LuftBlick Report 2019009

Fiducial Reference Measurements for Air Quality

TN on PGN products "correct use" guidelines

Version 1, 31th Dec 2019

	Name	Company
prepared by	Alexander Cede	LuftBlick
	Martin Tiefengraber	LuftBlick
	Manuel Gebetsberger	LuftBlick
	Axel Kreuter	LuftBlick

Contents

Document Change Record

Issue	Date	Section	Observations
1	31st Dec 2019	All	First version

Acronyms and Abbreviations

NO_2	Nitrogen dioxide
O_3	Ozone
AMF	Air Mass Factor
ATBD	Algorithm Theoretical Background Document
DQ	Data Quality
DQF	Data Quality Flag
DU	Dobson Units,
	Effective Ground Location
	European Space Agency
	Field of View
	Fiducial Reference Measurements for Air Quality
	Full Width Half Maximum
	Effective Height
	Level 2
	Multi-Axes Differential Optical Absorption Spectroscopy
	Minimum Langley Extrapolation
	Effective ozone temperature
	Pandonia Global Network
	Quality Control
	Root Mean Square
	Technical note
	Zenith angle

1 Introduction

This report is deliverable 7 (D7) of the ESA project "Fiducial Reference Measurements for Air Quality" (FRM4AQ) [1, 2]. It is a technical note (TN) giving guidelines on the correct use of official data products of the Pandonia Global Network (PGN). The document has main sections for each official data product. Key questions answered in this report include:

- What do the data represent?
- Which part of the atmosphere is sampled?
- How should the data should be used?
- What information on uncertainty and quality is given in the data?

In each section there is a table giving the retrieval details such as cross sections, wavelength range used etc. Parameters listed in this table including the word "limits" are threshold used to determine the data quality (DQ). Those entries have in general two values, which represent the limits for this parameter to change the last digit of the Data Quality Flag (DQF) from 0 to 1 and from 1 to 2 respectively (see section 5 for data quality).

Some data products use a soft-calibration technique called Minimum Langley Extrapolation (MLE) to determine the trace gas slant column amount in the reference spectra [6]. The main assumption of this method is that a certain subset of the data used for calibration only contains a background value of the gas (e.g. only stratospheric NO_2), which further is assumed to be constant over the day. More details on the algorithm can be found in the ATBD [3].

The native L2 data are in ascii format. The official PGN data products as of 31 Dec 2019 are:

- NO_2 total column amounts (section 2)
- O_3 total column amounts (section 3)

1.1 Applicable Documents

[1] Fiducial Reference Measurements for Air Quality [Proposal], LuftBlick Proposal 201805DEV, Issue 1, 2018.

[2] Fiducial Reference Measurements for Air Quality [Contract and Statement of Work], ESA Contract No. 4000125841/18/I-NS, 2018.

[3] A. Cede. *Manual for Blick Software Suite Version 12*, 2019. URL `http://pandonia.net/media/documents/BlickSoftwareSuite_Manual_v12.pdf`.

[4] M. Müller, M. Tiefengraber, and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016011: Validation reports, 2016.

1.2 Reference Documents

[5] G Bernhard, RD Evans, GJ Labow, and SJ Oltmans. Bias in dobson total ozone measurements at high latitudes due to approximations in calculations of ozone absorption coefficients and air mass. *Journal of Geophysical Research: Atmospheres*, 110(D10), 2005.

[6] J. Herman, A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan. NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation. *Journal of Geophysical Research (Atmospheres)*, 114: D13307, July 2009. doi: 10.1029/2009JD011848.

[7] Laura M Judd, Jassim A Al-Saadi, Scott J Janz, Matthew G Kowalewski, R Bradley Pierce, James J Szykman, Lukas C Valin, Robert Swap, Alexander Cede, Moritz Mueller, et al. Evaluating the impact of spatial resolution on tropospheric no 2 column comparisons within urban areas using high-resolution airborne data. *Atmospheric Measurement Techniques*, 12(11):6091–6111, 2019.

[8] Richard Meller and Geert K Moortgat. Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 k in the wavelength range 225–375 nm. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 105(D6):7089–7101, 2000.

[9] A. Serdyuchenko, V. Gorshelev, M. Weber, W. Chehade, and J. P. Burrows. High spectral resolution ozone absorption cross-sections - Part 2: Temperature dependence. *Atmospheric Measurement Techniques Discussions*, 6:6613–6643, July 2013. doi: 10.5194/amtd-6-6613-2013.

[10] M. Tiefengraber and A. Cede. More accurate Pandora total ozone columns by improved laboratory calibration and simultaneous retrieval of effective ozone temperature. Quadrennial Ozone Symposium 2016, Edinburgh, United Kingdom, 2016.

[11] A. C. Vandaele, C. Hermans, P. C. Simon, M. Carleer, R. Colin, S. Fally, M. F. Mérienne, A. Jenouvrier, and B. Coquart. Measurements of the NO_2 absorption cross-section from 42,000 cm^{-1} to 10,000 cm^{-1} (238-1000 nm)

at 220 K and 294 K. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 59:171–184, May 1998. doi: 10.1016/S0022-4073(97)00168-4.

[12] Ann Carine Vandaele, Paul C Simon, Jean Michel Guilmot, Michel Carleer, and Réginald Colin. So2 absorption cross section measurement in the uv using a fourier transform spectrometer. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 99(D12):25599–25605, 1994.

[13] Xiaoyi Zhao, Vitali Fioletov, Alexander Cede, Jonathan Davies, and Kimberly Strong. Accuracy, precision, and temperature dependence of pandora total ozone measurements estimated from a comparison with the brewer triad in toronto. *Atmospheric Measurement Techniques*, 9(12):5747–5761, 2016.

2 NO_2 total column amounts

PGN data product "NO_2 total column amount" gives the total NO_2 column amount between the surface and the top of the atmosphere. It is measured in direct sun observation mode (section 4.1). The following retrieval codes have been used for this data product:

- nvs0: official until 4 Nov 2019, now disused
- nvs1: official since 5 Nov 2019

The retrieval setting for these codes are given in table 1. Note that they only differ in the thresholds for the DQ. The NO_2 column amounts are expressed in DU, with 1 DU=2.687e20 NO_2 molecules per square meter. The DQF given in the L2 data can be used for filtering (section 5). The precision of the NO_2 column amounts is <0.005 DU for DQF 0 data. The accuracy is estimated to 0.05 DU for DQF 0 data at the 1-sigma level [6]. The user should especially consider the following points when using PGN NO_2 column amounts:

- Since NO_2 is retrieved above 400 nm, spectral and spatial stray light do not play a major role in the data quality. Therefore the thresholds for the allowed AMF limits are rather large, 7 and 14 respectively (table 1).
- NO_2 is extremely heterogeneous in time and space. Therefore parameters such as the sampled air mass and the effective location of the measurements should be considered when the data are compared to other measurements, e.g. surface in-situ data. MAX-DOAS data or satellite data (section 4). E.g. at solar ZA = 60° and an (true) effective height of the NO_2 of 4.2 km, the effective location is shifted for a distance of 7.3 km in the direction of the sun (equation 4).
- The different viewing geometry between the PGN direct sun observations and satellite data is the main driver for possible differences between these products. The magnitude of this difference typically grows with the size of the satellite footprint and the pollution level at the location. An excellent description of this can be found in *Judd et al.* [7].

- The effective height and temperature of the NO_2 profile are set to 7.2 km and 254.4 K respectively in the retrieval algorithm (table 1). These values assume an NO_2 profile with about 40% of the column in the stratosphere and 60% in the troposphere. However in real situations the tropospheric fraction can very from 0% to >90% which would alter theses estimations significantly.
- The accuracy of the data depends mostly on the quality of the calibration, i.e. the MLE. This method works better at stations, which are not always polluted, and consequently works worse at continuously polluted places such as large cities. This should also be considered when analyzing and interpreting the data.

Table 1: NO_2 total column retrieval settings

Parameter	Setting
r-code	nvs0 until 4 Nov 2019; nvs1 since 5 Nov 2019
Observation mode	Direct sun
Algorythm type	Real time direct sun algorithm, version 1
Filter used	None
Assumed effective layer height	7.2 km
Reference	Synthetic reference spectrum calibrated with MLE
Wavelength window	400 to 440 nm
Order of smoothing polynomial	4
Order of offset polynomial	0
Order of wavelength change polynomial	1
Fitted gases	NO_2 at 254.5 K from *Vandaele et al.* [11], O_3 at 225.0 K from *Serdyuchenko et al.* [9]
Fitted gas temperatures	None
Ring	Not fitted
Molecular scattering	Subtracted before fitting
Noise	Included in fitting
L1 correction steps applied	All but wavelength correction
Uncertainty limits	nvs0: 0.03 DU, 0.05 DU; nvs1: 0.0048 DU, 0.0133 DU
AMF limits	7, 14
wrms limits	nvs0: 0.001, 0.005; nvs1: 0.00093, 0.00195
Wavelength shift limits	0.2 nm, 0.5 nm
L1 quality limits	Standard

3 O_3 total column amounts

PGN data product "O_3 total column amount" gives the total O_3 column amount between the surface and the top of the atmosphere. It is measured in direct sun observation mode (section 4.1). The following retrieval codes have been used for this data product:

- out0: official up to present

The retrieval setting for this code are given in table 2. The O_3 column amounts are expressed in DU, with 1 DU=2.687e20 NO_2 molecules per square meter. The DQF given in the L2 data can be used for filtering (section 5). The precision of the O_3 column amount is <1 DU for DQF 0 data. The accuracy is estimated to 12 DU for DQF 0 data at the 1-sigma level [13]. The main reasons for this limited accuracy are listed here. They are discussed in more detail e.g. in *Müller et al.* [4] and *Tiefengraber and Cede* [10].

- The retrieval is purely based on laboratory calibration. Any non-smooth structure in the instrument's spectral sensitivity inside the O3 wavelength range (310 to 330 nm) can cause a systematic bias in the data of a few DU. This is especially true for units, which have not undergone absolute calibration in the laboratory.
- Elaborate stray light calibration is not applied to most instruments, which gives a negative bias in total O3 starting at SZA=70 deg and sometimes even at smaller values. Therefore the AMF limits in the retrieval are set to 5 and 7 respectively.
- The algorithm assumes a fixed effective O_3 temperature (O3temp) of 225 K, which commonly leads to underestimation (overestimation) of the total O_3 column in summer (winter). Therefore a seasonal difference between PGN data and other measurements, which take O3Temp into account, is seen.

Total O_3 columns are driven by the stratospheric O_3 amount. Therefore the different viewing geometries between measurement techniques (section 4) play a relatively small role and any differences are mostly caused by the aspects listed above.

Table 2: O_3 total column retrieval settings

Parameter	Setting
r-code	out0
Observation mode	Direct sun
Algorythm type	Real time direct sun algorithm, version 1
Filter used	U340
Assumed effective layer height	20.4 km
Reference	Theoretical reference spectrum
Wavelength window	310 to 330 nm
Order of smoothing polynomial	4
Order of offset polynomial	0
Order of wavelength change polynomial	1
Fitted gases	NO_2 at 254.5 K from *Vandaele et al.* [11], O_3 at 225.0 K from *Serdyuchenko et al.* [9], SO_2 at 259.2 K from *Vandaele et al.* [12]; HCHO at 256.9 K from *Meller and Moortgat* [8]
Fitted gas temperatures	None
Ring	Not fitted
Molecular scattering	Subtracted before fitting
Noise	Included in fitting
L1 correction steps applied	All but wavelength correction
Uncertainty limits	1 DU, 5 DU
AMF limits	5, 7
wrms limits	0.01, 0.02
Wavelength shift limits	0.5 nm, 1.0 nm
L1 quality limits	Standard

4 Viewing geometries

This section gives a short explanation of the air mass sampled in the different observation modes. It uses the "effective height" (h_{EFF}) of a trace gas, which is given by:

$$h_{EFFj} = \frac{\int_{SURF}^{ToA} n_j(h) \cdot h \cdot dh}{\int_{SURF}^{ToA} n_j(h) \cdot dh} \tag{1}$$

The integral runs from the surface SURF to the top of the atmosphere ToA along the vertical path h. $n_j(h)$ is the particle density of the trace gas at height h.

Another expression introduced is the "Effective Ground Location" (EGL). The EGL is defined as that location on the ground, for which the measurements are "most representative", which is in general NOT the location of the ground-based instrument or the center of the satellite's footprint projected to the ground.

4.1 Direct sun or moon

The viewing geometry of the direct sun or moon observation mode is outlined in figure 1. The sampled air mass is a circular cone with its apex at the entrance of the instrument and extending into the direction of the sun or moon. This means that for direct sun observations, the measurements sample air towards East in the morning, South around noon, and West in the afternoon (for Northern Hemispheric locations).

For PGN direct sun observations, the opening angle of the cone is about $\alpha = 2.6°$ = 0.045 rad, which is the FWHM angle of the Pandora FOV in direct sun mode. For direct moon observations, the FWHM of the FOV is $\alpha = 1.5°$ = 0.026 rad. However the vast majority of the light sampled in direct observation mode comes from an angle of 0.5° = 0.009 rad, since this is the angular size of both the sun and the moon. The direct light comes from inside these 0.5°, the forward scattered diffuse light is distributed over the entire FOV of the instrument. Only at large ZA and high aerosol content is the diffuse fraction significant (a few percent of the total signal) and can possibly introduce systematic errors in the retrieval. Each trace gas molecule inside the cone is equally "counted", i.e. there is no dependence of the vertical profile in the data.

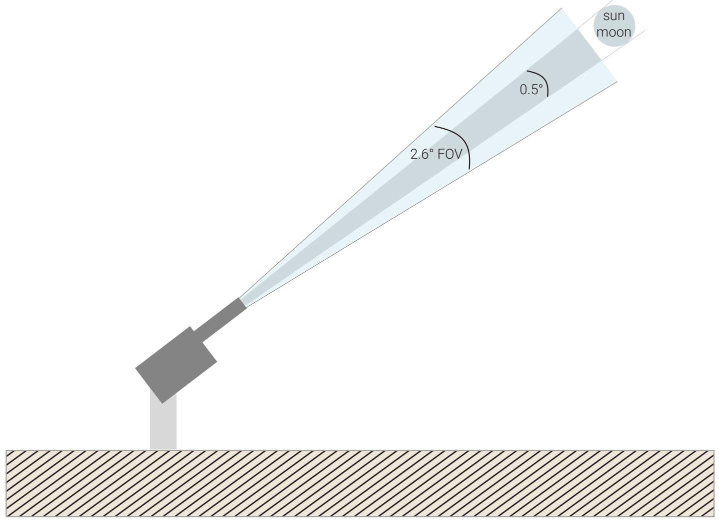

Figure 1: Direct sun observations

Figure 2 shows a simplification for the situation for direct sun observations. R is the distance from the center of the Earth to the measurement location (about 6370 km, refined based on the location's latitude), ZA* is the apparent solar (or lunar) zenith angle (i.e. the geometrical ZA corrected for refraction) and h_{EFF} is the effective height for the trace gas (from equation 1). s is the slant distance between the instrument and the point where the direct beam is at height h_{EFF} in the atmosphere. d is the ground distance between the location of the instrument and location underneath the point where the direct beam is at height h_{EFF} in the atmosphere. ZA' is the "reduced" zenith angle, which can be calculated by equation 2.

$$ZA' = \arcsin\left[\left(\frac{R}{R + h_{EFF}}\right) \cdot \sin(ZA^*)\right] \tag{2}$$

s and d can be calculated with equations 3 and 4 respectively.

$$s = R \cdot \frac{\sin(ZA^* - ZA')}{\sin(ZA')} \tag{3}$$

$$d = R \cdot (ZA^* - ZA') \tag{4}$$

The estimation of the direct AMF in the PGN retrievals, used to convert the measured slant columns into vertical columns, is based on some assumptions, e.g. that the vertical distribution of the trace gas is a delta function at h_{EFF}. It is given by equation 5. For details on this equation see e.g. *Bernhard et al.* [5]).

$$AMF_{DIR}(ZA^*) = \sec(ZA') \tag{5}$$

The accuracy of AMF_{DIR} depends mostly on how well h_{EFF} represents the "truth", but is in general better than 2% for ZA<80°.

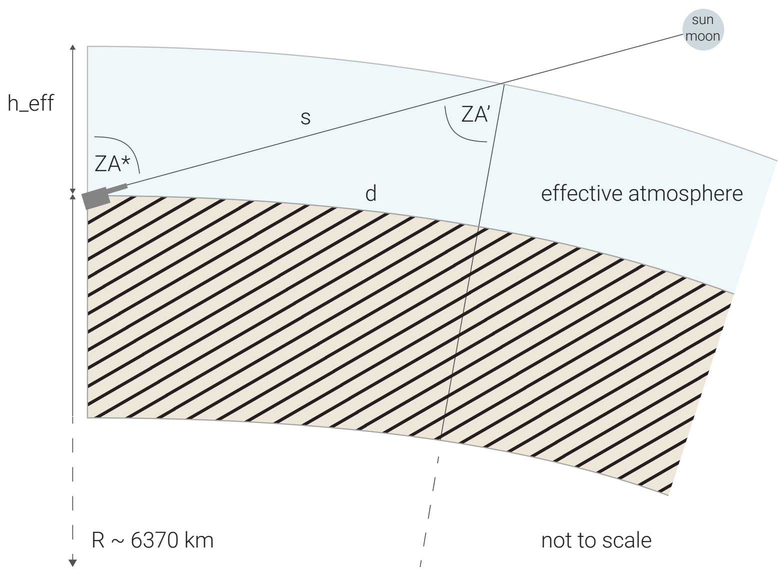

Figure 2: Direct sun geometry

4.2 MAXDOAS

The viewing geometry of MAXDOAS measurements is outlined in figure 3. The FOVs of MAXDOAS instruments is typically <3°. The photons captured by the instrument originate from the sun, enter the atmosphere in the direction of the solar ZA, and then make one or more interactions to be directed into the instrument. Some of them are reflected by the surface, others are scattered in the atmosphere to end up in the instrument's entrance optics. A smaller fraction does even multiple interactions, e.g. reflection plus scattering, or two scatter processes, etc. Therefore the sampled air mass is a rather large "region" above the instrument. The effective location depends on where the instrument is exactly pointing and therefore changes for each of the elevation angles sampled by the instrument. Hence for MAXDOAS observations, the effective location of the measurement is usually shifted relative to the instrument's location, namely into the direction of the pointing azimuth.

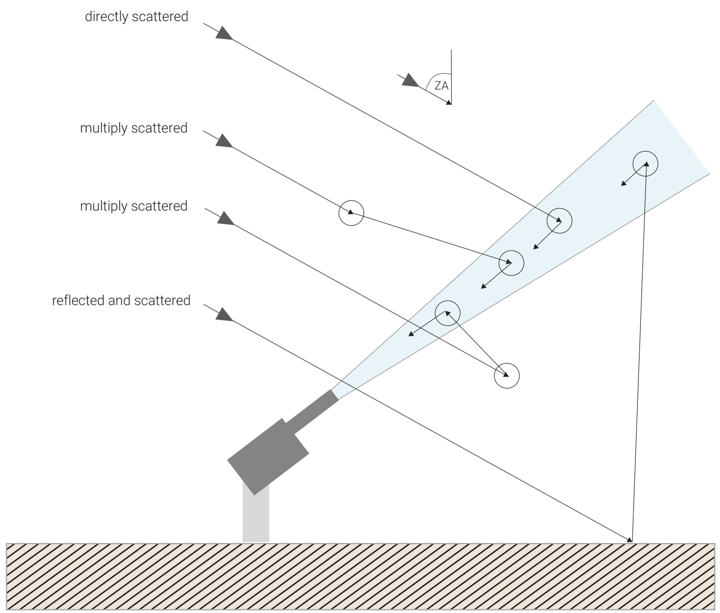

Figure 3: MAXDOAS observations

4.3 Satellite nadir view

The viewing geometry of near nadir observations from satellite (or aircraft) is outlined in figure 4. The FOV of the instrument projected to the ground gives the so-called footprint of the satellite instrument. In current satellites, the diameters of the footprint vary from about 5 km to more than 100 km. As for MAXDOAS measurements, the photons captured by the satellite instrument originate from the sun, enter the atmosphere in the direction of the solar ZA, and then make one or more interactions to be directed into the satellite instrument. The sampled air mass is a rather large "region" underneath the satellite and in the direction of the Sun. Hence for nadir observations, the effective locations is also shifted relative to the center of the footprint, namely towards East in the morning, South around noon, and West in the afternoon (for Northern Hemispheric locations). However the "displacement" into the direction of the sun is not as large as for direct observations.

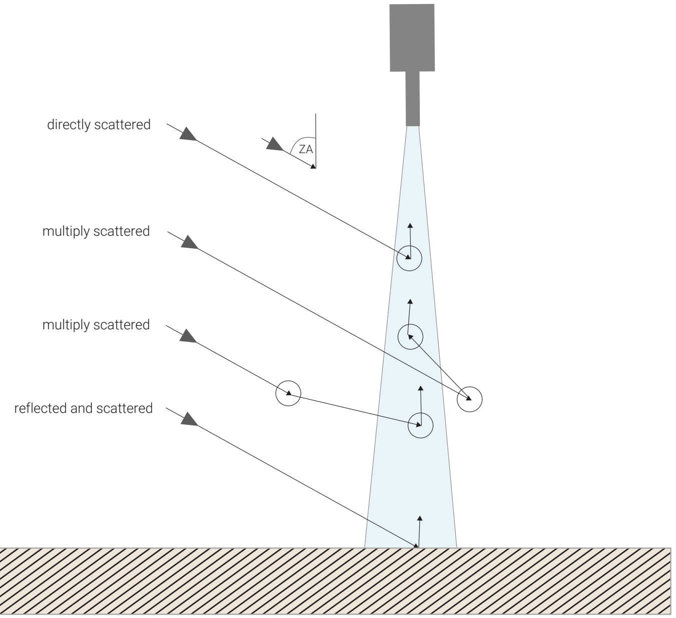

Figure 4: Nadir observations

5 Data quality information

The PGN L2 data contain several columns with information on the data quality for trace gas GAS. Columns labeled "Uncertainty of ..." give the noise in the data, which is propagated from the raw data to the L2 data. not that all uncertainties are given at the 1-sigma level.

- umeas: Column "Uncertainty of GAS total vertical column amount based on measured uncertainty". This is a combination of instrumental noise and atmospheric variation over the measurements duration, based on the measured standard error of the raw data.
- uinstr: Column "Uncertainty of GAS total vertical column amount based on instrumental uncertainty". This is the instrumental noise over the measurements duration, based on the calculated instrumental noise of the raw data, which is known from the calibration.
- urms: Column "Uncertainty of GAS total vertical column amount based on rms". This is umeas multiplied by the rms of the spectral fitting.

Another group of output columns indicating the data quality are the columns with "rms", i.e. the root mean square (rms) of the residuals in the spectral fitting procedure.

- wrms: Column "Normalized rms of spectral fitting residuals weighted with measured uncertainty". This is the obtained rms of the spectral fitting.
- wrmsm: Column "Expected normalized rms of weighted spectral fitting residuals based on measured uncertainty". This is a calculation of what rms of the spectral fitting should have been obtained, if it is purely determined by the measured (instrumental and atmospheric) noise in the data.
- wrmsi: Column "Expected normalized rms of weighted spectral fitting residuals based on instrumental uncertainty". This is a calculation of what rms of the spectral fitting should have been obtained, if it is purely determined by the instrumental noise in the data.

A combination of data quality indicators such as the above listed uncertainties and rms is used to product the DQF. The meaning of the difference DQFs is:

- DQF 0, assured high data quality. Those data can be used with high confidence.
- DQF 1, assured medium data quality. Depending on the application, the user should decide whether to use these data. Note that the reduced quality can origin from instrumental sources (e.g. too large wavelength shift) or atmospheric sources (e.g. clouds increasing the uncertainty in direct sun measurements).
- DQF 2, assured low data quality. For most purposes, the user should not use these data. As for DQF 1, the low quality can origin from instrumental or atmospheric sources.
- DQF 10 (11, 12), not-assured high (medium, low) quality. This means that all indicators suggest a high (medium, low) data quality, but no quality control (QC) has been performed yet. Unless the QC reveals some issues, this will then become DQF 0 (1, 2) after the QC is done.
- DQF 20 (21, 22), unusable high (medium, low) quality. This means that all indicators suggest a high data quality, but it is known that the data cannot be used. This can be due to a lack of calibration or due to known issues with he algorithm for this output parameter.