An Infrared Atmospheric Sounding Interferometer – New Generation (IASI-NG) channel selection for Numerical Weather Prediction

Francesca Vittorioso, Vincent Guidard, Nadia Fourrié

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

francesca.vittorioso@meteo.fr

Paper under review at QJRMS:
Vittorioso F., V. Guidard and N. Fourrie: *An Infrared Atmospheric Sounding Interferometer – New Generation (IASI-NG) channel selection for Numerical Weather Prediction*

I. Introduction

As the EUMETSAT Polar System-Second Generation (EPS-SG) is being prepared, a new generation of the hyperspectral Infrared Atmospheric Sounding Interferometer (IASI) has being designed. The IASI New Generation (IASI-NG), which is expected to be launched in 2022 on board the Metop second generation series, will measure at 16921 wavelengths in each sounding pixel with a spectral sampling of 0.125 cm^{-1} and a signal-to-noise ratio improved by a factor 2 compared to its predecessor. Measurement precision will be improved as well. The IASI-NG characteristics will lead to huge improvements in detection and retrieval of numerous chemical species and aerosols, and in thermodynamic profiles retrievals, as well as providing, like its predecessor, a huge contribution to Numerical Weather Prediction (NWP).

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 going to be performed aiming to select the most informative channels for NWP.

II. IASI-NG context

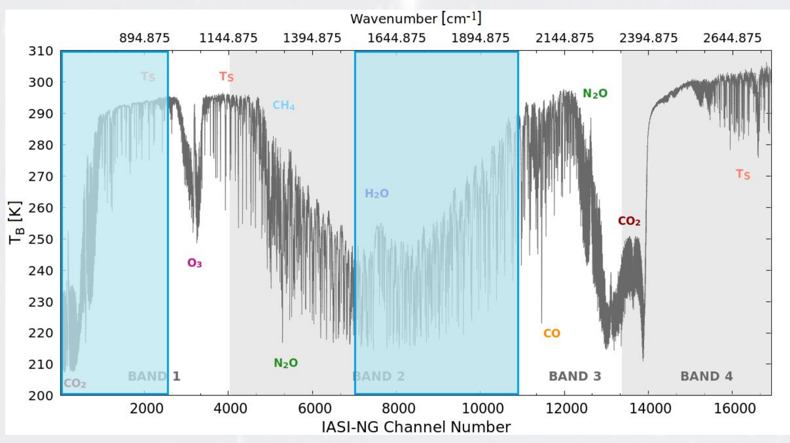

Figure 1: Example of simulated IASI-NG spectrum (divided into four spectral bands, which are here depicted together with the sensitivity of the different spectral regions). The blue boxes highlight the preselected areas to be examined during this study: the first 2448 consecutive channels of band 1 and the last 3601 consecutive channels of band 2 (from channel 6841 up to 10 441).

IASI-NG is a space-borne FTS, based on the concept of the Mertz interferometer, that will measure IR radiation emitted from the Earth within the range 645 and 2760 cm^{-1} (3.6 – 15.5 μm). It will benefit of a spectral sampling of 0.125 cm^{-1}, a spectral resolution of 0.25 cm^{-1} after apodization and a signal-to-noise ratio improved by a factor 2 compared to its predecessor [*Crevoisier et al.* (2014)]. Measurement precision will be improved starting from the 1 K in temperature and 10% in humidity IASI precision.

The IASI-NG spectrum will be split into four bands, compared to three bands for IASI. For this study, we focused on bands 1 and 2. We selected the first 2448 contiguous channels of band 1 (from 645.000 to 950.875 cm^{-1}) and the last 3601 contiguous channels of band 2 (from channel 6841 up to 10 441, namely from 1500.000 to 1950.000 cm^{-1}). **The total number of channels considered is 6049.**

This choice excludes from the study the ozone-related part of the band 1, since the minimisation of this quantity is not performed in the operational assimilation at Météo-France yet. On the other hand, the second half part of band 2 has been preferred to the first half since it will provide good information about water vapor, without being affected by the influence of trace gases (e.g. N_2O or CH_4). Moreover, this spectral area will be consistent with the one that the Infrared Sounder (IRS), which will fly on board the Meteosat Third Generation (MTG), will provide in the next few years.

The areas just described are highlighted on the spectrum of Figure 1.

III. Case study

Since the actual data from IASI-NG will only be available in a few years, we exploited the database from *Andrey-Andrés et al.* (2018) containing 5242047 simulated observations using the scan geometry of IASI for both IASI and IASI-NG and covering four dates in the middle of each season from 2013 in order to provide a maximum of meteorological variability.

For the purpose of this project we selected subsets of data to be used for the various steps of this study.

For the observation error estimation step, a subset of **6267** profiles has been selected matching the following criteria: Instrument Zenith Angle (IZA) < 2.8 (i.e. at nadir), over sea (water) clear sky day/night illumination conditions.

For the channel selection stage, on the other hand, an agglomerative hierarchical clustering technique has been used in order to isolate in the previous 6267 subset, a smaller group of **77** cases (both typical and extreme).

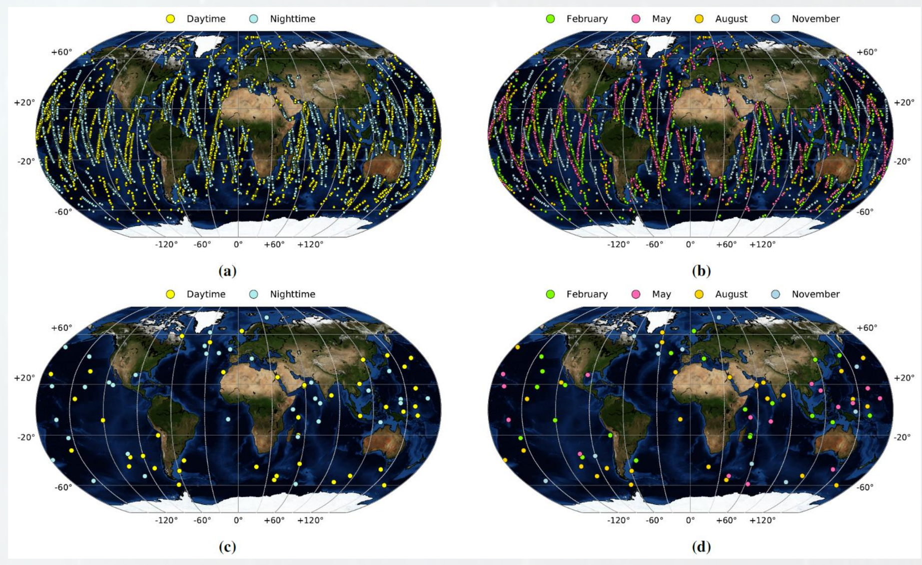

Figure 2: Geographical distribution of the profile sets taken as case study. The top panels show the 6267 profiles preselected to be used for the observation error evaluation. They are displayed in different colors depending on the illumination conditions (a) and the day of the year (b). The bottom panels (c) and (d) refer to the 77 profiles picked up, starting from the 6267 subset, through a hierarchical agglomerative technique in order to carry out the proper selection phase.

IV. Observation error estimation

For the present study we decided to use a full covariance matrix of the observation errors. The diagnostic procedure introduced by *Desroziers et al.* (2005) has been chosen to estimate the structure of a full R matrix. This method allows to obtain variances and covariances of observation errors *from observation-minus-background* $[\mathbf{y}-H(\mathbf{x_b})]$ and *observation-minus-analysis* $[\mathbf{y}-H(\mathbf{x_a})]$ statistics [y vector of the observations and H the observation operator, $\mathbf{x_b}$ a priori background state vector estimated before the analysis is carried out, $\mathbf{x_a}$ analysis state]. This matrix is given by the following expression:

$$\mathbf{R} = E\left\{[\mathbf{y}-H(\mathbf{x_a})][\mathbf{y}-H(\mathbf{x_b})]^T\right\}$$

[E statistical expectation operator].

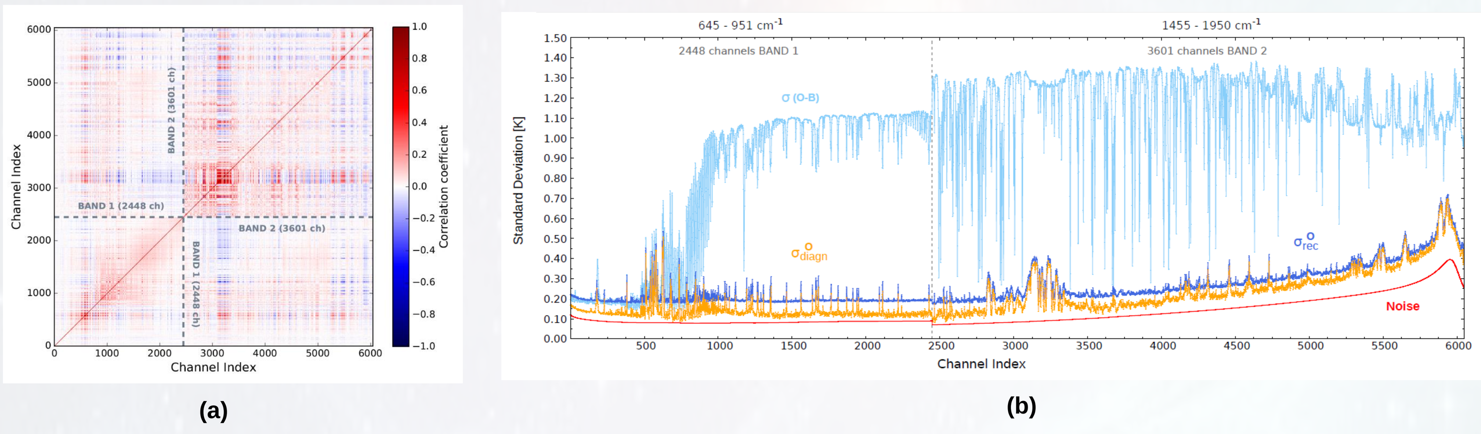

Figure 3:
(a) Diagnosed IASI-NG error correlations from 1D-Var output. The boxes outlined by dashed lines highlight the results corresponding to the 2448 channels of band 1 (bottom left) and to the 3601 channels of band 2 (top right). The bottom right and top left boxes, on the other hand, refer to the diagnosed crossed correlations.
(b) Diagnosed IASI-NG observation-error standard deviations from 1D-Var output (orange), standard deviations from 1D-Var output after reconditioning (dark blue), standard deviations of the innovations (light blue), instrument noise (red).

V. Optimal selection

The methodology applied for selecting channels is the one suggested by *Rodgers* (1996) and proved to be a good a priori method for the determination of an optimal channel set by *Rabier et al.* (2002). This method relies on evaluating the impact of the addition of one channel at time on a scalar figure of merit reflecting the improvement of the analysis error over the background one. The figure of merit we chose to iterate the selection process is the *Degrees of Freedom for Signal* (**DFS**) computed as follows: $\mathrm{DFS} = Tr\left[\mathbf{I} - (\mathbf{I} + \mathbf{B}\,\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\right]$, with **H** the Jacobian matrix, **I** the identity matrix and **B** background-error covariance matrix.

We performed a channel selection on each of the 77 profiles chosen for this study, stopping the iterative process at the threshold of 500 since currently this is the number distributed of IASI channels to NWP centres. An average of the DFS evolution is shown in Figure 4. From these 77 independent selections we chose the **500 channels most frequently selected** (channels chosen in 37.7% or more of the evaluated cases). The selection consists of 300 channels belonging to band 1 (277 temperature and 23 surface-sensitive channels) and 200 from the water vapour band) [Figure 5].

This selection, plus other groups of channels retained more or less frequently, have been tested on a larger set of 6267 atmospheric profiles, have been tested on a larger set of 6267 atmospheric profiles. The results obtained show an improvement in the Retrievals compared to the Background profiles when using the selected channels (Figure 6). Regarding temperature, the improvement is especially pronounced in the very low layers (from 960 to 990 hPa approximately) and in the upper troposphere, with improvements reaching up to 30% at 200 hPa. For water vapour, on the other hand, this improvement is more pronounced in the medium/high layers (between 600 and 200 hPa) with improvements up to 50%.

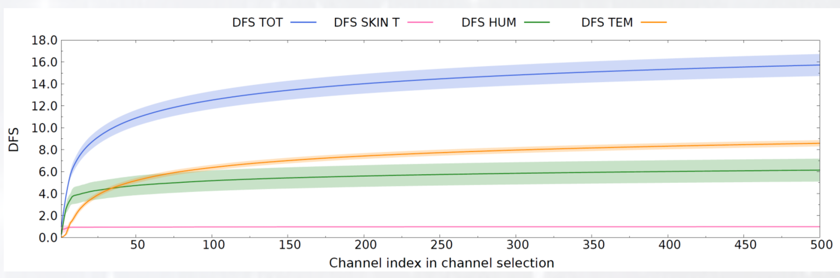

Figure 4: DFS trend values in channel selection averaged on the 77 case study profiles. The solid blue line corresponds to the total DFS. The pink, green and orange ones, on the other hand, refer to the the skin temperature, humidity and temperature terms respectively. The shaded areas associated to each curve report the standard deviations associated to that specific term.

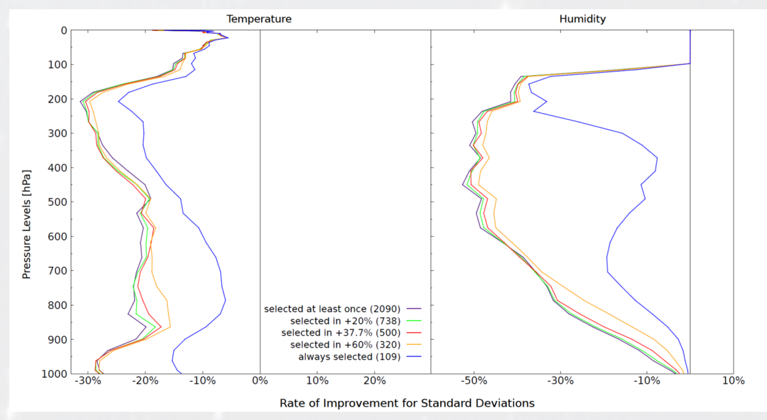

Figure 6: Rate of Improvement (ROI) of analysis-error standard deviations compared to the background ones averaged on 6267 profiles, for temperature and humidity. It is a comparison among the results obtained for the groups of channels chosen at least once, in 20, 60, 100% of cases and the ROI obtained through our 500 channel selection. Negative values mean Retrievals improved compared to the Background.

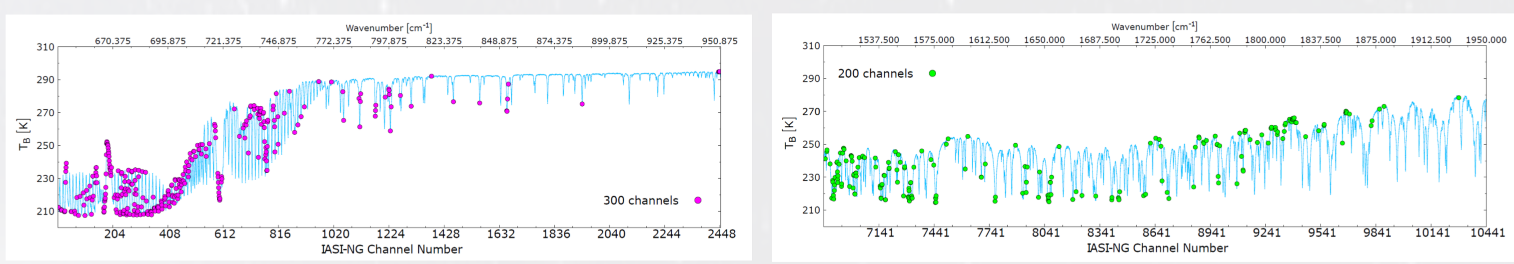

Figure 5: Spectral location of the 500 selected channels: 300 in band 1 (magenta) and 200 in band 2 (green dots). The left panel represents the preselected band 1 area (from 645.000 to 950.875 cm^{-1}), the bottom panel the band 2 one (from 1500.000 to 1950.000 cm^{-1}).

VI. Random selection

In order to evaluate the optimality of the already well-known channel selection methodology applied along this work, and as a good sanity check, a completely random selection was performed. First of all, 77 different 500 random channel selections were produced without any constraints or conditions except that the aforementioned channels belonged to the spectral areas covered during this study.

After that, 1D-Var assimilation experiments have been performed by using a different random channel selection for each of the 77 case study profiles. The aim was to compare these results with those from the assimilation experiments carried out using for each of the 77 cases its optimal selection obtained through the DFS method.

The optimal selection and the proposed 500 channel selection perform better than random ones.

The gain is particularly important for temperature in the upper part of the troposphere/stratosphere.

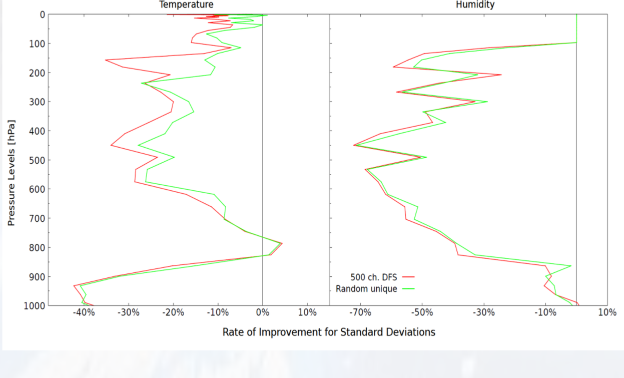

Figure 7: ROI of Retrieval Standard Deviations compared to the Background ones, for temperature (left panel) and humidity (right). In red is the average on the 77 case study profiles of the ROI obtained using for each of them the 500 channels most frequently selected through the DFS method. The green curve shows the results produced using the same random channel selection applied to all 77 profiles.

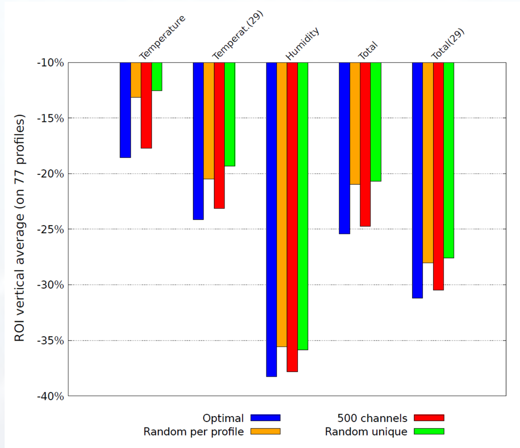

Figure 8: Histograms of ROI averaged on the 77 profiles and the vertical (total or partial) column, as a function of temperature, humidity and for the total. Blue bars refer to the optimal selection, the orange ones to a different random selection for each profile, the red ones to the 500 channel selection and the green bars to a single random selection applied over the whole 77 profile subset.

VII. Conclusions and perspectives

The selection proposed in this paper consists of the 500 channels most frequently selected by using the optimal selection method.

The selected channels depend on many components, such as the atmospheric profiles, the errors prescribed to the background and the observations.

In further studies, it could be interesting to compare several methods using the same inputs (**B**, **R**, observations) and to assess the impact on the metrics we used, such as the DFS statistics and the rate of improvement.

An intercomparison exercise using several **B** matrices provided by different NWP centres in the selection process may be of interest to evaluate the overall sensitivity of the selection to the background errors.

Future works may focus on spectral areas not yet explored in this study. Scanning and evaluating channels sensitive to other chemical species would make the selection even more complete and optimal. Cloudy and over land case studies could equally be analysed. Moreover, it would be interesting to evaluate the IASI-NG selection in a more realistic context as the one of a global model such as the Météo-France Action de Recherche petite Echelle Grande Echelle (ARPEGE).

References

Andrey-Andrés, J., N. Fourrié, R. Armante, P. Brunel, C. Crevoisier, V. Guidard, and B. Tournier, 2017: A simulated observation database to asses the impact of IASI-NG hyperspectral infrared sounder. *Atmos. Meas. Tech.*, ***2017***, 1-29. DOI: 10.5194/amt-2017-266.

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.

Rodgers, C. D.,1996: Information contentand opyimisation 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.

Acknowledgements

This research has received funding from CNES in the framework of the IASI-NG project.