8.   Data Processing and Products

G. Lichtenberg[1], K.-U. Eichmann[2], C. Lerot[3], R. Snel[4], S. Slijkhuis[1], S. Noël[2], R. van Hees[4], B. Aberle[1], K. Kretschel[1], M. Meringer[1], D. Scherbakov[1], H. Weber[1], A. von Bargen[5]

[1] Remote Sensing Technology Institute, German Aerospace Center (DLR-IMF), Oberpfaffenhofen, 82234 Wessling, Germany
[2] Institute of Environmental Physics / Institute of Remote Sensing (IUP-IFE), University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany
[3] Belgian Institute for Space Aeronomy (BIRA-IASB), 3 Avenue Circulaire, 1180 Brussels, Belgium
[4] SRON, Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands
[5] German Aerospace Center, Space Agency, Königswinterer Str. 522-524, 53227 Bonn, Germany

Abstract: SCIAMACHY data processing generates products on various levels as required by the user community. These products are either produced in the ENVISAT Payload Data Segment as *operational* products or by science institutes as *scientific* products or *value-added* products. Operational processing occurs in two steps – level 0-1b and level 1b-2. In both steps, different timescales may apply – near-realtime, fast delivery or offline. Level 0-1b processing generates geolocated and calibrated radiances from the raw atmospheric measurements, as well as from measurements for calibration and instrument monitoring. The algorithms convert measured signals into calibrated radiances. Therefore, a sequence of calibration steps has to be applied, starting with correcting the memory effect and non-linearity and ending with applying the radiometric instrument response. Particular attention has to be given to the correction of polarisation and degradation. The goal of level 1b-2 processing is to provide geophysical parameters such as column densities and profiles from trace gas species, as well as cloud and aerosol parameters. Nadir measurements permit retrieving total column densities or cloud and aerosol parameters. From limb observations, height resolved profiles of atmospheric parameters can be inferred. Together with the scientific and value-added products, the SCIAMACHY data can serve a wide range of applications.

Keywords: Operational products – Scientific products – Value-added products – Level 0-1b processing – Level 1b-2 processing

The processing of SCIAMACHY measurement data has to serve the requirements of different user communities. Therefore, products, differing in content, are provided at various levels. In addition, product generation may occur in two environments. Whereas the *operational* data processing follows the specific guidelines and rules of the ENVISAT Payload Data Segment (PDS) architecture and organisation, *scientific* product generation is based on science group algorithms which have not been developed within the framework of the PDS. This twofold approach, when fully implemented, has the advantage that standard products with well-defined requirements can be generated in a configuration controlled but also flexible processor regime. It relieves users from providing significant resources for basic processing steps, e.g. calibrating the measurement data. On the other hand, scientific products can utilise most recent, often advanced algorithms since the stringent requirements of the operational ground segment need not to be fulfilled. Also, new geophysical parameters which are not part of the operational chains are first introduced as scientific products demonstrating the potential of SCIAMACHY's measurement data. Later, these algorithms may be implemented in the operational environment, thus providing users with a continuous and up-to-date global dataset. This feedback between scientific and operational products ensures that algorithm improvements are rapidly and continuously fed into the operational system.

As outlined in chapter 2, the SCIAMACHY products follow the general ENVISAT data product rules and requirements. Thus, a level 0, level 1b and level 2 file is generated for each orbit. The SCIAMACHY specific definitions include:

  – Level 1b: The level 0 content plus all data needed for calibration together with the geolocation of the ground pixel measured.
  – Level 2: Geophysical quantities such as total trace gas columns, trace gas profiles and cloud and aerosol information.

Level 1b and level 2 products are distributed directly to the users. For calibration, additional products were defined:

  – Calibrated spectra (level 1c): The user has the choice to apply either the entire or only a limited set of calibrations (see below). Applying the full calibration results in Earth radiances as a function of wavelength.
  – Auxiliary Data Files (ADF): Data needed for calibration and geolocation are usually not measured as frequently as the Earth radiances. During the level 0-1b processing, the appropriate ADF are extracted and archived. Combining appropriate ADF and Earth measurement data yields a level 1b product.

Since the level 1b data contain all information required for calibration, including the relevant ADF, level 1c products can be generated with the specific software tool *SciaL1c*, which is available to the users. It permits filtering level 1b data and processing these to level 1c, i.e. calibrated radiances.

The operational processing is done in two steps starting with level 0 to level 1b and ending with level 1b to level 2. Each day of SCIAMACHY operations generates 14 or 15 level 1b and level 2 products, one for each orbit. ESA's responsibility for data generation, dissemination and archiving ends with level 2. This does not exclude experienced science users from processing their own scientific products of the same levels. It is up to the individual user how to specify and process products higher than level 2, known as *value-added* (VA) products. In the VA environment, geophysical parameters are often gridded on a global scale, i.e. these products are well suited to provide the interested public with SCIAMACHY's view of the Earth's atmosphere.

8.1   Operational Processing Overview

Operational processing of SCIAMACHY measurement data occurs in the PDS as part of the ENVISAT mission. It covers:

  – Facilities to serve the need of generating, archiving, and disseminating data products including the interface to the users, e.g. Processing and Archiving Centres (PAC)
  – Software to generate the data products, e.g. data processors and calibration facilities which run within the processing centres
  – Data products on level 0, level 1b, and level 2
  – Tools for monitoring and controlling the quality of the operational production chain.

Basic processing related information is provided via the file names of the products. A filename includes the sequence

  – Product ID: 'SCI_NL__0P', 'SCI_NL__1P', 'SCI_NL__2P' and 'SCI_OL__2P' for level 0, 1b or 2 data
  – Processing status flag: 'N' for NRT products, letters between 'N' and 'V' for consolidated products
  – Originator ID: 'PDK', 'PDE', 'LRA', 'DPA' for PDHS-K, PDHS-E, LRAC or D-PAC
  – Start date: year, month and day of measurement start

- Start time: hours, minutes and seconds of measurement start in UTC
- Duration of product coverage in seconds
- ENVISAT mission phase
- ENVISAT cycle number
- Relative orbit number within cycle
- Absolute orbit number
- File counter
- File extension: 'N1' corresponds to ENVISAT

This permits an unambiguous identification of files for a given orbit. A filename example is provided in Fig. 8-1. All ENVISAT files have two headers – the Main Product Header (MPH) which contains information common to all instruments of the platform such as orbit data, and the Specific Product Header (SPH) which holds instrument specific information.
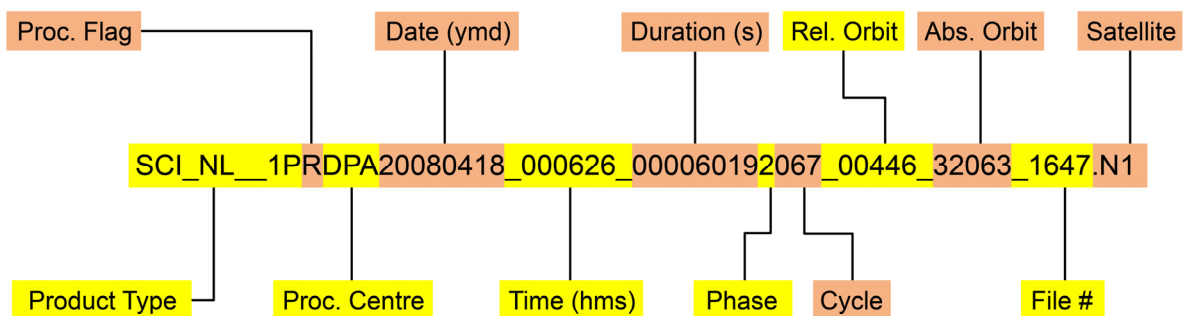

Fig. 8-1: An example for a SCIAMACHY level 1b product filename illustrating the individual filename fields. (Courtesy: DLR-IMF)

Level 1b Product

In addition to the MPH and SPH, the level 1 file contains the Measurement Data Sets (MDS) and the Annotated Data Sets (ADS) with supplemental information needed for e.g. calibration. Each type of measurement has its own data structure, giving four different MDS:

- Nadir MDS
- Limb MDS
- Occultation MDS
- Monitoring MDS

The atmospheric measurements for nadir, limb and occultation can be found in the corresponding MDS in raw format, i.e. uncorrected binary units. The MDS also contains calibration parameters such as e.g. size of memory effect, amount of stray light, polarisation parameters, for each measurement. Calibration measurements using the Sun, the Moon or the calibration lamps are stored in the monitoring MDS. Individual dark measurements are not stored in the product, thus avoiding too large file sizes. However, the average values and standard deviations are available for each dark state.

The ADS contain the calibration data, quality indicators, geolocation and instrument configuration parameters from the level 0 files. Two types of ADS exist, *global* ADS with data from the auxiliary data files (ADF, see above), and the *newly calculated* ADS: if calibration data are measured in a given orbit, these are stored in the newly calculated ADS. The global ADS are used in the standard calibration of the spectra, with the exception of the dark correction in the short-wave infrared (SWIR) channels for which the dark measurements performed during the same orbit are

always used (see also Fig. 8-3). The experienced user also has the possibility to apply the newly calculated ADS, but only if the corresponding measurements were performed in the same orbit.

Level 1c Product

The level 1c file has the same structure as the level 1b file. It contains the data which were filtered with the SciaL1c tool. Additionally, the applied calibration options are provided in an ADS for the sake of traceability.

Level 2 Product

In the level 2 product, four basic types of MDS are available:

- Cloud and aerosol MDS
- Nadir MDS
- Limb MDS
- Limb cloud MDS

The cloud and aerosol MDS combines cloud and aerosol information from nadir measurements. It is followed by multiple MDS corresponding to the different trace gas products retrieved from nadir and limb geometries, one for each trace gas. The limb cloud MDS contains height resolved information about clouds in the limb view. Additionally, the level 2 product contains several ADS with auxiliary information, e.g. data quality and geolocation.

Detailed information about SCIAMACHY's operational data products is provided via ESA's websites for the ENVISAT missions. The product descriptions can be found at

http://envisat.esa.int/instruments/sciamachy/data-app/

while product status, including disclaimers, are given via

http://envisat.esa.int/dataproducts/availability/.

Finally, a listing of the product performance on a daily and bi-monthly basis is provided by

http://earth.esa.int/pcs/envisat/sciamachy/reports/.

Processing Chains

Remote sensing data are used for different applications with different requirements. While forecasting requires data availability shortly after sensing, climate research and other longer-term scientific applications require the most accurate data calibration and retrieval. In order to satisfy both requirements, two separate processing chains were set up: the near-realtime (NRT) level 0-1b processing together with the fast delivery (FD) of level 2 products and the offline (OL) level 0-2 processing chain (see Fig. 8-2). While the processors used for both chains are identical, the input data are different. The tight temporal requirements for the NRT chain do not allow the processing and inventory of all the calibration data, including calculation of attitude corrections. However, experience shows that the content of NRT and OL products differs only by small amounts. Main differences between the processing chains comprise:

- Delivery time: NRT level 1b products are delivered a few hours after sensing, FD products within 24h after sensing. Offline data are delivered typically 1 week after sensing.
- Start/Stop times: OL products contain one orbit always starting at the ascending node crossing (ANX). NRT/FD data can contain data from different parts of the orbit.

– Geolocation: The geolocation differs slightly since the restituted attitude information for the ENVISAT platform is only available a few days after sensing.
– Calibration Data: In the NRT/FD chain the latest available in-flight calibration data are used while in the OL chain the most recent calibration data made before the measurement are applied.

All operational data products are subject to quality monitoring under the responsibility of the Payload Data Control Centre (PDCC). Its goal is to screen all products at the time of generation in order to identify anomalies or deviations from expected results. Quality monitoring includes content and consistency checks, e.g. formal correctness of the product or parameter limits. In the case of detected anomalies, data shall be flagged to initiate further actions. The PDCC executes quality monitoring activities on various timescales ranging from daily to multi-monthly. Level 0 products receive particular attention since they are the basis for all further products. The national agencies providing SCIAMACHY receive copies of the complete level 0 dataset for their own purposes, including instrument monitoring activities. In support of such functions, DLR as one of the instrument providers has generated a master data archive from the consolidated level 0 product stream. This repository, maintained throughout the mission, is free of occasional consolidation flaws; it keeps for each orbit a single consolidated level 0 product which reflects the instrument measurement sequences as best as possible. Whenever the PDS schedules reprocessing campaigns – after major algorithm and processor updates have become operational – the content of the master archive forms the applicable level 0 product basis.

8.2   Operational Level 0-1b Processing

SCIAMACHY level 1b data products comprise geolocated and calibrated radiances of the scientific measurements, as well as measurements for calibration and instrument monitoring. The algorithms used in operational level 0 to 1b processing are primarily driven by the scientific need to convert measured signals into calibrated radiances. However, constraints, particularly those imposed by the operational data processing environment or sometimes by instrument operations, may force the use of different strategies as would be employed for a ground-based instrument. The desire to obtain a level 1b data product which is not excessively large and complicated imposes an additional constraint. The principle processing cycle starting with instrument level 0 data and ending with the level 1b product is outlined in Fig. 8-2. A major requirement from the ENVISAT PDS architecture rules is that there may be only one output product per processing chain, in this case the level 1b product. As a consequence, the level 1b product must not only hold processed science data but also calibration measurements and instrument monitoring data, as well as newly calculated calibration parameters (see above).

In the operational processing from level 0 to 1b, all necessary calibration constants for each science measurement are processed from the input calibration data using both ground-based and in-flight measurements. The level 1b data product contains the raw detector signals of the science measurements plus calibration constants, mainly coded as single byte integers. In addition to measurement specific calibration constants, e.g. stray light or atmospheric polarisation, lookup tables are generated for globally applicable calibration constants, e.g. instrument polarisation sensitivity as a function of scan angle. The Instrument Engineering Calibration Facility (IECF) copies the calibration data from the level 1b product into the calibration database. The calibration data files themselves are generated by the *SciCal* tool using level 0 data as an input. In order to enable users to process level 1b products to fully calibrated data, the calibration application tool *SciaL1c* is provided (see above). For the user's convenience certain calibrations can be optionally omitted at extraction, or a subset of data can be filtered out.
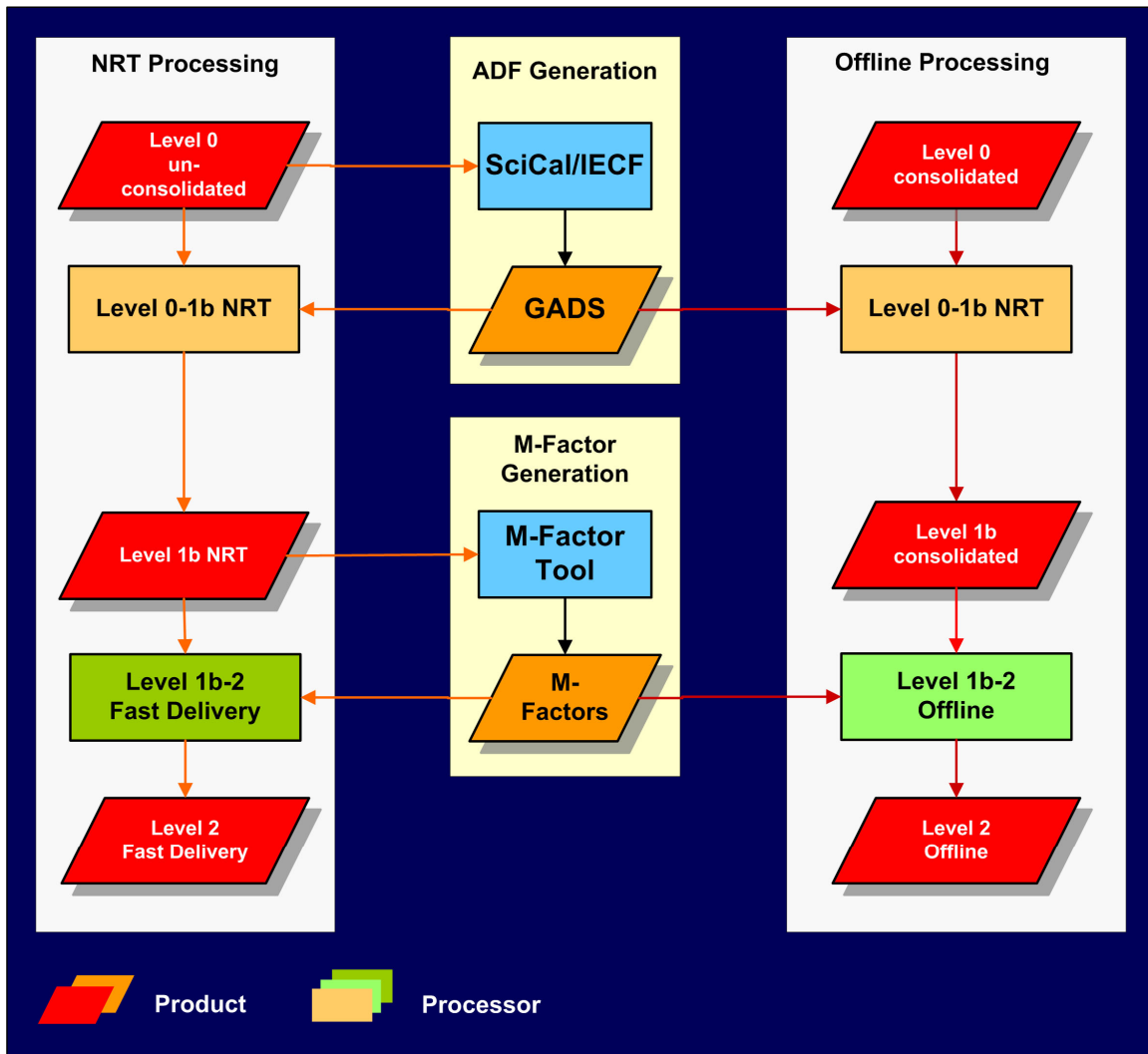
Fig. 8-2: General processing flow in NRT/FD (left) and OL (right). Both chains receive the calibration information from the same auxiliary flow. (Courtesy: DLR-IMF)

Fig. 8-3 provides an overview of the steps for level 1b to level 1c calibration of scientific measurements (nadir, limb and solar occultation) and the used measurement data and ADS. The individual calibration related steps as listed below reflect the calibration characteristics as described in chapter 5 and specified in detail by Slijkhuis (2004) in the Algorithm Theoretical Baseline Document (ATBD).
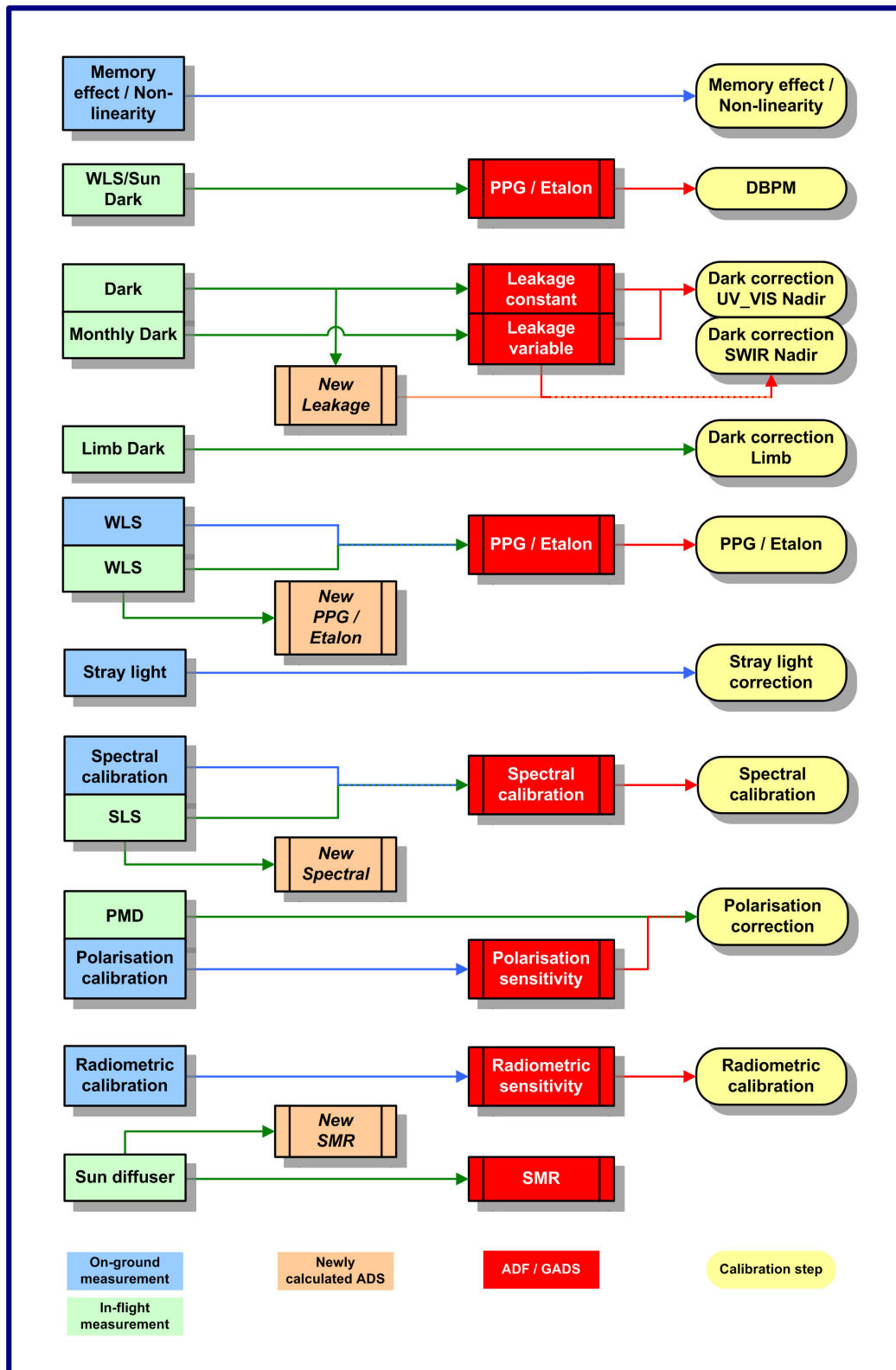
Fig. 8-3: Relations between calibration measurements (blue: on-ground, green: in-flight), ADF/GADS (red), newly calculated ADS (orange) and calibration steps (yellow). The calibration information is stored either directly in the newly calculated ADS of the level 1b file or written into an ADF and subsequently copied into the GADS of the level 1b file. Some calibration parameters, e.g. the stray light fractions, are contained in the measurement datasets of the level 1b file. Except for the *NEW LEAKAGE* ADS, the newly calculated ADS are not used in the standard calibration. (Courtesy: DLR-IMF)

Memory Effect and Non-Linearity

All channel detector readouts have to be corrected for the memory effect (in channels 1-5) or non-linearity (channels 6-8). The memory effect correction requires the value of the previous readout. If this is not available, e.g. for co-added measurements, an estimate has to be made. In case of co-adding, the value of each readout before co-adding is estimated on the basis of the PMD intensity which is sampled at a higher rate than the detector. Non-linearity correction is implemented similarly for co-added signals. A larger uncertainty in memory effect correction exists for limb measurements: at each new tangent height a reset readout of the detectors with an integration time of 31.25 msec is performed. This reset readout is discarded on-board, i.e. it is not available in the data. Similarly, the memory effect for the very first readout in a state can only be estimated. This is because before the first readout, the detector picks up an unknown signal as the mirrors are driven to their start position. For both cases an estimation of the memory effect is based on the measured signal and state setup parameters.

Dark Signal

The dark signal can be separated into three components:

- Analogue offset or fixed pattern noise: It is added by the electronics to each pixel to avoid negative or near zero signals. Thus, no dependence on the thermal environment or exposure times exists.
- Leakage current: It comprises the time dependent part of the dark signal mainly caused by thermally created electron-hole pairs in the detector. Its orbital phase independent component is derived from the orbital measurements on the dark side of the Earth. Different illumination conditions along the orbit with corresponding small thermal perturbations result in an orbital dependent part which is derived as a function of 12 orbit segments from the monthly dark measurements. This variable leakage current is only relevant for the SWIR channels (Fig. 8-4).
- Stray light: Near sunrise, the limb port of the instrument is directly exposed to sunlight leading to a significant amount of spatial stray light. This stray light is characterised during the monthly dark orbits. It is only relevant for orbit phases around sunrise, i.e. at high latitudes over the northern hemisphere. The stray light information is used for monitoring purposes only. For an actual correction of spatial stray light, the spatial coverage of the data is insufficient.

All orbital phase independent dark parameters are part of the *LEAKAGE CONSTANT* ADS. The *LEAKAGE VARIABLE* ADS contains the orbital phase dependent components of the dark signal. Both ADS are used in the dark correction for nadir measurements of channels 1-5. After launch, it turned out that the original dark signal correction using *LEAKAGE CONSTANT* ADS and *LEAKAGE VARIABLE* ADS had to be adjusted for limb data and the SWIR channels:

- Limb Data: At the end of each limb state, the Instantaneous Field of View (IFoV) is set to 250 km and a dedicated dark measurement is executed. This dark measurement is used for the dark correction of limb data.
- SWIR channels 6-8: Since these channels are sensitive to small changes of temperature and the ice layer in channels 7 and 8 leads to changes of the dark signal from one orbit to the next, the dark parameters from the same orbit (from the *LEAKAGE NEW* ADS*)* are used. The orbit phase dependence is taken into account by additionally exploiting data from the *LEAKAGE VARIABLE* ADS.
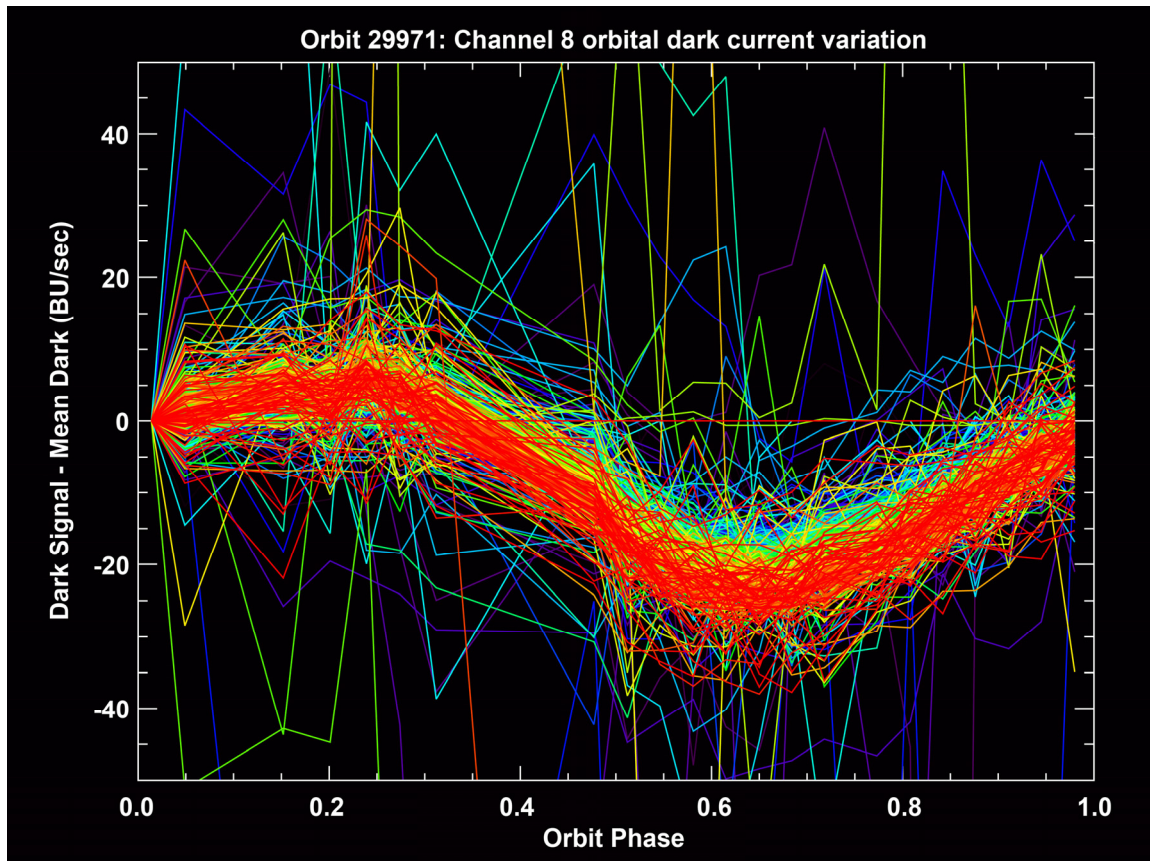
Fig. 8-4: Orbital variation of the thermally induced dark signal for each pixel in channel 8. Orbit phase = 0 corresponds to eclipse entry. The *Mean Dark* is the dark signal per pixel averaged over a complete dark current state close to phase 0. It permits to normalise the orbital variation at phase 0. (Courtesy: DLR-IMF)

Generally, no calibration measurements are used which fall into the South Atlantic Anomaly (SAA, Fig. 8-5). This is to reduce contamination of the spectrum from particle radiation which appear as spikes on the detector signal. For orbital dark measurements, this rule would imply that for many orbits per day the dark signal is not updated. Therefore, measurements from the SAA are incorporated into the orbital dark calibration for channels 6-8, even though this may imply a small degradation in quality. A 'hot pixel detection algorithm', which analyses a sequence of measurements for each detector pixel to detect and reject spikes in the data, will filter out large deviations. The individual dark measurements are not stored on the level 1b product, but state averaged mean values (after removal of spikes) and their standard deviations are stored in the *DARK AVERAGE* ADS.

Pixel-to-Pixel Gain and Etalon

The Pixel-to-Pixel Gain (PPG) correction aims at correcting relative gain changes with time. It is monitored in-flight using the White Light Source (WLS) measurements. However, observations show that the PPG does not change significantly in-orbit, thus the correction is currently not applied.

The detector etalon effect is caused by multiple reflections at the boundaries of the protective silicon oxide layer on top of the light detecting silicon. It leads to interference pattern in the measured spectrum. If material is deposited on top of the detector, the optical path and thus also the interference pattern changes when compared to the situation during the on-ground calibration. This change is corrected using data from the weekly WLS measurements and is only relevant for channels 1-5. Both PPG correction and the detector etalon correction are stored in the *PPG ETALON* ADS.
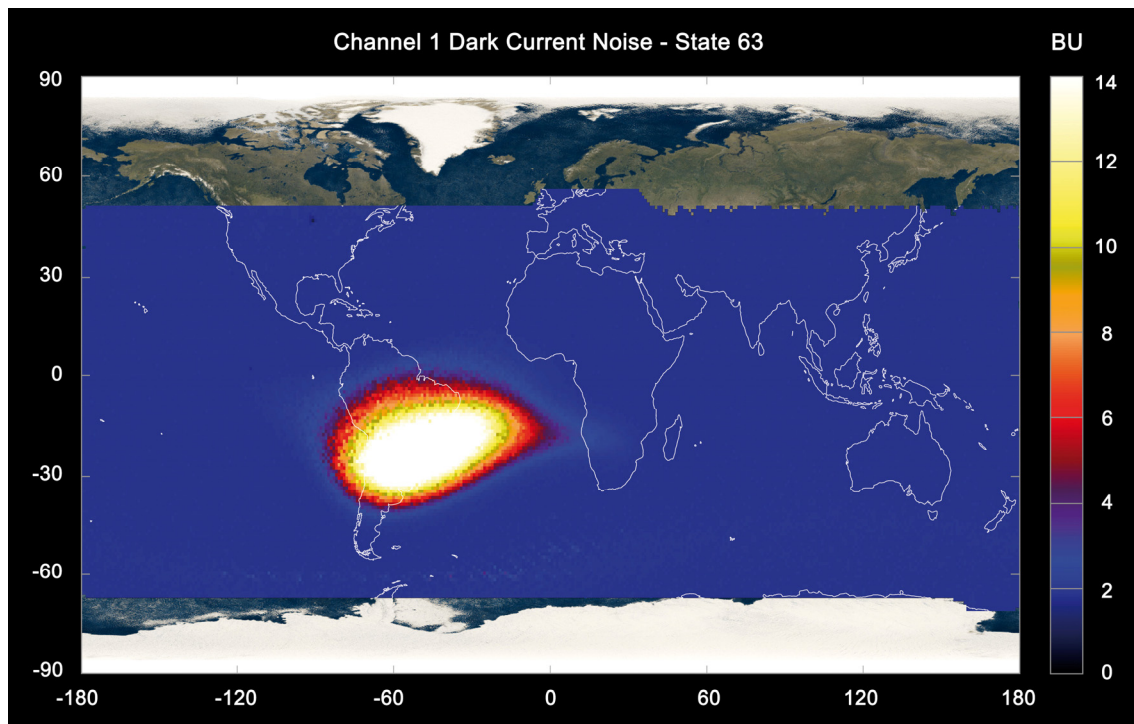
Fig. 8-5:   The SAA as derived from SCIAMACHY dark current measurements. (Courtesy: SRON)

Dead and Bad Pixel Mask

The SWIR channels 6+, 7 and 8 suffer from a rising number of bad or even dead pixels which are not – or only to a small extent – suitable for retrieval. The reason is a lattice constant mismatch between the substrate material and the light detecting material of the detectors (Kleipool et al. 2007). Such dead and bad pixels are detected using dark, WLS and Sun measurements. Several thresholds were determined to separate dead and bad from good pixels. The number of dead and bad pixels is not constant (Fig. 8-6). It increases with in-orbit operations time due to proton impacts. Therefore, a dead and bad pixel mask is calculated for each orbit and stored in the *PPG ETALON* ADS.

Spectral Stray Light

The corrections for spectral stray light are derived from each science measurement using calibration constants measured on-ground. Contrary to the situation on-ground, this correction has to cope with the clustered readout of the detectors (see chapter 3.2) in flight. The stray light at one detector pixel is the sum of stray light originating from wavelengths corresponding to all other detector pixels in a channel (Fig. 8-7), sometimes even from a neighbouring channel. If only part of the channel, i.e. one or more clusters, is read out, the intensity at other wavelengths is not known and must be estimated. For level 0-1b processing, the stray light calculation algorithm is applied after the spectrum covering the complete SCIAMACHY wavelength range has been estimated. Such a spectrum is calculated for each integration time (IT) in the state. After each longest IT, a complete spectrum has been measured. For shorter IT, intensity from missing clusters is estimated by assuming that the spectrum for shorter IT may only have a different intensity but retains the spectral shape of the longest integration time measurement, i.e. a complete spectrum for each IT is estimated by scaling the complete spectrum at the longest IT to the shorter IT. The stray light correction values are stored directly in each MDS of the level 1b product.
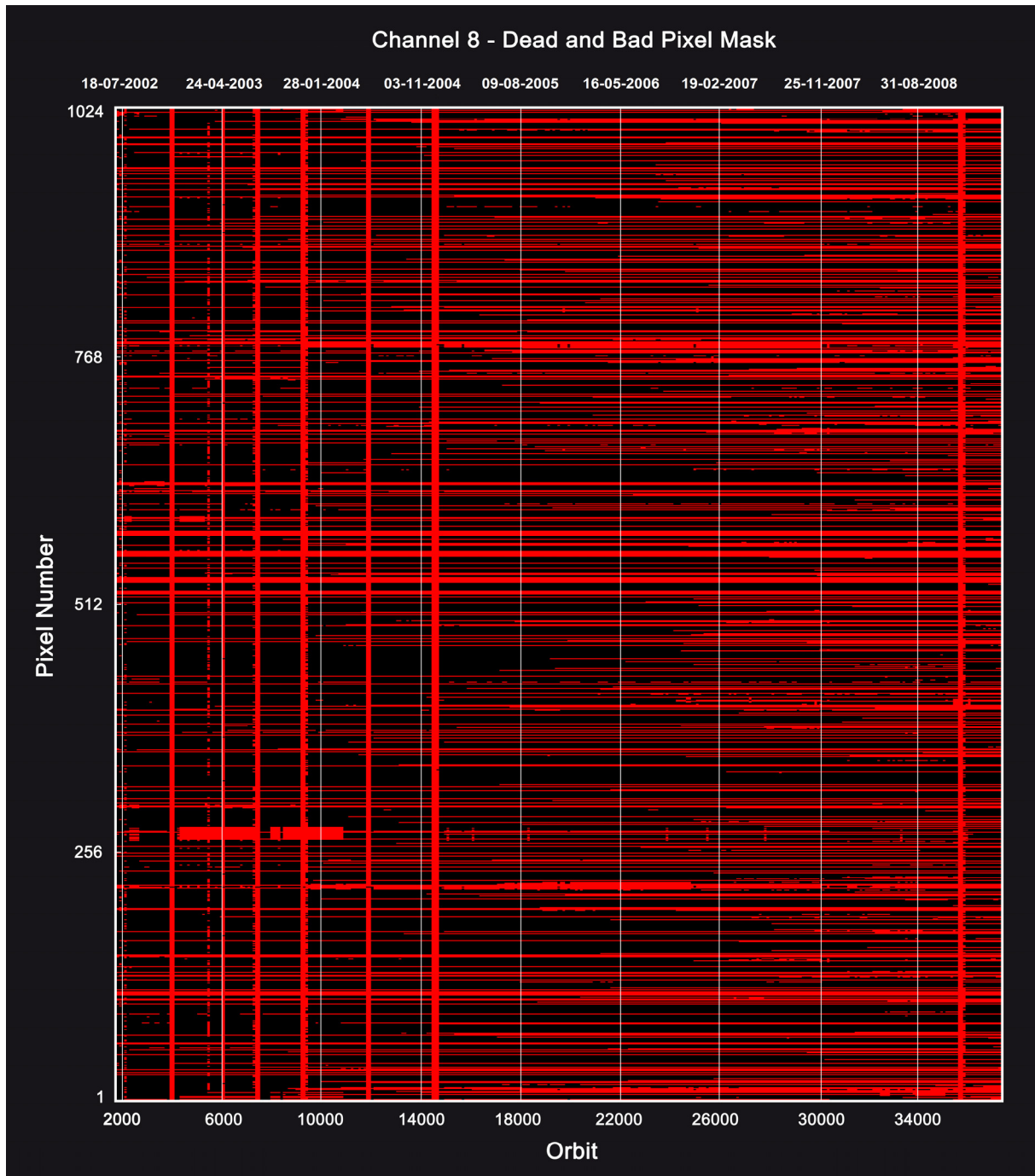
Fig. 8-6: DBPM for channel 8 between August 2002 and March 2009. Each row corresponds to a single pixel. When a pixel turns red it becomes either dead or bad. The vertical stripes correspond to decontamination periods. (Courtesy: DLR-IMF)
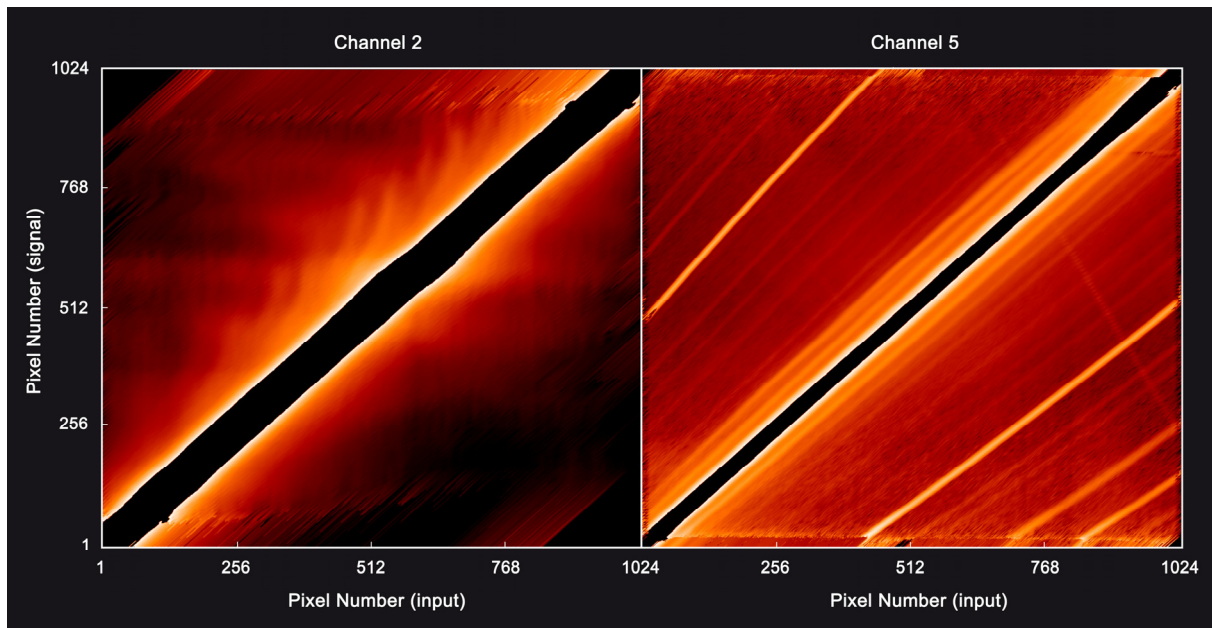
Fig. 8-7: Spectral stray light matrix (1024 × 1024) in channel 2 (left) and channel 5 (right). Each row shows the stray light spectrum (signal) due to unit input in the corresponding pixel (horizontal axis), with bright regions displaying the strongest stray light contribution. The region around the diagonal is defined as being free from stray light. The lines parallel and normal to the diagonal in the plot for channel 5 are reflected and refocused ghosts. These are absent in channel 2, which has a holographic instead of a classically ruled grating. (Courtesy: DLR-IMF)

Wavelength Calibration

The wavelength calibration uses on-ground and in-flight data for channels 1-6. From weekly and monthly in-flight measurements of the spectral line source, the difference to the on-ground calibration is determined. The line positions are derived using the Falk algorithm (Falk 1984). Subsequently, a wavelength for each pixel is stored in the product. Since the spectral line source does not have enough lines in the infrared and dead and bad pixels interfere with the line position determination, the on-ground wavelength calibration is applied for channels 7 and 8 without any changes.

Polarisation

The polarisation correction is required because the detected signal depends on the polarisation state of the incoming light. The instrument's polarisation response was determined on-ground for science and PMD channels using polarised calibration sources and is stored in the calibration Key Data. They contain the ratio of the PMD signal to the science channel signal for several instrument configurations and polarisation directions. As science and PMD channels are performing polarisation sensitive measurements in flight, the elements Q and U of the Stokes vector are derived from the obtained signals. By using the polarisation information from the Key Data, a linear polarisation correction factor is calculated for all wavelengths. A complication arises from the fact that the PMD channels are sampled independently and at a different rate from the channel detectors. Using the time stamps of both types of measurements, the PMD channels are resampled on a time grid synchronised with the detector readout, at twice the BCPS frequency with the BCPS corresponding to the maximum detector data packet generation rate of 62.5 msec. For channels 1-5, the time of detector readout depends on the pixel number, as each pixel is read out sequentially. The synchronisation is roughly performed for the part of the detector corresponding to the mean wavelength of each PMD bandpass. All data necessary to calculate the polarisation correction factors are contained in the level 1b product. The polarisation sensitivities from on-ground measurement are stored in the *POLARISATION SENSITIVITY* ADS for

nadir, limb and occultation while the polarisation fractions and the integrated PMD values for each observation are directly stored in the MDS. Raw PMD measurements – which are not further used in operational processing – are saved in the *PMD* ADS.

Radiometric Correction

The final calibration step concerns the radiometric calibration, i.e. applying the radiometric instrument response to the measured and corrected signals. Globally valid radiance calibration constants are taken from the calibration Key Data measured on-ground. The radiometric sensitivity for each observation is stored directly in the MDS of the level 1b product.

Sun Mean Reference Spectrum

For the purpose of trace gas retrieval in level 1b-2 processing based on reflectances, the level 0-1b processing generates a Sun Mean Reference (SMR) spectrum for each state of solar observations. Absolute irradiance calibration on-ground was only performed for the ESM diffuser mode. The irradiance calibration of all other solar observation modes is based on a rough estimate which leads to an incorrect absolute irradiance level and may leave residual instrument polarisation features in the spectrum. However, test trace gas retrievals showed good results with radiometrically uncalibrated ASM diffuser spectra. Therefore, it was decided to generate two SMR for each diffuser state: one with all calibrations applied and one without radiometric calibration. For monitoring purposes, SMR are also generated for the other solar observations, i.e. occultation and monitoring. All SMR are stored in the *SUN MEAN REFERENCE* ADS.

Degradation

Although correcting the degradation with the m-factors is implemented in the level 0-1b processing, it is not applied operationally at present. However the *SciaL1c* tool provides the capability for applying the radiometric degradation correction and thus producing degradation corrected level 1c data. Because the m-factors impact the science and PMD channels differently, the concept for calculating and providing them pursues a two-step approach. In a simple solution only the readouts of the science channels are corrected which comprises only multiplicative calculations. Thus, their application can be switched 'on' and 'off' as needed. The current baseline is to apply the simple correction during operational level 1b-2 processing for the products requiring radiometrically calibrated data (such as the Absorbing Aerosol Index). The full solution would include correcting the PMD readouts as well, a processor feature which is presently not activated. Since the calculation of m-factors is decoupled from the operational level 0-1b processing, updates of the external m-factor database are possible without interfering with operational PDS requirements. While all auxiliary data necessary to generate a calibrated level 1c product are stored in the level 1b product itself, the m-factors can be accessed via a separate location at

http://www.iup.uni-bremen.de/sciamachy/mfactors/ .

Here, they are generated and maintained as part of the SCIAMACHY optical performance monitoring.

8.3   Operational Level 1b-2 Data Processing

The goal of level 1b-2 data processing is to provide geophysical parameters such as column densities and profiles from atmospheric constituents as well as cloud and aerosol parameters. These data are given in the MDS of the level 2 product in combination with geolocation and additional auxiliary information (state geolocation, quality flagging, etc.) in the appropriate ADS.

The operational level 1b-2 data processing occurs in OL mode and fast delivery mode. Both chains use the same processor, i.e. SCIAMACHY Ground Processor (SGP) L12 OL V5. Since the level 2 products in the fast delivery chain are based on NRT level 1b data, they cannot utilise the most

accurate auxiliary and in-flight calibration data. The operational level 1b-2 product suite is comprised of retrievals from the nadir and limb observation geometry. Operational processing of occultation measurements is currently not foreseen, but the existing system can be extended to cope with it at a later stage. Table 8-1 summarises the level 2 product content.

Spectral Range	Nadir (FD& OL)		Limb (FD&OL)
UV-VIS-NIR	O_3 (vertical column)[1]	HCHO (vertical column)[2]	O_3 (profile)
	NO_2 (vertical column)[1]	cloud cover	NO_2 (profile)
	BrO (vertical column)	cloud optical thickness	BrO (profile)
	SO_2 (vertical column for anthropogenic and volcanic scenario)	cloud top height	
	OClO (slant column)	AAI	
	H_2O (vertical column)	CHOCHO (vertical column)[2]	
SWIR	CO (vertical column)	CH_4 (vertical column)[2]	cloud parameters

1) = constitute METEO products available for the meteorological community
2) = planned for the next processor version (SGP V.6)

Table 8-1: Operational SCIAMACHY level 2 product suite.

Fig. 8-8 (nadir DOAS) and Fig. 8-9 (limb) outline the general processing chain for the level 1b-2 data processors. It is implemented as a multi-processor Linux cluster system. Since scientific progress in retrieval techniques continues, an outline of level 1b-2 algorithms can only provide a snapshot. The reader is referred to the ATBD (Lichtenberg et al. 2010) for up-to-date algorithm information. Particularly experience gained from scientific product generation will be reflected in the operational chain stepwise, i.e. algorithm improvements that are validated will be introduced into the processor. Level 2 products will be generated anew whenever significant algorithm improvements are implemented into the processor. Even beyond the in-orbit mission lifetime, an ambitious atmospheric science project like SCIAMACHY requires continuing retrieval algorithm research and reprocessing.

Column Densities from Nadir Observations

UV-VIS data processing uses the DOAS approach for the retrieval of slant column densities from launch onwards. The DOAS concept, together with the derivation of vertical column densities, is described in chapter 7.2. The operational DOAS algorithm is based on the approach originally implemented for the GOME data processor (GDP, Thomas and Spurr 1999). Since DOAS algorithm development undergoes permanent improvements – currently GDP 4.0 is operational (Spurr et al. 2004, Van Roozendael et al. 2006) – SCIAMACHY's processor SGP L12 also follows these changes (Lerot et al. 2009). For O_3, the GDP 4.0 implementation yields vertical columns which are derived in an iterative way by successively updating the air mass factor calculated by using the vertical column provided by the previous iteration. It replaces the early standard DOAS approach where a distinction between AMF extraction, slant column density derivation, and the vertical column density calculation was made. Other UV-VIS products of the weak absorbing gases can rely on this standard method while the water vapour retrieval is based on a further DOAS variant, the AMC-DOAS approach (Noël et al. 1999, Noël et al. 2004), instead.

In the SWIR range the cross sections of line absorbers strongly depend on pressure, temperature and wavelength. Therefore, retrieval algorithms from the UV-VIS-NIR cannot be applied here. The BIRRA (Beer InfraRed Retrieval Algorithm, Schreier et al. 2009) retrieval scheme, which takes into

account the characteristics of SWIR absorption spectroscopy, handles the retrieval of CO. Further trace gases to be determined from the SWIR range in the near future shall rely on BIRRA, as well. BIRRA performs a nonlinear least squares fit of molecular vertical column densities together with some auxiliary parameters. When fitting the observed near infrared intensity/radiance spectrum, being essentially described by Beer's law with the Sun as source, the molecular absorption features are modelled by means of line-by-line calculations.

Cloud and Aerosol Parameters in the UV-VIS-NIR for Nadir Observations

The parameter cloud fraction is determined by the threshold algorithm OCRA (Loyola 2000). PMD measurements identified to be free of clouds from the red, green, and blue PMD are stored as RGB composite. The distance between the white point in the RGB reflectance space (= fully clouded) and the actually determined PMD reflectance is taken to determine the cloud fraction. Heuristically derived scaling factors allow the appropriate representation of cloud fraction between zero (cloud-free) and one (fully cloudy). In the initial processor versions, cloud top height or cloud top pressure was extracted from the ISCCP (International Satellite Cloud Climatology Project) database but it is replaced from version 3.0 onwards by an algorithm based on evaluation of the O_2 A-band in the visible spectral region (SACURA, Rozanov and Kokhanovsky 2004). This yields improved cloud top height and cloud optical thickness parameters since only one value for optical thickness served as fixed input in the previous implementation.

The Absorbing Aerosol Index (AAI) relies on the ratio of the reflectances between a spectral band which is not covered by spectral absorptions and a spectral band with ozone absorptions. The AAI is calculated from the difference between logarithmic ratios of measured and modelled reflectances. It indicates the presence of absorbing aerosols and depends strongly on an accurate calibration of the reflectance (Tilstra et al. 2007).

Profiles from Limb Observations

Retrieving trace gas profiles from limb measurements is a challenging task. The algorithm currently used for limb retrievals, the Iteratively Regularized Gauss-Newton Method (Doicu et al. 2002, Doicu 2005 – see also chapter 7.6), consists of mainly two parts, a forward model calculation followed by the inversion, using a Tikhonov regularisation with a variable regularisation parameter. In the forward model calculation the sequence of steps includes

- Calculation of the geometrical parameters describing the limb scans
- Calculation of the optical properties for each layer
- Computation of the radiation field in the framework of the independent pixel approximation
- Calculation of the multiple scattering correction factors by means of a spherical radiative transfer model using Picard iteration with open boundaries

The profiles are calculated from a predefined minimum altitude up to 100 km. For each state a maximum of four profiles in East-West direction is determined depending on the integration time of the state. Fig. 8-8 and Fig. 8-9 depict the s/w architecture of the limb retrieval algorithm with two preparative steps before the profile retrieval itself.
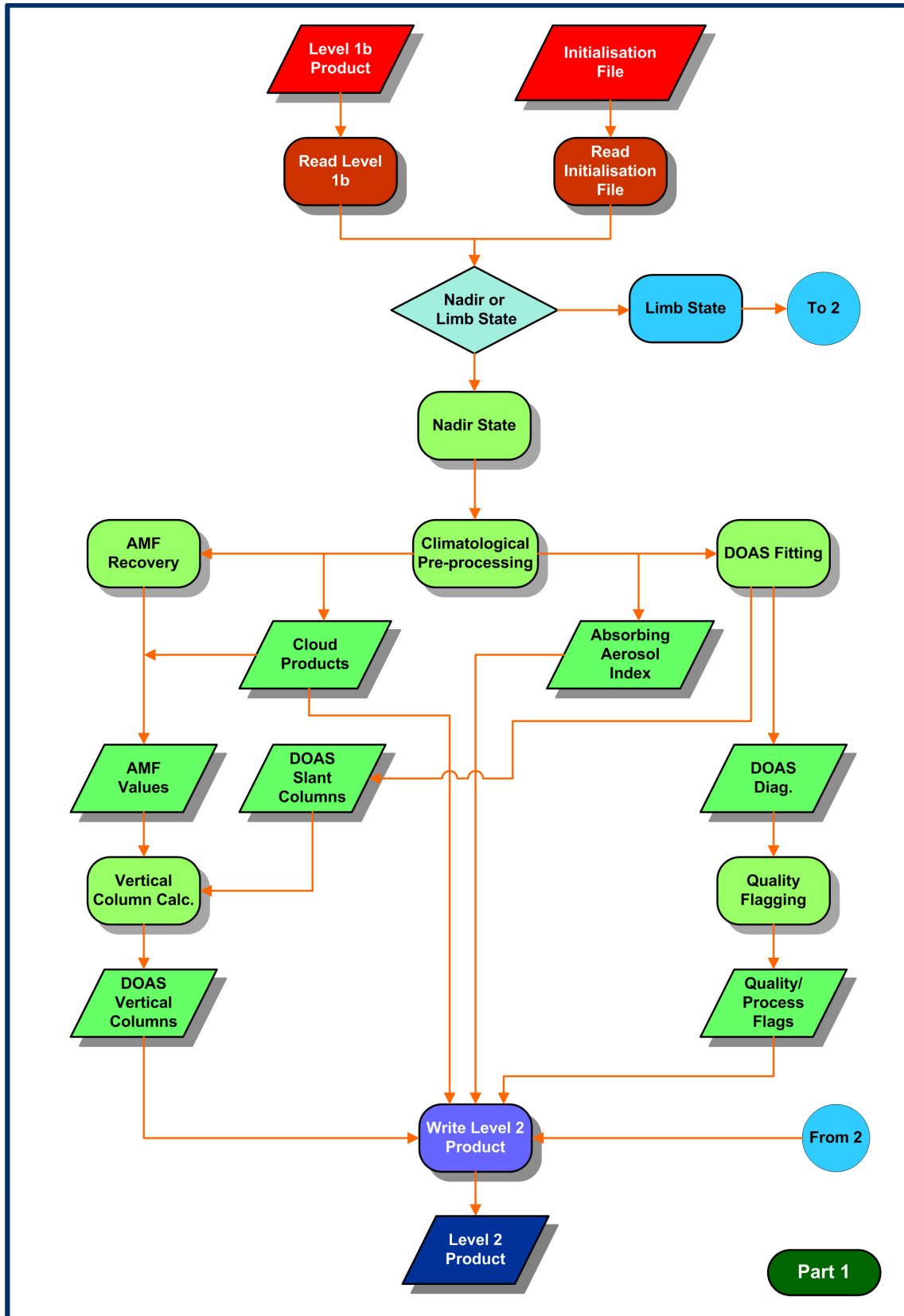
Fig. 8-8: Level 1b-2 data processing flow diagram illustrating the preprocessing and nadir DOAS algorithm. Limb algorithms are continued in Fig. 8-9. (Courtesy: DLR-IMF)
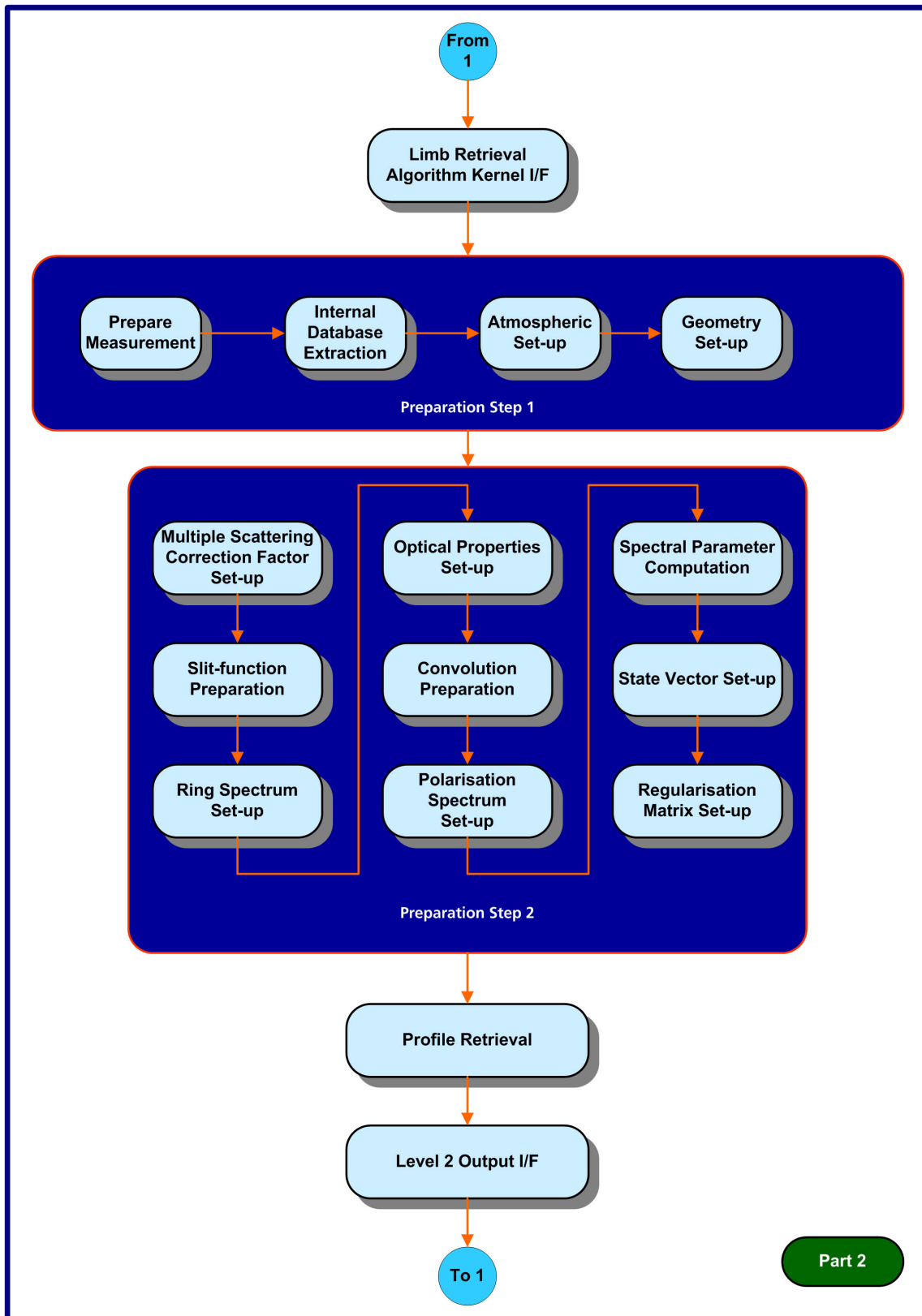
Fig. 8-9: Level 1b-2 data processing flow diagram for limb profile retrieval. At the end of the limb processing, the data flow returns to Fig. 8-8 for product output. (Courtesy: DLR-IMF)

Limb Cloud Detection

In limb mode SCIAMACHY measures light scattered along the line-of-sight. If the line-of-sight intersects a cloud at a certain height, the spectrally resolved measurements differ from cloud-free measurements. To eliminate systematic uncertainties and to enhance the sensitivity of the cloud determination, differences and/or ratios of spectral measurements in a certain wavelength region are exploited. Actually, radiance ratios at 750nm and 1090 nm are used for the detection of clouds. In an ideal Rayleigh scattering atmosphere populated only by molecules, radiances at two wavelengths about 300 nm apart differ considerably, while for larger particles like cloud droplets, this difference is reduced. The cloud top height (CTH) above the limb tangent point is then derived by analysing light from the cloud layer and from above it.

8.4   Scientific and Value-Added Products

Scientific Products

As outlined above, scientific products are a vital service of the mission. Although scientific products usually refer to level 2, scientific level 1c data are also generated from operational level 1b or level 0 data by applying calibration steps differing from what is done in the operational ground segment. Since scientific product generation does not necessarily follow the stringent requirements of the operational chains – although this does not preclude them from being as precise as possible and even permits the utilisation of the most accurate and evolved algorithms – they may be implemented in various ways. Therefore, a generic description of scientific processor architectures similar to the operational ones is not applicable. Table 8-2 presents a listing of geophysical parameters available as scientific products. Details can be found on SCIAVALIG's website at

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

with successive species or parameter dependent links.

Geophysical Parameters			
O_3	vertical column, profile stratosphere and mesosphere	IO	slant column
NO_2	slant/vertical/tropospheric column, profile	H_2O	vertical column
BrO	slant/vertical column, profile	HDO/H_2O	
SO_2	slant/vertical column	CO	vertical column
HCHO	vertical column	CH_4	vertical column
CHOCHO	vertical column	CO_2	vertical column
OClO	slant column, profile	$T_{mesopause}$	mesopause temperature
Clouds	cover, pressure, top height, optical thickness, liquid water path, phase index, top-of-atmosphere (TOA) reflectance, droplet effective radius, PSC index, NLC index		
Aerosol	AAI, AOT		
UV	index, dose		

Table 8-2: Scientific SCIAMACHY level 2 product suite (due to the nature of scientific products, the table is evolving and does not claim to be complete).

Value-Added Products

The value-added (VA) products are generated on a routine basis for various applications. Requirements for value-adding originate from very different user groups. Whereas scientific applications – if not based on level 2 data – usually need data sets on regular grids with complete information on pixelwise data quality, applications for governmental monitoring duties, public services or education purposes ask for synoptic maps, i.e. overviews at one point in time, on different scales without any gaps and for long time series of continuous observations. In some cases, NRT data provision is also requested to initialise forecast services.

VA products are based on the operational products generated within the PDS or on scientific products. In most cases value adding transforms level 2 products from their satellite projection and the time of the satellite overpasses to integrated datasets (level 3 and 4) which contain a geophysical parameter as observed by one instrument. Averaging level 2 data over a certain period and grid cell generates level 3 products yielding daily, weekly, monthly or annual mean datasets. Level 4 products are based on the combination of satellite observations (level 2) and atmospheric models such as e.g. meteorological circulation and chemistry. In addition, synergistic value-adding is conducted by either combining level 2 products of several sensors into one comprehensive level 3 and 4 data set or by exploiting level 1 products (top-of-atmosphere radiances or spectra) from different sensors to retrieve a level 2 product with new information.

A wide set of requirements on atmosphere-related monitoring and forecasting are targeted within the ESA GMES service element project MACC, previously PROMOTE, where atmospheric satellite observations are exploited. Topics are e.g. distribution and trends of the stratospheric ozone layer, UV irradiance at the surface, near-surface air quality, greenhouse gases and climate. Application fields range from global and European protocol monitoring (Montreal, Kyoto, CLRTAP – Convention on Long-range Transboundary Air Pollution) to public information services such as skin exposure sunburn time warning, street level air quality forecasts or flight route warning of dust/volcanic ash events. Further currently developed applications include services to support planning and management of solar power plants (ESA market development project ENVISOLAR – Environmental Information Services for Solar Energy Industries) or to improve the quality of satellite land images by providing an automatic atmospheric correction with actual information on the state of the atmosphere. In all these application areas, quality information is essential to improve confidence in the VA products. In the future, the use of value-adding will most likely increase as new requirements call for integration of all observations – satellite, aircraft, balloons, and ground – together with models to extract the best possible and comprehensive knowledge on the state of the atmosphere. For more details the reader is referred to the IGACO (Integrated Global Atmospheric Chemistry Observations) theme strategy.

One key tool for achieving the goals described above is the use of data assimilation techniques which provide a mathematically consistent technology to integrate measurements and models together with their statistical errors. Data assimilation generates synoptic trace gas fields from asynoptic spaceborne measurements. It enables the derivation of interpolated concentration fields as well as information about transport mechanisms.

A major application of assimilation schemes to SCIAMACHY data concern the combination of stratospheric modelling and nadir column data. This is ideally suited for applications such as scientific studies of the evolution of the ozone layer and of special events (e.g. ozone hole or low ozone episodes) or inter-comparisons with models (e.g. study of dynamical and chemical processes). Since assimilated fields are globally available, comparison with independent observations can be performed without space/time mismatches. Stratospheric O_3 is also of particular interest as it can be assimilated into operational weather forecasts, thus improving the model representation of stratospheric wind fields and thereby the quality of the forecast. Since forecast services are provided on short timescales, the fast availability of O_3 data – via the operational fast delivery service or NRT scientific products – is essential (Eskes et al. 2005). Such ozone columns serve as an input for the assimilation analysis and a subsequent forecast of how the stratospheric ozone layer will develop in the upcoming nine days. The assimilation yields a complete picture of the global ozone distribution (Fig. 8-10). Within the error margins of both model and observations, it is consistent with observations and our knowledge of atmospheric transport and chemistry (Eskes et al. 2002, Eskes et al. 2003). SCIAMACHY ozone

columns are currently also assimilated operationally in the numerical weather prediction model of the ECMWF. Several centres use the ozone forecasts for UV radiation predictions.
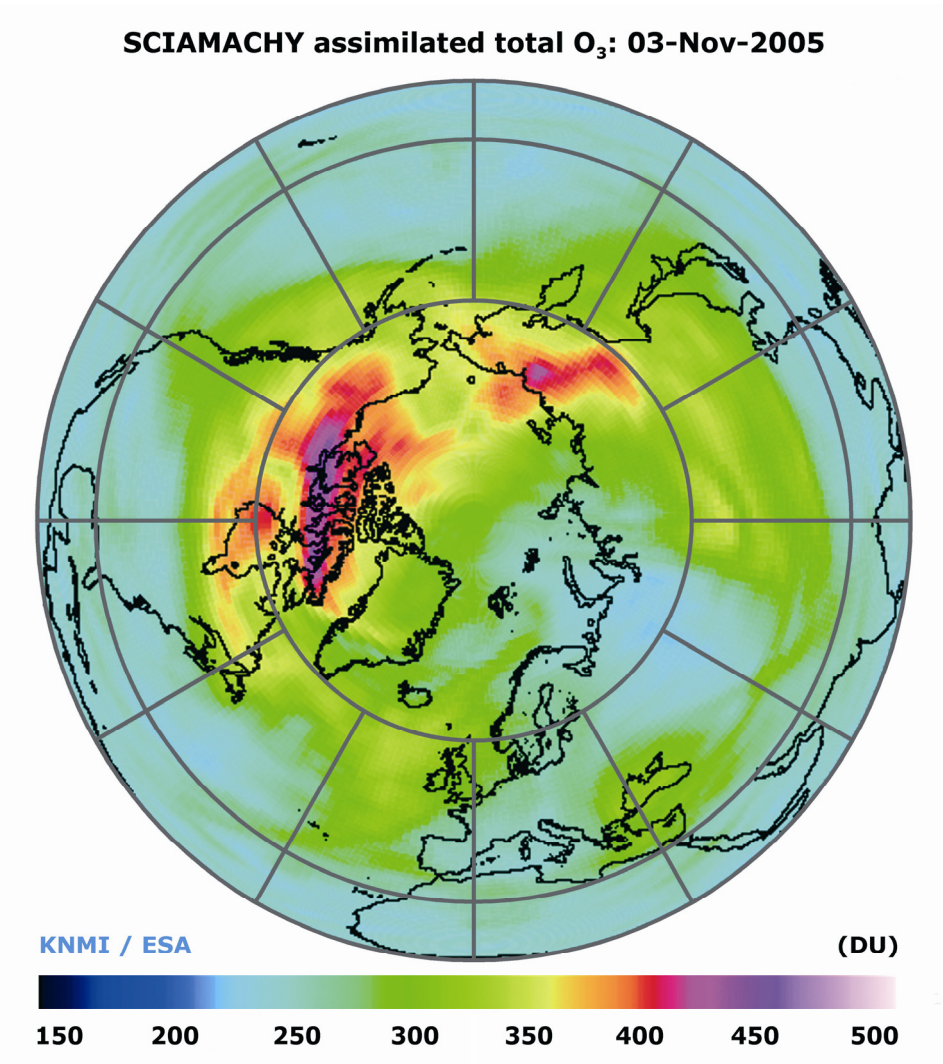

Fig. 8-10: A forecasted North Pole view of the assimilated total ozone column field for 3 November 2005 at 12:00 UTC based on SCIAMACHY data. (Courtesy: KNMI/ESA)

References

Doicu, A., Schreier, F. and Hess, M. 2002. Iteratively regularized Gauss-Newton method for atmospheric remote sensing. *Comp. Phys. Comm.*, 148, 214-226.

Doicu, A. 2005. SCIAMACHY Level 1b to 2 OFL Processing Algorithm Change Description Document Limb Profile Retrieval Algorithm (ENV-TN-DLR-SCIA-0062). *Technical Document*, DLR.

Eskes, H.J., van Velthoven, P.F.J. and Kelder, H.M. 2002. Global ozone forecasting based on ERS-2 GOME observations. *Atmos. Chem. Phys.*, 2, 271-278.

Eskes, H. J., van Velthoven, P.F.J., Valks, P.J.M. and Kelder, H.M. 2003. Assimilation of GOME total ozone satellite observations in a three-dimensional tracer transport model. *Quart. J. Roy. Meteorol. Soc.*, 129, 1663-1681.

Eskes, H.J., van der A, R.J., Brinksma, E.J., Veefkind, J.P., de Haan, J.F. and Valks, P.J.M. 2005. Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on ENVISAT. *Atmos. Chem. Phys. Discuss.*, 5, 4429-4475.

Falk, W.R. 1984. Data Reduction from Experimental Histograms. *Nuclear Instruments and Methods in Physics Research*, 220, 473-478.

Kleipool, Q.L., Jongma, R.T., Gloudemans, A.M.S., Schrijver, H., Lichtenberg, G.F., van Hees, R. M., Maurellis, A.N. and Hoogeveen, R.W.M. 2007. In-flight proton-induced radiation damage to SCIAMACHY's extended-wavelength InGaAs near-infrared detectors. *Infrared Phys. & Techn.*, 50, 30-37, doi:10.1016/j.infrared.2006.08.001.

Lerot, C., Van Roozendael, M., van Geffen, J., van Gent, J., Fayt, C., Spurr, R., Lichtenberg, G. and von Bargen A., 2009. Six years of total ozone column measurements from SCIAMACHY nadir observations. *Atmos. Meas. Tech.* 2, 87-98.

Lichtenberg, G., Bovensmann, H., Van Roozendael, M., Doicu, A., Eichmann, K.-U., Hess, M., Hrechanyy, S., Kokhanovsky, A., Lerot, C., Noël, S., Richter, A., Rozanov, A., Schreier, F. and Tilstra, L.G. 2010. SCIAMACHY Offline Level 1b-2 Processor Algorithm Theoretical Baseline Document (ENV-ATB-QWG-SCIA-0085). *Technical Document*, SQWG/DLR-IMF.
available at http://atmos.caf.dlr.de/sciamachy/documentation.html

Loyola, D. 2000. Cloud retrieval for SCIAMACHY, Proceedings of ERS-ENVISAT symposium, Gothenburg, Sweden, ESA-SP-641

Noël, S., Buchwitz, M., Bovensmann, H., Hoogen, R. and Burrows, J.P. 1999. Atmospheric Water Vapor Amounts Retrieved from GOME Satellite Data. *Geophys. Res. Lett.*, 26(13), 1841-1844.

Noël, S., Buchwitz, M. and Burrows, J.P. 2004. First retrieval of global water vapour column amounts from SCIAMACHY measurements. *Atmos. Chem. Phys*. 4, 111-125.

Rozanov, V.V. and Kokhanovsky, A.A. 2004. The semi-analytical cloud retrieval algorithm as applied to the cloud top altitude and the cloud geometrical thickness determination from the top of atmosphere reflectance measurements in the oxygen absorption bands. *J. Geophys. Res.*, 109, D05202, doi: 10.1029/2003JD004104.

Schreier, F., Gimeno-Garcia, S., Hess, M., Doicu, A. and Lichtenberg, G. 2009. Carbon monoxide vertical column density retrieval from SCIAMACHY infrared nadir observations. In IRS 2008:

Current Problems in Atmospheric Radiation, eds. T. Nakajiama and M.A. Yamasoe, *American Institute of Physics*, doi: 10.1063/1.3116983.

Slijkhuis, S. 2004. ENVISAT-1 SCIAMACHY Level 0 to 1c Processing, Algorithm technical basis document (ENV-ATB-DLR-SCIA-0041). *Technical Document*, DLR.
available at http://atmos.caf.dlr.de/sciamachy/documentation.html

Spurr, R.J.D, Van Roozendael, M. and Loyola, D.G. 2004. Algorithm theoretical basis document for GOME total column densities of ozone and nitrogen dioxide (ERSE-DTEX-EOPG-TN-04-0007). *Technical Document*.
available at http://atmos.caf.dlr.de/gome/documentation.html

Thomas, W. and Spurr, R.J.D. 1999. GOME Level 1 to 2 Algorithms description (ER-TN-DLR-GO-0025). *Technical Document*, DLR.

Tilstra, L.G., De Graaf, M., Noël, S., Aben, I. and Stammes, P. 2007. Sciamachy's absorbing aerosol index and the consequences of instrument degradation. *Proc. 3rd Workshop on the Atmospheric Chemistry Validation of ENVISAT (ACVE-3),* Frascati, Italy, ESA-SP-642.

Van Roozendael, M., Spurr, R.J.P., Lerot, C., Loyola, D., Schröder, T. and von Bargen, A. 2006. SCIAMACHY Level 1b-2 Off-line Data Processing: Algorithm theoretical baseline document for trace gas retrieval from UV/VIS nadir spectra (ENV-ATB-BIR-SCIA-0074), *Technical Document*, BIRA-IASB