Impact of PCA on the assimilation of hyperspectral infrared sounder data in the frame of the AROME mesoscale convection-permitted NWP model

Javier Andrey-Andrés, Nadia Fourrié, Vincent Guidard
CNRM, Météo-France and CNRS
March 6th, 2016

Outline

Plan

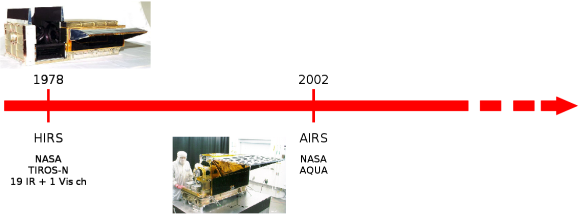
Evolution of IR sounders

Instrument	Year	Sat.	Orbit	Channels	Spec. resolution

Evolution of IR sounders

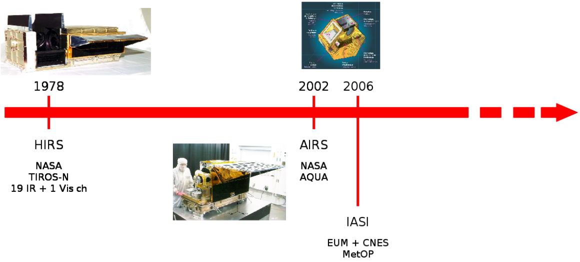
Instrument	Year	Sat.	Orbit	Channels	Spec. resolution
HIRS	1978	TIROS-N		19 + 1	3 to 40 cm^{-1}

Evolution of IR sounders



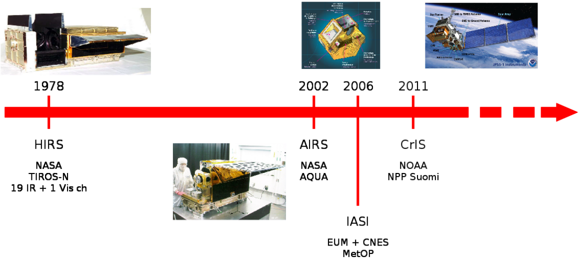
Instrument	Year	Sat.	Orbit	Channels	Spec. resolution
HIRS	1978	TIROS-N		19 + 1	3 to 40 cm^{-1}
AIRS	2002	Aqua	LEO	2,378	0.25 to 1 cm^{-1}

Evolution of IR sounders



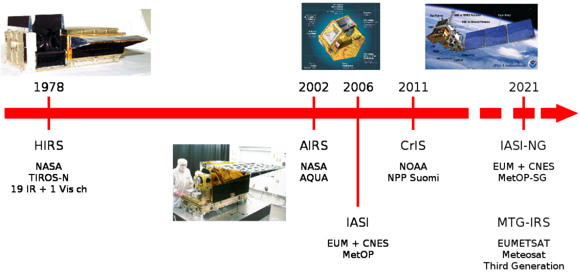
Instrument	Year	Sat.	Orbit	Channels	Spec. resolution
HIRS	1978	TIROS-N		19 + 1	3 to 40 cm^{-1}
AIRS	2002	Aqua	LEO	2,378	0.25 to 1 cm^{-1}
IASI	2006	Metop	LEO	8,461	0.25 cm^{-1}

Evolution of IR sounders



Instrument	Year	Sat.	Orbit	Channels	Spec. resolution
HIRS	1978	TIROS-N		19 + 1	3 to 40 cm^{-1}
AIRS	2002	Aqua	LEO	2,378	0.25 to 1 cm^{-1}
IASI	2006	Metop	LEO	8,461	0.25 cm^{-1}
CRIS	2011	JPSS	LEO	1,305	0.625 cm^{-1}

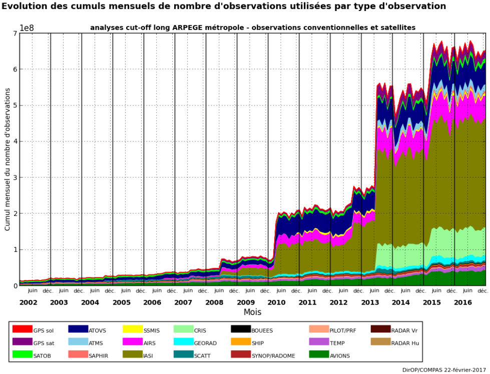
Evolution of IR sounders



Instrument	Year	Sat.	Orbit	Channels	Spec. resolution
HIRS	1978	TIROS-N		19 + 1	3 to 40 cm^{-1}
AIRS	2002	Aqua	LEO	2,378	0.25 to 1 cm^{-1}
IASI	2006	Metop	LEO	8,461	0.25 cm^{-1}
CRIS	2011	JPSS	LEO	1,305	0.625 cm^{-1}
IASI-NG	2021	Metop-SG	LEO	16,921	0.125 cm^{-1}
IRS	2021	MTG	GEO	1,738	0.625 cm^{-1}

Use of hyperspectral IR data in NWP models

- Data from IR sounders are the most used by NWP models in terms of number
- This tendency will continue with the arrival of new hyperspectral IR sounders



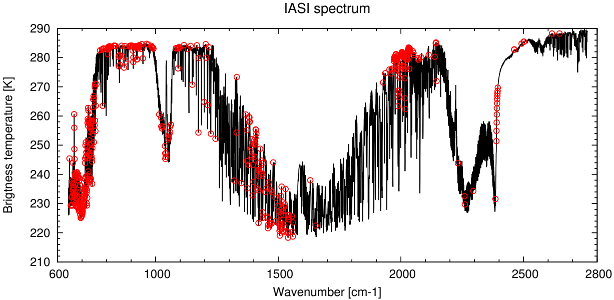
	IASI	MTG-IRS
Spectral sampling	0.25 cm^{-1}	0.625 cm^{-1}
Samples per spectrum	8,461	1,808
Spatial sampling at nadir	12 km	4 km
Samples per hour	54,000	8.0 10^6
Estimation of data volume	0.92 GB/h	28 GB/h

(Atkinson, 2013)

Inpact of Hyperspectral IR data

- Atmospheric profiling errors are improved
- More chemical compounds can be profiled

- Data dissemination becomes impossible (costs) and data storage needs explose
- Inter-channel redundancy becomes more and more important. NWP center keep just 500 IASI channels from the 8461



There is a need to reduce the data volume. . .

Why not to compress the data?

Compression types definitions (from Atkinson, 2013):

- Lossless
  - Exact reconstruction of the input (with machine precision)
- Near-lossless
  - Input reconstruction with a maximum defined error
  - Error typically a defined (small) fraction of instrument noise
  - Example: digitisation error (or quantisation error)
- Lossy
  - e.g. compression algorithms for images (jpeg, etcetera)
  - e.g. Principal Components Analysis technique for hyperspectral sounders

For IASI, the best performances were obtained using PCA plus residuals quantisation

The PCA compression technique

PCA allows the reduction of the dimensionality of a problem by examining the linear relationship between all the variables contained in a multivariate dataset

- The original set of correlated variables, y^{obs}, is replaced by a smaller number of uncorrelated variables called principal component scores (PCS, x^{pcs}). E corresponds with the eigenvectors matrix:

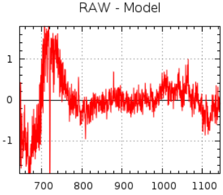
$$x^{pcs} = E * y^{obs}$$

- To return to the original space it is only need to make the following multiplication:

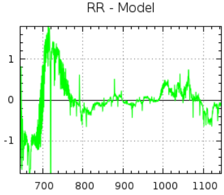
$$y^{pcs} = E^T * x^{pcs}$$

- These new variables retain most of the information contained in the original dataset (most of the gaussian noise is filtered):

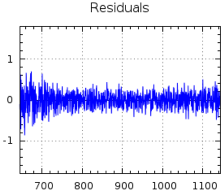
$$y^{obs} \quad = \quad E * x^{pcs} + residuals \quad = \quad y^{pcs} + residuals$$



=



+



IASI PC scores and quantised residuals

From T. Lee and S. Bedford (2004), slide from N. Atkinson (2013)

It is a lossy compression, but most of the "loss" is noise

1. Noise-normalised radiance: $y = \frac{r - y_0}{n}$

   y Normalised radiance
   r Observed spectrum
   n Noise
   y_0 Mean radiance

2. PC Score (integer): $s = NINT(\frac{E^T y}{f_s})$

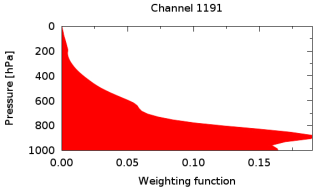
   s PC score
   E^T Eigenvectors matrix transposed
   f_s Quantisation factor, typically 0.5

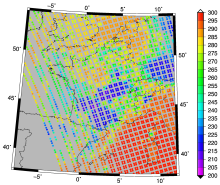
3. Residual (integer): $\Delta y = NINT(\frac{y - f_s E s}{f_r})$

   f_r Noise quantisation factor, typically 0.5. Gives 1% noise increase

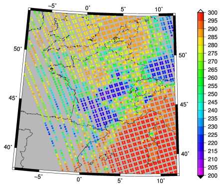
Example of PCA compression

IASI channel 1191, @942.5 $\text{cm}^{-1} \Rightarrow$
Surface sensitivity channel

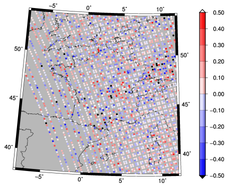




(a) Raw radiances



(b) Raw radiances

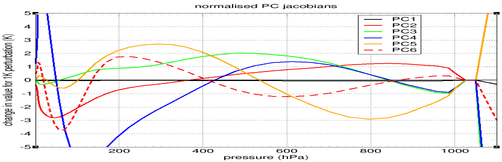


(c) Raw radiances

How can we assimilate PCA compressed data

1. We can use reconstructed radiances from PCs. . .
   - + No much work to adapt current assimilation systems
   - + Channel noises are filtered by PCA
   - - Interchannel correlations are heavily increased

2. We can assimilate PCs directly
   - + We can use all the information registered in the observation
   - - More difficult to understand. PCs are a mathematical representation
   - - Some PCs Jacobians present structures peaking low and high in the atmosphere ⇒ What happen for low-top models?



(Figure from McNally (2013))

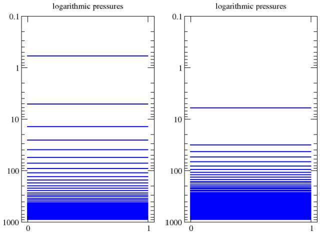
Plan

AROME main figures

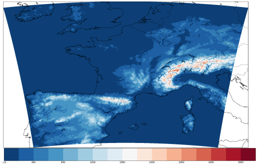
Upgrade from 04/2015

- Increase of vertical resolution in lower atmosphere
- Increase of the number of radar observations
- Bias correction for satellite observations not modified (taken from ARPEGE global model)

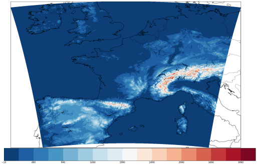
AROME version	Old	OPER	EXP
Mesh grid	2.5 km	1.3 km	1.3 km
Assim. cycle	3 h	1 h	3 h
Levels	60	90	90
Model top	1 hPa	10 hPa	10 hPa
Levels $< 2\ km$	21	33	33
Lowest model lev.	10 m	5 m	5 m



(a) AROME vertical resolutions



(b) Old AROME domain



(c) New AROME domain

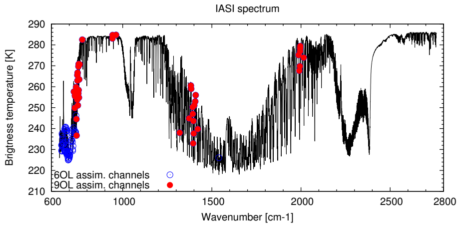
AROME main figures

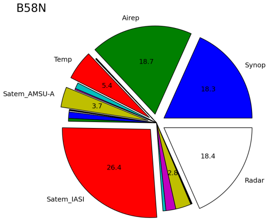
Weight of IASI observations and data usage

- Weight of different observation types in AROME versions:

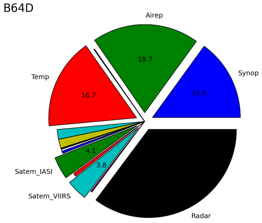
Obs	Old [%]	OPER [%]
All Sats.	37.7	11.3
IASI	26.5	4.2
Radar	18.4	36.7
RS	5.4	16.0

- Additionally satellite channels with contributions above the model top need to be removed
- The number of assimilated IASI channels have decreased from 123 to 44 channels





(a) Old version



(b) OPER version

ARPEGE main figures

Upgrade from 04/2015

Horizontal resolution: T1198c2.2 (7.5km over France, 36km over antipodes)

Vertical: 105 levels

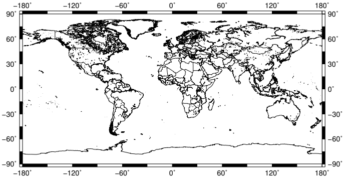
1st Minimization: T149c1L105 ($\sim$135km) with 40 iterations (T107c1L70 with 25 it)

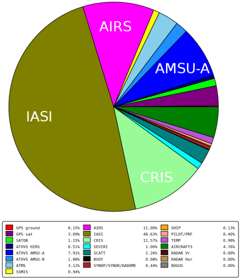
2nd minimization: T399c1L105 ($\sim$50km) with 40 iterations (T323c1L70 with 30 it)

Timeslots of 30 minutes (vs $\sim$1 hour)

Assimilation of SSMIS sounding channels, more radiances in the screening (+10% assimilated observations), use of swath ATMS data, preparation of the SAPHIR data assimilation.

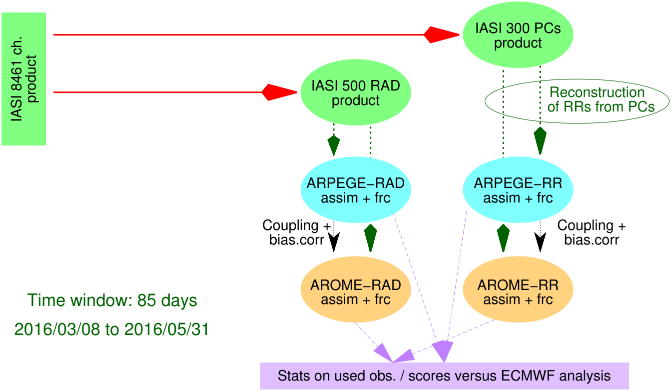
Provides lateral coupling conditions and satellite bias corrections to AROME regional model





Plan

Schema of experiments, period of study



Time window: 85 days

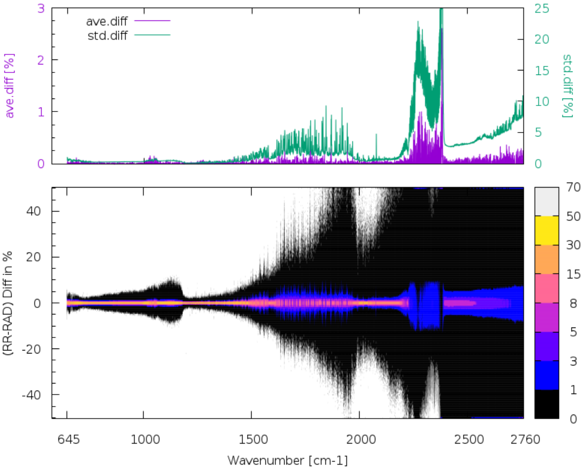
2016/03/08 to 2016/05/31

** Same scheme for IASI diagonal and full error matrix experiments*

Plan

Differences between 'RAD' and 'RR' radiances

From 8461 channels and PCs EUMETSAT products



$$dif = \left[\frac{RR - RAD}{RAD} - 1\right] \times 100$$

Calculated using a full IASI single orbit (91,800 obs.)

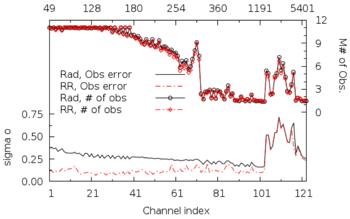
IASI orbit: 20130805, from 20:08:58 to 21:50:58

Three IASI bands well observed, bigger noise at the beginning of each band

PCA compresison removes the *ghost effect* observed in CO_2 IASI band 3, between 2,200 and 2,400 cm^{-1}

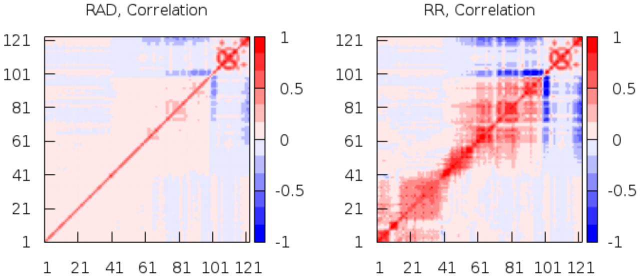
Results from ARPEGE assimilation experiments

IASI interchannel correlation



(a) Error and obs. number

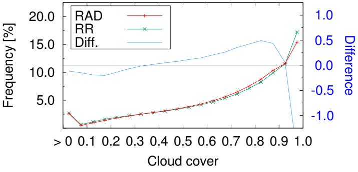
- Error matrices computed using Desroziers methodology
- Similar number of observations of both RAD and RR matrices
- IASI error is sensibly reduced but only for temperature and surface channels



(b) Desroziers error matrices

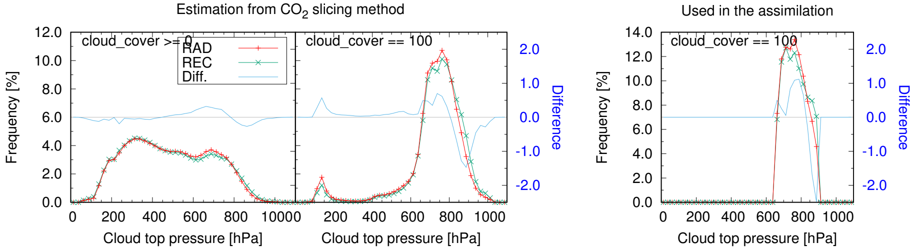
Results from ARPEGE assimilation experiments

Cloud detection from CO2 slicing method, sea observations



(a) Cloud cover

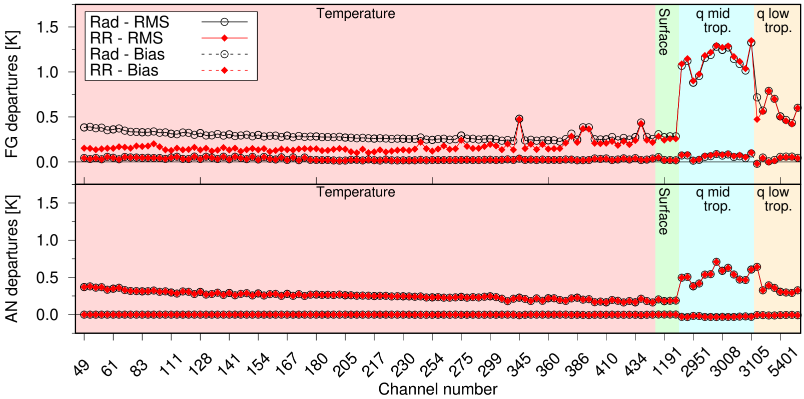
- Small but negligible differences in cloud cover retrievals from CO_2 slincing method
- Similar results for cloud top pressure retrievals either in CO_2 slicing diagnostics or used in model data assimilation



(b) Cloud top pressure

Results from ARPEGE assimilation experiments

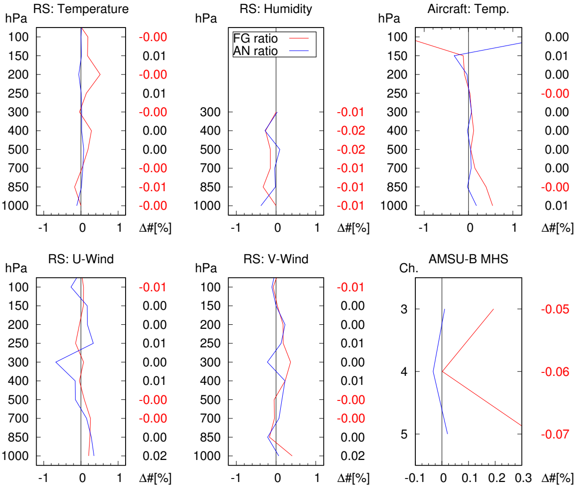
Statistics on IASI used observations



- Lower rms error for temperature channels, not for humidity ones in First Guess (FG) departures
- Very close results for Analysis Departures (AN)

Results from ARPEGE assimilation experiments

Statistics on used observations



$$ratio = \left[\frac{rms_{RR}}{rms_{RAD}} - 1 \right] \times 100$$

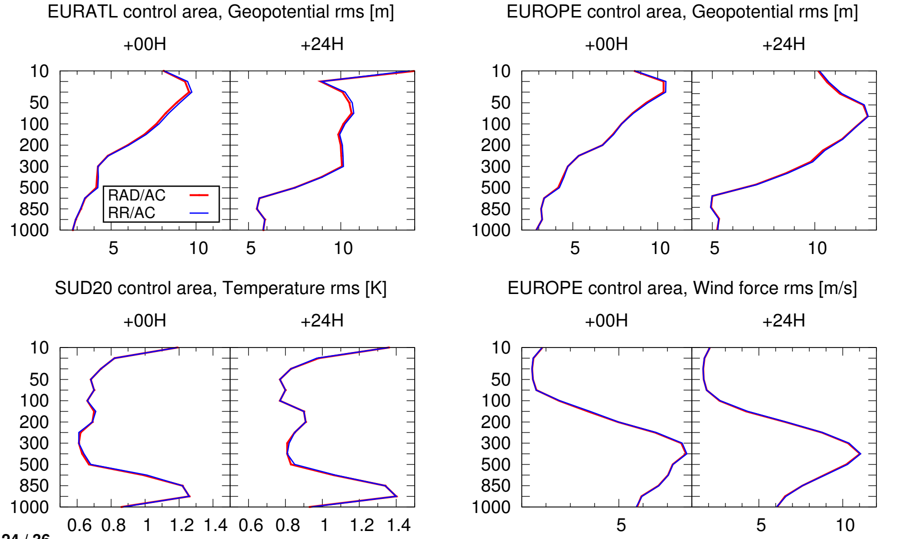
- Negative values mean an improvement
- Positive values point a degradation of the system
- Columns of numbers indicate the cariation in percentage of the observations number used

Similar results for both *RAD* and *RR* experiences

Negligible differences in used observation number

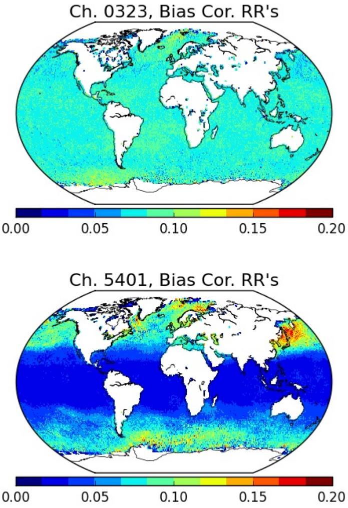
Results from ARPEGE assimilation experiments

ARPEGE scores against ECMWF analysis



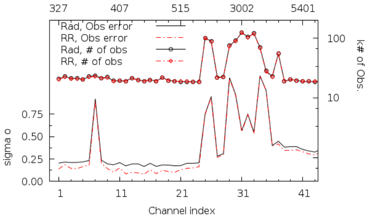
Results from ARPEGE assimilation experiments

Differences in bias corrections



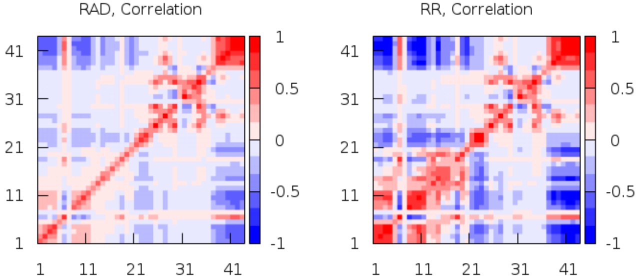
Results from AROME assimilation experiments

IASI interchannel correlation



(a) Error and obs. number

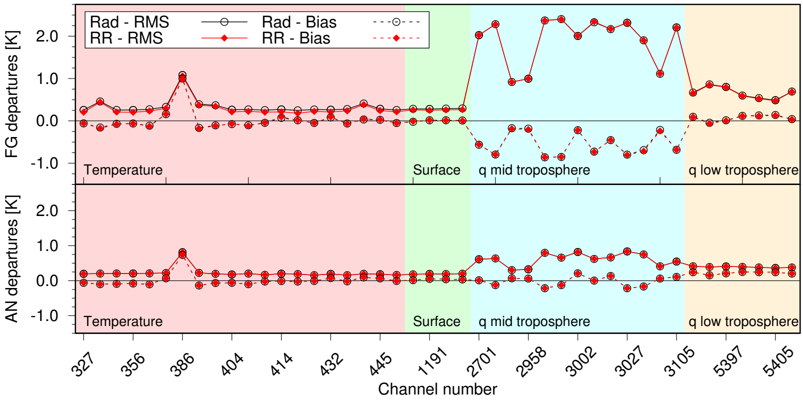
- Error matrices computed using Desroziers methodology
- Similar number of observations of both RAD and RR matrices
- IASI error is sensibly reduced but only for temperature and surface channels



(b) Desroziers error matrices

Results from AROME assimilation experiments

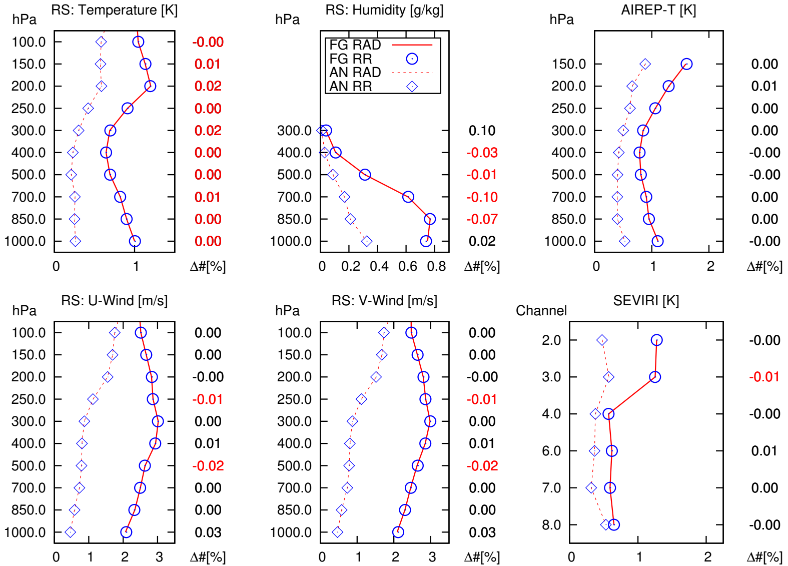
Statistics on IASI used observations



- Lower rms error for temperature channels, not for humidity ones in First Guess (FG) departures. Very close results for Analysis Departures (AN)
- Reduction of FG rms between *RAD* and *RR* is smaller that the observed for ARPEGE experiments

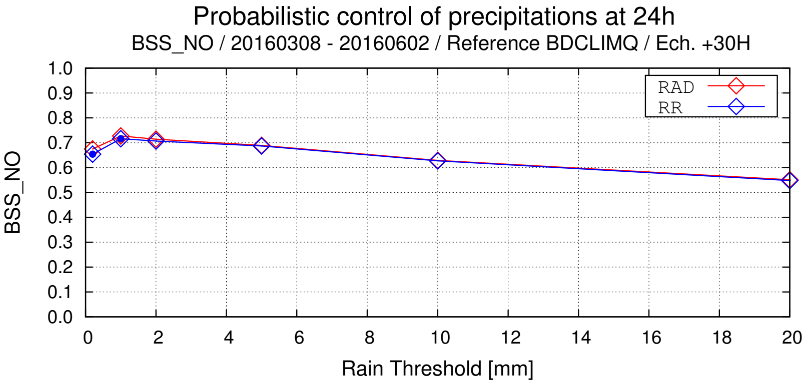
Results from AROME assimilation experiments

Statistics on used observations, First Guess (FG) and Analysis departures (AN)



Results from AROME assimilation experiments

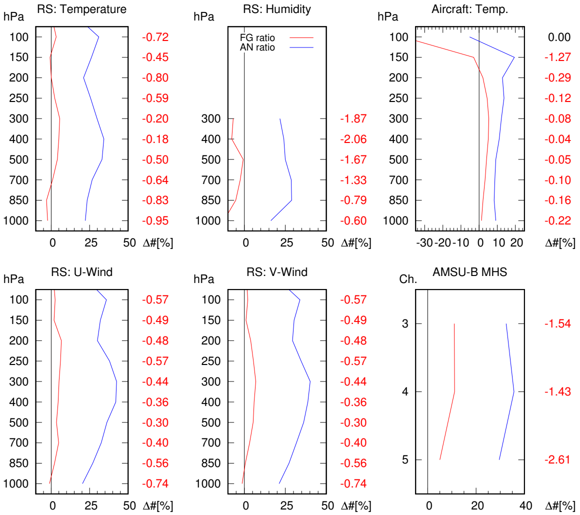
Precipitation scores for 20.6 km neighbourhood



- Differences are significant only for small rain thersholds, but very small
- From a global point of view, *RAD* and *RR* data provide equivalent results

Using a full error matrix in ARPEGE

Reconstructed radiances. Full vs diagonal error matrix



$$ratio = \left[\frac{rms_{RR}}{rms_{RAD}} - 1\right] \times 100$$

- Negative values mean an improvement
- Positive values point a degradation of the system
- Columns of numbers indicate the cariation in percentage of the observations number used

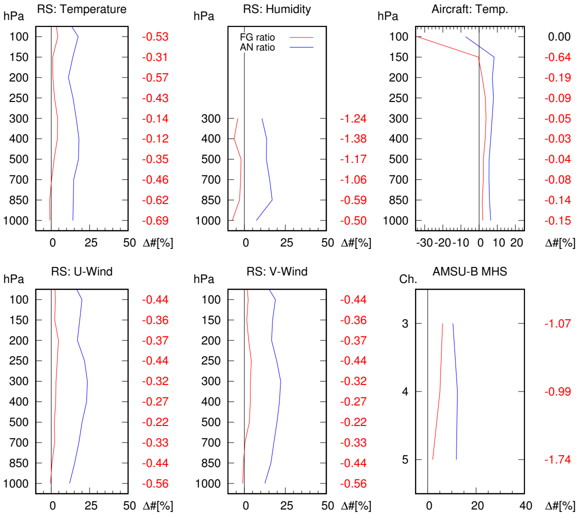
Degradation of the system and reductions in the used observations number

Need of adjust observation error weigths

ECMWF: Best results for
$R = 1.65 \times R_{desroziers}$

Using a full error matrix in ARPEGE

Full IASI error matrix. Preliminary RR vs RAD obstats



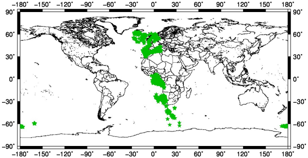
$$ratio = \left[\frac{rms_{RR}}{rms_{RAD}} - 1\right] \times 100$$

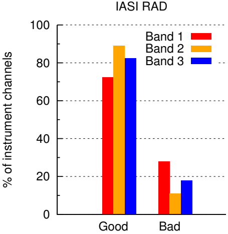
- Negative values mean an improvement
- Positive values point a degradation of the system
- Columns of numbers indicate the cariation in percentage of the observations number used

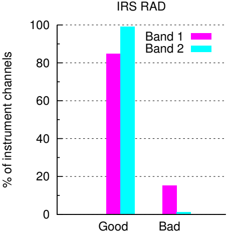
Degradation of the system and reductions in the used observations number

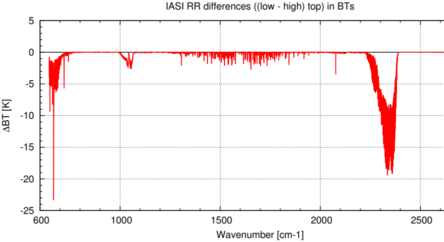
Worst results on RR because of the higher correlation

Impact of having a low-top model









Plan

Take home messages

- RAD/RR differences are small in general terms. Bigger differences in the last IASI band, where IASI is more noisy.
- MF NWP models present no significant impact when using RAD or RR radiances. No significant differences in cloud detection, obstats, scores or rain scores
- AROME bias correction needs to be re-adapted for all satellites because of having lowered the model top
- RAD and RR radiances give different results when using a full error matrix. First examples from MF NWP model are given. There is need of tuning the observation errors in NWP MF models before using a full error matrix for IASI
- Low-top model constraint has a huge impact in the simulation of the IASI radiance or PCs. Lower impact for IRS. IASI PC are heavily affected by the low-top contraint, around 50 PC of the first 100 present statistically different values when using low- or high-top models. Need to adapt the PC eigenvectors rejecting bad channels or developing a top coupling for this kind of models

Acknowledgements

- Thanks to EUMETSAT for funding my fellowship
- Also thanks to Météo France and my colleagues for their help and support xhich has made this work possible