Preparing the assimilation of IASI-NG in NWP models: a first channel selection

Francesca Vittorioso, Vincent Guidard, Nadia Fourrié, J Andrey-Andrés

Preparing the assimilation of IASI-NG in NWP models: a first channel selection

F. Vittorioso, V. Guidard, N. Fourrié, J. Andrey-Andrés

CNRM UMR 3589, Météo-France & CNRS, GMAP/OBS, Toulouse, France

Abstract

As the EUMETSAT Polar System-Second Generation (EPS-SG) is being prepared, a new generation of the IASI instrument has been designed. The IASI New Generation (IASI-NG) will measure at a spectral resolution and signal-to-noise ratio improved by a factor 2 compared to its predecessor. Measurement precision will be improved as well.

The high amount of data resulting from IASI-NG will present many challenges in the areas of data transmission, storage and assimilation and the number of individual pieces of information will be not exploitable in an operational Numerical Weather Predictions (NWP) context. For these reasons, an appropriate IASI-NG channel selection is needed, aiming to select the most informative channels for NWP models.

In this paper the preparatory studies towards the channel selection will be displayed, as well as methodology and results concerning a very first channel selection so far obtained.

1 Introduction

The hyperspectral Infrared Atmospheric Sounding Interferometer (IASI), key payload element of the European Meteorological Operational Satellites (MetOp) series, provides since 2007 a huge contribution to Numerical Weather Prediction (NWP), pollution monitoring and climate research.

In the framework of the preparation for the next European polar-orbiting program (EPS-SG), a new generation of the IASI instrument has been designed. The IASI-NG, which will be launched on board the Metop-SG series in 2021, will be characterized by an improvement of both spectral and radiometric characteristics in comparison with IASI. It will measure at 16 921 wavelengths (or channels) in each sounding pixel benefiting of a spectral resolution and a signal-to-noise ratio improved by a factor 2 compared to its predecessor [*Crevoisier et al.* (2014)]. Measurement precision will be improved as well starting from the 1 K in temperature and 10% in humidity IASI precision. IASI-NG characteristics will lead to huge improvements in detection and retrieval of numerous chemical species and aerosols, and in thermodynamic profiles retrievals.

The high amount of data resulting from IASI-NG will present many challenges in the areas of data transmission, storage and assimilation. Moreover, the number of individual pieces of information will be not exploitable in an operational NWP context and the choice of an optimal data subset will be needed. For all these reasons, an appropriate IASI-NG

channel selection is needed, aiming to select the most informative channels to be used in global and mesoscale NWP models.

The work so far has been carried out on a simulated observation database, containing simulated data for IASI and IASI-NG. Additionally, one-dimensional variational (1D-Var) assimilation experiments have been carried out as well in order to evaluate the impact of IASI-NG with respect to IASI on temperature and humidity retrievals.

The need to take into account the error correlations among channels has required a study for the evaluation of full covariance matrices of the observation errors (**R** matrices). This propaedeutic study will be here below described in Section 4.

Finally, relying on the data thus obtained, a first selection of channels was performed by assessing the contribution of each individual channel to to improving the analysis errors over the background ones. The methodology employed and the results hence obtained will be totally illustrated in Section 5 of this paper.

2 Simulated observation database construction and 1D-Var retrievals

In order to perform the present work, during a previous project by *Andrey-Andrés et al.* (2018), a database of simulated observations has been build, containing simulated profiles for both IASI and IASI-NG. Four dates in the middle of each season from 2013 have been selected: February the 4th, May the 6th, August the 6th and November the 4th. The full IASI orbit for each one of these dates has been computed for a total of 5 242 448 simulations for each instrument (IASI scan geometry is used for IASI-NG).

The experiment has been carried out into four steps as depicted in Figure 2.1: **(1)** construction of the input database containing all the required information; **(2)** spectra simulation for both IASI and IASI-NG; **(3)** data perturbation through Random Gaussian Noise; **(4)** introduction of perturbed input data into 1D-Var to obtain Temperature and Humidity profiles.

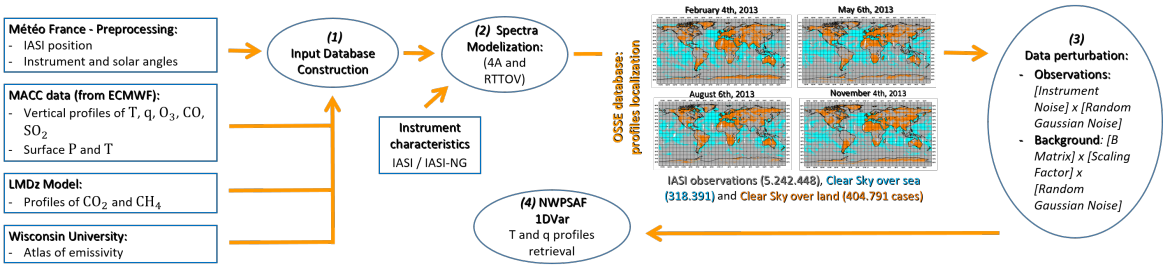


Figure 2.1: Construction of the simulated observation database by *Andrey-Andrés et al.* (2018) and 1D-Var retrievals.

3 Case Study

The firts phase of the selection in question has been planned to be carried out on **nadir - over sea - clear sky conditions**.

For this purpose, a subset of observations matching these criteria has been derived from the starting simulated database. More in details, 1099 of the 318 391 exploitable profiles, have been judged to be a representative sample of the overall data. The new subset contains observations for:

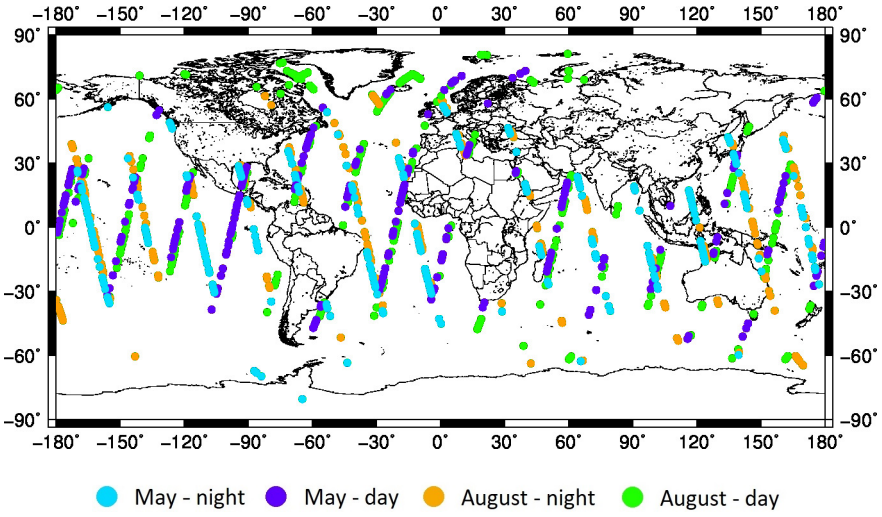


Figure 3.1: Subset of 1099 simulated observations, selected from the database by *Andrey-Andrés et al.* (2018), for preparing the IASI-NG channel selection.

- May and August
- polar, mid-latitudes and tropical regions
- day and night

located as in Figure 3.1.

4 Preparatory study towards the channel selection

A series of tests has been carried out on 1D-Var output retrieved profiles to evaluate the behavior of IASI-NG with respect to IASI. For this purpose 123 IASI channels (99 longwave CO_2 channels for temperature, 4 window channels and 20 water vapour channels), already selected at Météo-France in the operational context of the global model ARPEGE, have been used. The 123 IASI-NG corresponding channels at the same central wavenumbers have been taken into account as well for this preliminary study. In this first stage, a diagonal covariance matrix of the observation errors ($\mathbf{R}$ matrix) filled in with the instrument noise, has been provided as input to the 1D-Var.

Nevertheless, using a diagonal $\mathbf{R}$ matrix that only contains information about the instrument noise error can be quite unrealistic, since it does not take into account all the other sources of observation errors. This significantly contributes to correlations between different channel errors, which should be taken into account in this kind of study. For this reason, a diagnostic procedure introduced by Desroziers et al. (2005) has been used to estimate the structure of the full $\mathbf{R}$ matrix.

A code has been developed in order to implement the diagnostic until its convergence. Through this tool, the different regions of the spectrum that IASI-NG is able to characterize and therefore the bands in which it is divided, have been explored. The attention has been mainly focused on the BAND 1 and BAND 2, which are the most relevant for the assimilation in the NWP context.

Successful trials have been performed in the initial part of BAND 1 (645.000 up to 950.875 cm^{-1}, i.e. the first 2448 channels). This spectrum area, highlighted in blue in Figure 4.1, was approached by applying 8 parallel diagnostic procedures to the same amount of channel groups (of 306 channels each) and obtained by sampling the entire range at a

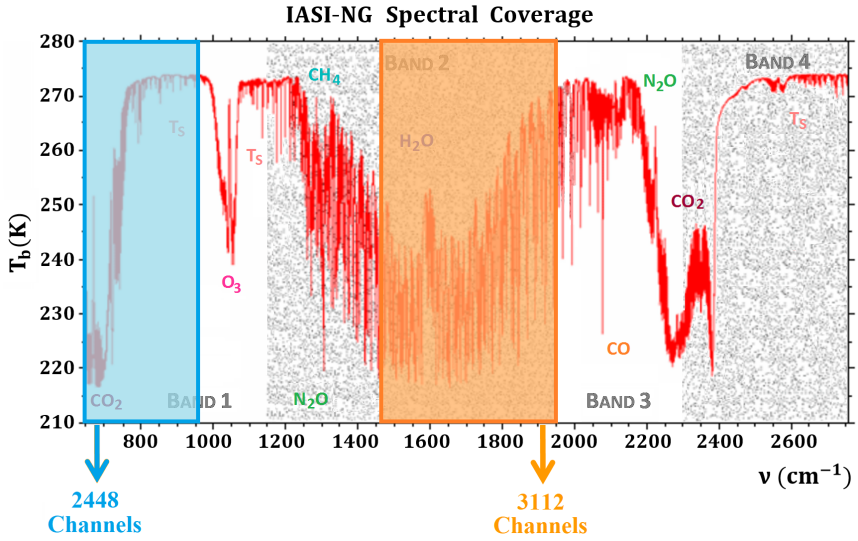


Figure 4.1: here highlighted are the IASI-NG spectrum portions, which have been examined during this study. 2448 and 3112 channels have been considered for BAND 1 and BAND 2, respectively.

distance of 1 cm^{-1}. Subsequently, using the data thus obtained a quasi-full **R** matrix has been built for the 2448 total amount of channels and then used to reiterate the diagnostics.

Figure 4.2 shows the diagnostic error Standard Deviations associated at each iteration of the diagnostic procedure up to its convergence, occurring after 7 iterations.

For what concerns the correlations among different channel errors, the matrices corresponding to the first and the last (7^{th}) step in diagnostic show comparable structures, with the same macro-areas of strong correlation coefficient values (Figure 4.3).

Looking at the 1D-Var retrievals (Figure 4.4), we can see different behaviors for temperature and humidity statistics computed on the 1099 case study profiles. More in details, the Standard Deviations values for temperature show, on most of the atmospheric column, a difference in the usage of a quasi-full or a full **R** matrix. Between 1 and 0.1 hPa, on the other hand, the improvement in the use of full-diagnosed matrices is rather evident if compared to the other one (just imperceptible changes between full first and last iteration matrices). In terms of humidity, the improvement in results is clear between the surface and 500 hPa. Even in this case, there are no big differences between first and last iteration in diagnostics.

The *Degrees of Freedom for the Signal* (DFS) has been computed as well. It is defined as DFS := Tr ($\mathbf{I} - \mathbf{AB}^{-1}$), where **I** is the identity matrix, **A** the covariance matrix of the observation errors, **B** the covariance matrix of the background errors and Tr represents the Trace operator. The retrieved profiles resulting from the usage of the full **R** matrix from diagnostic convergence iteration (7^{th}) have been used with this purpose. The results are illustrated in Table 4.1, where we can notice how the Total DFS (Temperature + Humidity + Skin Temperature DFS), whose Average value is equal to 9.0, is more predominantly influenced by the Temperature DFS components.

For what concerns BAND 2, the diagnostics has been successfully iterated on a set of 3112 approximately contiguous channels, located in the second half part of the band (1455.000 up to 1950.000 cm^{-1} - channels from 6481 to 10 441). The process has been initialized with a diagonal **R** matrix filled in with the instrument noise values.

In this case, we can observe a slower convergence of the diagnostic compared to the previous one (it takes 10 iterations - Figure 4.5). The converged matrix, on the other hand, shows less strong correlations at the first iteration if compared to the last one, but always the same kind of structures (Figure 4.6).

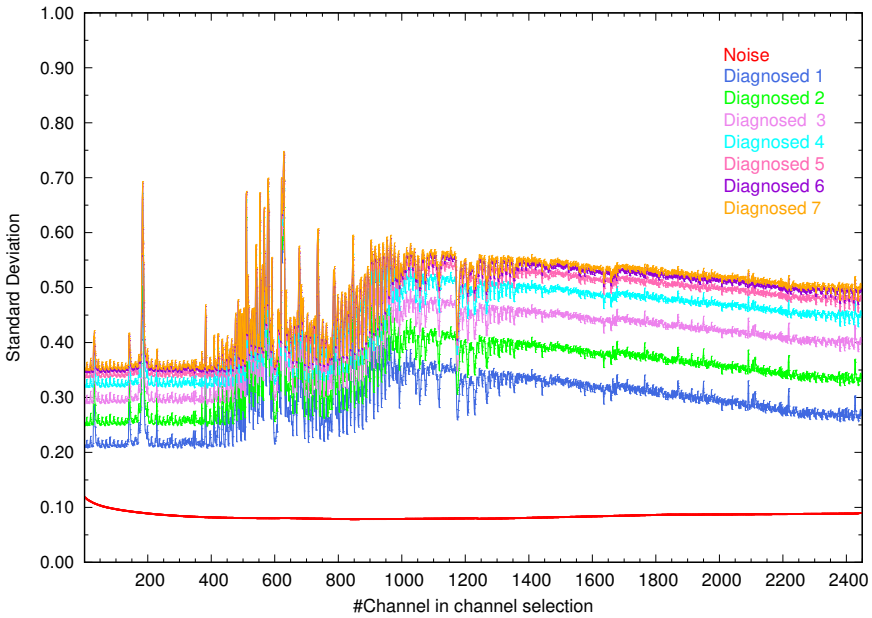
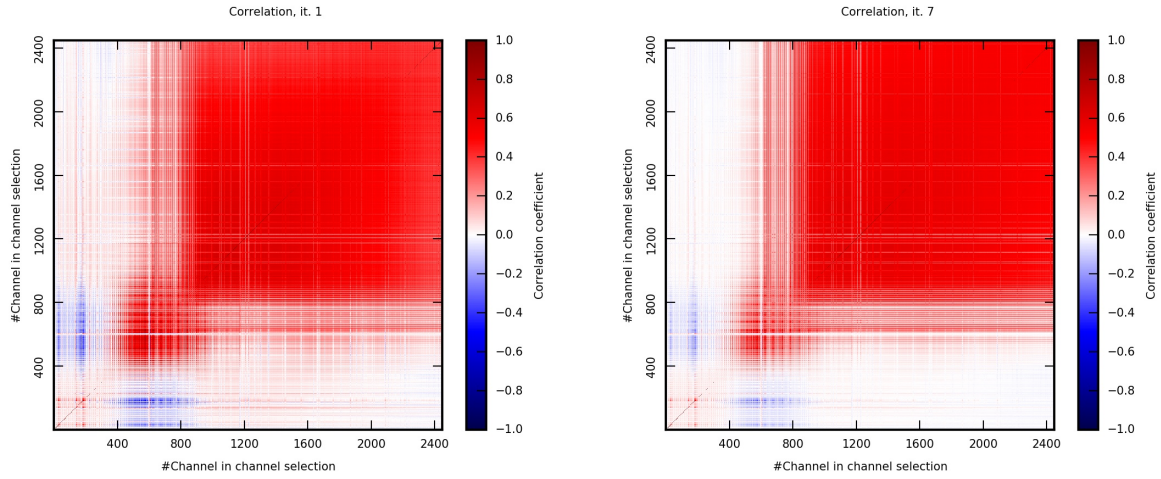


Figure 4.2: diagnosed error Standard Deviations from 1D-Var output and instrument noise for the BAND 1 set of channels.



(a) Correlation, 1^{st} iteration

(b) Correlation, 7^{th} iteration

Figure 4.3: BAND 1 diagnostic for correlation matrix from 1D-Var output; first (a) and last (b) iteration in the process.

DFS	Total	Temperature	Humidity	Skin Temperature
Min	6.9	5.1	0.5	0.98
Max	10.1	6.0	3.6	1.02
Average	9.0	5.6	2.4	1.00

Table 4.1: DFS values computed for retrieved profiles resulting from the use of the full **R** matrix from diagnostic convergence (7^{th}) iteration in the 2448 channels of BAND 1.

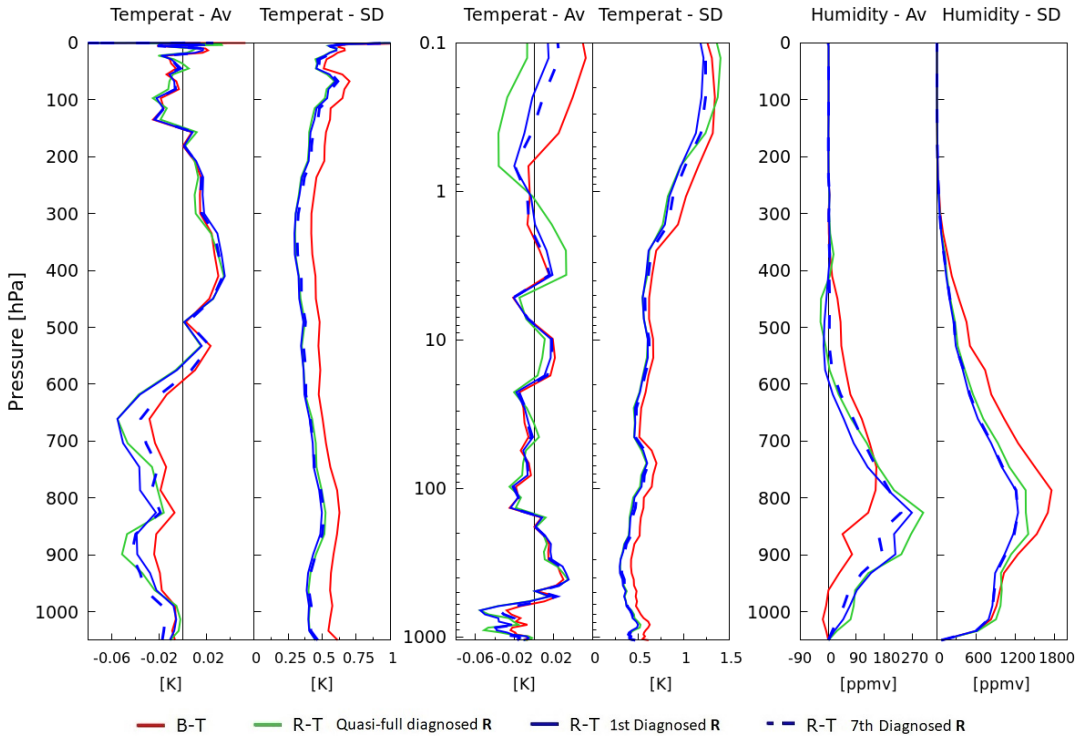


Figure 4.4: Background and Analysis Arrors averaged on the 1099 profiles case study, for temperature (linear scale on the left and log scale in the middle) and humidity (right). In each box are Average on the left and Standard Deviation on the right.

For this part of the study, the 1D-Var has been run using the diagonal **R** matrix, the **R** matrix corresponding to the first iteration of the Desroziers's diagnostic, and the **R** matrix from the last iteration (10^{th}). Looking at the Background and Analysis Error results in Figure 4.7, we can ascertain a big advantage in the usage of the full 10^{th}-iteration **R** matrix. This improvement is evident in the case of both Temperature and Humidity Standard Deviations and it is evident if compared to the results obtained via a diagonal **R** matrix of instrumental noise.

As for BAND 1, the DFS values have been computed and are here shown in Table 4.2. Here we can observe how a major contribution to the Total DFS values is given by humidity fields.

Finally, the *Rate of Improvement* of the Retrieval Standard Deviations compared to the Background ones has been evaluated for both bands. Note that the 1D-Var retrievals here employed have been obtained providing in input the last Desroziers diagnostic iteration **R** matrix for each band.

Figure 4.8 certifies the good improvement in results thus obtained. In more details, for what concerns temperature, the improvement is more pronounced for BAND 1, that is the "temperature band". A maximum score is achieved around -0.3 at the height of about 950 hPa. Concerning humidity, the best results are clearly for BAND 2 in the higher troposphere, over 650 hPa. On the other hand, due to the different sensitivity of the involved channels, in the lower troposphere the best performance comes from the usage of BAND 1 data. The best evidence in this case is registered around -0.5 at 450 hPa (i.e. 50% of improvement) on the curve representing BAND 2.

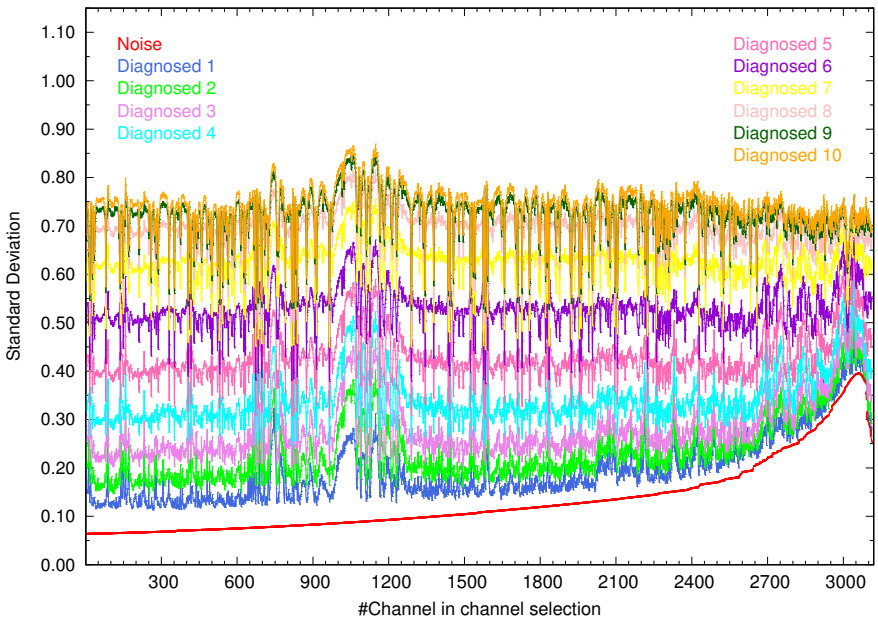
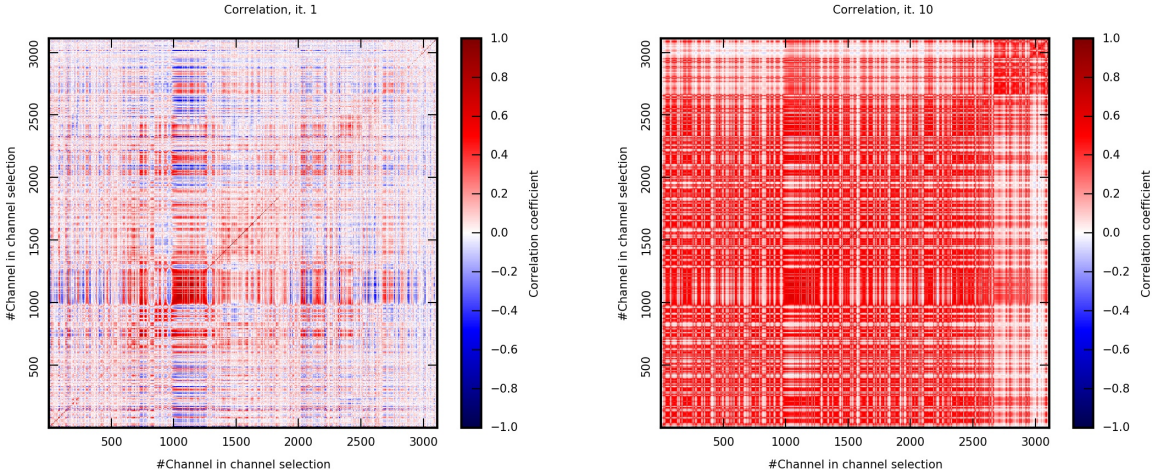


Figure 4.5: diagnosed error Standard Deviations from 1D-Var output and instrument noise for the BAND 2 set of channels.



(a) Correlation, 1^{st} iteration

(b) Correlation, 10^{th} iteration

Figure 4.6: BAND 2 diagnostic for correlation matrix from 1D-Var output; first (a) and last (b) iteration in the process.

DFS	Total	Temperature	Humidity	Skin Temperature
Min	6.4	1.6	3.2	0.09
Max	8.7	2.9	6.2	0.97
Average	7.7	2.2	5.0	0.51

Table 4.2: DFS values computed for retrieved profiles resulting from the use of the full R matrix from diagnostic convergence (10^{th}) iteration in BAND 2 (3112 channels)

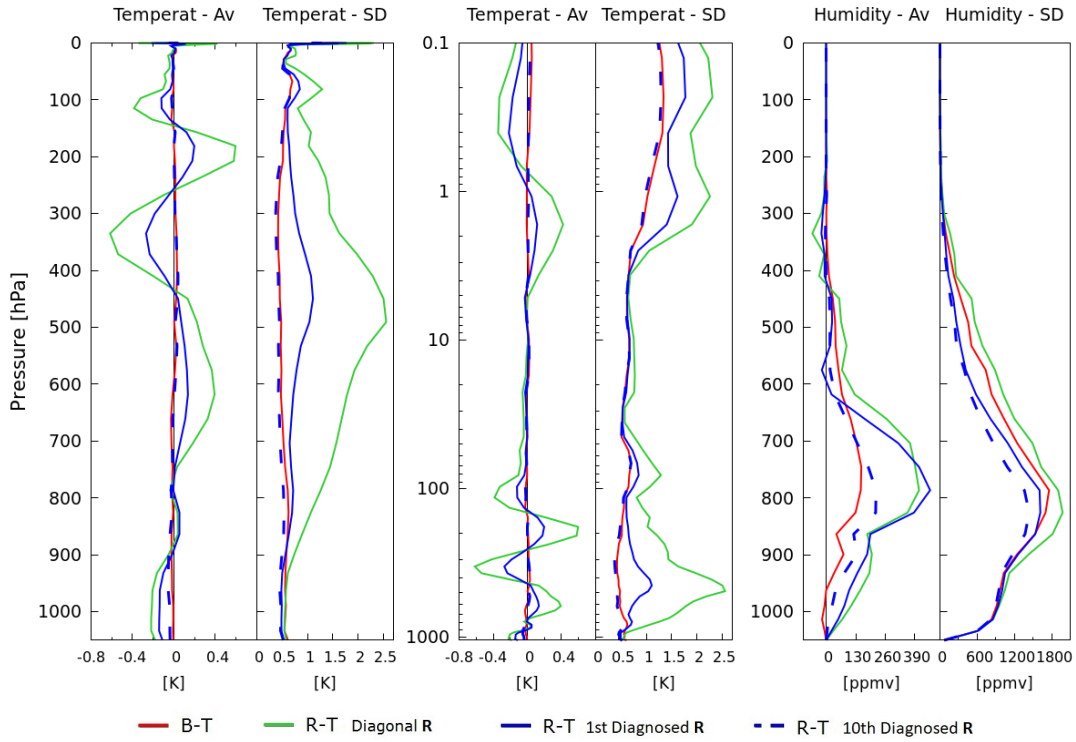


Figure 4.7: Background and Analysis Errors averaged on the 1099 profiles case study, for temperature (linear scale on the left and log scale in the middle) and humidity (right). In each box are Average on the left and Standard Deviation on the right.

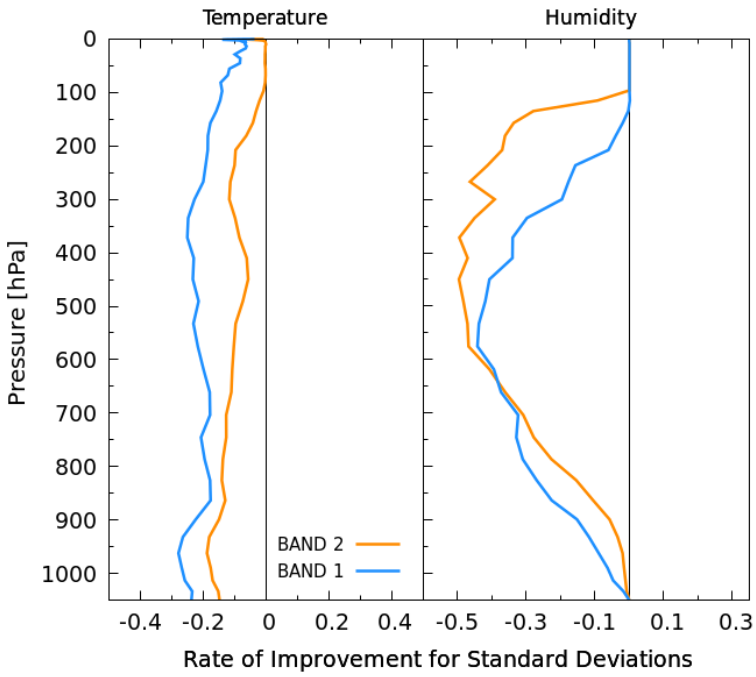


Figure 4.8: Rate of Improvement $= [(R-T)-(B-T)] \cdot (B-T)^{-1}$ for Standard Deviations (R = Retrieval, B = Background, T = Truth). It has been computed using Retrievals obtained via the last iteration diagnostic matrices for both BAND 1 and 2.

5 A first channel selection

This step has been based on a methodology suggested by *Rodgers* (1996) and proved to be a good a priori method for determination of an optimal channel subset by *Rabier et al.* (2002). The method relies on evaluating the impact of the addition of single channels on a figure of merit. This latter is normally a quantity reflecting the improvement of the analysis errors over the background ones (e.g. DFS or Entropy Reduction - ER).

In this very first stage, only one among the case study profiles has been picked up among the *nocturnal* data of *August*, at *mid-latitudes*. The figure of merit chosen to iterate the procedure is the Total DFS.

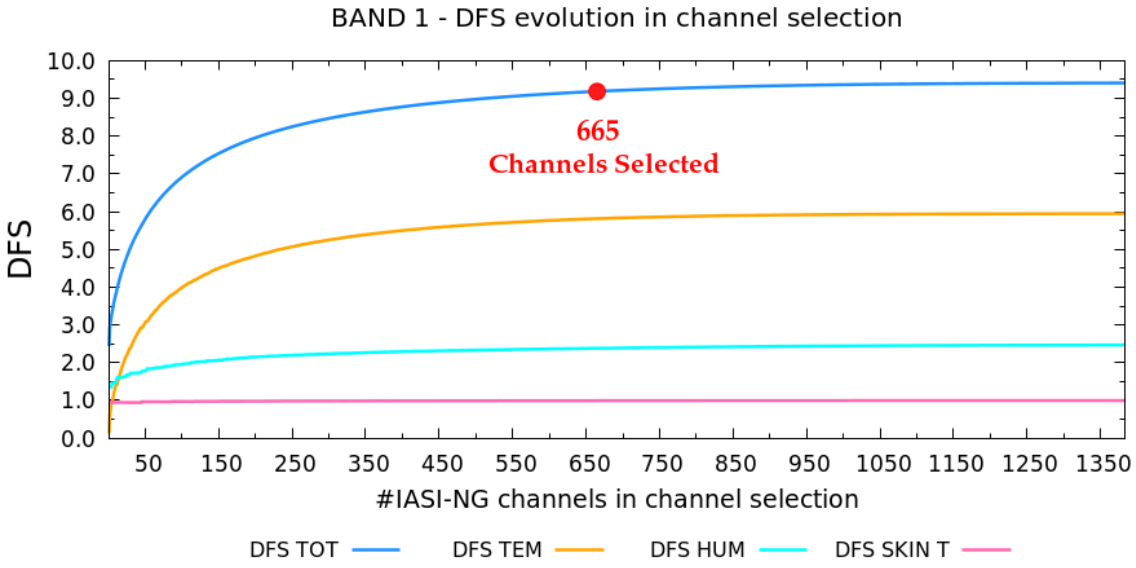
The iterative method has been implemented at this stage so that it stops when the maximum DFS does not reach any significant variation anymore. The growth of the Total DFS with the amount of selected channels is displayed in Figure 5.1 for both BAND 1 and BAND 2. The growth is, in both cases, quite pronounced in the initial phase and it tends to slow asymptotically as the number of chosen channels increases. In the BAND 1 case, since the involved part of the spectrum is the temperature sensitive one, the best contribution is given by the temperature field. Concurrently, in the BAND 2 case, the component that mainly influences the curve of Total DFS is the humidity one.

Based on an analysis of the Total DFS growth rate at each iteration, it has been decided to narrow the choice down to the channels showing a percentage of this parameter greater than 0.01%. This threshold leads to 665 channels selected for BAND 1 and 758 for BAND 2.

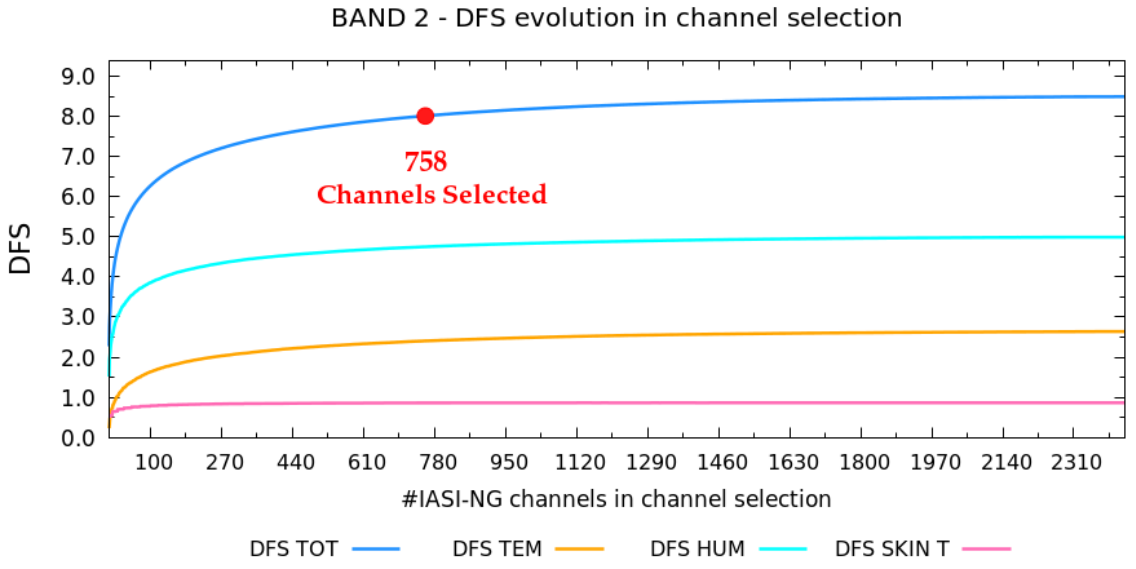
In Figure 5.2 we can observe the distribution of the selected channels along the involved spectrum areas. In the BAND 1 case we succed in selecting a whole series of temperature channels, but also a sub-set of surface-sensitive ones (Figure 5.2(a)). In the matter of BAND 2, a homogeneus distribution of the chosen channels is ascertained, except for the area between 260 and 275 K (Figure 5.2(b)). Here, due to a computational limit in running the 1D-Var, a series of almost 800 channels has been excluded at the stage of a complete $\mathbf{R}$ matrix construction. However, further tests have shown that these channels can be safely recovered by increasing the computational power of the calculating machine.

6 Future works

The next step of the present study aims to refine and test the very first channel selection shown in Section 5. The procedure outlined will be spread to the whole dataset of atmospheric profiles selected as case study in Section 3. Thereafter, a further method will be applied as well. This second approach will be based on selecting the most informative channels relying on the characteristics of their weighting functions [*Gambacorta and Barnet* (2012)].

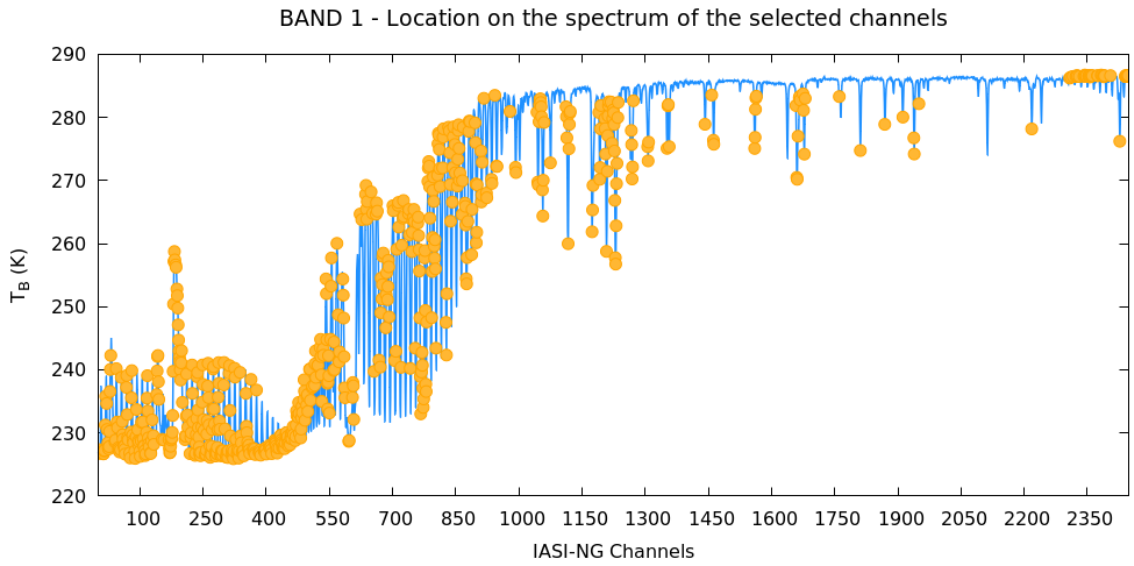


(a)

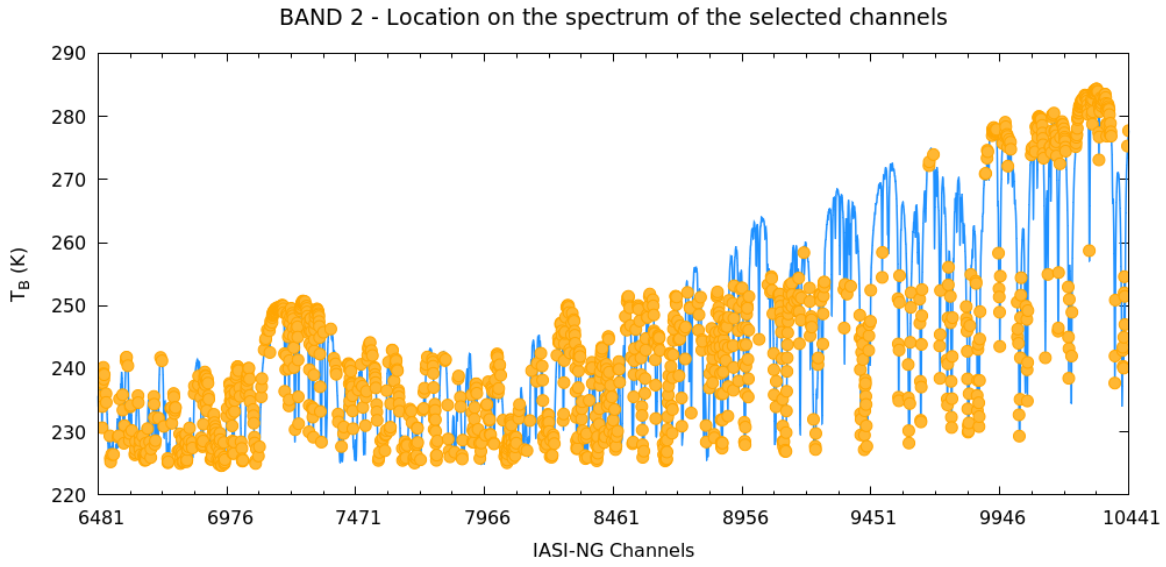


(b)

Figure 5.1: Total DFS growth with the amount of channels selected in BAND 1 (a) and BAND 2 (b). The contribute of Temperature, Humidity and Skin Temperature DFS at each iteration are shown as well. In each panel the threshold of channels selected according to the growth rate $\geqslant 0.01\%$ criterion is highlighted in red.



(a)



(b)

Figure 5.2: Position on the spectrum of the 665 channels selected for BAND 1 (a) and of the 758 ones picked up in BAND 2 (b) accorng to the growth rate $\geqslant 0.01\%$ criterion.

Acknowledgements

This research has received funding from the Centre National d'Études Spatiales (CNES) in the framework of the IASI-NG project.

References

Andrey-Andrés, J., N. Fourrié, R. Armante, P. Brunel, C. Crevoisier, V. Guidard, and B. Tournier, 2018: A simulated observation database to assess the impact of IASI-NG hyperspectral infrared sounder. *Atmos. Meas. Tech.*, **11**, 803-818. DOI: 10.51941/amt-11-803-2018.

Crevoisier, C., and Coauthors, 2014: Towards IASI-New Generation (IASI-NG): impact of improved spectral resolution and radiometric noise on the retrieval of thermodynamic, chemistry and climate variables. *Atmos. Meas. Tech.*, **7**, 4367-4385. DOI: 10.5194/amt-7-4367-2014.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**, 3385-3396. DOI: 10.1256/qj.05.108.

Gambacorta, A., and C.D. Barnet, 2012: Methodology and Information Content of the NOAA NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS). *IEEE Trans. Geosci. Remote Sens.*, **51**, 3207-3216. DOI: 10.1109/TGRS.2012.2220369.

Rodgers, C. D., 1996: Information contentand optyimisation of high spectral resolution measurements. Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II, SPIE, **2830**, 136-147.

Rabier, F., N. Fourrié, D. Chafai, P. Prunet, 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteorol. Soc.*, **128**, 1011-1027. DOI: 10.1256/0035900021643638.