Evaluation of a first IASI-NG channel selection for Numerical Weather Prediction

Francesca Vittorioso, Vincent Guidard, Nadia Fourrié

CNRM, Météo-France & CNRS, GMAP/OBS
francesca.vittorioso@meteo.fr
Objectives

Why a channel selection

- High amount of data resulting from IASI-NG
- Many challenges in the areas of data storage and assimilation
- The number of individual pieces of information will be not exploitable in an operational Numerical Weather Predictions (NWP) context

An appropriate IASI-NG channel selection is needed, aiming to select the most informative channels for NWP models
3601 channels

✓ no trace gases
✓ consistent with IRS

2448 channels

✓ no Ozone
31 October - 6 November 2019
IASI-NG: a first channel selection

Context

2448 channels
✓ no Ozone

3601 channels
✓ no trace gases
✓ consistent with IRS

6049 channels
Case Study

Full IASI orbit computed, for a total amount of 5242448 simulations for each instrument (IASI scan geometry used for IASI-NG) [Andrey-Andrés et al. (2018)]

- Nadir
- Over sea
- Clear sky
- Day/night illumination conditions

30210 observations matching these criteria

6267 profiles judged to be a representative sample
Full IASI orbit computed, for a total amount of 5242448 simulations for each instrument (IASI scan geometry used for IASI-NG) [Andrey-Andrés et al. (2018)]

- Nadir
- Over sea
- Clear sky
- Day/night illumination conditions

30210 observations matching these criteria

6267 profiles judged to be a representative sample
Preparatory study towards the channel selection

1D-Var retrievals

\[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} [y - H(x)]^T R^{-1} [y - H(x)] \]

- \( x = \) model state vector (Temperature, Humidity, Skin Temperature)
- \( x_b = \) background state vector
- \( y = \) vector of observations
- \( B = \) background-error covariance matrix (provided in NWPSAF 1D-Var package)
- \( R = \) observation-error covariance matrix
- \( H = \) observation operator
Observation error estimation

The previous generation of channel selection schemes ignored, or represented just roughly, spectrally correlated errors [e.g. Rabier et al. (2002), Collard (2007)]. This often limited the accuracy of the assimilation process.

To estimate full $\mathbf{R}$ matrices, a diagnostic procedure introduced by Desroziers et al. (2005) was implemented.

$$J(x) = \frac{1}{2} (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + \frac{1}{2} [y - H(x)]^T \mathbf{R}^{-1} [y - H(x)]$$

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

(Observ - Analysis) (Observ - Backg)
Preparatory study towards the channel selection

Diagnostic error standard deviations from 1D-Var output and instrument noise

6049 channels

Correlation coefficient

Diagnostic for correlation matrix

6049 channels from BAND 1 and BAND 2
Case Study for Channel Selection

6267 profiles case study set can be further thinned

**Agglomerative hierarchical clustering technique:**

profiles grouped by evaluating distance in terms of temperature and humidity

77 profiles representative of different atmospheric conditions (typical and extreme cases)
Case Study for Channel Selection

77 profiles representative of different atmospheric conditions (typical and extreme cases)
Case Study for Channel Selection

77 profiles representative of different atmospheric conditions (typical and extreme cases)
A methodology relying on evaluating the impact of the addition of single channels on a figure of merit has been applied in order to determinate the optimal channel set [Rodgers (1996), Rabier et al. (2002)].

The figure of merit chosen to implement the selection is the **Total DFS** (*Degrees of Freedom for the Signal*)

\[
\text{DFS} := Tr \left( I - AB^{-1} \right) = Tr(I) - Tr\left( (I + BH^TR^{-1})^{-1} \right)
\]

\[
\text{DFS}_{\text{Tot}} = \text{DFS}_{\text{Temp}} + \text{DFS}_{\text{Hum}} + \text{DFS}_{\text{Skin Temp}}
\]

- **A** = analysis-error covariance matrix
- **B** = backg-error covariance matrix
- **R** = analysis-error covariance matrix
- **H** = matrix of the Jacobians
The choice has been narrowed down to the first 500 channels selected on each profile (77 different channel selections).

500 is the amount of channels distributed for IASI (if the PC scores will be used, there is a good chance that this amount will be the same for IASI-NG).
We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is never selected (3959 out of 6049)
We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is never selected (3959 out of 6049)
- **35%** (2090 channels) selected at least once
We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is **never selected** (3959 out of 6049)
- **35%** (2090 channels) are **selected at least once**

**Distribution in percentage of selected channels** (from 6049 channels - **BAND 1** and **BAND 2**)

- **BAND 1**
- **BAND 2**

<table>
<thead>
<tr>
<th>Amount of profiles</th>
<th>Amount of selected channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>780 ch</td>
</tr>
<tr>
<td>5%</td>
<td>330 ch</td>
</tr>
<tr>
<td>10%</td>
<td>162 ch</td>
</tr>
<tr>
<td>15%</td>
<td>85 ch</td>
</tr>
<tr>
<td>20%</td>
<td>80 ch</td>
</tr>
<tr>
<td>25%</td>
<td>74 ch</td>
</tr>
<tr>
<td>30%</td>
<td>65 ch</td>
</tr>
<tr>
<td>35%</td>
<td>41 ch</td>
</tr>
<tr>
<td>40%</td>
<td>48 ch</td>
</tr>
<tr>
<td>45%</td>
<td>48 ch</td>
</tr>
<tr>
<td>50%</td>
<td>29 ch</td>
</tr>
<tr>
<td>55%</td>
<td>28 ch</td>
</tr>
<tr>
<td>60%</td>
<td>22 ch</td>
</tr>
<tr>
<td>65%</td>
<td>33 ch</td>
</tr>
<tr>
<td>70%</td>
<td>23 ch</td>
</tr>
<tr>
<td>75%</td>
<td>26 ch</td>
</tr>
<tr>
<td>80%</td>
<td>14 ch</td>
</tr>
<tr>
<td>85%</td>
<td>18 ch</td>
</tr>
<tr>
<td>90%</td>
<td>25 ch</td>
</tr>
<tr>
<td>95%</td>
<td>50 ch</td>
</tr>
<tr>
<td>100%</td>
<td>109 ch</td>
</tr>
</tbody>
</table>

**Always Selected**

**Channels Selected in (0%-5%) of cases**

**BAND 1**

- 0% of profiles
- 5% of profiles

**BAND 2**

- 10% of profiles
- 15% of profiles
- 20% of profiles
- 25% of profiles
- 30% of profiles
- 35% of profiles
- 40% of profiles
- 45% of profiles
- 50% of profiles
- 55% of profiles
- 60% of profiles
- 65% of profiles
- 70% of profiles
- 75% of profiles
- 80% of profiles
- 85% of profiles
- 90% of profiles
- 95% of profiles
- 100% of profiles
We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is never selected (3959 out of 6049)
- **35%** (2090 channels) selected at least once

**Distribution in percentage of selected channels** (from 6049 channels - BAND 1 and BAND 2)
Channel Selection

We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is never selected (3959 out of 6049)
- **35%** (2090 channels) selected at least once

Distribution in percentage of selected channels (from 6049 channels - BAND 1 and BAND 2)
We performed an analysis to evaluate the percentage of choice of the different channels:

- **65%** of the total amount of channels is **never selected** (3959 out of 6049)
- **35%** (2090 channels) **selected at least once**
The groups of channels just highlighted (selected on 77 profiles) have been used to perform 1D-Var experiments over the 6267 initial case study profiles, in order to evaluate the impact of the selection.
The groups of channels just highlighted (selected on 77 profiles) have been used to perform 1D-Var experiments over the 6267 initial case study profiles, in order to evaluate the impact of the selection.
Channel Selection

![Graph showing spectral lines and channel selection]

Wavenumber (cm⁻¹)

T_B (K)

IASI-NG Channel Number

301 channels

199 channels

CO₂

H₂O

N₂O

CH₄

O₃

TS

TS

TS

TS

CO

31 October - 6 November 2019

IASI-NG: a first channel selection
Conclusions and Future Works

- The selection we offer consists of 500 channels.
- This selection has been tested with very good results.

- Study bands and spectrum areas that we have not treated in this work (ozone, short waves, etc.)
- Bring the selected channels into the global model ARPEGE data assimilation, with all the observations available and over several weeks (OSSE)
- Vittorioso et al. in preparation for QJRMS