Special processing of TROPOMI NO_2 for lightning NO_x research

Jos van Geffen (KNMI)

(01 September 2020)

Saturation of the TROPOMI detector in Band 4 (used for the NO_2 retrieval) and Band 6 (used for FRESCO cloud retrieval) occurs mostly over bright clouds and may cause large areas of ground pixels with missing NO_2 vertical column data, either because of problems with the NO_2 retrieval or because of missing cloud data or both.

This is a pity in view of research into NO_2 in combination with lightning. For these special cases it may be useful to start from the NO_2 slant column data, but this should be done with care. This document describes some issues related to dealing with saturation in the NO_2 processor and discusses the additional data variables made available in the specially processed orbit files.

Initial note on the processing_quality_flags and the qa_value

The variable processing_quality_flags is a combination of error flags and warnings. The error flags are are listed in the first 7 bits of the processing_quality_flags. If somewhere in the retrieval algorithm an error occurs for a given ground pixel, the corresponding value is set in processing_quality_flags and the rest of the retrieval chain is skipped for that ground pixel and variables with data from the rest of the retrieval will have a FillValue. In case of warnings one or more of the higher bits of processing_quality_flags may be set – these do not in themselves cause the ground pixel to be missing from the data.

Warnings in the processing_quality_flags or other criteria in the retrieval code (e.g. presence of clouds) may set the qa_value to something other than 1 (fully successful retrieval), while errors set in the processing_quality_flags automatically lead to qa_value = 0.

Since we are going to do special processing in order to try to salvage more ground pixels despite some error would under normal usage prevent the usage of the ground pixel, the qa_value cannot be used. Instead we will be looking at additional, and to some extent not available in regular files, information from the processor. Some of the error values of the processing_quality_flags are useful for this.

Since the errors in the processing_quality_flags are given in the first 7 bits of that variable, it is important to filter the variable when reading it. In my Python script I am using the following code:

```
datFile = netCDF4.Dataset(dataFile,'a')
dtMain = datFile.groups['PRODUCT']
```

```
dtPSD  = dtMain.groups['SUPPORT_DATA'].groups['DETAILED_RESULTS']
pqf_in = dtPSD.variables['processing_quality_flags'][0,:,:]
mask = 0b1111111
pqf = pqf_in & mask
```

The variable ‘pqf’ contains the error code; below ‘pqf’ is used as error code identifier.

One of the more common pqf to occur is 36, due to missing cloud information – that error is not relevant in case we consider SCD values to start from and here we will therefore ignore it; the error does mean there will not be any vertical column data for this ground pixel.

Dealing with saturated wavelength pixels

Saturation leads to one or more wavelength pixels in the radiance spectrum being flagged. These wavelength pixels are omitted from the spectrum when performing the NO_2 slant column (SCD) fit. Since this reduces the number of wavelength pixels available for the DOAS fit, we have to set a maximum on the number of wavelength pixels that may be flagged as saturated.

Since we know that in the L1B-v1.0 data currently in use the saturation flagging is not optimal and does not cover cases of so-called blooming (wavelength pixels that are saturated may receive so much light that part of the electrons in that particular detector pixel spill over to neighbouring wavelength pixels), the current NO2-v1.3 retrieval has a very conservative limit set – see Table below – and if that limit is exceeded, an error is raised (processing_quality_flags = 54)..

After all, wavelength pixels affected by saturation or blooming but not flagged as such (and thus not removed from the spectrum before the DOAS fit) appear as spikes in the reflectance and cause problems in the fit, i.e. give SCD value that make no sense.

Spikes in the reflectance may also occur due to other processes, e.g. the impact of high energy particle on the detector (notably over the South Atlantic Anomaly, but in principle this can happen anywhere along the orbit).

To better deal with such spikes, the NO2-v2.1 algorithm will be equipped with a spike, or: outlier, removal scheme: wavelength pixels with fit residual values larger than a given threshold are marked as outlier and removed from the spectrum, after which the DOAS fit is done again. (To avoid ending up in a loop, there is no second check on outliers after that.) In view of spectral quality we have to set a limit to the number of outliers we allow to be removed before issuing an error (processing_quality_flags = 55) – see table below.

If the maximum number of outlier allowed is exceeded, the DOAS fit is *not* performed again and the SCD value given in the variable is the outcome of the retrieval with those outliers, i.e. a highly untrustworthy value.

For the special processing discussed here we allow for more outliers than we would normally do in the processing – see table below – but obviously the fit results of these extra ground pixels have to be treated with care.

In the forthcoming L1B-v2.0 data the flagging for saturation and blooming will be much improved (both being captured by the same spectral pixel flag). As a result of this we may relax the limits on the number of wavelength pixels that is allowed to be flagged as saturated and those identified as outlier – see table below.

config parameter	NO2-v1.3 L1B-v1.0 default	NO2-v2.1 L1B-v1.0 default	NO2-v2.1 L1B-v1.0 special	NO2-v2.1 L1B-v2.0 default	pqf code
The maximum fraction of the radiance spectrum that is allowed to be flagged as saturated before the ground pixel is skipped	0.01	0.12	0.12	0.25	54
The maximum number of outliers that is allowed to be in a spectrum before the ground pixel is skipped	N/A	5	15	10	55

Note here that the NO_2 fit window, 405 – 465 nm, has 304 or 305 wavelength pixels, and that the first config parameter is a fraction of that, while the second one is an absolute number (for historic reasons). Outliers are detected only in the part of the spectrum left over after the saturated wavelength pixels have been removed.

Regular and additional variables in the special processing

With a special option set the NO_2 processor writes more details about the retrieval to the orbit file, data that is not in the regular orbit files, and some of these are useful for this particular case. In addition I have done a kind or post-processing step to add some more data variables that may be useful. All these possible useful variables are in the group PRODUCT/SUPPORT_DATA/DETAILED_RESULT

Regular variables

processing_quality_flags

➔ giving the 'pqf' as mentioned above

nitrogendioxide_slant_column_density

nitrogendioxide_slant_column_density_precision

- ➔ these contain values for every ground pixel where the DOAS retrieval actually started
- ➔ as mention these values is cannot be trusted in case the max number of outliers is exceeded
- ➔ if the SCD error is > 3.3e-5 mol/m2 (=2e15 molec/cm2) the fit results are considered unreliable in the routine processing and the qa_value is set to 0.15 (and the pixel data is not used in the data assimilation system), but vertical column data is provided

cloud_fraction_crb_nitrogendioxide_window

cloud_radiance_fraction_nitrogendioxide_window

- ➔ these say something about the cloudiness of the ground pixel; in the cases we are interested here (lightning above clouds), these are probably not so useful
- ➔ if the latter of these two is > 0.5 the qa_value is set to 0.74

Additional variables

reflectance_nitrogendioxide_window

- ➔ this is the reflectance value at 440 nm we reconstructed from the DOAS polynomial and the Ring correction as input to the routine that calculates the cloud (radiance) fraction in the NO_2 window
- ➔ it might provide additional information in this special case, though I have not investigated how precisely
- ➔ in the problem cases we are interested in here this reflectance may be > 1 and it probably has a unreliable value in case of pqf=55 or 41 (see below)

number_of_outliers

- ➔ the number of outliers detected / removed; if this exceeds the number given in the above table an error is issued

Post-processing variables (which do not have official names)

amf_geo

- ➔ geometric air-mass factor, AMFgeo, i.e. depending only on SZA and VZA

no2_scd_div_amfgeo

- ➔ SCD divided by AMFgeo, i.e. the “geometric column density” (“GCD”)

no2_scd_flag

- ➔ see below

The no2_scd_flag is introduced to make usage of the SCD data easier, by gathering information from a few variables into one flag. This flag can thus be used for filtering, though with care as it probably does not cover all possible situation. Here ‘Delta’ refers to the SCD error (in mol/m2).

As mentioned pqf=55 refers to too many outliers. If that limit is not exceeded there can still be problems with the fit results (even though the DOAS fit is re-done), which

usually will lead to the (cryptic) “generic range error” with pqf=41, though other errors may occur too (my script this not check for those; it only makes sure to skip pqf=36, since that refers to missing cloud data).

value	meaning
-1	no SCD value due to saturation limit exceeded, i.e. pqf = 54
0	SCD with Delta < 3.3e-5 & no error reported
1	SCD with Delta < 3.3e-5 & error reported: pqf=55
2	SCD with Delta < 3.3e-5 & other error reported, e.g. pqf=41
3	SCD with Delta >= 3.3e-5 & no error reported
4	SCD with Delta >= 3.3e-5 & error reported: pqf=55
5	SCD with Delta >= 3.3e-5 & other error reported, e.g. pqf=41
FillValue	no SCD due other error (prior to the DOAS fit)

Only SCD values with flag ‘0’ are in general reliable, though values with flag ‘3’ may still be used with care. But keep here in mind the limit set on the number of allowed outliers for L1B-v1.0 data: we would like to set that at 5 but use for this special processing 15 (first table). Hence it would for these cases be good idea to also take note of the value given in number_of_outliers: if that is above 5, take care, if that is above 10 be even more careful …

SCD values with flags ‘1’ and ‘4’ are highly unreliable as there were too many outliers, so that the DOAS fit is not redone (i.e. the given values are the result of the DOAS fit with spikes in the reflectance!).

SCD values with flag ‘2’ (and ‘5’) might be useful, that is not possible to say beforehand as we do not know where exactly the pqf=41 comes from (this could be related e.g. to reflectance values > 1).