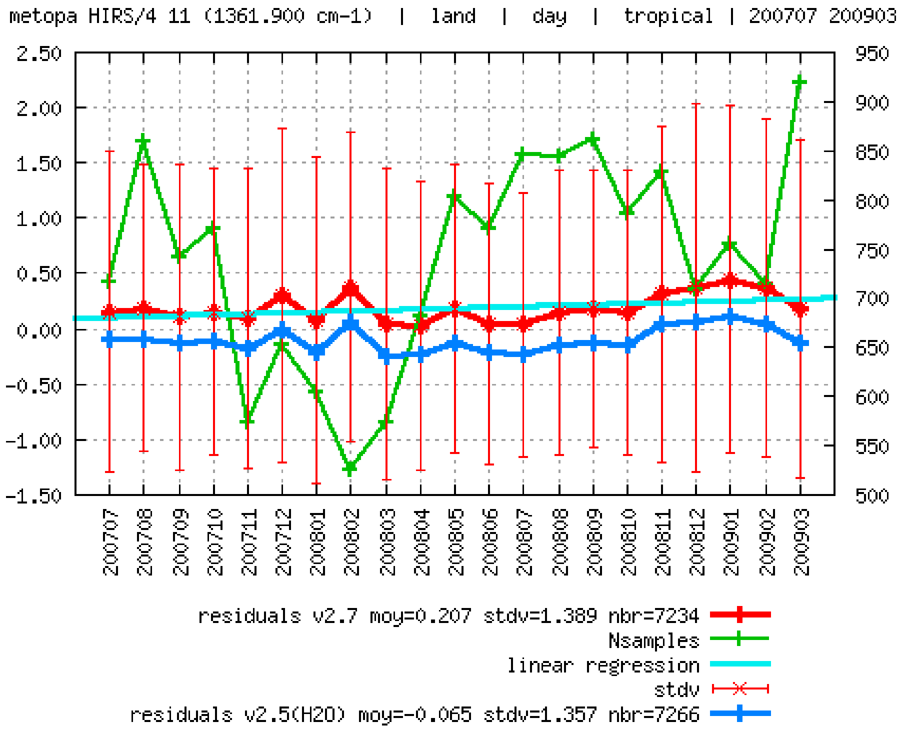


Figure 4-6: *Figure showing the positive impact (bias : top, standard deviation : bottom) on the residuals (simulated-observed) IASI Brightness temperatures obtained after i) extension of RAOBs from 350hPa to 0.1 hPa with ERA_Interim profiles (red) ii) empirical adjustment of ERA_Interim profiles between 350 and 100 hPa (blue).*

We have also verified that this improvement was also present in the computation of the residuals for HIRS4 (channels 11 and 12, sensitive to water vapor in the 6.3 microns spectral region) and MHS (channels 3, 4 and 5), for different latitude zones: -30 +30, (Figures 4.7) +30 +90, -30 -90 (Figures 4.8.a and 4.8.b).



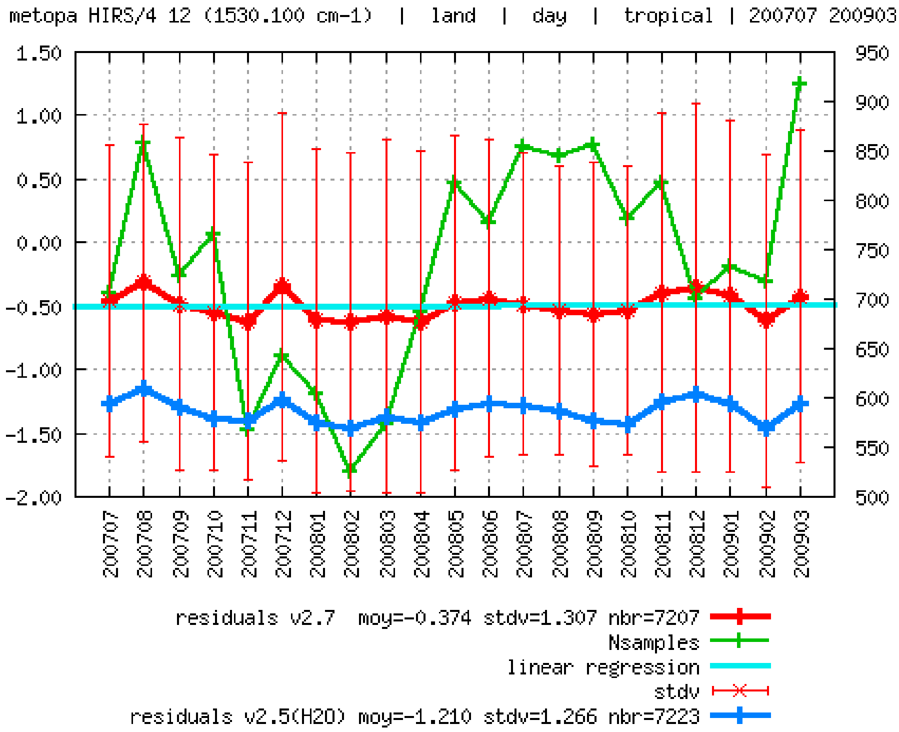


Figure 4-7: *Time series (July 2007 to March 2009) of monthly mean of simulated-observed BTs residuals. Impact on BTs residuals of two successive versions of ARSA for water vapor profiles METOPA/HIRS channel 11 (up) and channel 12 (bottom).*

V2.7: Residuals computed with the current version of ARSA (Red).
V2.5: Residuals computed with ARSA, without the empirical correction made on ERA_Interim H2O profiles (Blue). Nb of ARSA profiles in collocation with MetOpA observations(Green) – right y-axis.

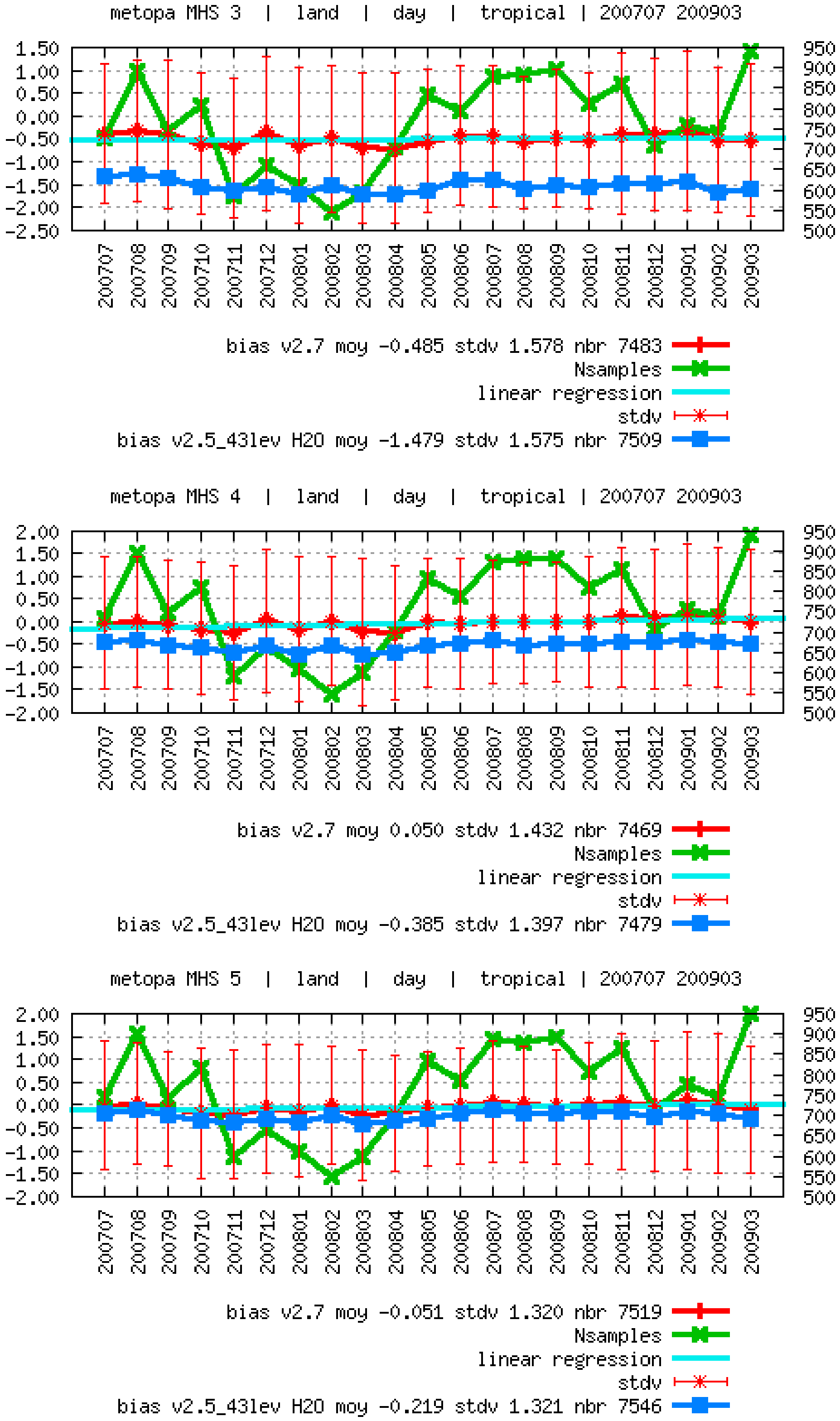
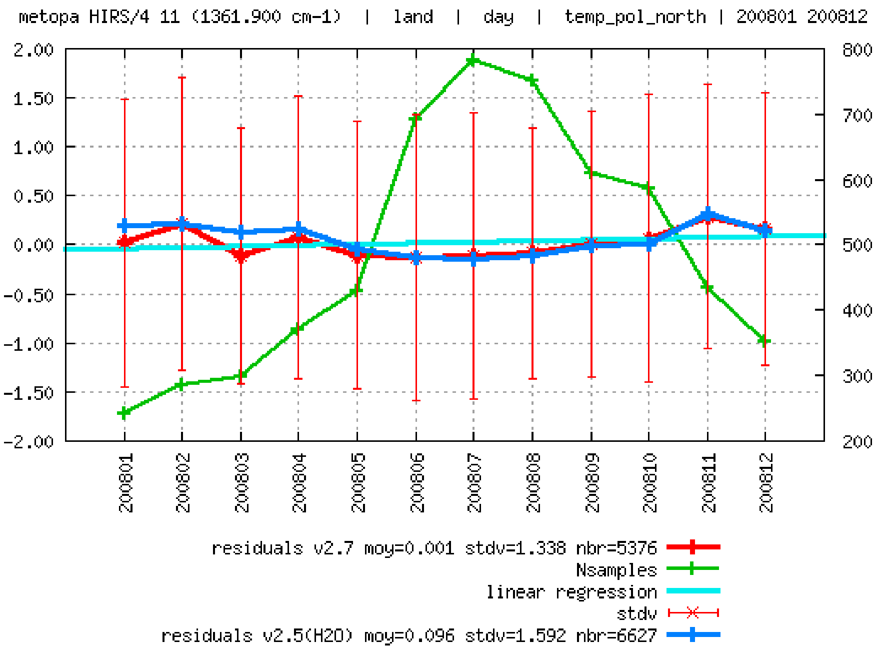


Figure 4.7.b: Same as Figure 4.7.a for MHS/MetOpA.
Time series (July 2007 - March 2009) of monthly mean of BTs residuals for Land/Day Case:

channels 3 (up), 4 (middle), 5 (bottom).

While the preceding results concern the tropical air mass, the following two figures 4.8.a and 4.8.b illustrate results corresponding to resp; +30-+90 latitude zone and -30, -90° latitude zone. One can see the same kind of improvement as in the case of tropical air-masses (cf Figures 4.7). This also demonstrates the appropriateness of the modifications (extrapolation by the ERA_Interim water vapor profiles and empirical correction) for air masses other than the tropical air mass for which these modifications were first performed and validated.



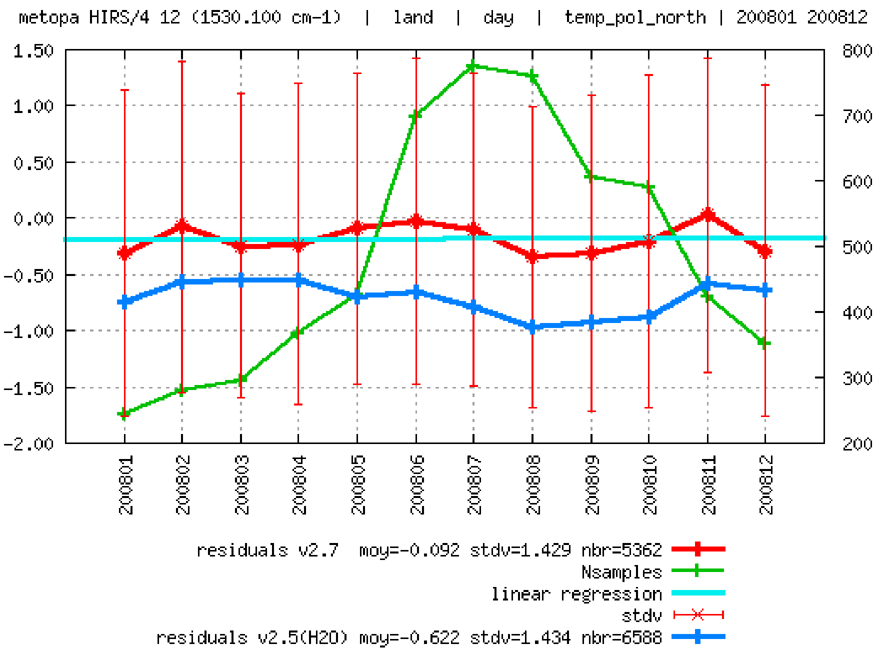
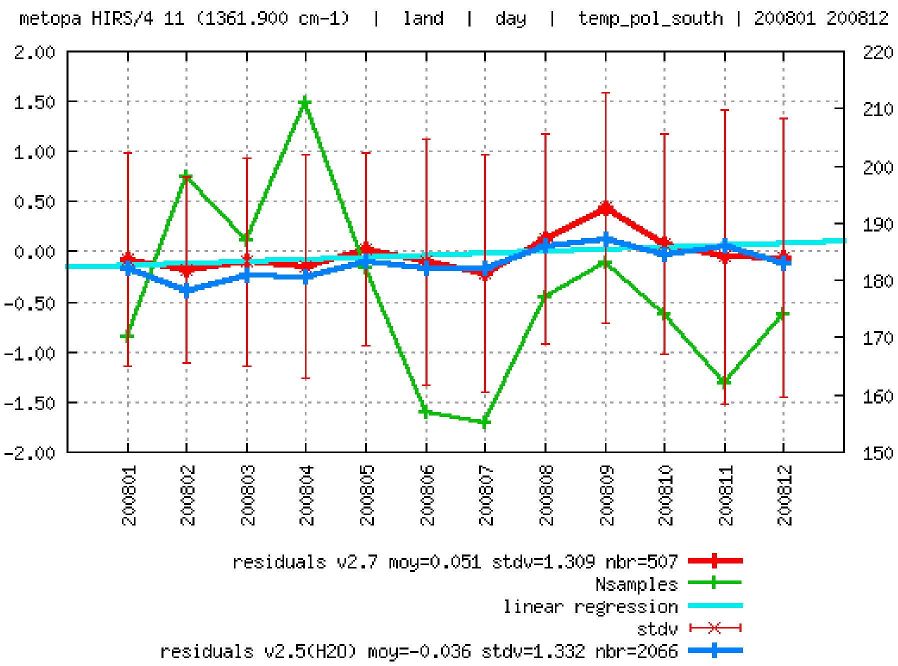


Figure 4-8a: *Time series (January 2008 to December 2008) of monthly mean of residuals for MetOpA. In the +30)-+90° latitude zone HIRS4 channel 11 (top), HIRS channels 12 (bottom).*

LEGEND: V2.7 Residuals computed with the current version of ARSA (Red); V2.5 Residuals computed with ARSA, prior to the empirical correction made on ERA_Interim H2O profiles (Blue); Nb of Items (Green) – right y-axis



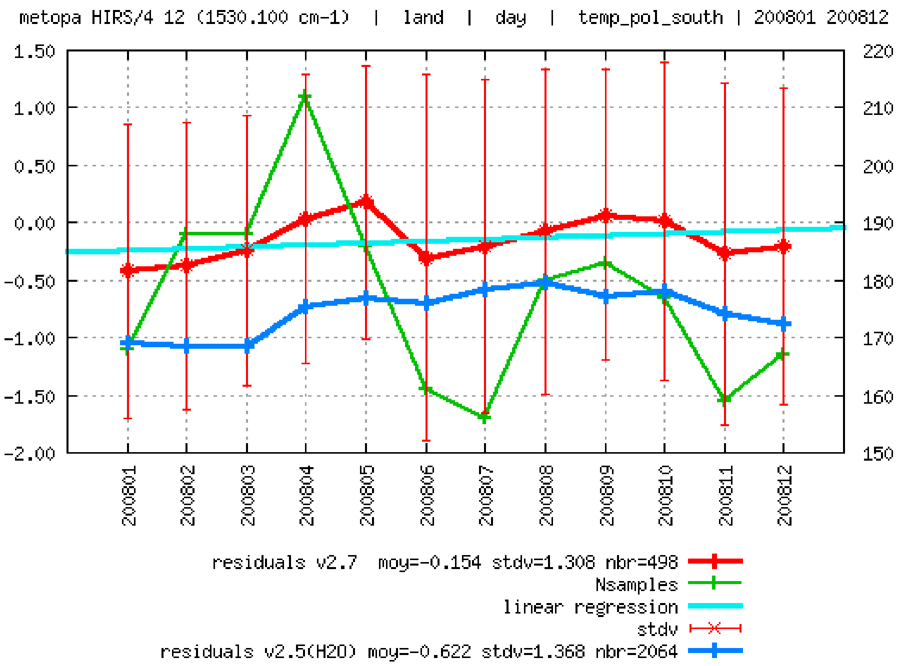


Figure 4.8.b: Same as Figure 4.8.a for the-30°,-90° latitude zone.

The table 4.1 displays statistics (biases, standard deviations, number of collocations) concerning the residuals of HIRS (Channels 11 and 12) and MHS (channels 3, 4, 5) onboard MetOpA, for different cases of Land/Sea/Day/Night cases, for tropical atmospheres, for a whole year of observations (2008). These residuals have been generated with the 4A/OP forward model using as input i) the current version of the ARSA database (results in red); ii) a modified version of the ARSA database which does not include the empirical correction described above (results in blue).

It is seen from values given in this Table (as well from the preceding Figures 4.6, 4.7, 4.8) that the empirical correction brought to the ECMWF water vapor profiles (see above in this same Section) is extremely beneficial in cases considered here: HIRS infrared radiometer and MHS microwave sounder.

MetOpA Residuals	Bias				Stdv				Nbr of collocations			
	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N
HIRS, CH. 11	0.207	0.207	-0.073	-0.233	1.389	1.710	1.238	1.393	7234	3303	4951	2769
HIRS, CH. 11	-0.065	-0.048	-0.335	-0.495	1.357	1.682	1.214	1.376	7266	3321	4974	2787
HIRS, CH. 12	-0.374	-0.290	-0.525	-0.470	1.307	1.614	1.278	1.447	7207	3273	4950	2738
HIRS, CH. 12	-1.210	-1.121	-1.378	-1.302	1.266	1.560	1.231	1.400	7223	3291	4967	2754
MHS, CH. 3	-0.485	-0.555	-0.608	-0.807	1.578	1.955	1.615	1.794	7483	3396	5144	2857
MHS, CH. 3	-1.479	-1.524	-1.608	-1.783	1.575	1.920	1.596	1.765	7509	3413	5170	2876
MHS, CH. 4	0.050	-0.045	-0.058	-0.377	1.432	1.768	1.472	1.614	7469	3401	5130	2855
MHS, CH. 4	-0.385	-0.464	-0.490	-0.810	1.397	1.743	1.450	1.585	7479	3422	5151	2873
MHS, CH. 5	-0.051	-0.200	-0.243	-0.613	1.320	1.570	1.276	1.372	7519	3464	5154	2880
MHS, CH. 5	-0.219	-0.368	-0.404	-0.786	1.321	1.558	1.277	1.363	7546	3473	5183	2889

Table 4.1: Comparison of the biases, standard deviations (Stdv) and number of collocations of monthly means of residuals of MetOpA HIRS channel 11 and 12 and MHS channels 3, 4, 5. In red: current version of ARSA. In blue a version of ARSA which does not include the empirical correction described in this Section. All cases in this Table are for tropical situations and for Land/Day (L/D), Land/Night (L/N), Sea/Day (S/D), Sea/Night (S/N).

4.4.5 Validation of the ARSA database: impact on Ozone profiles

NB: Although the content of this paragraph is an aspect aside of this study - dedicated to the water vapor - it demonstrates the capacity of our method of validation (see Section 4.4.1) to detect unexpected jumps in the time series of residuals and, accordingly, to contribute to an homogenization of the database ARSA, including the elements which constitute it.

As already stated in Section 4.1.2, ozone profiles included in the current version (v2.7) of the ARSA database are extracted from the ERA_Interim reanalyses. The ozone profile is selected according to a time space collocation of the ERA_Interim profile with the ARSA radiosonde report.

Previous versions of ARSA were based on ozone profiles given in Mc Peters (1994). In 1999, we moved to climatological profiles from the 1985-1989 Ozone UGAMP climatology (Li et al., 1995).

http://www.badc.rl.ac.uk/cgi-bin/data_browser/mget/badc/ugamp-o3-climatology/data

Figures 4.9 and 4.10 demonstrate the spectacular benefit gained on the IASI residuals when using ERA_interim ozone profiles (space and time collocated with the radiosonde report) – green curves - instead of our previous source of ozone (climatology from UGAMP) – red curves. This is spectacular both in the nu_1 and nu_3 bands (Figure 4.9) and nu_1+nu_3 band (Figure 4.10): on these figures it is shown that both the bias (top) and the standard deviation (bottom) are improved.

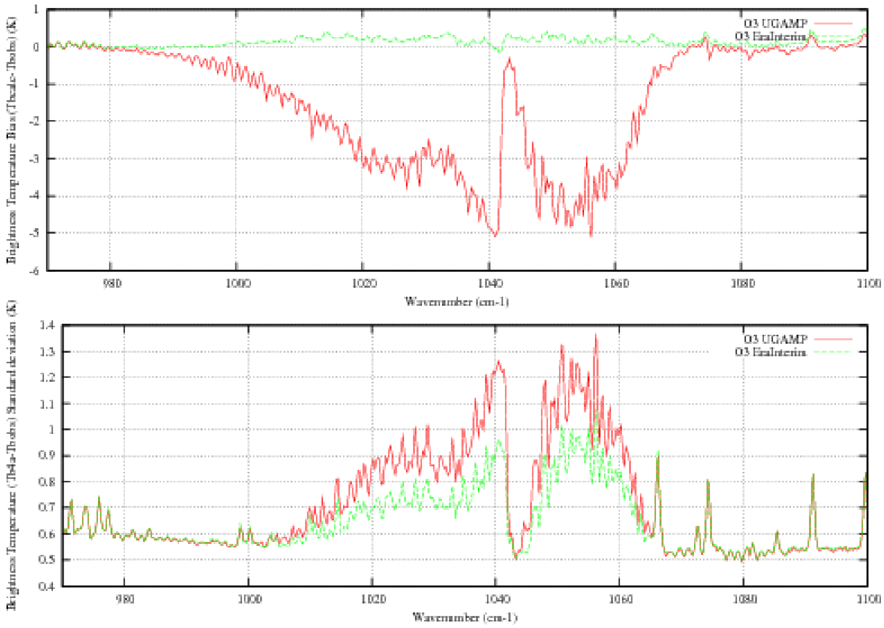


Figure 4-9: *"Simulated-Observed" Bias (top) and Standard deviation (bottom) in the nu1 and nu3 bands of Ozone (in Band I of IASI). Red curves are for our previous source of Ozone profiles, green curves are the current ARSA version based on the ERA_interim results.*

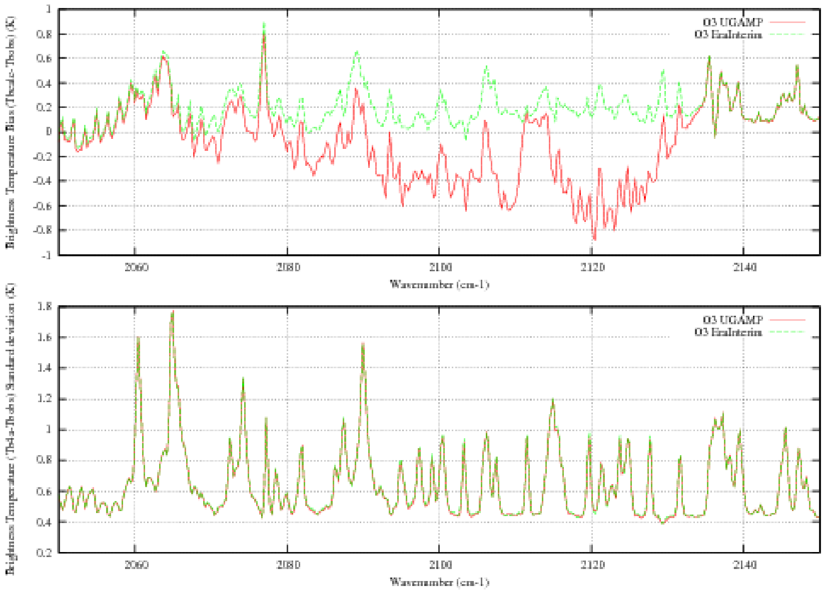


Figure 4-10: *Same as Figure 4.9 for the nu_1+nu_3 band of Ozone, in Band 3 of IASI.*

Always concerning ozone: under study is an a posteriori correction of the current ozone profiles related to an unexpected behaviour of the time series (July 2007-June 2013) of the IASI residuals for a given the wavenumber (not shown here).
From these time series, it appears a discontinuous behaviour occurring in February 2009 leading to as high as 1K difference in the high absorbing spectral regions (less important in the less absorbing ones) between the beginning (July 2007) and the end of the processed period. Since no other spectral regions than ozone display such a behaviour, we came to the conclusion that this discontinuity could be associated to discontinuities in the ERA_Interim ozone profiles, themselves being related to “different assimilation of different satellite data” by ERA_Interim over this period. This latter assumption has been confirmed by Rossana Dragani, ECMWF, Private communication, 2012.

5 Igra_Homogenized

As stated in "Overview of the Integrated Global Radiosonde Archive" by Imke Durre et al - Journal of Climate, vol. 19, Issue 1, pp.53-68): "Radiosonde Archive (IGRA), is a radiosonde dataset from the National Climatic Data Center (NCDC). It consists of radiosonde and pilot balloon observations at more than 1500 globally distributed stations with varying periods of record, many of which extend from the 1960s to present. Observations include pressure, temperature, geopotential height, dewpoint depression, wind direction, and wind speed at standard, surface, tropopause, and significant levels. IGRA contains quality-assured data from 11 different sources. Rigorous procedures are employed to ensure proper station identification, eliminate duplicate levels within soundings, and select one sounding for every station, date, and time. The quality assurance algorithms check for format problems, physically implausible values, internal inconsistencies among variables, runs of values across soundings and levels, climatological outliers, and temporal and vertical inconsistencies in temperature. The performance of the various checks was evaluated by careful inspection of selected soundings and time series. Its temporal and spatial coverage is most complete over the United States, Western Europe, Russia, and Australia. IGRA data are updated on a daily basis and are available online from NCDC as both individual soundings and monthly means."

Later on, the IGRA database has been submitted to a homogenization process (Dai et al., 2011) leading to the IGRA_Homogenized data set. Biases stratification by variables such as location, time-of-day, season, and pressure level has been performed to apply meaningful adjustments to data time series.

6 Methodology and upstream activities to generate the needed datasets

6.1 Pre-processing of the IGRA_Homogenized data and the of the ARSA database

Following next Sections we describe the required steps of processing and of filtering in order to obtain a coherent dataset for the inter-comparison

6.1.1 Desarchive the IGRA_Homogenized dataset from DWD archive

After contacting IGRA and homogenised IGRA PIs and receiving the permission to provide the data , IGRA and IGRA_Homogenised have been provided by DWD (Maarit Lockhoff and Marc Schröder) to LMD (cf mail of November 14th, 2013).

6.1.2 The Station Identification in IGRA_Homogenized and ARSA datasets

The station identification procedure is crucial in order to be able to intercompare the two datasets. This step has been relatively straightforward since the two databases share the World Meteorological Organization (WMO) ID numbers and coordinates for all its station and records in each station. However, to make easier the intercomparison, the ARSA station ID, date, time have been reformatted into the IGRA_Homogenized format. For example:

#010041992100212

Header Record Indicator	1- 1 # character
Station Number	2- 6 WMO station number
Year	7- 10
Month	11- 12
Day	13- 14
Observation Hour	15- 16 00-23 UTC

6.1.3 Reformatting

As indicated in the readme file of IGRA_Homogenized, certain characters give information on the presence or the quality of the variable to which they are associated.

They write:

"1. Cases with -8888 indicate the values were bad and were removed.

- For this study they have been obviously removed.

2. For each pressure, geopotential height, and temperature value, a one-character quality assurance flag indicates whether the corresponding value was checked by procedures based on climatological means and standard deviations. Possible flag values are: blank = no climatological check applied due to an insufficient number of data values for computing the relevant statistics, A = value falls within "tier-1" climatological limits based on all days of the year and all times of day at the station, and B = value passes checks based on both the tier-1 climatology and a "tier-2" climatology specific to the time of year and time of day of the data value."

Based on the explanation given above in bullet 2, we have considered these flags as a warning and not as an obligation to reject the associated value. For this study, we have decided to accept all the variables with such signals as "A", "B", "Blank".

Besides, by accepting these values led us to increase considerably the number of potential comparisons with ARSA because there is in IGRA_Homogenized, a considerable number of B and A flags.

6.1.4 Identify an overlapping period of time for the ARSA and IGRA_Homogenized databases

Due to the fact that the two databases ARSA and IGRA_Homogenized have different "begin-end" periods, an overlapping period had to be found. Once performed all the rejections of uncoherent situations, the period when comparisons become possible extend from January 1st 1979 to 31st December 2010.

6.1.5 Collocation of the ARSA and IGRA_Homogenized databases

The space collocation between the two datasets is perfect due to the same origin of the data.

Then, a rejection is made of IGRA_Homogenized reports which do not reach 300 hPa. This filtering is based on the fact that this specific criterion *is a prerequisite for the water vapor profile in ARSA*: it is thus obvious that none of the IGRA_Homogenized reports which do not reach 300 hPa will find a companion in the ARSA database.

The constraints (vertical extent, density of the measures along the pressure axis) applied to a raw radiosonde report in order to be accepted in ARSA, result in a more severe selection of the raw radiosonde measurements than in IGRA_Homogenized. As a result, the ratio "number of profiles accepted in ARSA / number of raw radiosonde reports" is smaller than the equivalent ratio for IGRA_Homogenized.

As a result, we have, common to the two databases, 940 stations and, for all these stations, more than 2,800000 radiosondes reports.

6.1.6 The units

Three type of variables have been considered in this study: pressure, temperature, water vapor. The table below gives the units in both ARSA and IGRA_ Homogenized databases and the common unit chosen for this study:

	ARSA	IGRA_Homogenized	Units for this study
Pressure	hPa	units of Pa (mb * 100)	hPa
Temperature	°K	units of 0.01degrees C	°K
Water vapor	g/g	units of 0.00001 g/kg	g/g
Precipitable Water vapor	-	-	cm

6.1.7 Towards a common ARSA IGRA_Homogenized pressure grid: standard pressure levels

ARSA profiles are described on a nominal 43-level pressure grid between surface and 0.0026hPa (see Table in Section 14).

Concerning the IGRA_Homogenized database, values of temperature, water vapor, winds, … are generally given at standard pressure levels. As indicated in Section 11, some values are given at additional “significant thermodynamic levels”. We have decided to work with the standard pressure levels. Throughout the IGRA_Homogenized database, we have identified 17 such levels between surface and 1 hPa. We have retained the surface as a 18th level.

They are given in the next table.

Level 1 : 1 hPa	Level 2 : 5 hPa	Level 3 : 10 hPa	Level 4 : 20 hPa
Level 5 : 30 hPa	Level 6 : 50 hPa	Level 7 : 70 hPa	Level 8 : 100 hPa
Level 9 : 150 hPa	Level 10 : 200 hPa	Level 11 : 250 hPa	Level 12 : 300 hPa
Level 13 : 400 hPa	Level 14 : 500 hPa	Level 15 : 700 hPa	Level 16 : 850 hPa
Level 17 : 1000 hPa	Level 18 : SURFACE		

NB: Every time the surface pressure is smaller than 1000hPa, the “surface level” identification number is renamed from 18th to 17th or 16th, … according to its value at the time of the measurement. Note that, in this case, the original specific values of the temperature and water vapor variable associated to this surface pressure level are kept unchanged.

In order to get a uniform pressure grid, temperature and water vapor profiles of ARSA have been linearly interpolated to these 18 standard pressure levels.

ARSA 43 Pressure Levels ➔ Linear Interpolation ➔ 18 Standard Pressure Levels

Built on these 18 Standard Pressure level grid, a 17 layer system has been built, the limits of which are two successive adjacent standard pressure levels.

For each layer, a mean temperature and a mean water vapor specific humidity are computed as the mean values of the temperatures and the water vapor specific humidity associated with each of the two pressure levels which limit the layer.
Also, for each layer, a precipitable water (PW) amount is computed according to the following formula:

$$\text{Specific humidity} * (P_2\text{-}P_1)/g$$

6.1.8 The deep layers system

For the water vapor product, seven deep layers have been defined, built from the nominal standard pressure levels.

Layer 1 1.00 30.0	Layer 2 30.0 100.0	Layer 3 100.0 300.0	Layer 4 300.0 500.0
Layer 5 500.0 700.0	Layer 6 700.0 850.00	Layer 7 850.0 1000.00	

For each of these deep layers limited by pressures $P_L$and P_U, a precipitable water vapor amount is computed according to the formula:

$$(1/g)\ \Sigma\ \rho_i * (P_i\text{-}P_j)$$

where ρ_i is the mean value of the specific humidity in the layer limited by pressures P_i and P_j.
The sum is performed for all intermediate layers (P_i,P_j)contained within the interval P_L to P_U

6.1.9 An auxiliary dataset: the raw radiosonde dataset

As auxiliary dataset to serve as “historical” reference, we have also included “in the loop” of this study the raw radiosonde report archive we get from the ECMWF.

6.1.10 Towards the final merged ARSA, IGRA_Homogenized, Raw radiosonde reports dataset

At this stage and based on the long and rigourous QC process , *all the unphysical values have been eliminated from the merged ARSA, IGRA_Homogenized, Raw Radiosonde reports data base and all missing values replaced by a “missing data” flag and a coded value*.

However to be even more sure to detect “unexpected event” (misplaced values, or incorrectly formatted values, …, or gaps in the three different datasets) we apply a “minmax” type procedure on each column of the dataset. “Minmax” is a very simple LINUX command that reads ASCII files and finds the gross extreme values in each of the columns. It recognizes NaNs and prints warnings if the number of columns varies from record to record. When minmax control detects “unexpected” values this generally lead to the revision of one of the previous steps of the quality control process, or of the modification of some threshold values, or correction of bad format. This process is iteratively executed in order until obtain one "coherent with physics" diagnosis at the end.

After all the previous steps, ARSA, IGRA_Homogenized and the RAW_RADIOSONDE file are put together into a unique file.

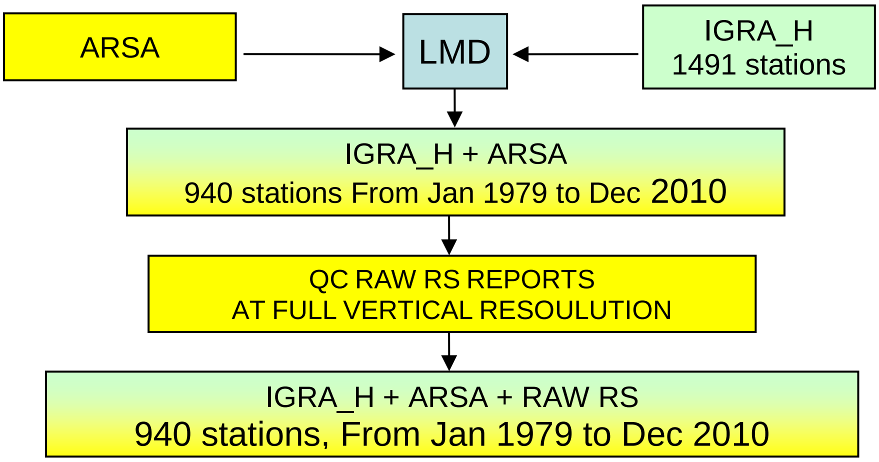


Figure 6.1 gives a representation of the number of collocated ARSA and IGRA_Homogenized data (y-axis) as a function of the pressure level (x-axis). This figure illustrates the fact that above the 300 hPa levels, the number of values measured by the radiosonde - found in IGRA_Homogenized - become smaller and smaller.

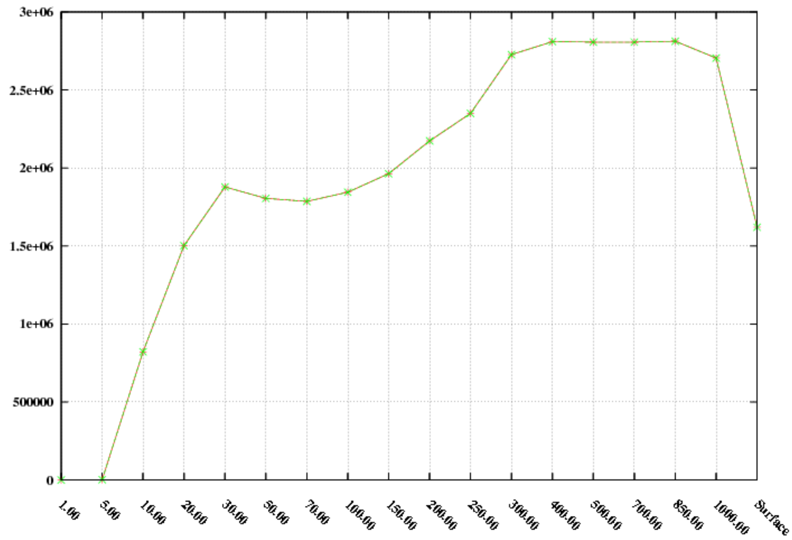


Figure 6-1: *Number of collocated ARSA and IGRA_Homogenized data (y-axis) as a function of the pressure level (x-axis).*

6.2 Satellite data

This Section addresses work described in Section 2, task 3: “Simulate satellite radiances for sub-periods such as 2 years in each decade between 1980 and 2010. In recent years IASI can serve as reference. The HIRS and MHS instrument on board the same satellite will also be used in order to ease the interpretation”.

NB: This part of the work only concerns the ARSA database. Eventually, due to a lack of sufficient vertical resolution and missing information on certain pressure levels, the majority of the IGRA_Homogenized database cannot be used for such simulations. The work to complete or interpolate between the existing IGRA_Homogenized standard pressure levels in order to get a coherent representation of the atmospheric thermodynamic state would have probably introduced too high a noise to lead to reliable conclusions.

6.2.1 The choice of vertical sounding instruments

For the purpose of this contract, several *passive vertical sounding* instruments - all observing the earth atmosphere in spectral regions particularly sensitive to water vapor - have been considered. The observations are made either in the infrared or in the microwave.

These instruments are:

- IASI, HIRS4 and MHS onboard MetOpA,
- HIRS on board NOAA 10, NOAA 11, NOAA15.

Among the 19 infrared channels of each of the HIRS instruments, Channels 10, 11 and 12 are particularly sensitive to water vapor in various layers along the pressure range. This is illustrated by figure 6.2 below which displays the Jacobians with respect to water vapor for three IASI channels “companion” (ie radiatively close to) of HIRS channels.

For the microwave spectral regions, channels 3, 4, 5 of MHS have been selected.

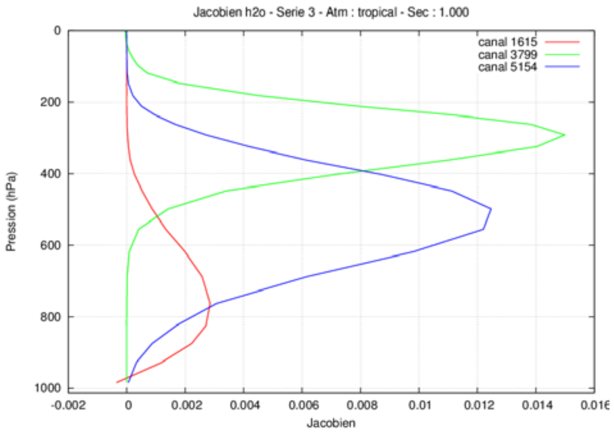


Figure 6-2: *Jacobians with respect to water vapor for three IASI channels “companion” (ie radiatively close to) of HIRS channels (Jacobians computed with the 4A/OP model).*

Figure 6.3 (and 6.4) display, superimposed on a IASI spectrum in the 6.3 microns spectral region, the filter functions of HIRS channel 11 (and HIRS channel 12) of NOAA-10, NOAA-11, NOAA-15, MetOpA.

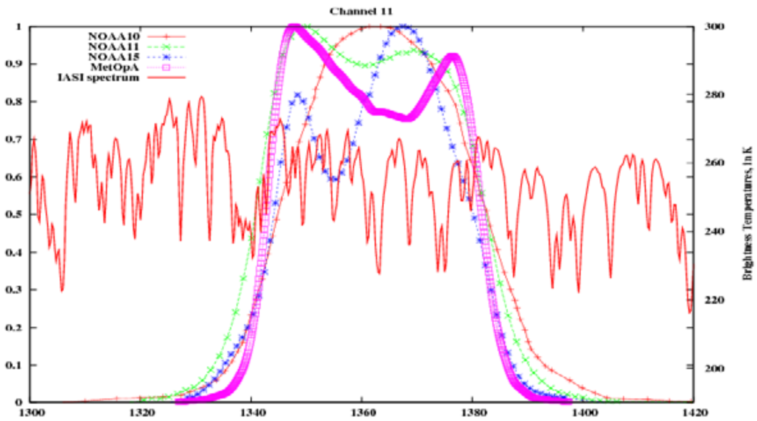


Figure 6-3: *Comparison of the channel 11 filter functions of HIRS/3 onboard NOAA-10, NOAA-11, NOAA-15 and HIRS/4 onboard MetOpA.*

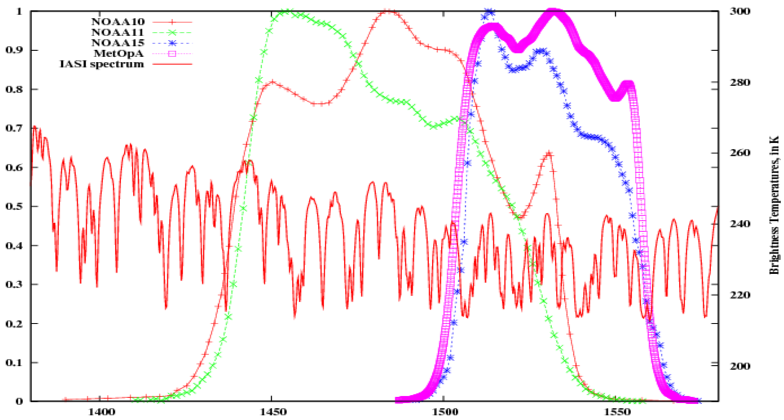


Figure 6-4: *Same as Figure 6.3 for channel 12 filter functions.*

6.2.2 The choice of the periods

LMD and DWD agreed to have several 2-year sub-periods in each decade – between 1980 and 2010 - to check the stability of the ARSA database.

In addition to the IASI/MetOpA July 2007 to March 2009 period, the following sub-periods have been finally chosen:

- NOAA 10 May 1989 to April 1991
- NOAA 11 January 1990 to December 1991

- NOAA 15 February 2001 to January 2003

6.2.3 Origin of the satellite data

All the satellite data used in this work have been extracted from the ARA/LMD archive.

7 The different data/metadata files generated during the study

This study has generated a certain number of

- Files
- Statistics
- Graphs

To better understand the number of files/graphs generated, it is worth recalling the following numbers:

- Three databases intercompared: ARSA, IGRA_Homogenized, RAW_Radiosonde reports
- Three variables: Temperature and Water Vapor and precipitable water vapor
- 940 radiosonde stations
- ~2,8 Millions of measurements
- 17 standard pressure levels and 1 surface pressure level
- 7 deep layers (built from the 18 standard pressure levels)

7.1 The merged ARSA, Raw radiosondes reports and IGRA_Homogenized dataset

As stated in Section 6.1.9, we have added to the ARSA and IGRA_Homegenized dataset, the space time collocated dataset of raw radiosonde reports.
As indicated in Section 6.7, the pressure grid is a grid of 17 standard pressure levels with an additional pressure level corresponding to the surface.
The Raw radiosondes reports and IGRA_Homogenized datasets did not require any specific processing since they are nominally described on this pressure grid.
To get ARSA values of temperature and water vapor at this standard levels pressure grid, we made a linear interpolation between the two ARSA pressure levels below and above the considered standard pressure level.

7.2 The different statistics generated during the study

Statistics (bias, standard deviations) have been performed on the 18 standard levels (surface to 1 hPa) for water vapor and temperature products and also on 7 deep layers for water vapor products.

NB 1: All the files generated by this study and included as deliverables are either in ASCII or .txt format.

NB 2: Details on the format and variables in each file are given in the user guide.

7.2.1 Statistics per station, per level for water vapor product

Statistics are performed for:

- ARSA in stand alone
- IGRA_Homogenized in stand alone
- ARSA versus IGRA_Homogenized
- IGRA_Homogenized versus raw radiosonde data
- ARSA versus raw radiosonde data

The output products are:

- Files of results
- Graphs

7.2.2 Statistics per station, per level for temperature product

Statistics and outputs same as Section 7.2.1

7.2.3 Statistics per station per deep layer for water vapor product

Statistics and outputs same as Section 7.2.1

7.2.4 Time series of water vapor products per station per level: graphs

Statistics and outputs same as Section 7.2.1

7.2.5 Time series of temperature per station, per level: graphs

Statistics and outputs same as Section 7.2.1. Below is an example of such an output.

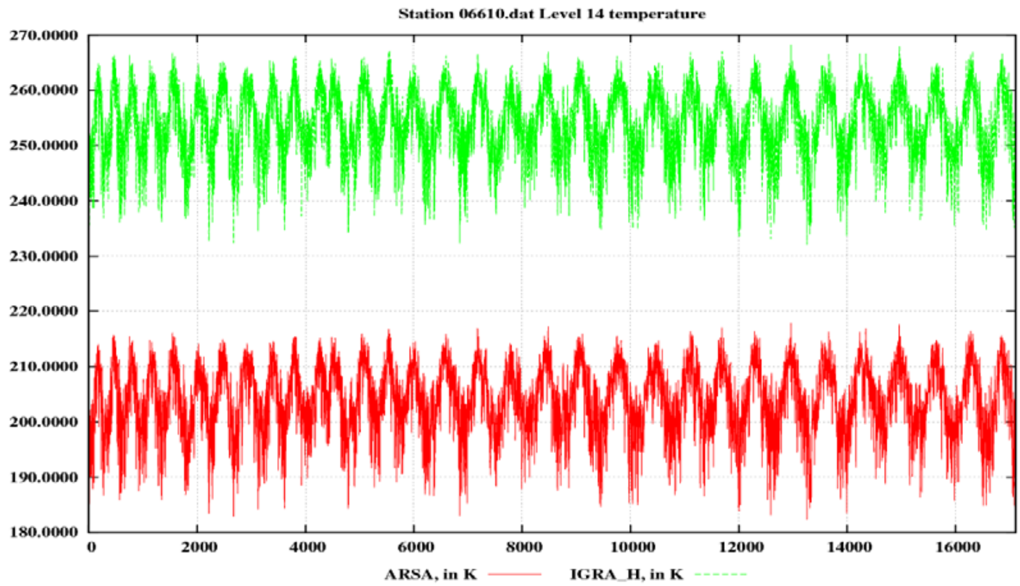


Figure 7-1*: Time series of temperature at 500 hPa standard pressure level "IGRA_Homogenized" (red) and "ARSA" (green) for station 04339. For sake of clarity, one of the archive has been arbitrarily shifted (50K) on the y-axis with respect to the other.*

7.2.6 Map of the statistics on the water vapor product: all stations, full period, per deep layer

These statistics (bias between ARSA and IGRA_Homogenized water vapor product) are made for every station, every deep layer (see Section 6.1.8), over the whole period: the results are shown as a 3D (latitude, longitude, bias) representation on a map. Examples of such maps are given below for layer 850 hPa-Surface (left); Layer 300-100 hPa (right).

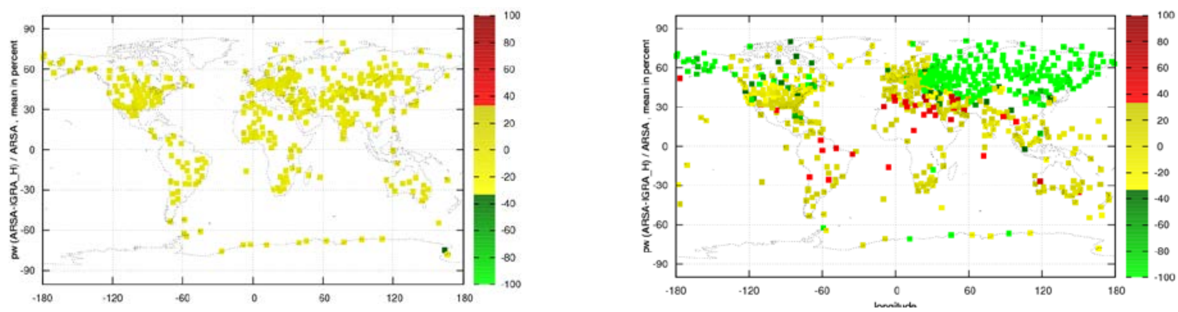


Figure 7-2*: Map projection of the bias on water vapor (full period). Units are in percentage of ARSA-IGRA_Homogenized/ARSA. Layer 850hPA-Surface (left); Layer 300-100hPa (right).*

Associated to each station represented on this map are graphs showing the time series for this station and each available deep layer of this station (see below Section 7.2.10).

7.2.7 Map of the statistics on the temperature product: all stations, full period, per level

These statistics (bias between ARSA and IGRA_Homogenized temperature product) are made for every station, every level, over the whole period: the results are shown as a 3D (latitude, longitude, bias) representation on a map. Associated to each station represented on this map are graphs showing the time series for this station and each available level of this station (see below Section 7.2.11).

7.2.8 Time series of ARSA vs IGRA_Homogenized water vapor products

These graphs represent time series of the difference “ARSA- IGRA_Homogenized” for the water vapor product. There is one graph per station and one per pressure level.

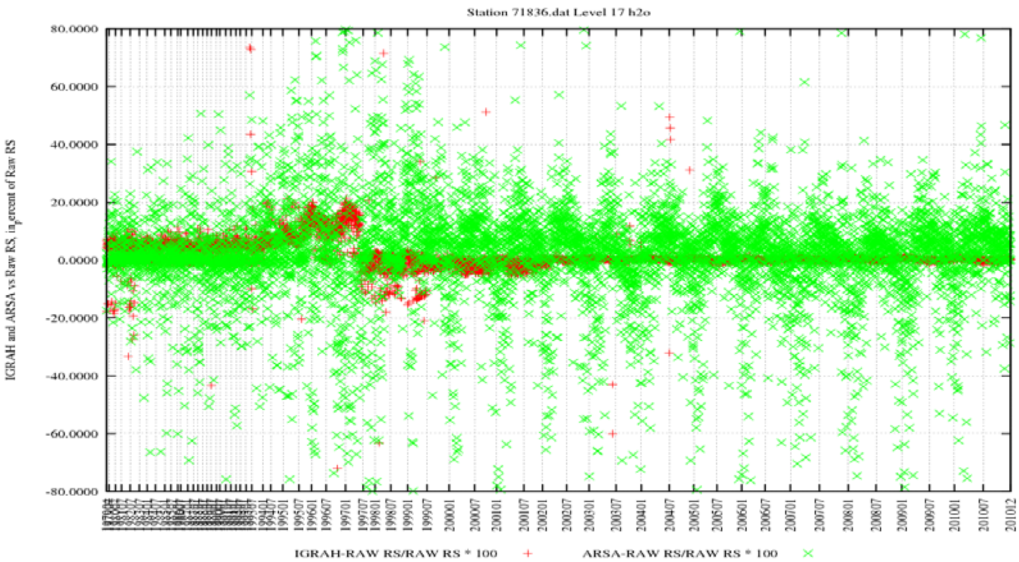
7.2.9 Time series of ARSA vs IGRA_Homogenized temperature products

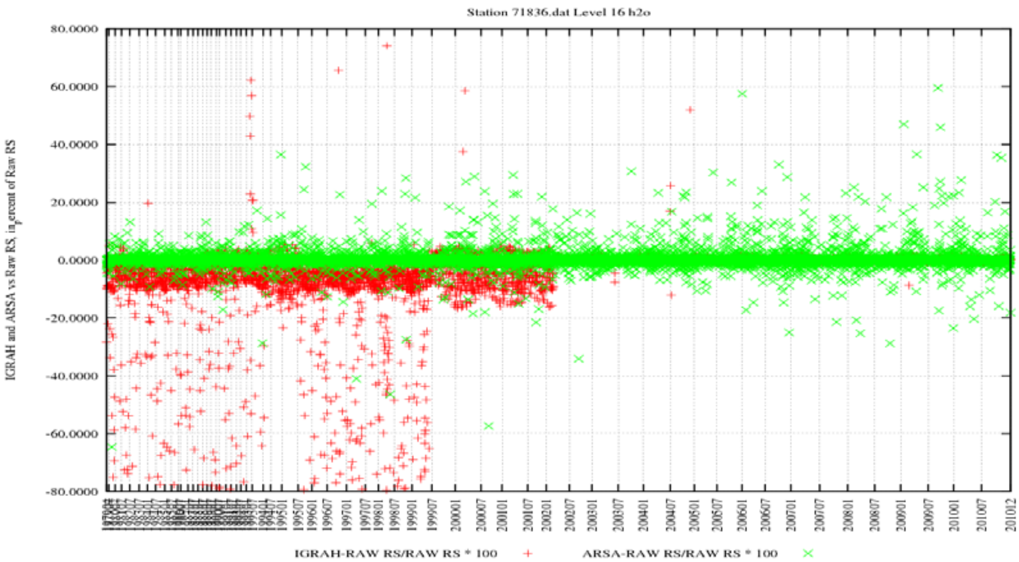
These graphs represent time series of the difference “ARSA- IGRA_Homogenized” for the temperature product. There is one graph per station and one per pressure level.

7.2.10 Time series of ARSA and IGRA_Homogenized water vapor products compared to raw radiosondes reports

These graphs represent, on the same graph, time series of the difference “ARSA-Raw Radiosonde report” (in green) and “IGRA_Homogenized- Raw Radiosonde report” (in red) for the water vapor product for different pressure levels. There is one graph per station and one per pressure level. Along the x-axis is a running index associated to the date of the measurement.

Examples of such graphs are given below for levels 17, 16 and 13 and for station 71836.
These graphs illustrate the different behaviour of the two databases with respect to the raw radiosonde report.
They also illustrate, from one pressure level to the other, the difference in the homogenization process for IGRA.


Station 71836.dat Level 17 h2o
IGRAH and ARSA vs Raw RS, in percent of Raw RS
IGRAH-RAW RS/RAW RS * 100
ARSA-RAW RS/RAW RS * 100


Station 71836.dat Level 16 h2o
IGRAH and ARSA vs Raw RS, in percent of Raw RS
IGRAH-RAW RS/RAW RS * 100
ARSA-RAW RS/RAW RS * 100

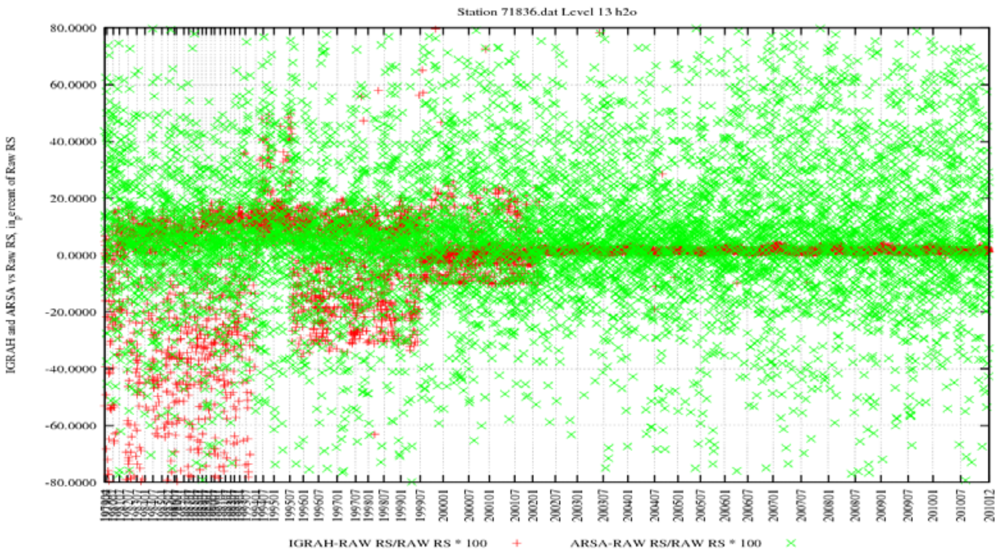


Figure 7-3: *a, b, c Time series of bias on water vapor "IGRA_Homogenized – Raw Radiosonde reports" (red) and "ARSA – Raw Radiosonde reports" (green) for station 71836. Along x-axis is a running index associated to the date of the measurement. Figures from top to bottom are for standard pressure levels 17, 16, 13, respectively.*

7.2.11 Time series of ARSA and IGRA_Homogenized temperature products compared to raw radiosondes reports

These graphs represent time series of the difference "ARSA-Raw Radiosonde report" and "IGRA_Homogenized- Raw Radiosonde report" for the temperature product. The results concerning the two comparisons are on the same graph. There is one graph per station and one per pressure level.

8 Other relevant metadata files generated during the study

8.1 List of problems encountered with some stations in IGRA_Homogenized. Identification of the corresponding stations

A few technical problems appeared when trying to use the IGRA_Homogenized files.

- « A », « B », « I », « ***** » characters in the files
- A few stations contain records which are not in chronological order.
- A few files contain two different stations in the same radiosonde file. The corresponding station numbers are the following.
  - 71203 (71151) → rejected
  - 60018 (60020) → rejected
  - 40437 (40438) → rejected
  - 08522 (08521) → rejected

Every such situation has led to the rejection of the concerned station(s).
This does not mean that they have to be ignored at all times. Eventually, as soon as they are correctly labelled by IGRA_H team, it will be possible to reintroduce them in the QC process.

8.2 List of problems encountered with some stations in ARSA. Identification of the corresponding stations

Unexpectedly, one station among the 940 has no data for year 2009.

8.3 List of the 940 stations finally retained for the ARSA IGRA_Homogenized inter-comparison

This file gives the identification of the stations finally retained.

8.4 Starting and ending date of each station participating to the inter-comparison.

This file gives the starting and ending date of the stations finally retained.

8.5 Number of radiosonde reports retained for the inter-comparison: station per station

This file gives station per station, the number of radiosonde reports. This is illustrated on Figure 8.1 below.

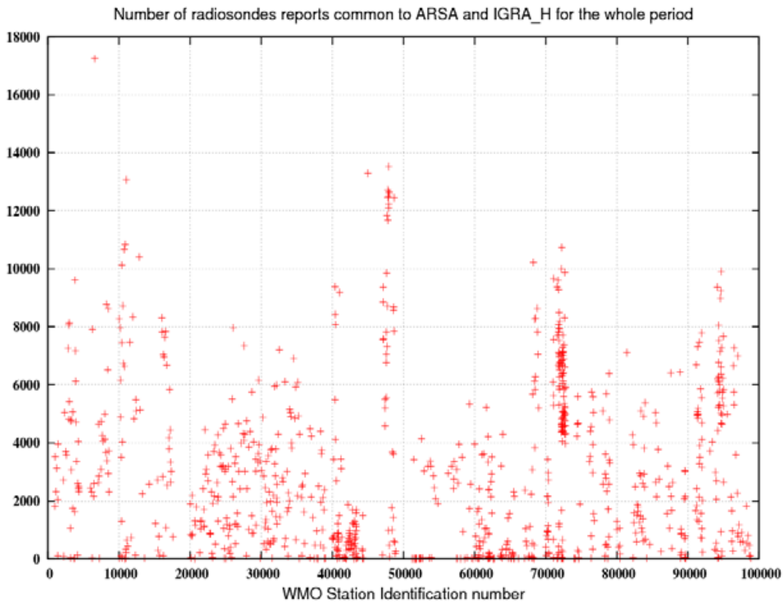


Figure 8-1: *Number of radiosonde reports common to ARSA and IGRA_Homogenized for the whole period (January 1979 – December 2010).*

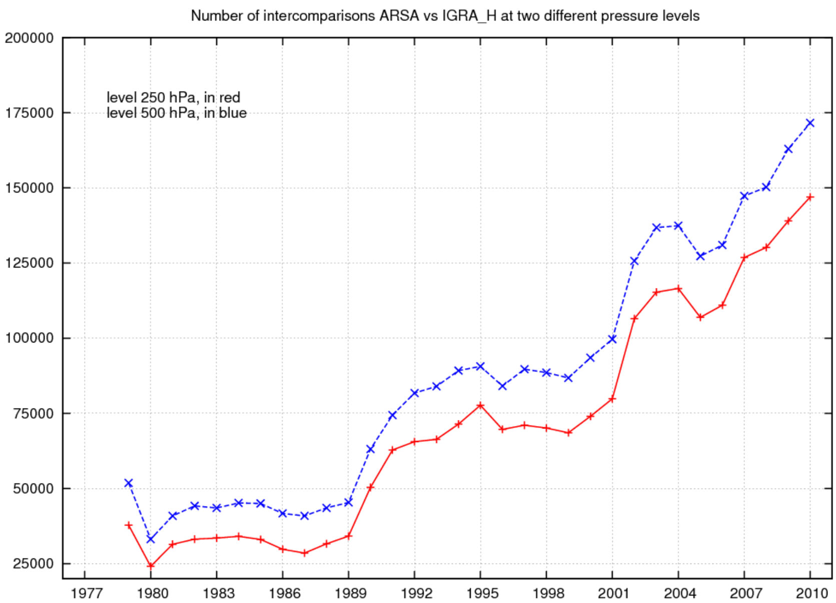


Figure 8-2: *Number of radiosonde reports common to ARSA and IGRA_Homogenized at two different pressure levels: 250 hPa (in red) 500hPa (in blue), between 1979 and 2010 (x-axis).*

8.6 Map projection of the number of items used to compute the statistics on water vapor product: all stations, per deep layer

This map projection is another way to represent the Figure 8.1. It is a 3-D representation of the number of radiosonde reports common to ARSA and IGRA_Homogenized for the whole period (January 1979 – December 2010) as a function of latitude and longitude.

8.7 Map projection of the number of items used to compute the statistics on temperature product: all stations, per level

Same as 8.5 for temperature product.

8.8 Characteristics of situations finally retained for the ARSA IGRA_Homogeneized inter-comparison, per level, all stations

8.9

Such characteristics are given in an ASCII file:

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Number of situations In common	Latitude of the station	Longitude of the station	Identification of the station (WMO nb)	Begin date (YYYYMM)	End date (YYYYMM)

8.10 Identification of situations finally retained for the ARSA IGRA_Homogeneized inter-comparison, per level, all stations, created in ascending order of the number of situations

Same as above, but rearranged in ascending order of the number of situations in common (column 1). These results – including their associated 3-D graph below - can be used for the selection of a “reference stations list” offering the larger number of observations over the considered period. In the representation given in the 3-D map of Figure 8.3, it appears that 150 to 200 (over the 940 nominal) stations have at least 6,000 measurements at standard pressure levels 700 hPa and 300 hPa.

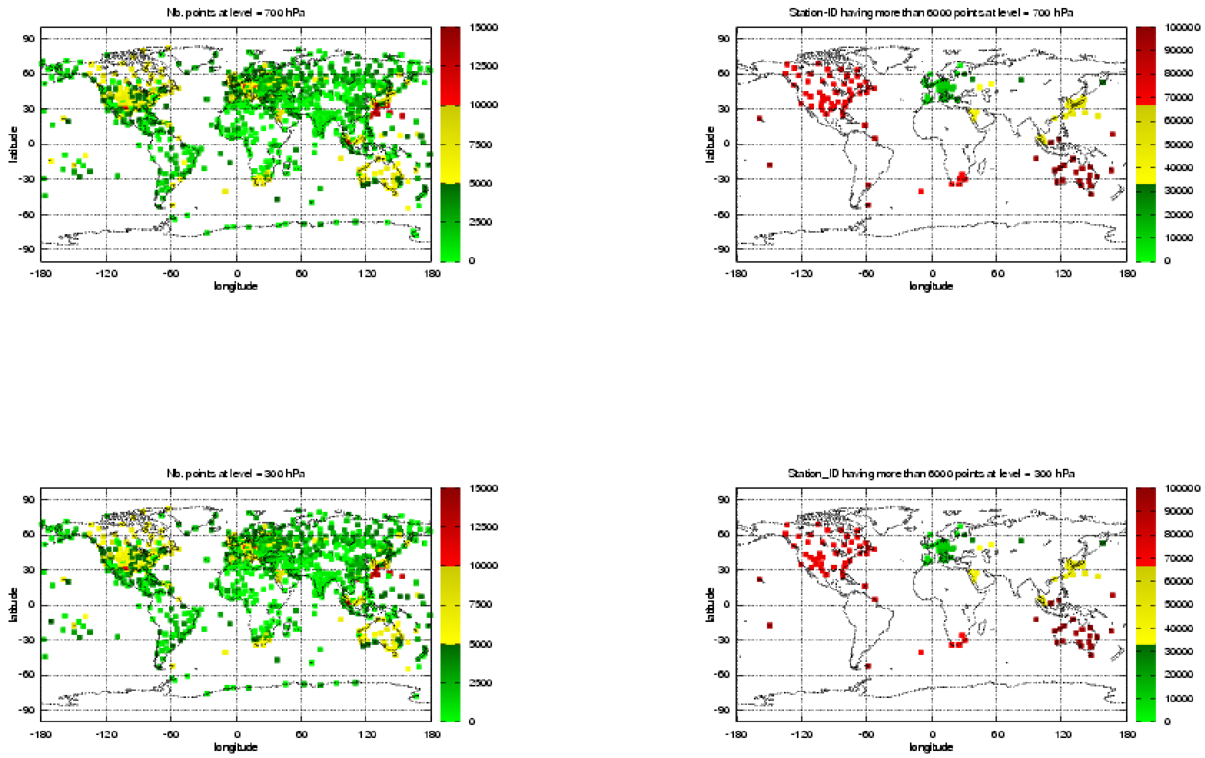


Figure 8-3: *Number of points per station at 700 hPa (top left) and 300 hPa (bottom left) standard pressure levels. Graphs on the right side of the figure show the ID of the stations (WMO code) having more than 6,000 points over the period considered (Jan 1979 – Dec. 2010); top right for 700 hPa, bottom right is for 300 hPa).*

9 Results

9.1 Statistics for the intercomparison of ARSA and IGRA_Homogenized

As indicated in the preceding Sections, many kinds of statistics have been performed either on the water vapor products or on the temperature products.

The two bases ARSA and IGRA_Homogenized have been considered either

- in stand alone or
- in inter-comparison or also,
- each of the two in comparison with the raw radiosonde reports.

Preliminary remarks – however obvious – are required:

For the present study, an advantage is that there is no misfit in the identification or space time collocation of stations.

A great difficulty comes from the totally different approach and the way the products have been obtained and designed for.

Another difficulty comes from the differences in the pressure grids: indeed it has to be recalled that, *for the present study*, ARSA profiles have been submitted to an interpolation between pressure levels of its nominal pressure grid (see Section 6.1.7) in order to get products (temperature and water vapor) at the same (standard) pressure levels values found in IGRA_Homogenized.

Having this in mind, we can however make the following remarks:

In standalone, ARSA and IGRA_Homogenized have very comparable statistics (mean and standard deviations) from the 1000 to ~400 hPa pressure levels. For the differences occurring above this level, it has to be recalled that ARSA has been extrapolated with ERA_Interim values and furthermore that its water vapor is the result of an empirical adjustment (see Section 4.4.4) of the ERA_Interim water vapour profiles based on the study of simulated-observed brightness temperatures values (residuals). Both these modifications lead to noticeably improved results in the statistics on residuals as can be seen in figures of Section 4.4.2 and 9.2.

When inter-compared, ARSA and IGRA_Homogenized differences occur at two places of the pressure grid: one above 400 hPa the other near the surface. For the "above 400 hPa" vertical range, the explanation given above for ARSA is still valid. For the pressure region near the surface, interpolation to e.g. 1000 hPa and extrapolation of ARSA profiles to the surface pressure may present some weaknesses, resulting in an increased standard deviation with respect to IGRA_Homogenized or with respect to raw radiosonde reports.

Concerning the stability, we have based our study on such figures as time series representation of water vapor from ARSA, from IGRA_Homogenized and from the Raw Radiosonde reports. These Figures are available for all stations and all standard pressure levels. Examples of such figures are Figure 9.1 as well as Figures 7.4. These Figures give an illustration not only of the behaviour of the two databases in “stand-alone”, but also of their behaviour with respect to the raw radiosonde reports.

The interpretation of such figures in terms of stability is quite difficult. In particular, in Figure 9.1, (as well as in Figures 7.4) we see very well the points where corrections (or no corrections) have been made for the homogenization of IGRA. To the best of our knowledge and based upon hundreds of such graphs obtained during this study, the most recent period of time has been often selected to be the reference to which other periods of time have been adjusted.

This intercomparison makes quite impossible to draw a conclusion on the stability of the two databases.

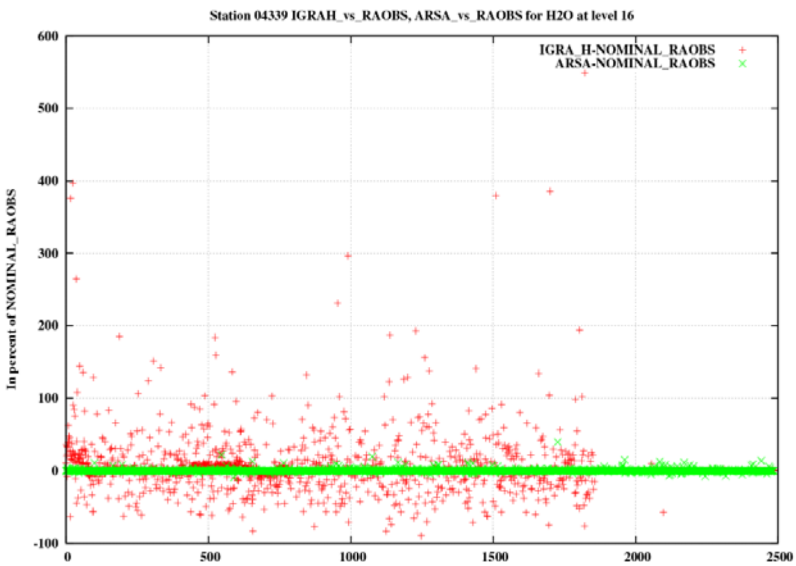


Figure 9-1: *Time series of bias on water vapor at 850 hPa standard pressure level “IGRA_Homogenized – Raw Radiosonde reports” (red) and “ARSA – Raw Radiosonde reports” (green) for station 04339. Along x-axis is a running index associated to the date of the measurement.*

9.2 Simulations of Satellite radiances based on ARSA

Preliminary statement: As pointed out in Section 6.2, due to a lack of sufficient vertical resolution and missing information on certain pressure levels, IGRA_Homogenized database cannot be used for such simulations. The work to complete or interpolate between the existing IGRA_Homogenized standard pressure levels in order to get a coherent representation of the atmospheric thermodynamic state would have probably introduced too high a noise to lead to reliable conclusions.

As indicated in Section 6.2, several periods, satellites, instruments have been chosen. The computation of the residuals has been performed for the corresponding cases

Satellite	Instrument	Period	Channel(*)	Case studies L,S,N,D (**)	Air Mass
NOAA 10	HIRS 3	May 1989 - April 1991	Chan. 11, 12	L/D L/N S/D S/N	-30° + 30°
NOAA 11	HIRS 3	January 1990 - December 1991	Chan. 11, 12	L/D L/N S/D S/N	-30° + 30°
NOAA 15	HIRS 3	February 2001 - January 2003	Chan. 11, 12	L/D L/N S/D S/N	-30° + 30°
MetOpA	HIRS4	July 2007 - March 2009	Channels 11,12	L/D L/N S/D S/N	-30° + 30°
MetOpA	MHS	July 2007 - March 2009	Channels 3,4,5	L/D L/N S/D S/N	-30° + 30°

Table 9.1 Summary of the periods and numerical experiments chosen for this study.

(*) For each of these instruments, all the channels are processed. Only channels particularly sensitive to tropospheric water vapor are indicated here.

(**) Characters "L", "S", "D", "N", resp. stand for Land, Sea, Day and Night. The combination L/N stands, for example, for Land+Night Situations.

The following figures (9.2, 9.3, 9.4,) confirm, for the processing of other than MetOp periods and other than MetOP/IASI satellite instruments, the relevance of the empirical corrections (see Section 4.4.4) performed on the ERA_Interim water vapor profiles, required after comparisons between simulated and observed IASI brightness temperatures.

In the following figures we see time series of such residuals over the period selected for each satellite (see Section 6.2.2). These figures compare residuals obtained with the *current version of ARSA (in red, referred to as version v2.7) to a modified version of ARSA not taking into account the empirical corrections (in blue referred to as version v2.5) with all else equal*. It can be noticed a general improvement of the biases. Standard deviations more or less remain the same: it has to be noticed that since they mainly come from the noise of the collocations of the "closest" ARSA to the satellite observation, they are expected not to differ from one experiment to the other.

In addition to these three Figures, the Table 9.2 (see after figure 9.4) summarizes the results for these three satellites.

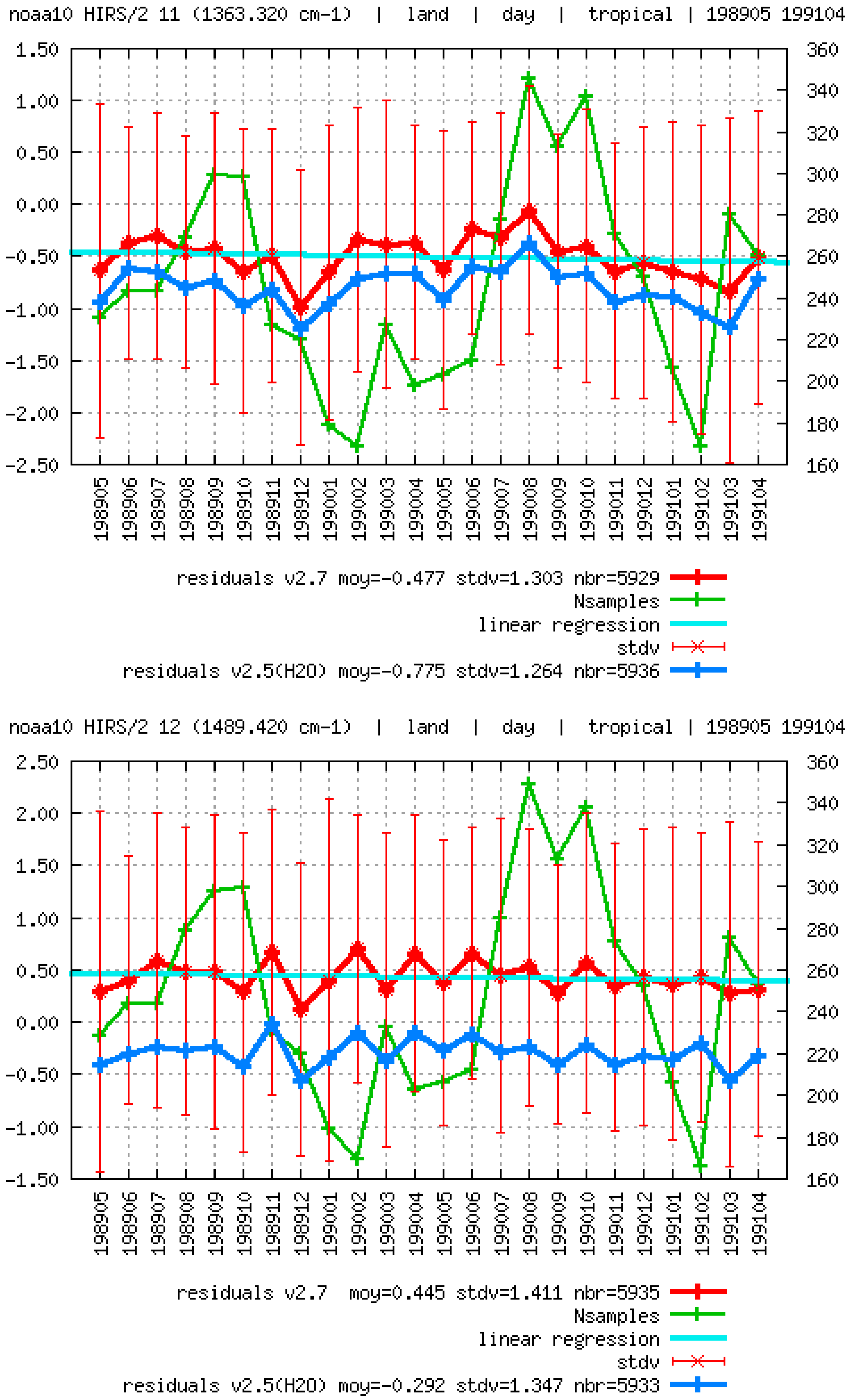
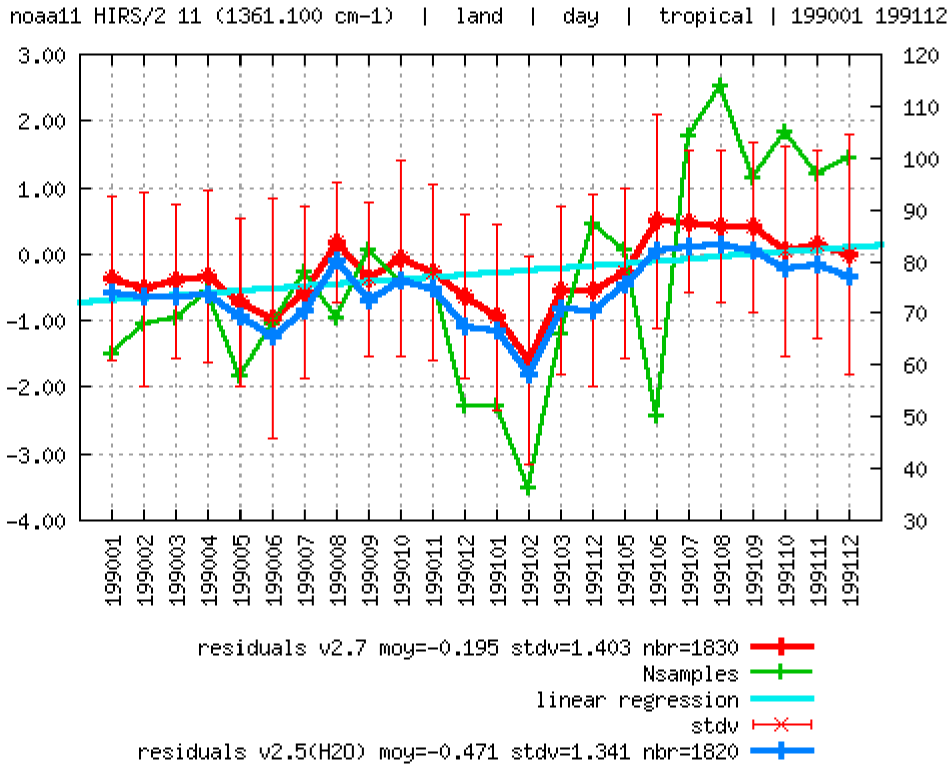


Figure 9-2: *Time series of residuals for NOAA-10. HIRS channel 11 (top), HIRS channels 12 (bottom).*



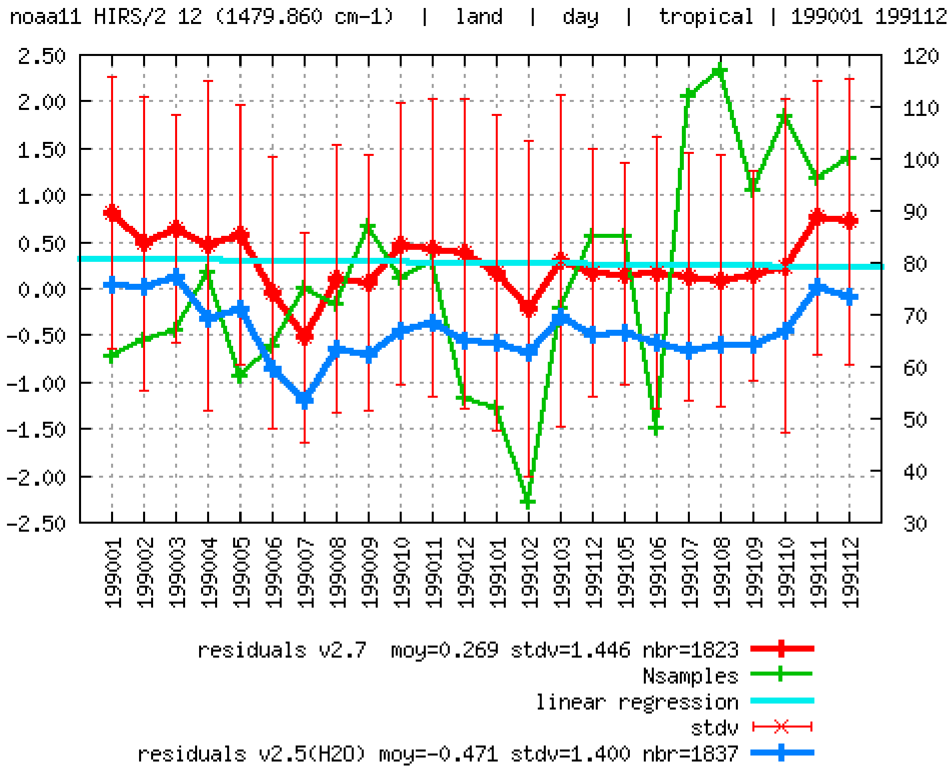
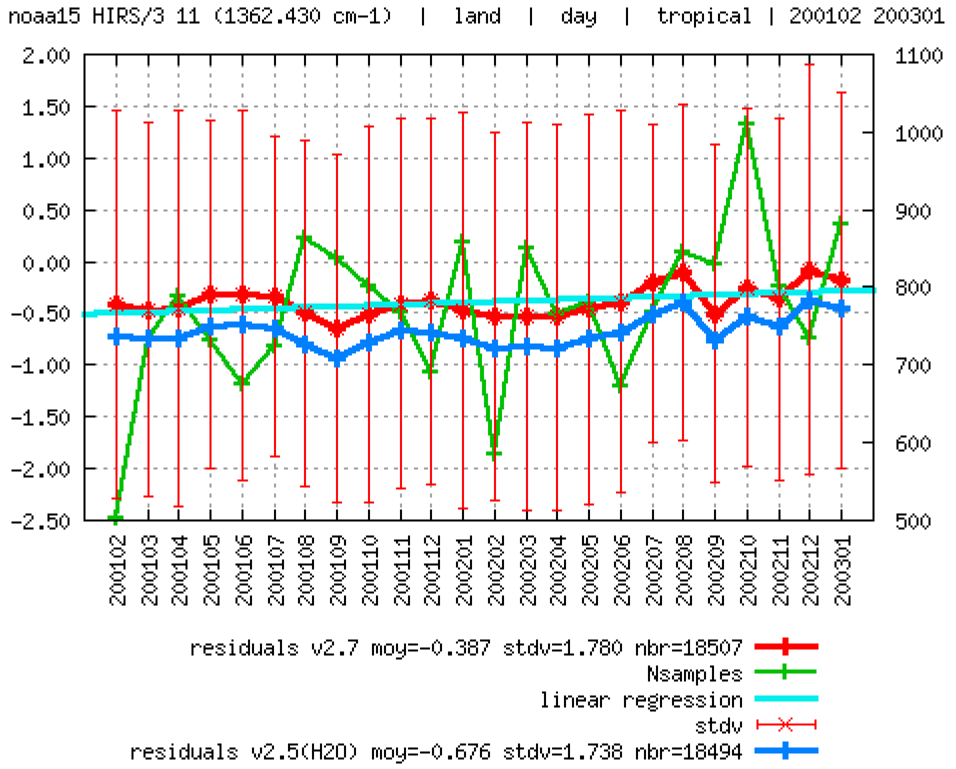


Figure 9-3: *Time series of residuals for NOAA-11. HIRS channel 11 (top), HIRS channels 12 (bottom).*



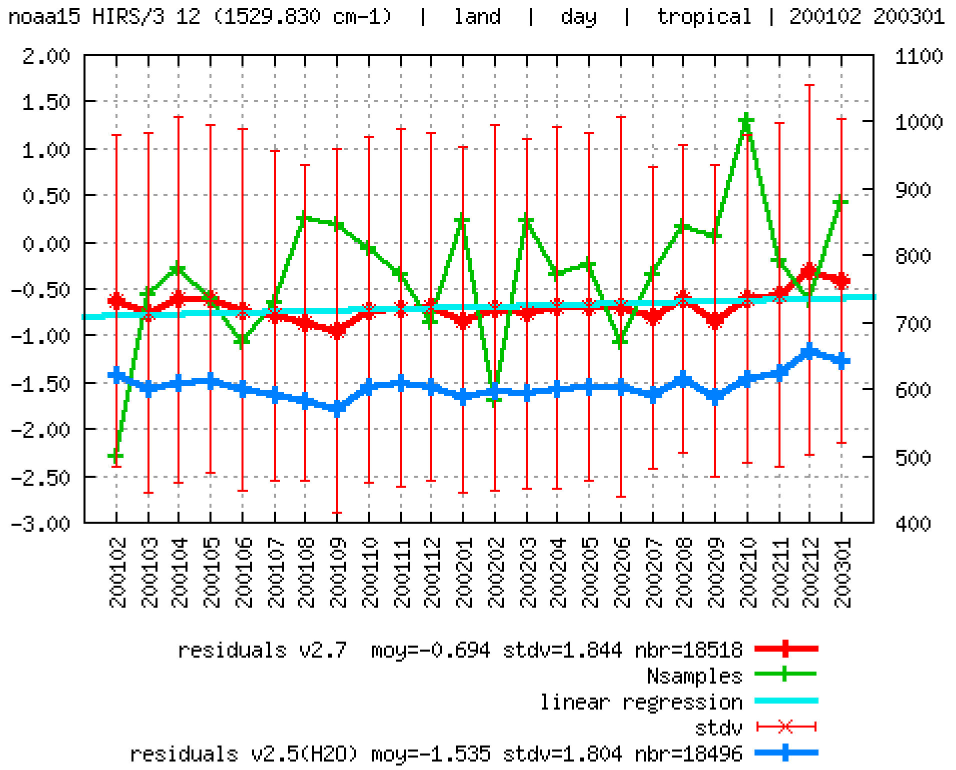


Figure 9-4: *Time series of residuals for NOAA-15. HIRS channel 11 (top), HIRS channels 12 (bottom).*

Table 9.2 displays statistics (biases, standard deviations, number of collocations) concerning the residuals of HIRS (Channels 11 and 12) and MHS (channels 3, 4, 5) onboard four satellites (NOAA10, NOAA11, NOAA15, MetOpA) for different cases of Land/Sea/Day/Night cases, for tropical atmospheres. These residuals have been generated with the 4A/OP forward model using as input i) the current version of the ARSA database (results in red) ii) a modified version of the ARSA database which does not include the empirical correction described in Section 4.4.4. These results are an extension of results presented in Section 4.4.4 for MetOpA and tropical situations only.

NOAA10 Residuals	Bias				Stdv				Nbr of collocations			
	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N
HIRS, CH. 11	-0.477	-0.683	-0.477	-0.696	1.303	1.423	1.262	1.231	5929	3203	5045	2507
HIRS, CH. 11	-0.775	-0.964	-0.790	-1.017	1.264	1.365	1.212	1.181	5936	3202	5046	2511
HIRS, CH. 12	0.445	0.201	0.282	0.011	1.411	1.536	1.339	1.335	5935	3210	5041	2515
HIRS, CH. 12	-0.292	-0.522	-0.466	-0.750	1.347	1.486	1.274	1.298	5933	3226	5057	2535
NOAA11 Residuals	Bias				Stdv				Nbr of collocations			
	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N
HIRS, CH. 11	-0.195	-0.276	-0.602	-0.353	1.403	1.442	1.550	1.334	1830	2012	2188	1343
HIRS, CH. 11	-0.471	-0.558	-0.932	-0.649	1.341	1.414	1.492	1.289	1820	2018	2192	1340
HIRS, CH. 12	0.269	0.332	-0.002	0.314	1.446	1.543	1.563	1.489	1823	2018	2181	1347
HIRS, CH. 12	-0.471	-0.396	-0.761	-0.488	1.400	1.477	1.484	1.411	1837	2026	2188	1347
NOAA15 Residuals	Bias				Stdv				Nbr of collocations			
	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N	L/D	L/N	S/D	S/N
HIRS, CH. 11	-0.387	-0.717	-0.532	-1.058	1.780	1.725	1.798	1.778	18507	6884	17118	7411
HIRS, CH. 11	-0.676	-1.003	-0.827	-1.358	1.738	1.683	1.772	1.740	18494	6878	17145	7421
HIRS, CH. 12	-0.694	-0.846	-0.833	-1.046	1.844	1.825	1.862	1.958	18518	6877	17145	7437
HIRS, CH. 12	-1.535	-1.659	-1.675	-1.859	1.804	1.782	1.828	1.936	18496	6870	17158	7449

Table 9.2: Comparison of the biases, standard deviations (Stdv) and number of collocations for two versions of HIRS channel 11 and HIRS channel 12 residuals of NOAA10 (top), NOAA11 (middle), NOAA15 (bottom). For each case Land/Day (L/D), Land/Night (L/N), Sea/Day (S/D), Sea/Night (S/N) values are given. All cases in this Table are for tropical situations. In red are the values of the

residuals for the current version of ARSA. In blue, results obtained with a version of ARSA which does not include the empirical correction of water vapor as described in Section 4.4.4.

Summary of the results

- Concerning IGRA Homogenized, its strongest advantage is the homogenization, - adjustment of the complete time series to remove the discontinuities – based on a careful analysis and a deep knowledge of the history of each radiosonde at global scale. To the best of our knowledge and based upon various graphs obtained during this study, the most recent period of time has been often selected to be the reference to which other period of time s have been adjusted. Such an adjustment "relative to" a given time series is not a guarantee that all the biases have been removed.
- Concerning ARSA, its main advantage is to give a complete, validated description of the atmosphere and the surface appropriate to be used in applications involving forward and inverse radiative transfer. Complete because, once severe quality controls performed on each raw radiosonde data, the gaps of these radiosonde measurements have been filled: above 300 hPa for water vapor, above 30 hPa for temperature, on the full pressure range for ozone. Validated because elaborated through the analysis of the impact of each modification on the brightness temperatures residuals (simulation-satellite observation) followed by an interactive adjustment of the water vapor vertical profiles.

For the assessment of the climate quality of either ARSA or IGRA_Homogenized water vapor product no final conclusions can yet been drawn.

- Concerning ARSA, because the current version starts from raw radiosonde measurements, not yet homogenized. However, the validation process of ARSA based on the analysis of the sea/land/day/night brightness temperature residuals remains a very powerful qualifier for climate applications, as demonstrated in the past with the detection of "jumps" in ozone profiles time series due to changes in the assimilation process of ozone data at ECMWF. Validations will here again take advantage of the vertical (associated to spectral) resolution of IASI as well as its unanimously recognized radiometric stability.
- Concerning IGRA_Homogenized and associated to its homogenization process itself which is *relative to* a given period of time, it remains to be checked - through e.g. an approach similar to the one which consists in simulating satellite observed brightness temperatures – that the biases are really removed.

10 Conclusion

For this study, 940 radiosonde stations have been retained from January 1979 (beginning of ARSA) to December 2010 (end of IGRA Homogenized). As a consequence, several types of radiosondes (Vaisala, VIZ, former USSR …), several periods associated to their evolution, several variables (temperature, water vapor …) are concerned.
This represents a total of several millions of collocated radiosonde reports/levels to be compared.
Several types of statistics, maps, graphs have been performed:

- concerning the behaviour of each database, in stand alone,
- concerning the inter-comparison of water vapor products and temperature products.

Comparing or inter-comparing products like water vapor with such a high 4-D (latitude, longitude, altitude, time) natural variability (high standard deviation at each level, for each station, along any period of time) is very difficult.

When these products are obtained i) from radiosondes, ii) from radiosondes of different types, iii) from different stations iv) over long period of times - inherently including changes in the techniques and performances of the radiosondes, the comparison becomes more and more delicate when it is necessary to generalize or to conclude.

The following remarks can be done on the sample of the selected 940 stations:

- It follows from our analysis of *spatial* coincident ARSA and IGRA Homogenized temperature and water vapor profiles data can substantially agree. Main reason is that both datasets rely upon the same raw radiosonde observations.
- Large discontinuities exist in the availability of daily measurements for a given station. These discontinuities which are clearly seen in the graphs displaying times series of temperature, water vapor, …- make problematic to quantitatively detect, study or compare temporal trends of each database either in standalone or relative to the other one or relative to the raw radiosonde archive.
- It was found that the number of common measurements to ARSA and IGRA become smaller and smaller when the height increases. This is due i) to the difficulty of radiosondes to deliver measurements of temperature and water vapor at the upper levels of the atmospheric column ii) to the fact that, in ARSA, any such gaps are filled by procedures to complete and extrapolate (above 300hPa as far as water vapor is concerned). This number of common measurements starts decreasing around the 300hPa pressure range that represents the greatest challenge to radiosonde relative humidity measurements.
- It comes as no surprise from our analysis of *vertical* coincident ARSA and IGRA Homogenized temperature and water vapor profiles data that the agreement between the two datasets is optimal in the lower part of the atmosphere (high pressure levels) and worsens as the altitude increases (see preceding topic). However, our results indicate that the significant discrepancies (mainly standard deviation) between the ARSA and IGRA Homogenized can be attributed i) to sampling effects along the vertical ii) interpolation/extrapolation effects for ARSA iii) homogenization according to the pressure level for IGRA_Homogenized.
- It follows from our analysis of *temporal* coincident ARSA and IGRA Homogenized temperature and water vapor profiles data may substantially agree within periods corresponding to the “reference period for homogenization”. Outside this “reference period of IGRA_Homogenized”, biases appear, inherently related to the fact that ARSA takes as input raw radiosonde (ie non homogenized) reports.

- Accordingly, it was also found that time series of temperature and water vapor may agree (within the limit of the characteristics of natural variability of each products) in certain period and not in others.
- It turns out very difficult to assess the stability of the two databases: in principle, IGRA_Homogenized is supposed to present the best stability due to its homogenization. However, the fact that homogenization seems to be performed relative to a “period of reference” is a guarantee that all biases have been removed, relatively to this period of reference, not in the absolute.

The following remarks can be done on the study of the residuals (Simulated_observed brightness temperatures):

As regards the impact of the ARSA database on the residuals, it was confirmed that the current version of ARSA (referred to as v2.7) contributes to decrease their biases: this turns out to be true not only in the case of MetOpA and for IASI whose observations were used for the validation of ARSA, but also for different satellites and instruments. If it is certainly premature to consider this as a proof of the stability of ARSA, on the other hand we can say that the empirical correction brought to the water vapor profiles of the ECMWF remains coherent for other instruments and other periods of time.

The following remarks can be done on “How to select representative and or reference sites?”

It is clear that the ideal “lot of reference stations” would answer the following criteria:

- having uninterrupted radiosonde flight series with the more stable and more reliable material,
- being global, or at least covering all types of air masses, as well as land/sea/day/night cases.

Our feeling is that the selection of the 940 stations made for this study is already a good pre selection. However, the number of measurements for each station is very unevenly distributed and this makes rather difficult for all of them the identifications of trends or biases through time series studies.
Our approach would be to select among the 940 stations which have participated to this study

- the ones which have the longer time series,
- the ones which have the highest number of measured levels between the surface and 100 hPa,
- the ones that offer the wider coverage of the earth surface.

Moreover, without a careful description of the “beginning-end” period chosen for the homogenization of IGRA_Homogenized – for each station and for each standard pressure level of this station-, such a representative subset of stations will not be fully useful for the purpose of studying the stability.

11 Recommendations

The recommendation that comes first is:

- Constantly improve the quality and the quantity of the radiosonde measurements over the world

Generally speaking:

- Overall, our results confirm, if necessary, the importance of continuing a dedicated program of homogenization of the radiosondes for applicability to operational radiosonde data on a global scale.
- Radiosonde continuity testing globally is a challenging if not insurmountable task. Encourage open community cooperation and results sharing. This is essential for inter-comparison and testing of various approaches to validate climate products directly or remotely obtained.

As a follow on to this work:

Our findings on the whole illustrate the complexity of the problem of comparison of two databases sharing the same basic information (raw radiosonde data) but intended for different uses/users: for this reason, and due to their inherent processing method, IGRA_Homogenized and ARSA (this latter, based on non homogenized raw radiosonde data) exhibit differences which by far exceed differences that one would expect with respect to raw radiosonde measurements.

Results obtained in the frame of this contract, give guidance but cannot be used to definitely determine which dataset yields a more accurate reference for water vapor.

Recommendations for further work on IGRA_Homogenized

- Continue this very important and very useful work.
- Continue distributing the results of the homogenization: add the information of the period chosen as reference.
- In the IGRA file, as it is distributed to users, make available the type of radiosonde used for each station.

Recommendations for further work at LMD on ARSA

At LMD devote more resources to research involving ARSA to properly reflect its relevant value for climate applications. In particular:

- Improve the vertical description of the atmosphere in the “near surface” pressure grid by e.g. taking into account the significant and intermediate pressure levels available in the radiosonde reports.
- Increase the vertical resolution of the ARSA pressure grid at these levels of the upper troposphere and lower stratosphere transition.
- As the ARSA database appears to be very well suited to improve on the high accuracy climate record from high resolution IR spectra begun with the IASI instrument, further examine a process of homogenization based on the remarkable radiometric stability of the

MetOpA/IASI instrument – and later on of the MetOp series B, C, IASI-NG - and the accuracy of the 4A/OP model.

- Perform the homogenization of the ozone profiles (although not really relevant to this study, see Section 4.4.5).

12 Deliverables: Files, Graphs, MetaData

The material used or generated in this study – as described in preceding Sections - will be made available via a ftp site to the PI of this QUASAR contract, together with a copy of the Final report (after revisions suggested at the occasion of the Final Presentation and by email).

All Files, Graphs, …, will be described in a Technical Memo made available to the PI of this study.

13 References

All references concerning the databases and model used by LMD for this study can be found at the following web address:

http://ara.abct.lmd.polytechnique.fr/index.php?page=publications

Specific to ARSA, see also:

http://ara.abct.lmd.polytechnique.fr/index.php?page=arsa

Specific to 4A at LMD see also:

http://ara.abct.lmd.polytechnique.fr/index.php?page=4a

and for the operational version of 4A called 4A/OP:

http://4aop.noveltis.com/

14 More on the ARSA database

14.1 ARSA variables description

Description of the variables

1(#)	'latitude of the radiosounding (degree)'
2(#)	'longitude of the radiosounding (degree)'
3	'date (year+month)of the radiosounding (yymm)'
4	'date (day+hour) of the radiosounding (ddhh)'
5	'date (minutes+seconds) of the radiosounding (mnss)'
6	'land/sea flag of the radiosounding (1➔land, 0➔sea)'
7(##)	'number of levels: "**nlevel**"'
8, 9, 10, 11	'pressure, temperature, water vapor specific humidity, ozone specific humidity' at level 1 of the atmosphere
12, 13, 14, 15	'pressure, temperature, water vapor specific humidity, ozone specific humidity' at level 2 of the atmosphere
……	Etc…
	'pressure, temperature, water vapor specific humidity, ozone specific humidity' at level "nlevel" of the atmosphere
nlevel*4+8	'surface temperature (K)'
nlevel*4+9	'surface pressure (hPa)'
nlevel*4+10	'cloud cover (%)'
nlevel*4+11	'raw altitude of the radiosounding (m)'
nlevel*4+12	'WMO station number' – see Appendix 8 -.
nlevel*4+13	'WMO block number' – see Appendix 8 -.
nlevel*4+14	'latitude of the surface station (degree)'
nlevel*4+15	'longitude of the surface station (degree)'
nlevel*4+16	'hour of the surface temperature record'
nlevel*4+17	'distance between the radiosounding and the surface station (< 300 km)'
nlevel*4+18	'hours between the radiosounding and the surface station (< 3 hours)'
nlevel*4+19	'day/night index of the radiosounding (1➔day, 0➔night)'
nlevel*4+20	'day/night index of the the surface station (1➔day, 0➔night)'
nlevel*4+21	'altitude of the radiosounding station from the topography database (m)'
nlevel*4+22	'altitude of the surface station from the topography database (m)'
nlevel*4+23	Identification of the pressure level corresponding to the last (highest altitude) measured value of the temperature profile
nlevel*4+24	Identification of the pressure level corresponding to the last (highest altitude) measured value of the water vapor profile

(#) longitude : -180°,180° (west to east)
latitude : -90°, 90° (south to north)
Time information is GMT

(##) in the current version, the maximum value taken by nlevel is 43. (It used to be 40 in the preceding version of ARSA). This value is obviously adjusted to the true value of the surface pressure given by the radiosounding. See the table below, which gives the 4A pressure grid.

14.2 The ARSA 43-level Pressure grid

As stated at the end of the preceding Section 15.1, the maximum number of points on the pressure grid is 43. The corresponding pressure values are given in the Table below. It has to be noticed that, for every profile found in the ARSA database, this nominal value of 43 is adjusted in order to take into account the surface pressure of the considered radiosonde measurement.

ARSA level number	pressure (hPa)		ARSA level number	pressure (hPa)
43	1013.25		20	56.46
42	955.12		19	45.73
41	900.33		18	37.04
40	848.69		17	24.79
39	800.00		16	16.60
38	724.78		15	11.11
37	651.04		14	7.43
36	584.80		13	4.98
35	525.00		12	3.33
34	471.86		11	2.23
33	423.85		10	1.50
32	380.73		9	1.00
31	341.99		8	0.55
30	307.20		7	0.30
29	275.95		6	0.17
28	247.87		5	0.09
27	222.65		4	0.05
26	200.00		3	0.024
25	161.99		2	0.0089
24	131.20		1	0.0026
23	106.27			
22	86.07			
21	69.71			

14.3 The naming conventions of the various ARSA data and metada

As stated above in Section 3.2, several by-products exist and their naming convention is as follows:

ARSA_yyyymmdd1_yyyymmdd2_type.data

With :

yyyymmdd1 characterizing the beginning of the considered period
yyyymmdd2 characterizing the end of the considered period
yyyy=year,
mm=month,
dd1= first day of the period
dd2= last day of the period

and

type =

- ara_QC_rej (for rejected raw radiosonde reports),
- ara_QC_kept (for kept raw radiosonde reports),
- ara_QC_kept_to_4A (for kept radiosonde reports, which are inter/extrapolated on a multi-level pressure grid between surface and 0.0026hPa)

Example for this latter type:
ARSA_20090401_20090430_ara_QC_kept_to_4A.data
For April 2009 1st to 30th period.

14.4 Reading the ARSA files

The ARSA database is in binary (unix big endian convention). Therefore, on a linux system, a "byteswap" option is required when compiling:
a. For pgf90, you can use "-byteswapio"
b. For ifort, you can use "-convert big_endian"

Name of the fortran source: lec_arsa_ara_QC_kept_to_4A.f90
Compiling : pgf90 -fast -byteswapio -o lec_arsa_ara_QC_kept_to_4A lec_arsa_ara_QC_kept_to_4A.f90
Executing : lec_arsa_ara_QC_kept_to_4A /nomdir/ARSA_yyyymmdd_yyyymmdd_ara_QC_kept_to_4A.data

Logical Unit of the ARSA input file: 1
Logical Unit of RAOB (RAdiosonde OBservation) station list input file: 10
Logical Unit of the output file (in ASCII) containing – as an example – P, T, H2O and O3 profiles of the first and last profile of the ARSA datafile extracted by the user: 21
Logical Unit of the output file containing information on the corresponding RAOB stations: 20

Logical unit	Name	Input/ output	Access	ASCII
1	ARSA_20080801_20080831_ara_QC_kept_to_4A.data	input	direct	no
10	liste_stations.txt	input	sequential	yes
20	out_"name_of_file_1".txt	output	sequential	yes
21	out_"name_of_file_1"_profiles.plt	output	sequential	yes

NB: In this example, name_of_file_1 =
ARSA_20080801_20080831_ara_QC_kept_to_4A.data.

14.5 Day/Night Index

If the solar zenith angle is greater than or equal to 90° then the day / night index is set to “0” (night).
If the solar zenith angle is smaller than 90° then the day / night index is set to “1” (day).
As stated in Section 3, RAOB are “time and space” collocated with the surface station archive information of the ECMWF Data Server. This explains why two day/night index values are usually found in the ARSA database: one for the RAOB, the other one for the surface. In case collocation was impossible, then the day/night index for surface is set to default value -999.0.

14.6 Land/sea index and Elevation

The RAOB altitude is the altitude at the corresponding station. As stated above, RAOB are “time and space” collocated with the surface station archive information of the ECMWF Data Server. The US Navy 1/6° resolution database gives the elevation and the percentage of water at the latitude and longitude of the RAOB and of its collocated surface station. In case the collocation was impossible (too large a difference in altitude or in percentage of water), then the altitude of the surface station is set to the default value of -999.0.

14.7 The radiosonde stations list

Given below is an example of the information found in the radiosonde stations list, elaborated from the nominal WMO station list (as in 2008).
The corresponding ASCII file name is: “liste_stations.txt”.
In this file, the station identification number is represented by the so-called 5-digit WMO index number “IIiii”, composed of the block number (II) and the station number (iii).

Station index Number	Name of Location	Latitude	Longitude	Elevation in Meters
60369	Alger-Port	36.46N	03.06E	12
60387	Dellys	36.55N	03.57E	8
60390	Dar-El-Beida	36.41N	03.13E	29
60395	Tizi-Ouzou	36.42N	04.03E	189
60401	Bejaia-Port	36.45N	05.06E	4
60403	Guelma	36.28N	07.28E	228
60405	Bouchegouf	36.30N	07.43E	111
60410	Tenes	36.30N	01.20E	18
60415	Ain-Bessam	36.19N	03.32E	0
…	…	…	…	…

In this Table, Latitude and Longitude are given in “degrees**.**minutes”. For example, the latitude of station 60390 (Dar-El-Beida) is 36°41’ and its Longitude is 03°13’.
User has to be aware that, in the ARSA files as described in Section 15.1, the latitude and longitude are given in decimal units, as is done in the ECMWF nominal files. For example, for this same station 60390 (Dar-El-Beida) we find in the ARSA records: 36.68 for the latitude and 3.22 for the longitude.

14.8 References

This work has been performed using the satellite archive and the radiative transfer tools derived, maintained and regularly validated at LMD: the forward radiative transfer model 4A; the spectroscopic database GEISA, the climatic database, TIGR, and high spectral resolution infrared emissivity databases. Description of each of these tools, databases or satellite archive, and a list of related references in the open literature, may be found at the following address: http://ara.abct.lmd.polytechnique.fr/

15 More on the IGRA_Homogenized database

Header Record Format:

Variable Name	*Columns Description*
Header Record Indicator	1- 1 # character
Station Number	2- 6 WMO station number
Year	7- 10
Month	11- 12
Day	13- 14
Observation Hour	15- 16 00-23 UTC
Release Time	17- 20 0000-2359 UTC, 9999 = missing
Number of levels	21- 24 number of subsequent data records
PW surface-500mb	25- 36 in mm
PW 500-300mb	37- 48 9999 = missing
PW 300-100mb	49- 60

Data Record Format:

Variable Name	*Columns Description*
Major Level Type	1- 1 1 = standard pressure level 2 = significant thermodynamic level 3 = additional wind level
Minor Level Type	2- 2 1 = surface, 2 = tropopause, 0 = other
Pressure	3- 8 units of Pa (mb * 100)
Pressure Flag	9- 9 A, B, or blank (see note 2 below)
Geopotential Height	10- 14 units of meters
Geopotential Height Flag	15- 15 A, B, or blank (see note 2 below)
Temperature	16- 21 units of 0.01degrees C
Temperature Flag	22- 22 A, B, or blank (see note 2 below)
Dewpoint Depression	23- 28 units of 0.01degrees C
Wind Direction	29- 34 units of degrees (0-360, inclusive)
Wind Speed	35- 40 units of (m/s)*10
Specific Humidity	41- 48 units of 0.00001 g/kg
Relative Humidity	49- 54 units of 0.01%

Notes:

1. Cases with -8888 indicate the values were bad and were removed.

2. For each pressure, geopotential height, and temperature value, a one-character quality assurance flag indicates whether the corresponding value was checked by procedures based on climatological means and standard deviations. Possible flag values are: blank = no climatological check applied due to an insufficient number of data values for computing the relevant statistics, A = value falls within "tier-1" climatological limits based on all days of the year and all times of day at the station, and B = value passes checks based on both the tier-1 climatology and a "tier-2" climatology specific to the time of year and time of day of the data value.

16 Acknowledgements

This work has been performed within the frame **QUASAR** (**QU**ality **A**ssessment of **SA**tellite and **R**adiosonde data, SAF_CM_CDOP2_nnn_Quasar_v2.4k). Quasar was funded through the visiting scientist scheme of EUMETSAT's Satellite Application Facility on Climate Monitoring. EUMETSAT member states are kindly acknowledged for their financial support of this VS study. This work was dedicated to support CM SAF and the GEWEX water vapor assessment.

Many thanks to Marc Schröder for suggesting this cooperation and for very constructive discussions throughout this contract.

The IGRA Radiosonde data were obtained from Imke Durre (IGRA) and June Wang (IGRA_Homogenized, HOMORS92). We will be very happy to send and discuss the results obtained during this study and hopefully share experience with the IGRA Team, as soon as the present report is available in final version.

ARSA data as well as satellite data archive are from the ARA/ABC(t)/LMD group. From inputs to very fruitful discussion, ARSA has been elaborated with the help of A. Chédin, L. Crépeau, J. Pernin, R. Armante, V. Capelle, C. Crevoisier.

The LMD multi year satellite data (from TIROS-N to NOAA-14) including the TOVS and ATOVS instruments onboard the NOAA operational satellite system have been obtained within the frame of the NOAA/NASA Pathfinder Programme and later on from the Comprehensive Large Array-data Stewardship System (CLASS), the electronic library of NOAA. The MetOp IASI, AMSU, MHS and HIRS data are taken (antenna and NRT acquisition) from Ether, the IPSL Thematic expertise group for atmospheric chemistry, via the system EUMETCAST. Many thanks to NOAA/CLASS, NASA/GSFC, Eumetsat and Ether for archiving and making these data available to us.

LMD computing and archiving resources for this study are from IDRIS (Institut du développement et des ressources en informatique scientifique) the main CNRS Computer Center, from the ECMWF Computer Center and from ClimServ our local computer system.

LMD wants to warmly thank ECMWF for making available the ERA-Interim outputs, the radiosonde archive and the surface station archive through the ECMWF Data Server.