# 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**

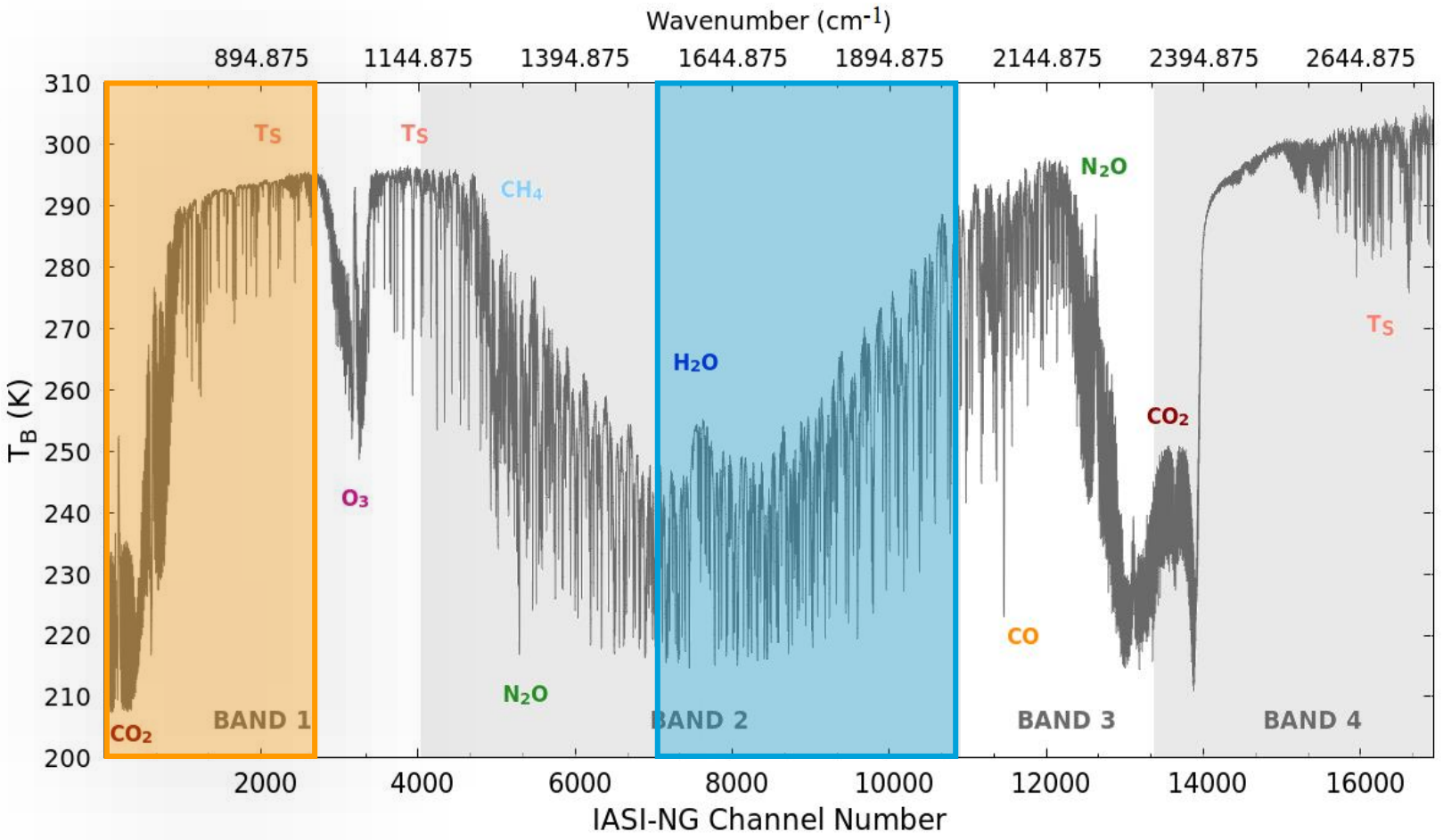
# Context

**2448 channels**

- ✓ no Ozone

**3601 channels**

- ✓ no trace gases
- ✓ consistent with IRS



# Context

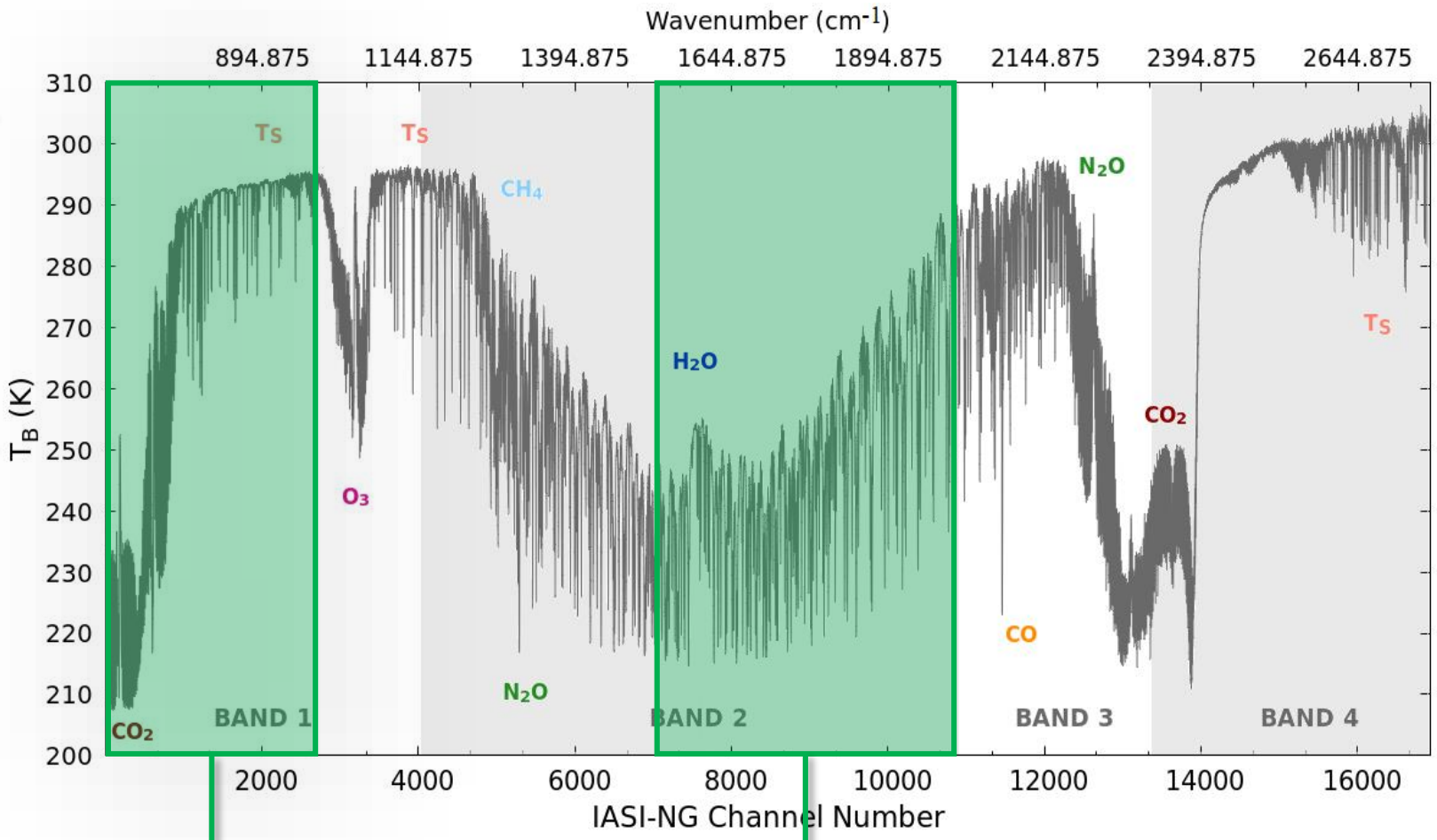
**2448 channels**

- ✓ no Ozone

+

**3601 channels**

- ✓ no trace gases
- ✓ consistent with IRS




**6049 channels**

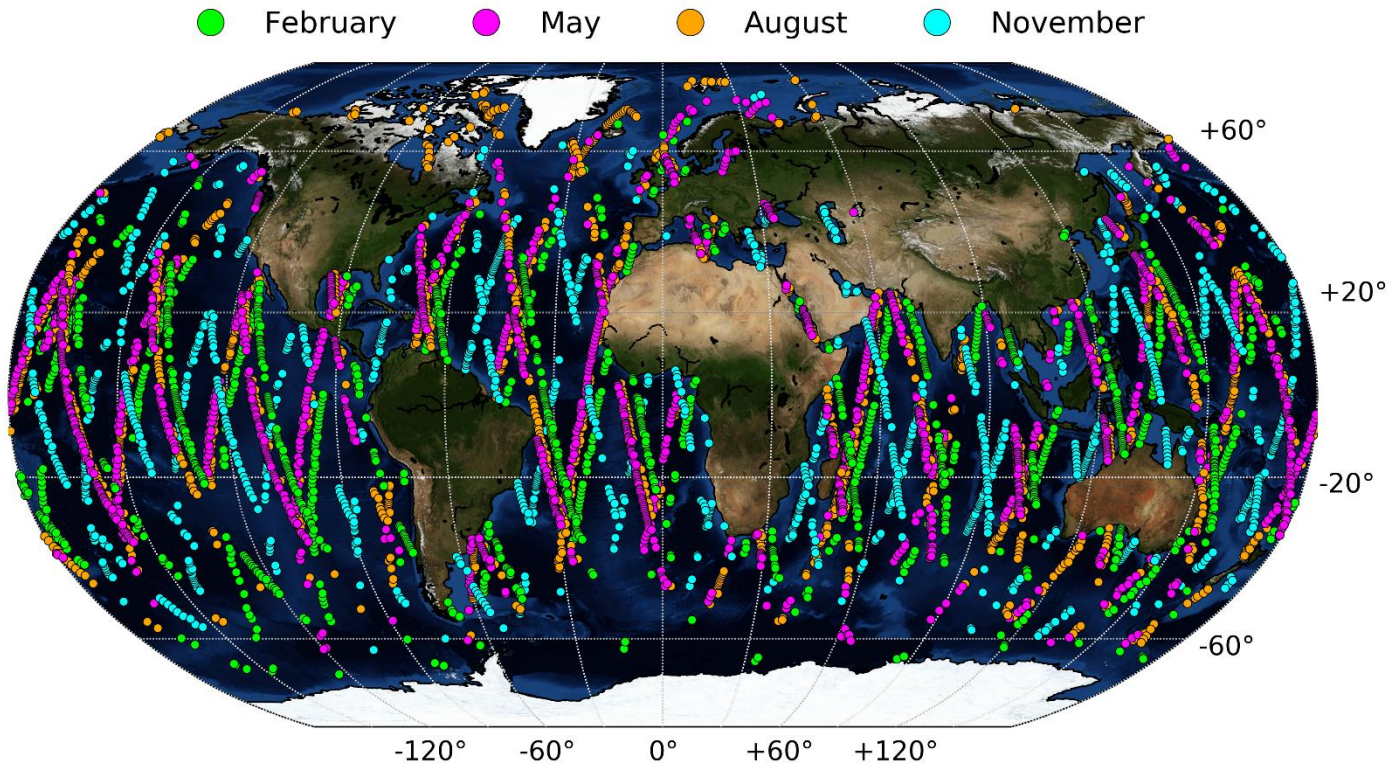
# Case Study

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**



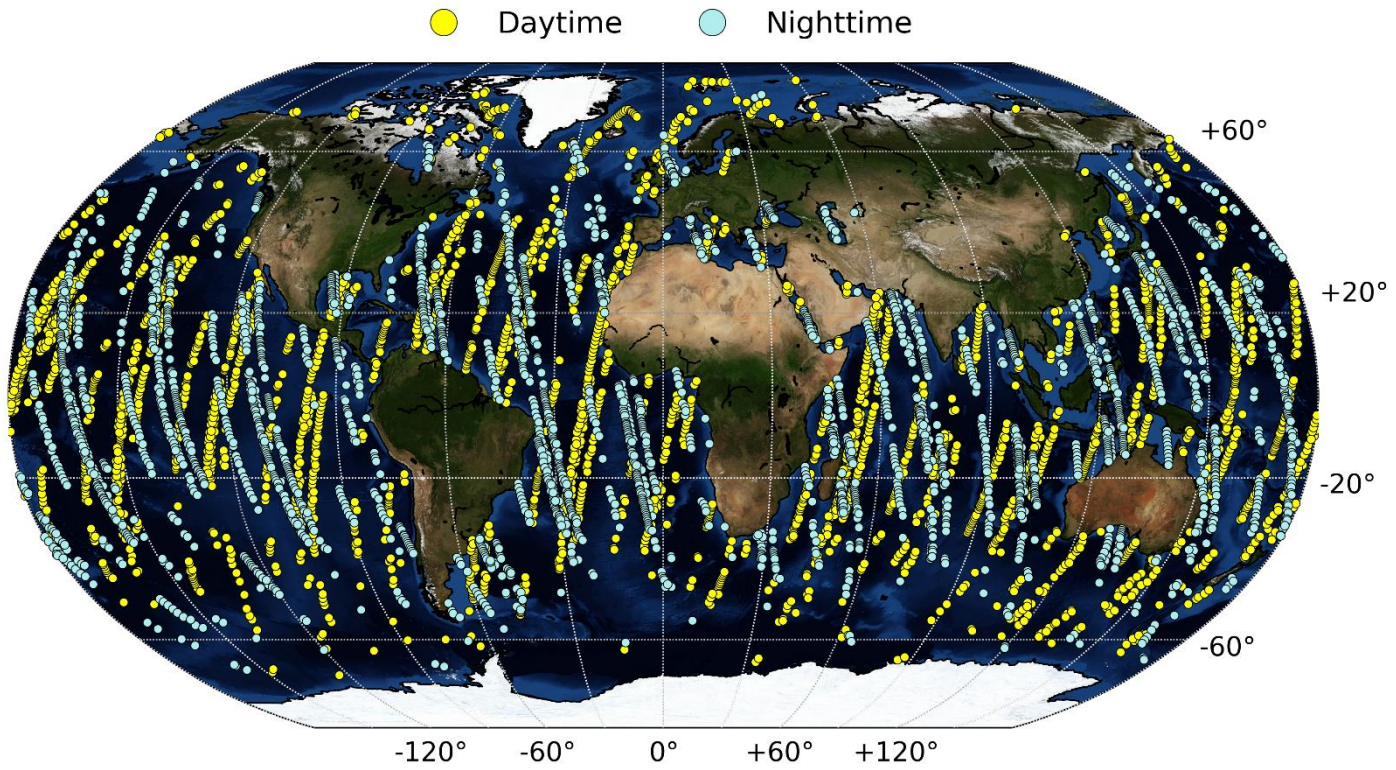
# Case Study

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**



# Preparatory study towards the channel selection

## 1D-Var retrievals

Background

Observations

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

$\mathbf{x}$ = model state vector

(Temperature, Humidity, Skin Temperature)

$\mathbf{x}_b$ = background state vector

$\mathbf{y}$ = vector of observations

$\mathbf{B}$ = background-error covariance matrix

(provided in NWPSAF 1D-Var package)

$\mathbf{R}$ = observation-error covariance matrix

$H$ = observation operator

# Preparatory study towards the channel selection

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

## 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.

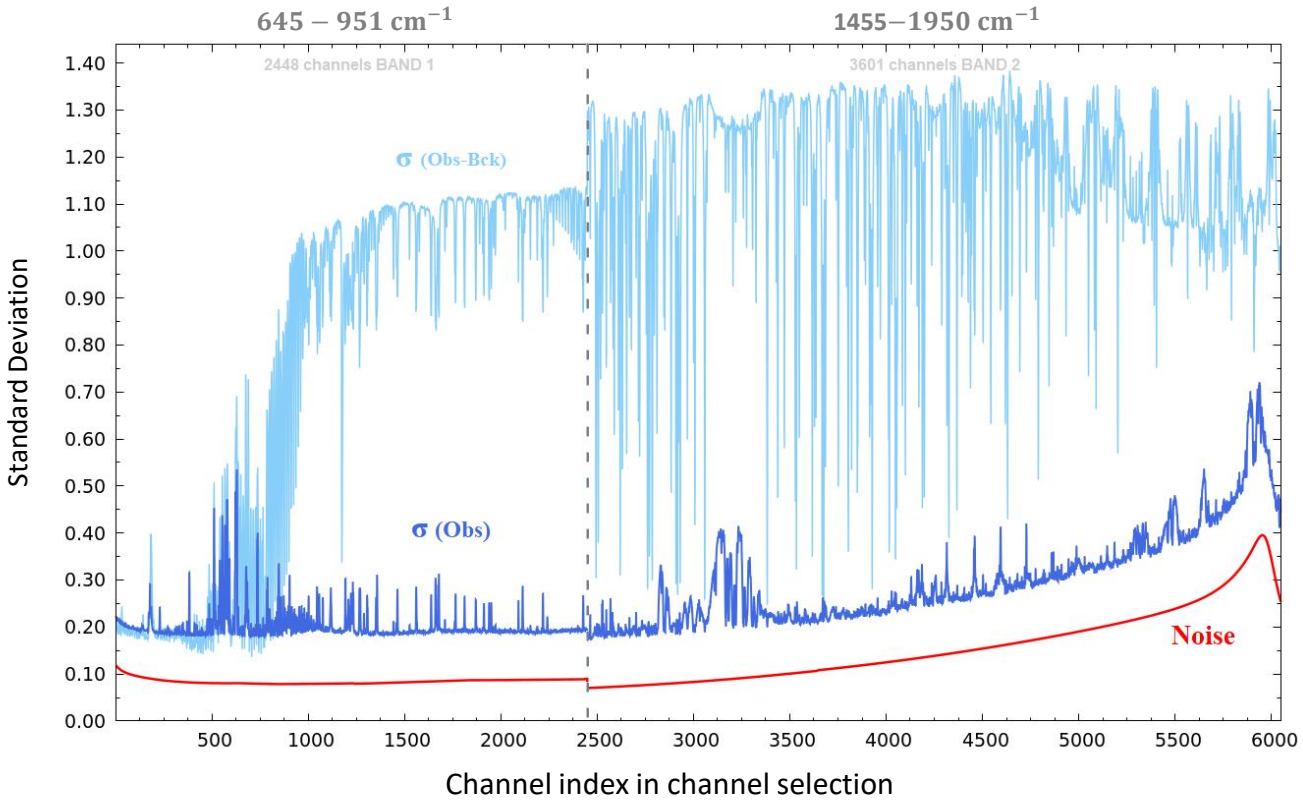
To estimate full $\mathbf{R}$ matrices, a diagnostic procedure introduced by Desroziers et al. (2005) was implemented.

$$\mathbf{R} = E\{[\mathbf{y}-H(\mathbf{x}_a)]\,[\mathbf{y}-H(\mathbf{x}_b)]^{\mathrm{T}}\}$$

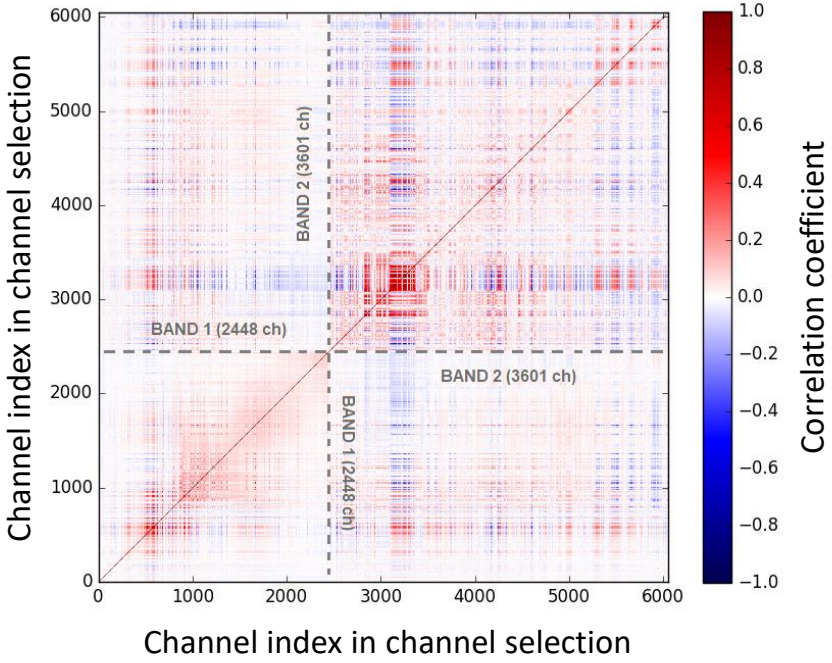
*(Observ – Analysis)* *(Observ – Backg)*

# Preparatory study towards the channel selection

**Diagnostic error standard deviations from 1D-Var output and instrument noise**
6049 channels




**Diagnostic for correlation matrix**
6049 channels from BAND 1 and BAND 2



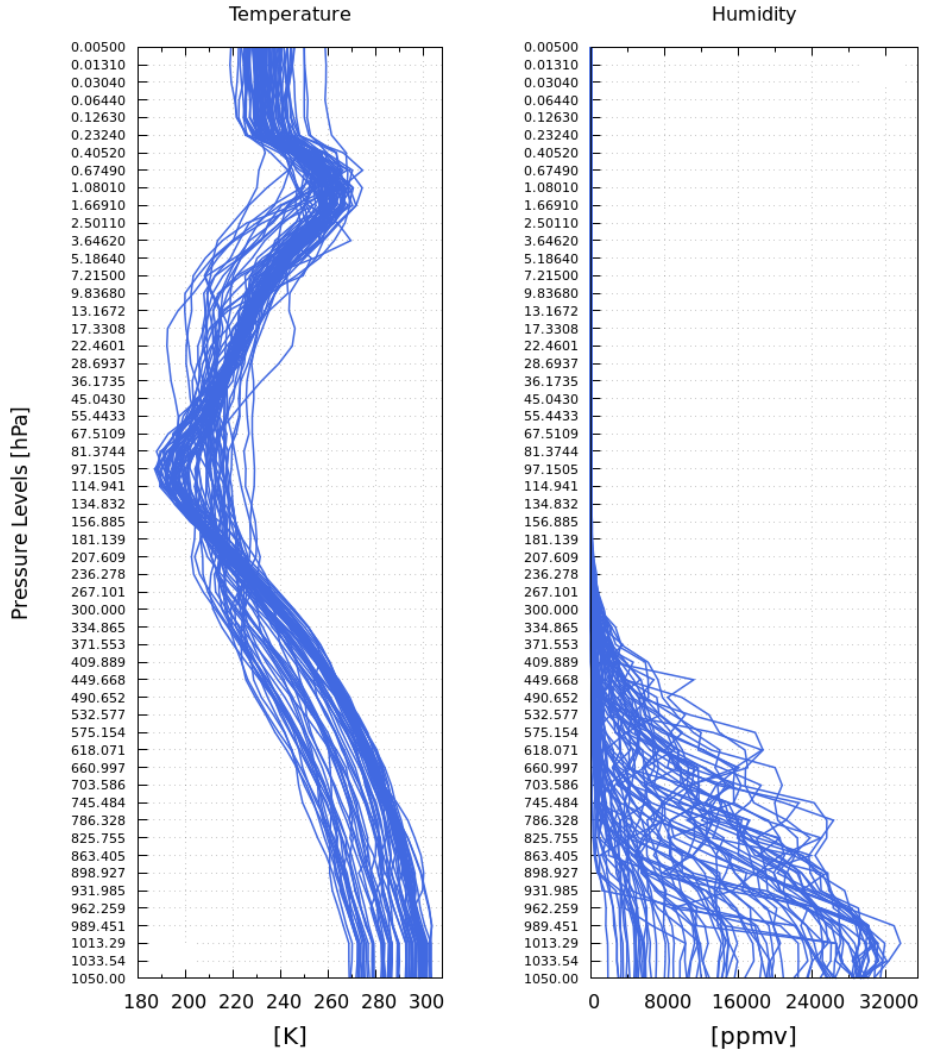
# Case Study for Channel Selection

6267 profiles case study set can be further thinned

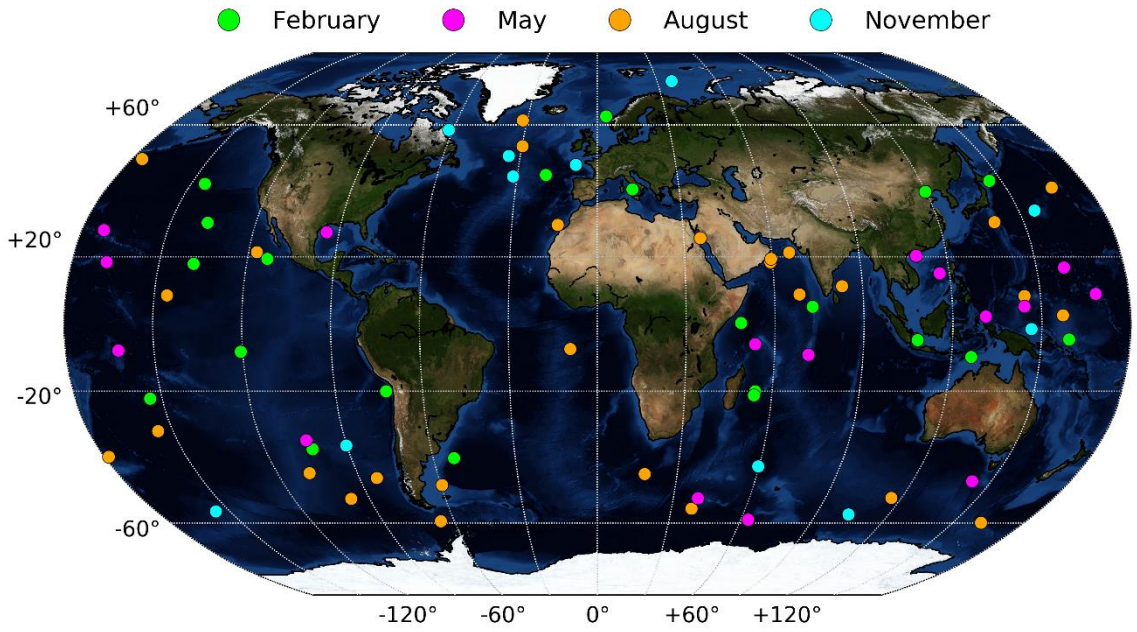
**Agglomerative hierarchical clustering technique:**

profiles grouped by evaluating distance in terms of temperature and humidity

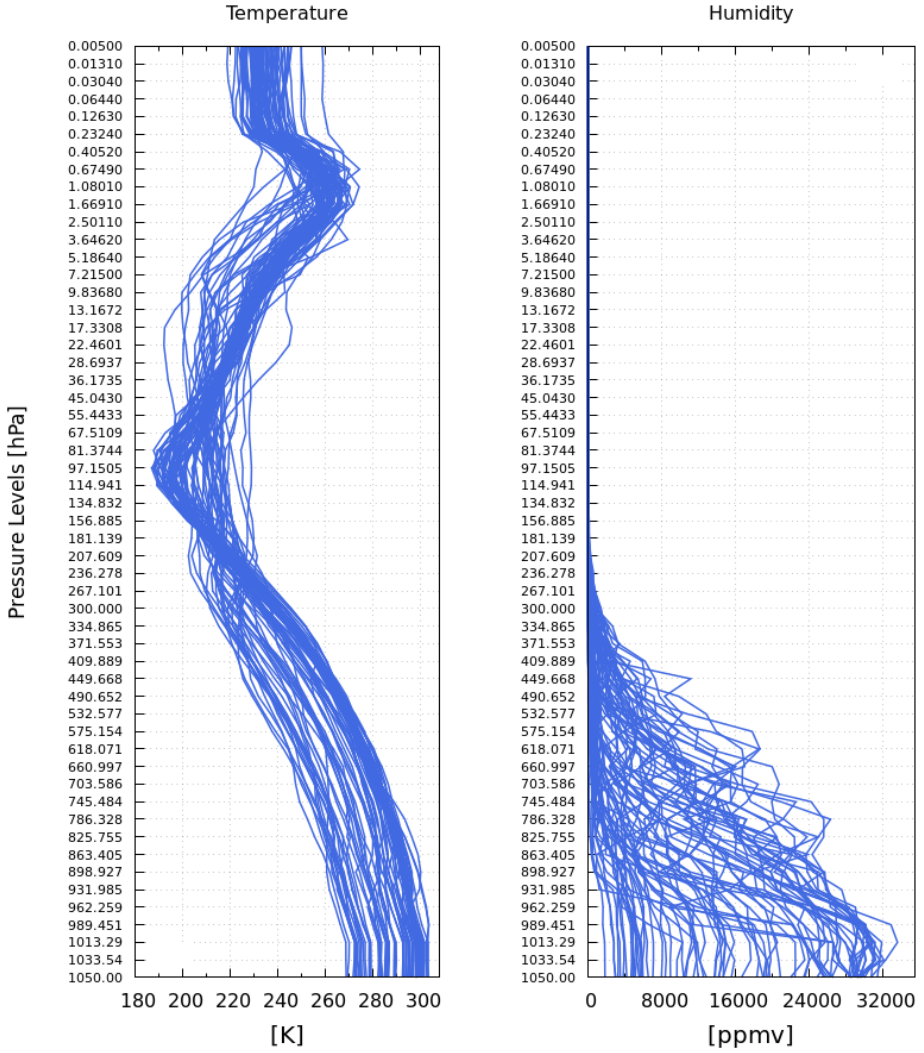
**77 profiles representative of different atmospheric conditions**
(typical and extreme cases)



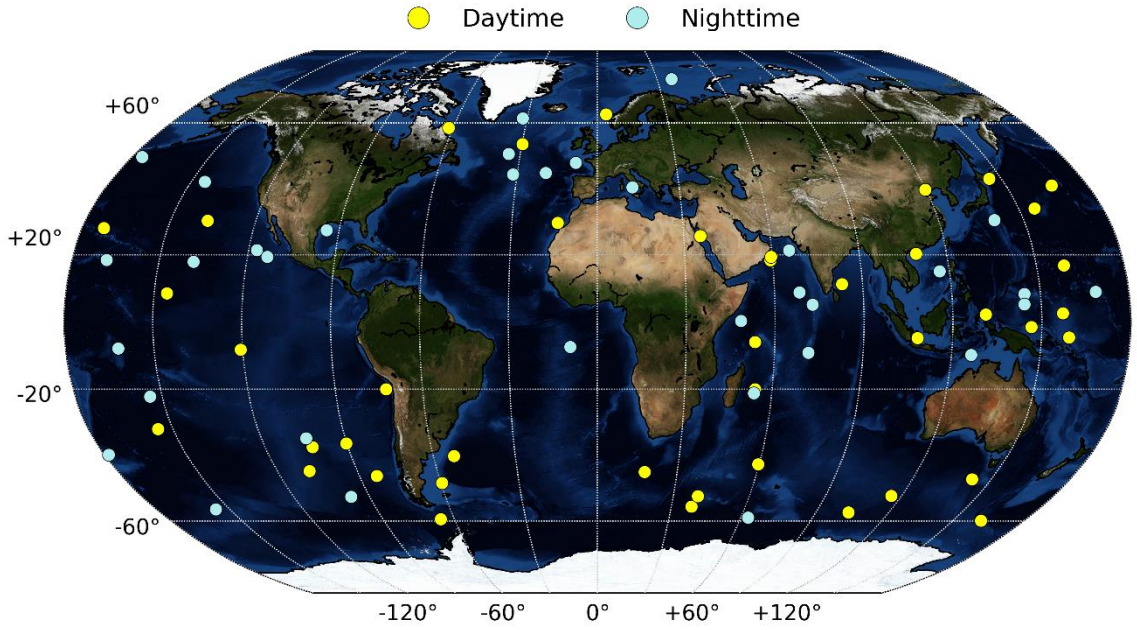
# Case Study for Channel Selection



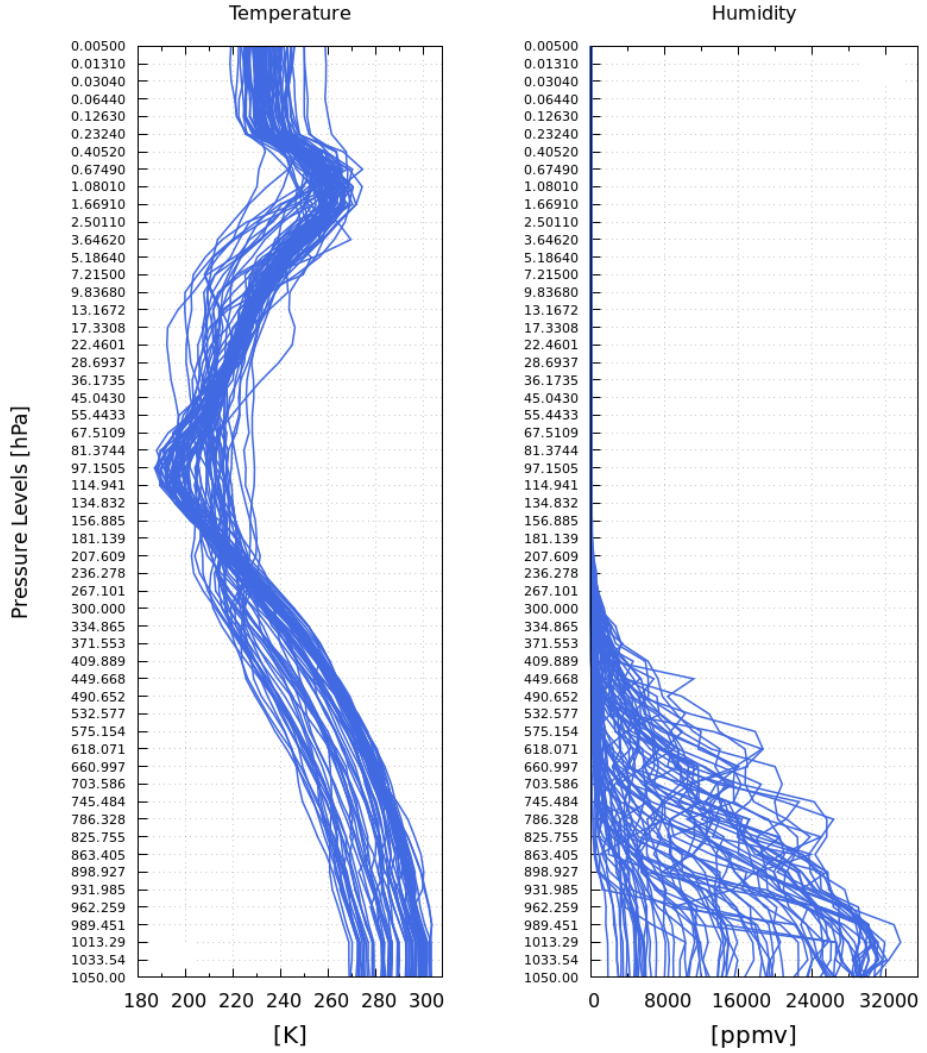

**77 profiles representative of different atmospheric conditions**
(typical and extreme cases)



# Case Study for Channel Selection




**77 profiles representative of different atmospheric conditions**
(typical and extreme cases)



# Methods

- ✓ 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***)

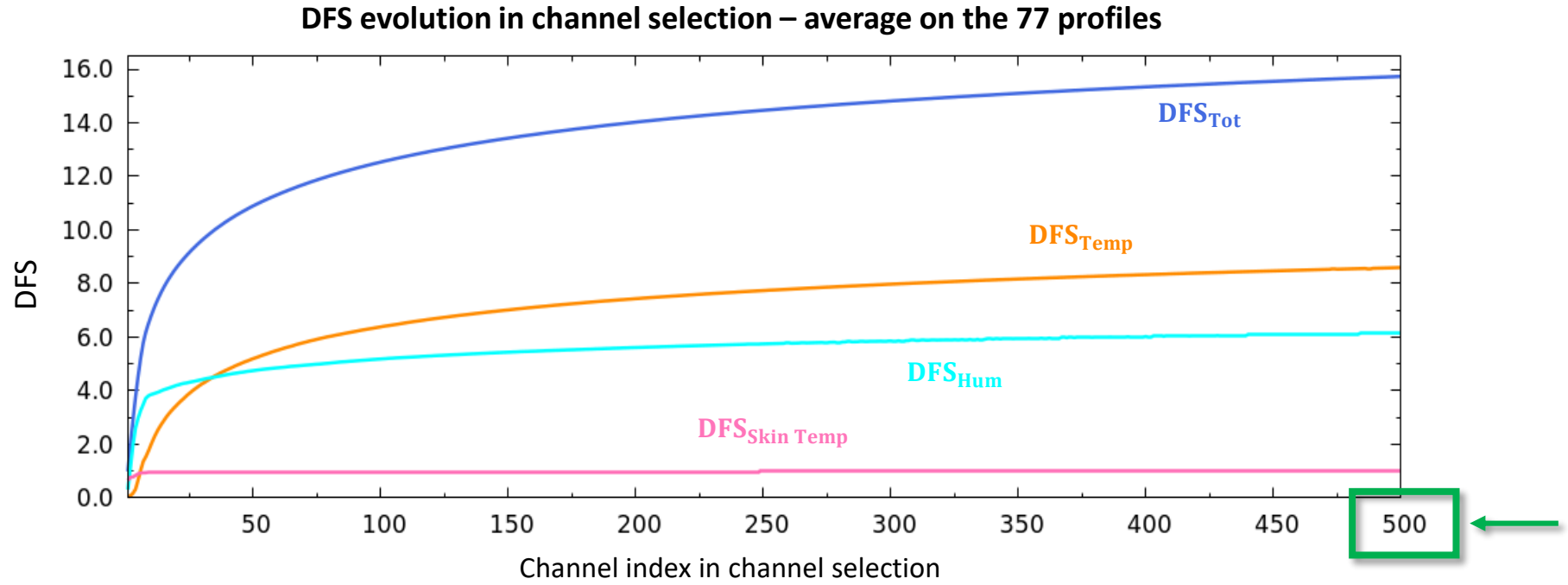
$$\mathrm{DFS} \coloneqq Tr\left(\mathbf{I} - \mathbf{AB}^{-1}\right) = Tr(\mathbf{I}) - \mathrm{Tr}[(\mathbf{I} + \mathbf{BH}^{\mathrm{T}}\mathbf{R}^{-1})^{-1}]$$

$$\mathrm{DFS}_{\mathrm{Tot}} = \mathrm{DFS}_{\mathrm{Temp}} + \mathrm{DFS}_{\mathrm{Hum}} + \mathrm{DFS}_{\mathrm{Skin\ Temp}}$$

$\mathbf{A}$ = analysis-error covariance matrix  $\mathbf{B}$ = backg-error covariance matrix  $\mathbf{R}$ = analysis-error covariance matrix

$\mathbf{H}$ = matrix of the Jacobians

# Channel Selection




- ✓ 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).

# Channel Selection

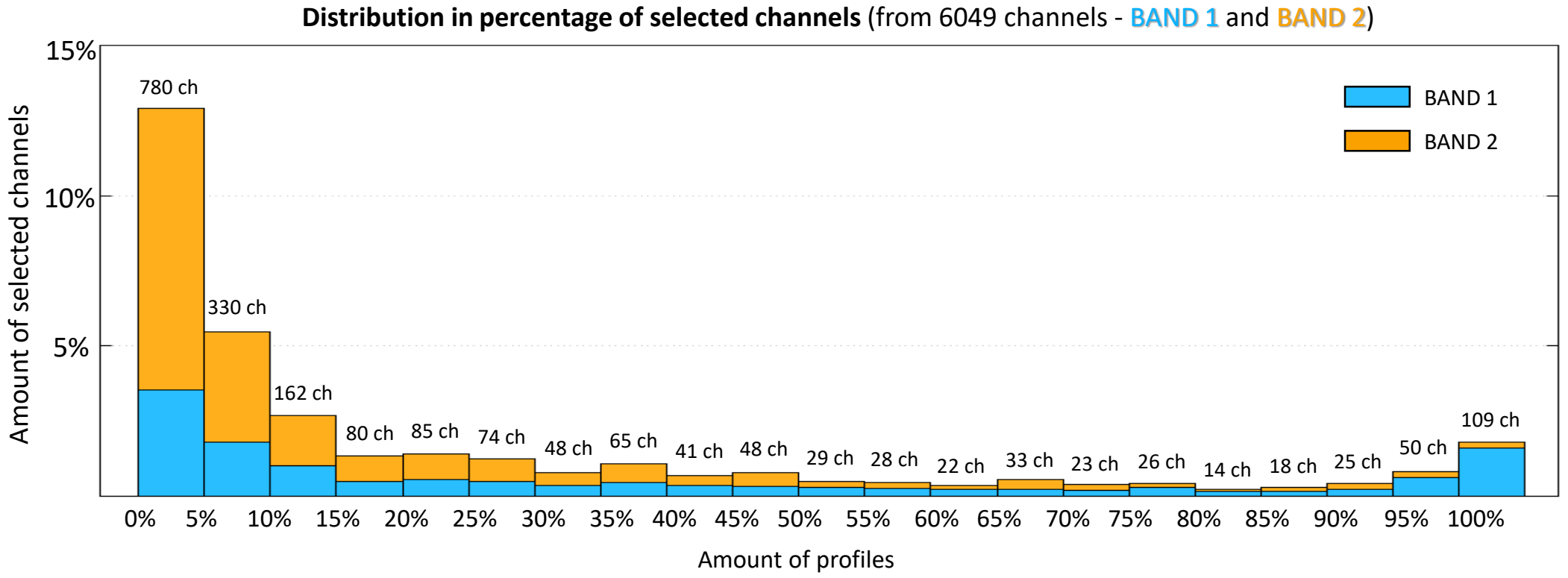
We performed an analysis to evaluate the percentage of choice of the different channels:

- ✓ **65%** of the total amount **of channels** is **never selected** (3959 out of 6049)

# Channel Selection

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

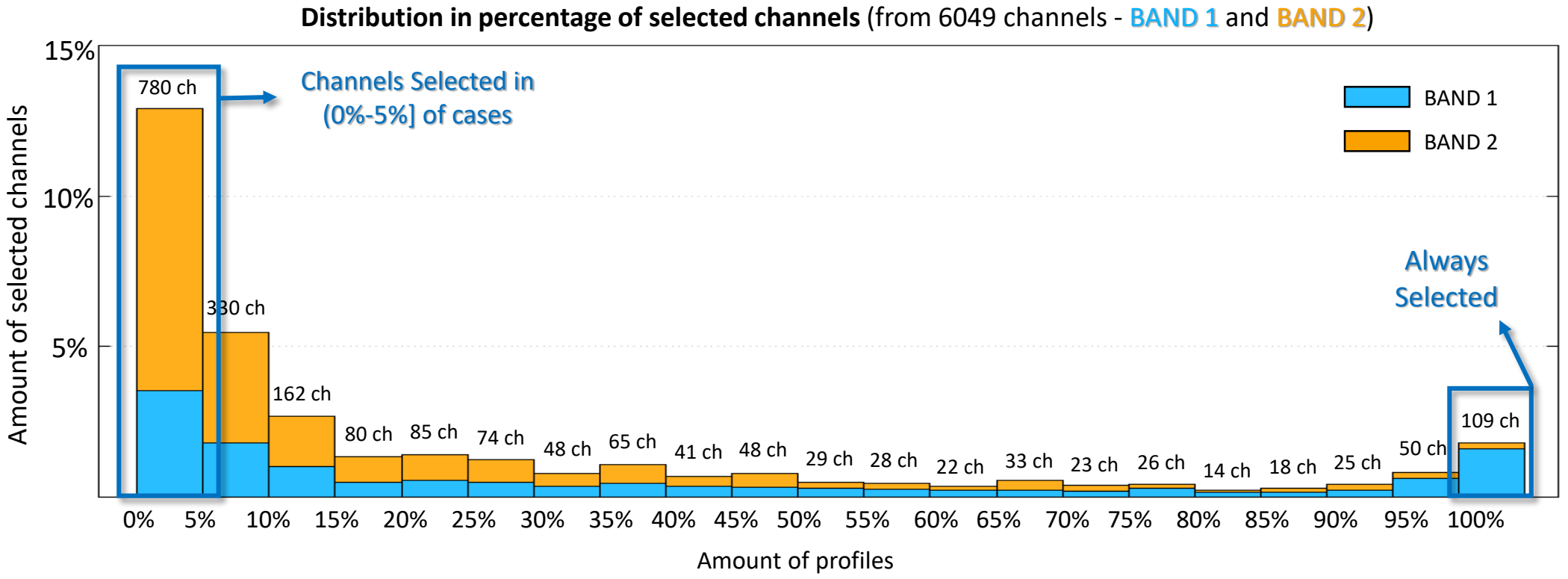
- ✓ **65%** of the total amount **of channels** is **never selected** (3959 out of 6049)
- ✓ **35%** (2090 channels) **selected at least once**



# Channel Selection

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

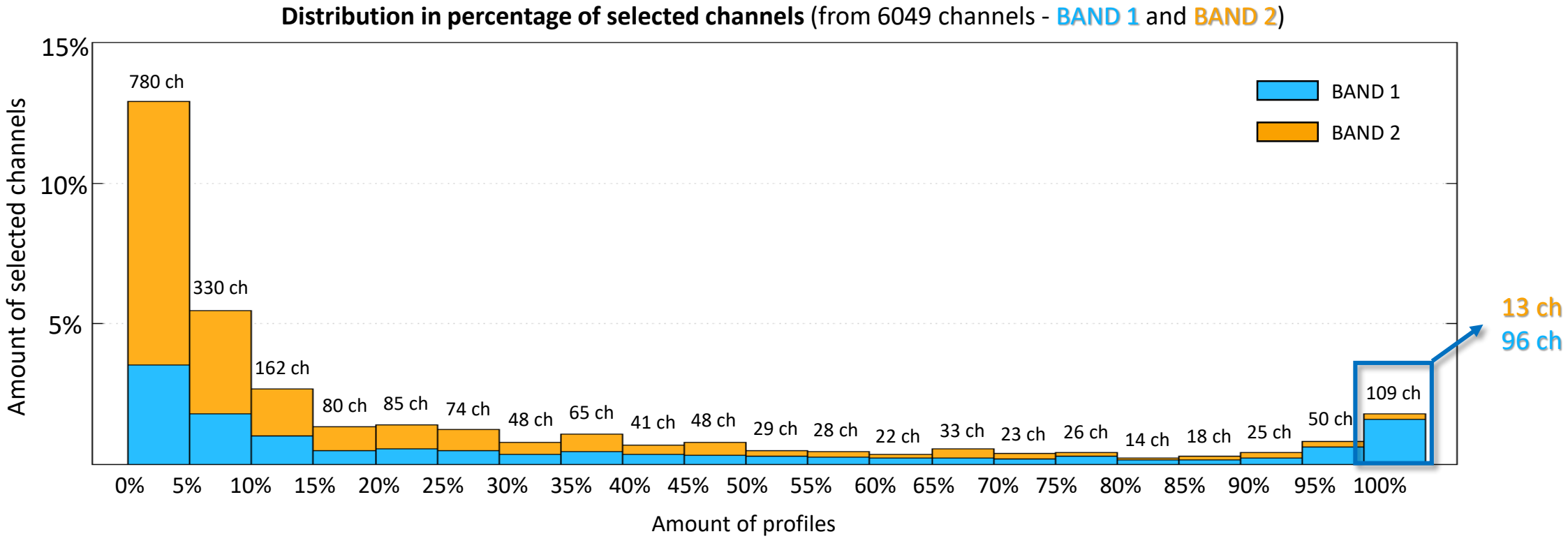
- ✓ **65%** of the total amount **of channels** is **never selected** (3959 out of 6049)
- ✓ **35%** (2090 channels) **selected at least once**



# Channel Selection

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

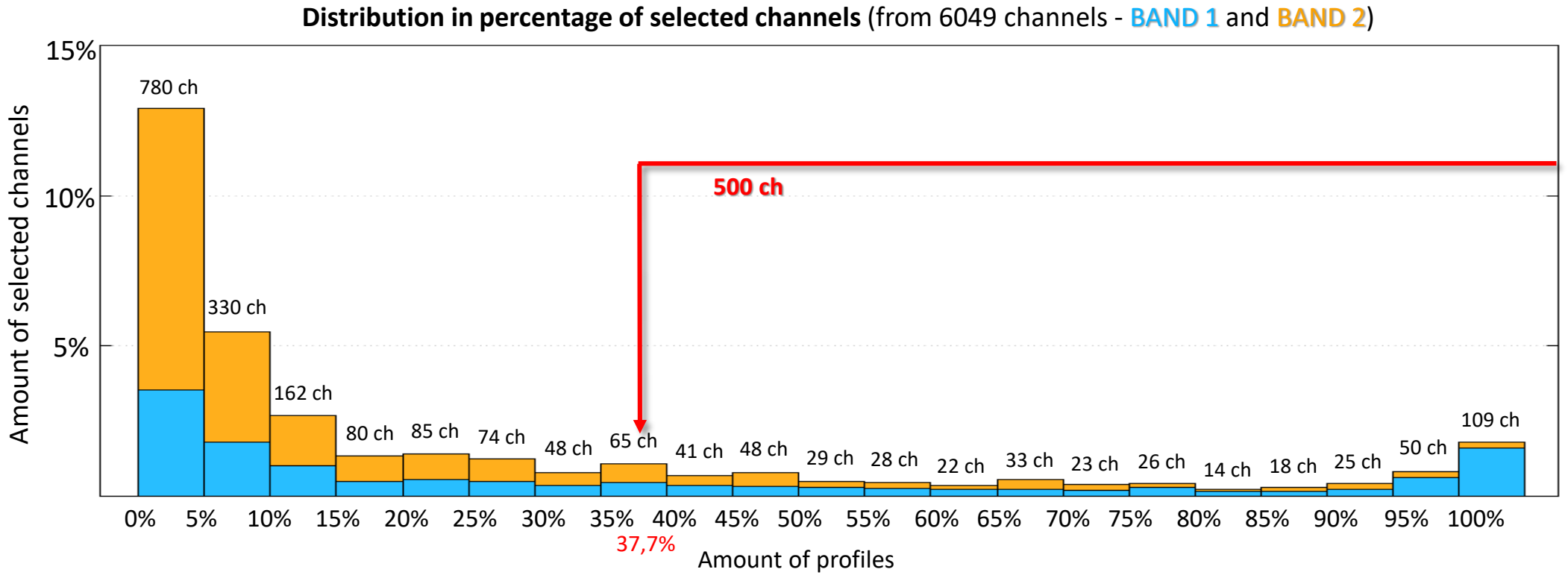
- ✓ **65%** of the total amount **of channels** is **never selected** (3959 out of 6049)
  - 2329 ch
  - 1630 ch
- ✓ **35%** (2090 channels) **selected at least once**



# Channel Selection

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

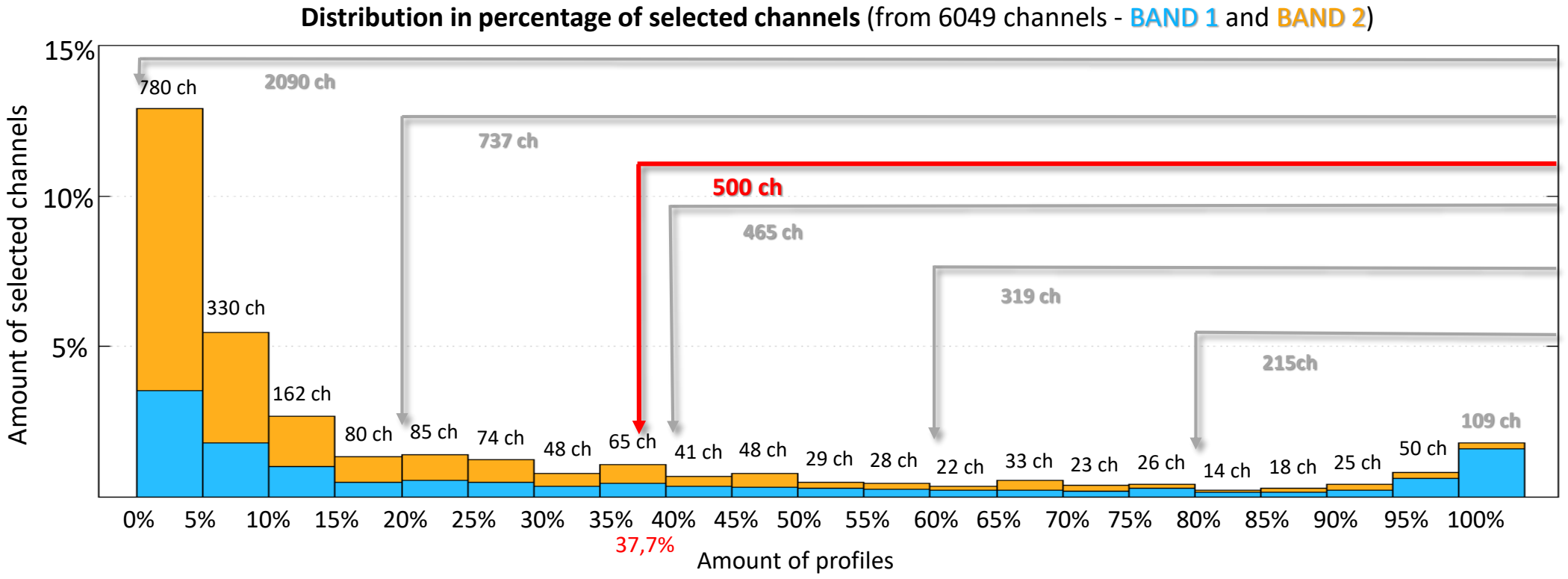
- ✓ **65%** of the total amount **of channels** is **never selected** (3959 out of 6049)
- ✓ **35%** (2090 channels) **selected at least once**



# Channel Selection

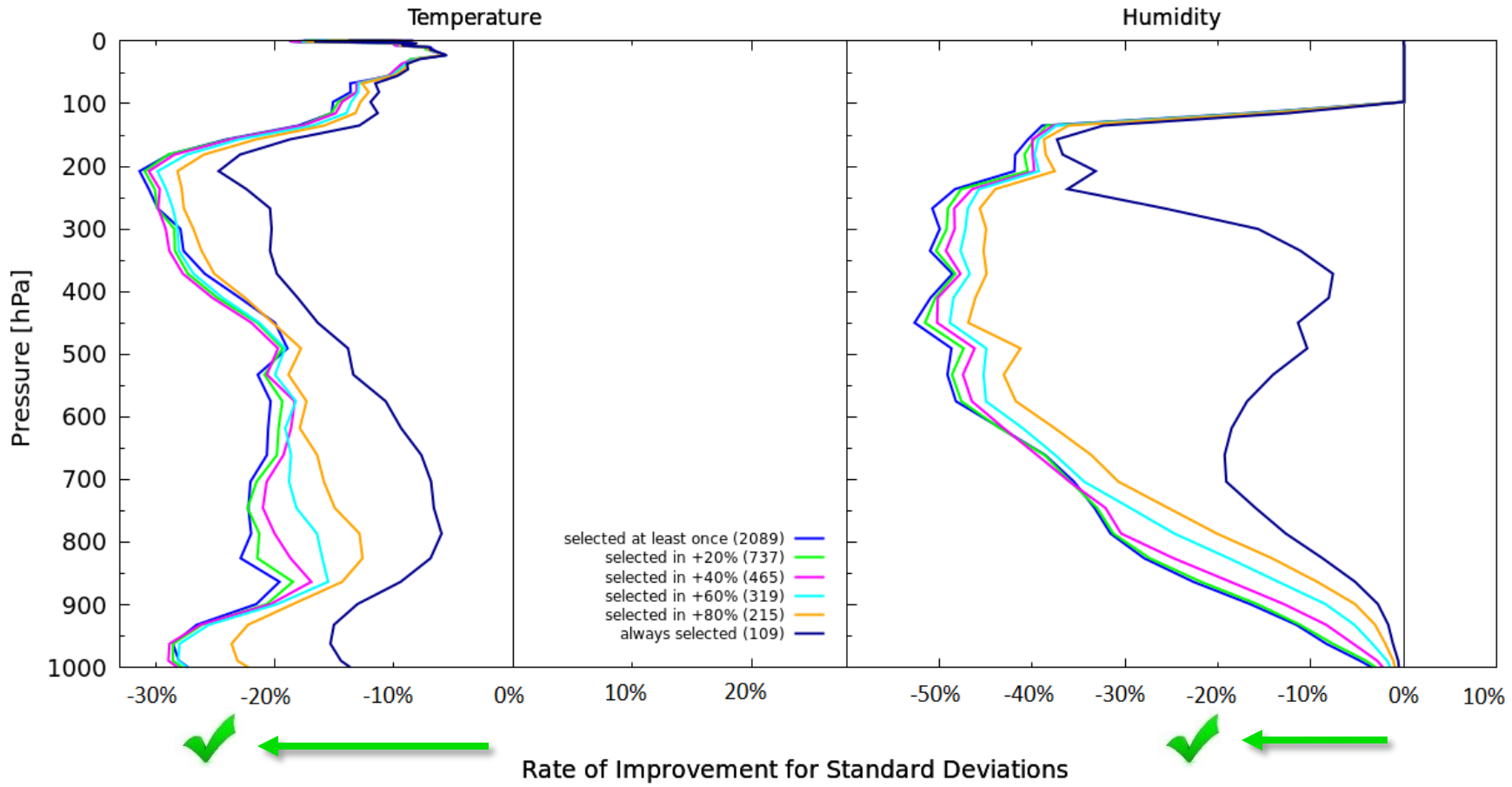
We performed an analysis to evaluate the percentage of choice of the different channels:

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



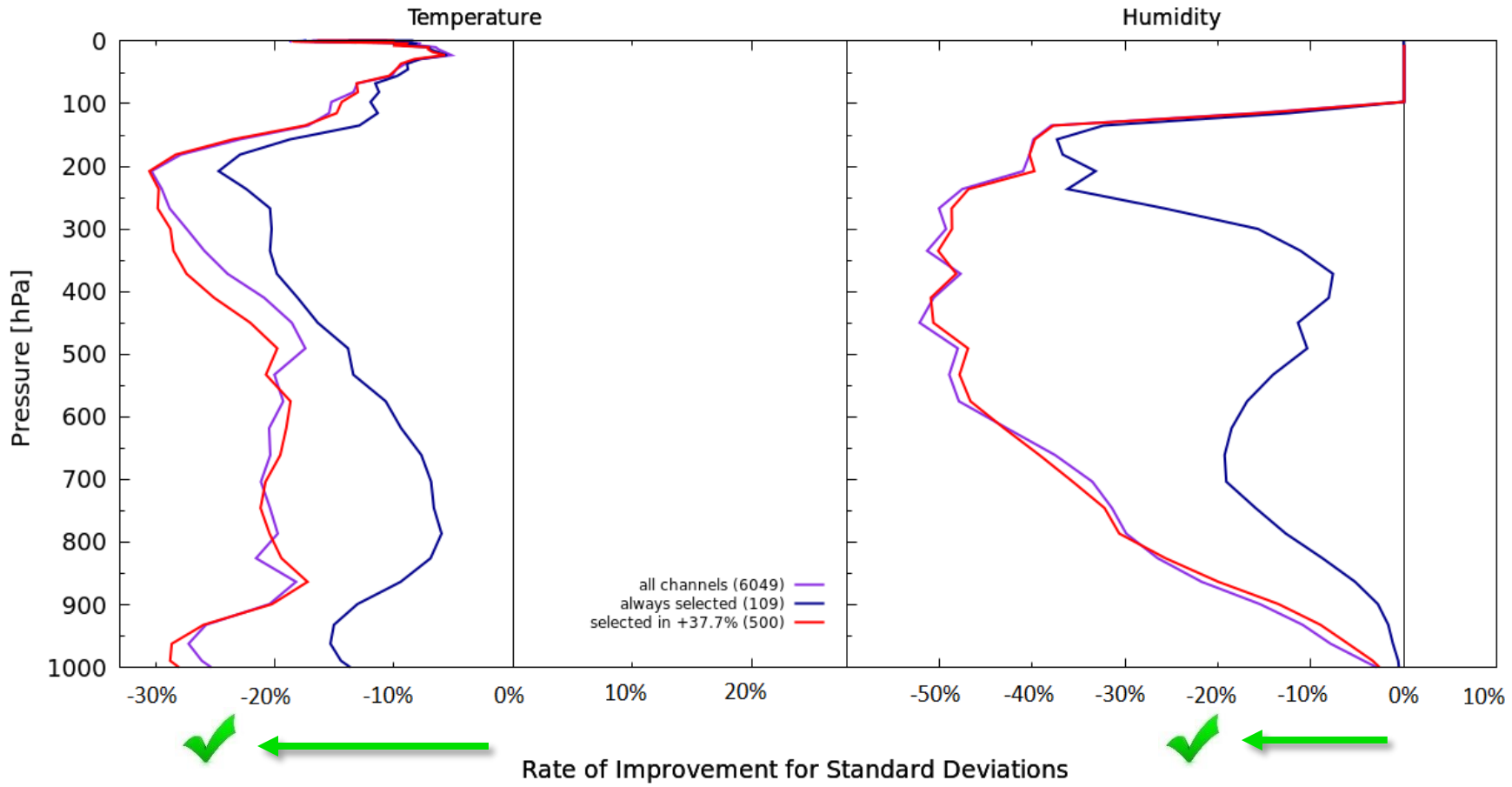
# ROI of Retrieval SD compared to the Background

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.

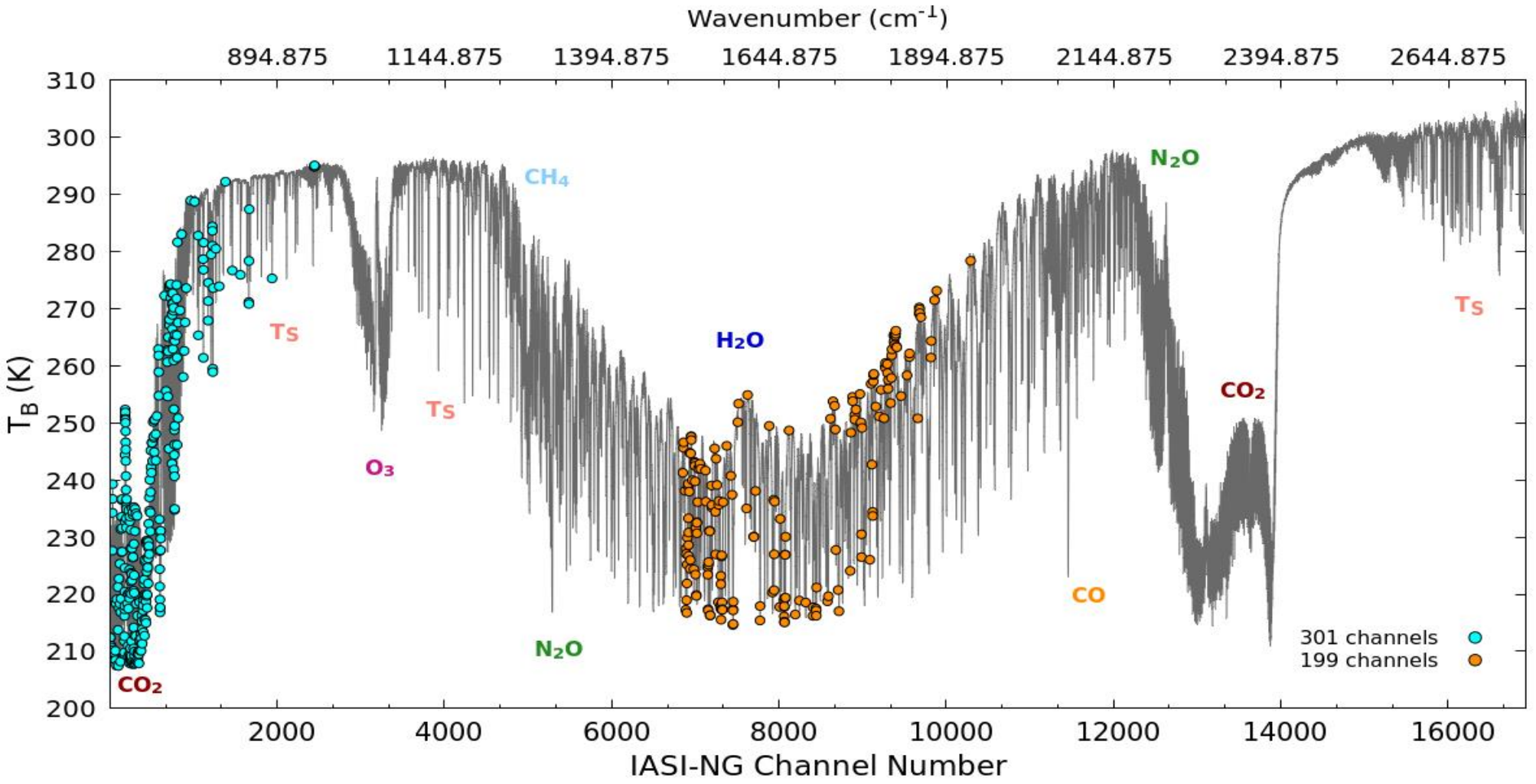


# ROI of Retrieval SD compared to the Background

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



# 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