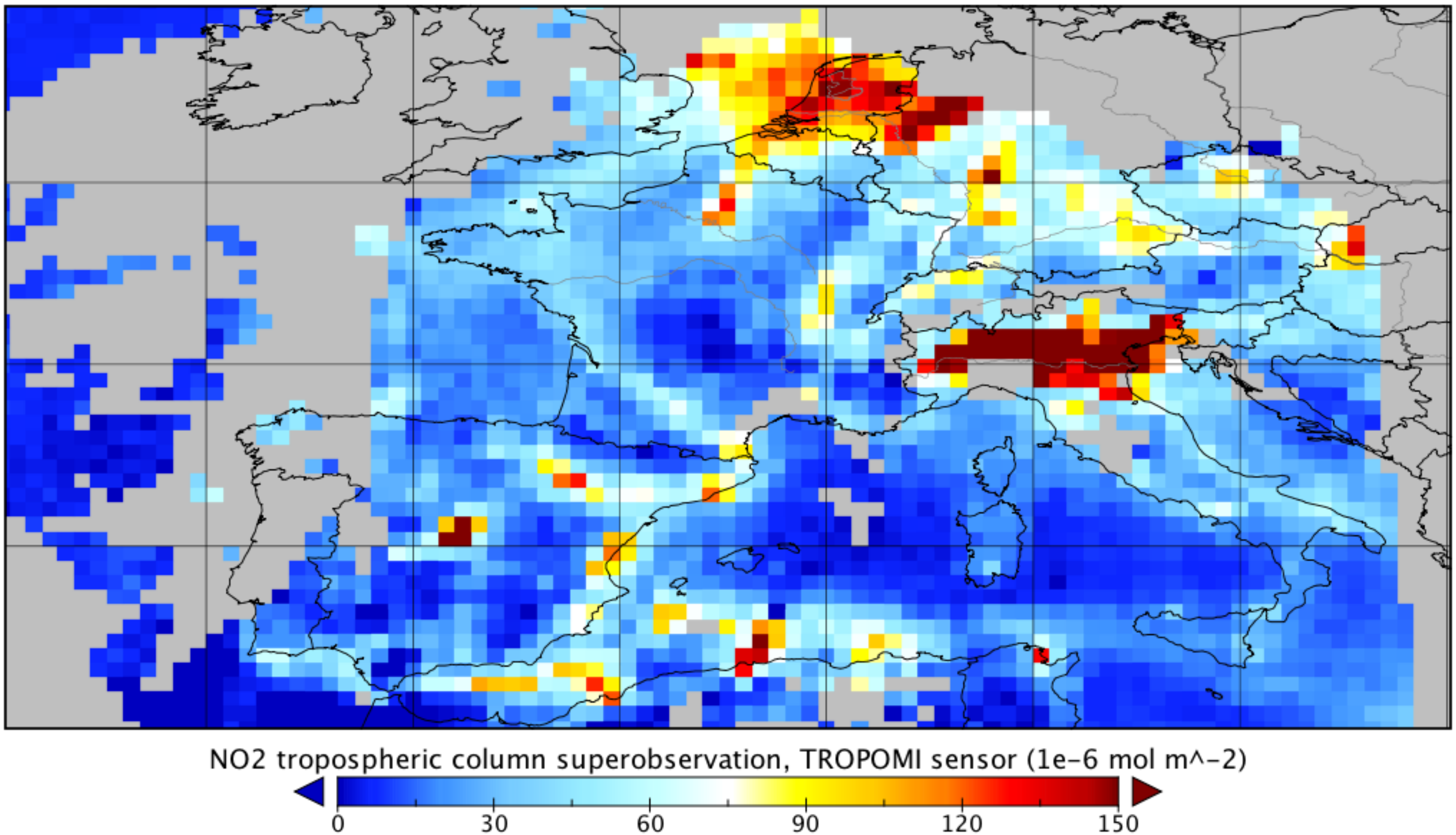
Making better use of high-resolution satellite data in data assimilation



S5P-VT meeting ESTEC, Feb 2018

Question raised:

"How can data assimilation deal with measurements showing a large variability (e.g. NO2) on the scale of a few km"

Possible approach: **superobservations**

Reference:

Boersma, Vinken, Eskes, GMD 2016

What are superobservations ?

Combination of individual observations into a single effective observation representing the horizontal length scale resolved by a model / assimilation system.

In practice this could be
one combined observation per model grid cell.

Superobservations are assimilation system dependent.

CAMS-global versus TROPOMI

CAMS resolution is about 0.4x0.4 degree,
or about 2000 km^2

TROPOMI resolution is about 3.5x5.5 km^2
or about 20 km^2

Meaning there are about 100 observations / grid cell

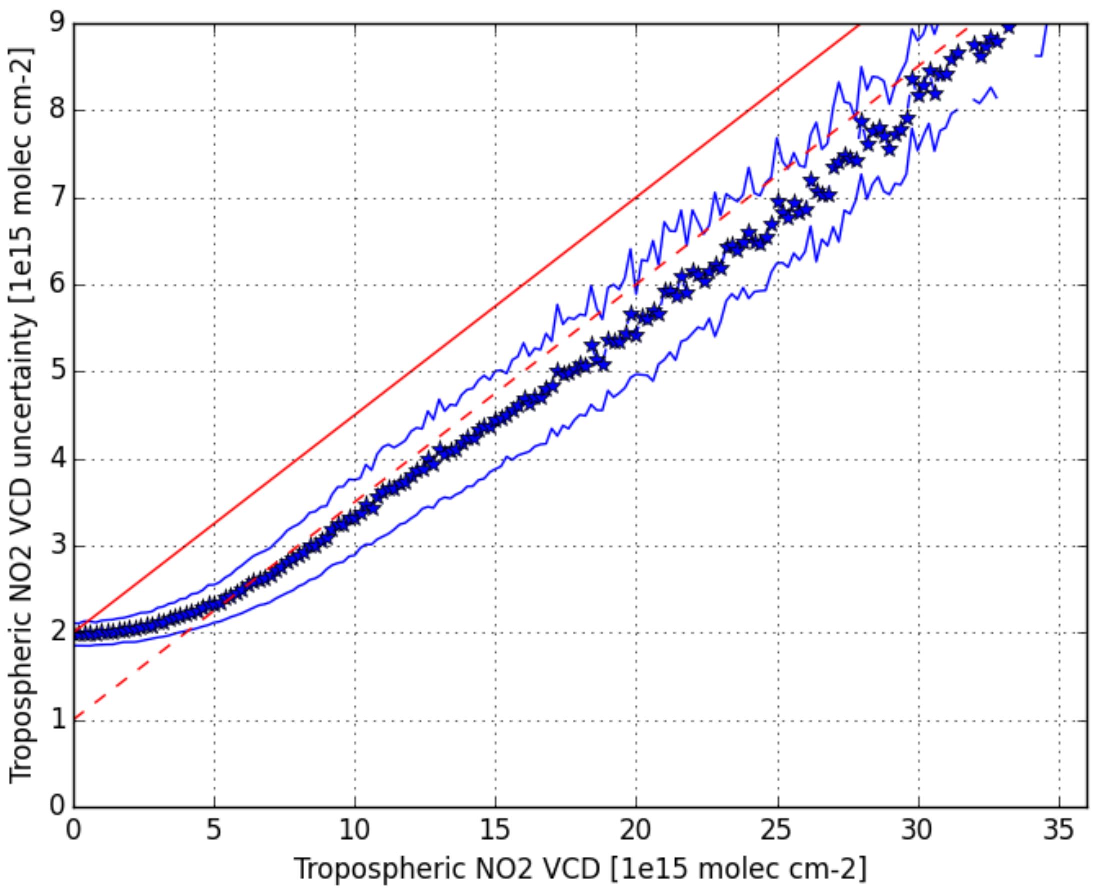
Notes:

- An instrument like TROPOMI provides about 20 million observations per day. Assimilating this amount is computational intensive / not desired / not acceptable.

- Common data assimilation practice is **thinning**: e.g. random selection of 1% of all observations. This is a **bad approach** for strongly varying concentrations of short-lived tracers. Leads to very large representativity errors.

Notes:

✦ Assimilating multiple observations per grid cell may be a **bad approach** as well: the assimilation system will give weights to the observations according to the uncertainty of the individual measurements which may lead to **biased results.**



How to construct superobservations ?

Individual observations inside one model grid cell should *not* be interpreted as
estimates of the mean of that grid cell, but as fully independent observations for their own respective footprint.

So: do **not** provide weights to the individual observations according to their uncertainty

Rather: construct a superobservation with a
weigth = area of overlap with the model grid cell (km^2)

How to construct superobservations ?

$$\hat{y}_o = \frac{\sum_i w_i \hat{y}_i^o}{\sum_i w_i}$$

w_i = overlap area between footprint and grid cell (km^2)

Note: the averaging kernels are averaged in the same way

Superobservations

Advantage:

Number of observations offered to the assimilation is strongly reduced

factor 100 for TROPOMI-CAMS case

Superobservations uncertainty

Advantage: we can (partially) account for
spatial correlations between observation uncertainties

$$\sigma_N^2 = \langle \varepsilon_N^2 \rangle = (1-c)\sum_i w_i^2 \sigma_i^2 + c\left(\sum_i w_i \sigma_i\right)^2$$

c = spatial correlation between individual observation uncertainties (not part of the data product)

Systematic part

The uncorrelated, random part reduces like 1/sqrt(n)

Correlations between observation uncertainties ?

Correlation depends on the error source

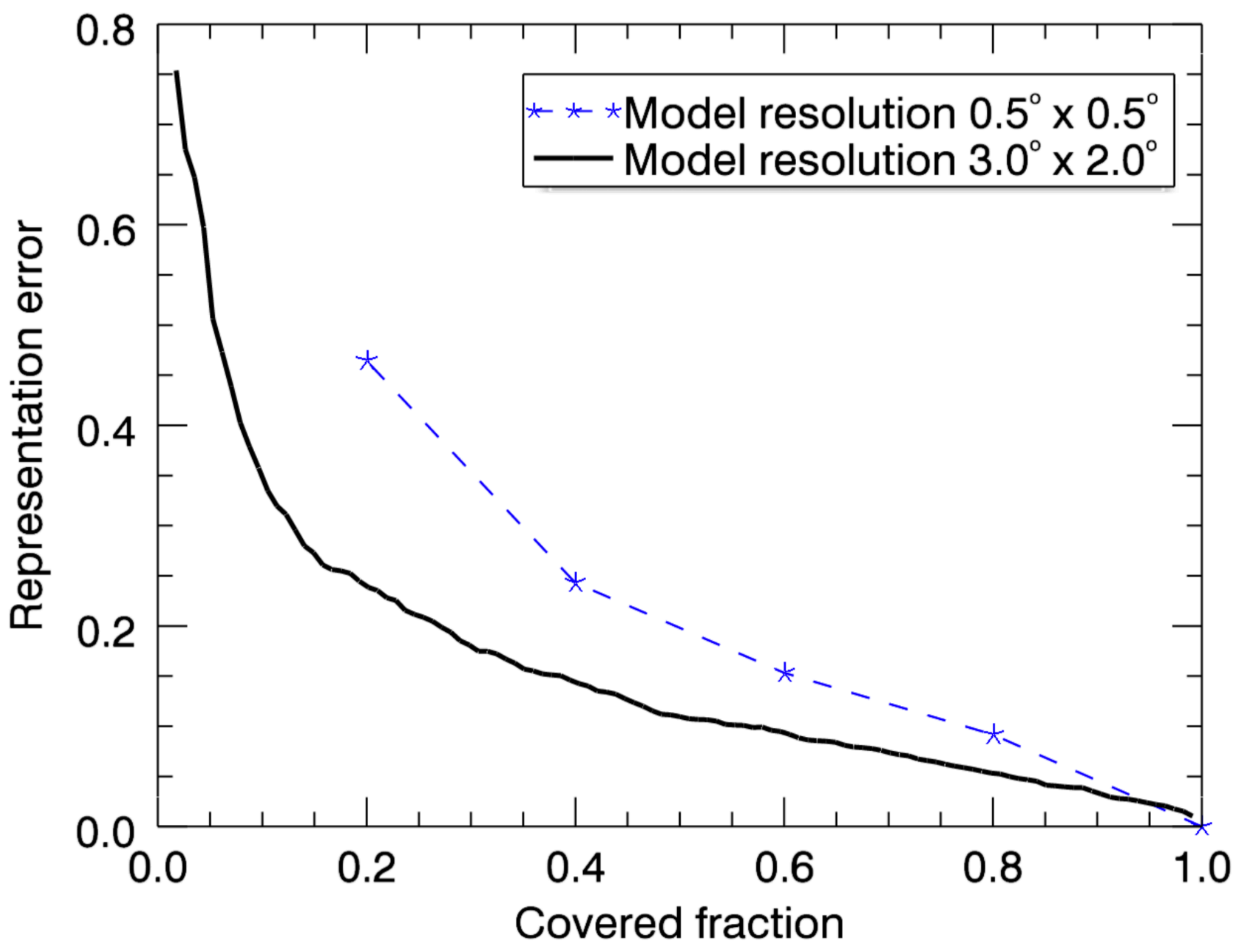
Example: NO2, TROPOMI

Uncertainty due to	Random	Systematic
Slant column	**X**	X
Cloud fraction	X	X
Cloud pressure	X	X
Surface albedo	X	X
Stratosphere	x	**X**

Superobservations uncertainty

Advantage:
we can account for representativity error

Note:
This scales with the variability within the grid cell.



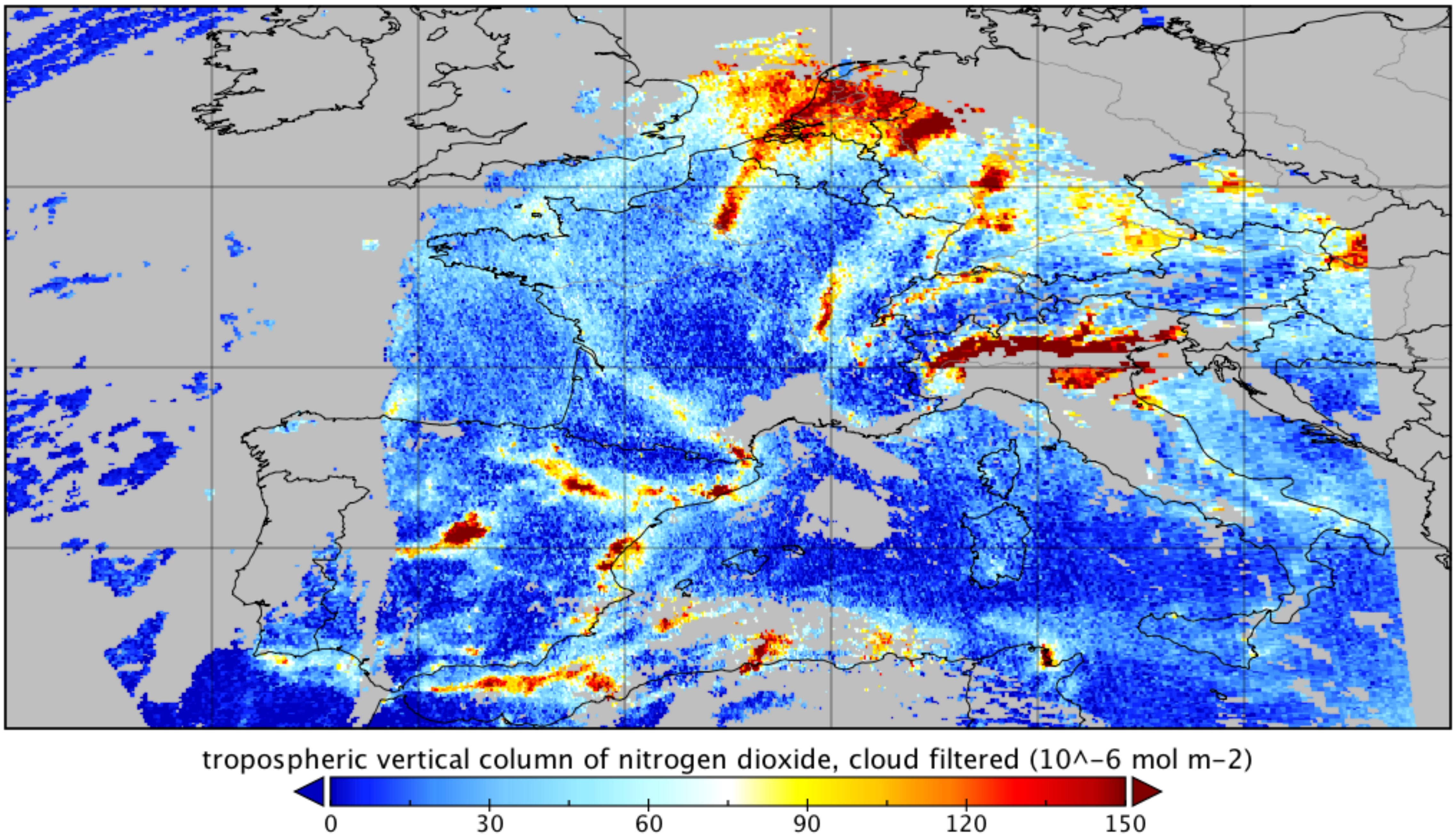
Remark: additional filtering

Depending on the application, we may apply additional pre-filtering.

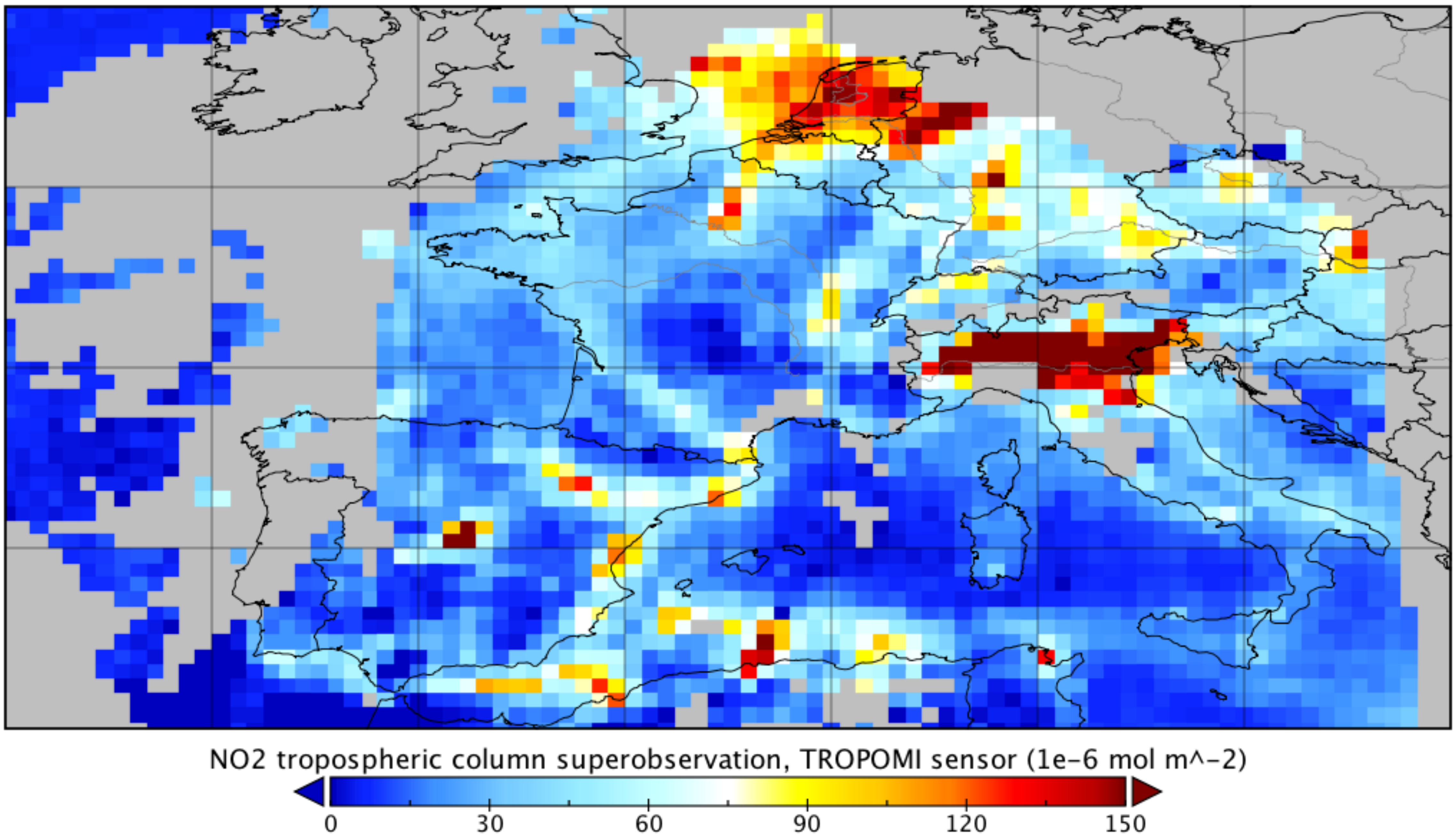
Common filter for NO_2:

cloud radiance fraction < 0.5

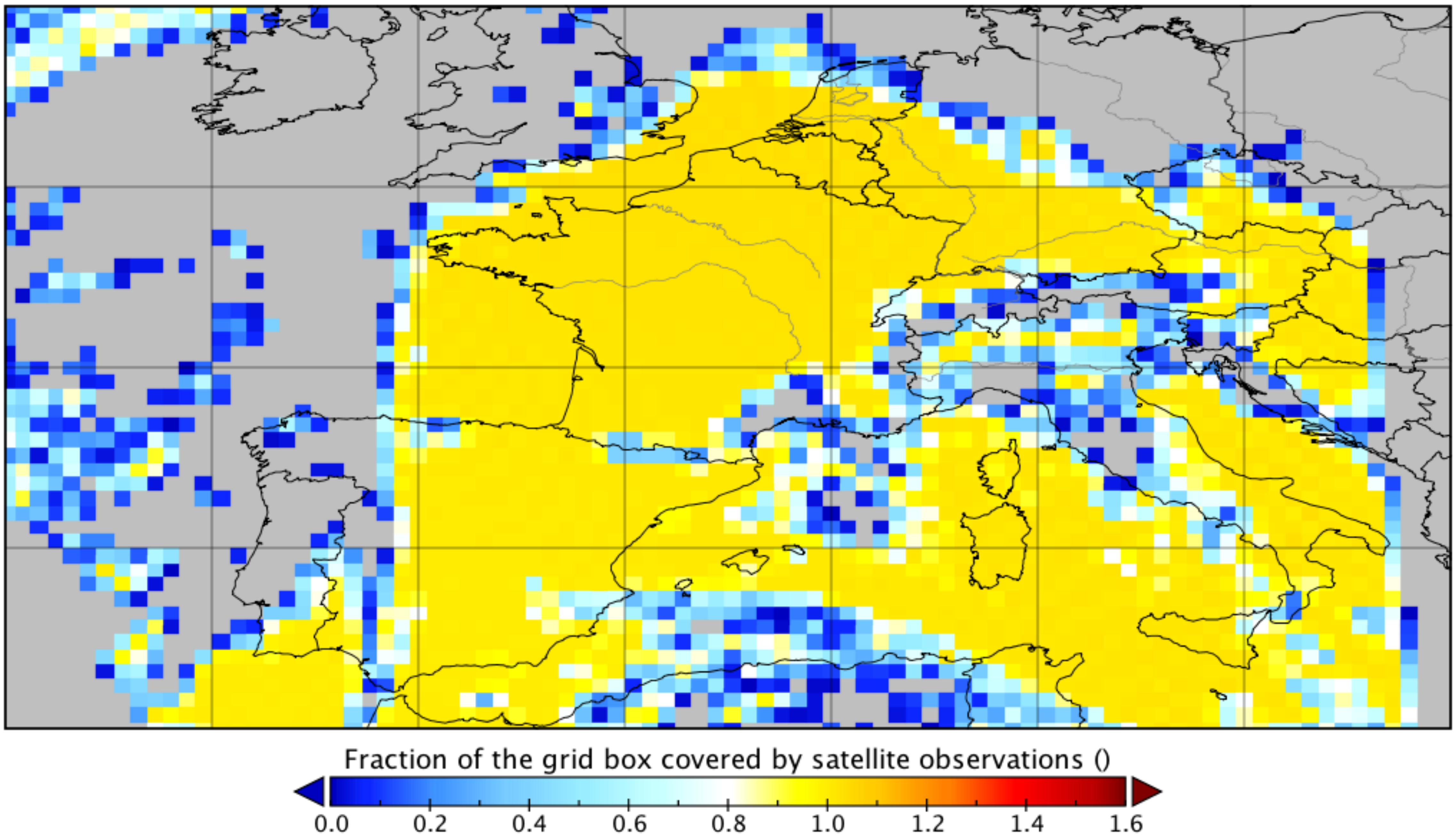
TROPOMI NO2 observations



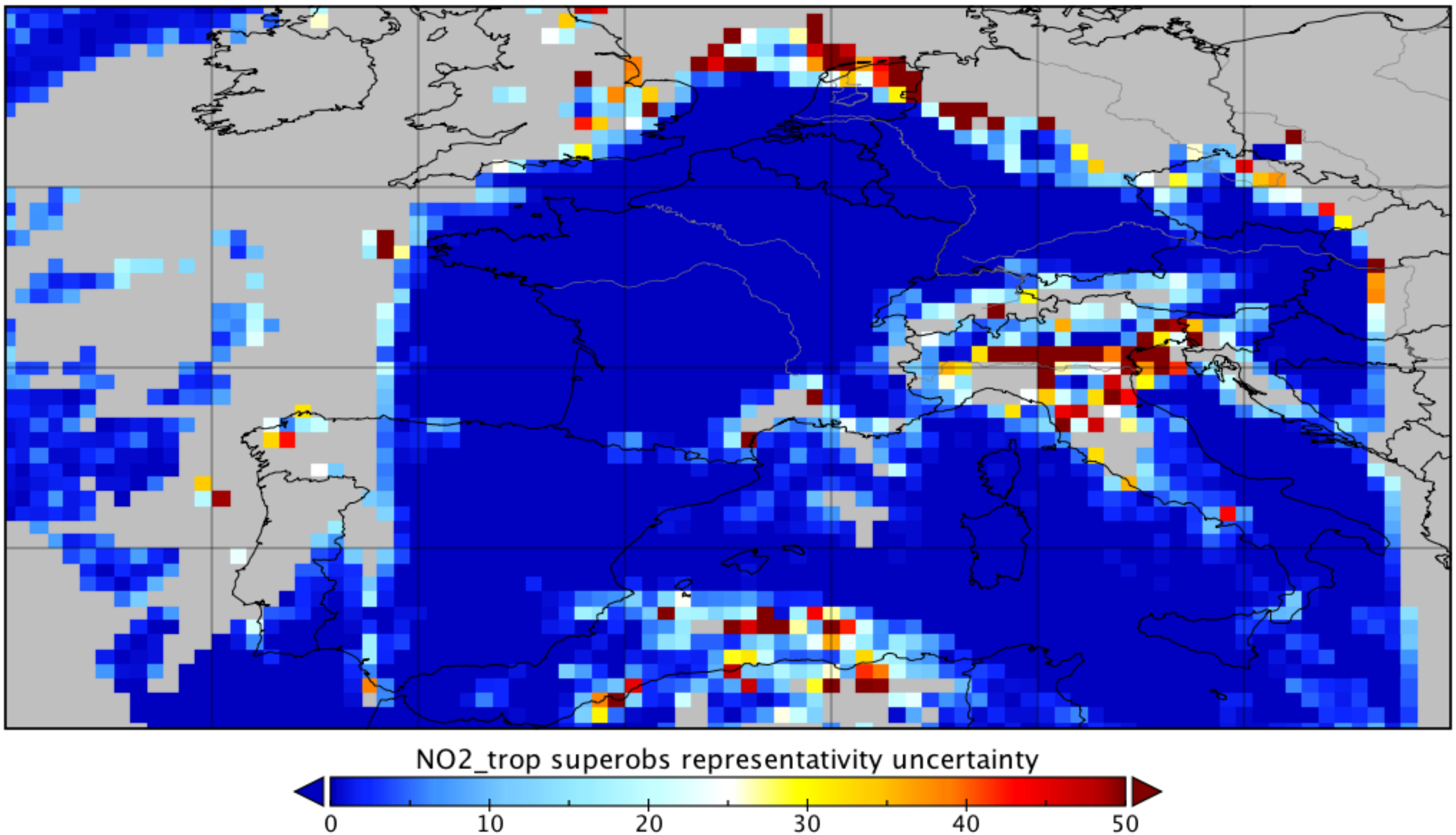
TROPOMI NO2 superobservation on 0.4x0.4 degree grid



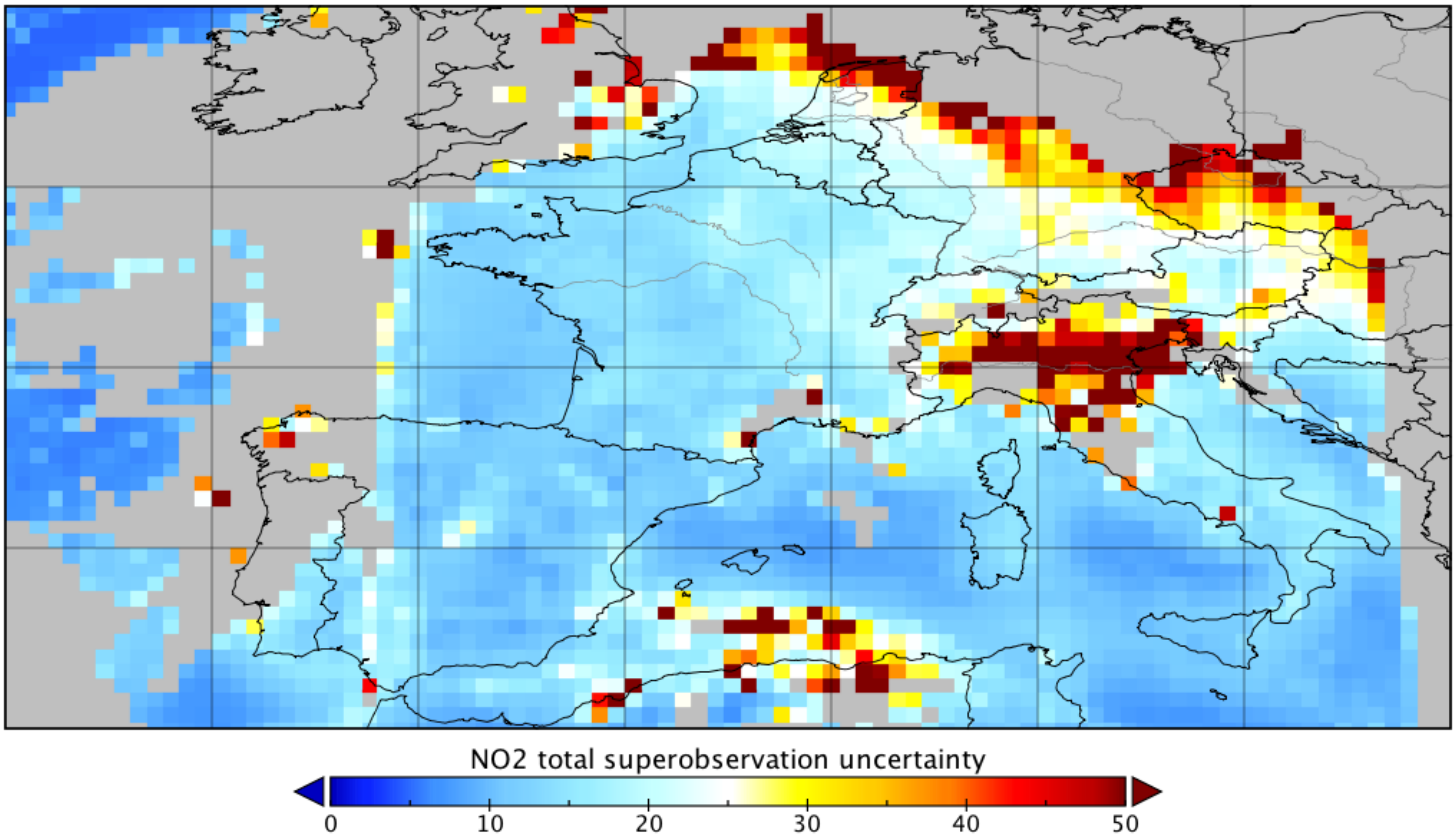
Fraction grid box covered by observations



TROPOMI NO2 superobservation representativity error

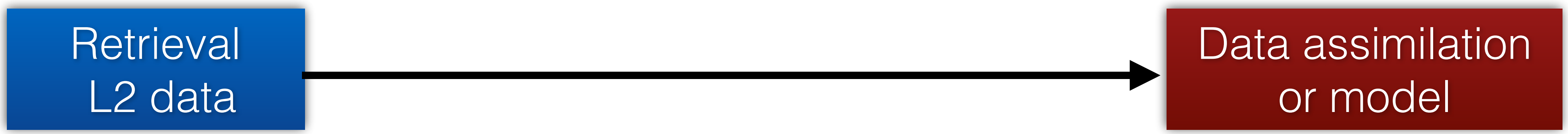


TROPOMI NO2 superobservation total uncertainty



Conclusion:

Proposal to replace:



by:

This idea is discussed with ECMWF for TROPOMI NO2