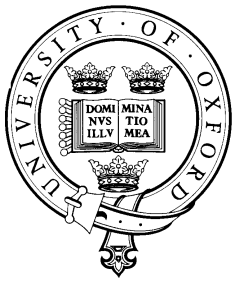
GRAPE Output Products Version 3.2

Andrew Sayer, Caroline Poulsen, Don Grainger
June 2010



Science & Technology Facilities Council

Contents

Abstract

This document serves as an overview of the data products available from version 3 of the Global Retrieval of ATSR Cloud Parameters and Evaluation (GRAPE) project. It describes the content of Level 2 (orbit-based) data products, and provides examples of what the user should expect to see when mapping the data. Finally there is an explanation of the quality control information supplied with the data, and a discussion of known issues and the thresholds which users are advised to apply to the data before use in scientific applications.

1 Introduction

The GRAPE[1] processor, using the ORAC[2] retrieval engine, produces two types of scientific data sets. The first is known as the Level 2 output product (Level 1 being the calibrated radiance and albedo input data) and is generated for the measurements from a single orbit, providing the cloud (and aerosol) retrieval results for data points with unique temporal and spatial locations. The second product is known as the Level 3 output, which is derived from the Level 2 dataset, and provides a global climatology of the retrieved cloud parameters on a monthly mean basis. The Level 3 format is currently being revised into a format matching those of other satellite cloud climatologies taking part in the GEWEX cloud assessment[3]. This will be made available shortly.

1.1 Version 3.X history

This section describes the history of the products in the Version 3 collection. The dates indicate the date at which the products were uploaded to the BADC server; if in doubt as to which version you have, examine the date of product acquisition or contact the authors (see Section 5). The BADC only holds the most recent version of the product.

- Version 3.0, February 2009. Initial processing using the Version 3 algorithm.
- Version 3.1, October 2009. Bug identified and fix in Ice Water Path calculations (assumed density was incorrect). All products regenerated. Time series extended to end of ATSR-2 available record (June 2003), and AATSR products from July 2002-December 2003 included.
- Version 3.2, May 2010. Time series extended with AATSR to the end of 2009. Retrieval quality flag reinstated, and data comments updated to indicate data version. All products regenerated.

1.2 Dataset coverage

The Version 3 GRAPE products process the ATSR-2 daylight data from June 1995 to June 2003. Note that January 1996 to June 1996 are missing due to instrument problems during this periods. Data from the successor instrument AATSR (on ENVISAT) are also included from the beginning of the available record (July 2002) through to the end of 2009.

[1] Global Retrieval of ATSR (Along Track Scanning Radiometer) Cloud Parameters and Evaluation
[2] Oxford and RAL (Rutherford Appleton Laboratory) Aerosols and Clouds
[3] See http://climserv.ipsl.polytechnique.fr/gewexca/ for details and data.

1.3 File type and naming convention

The general naming convention of the GRAPE products follows the structure:

`project-instrument_location_date+time_level+version_number.filetype`

The output products are stored in the HDF5 file format.

1.4 Known bugs

The following bug has been identified in the Level 2 products. The 0.67 μm surface albedo is written where the 0.55 μm surface albedo should be. The 0.55 μm surface albedo is consequently missing from Level 2 files.

2 Level 2 - Orbital products

An example name is:

`grape-atsr2_25000_200001311346_l2v03.hdf`

The Level 2 product is produced from the visible and infrared measurements for each ATSR-2 orbit. The 'location' part of the filename indicates the absolute orbit number. The Level 2 data can be thought of as three parts:

1. The metadata (or header) information
2. The geographic location data
3. The ORAC scientific data.

2.1 Metadata

Variable name	Data type
Title	String, length=70
Data version	32-bit floating point
Institution	String, length=70
Source	String, length=70
History	String, length=70
Comments	String, length=70
References	String, length=70
Datetime	String, length=24
Super-pixel size X	String, length=3
Super-pixel size Y	String, length=3
Image size	32-bit integer (array dimension 2)
Missing value	32-bit floating point

The image size gives the dimensions of the regularly gridded data, i.e. the number of along track and across track super-pixels. Any pixels with missing data (or where the data is not appropriate) will be set to the missing value, typically -999.

2.2 Geolocation

As the ATSR-2 instrument is an imaging radiometer, data is recorded along track in the nadir by scanning a 575 pixel swath across-track (which is approximately 500 km wide). For reasons of processing speed the ATSR data is down-sampled to create a grid of super-pixels. The GRAPE processing uses a 4 by 3 super-pixel (4 pixels along track and 3 pixels across track) so the input data becomes 191 super-pixels wide across-track. For a typical orbit the day data gives approximately 4000 super-pixels along track. The image size is given in the metadata. The geolocation data for an orbit comprises of two arrays:

Variable Name	Description	Dimensions	Valid range
Latitude	Latitude of the centre of each super-pixel	regularly gridded: $\sim$ 4000 by 191	-90 to 90
Longitude	Longitude of the centre of each super-pixel	regularly gridded: $\sim$ 4000 by 191	-180 to 180

2.3 ORAC scientific data

The ORAC Scientific data itself has two parts:

1. The original input data from the ATSR-2 measurements and ancillary information

2. The output data from the ORAC processing.

For each part the variables stored in the Level 2 product are now described in tabular form. Each of the variables in the tables below are arrays with size given by the image size in the metadata and the dimensions are generally 191 by ~ 4000.

2.3.1 Input data

The following input data is regularly gridded such that each variable is an array of size given by the image size. Each array element corresponds to a single super-pixel.

A. Observed reflectances and brightness temperatures

The visible channel reflectances are expressed as sun-normalised reflectances (expressed as a percentage). Note all measurements correspond to ATSR-2's nadir view.

Variable name	Units	Valid range
0.55 μm reflectance	[%]	0-100
0.55 μm reflectance variance		0-1000
0.67 μm reflectance	[%]	0-100
0.67 μm reflectance variance		0-1000
0.87 μm reflectance	[%]	0-100
0.87 μm reflectance variance		0-1000
1.6 μm reflectance	[%]	0-100
1.6 μm reflectance variance		0-1000
3.7 μm brightness temperature	[K]	positive real
3.7 μm BT variance	[K]	positive real
11.0 μm brightness temperature	[K]	positive real
11.0 μm BT variance	[K]	positive real
12.0 μm brightness temperature	[K]	positive real
12.0 μm BT variance	[K]	positive real

B. Observation angles and time

Again, these arrays pertain to the nadir-viewing geometry of ATSR-2.

Variable name	Units	Valid range
Solar zenith angle	degrees	0-90
Satellite zenith angle	degrees	0-90
Relative azimuth angle	degrees	0-180
Julian time	days since Jan 1 1990 (from noon)	positive real

2.3.2 Output data

The dimensions of the output data arrays match those of the input, such that there is one element for each retrieval scene (superpixel).

A. Quality control information

Variable name	Description	Valid range
Flag	Flag to indicate the type of data	1=cloud 2=aerosol 3=no data
Total cost	Cost indication of forward model fit to measurements given the solution	positive real
A priori cost	A priori contribution to the total cost	positive real
Iterations	Number of iterations the retrieval took to converge	-1-25
Retrieval quality flag	Flag to be used as indication of retrieval quality 0: Failed to converge, or very poor fit (cost>100) 1: Poor fit (30<cost<=100) 2: Moderate fit (10<cost<=30) 3: High quality fit (cost<=10)	0-3

All quality control information is provided to users, as appropriate thresholds may depend on the desired application of the data. For general use, it is recommended to check for the quality flag being 3 to ensure only the highest quality data are used. To improve coverage, data with a quality of 2 may be used. It is generally not recommended to use data with a quality flag of 1, and 0 should be avoided as it indicates no satisfactory fit could be found.

B. Retrieved properties

With the exception of phase, the arrays contain data when a cloud retrieval was attempted. When there was no cloud (and an aerosol retrieval was attempted), they contain the missing data value. For aerosol retrievals the cloud phase/aerosol type array contains an indicator of the aerosol model used in the retrieval (based upon a global map of typical aerosol type). Note that for aerosol this is a prescribed, not a retrieved, quantity.

The cloud water path (both liquid and ice) is presented as the water path *for the cloudy portion of the pixel*. To calculate the water path of the pixel as a whole, it should be mulitplied by the retrieved cloud fraction.

As of the current version of GRAPE (v03 in filenames), the following bug has been identified in the level 2 products. The 0.67 μm surface albedo is written where the 0.55 μm surface albedo should be. The 0.55 μm surface albedo is consequently missing.

For the phase variable, values of 3, 5, 7, and 9 existed in previous versions of the code corresponding to other aerosol models. They are currently unused although the numbering system has not been changed to maintain consistency.

Variable name	Description	Valid range
Cloud optical depth	$\log_{10}$ optical depth referenced to 0.55μm	0 to 2.4
Cloud effective radius	Effective particle radius (given particle shape assumption, i.e. sphere for water and hexagonal column for ice)	0 to 100 [μm]
Cloud top pressure	Pressure at cloud top	0 to 1000 [hPa]
Cloud fraction	Fraction of super-pixel with cloud	0 to 1
Surface temperature	Temperature of the ground	20 to 350 [K]
Cloud top height	Geometric cloud top height	0 to 20 [km]
Cloud top temperature	Temperature at cloud top	20 to 350 [K]
Cloud/aerosol phase	Cloud water phase or aerosol type	Cloud: 1-water/liquid; 2-ice/solid Aerosol: 4-Continental average; 6-Desert; 8-Maritime clean; 10-Arctic; 11-Antarctic
Cloud water path	Cloud water path	0 to $\sim$ 16,000 [g/m^2]

The arrays described in the table below contain data when an aerosol retrieval was attempted. When there was cloud, they contain the missing data value.

Variable name	Description	Valid range
Aerosol optical depth	$\log_{10}$ optical depth referenced to 0.55 μm	-2 to 0.301
Aerosol effective radius	$\log_{10}$ effective particle radius (given spherical particle shape assumption),	-2 to 1 [$\log_{10}$ μm]
Surface albedo (CURRENTLY INCORRECT)	Retrieved white-sky surface albedo referenced to 0.55 μm channel	0 to 1

C. Error on retrieved properties

Analagous to the retrieved data arrays are the arrays of uncertainty estimates on the retrieved state.

Variable name	Valid range
Error on $\log_{10}$ cloud optical depth	Real
Error on cloud effective radius	Positive real [μm]
Error on cloud top pressure	Positive real [hPa]
Error on cloud fraction	0 to 1
Error on surface temperature	Positive real [K]
Error on cloud top height	Positive real [km]
Error on cloud top temperature	Positive real [K]
Error on cloud liquid water path	Positive real [g/m^2]

Variable name	Valid range
Error on $\log_{10}$ aerosol optical depth	Real
Error on $\log_{10}$ aerosol effective radius	Real [$\log_{10}$ μm]
Error on surface albedo at 0.55 μm	Positive real

Due to the bug in the recorded surface albedo value, the error on the retrieved white-sky surface albedo should not be used with the current version of the dataset.

D. Surface albedos

Where an aerosol retrieval has been attempted, the surface albedos are derived from the retrieved surface albedo at 0.55 μm output from ORAC. Where a cloud retrieval is attempted the surface albedos are those which have been input to ORAC as they are necessary for the forward model. The input albedos used are obtained from the MODIS albedo products. The surface albedos are not available for super-pixels where no retrieval is attempted, and in these cases are set to the appropriate missing data value.

Note that all of the albedos presented in GRAPE are the white-sky (Lambertian) surface albedo and hence have no geometric dependence.

Variable name	Valid range
Surface albedo at 0.67 μm	0 to 1.0
Surface albedo at 0.87 μm	0 to 1.0
Surface albedo at 1.6 μm	0 to 1.0

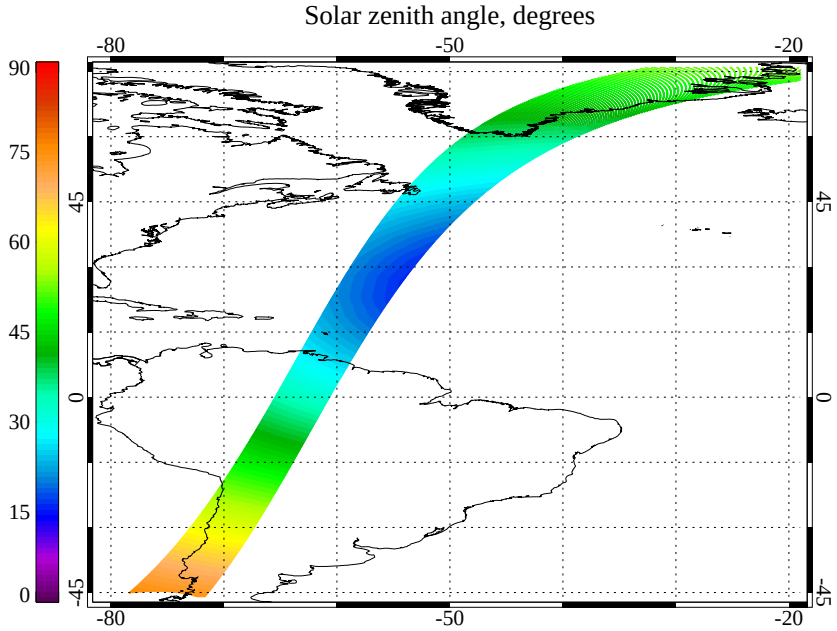
3 Reading GRAPE Products

As the Level 2 products are stored as HDF5 files, they can be accessed through HDF5 library subroutine and function calls, which can be downloaded from the HDF5 web site[4]. The meta-data is stored as attributes of the scientific data arrays. IDL code to read GRAPE files into data structures can be downloaded from the GRAPE website at:

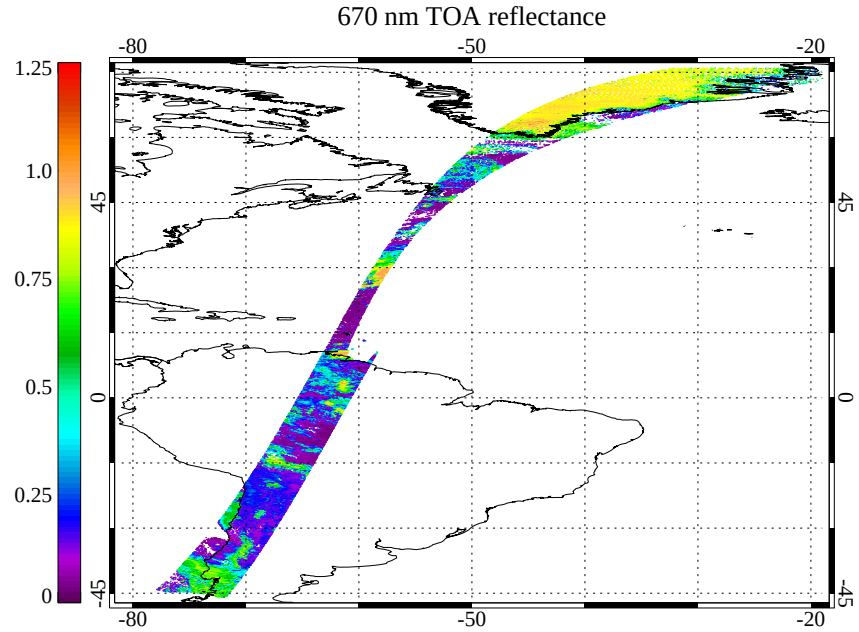
`http://www.atm.ox.ac.uk/project/grape/code/code_con.html`

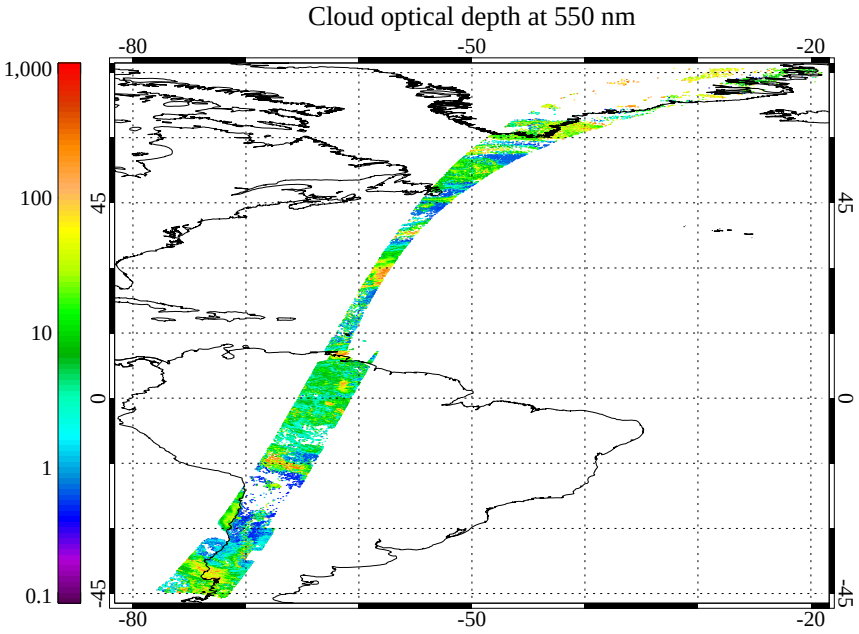
3.1 Example Level 2 products

This section contains some images of Level 2 data products, as a guide to what the user should expect to see. All are taken from Orbit 995 on June 29, 1995. The data for retrieved quantities are filtered to exclude failed retrievals. The images show the solar zenith angle, measured TOA reflectance at 670 nm, retrieved cloud optical depth, uncertainty on retrieved cloud optical depth, retrieval cost, and cloud phase/aerosol type.

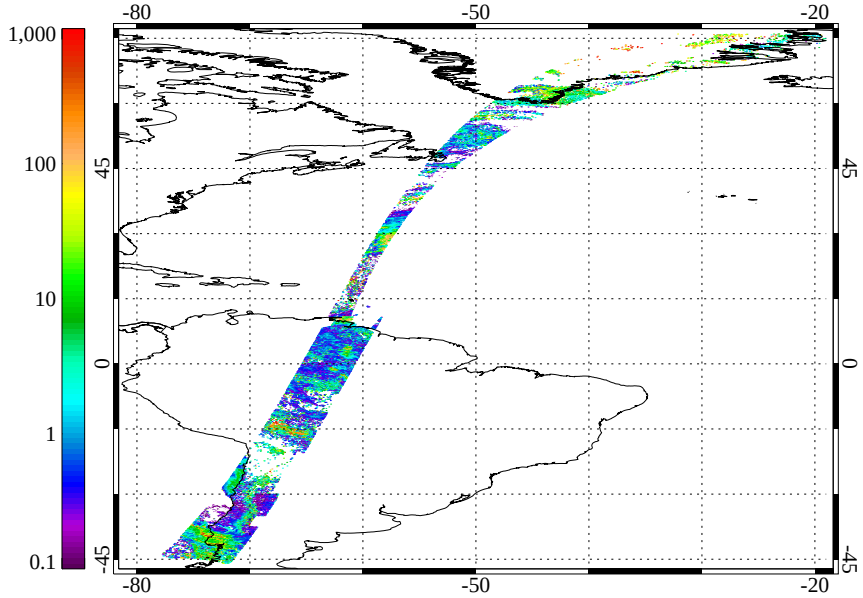


[4]See http://www.hdfgroup.org/HDF5/

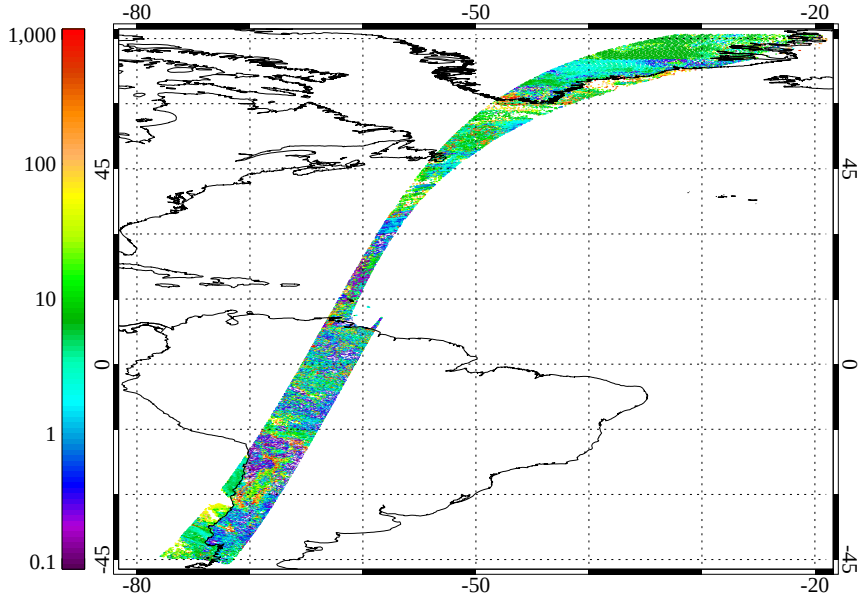

670 nm TOA reflectance
-80
-50
-20
1.25
1.0
0.75
0.5
0.25
0
45
0
-45

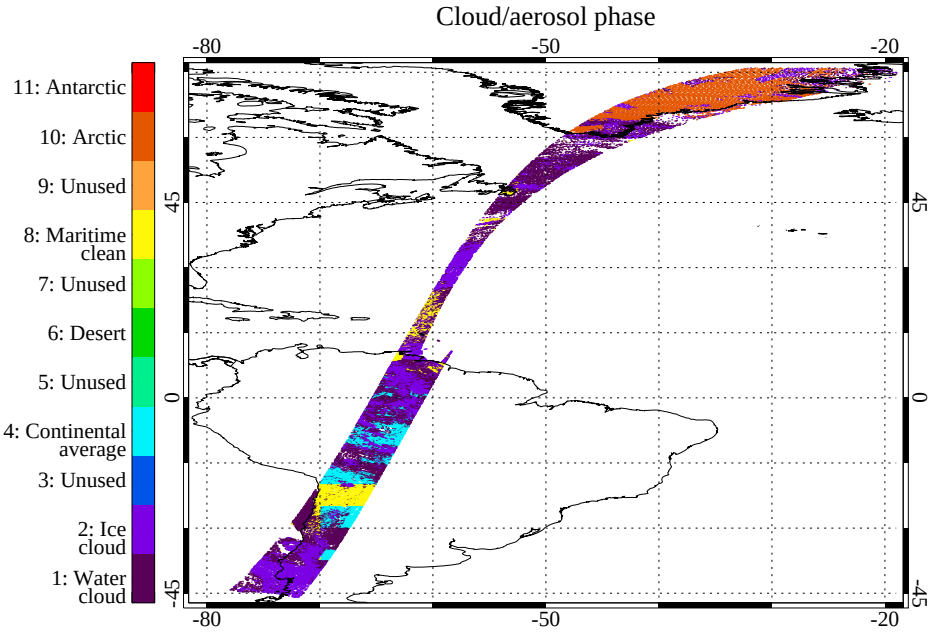

Cloud optical depth at 550 nm
-80
-50
-20
1,000
100
10
1
0.1
45
0
-45

Uncertainty on cloud optical depth at 550 nm



Retrieval cost





4 Quality control on retrieved data

A strength of the optimal estimation framework used in the ORAC retrieval scheme is that it can provide information on the quality of each retrieval fit (the 'cost') and the uncertainty on each retrieved parameter. This arises given knowledge of the uncertainty on the measurements, *a priori* information and forward model as well as the gradient of the forward model in state space.

Attempting to use the GRAPE dataset without applying some quality control is strongly discouraged. In some cases the retrieval algorithm is not able to describe accurately the measurements taken by ATSR-2, which may lead to artefacts in the retrieved data products.

4.1 Guidelines and known performance issues

All quality control information is provided to users, as appropriate thresholds may depend on the desired application of the data. For general use, it is recommended to check for the quality flag being 3 to ensure only the highest quality data are used. To improve coverage, data with a quality of 2 may be used. It is generally not recommended to use data with a quality flag of 1, and 0 should be avoided as it indicates no satisfactory fit could be found. Several considerations should be made when making use of Level 2 GRAPE data products:

- The first and most important quality control setting which should be applied is to check for convergence of data (iterations between 1 and 24 for aerosol or 1 and 14 for cloud).

A failure to converge indicates that no solution was reached by the retrieval and so the final state vector is unlikely to be the correct one.

- The second parameter which should be considered is the retrieval cost; this is essentially a χ^2 statistic of the fit of the retrieval solution to the measurements and *a priori* data. It is suggested that for a reasonably inclusive cut of data an appropriate cost threshold is 5 for aerosol retrievals and 10 for cloud. For a less restrictive cut cut to increase coverage, a threshold of 30 is appropriate, although this will include more cases where the retrieval failed to find a high-quality solution (typically multi-layer clouds).

The above are taken into account when quality flags are assigned. Further points to consider are listed below:

- The ERS-2 platform on which ATSR-2 flies suffered a partial failure in January 2001, which is thought to have negatively impacted cloud flagging over land. This is under investigation by the ATSR quality working group. For the moment users are advised to be aware of the potential for reduced quality of ATSR-2 retrievals from this period onwards. This does not affect ATSR-2 before this data, or AATSR at all.

- Cloud retrievals over snow and sea ice are difficult due to the very bright surfaces. Additionally, retrieval of cloud properties is difficult for broken cloud fields. Users are advised to exercise caution over snow/ice surfaces, or for scenes where the cloud fraction is less than 1.

- Retrievals where state variables lie on the limits of the permitted range should also be treated with caution, as these may indicate the scheme attempting to retrieve an unphysical value. An example of this is unflagged cloud contamination in aerosol retrievals, which can typically lead to an apparant aerosol optical depth of the maximum permitted value. However factors such as the retrieval cost should be considered, as in some cases extreme values like this may in fact be the true solution.

- The optimal estimation framework provides uncertainty estimates on each retrieved quantity. If the measurements are only weakly sensitive to a particular parameter in a particular situation then this may result in the resulting uncertainty being large compared to the retrieved value. Although not necessarily an indicator of a bad retrieval, this can show where the state is not well-defined by the measurements and retrievals with poor precision may be unsuitable for some applications.

- As the GRAPE algorithm currently makes use of only the nadir view of ATSR-2, the aerosol retrieval is comparatively weaker over the land as compared to the sea. This is

because with a single-view there can be insufficient information to decouple the surface and atmospheric contributions to the observed TOA reflectance unambiguously, as the contribution from the surface is large. Users interested in aerosol properties are advised to use data from the GlobAerosol project[5], which provides a global aerosol record from 1995 to 2007, including a dual-view version of the ORAC algorithm applied to ATSR-2 and AATSR data which performs better over both land and sea.

- The cloud retrieval currently assumes a single-layer plane-parallel atmosphere. Hence multilayer cloud systems are not well-modelled by ORAC, which can lead to retrievals failing, converging with a higher-than-expected cost, or retrieving some 'merged' properties (such as a mid-level cloud-top pressure when in reality there is a thinner high cloud above a thicker low cloud).

5 Contact

For enquiries related to GRAPE data, or comments on this document, please contact either:

- Andy Sayer: *sayer@atm.ox.ac.uk*
- Caroline Poulsen: *caroline.poulsen@stfc.ac.uk*
- Don Grainger (Project PI): *grainger@atm.ox.ac.uk*

[5]See http://www.globaerosol.info