Aeronomy Level 2 data products

F. Merino, P. Baron
Stockholm University Stockholm SE-106 91 Sweden

P. Eriksson, D. Murtagh, C. Jimenez
Department of Radio and Space Science, Chalmers University of Technology

E. LeFlochmoen, J. Urban, N. Lautie, P. Ricaud, J. de la Noe
Bordeaux Observatory, CNRS/INSU, BP 89, 3370 Floirac

M. Olberg
Onsala Space Observatory, Onsala, SE-43992 Sweden

Version 1.2 – 24 February 2003

The following outlines the format and storage of the various data products from the Level 2 processing software. Level 2 is basically any data product derived from Level 1b. Data in a Level 2 file is stored by orbit.

SMR Aeronomy Level 2 data will be stored using the HDF-EOS 4 Point object data format. HDF-EOS is an extension of NCSA (National Center for Supercomputing Applications) HDF (Hierarchical Data Format) selected by NASA as the baseline standard for EOS (Earth Observing System).

The version of the HDF-EOS library used is 2.5. For more information download the documents "HDF-EOS Library User's Guide for the ECS Project, Volume 1: Overview and Examples; Volume 2: Function Reference Guide". This document is available on the web at http://hdfeos.gsfc.nasa.gov.

Scientific product file name convention

The official product from MISU will have the filename,

SMR_XXXX_MPPPP_VVJ.L2P

- SMR means sub-millimeter instrument[1].
- XXXX is a four digit number in GHz representing the center frequency of the band that was used in the retrieval. e.g. 5014 is read as 501.4 GHz. In the case of both sub-bands inverted together, the centre frequency is defined as the centre of the band considering both sub-bands, e.g. 5018 for a simultaneous inversion of 501.4 GHZ and 502.2 GHz bands.
- MPPPP, here M is the spectrometer that was used (A , B or C) and PPPP is the 4 digit hexadecimal number of the orbit.
- VVJ is the version number of the product. VV is the major version number and J is the minor.
- version number. e.g. 013 is read as version one point three.
- The extension .L2P means level two product. This level two product is what is known as the scientific product.

Level 2 Retrieved Profiles Data Product Structure

Reading an SMR Level 2 data product retrieves the data needed to write the Level 2 Retrieved Profiles Data Product Structures. The history implementation of the structure, the structure fields with some commnets about their HDF-EOS coding, and a description of the fields follow. See next section for further explanations about the HDF-EOS coding.

[1]For the validation institues of Bordeaux and Chalmers the filename will be the same except that SMR is replaced by SOB ans SCH respectively

```
#ifndef SMRLEVEL2_H
#define SMRLEVEL2_H

/*
        History:

        Version 0.1:
        First version created March 11 1999

        Version 0.2
        August 16 1999
        Major changes concerning new and removed structure members.

        Version 0.3
        August 3 2000
        More fields added.

        Version 0.4
        March 21 2001
        Changes made after retrieval group meeting 1 at MISU
        Constituents field changed to Profiles
        Field double Orbit changed to char *OrbitFileName
        Variances field is removed.
        AprioriContr field is removed.
        Added field ZPTSource.
        Added field ZPT
        Added fields NullSpaceError, MeasError, MeasResp and TotalError.
        Added fields StartTan and EndTan.

        Version 0.5
        March 26 2001
        Removed fields Level1Bfile, ClimTP
        Added fields Nzpt, ZPTNames and ZPT.
        Changed NullSpaceError to SmoothingError
        Changed AprioriUncert from a vector to a matrix.

        Version 0.6
        July 17 2001
        Added #define MAXSTRING  32, all strings in the structure
        are now of length MAXSTRING
        Changed AprioriUncert from float to a double.
        Changed SmoothingError from float to a double.
        Added field NzptParam
        Fields Day, Month, Year, Hour, Min and Secs changed from int to short

        Version 0.7
        Aug 30 2001
        Extra parameters added to the diagnostic product.
        (see diagnostics product below)

        Version 0.8
        Jun 27 2002
        Revision of fields before implementatioon of Level2 files
        in Bdx and Chalmers.
        Remove AprioriUncert
        Change all double fields to float.

        Version 0.9
        Sep 5 2002
        Adding modified julian date field

        Version 1.0
        Sep 23 2002
        Changing char to short for fields Nspecies and NzptParam
        Changing float to double for field MJD and LST

        Version 1.1
        Sep 30 2002
        Changing short to char for fields Version1b and Version2
        Specifying latitude and longitude in ranges [-90,90] [-180,180] and
        adding Time field, all changes to comply with reserved field names
        for geolocation
        Changing units of LST to hours.
```

```
        Especyfing units for altitude to km, pressure in hPa and temperature
        in Kelvin.

        Version 1.2
        Feb 24 2003
        ZPT is from now our best knowledge of these quantitites, not the pure
        apriori data.
*/


#define VERSION_L2 0x0008 /* read: version (high byte).(low byte), here: 0.6 */
#define L2VERSION "080"
#define MAXSTRING  32
#define UNDEFINED  0

struct SMRLevel2Data {

        char Version1b;           // level 0, Retrieval point grid
        char Version2;            // level 0, Retrieval point grid
        unsigned long Quality;    // level 0, Retrieval point grid
        char *Source;             // level 0, Retrieval point grid

        char *ZPTSource;          // level 0, Retrieval point grid
        char *OrbitFilename;      // level 0, Retrieval point grid
        float SunZD;              // level 0, Retrieval point grid
        double LST;               // level 0, Retrieval point grid
        int ScanNo;               // level 0, Retrieval point grid also T/P point grid
        short Nspecies;           // level 0, Retrieval point grid
        short NzptParam;          // level 0, T/P point grid
        short Day;                // level 0, Retrieval point grid
        short Month;              // level 0, Retrieval point grid
        short Year;               // level 0, Retrieval point grid
        short Hour;               // level 0, Retrieval point grid
        short Min;                // level 0, Retrieval point grid
        short Secs;               // level 0, Retrieval point grid
        float Ticks;              // level 0, Retrieval point grid

        float Latitude;           // level 0, Retrieval point grid also T/P point grid
        float Longitude;          // level 0, Retrieval point grid also T/P point grid
        float StartLat;           // level 0, Retrieval point grid
        float EndLat;             // level 0, Retrieval point grid
        float StartLong;          // level 0, Retrieval point grid
        float EndLong;            // level 0, Retrieval point grid
        float StartTan;           // level 0, Retrieval point grid
        float EndTan;             // level 0, Retrieval point grid
        double MJD;               // level 0, Retrieval point grid
        double Time;              // level 0, Retrieval point grid

        char  **Species;          // level 1, Retrieval point grid
        short *Naltitudes;        // level 1, Retrieval point grid
        char  **ZPTNames;         // level 1, T/P point grid
        short *Nzpt;              // level 1, T/P point grid

        float **Altitudes;        // level 2, Retrieval point grid
        float **Profiles;         // level 2, Retrieval point grid
        float **MeasError;        // level 2, Retrieval point grid
        float **MeasResp;         // level 2, Retrieval point grid
        float **TotalError;       // level 2, Retrieval point grid
        float **SmoothingError;   // level 2, Retrieval point grid
        float **ZPT;              // level 2, T/P point grid

};

#endif
```

- char Version1b
  The version number of the level 1b processing chain that was used to create the level 1b file.

- char Version2

The version number of the level 2 processing chain that was used to create the level 2 file.

- unsigned long Quality
  Up to 32 bit of status information for various error, warning or informative purposes. Quality = 0 means a good retrieval. i.e. no problems. Quality = 1 means a bad retrieval. i.e. should not be used.

- char *Source
  The observation mode. e.g Stratospheric, Odd Hydrogen, Odd Nitrogen, Water Isotope.

- char *ZPTSource
  A character string containing the source of the temperature and pressure files.

- char *OrbitFileName
  The full orbit filename (excluding the extension .HDF) including the spectrometer that was used.
  e.g. OB1B13A4

- float SunZD
  The average solar zenith angle for the scan. Taken from the log file.
  Unit: degrees.

- double LST
  The average local (mean) sidereal time for the scan.
  Calculated from the values of the scan in the level 1b file.

$$\text{LST} = \frac{(\text{LST from the start of the scan} + \text{LST at the end of the scan})}{2} \tag{A-1}$$

  Unit: hours.

- double MJD
  The average modified julian date for the scan.
  Calculated from the values of the scan in the level1b file.

$$\text{MJD} = \frac{(\text{MJD from the start of the scan} + \text{MJD at the end of the scan})}{2} \tag{A-2}$$

  Unit: days.

- double Time
  A time field based on average modified julian data. Calculated as TAI93: seconds until(-)/since(+) midnight,1/1/93.

$$\text{Time} = 24 * 60 * 60 * (\text{MJD} - 48988) \tag{A-3}$$

  Unit: seconds.

- int ScanNo
  The scan number within this orbit. This number always refers to complete scans, it is taken from the log file.

- short Nspecies
  The number of species that were retrieved.If temperature is a retrieval product, Nspecies is the number of species plus one, see below the field Species.

- short NzptParam
  The number of ZPT parameters. At the moment this is set to 3. i.e. Altitude, Pressure and Temperature.

- short Day
  Day of observation. Calculated from the average MJD in the log file.

- short Month
  Month of observation. Calculated from the average MJD in the log file.

- short Year
  Year of observation. Calculated from the average MJD in the log file.

- short Hour
  Hour of observation. Calculated from the average MJD in the log file.

- short Min
  Minute of observation. Calculated from the average MJD in the log file.

- short Secs
  Seconds of observation. Calculated from the average MJD in the log file.

- float Ticks
  Ticks of observation. Calculated from the average MJD in the log file.

- float Latitude
  Reference latitude for the scan. Calculated as,

  $$\text{Latitude} = \frac{(\text{Latitude from the start of the scan} + \text{Latitude at the end of the scan})}{2} \quad \text{(A-4)}$$

  Calculated from the values in the log file.
  Unit: decimal degrees on the range [-90,90].

- float Longitude
  Reference longitude for the scan. Calculated as,

  $$\text{Latitude} = \frac{(\text{Attitude from the start of the scan} + \text{Latitude at the end of the scan})}{2} \quad \text{(A-5)}$$

  Calculated from the values in the log file.
  Unit: decimal degrees on the range [-180,180].

- float StartLat
  The geodetic latitude of the tangent point for the first spectra in the scan.
  Taken from the log file.
  Unit: degrees on the range [-90,90].

- float EndLat
  The geodetic latitude of the tangent point for the last spectra in the scan. Taken from the log file.
  Unit: degrees on the range [-90,90].

- float StartLong
  The geodetic longitude of the tangent point for the first spectra in the scan. Taken from the log file.
  Unit: degrees on the range [-180,180].

- float EndLong
  The geodetic longitude of the tangent point for the last spectra in the scan. Taken from the log file.
  Unit: degrees on the range [-180,180].

- float StartTan
  The start tangent altitude of the scan. Taken from the log file.
  Unit: Kilometers.

- float EndTan
  The end tangent altitude of the scan. Taken from the log file.
  Unit: Kilometers.

- char **Species
  The name of each retrieved species. The convention here is to write the species name followed by the center frequency number of the band. e.g O3_5014, this is read as ozone using the frequency band with center frequency 501.4 GHz. If temperature is a retrieval product, the field can also have the string TEMP followed by the center frequency as before, e.g TEMP_5014. The string is of length MAXSTRING

- short *Naltitudes
  The number of retrieval grid altitude points for each retrieved species.

- char **ZPTNames
  This is an array that contains the following strings:
  ZPTNames[0] = "Altitude ";
  ZPTNames[1] = "Pressure ";
  ZPTNames[0] = "Temperature";
  They are always called this and always in the same order. The strings are of length MAXSTRING.

- short *Nzpt
  The number of altitude points for the *a priori* temperature, pressure and altitude grid. This should be equal to the number of retrieval layers that was used in the inversion.

- float **Altitudes
  A two dimensional array, where each row corresponds to the retrieval altitude grid for each retrieved species.
  Unit: Kilometers.

- float **Profiles
  The molecular volume mixing ratio (VMR) for each retrieved constituent. If temperature is a retrieval product, it can also be the temperature distribution in K. A two dimensional array, where each row corresponds to the VMR of a retrieved species and the temperature if it is retrieved.
  Unit: VMR and Kelvin.

- float **ZPT
  The best knowledge of the temperature and pressure after the retrieval process in the retrieval altitude grid. Whether this is the *a priori* pressure and temperature used during the inversion or not depends on the retrieval procedure. For instance, if temperature is retrieved the temperature here corresponds to the retrieved temperature.
  Unit: Kilometers, hPascal and Kelvin.

- float **SmoothingError
  This is the retrieval error covariance matrix (S_n) due to the error between the true and *a priori* states. It is calculated as,

$$S_n = (A - I)S_x(A - I)^T \tag{A-6}$$

  here A is the averaging kernels matrix, I is the identity matrix, S_x is the *a priori* covariance matrix that was used when performing the inversion and T means transpose. Only the diagonal terms are kept.

- float **MeasError
  This is the retrieval error covariance matrix (S_m) due to errors in the measurement. It is calculated as,

$$S_m = D_y S_y D_y^T \tag{A-7}$$

  here D_y are the contribution functions, S_y is the error covariance matrix, and T means transpose. Only the diagonal terms are kept.

- float **MeasResp
  This is defined as,

$$w^p(i) = \sum_j A(i,j)\frac{x_a(i)}{x_a(j)} \tag{A-8}$$

  Here A(i,j) describes the submatrix from species j to i. Only the rows of the submatrices from species i to j are stored.

- float **TotalError
  This is the total retrieval error covariance matrix (S_δ), given by

$$S_\delta = S_n + S_m \tag{A-9}$$

  Only the diagonal terms are kept.

Data Format with HDF-EOS

The SMR data is stored in a HDF-EOS file containing two point grid structures. One point grid structure is used to represent the data retrieved from a particular

frequency band, the other Point Grid structure holds the *a priori* temperature and pressure profiles. The point grid structure name for the retrieved profiles is given by the frequency band name, see table 1. For example '501.180 - 501.580 GHz' will be the Point Grid name for all the retrieved parameters in this band [2] . The *a priori* temperature and pressure Grid Point regardless of the frequency band is always called "T/P apriori".

Each point grid structure is implemented with 3 levels. Tables 2 - 4 illustrate the hierarchical table structure for the profiles grid-structure . Level 0 contains all the information on a per profile basis. The next level, Level 1, contains the information on the species that were retrieved. Level 0 and 1 are linked through the ID1 field. Level 2 contains the actual retrieved data. Level 1 and 2 are linked through the ID2 field. Similar hierarchical structure is implemented for the *a priori* point grid structure, see Tables 5 - 8.

Frequency Bands	
ID	Point grid name
1	501.180 - 501.580 GHz
2	501.980 - 502.380 GHz
3	501.180 - 502.380 GHz
4	544.202 - 545.002 GHz
5	488.950 - 489.350 GHz
6	488.350 - 488.750 GHz
7	488.350 - 489.350 GHz
8	556.598 - 557.398 GHz
9	576.062 - 576.862 GHz
10	489.950 - 490.750 GHz
11	556.550 - 557.350 GHz

Table 1: Frequency modes, also used as point grid names.

Geolocation			
Latitude	Longitude	NSpecies	ID1
y1	x1	3	0
y2	x2	4	1
y3	x3	3	2

Table 2: Level 0, retrieved parameters (product/diagnostics) point grid

[2]Only the frequency bands used presently during the observations are listed here. If further bands are used, the table and document will be consequently updated. Notice that bands with IDs 3 and 7 correspond to a joint retrieval of bands with IDs 1-2 and 5-6 respectively

Retrieval			
ID1	Species	Naltitudes	ID2
0	O3_501	2	0
0	NO2_501	3	1
0	H2O_501	2	2
1	O3_501	2	3
1	NO2_501	6	4
1	BrO_501	4	5
1	OClO_501	5	6

Table 3: Level 1, retrieved parameters product point grid

Data			
ID2	Altitude	Profiles	SmoothingError
0	20	5.4	2
0	30	6.3	6
1	20	5.2	4
1	40	7.9	3
1	50	5.8	1
2	50	4.9	7
2	60	6.9	7

Table 4: Level 2, retrieved parameters product point grid

Geolocation				
ScanNo	Latitude	Longitude	NzptParam	ID1
1	y1	x1	3	0
2	y2	x2	3	1
3	y3	x3	3	2

Table 5: Level 0, *a priori* temperature and pressure point grid

Retrieval			
ID1	ZPTNames	Nzpt	ID2
0	Altitude	2	0
0	Pressure	2	1
0	Temperature	2	2
1	Altitude	2	3
1	Pressure	2	4
1	Temperature	2	5
2	Altitude	2	6
2	Pressure	2	7
2	Temperature	2	8

Table 6: Level 1, *a priori* temperature and pressure point grid

ZPTinfo	
ID2	ZPT
0	20
0	30
1	20
1	40
2	50
2	60
3	20
3	30
4	20
4	40
5	50
5	60
6	20
6	30
7	20
7	40
8	50
8	60

Table 7: Level 2, *a priori* temperature and pressure point grid

TemperatureBrightness		
ID2	TB	Frequencies
0	200	501
0	300	502
1	250	501
1	100	502
2	120	501
2	180	502
3	200	490
3	250	491
3	204	492
4	100	490
4	150	491
4	260	492
5	320	490
5	301	491
5	160	492

Table 8: Level 2, Measurement or Fit point grid