QA4ECV

Quality Assurance for Essential Climate Variables

Project Number 607405

Deliverable:	D2.3 Review of Pixel Level Quality Flags Provided with ECV Data Products
Responsible Partner:	National Physical Laboratory
Delivery date:	August 2017

NPL National Physical Laboratory

Review of Product and Pixel Level Quality Indicators Provided with ECV Data Products

(Version 1.0)

Tracy Scanlon [1]
Joanne Nightingale [1]
Steven Compernolle [2]
Jean-Christopher Lambert [2]

[1] National Physical Laboratory, UK
[2] Belgian Institute for Space Aeronomy, Belgium

CONTENTS

TERMS & ACRONYMS

ATBD	Algorithm Theoretical Basis Document
BHR	Bi-hemispherical reflectance (also called white sky albedo)
BRDF	Bi-directional Reflectance Distribution Function
C3S	Copernicus Climate Change Service
DHR	Directional Hemispherical Reflectance (also called black sky albedo)
ECV	Essential Climate Variable
EM	Electromagnetic spectrum
EO	Earth Observation
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
L1B	Level 1B
LAI	Leaf Area Index
MFG	Meteosat First Generation
MODIS	Moderate Resolution Imaging Spectroradiometer
MSSL	Mullard Space Science Laboratory
NASA	National Aeronautics and Space Administration
NIR	Near Infra-Red
NPL	National Physical Laboratory
PUG	Product User Guide
PUM	Product User Manual
QA4ECV	Quality Assurance for Essential Climate Variables (EU FP7 Project)
QA4EO	Quality Assurance for Earth Observation
QF	Quality Flag
QI	Quality Indicator
RMSE	Root Mean Squared Error
RT	Radiative Transfer
SI	International System of Units
UCL	University College London

EXECUTIVE SUMMARY

In support of the European Union's Earth Observation Programme's Copernicus Climate Change Service (C3S), the Quality Assurance for Essential Climate Variables (QA4ECV) project aims to fulfil a current gap in the delivery of satellite derived climate data products. The project will prototype a system for the implementation and evaluation of QA measures for 6 satellite-derived ECV and ECV precursor datasets, thus providing confidence in their application for climate monitoring studies and climate change assessments.

The purpose of developing and implementing a QA4ECV system is two-fold:

1. To provide **ECV data product producers / science teams** with the necessary resources (internationally accepted tools, standards, methodologies) to develop products with embedded QA information that is presented in a clear and common format throughout the Earth Observation (EO) community, and,
2. To provide **ECV data users (scientists – policy-makers)** with robust QA information as a means to quantitatively assess uncertainty and fitness-for-purpose of the data and derived products.

This report provides an outline of the documentary framework for the QA service which is to be provided under the QA4ECV project. The aim of the documentary framework is to provide a robust basis for the QA4ECV QA service.

The framework will consist of a set of documentation including procedures, good practice guidance and training. These documents will be underpinned by tool specifications and appropriate scientific studies to ensure that the framework is robust.

The current document undertakes a review of the pixel-level Quality Indicators (QIs), i.e. Quality Flags (QFs) currently utilised in developed ECV products for both the land and atmospheric domains. The review is undertaken with the intent of representing a cross section of products for all of the ECVs / ECV precursors which are being produced within the QA4ECV project; this includes products from the US and Europe.

The output of the review is utilised in the design of the Quality System developed under the QA4ECV project. In addition, some suggestions are made on requirements for QFs associated with ECV data products.

1 Introduction

1.1 Quality Assurance for Essential Climate Variables

Climate change mitigation and adaptation has risen to the top of the agenda for many governments and international organisations. This has led to the establishment of projects and programmes dedicated to the development of long-term global records of Essential Climate Variables (ECVs) using space-borne assets.

In support of the European Union's Earth Observation Programme's Copernicus Climate Change Service (C3S), the Quality Assurance for Essential Climate Variables (QA4ECV) (http://www.qa4ecv.eu/) project aims to fulfil a current gap in the delivery of satellite derived climate data products. The project will prototype a system for the implementation and evaluation of QA measures for 6 satellite-derived ECV and ECV precursor datasets, thus providing confidence in their application for climate monitoring studies and climate change assessments.

The purpose of developing and implementing a QA4ECV system is two-fold:

1. To provide **ECV data product producers / science teams** with the necessary resources (internationally accepted tools, standards, methodologies) to develop products with embedded QA information that is presented in a clear and common format throughout the Earth Observation (EO) community, and,
2. To provide **ECV data users (scientists – policy-makers)** with robust QA information as a means to quantitatively assess uncertainty and fitness-for-purpose of the data and derived products.

Provision of such QA information will demonstrate traceability of products and simplify comparisons, including round-robin selection, between the same ECV produced by independent science teams. It will also provide data users with evidence-based confidence in the products and enable judgement on the fitness-for-purpose of various ECV Climate Data Records (CDRs) for their specific applications.

1.2 QA Service and System

One of the main aims of the QA system being developed under QA4ECV is to bridge the gap between data users and data producers, i.e. allow the transfer of information between the two in easy-to-use and consistent formats (as far as practicable). Taking this into account, the QA system is split into two main "paths":

- Data providers will follow a series of pages through the QA system to provide evidence relating to their ECV data product. This evidence will be included in a central repository.
- Data users will be able to search and download quality reports about a range of ECV data products from the central repository for comparison.

The ***QA service*** includes:

- The ***QA system*** is a physical system implemented is an interactive web-service through which data products will be assessed, and,
- The ***documentary framework*** is a series of documentation including procedures, good practice guidance and training which support the QA system (and are linked to

Details of the Copernicus climate change service are available at: https://climate.copernicus.eu/

Note that several other EU funded projects including EUPORIAS and Core Climax have identified this as a key factor which could improve the overall use of climate data sets. See [28].

throughout the QA system).

The structure of the ***QA service*** is summarised in Figure 1 (in green) within the context of ECV production (yellow) and dissemination (blue and purple). The ***documentary framework*** which is the associated guidance etc. is within the "Tools & Guidance to establish and evaluate QA" box (in green).

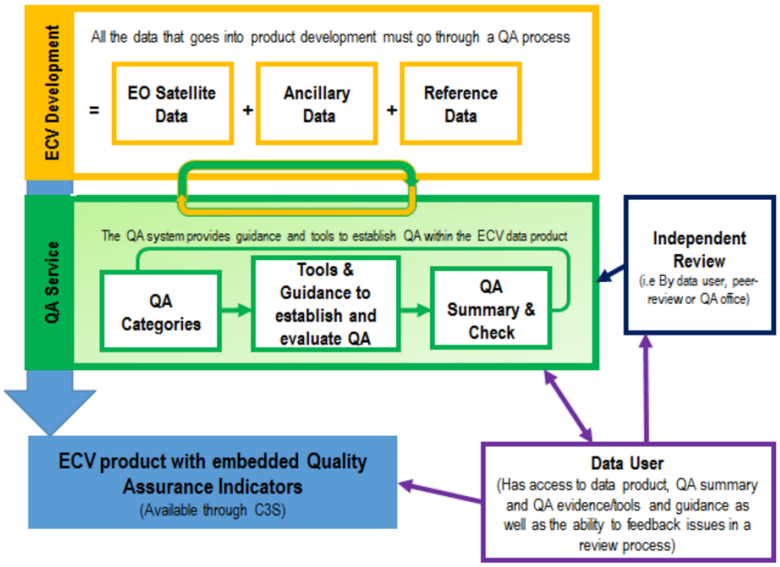

Figure 1: Diagram of system.

1.3 Objective

The main objective of this report is to allow an understanding of the current practices implemented in the land and atmosphere community with regards to how quality information is disseminated in products through pixel-level Quality Flags (QFs) (see Section 1.4 for definitions).

This information is utilised in determining the design of the QA system (see Section 1.2), i.e. what information is requested from the data providers regarding QFs.

1.4 Quality Flags Definitions

QA information (also termed "Quality Indicators") is central in the Quality Assurance Framework for Earth Observation (QA4EO) established to facilitate harmonisation and interoperability of EO data, derived products and the associated tools and assets. The QA4EO core principle reads as follows [1]:

> *The fundamental principle of the Quality Assurance Framework for Earth Observation (QA4EO) is that all Earth Observation (EO) data and derived products have associated with them a documented and fully traceable Quality Indicator (QI).*

Quality Flags (QFs) are a subset of Quality Indicators (QIs). QFs are specific data provided within the product (either as a data layer or a field within the main product). A QF generally provides basic information about the performance and quality attributes of the data set.

Note this diagram is also provided at www.qa4ecv.eu.

1.5 Method

The understanding of the different current practices is derived from consideration of several different land (see Section 2) and atmospheric (see Section 3) ECV data products. The specific products are chosen because they allow the practices of different organisations to be captured as well as allowing consideration of both ECV CDRs and short term monitoring data sets.

For each product, the QFs provided with the product are listed and categorised into one of the following:

1. Uncertainty estimate,
2. Designated quality statement, or,
3. Ancillary data and other information which indicates quality

A summary of the available quality flags and the conventions employed for the products associated with each ECV is provided in the following sections.

2 Land Products

2.1 Albedo

2.1.1 Products Studied

The albedo products listed in Table 1 have been studied within this document. The QFs provided with albedo products are listed in Annex A, Tables A.1 to A.4.

Each of the albedo products considered provide either “white” or “black” sky albedo (or both). These parameters are defined as follows (from [2]):

> **White Sky Albedo**: BHR (bi-hemispherical reflectance) is the ratio of the radiant flux reflected from a unit surface area into the whole hemisphere to the incident radiant flux of hemispherical angular extent.
>
> **Black Sky Albedo**: DHR (directional hemispherical reflectance) pure direct illumination.

Table 1: Albedo Products Studied

Product Name / Organisation	Description	QF table
GlobAlbedo MSSL / UCL	Provides albedo (DHR and BHR) from 1998 to 2011 at 1-day, 8-day or monthly intervals. The product is derived from European satellites using MODIS (Moderate Resolution Imaging Spectroradiometer) as a prior to constrain the solution. The product is available at various spatial scales: 500 m, 1 km, 5 km, 25 km, 0.05° or 0.5° pixels and covers the visible, short wave and near infra-red (NIR) regions of the electromagnetic (EM) spectrum. The algorithm is described in the ATBD [3] and details of the QFs are provided in the PUG [4].	A.1
Copernicus Land Surface Albedo VITO / University of Hamburg	Provides DHR and BHR from SPOT data retrieved through inversion of a semi-empirical linear kernel model which favours the most recent observations from a 30-day composite. The product is provided for the visible ([0.4µm-0.7µm]), near-infrared ([0.7µm-4µm]), and total shortwave ([0.3µm-4µm]) bands [5]. Further described at [6].	A.2
Meteosat Surface Albedo EUMETSAT	Provides albedo (DHR and BHR) derived from Meteosat First Generation (MFG) satellites for each day. The algorithm is described in the ATBD [7] and details of the QFs are provided in the PUG [8].	A.3
MODIS BRDF / Albedo (MOD43) (Version 6.0) NASA	Both DHR and BHR products are provided at threes spatial scales (500 m, 1000m, and 5600 m) on a daily basis. Several MODIS albedo products are available; a full list can be found at the MODIS products table [9]. The products considered in this document are those from MODIS collection 6.0 (there are 76 products). The algorithm is described within the ATBD [10] and the QFs and their meanings are provided on the webpage for each of the individual products (listed at [9]).	A.4

Details available at: http://www.globalbedo.org/
Details available at: http://land.copernicus.eu/global/products/sa

2.1.2 Summary of Quality Flags

A summary of the QFs included in each albedo product is provided at Table 2.

Table 2: Albedo Quality Flags Summary Table

QF Field	QF Description	GlobAlbedo	Copernicus	Meteosat	MODIS
Uncertainty	A measure of uncertainty associated with the product.	Y	Y	Y	Y
Quality Statement	Statements about the quality of the dataset provided.	N	Y	Y	Y
Ancillary Datasets	Various ancillary datasets including number of observations, snow cover and land cover.	Y	Y	Y	Y

Uncertainty Estimate

All of the albedo products studied include an indicator of the uncertainty associated with the product. The uncertainty is referred to either the “sigma”, the “error” or the “uncertainty” depending on the product. For each of the products, the uncertainty refers to slightly different quantities, which may be related to the inversion method or how many observations have been utilised. In all cases, it is unclear to the reader of the documentation how the uncertainty has been derived and it is clear that the uncertainty contributors from other parts of the processing chains have not been considered. In the case of the Meteosat product, various uncertainty contributor values are provided including radiometric noise.

Quality Statement

Some form of quality statement is provided in all products. These statements take several forms including simple, use / do not use statements to detailed information on the probability of the correctness of the data (allowing users to apply thresholds on the use of the data).

In many cases, values of 0 and 1 or 0 to x are assigned to various different qualitative descriptions of quality. In all but one case, “0” is utilised to mean “good, ok, best solution”, with higher values representing worse data.

The other fields directly related to quality which are provided include “goodness-of-fit”, number of observations, details of the success of the inversion, etc.

Ancillary Data

The albedo products include many different ancillary datasets, which may have some effect on the quality of the retrieval. These include land cover and snow cover in most cases. Whilst these ancillary data are very important to the derivation of the data sets, their definition, presentation and presence are dependent on the product being retrieved.

2.2 LAI and FAPAR

2.2.1 Products Studied

The LAI (Leaf Area Index) and FAPAR (Fraction of Absorbed Photosynthetically Active Radiation) products listed in Table 3 have been studied within this document. The QFs

provided with LAI and FAPAR products are listed in Annex A, Tables A.5 to A.8.

LAI and FAPAR products are treated together in this document due to the commonality between the methods used to generate the products, the fact that many products include both parameters and the QFs are usually similar to one another. For clarity, LAI and FAPAR are defined as follows (from [11]):

> **LAI** is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as one-half the total needle surface area per unit ground area in coniferous canopies.
>
> **FPAR** is defined as the fraction of incident photosynthetically active radiation (400-700nm) absorbed by the green elements of a vegetation canopy.

Table 3: LAI Products Studied

Product Name / Organisation	Description	QF table
Copernicus LAI INRA, CREAF, VITO	The Copernicus LAI product has been derived from PROBA-V data and covers the period May 2014 – present for Europe. It is a global product provided at a spatial resolution of 1 / 3 km. The product is described within the ATBD [12] and information on the QFs associated with the product are given in the PUG [13].	A.7
MODIS FPAR / LAI (MOD15) NASA	The MODIS combined LAI / FPAR product is an 8-day composite dataset at a spatial scale of 500 m, 1000 m or 5600 m. The algorithm chooses the best available pixel with the 8-day period. The algorithm is described within the ATBD [14] and the QFs and their meanings are provided on the webpage for each of the individual products (listed at [9]).	A.8
QA4ECV TIP LAI / FAPAR FastOpt	LAI and FAPAR and the associated uncertainties are retrieved from LUTs of inversions of the two-stream model. The model relates BHRs to various canopy parameters and fluxes. The product is derived from the QA4ECV albedo product and therefore has the same temporal and spatial resolutions. The algorithm and QFs are described within the ATBD [15].	A.9
Copernicus FAPAR INRA, CREAF, VITO	See description of “Copernicus LAI” above.	See A.7
QA4ECV FAPAR JRC	The product provides instantaneous, green FAPAR. Theoretical FAPAR values are computed using the closure of the energy balance inside the plant canopy in the spectral range 400 to 700 nm. The output of the algorithm is computing using a ratio of polynomial for which a series of coefficients is pre-computed and depend on NOAA platforms [16].	A.11

(1) Note that both the LAI3g and FAPAR3g products (created by the University of Boston) were investigated as part of this study (described at [17] and [18]), however, these contain no QFs and therefore were not considered further.

2.2.2 Summary of Quality Flags

A summary of the QFs included in each LAI / FAPAR product is provided at Table 2.

Table 4: LAI / FAPAR QFs – Summary Table

QF Field	QF Description	Copernicus FAPAR / LAI	MODIS FPAR / LAI	QA4ECV TIP LAI / FAPAR	QA4ECV FAPAR
Uncertainty	A measure of uncertainty associated with the product.	N	Y	Y	Y
Quality Statement	Statements about the quality of the dataset provided.	Y	Y	Y	Y
Ancillary Datasets	Various ancillary datasets including number of observations, snow cover and land cover.	Y	Y	N	Y

Uncertainty Estimate

Uncertainty estimates are provided in each of the products except the Copernicus one. In some cases, such as the QA4ECV FAPAR product, uncertainty are contributors added in quadrature; for the other two cases, the uncertainties are taken as the distribution of values around the solution.

Quality Statement

Quality statements are provided in all products. In all cases, 0 as a value of an integer field corresponds to a qualitative description of “good” data, with other values relating to various descriptions. The variation to this rule is the LAI / FAPAR product generated from TIP (QA4ECV product). In this case, the various encoded data is provided, allowing a range of other information to be communicated to the users.

Ancillary Data

Ancillary datasets are only provided in the Copernicus product where the number of observations utilised and the model fit are provided.

3 Atmospheric Pre-Cursor Products

3.1 Products Studied

The LAI and FAPAR products listed in Table 3 have been studied within this document. The QFs provided with albedo products are listed in Annex C, Tables B.1 to B.4.

Table 5: Atmospheric Products Studied

Product Name / Organisation	Description	QF table
MOPITT CO v5 product NCAR	Profile and column provided. More details are provided in the product user guide [19].	B.1
IASI FORLI-CO product LATMOS, ULB	Column provided. More information about the retrieval principles is provided in [20]. The current ASCII product files contain CO total column relative uncertainty, Degrees of Freedom of Signal (for the vertical profile), fit residual RMS, fit residual bias, and "super" quality flag, indicating use-only if equal to zero. There will be a new netCDF-CF CCI format for this product in the future therefore this product has not been studied in detail here.	N/A
DOMINO v2 NO2 product KNMI	Total, stratospheric and tropospheric column provided. The total column stored is however not recommended: use sum of stratospheric and tropospheric instead. For details, see	B.2
QA4ECV NO2 product KNMI	Total, stratospheric and tropospheric column provided.	B.3
QA4ECV HCHO product (1) BIRA-IASB	This is the product being developed under the scope of the QA4ECV project. Column provided.	B.4

(1) Note: there is a TROPOMI HCHO product for which the data is not yet available, however the ATBD and PUG is already available (http://www.tropomi.eu/documents). However, due to similarities between this and the QA4ECV HCHO product (both designed by the same groups), this product is not studied here.

Chemical species in atmospheric satellite products are typically quantified as vertical profiles (in volume mixing ratio, or number density, or partial column) or as vertical columns (in number of molecules per surface unit).

The formaldehyde (HCHO) and nitrogen dioxide (NO2) are discussed here together because of commonalities between the products and their derivation. In both cases the differential optical absorption spectroscopy (DOAS) method is applied to derive a column product (molec/cm^2) starting from (ir)radiances in the UV-Vis domain. The DOAS method, the use of air mass factors and averaging kernels, and the associated uncertainty analysis is described by [21], [22] and [23]. Shortly, the derivation of the HCHO and NO2 products can be summarised as follows:

- In a first step, a spectral fit is performed to reference spectra, from which a slant column is derived.
- Specifically for NO2, the stratospheric contribution to this slant column is estimated using data assimilation, and subsequently subtracted to obtain a tropospheric slant column.

- Specifically for the weak absorber HCHO, a remote background correction is applied, based on a chemical transport model.
- In a final step, an air mass factor, obtained from radiative transfer modelling (using input data such as viewing geometry, cloud properties and surface albedo), is applied to derive a vertical column from the slant column.

Both the HCHO and the NO2 products are provided with column averaging kernels. These are vector quantities associated with the retrieved columns. They describe the sensitivity of the retrieved column to the true atmospheric profile.

The carbon monoxide (CO) product is derived from radiance measurements in the infrared, using the Maximum A Posteriori (MAP) (also called Optimal Estimation) method, of which the generalities are described in [24]. The retrieved quantity (column or profile) depends not only on measurement information but also on an 'a priori' profile and an associated a priori covariance matrix.

3.2 Summary of Quality Flags

Uncertainty Estimate

Uncertainty estimates areprovided with all products. In some, (QA4ECV products and DOMINO), these are explained as the error assocaited with the retrieval. In the CO product, there are various elements of the uncertainty provided including details of the covarianace.

Quality Statement

For the QA4ECV products, a simple indiactor is provided with each dataset stating whether or not th eretreival was sucessgul (0 being used to indiactate this). For the CO product, a quality indicator can be calculated ysing information provided and for the DOMNIO (NO2) product, a quality indicator is provided stating whether or not the model or actual observations are used.

Ancillary Data

Ancillary datasets are provided in each of the data products. These are either land surface, albedo or snow / ice masks.

4 Summary

The community utilise many different conventions for communicating quality information (as is demonstrated in this document) and the QA4ECV project does not aim to provide recommendations on what should / should not be utilised. However, from the above review, it is clear that some commonly used conventions are very useful and it may be prudent to consider these for use. These recommendations will help to shape the QA system being developed for QA4ECV.

Uncertainty

A measure of uncertainty has been provided with the majority of the products. However, in many cases the description of what the uncertainty refers to is not clear in the documentation. In addition, it appears that for the same ECV (particularly in the case of albedo), the uncertainty will mean different things for products and therefore cannot be equally utilised in applications.

Quality Flags

There appears to be two different approaches to quality statements:

1. Either a go / no go flag which simply states whether or not a data point should be utilised, or,
2. A series of other fields which the data users can utilise according to their own needs.

Both are very useful for communicating the usefulness of the data and should both be made available where possible.

The go / no go flag generally utilises integer values; in most cases studied, “0” is utilised to mean “good data”.

No recommendations can be made on the other fields provided as these are specific to the retrieval methods utilised.

Ancillary Data

The ancillary datasets provided vary depending on the ECV and the product. No recommendations can be made on what might be included, however, it would be useful to have details of how each field might affect the data product clearly stated within the Product User Manual (PUM).

5 References

[1] QA4EO, “QA4EO: Principles,” 2010.

[2] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik, “Reflectance quantities in optical remote sensing—definitions and case studies,” *Remote Sens. Environ.*, vol. 103, no. 1, pp. 27–42, Jul. 2006.

[3] P. Lewis, G. López, J. P. Muller, D. Potts, N. Shane, S. Kharbouche, D. Fisher, C. Brockmann, O. Danne, O. Krueger, A. Heckel, P. North, and L. Guanter, “GlobAlbedo Algorithm Theoretical Basis Document,” 2013.

[4] O. Danne, M.Zuehlke, and U. Krämer, “GlobAlbedo Product User Guide,” 2013.

[5] Hamburg Universitat, “Integrated Climate Data Center - ICDC: Land Surface Albedo (SPOT),” 2015. [Online]. Available: https://icdc.zmaw.de/1/daten/land/spot-landalbedo.html#c8515.

[6] “Copernicus Global Land Service: Surface Albedo,” 2016. [Online]. Available: http://land.copernicus.eu/global/products/sa.

[7] EUMETSAT, “Meteosat Surface Albedo Retrieval Algorithm Theoretical Basis Document,” no. Version v2B, 2014.

[8] EUMETSAT, “Meteosat Surface Albedo Product User’s Manual,” no. Version v.2H, 2014.

[9] NASA, “LP DACC: MODIS Products Table.” [Online]. Available: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table.

[10] A. H. Strahler, J. P. Muller, W. Lucht, C. Barker-Schaaf, T. Tsang, P. Lewis, and M. J. Barnsley, “MODIS BRDF / Albedo Product: Algorithm Theoretical Basis Document (MOD43),” 1999.

[11] “MODIS LAI / FPAR Product User’s Guide,” 2012. [Online]. Available: https://lpdaac.usgs.gov/sites/default/files/public/modis/docs/MODIS-LAI-FPAR-User-Guide.pdf.

[12] F. Baret and M. Weiss, “Algorithm Theoretical Basis Document: LAI, FAPAR and FCover,” 2015.

[13] B. Smets, T. Jacobs, and R. Lacaze, “Gio Global Land Component - Lot I ”Operation of the Global Land Component”: Product User Manual - LAI, FAPAR., FCover - Collection 300m,” 2016.

[14] R. B. Myneni, Y. Knyazikhin, Y. Zhang, Y. Tian, Y. Wang, A. Lotsch, J. L. Privette, J. T. Morissette, S. W. Running, R. Nemani, J. Glassy, and P. Votava, “MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis,” 1999.

[15] S. Blessing, “QA4ECV Algorithm Theoretical Basis Document for TIP-FAPAR and TIP-LAI,” 2016.

[16] N. Gobron, “QA4ECV: Algorithm Theoretical Basis Document for JRC FAPAR,” 2016.

[17] Z. Zhu, J. Bi, Y. Pan, S. Ganguly, A. Anav, L. Xu, A. Samanta, S. Piao, R. R. Nemani, and R. B. Myneni, “Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3G) for the period 1981 to 2,” *Remote Sens.*, vol. 5, no. 2, pp. 927–948, 2013.

[18] Z. Zhu, J. Bi, Y. Pan, S. Ganguly, and A. Anav, “Fraction of Photosynthetically Active Radiation (FPAR) 3g,” *Re*, pp. 1–16, 2011.

[19] M. N. Deeter, “MOPITT (Measurements of Pollution in the Troposphere): Product User’s Guide (Version 5.0),” 2011.

[20] D. Hurtmans, P.-F. Coheur, C. Wespes, L. Clarisse, O. Scharf, C. Clerbaux, J. Hadji-Lazaro, M. George, and S. Turquety, “FORLI radiative transfer and retrieval code for IASI,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 113, no. 11, pp. 1391–1408, Jul. 2012.

[21] P. I. Palmer, D. J. Jacob, K. Chance, R. V. Martin, R. J. D. Spurr, T. P. Kurosu, I. Bey, R. Yantosca, A. Fiore, and Q. Li, “Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global

Ozone Monitoring Experiment," *J. Geophys. Res. Atmos.*, vol. 106, no. D13, pp. 14539–14550, Jul. 2001.
[22] K. F. Boersma, H. J. Eskes, and E. J. Brinksma, "Error analysis for tropospheric NO 2 retrieval from space," *J. Geophys. Res.*, vol. 109, no. D4, p. D04311, 2004.
[23] H. J. Eskes and K. F. Boersma, "Averaging kernels for DOAS total-column satellite retrievals," *Atmos. Chem. Phys.*, vol. 3, no. 5, pp. 1285–1291, Sep. 2003.
[24] C. D. Rodgers and C. D., *Inverse Methods for Atmospheric Sounding - Theory and Practice*, vol. 2. World Scientific Publishing Co. Pte. Ltd., 2000.
[25] O. Leonard, S. Pinnock, J. P. Muller, T. Kennedy, P. Lewis, S. Kharbouche, P. North, J. Fisher, R. Preusker, C. Brockmann, U. Krämer, O. Danne, and N. Fomferra, "GlobAlbedo Algorithm Theoretical Basis Document," pp. 1–313, 2013.
[26] C. L. Parkinson and R. Greenstone, "EOS Data Products Handbook," 2000.
[27] Q. L. Kleipool, M. R. Dobber, J. F. de Haan, and P. F. Levelt, "Earth surface reflectance climatology from 3 years of OMI data," *J. Geophys. Res.*, vol. 113, no. D18, p. D18308, Sep. 2008.
[28] T. Scanlon, S. Douglas, J. Nightingale, S. Compernolle, and J.-C. Lambert, "QA4ECV: Linkages between the QA Service of QA4ECV and Other Projects and Initiatives," 2017.

Annexes

A Land Quality Indicators

This annex provides details of the QFs used in the land products studied within this document. The colour coding is as follows:

Uncertainty Estimate	Designated Quality Flag	Ancillary data and other information which indicates quality

A.1 Albedo QFs: GlobAlbedo

Field name	Meaning	Values (if applicable)
<a>_sigma<n>()	This is the uncertainty associated with the albedo parameter.	
Weighted_ number_ of_samples	The number of measurement samples in the inversion period.	
Relative_Entropy	A measure of the difference between the prior and posterior uncertainty PDFs used in the optimal estimation framework for calculating the Bi-directional Reflectance Distribution Function (BRDF) parameters. If there are no measurements, the relative entropy will be 0 (i.e. the prior estimate is used). This is the primary metric for passing information to users on the impact of the assumed prior information on the product. Information from [25].	
Goodness_of_Fit	This field provides information on the quality of the inversion model and is provided per pixel. Information from [25].	
Snow_fraction	This is the fraction of snow associated with the pixel.	
Data_mask	The data mask states whether or not the BRDF / albedo values have come from the GlobAlbedo retrieval or the MODIS prior.	

() <a> refers to the different albedo product parameters, i.e. BHR or DHR.
() <n> refers to data either associated with bands of visible, NIR and broadband wavelength ranges in the corresponding products.

A.2 Albedo QFs: Copernicus Land Surface Albedo

Field name	Meaning	Values (if applicable)
AL-<a>-<n>-ERR	This is the uncertainty associated with the DHR or BHR.	
AL-<a>-QFLAG	Bitwise quality flag coded on 2 bytes. Includes information on several parameters.	*Bit #: parameter* / *Bit = 0* / *Bit = 1* Bit 1: Land/Sea — Land — Sea Bit 2: Snow status — Clear — Snow Bit 3: Suspect — Not suspect — Suspect Bit 4: Aerosol status — Pure — Mixed Bit 5: Aerosol source — Modis — Climatology Bit 6: Input status — OK — Out of range or invalid Bit 7: AL-XX-VI status[1] — OK — Out of range or invalid Bit 8: AL-XX-NI status[1] — OK — Out of range or invalid Bit 9: AL-XX-BB status[1] — OK — Out of range or invalid Bit 10: B2 saturation status — OK — Saturated Bit 11: B0 saturation status — OK — Saturated *Other Values:* 65535 invalid pixels 65533 pixel value higher than max physical value 65534 pixel value lower than minimum physical value
NMOD	Number of measurements used to calculate the albedo, hence reflects the number of valid observations during the synthesis period. The physical values are equal to the digital number within the range [0, 60].	
LMK	Land-sea mask	GLC2000 values used.

(8) GLC2000 legend available here: http://forobs.jrc.ec.europa.eu/products/glc2000/legend.php

A.3 Albedo QFs: Meteosat Surface Albedo

Field name	Meaning	Values (if applicable)
DHR30 error 10D	The DHR error estimated using the information on all available retrieval during a 10-day period.	
Probability	The probability of the solution on the day selected. Pixels with a probability smaller than 80 or 90 % should be disregarded.	
QUALITYFLAG	Flag indicating the quality of the solution.	*Value* — *Meaning* 0 — OK 1 — No valid days in the period 2 — No valid samples in the period 3 — No likely days 4 — Invalid solution index 5 — Dubious solution 6 — Weak solution
NUMSOL	The number of acceptable solutions.	*Value* — *Meaning* 255 — Invalid
NSLOT	Number of input lots before cloud screening.	
NSLOTASM	Number of clear sky input slots.	
ERR_R0	Estimated error of the reflectance.	
BESTDAY	Normalised cost function of the best solution.	
Chi2DCP	Cost function of the cloud screening.	
Chi2ASM	Cost function of the inversion.	
DHRError	Estimated error of the DHR for the best day.	
ERR_K	Estimated error of K.	
ERR_T	Estimated error of T.	
ERR_OPT	Estimated error of OPT.	
ERR_AVG_ERR	Standard deviation of AVGOPT (which is the average value of tau hat during the compositing period.	
RADIOMETRIC NOISE	Mean daily radiometric noise of the best day.	

() First two fields are from the “scientific” dataset, remainder are from the ancillary dataset.

A.4 Albedo QFs: MODIS BRDF / Albedo (MOD43)

Field name	Meaning	Values (if applicable)
MCD43A products: The "A" products are the main products available. These products are provided at the native spatial resolution of the instrument in a sinusoidal grid. There are three main products (1, 3 and 4) with data quality for 1 being provided in 2.		
Several albedo products utilise the mandatory quality band for MODIS products. These albedo products are: *Product Number* — *Product Name* MCD43A1 — MODIS/Terra and Aqua BRDF/Albedo Model Parameters Daily L3 Global 500 m SIN Grid V006 MCD43A3 — MODIS/Terra and Aqua Albedo Daily L3 Global 500 m SIN Grid V006 MCD43A4 — MODIS/Terra and Aqua Nadir BRDF-Adjusted Reflectance Daily L3 Global 500 m SIN Grid V006		
BRDF_Albedo_Band_Mandatory_Quality_Band<n>	Mandatory QA bit. This is a minimum requirement for MODIS products and is aimed at giving the user a quick use / do not use statement. This is a summary of the main quality layer "MCD42A2".	*Value* — *Meaning* 0 — processed, good quality (full BRDF inversions) 1 — processed, see other QA (magnitude BRDF inversions) 255 — Fill Value
The following is the BRDF / Albedo quality product which is used with the main product, MCD43A1. This layer contains all the available information about the product. The data producers advise that users use this quality layer where possible. *Product Number* — *Product Name* MCD43A2 — MODIS/Terra and Aqua BRDF/Albedo Quality Daily L3 Global 500 m SIN Grid V006		
Snow_BRDF_Albedo	Snow-free BRDF/Albedo retrieved	*Value* — *Meaning* 0 — Snow-free Albedo retrieved 1 — Snow Albedo retrieved
BRDF_Albedo_LandWaterType	Land Water Type – This is a map provided as ancillary data and is not derived from the product.	*Value* — *Meaning* 0 — Shallow ocean 1 — Land (Nothing else but land) 2 — Ocean coastlines and lake shorelines 3 — Shallow inland water 4 — Ephemeral water 5 — Deep inland water 6 — Moderate or continental ocean 7 — Deep ocean

() Note: there are no "MCD43B" products.

Field name	Meaning	Values (if applicable)
BRDF_Albedo_Valid Obs_Band<n>	Valid Observation for Bands 1 – 7	*Bit #: parameter* — *Bit = 0* — *Bit = 1* Bit 1: Day 1 quality — Observation not used — Valid clear observation … Bit 16: Day 16 quality — Observation not used — Valid clear observation
BRDF_Albedo_Uncertainty	BRDF Inversion Information	No information is provided on how this was calculated within the documentation.
BRDF_Albedo_Band_Quality_Band<n>	BRDF Inversion Information	*Value* — *Meaning* 0 — best quality, full inversion (WoDs, RMSE majority good) 1 — good quality, full inversion (also including the cases that no clear sky observations over the day of interest or the Solar Zenith Angle is too large even WoDs, RMSE majority good) 2 — Magnitude inversion (numobs >=7) 3 — Magnitude inversion (numobs >=2&<7) 4 — Fill value
MCD43C products: The "C" products are the products at the CMG resolution. There are three main products (1, 2, 3) which each include QFs (varying by product); no independent quality layer is provided. These coarser resolution products are retrieved from the original data rather than a spatial average of the higher resolution product. *Key* — *Product Number* — *Product Name* C1 — MCD43C1 — MODIS/Terra and Aqua BRDF/Albedo Model Parameters Daily L3 Global 0.05Deg CMG V006 C2 — MCD43C2 — MODIS/Terra and Aqua BRDF/Albedo Snow-free Parameters Daily L3 Global 0.05Deg CMG V006 C3 — MCD43C3 — MODIS/Terra and Aqua BRDF/Albedo Daily L3 Global 0.05Deg CMG V006		
BRDF_Albedo_Uncertainty	BRDF Inversion Information	*Product(s): C1, C2, C3* Calculation from how much variability there is in the retrieval over the area being considered using weights of determination as uncertainty. Weights of determination is a measure of whether the looks are well distributed over the hemisphere (not if all in one sample space).
BRDF_Quality	BRDF and albedo quality information	*Product(s): C1, C2* *Value* — *Meaning* 0 — best quality, 100% with full inversions 1 — good quality, 75% or more with best full inversions and 90% or more with full inversions 2 — relative good quality, 75% or more with full inversions 3 — mixed, 75% or less full inversions and 25% or less fill values 4 — all magnitude inversions or 50% or less fill values 5 — 50% or more fill values 255 — Fill Value
Albedo_Quality	Global albedo quality	*Product(s): C3* See "BRDF_Quality"

Field name	Meaning	Values (if applicable)
Percent_Inputs	Percent finer resolution data which contributed to the CMG	*Product(s): C1, C2, C3*
Percent_Snow	Percent of underlying data flagged as snow	*Product(s): C1, C3*
MCD43D products: The "D" products are provided at 1000 m spatial resolution. These products are retrieved from the original MODIS data rather than being an average of the 500 m products. Each band / parameter / quality indicator is provided as a separate product (i.e. only one field provided per product). *Key* *Product Number* *Product Name* D31 MCD43D31 MODIS/Terra and Aqua BRDF/Albedo QA BRDF Quality Daily L3 Global 30 ArcSec CMG V006 D33-39 MCD43D33 – 39 MODIS/Terra and Aqua BRDF/Albedo ValidObs Band 1 - 7 Daily L3 Global 30 ArcSec CMG V006 D41 MCD43D41 MODIS/Terra and Aqua BRDF/Albedo Uncertainty Daily L3 Global 30 ArcSec CMG V006		
BRDF_Albedo_Valid Obs_Band<n>	Valid Observation for Bands 1 – 7	*Product(s): D33-39* *Bit #: parameter* *Bit = 0* *Bit = 1* Bit 1: Day 1 quality Observation not used Valid clear observation … Bit 16: Day 16 quality Observation not used Valid clear observation
BRDF_Quality	BRDF	*Product(s): D31* See "BRDF_Albedo_Band_Quality_Band<n>"
BRDF_Albedo_Uncertainty	BRDF Inversion Information	*Product(s): D31, D41*

A.5 LAI and FAPAR QFs: Copernicus LAI and FAPAR

Field name	Meaning	Values		
NOBS	Number of available valid instantaneous estimates in the compositing window			
QFLAG	Bitwise quality flag	*Bit #: parameter*	*Bit = 0*	*Bit = 1*
		Bit 1: Land/water	Water	Land
		Bit 2: Vegetation class	No EBF case	EBF case
		Bit 3: Missing dekad interp.	Not interpolated	Interpolated
		Bit 4: Not used	-	-
		Bit 5: Obs. used for product computation in EBF case	Daily observations	Dekadal product
		Bit 6-7: Method used for product computation in no-EBF case	00: Second degree polynomials fit 01: Linear fit 10: Interpolation between the nearest dates 11: Nearest data or missing value	
		Bit 8: Instantaneous EBF classification flag	None	Value
RMSE	RMSE on LAI.			

A.6 LAI and FAPAR QFs: MODIS LAI and FAPAR (MOD15)

Field name	Meaning	Values		
FparLai_QC	Class flag	*Bit #: parameter*	*Value*	*Meaning*
		MODLAND_QC bits	0	Good quality (main algorithm)
			1	Other Quality (back-up algorithm)
		Sensor	0	Terra
			1	Aqua
		DeadDetector	0	Detectors apparently fine for up to 50% of channels 1,2
			1	Dead detectors caused >50% adjacent detector retrieval
		CloudState	00	Significant clouds NOT present (clear)
			01	Significant clouds WERE present
			10	Mixed cloud present on pixel
			11	Cloud state not defined, assumed clear
		SCF_QC	000	Main (RT) method used, best result possible (no saturation)
			001	Main (RT) method used with saturation. Good,very usable
			010	Main (RT) method failed due to bad geometry, empirical algorithm used
			011	Main (RT) method failed due to problems other than geometry, empirical algorithm used
			100	Pixel not produced at all, value couldn't be retrieved (reasons: bad L1B data, unusable data)

Field name	Meaning	Values		
FparExtra_QC	Class flag Fill = 255; Valid range = 0-254; Scale factor = na	*Bit #: parameter*	*Value*	*Meaning*
		Bit 0–1: LandSea Pass-Thru	00	Land
			01	Shore
			10	Freshwater
			11	Ocean
		Bit 2: Snow_Ice	0	No snow/ice detected
			1	Snow/ice detected
		Bit 3: Aerosol	0	No / low atmos. aerosol detected
			1	Average / high aerosol detected
		Bit 4: Cirrus	0	No cirrus detected
			1	Cirrus was detected
		Bit 5: CloudMask	0	No clouds
			1	Clouds were detected
		Bit 6: Cloud_Shadow	0	No cloud shadow detected
			1	Cloud shadow detected
		Bit 7: SCF_Biome_Mask	0	Biome outside interval <1, 4>
			1	Biome in interval <1, 4>
LaiStdDev_1km	For the MODIS derived LAI product, the mean and dispersion values of the LAI solution distribution function are taken as the retrieved LAI accuracy [26].	*Value*	*Meaning*	
		248	No standard deviation available, pixel produced using backup method	
		249 – 255	See fill values for main product parameters	

A.7 LAI and FAPAR QFs: QA4ECV TIP LAI / FAPAR

Field name	Meaning	Values (if applicable)		
Sigma_Lai_Fo Sigma_fapar_Fo	The uncertainty provided is one standard deviation of the Gaussian distribution. Tails outside the physically meaningful range reflect the possibility that the value cannot be consistently retrieved under the assumptions made.			
retrieval_flag_Fo	Quality flag provided with the product giving information on the usability of the product and the traceability (in terms of LUT used to generate the product)	*Bit #: parameter*	*Value*	*Meaning*
		Bit 0: Obs is fillvalue	1U	Missing input BHR
		Bit 1-5:	Not used	
		Bit 6: TIP untrusted	64U	Probably wrong, do not use
		Bit 7: Obs unusable	128U	Input BHR was unusable or snow fraction was out-of-range
		Bit 8: Obs inconsistent	256U	Input BHRs failed in basic consistency check
		Bit 9 + 10: Obs <s> high uncerts		At least one <s> input BHR uncerts above 20%
		Bit 12 + 13: Obs <s> only		Only <s> albedos were used
		Bit *various*: Masks		Marks the bits used for the LUTS used for different <s> <%> (under bits 14 – 25)
		Bit 14-25: Applied tip <s> <b> <%>		Marks the LUT used for the different uncertainty levels of <s> in each <b>
		Bit *various*: Masks		Marks the bits used for the LUTS used for different correlations between the <s> for BHR VIS and NIR uncertainties (under bits 26 - 30)
		Bit 26-30: Applied tip <s> <b> <%>		Marks the LUT used in relation to the correlation between the <s> for BHR VIS and NIR uncertainties

Various abbreviations used. <s> means snow / no snow. <b> means the band, in this case, VIS or NIR. <%> means a percentage, in this case, 5, 7. 10 or 20%

A.8 FAPAR QFs: QA4ECV FAPAR

Field name	Meaning	Values (if applicable)
FAPAR Uncertainty	The uncertainty values provided are a combination (in quadrature) of the FAPAR uncertainties from the retrieval algorithm and those associated with the assumptions coming from RT models and the optimisation step. In addition the accuracy and precision of the data are evaluated using simulated data. Information from [16].	
Categorisation Flag	Surface type categorisation flag. Indicates how the pixel has been processed.	*Value* — *Meaning* 0 — Land surface. FAPAR valid 1 — "Bad data" ". FAPAR not computed. 2 — "cloud, snow and ice" ". FAPAR not computed. 3 — "water body and deep shadow". FAPAR not computed. 4 — Bright surface. FAPAR set to 0. 5 — Surface undefined. FAPAR error value used 6 — FAPAR value less than zero; FAPAR set to error value. 7 — FAPAR value greater than 1; FAPAR set to 1.

B Atmosphere Quality Indicators

This annex provides details of the QFs used in the atmosphere products studied within this document. The colour coding is as follows:

Uncertainty Estimate	Designated Quality Flag	Ancillary data and other information which indicates quality

B.1 CO QFs: MOPITT CO column & profile

Field name	Meaning	Values (if applicable)
"Retrieved CO Total Column" (index 2 of array)	Standard uncertainty (standard deviation). It represents the total retrieval error for the total column.	
"Retrieved CO Total Column Diagnostics". Array size: (2,ntime)	Diagnostics related to the retrieved CO total column are provided in the L2 field 'Retrieved CO Total Column Diagnostics'. For each retrieval, this two-element vector contains both the estimated smoothing error and measurement error contributions to the total retrieval error (in that order). Smoothing error represents the uncertainty in the retrieved total column due to the departure of the actual total column averaging kernel from the ideal total column averaging kernel. Measurement error describes the uncertainty due to errors in the measured radiances. (molec./cm^2).	
"Retrieved CO Mixing Ratio Profile" (index 2 of array) & "Retrieved CO Surface Mixing Ratio Profile" (index 2 of array)	Standard uncertainty on retrieved vertical VMR profile.	
Retrieval Error Covariance Matrix	A posteriori covariance matrix on log10(VMR) profile.	
Degrees of Freedom for Signal	Trace of the averaging kernel matrix. Pieces of independent measurement information for the vertical profile.	

Vertical profile is partitioned in two fields: one for the surface level and one for the other 9 levels.

Field name	Meaning	Values (if applicable)
Surface Index [Ancillary data and other….]	Surface type categorisation flag.	*Value* — *Meaning* 0 — water 1 — land 2 — mixed (e.g., coastline)
Level 1 Radiances and Errors (W/m 2 Sr) [Ancillary data and other….]	Radiances (per channel) and their uncertainties. Allows to calculate an observation quality index on which can be filtered. Example for TIR-only product: $OQI_T = ((\sigma_{5A}/R_{5A})^2 + (\sigma_{5D}/R_{5D})^2)^{-1/2}$ Where A, D refer to channels, R to radiance and sigma to uncertainty.	
Signal Chi2 [Ancillary data and other….]	Signal chi-squared diagnostic, dimensionless.	

B.2 NO2 QFs: DOMINO v2 NO2 tropospheric column

Field name	Meaning	Values (if applicable)	
TroposphericVertical ColumnError	Standard uncertainty in the tropospheric NO2 vertical column.		
VCDTropErrorUsing AvKernel	Standard uncertainty in the tropospheric NO2 vertical column, without profile error contribution.		
TroposphericColumn Flag	Flag to indicate when the retrieved tropospheric column is unreliable.	*Value*	*Meaning*
		0	Tropospheric column for more than 50% determined by observed information.
		-1	Tropospheric Column for more than 50% determined by forward model parameter assumptions (cloud radiance fraction > 50%), or row anomaly.
		-127	Missing data

Field name	Meaning	Values (if applicable)		
GroundPixelQualityFlag	Additional information on the viewing scene.	*Bit #: parameter*	*Value*	*Meaning*
		Bit 0-3: land / water flags	0	Shallow ocean
			1	Land
			2	Shallow inland water
			3	Ocean coastline / lake shoreline
			4	Ephemeral (intermittent) water
			5	Deep inland water
			6	Continental shelf ocean
			7	Deep ocean
			15	Error Flag for land / water
		Bit 4: Sun glint possibility flag		
		Bit 5: Solar eclipse possibility flag		
		Bit 6: Geolocation error flag		
		Bit 8 – 14: Snow / ice	0	Snow free land
			1 – 100	Sea ice percentage
			101	Permanent ice (Greenland / Antarctica)
			103	Dry snow
			104	Ocean
			124	Mixed pixels at coastline
			125	Suspect ice value
			126	Corners (undefined)
			127	Error
		Bit 15: NISE nearest neighbour filling flag	0	Not set
			1	Set
SurfaceAlbedo	Surface albedo from [27]. Values hold for 439 nm. Recommendation: discard pixels with surface albedo > 0.3.			

B.3 NO2 QFs: QA4ECV NO2 column

Field name	Meaning	Values (if applicable)
tropospheric_no2_vertical_column_precision	Standard uncertainty in the tropospheric NO2 vertical column.	
tropospheric_no2_vertical_column_precision_kernel	Combined contribution of uncertainties in surface albedo, cloud fraction and cloud pressure to the uncertainty in tropospheric NO2 vertical column.	
processing_error_flag	Flag indicating whether the processing was successful (0) or failed (1).	*Value* *Meaning* 0 Successful 1 Failed
INPUT_DATA/snow_ice_flag	Flag indicating snow/ice at centre of ground pixel.	*Values* snow-free_land sea_ice_[x]_percent permanent_ice snow mixed_pixels_at_coastlines suspect_ice_value corners_or_other_error ocean
DETAILED_RESULTS/number_of_spectral_points_in_retrieval	Number of spectral points used in the retrieval.	

B.4 HCHO QFs: QA4ECV HCHO column

Field name	Meaning	Values (if applicable)
tropospheric_hcho_vertical_column_uncertainty_random	Random standard uncertainty on formaldehyde vertical column.	
tropospheric_hcho_vertical_column_uncertainty_systematic	Systematic standard uncertainty on formaldehyde vertical column.	
processing_error_flag	Flag indicating whether the processing was successful (0) or failed (1).	*Value* *Meaning* 0 Successful 1 Failed
INPUT_DATA/snow_ice_flag	Flag indicating snow/ice at centre of ground pixel.	*Values* snow-free_land permanent_ice dry_snow wet_snow mixed_pixels_at_coastlines suspect_ice_value corners ocean
DETAILED_RESULTS/number_of_spectral_points_in_retrieval	Number of spectral points used in the retrieval	