

# Quality criteria for SCIAMACHY products and procedures for quality assessment and validation

Written by: SCIAVALIG, August 2006 (update May 2008)

## 1. Introduction

The first four years of its mission, SCIAMACHY has proven to be a remarkable stable instrument. Scientific products, generated at several institutes, have shown the great scientific potential of SCIAMACHY. Even the SWIR spectrum, although suffering from an ice layer, is shown to be good enough to retrieve the planned species with improving precision, thanks to the creativity of the highly motivated and competing scientific teams.

To increase the scientific output from these developments a system is set up for registration of SCIAMACHY products, and for reporting on the quality of these products. The anticipated benefits are:

- Developing institutes are encouraged to generate larger dataset;
- Developing institutes are encouraged to improve their algorithms;
- Datasets become accessible more easily;
- Datasets will be validated by more groups;
- The quality of datasets becomes clearer;
- The use of SCIAMACHY data will increase for scientific and other applications.

After a product has been registered, its content, documentation and availability will be assessed by SCIAVALIG with respect to the criteria specified in Section 2. When the product is fulfilling these criteria, the validation of the product will be stimulated and monitored, and validation results will be reported. In Section 3 the different stages of validation are defined, and in Section 4 the procedures are detailed.

## 2. Criteria for product content, documentation, and availability

### 2.1 Product content

#### 2.1.1 General

##### Required

- The filenames should be unique and should have a time description.
- The product should contain a header with at least:
  - software version number
  - point of contact (e.g. e-mail of data originator)

##### Recommended

- A software version identifier is in the filename.
- Date/time of product generation
- Input files are traceable by the user, i.e. unique level 0, level 1b, or level 2 identifier, including the software version numbers, and the version numbers

for calibration parameters, like m-factors, dead and bad pixel mask, keydata, etc

- All retrieval method related numerical data are given, like intermediate results and retrieval input data that vary from scene to scene. This will allow the user to fully understand the possible applications of the product, and to make selections on the basis of retrieval sensitivities, if necessary.

### 2.1.2 Level 1

#### Required

- date/time of measurement
- 4 corner coordinates (latitude, longitude) of ground scene
- solar zenith angle, viewing angle, azimuth angle
- wavelength or detector pixel identifiers
- radiance, irradiance or reflectance with error per wavelength or detector pixel

### 2.1.3 Level 2 from nadir

#### Required

- date/time of measurement
- 4 corner coordinates (latitude, longitude) of ground scene
- either: geophysical value with retrieval error, or: index value

#### Recommended

- center coordinates of ground scene
- state id
- identifier for forward and backward scan
- solar zenith angle
- viewing angle
- azimuth angle
- intermediate results (e.g. slant column, air mass factor)
- other retrieved parameters (e.g. slant columns of other trace gases, cloud cover)
- retrieval input parameters (e.g. assumed albedo, surface pressure, temperature, ghost column)
- retrieval quality flags

### 2.1.4 Level 2 from limb

#### Required

- date/time of measurement
- corner coordinates (latitude, longitude) of the ground scene (see also under 'Recommended')
- for height-dependent products:
  - altitudes or pressure levels
  - geophysical value on or between these altitudes or pressure levels
  - retrieval error on geophysical value
  - specific additional data, depending on retrieval method, needed to use the data for research applications, e.g.: averaging kernels, a-priori profiles, weighting functions
- for 'integrated' products, either: geophysical value with retrieval error or: index value

Recommended

- state id
- solar zenith angles
- azimuth angles
- intermediate results (e.g. slant columns per altitude)
- instead of the corner coordinates of the ‘ground scene’: the corner coordinates of the volume probed by the limb observation, e.g. of the volume (or area) from which 50% and/or 90% of the observed photons originate
- other retrieved parameters (e.g. pointing correction)
- retrieval input parameters (e.g. assumed albedo, surface pressure, temperature profile)
- retrieval quality flags

*2.1.5 Level 3 and 4 gridded products*Required

- date/time of product
- grid definition
- geophysical values or index per grid cell

Recommended

- estimated errors per grid cell
- averaging or interpolation method in header

**2.2 Documentation***2.2.1 Algorithm Document (AD)*

The retrieval algorithm for the product is well described, i.e., mathematical methods, assumptions, input data, and auxiliary data are all listed and properly referenced. The AD is available to the users. It either contains all details about the used algorithm or it refers to the documentation in which these details can be found. However, it is not enough to only refer to other documentation, at least a few sentences per section are given in the AD so that the user has an overview of the algorithm specifics from this document alone. The table of contents includes:

1. Forward model

The level of detail depends on what can be found in the literature. In general, the user should be able to understand the characteristics of the product, its limitations regarding accuracy, precision, and resolution for all possible situations. E.g., if an RTM is used which is fully described in a paper, the details need not be in the AD. However, specific adaptations and assumptions regarding, e.g., geometry, scattering, clouds, aerosols, atmospheric constituents, etc, which are not in the paper, should be addressed in the AD.

2. Inversion procedure3. Auxiliary data

The actual data bases used can also be given in the Product Specification Document. Here it should be clear what data is taken from other sources as input to the retrieval.

4. Sensitivity and error analysis5. Algorithm validation

This section should demonstrate that the algorithm works in a simulated situation and/or when comparing to other algorithms or models and/or when applying the

algorithm to other instruments. This section can be combined with the 'Sensitivity and error analysis' section.

6. Recommendations for product validation

Recommendations can be given on the kind of validation measurements (e.g. instruments, locations, season, time-difference, special events), on the use of models, and on what issues should be addressed (e.g. difference between cloudy and cloud-free retrievals, dependence on solar zenith angle, aerosol load, surface albedo).

2.2.2. *Product Specification Document (PSD)*

The product is well described and the description is available to the users. There should be one document for each new software version. It should contain all details about the product for this particular release. The PSD should always be released together with the data products. The table of contents should include:

1. Product description

A short description of the product. This description should include details on the geophysical value and the geographical extend. Also a short description of all data accompanying the product should be given (e.g. a-priori profiles, retrieved albedo, pointing correction number). For trace gas profiles, for instance, it should include information on number of profiles per state, units, geophysical value (e.g., partial column between two height levels), altitude range, number of retrieval levels and their spacing, and the actual resolution. A reference to the AD should also be given in this section.

2. Product format specification

3. Software release history

This can be limited to changes with respect to the previous version, including a reference to the previous PSD.

4. Implementation details

Here all details that are not in the AD should be given. For example: initialisation parameters, data bases with references, fitting windows, selection criteria, thresholds.

5. List of known issues

6. Data quality assessment

What is already known from validation studies, and what improvements you expect from this software release. The sections on 'List of known issues' and 'Data quality assessment' together are comparable to the ESA data disclaimers for SCIAMACHY. Both sections can be combined.

## 2.3 Availability

- The data set is generated with one software version for a consecutive period of at least 1 year, and for at least 50% of the orbits in that period;
- The data set is available for scientific research at no cost.

It is recommended to make the data available without any restrictions for use, such as offering co-authorship when publishing the data. Such restrictions may limit the use of the data.

### **3. Validation stages**

The SCIAMACHY products can be in a different stage of validation. They can be 'validated', 'partly validated', or 'demonstrated'. These terms do not give information about the quality of the product, but about the completeness of the knowledge of the quality. Also for these terms we need criteria. The criteria, listed here, are loosely based on the criteria in the SCIAMACHY validation requirements document (SCIAVALIG, 1998):

#### **Validated**

A product is called "validated", when:

- comparisons with other data have been performed on a global scale, covering at least the latitude bands -90/-60, -60/-30, -30/30, 30/60, 60/90;
- comparisons with other data have been performed for different times, covering at least the four seasons;
- comparisons have been performed with different independent measurement techniques or models of known (i.e., traceable) quality.

For a "validated" product estimated accuracy and precision values can be given, including their seasonal and latitudinal dependencies. Specific regimes where the product is less accurate can be identified.

#### **Partly validated**

A product is called "partly validated" when at least two of the criteria under "validated" are met.

#### **Demonstrated**

A product is called "demonstrated" when comparisons have been performed with at least one measurement technique or model of known quality.

#### **4. Procedures**

The procedure for registration, validation, and reporting is as follows:

1. The product developer fills in a questionnaire for registration (available on the web).
2. The product is directly listed on the central web site, but without any report on quality.
3. SCIAVALIG (via the product coordinator) checks whether the criteria for product content, documentation and availability are fulfilled. The result is reported on the web site via quality flags.
4. If the product fulfils these criteria, SCIAVALIG establishes the initial stage of validation, advises on further validation of the product, monitors the results, and reports on its quality (mostly via the product coordinator).
5. The quality flags on the web site are adapted whenever appropriate.

The web site is maintained by the SCIAVALIG support office. Apart from the list of products and their quality flags, it will contain links to the data, the contact person, and to all relevant documentation. The URL of the web site is:

<http://www.sciamachy.org/products>