

Development of Level 3 (gridded) Products for the Atmospheric Infrared Sounder (AIRS)

Stephanie L. Granger, Stephen S. Leroy, Evan M. Manning, Eric J. Fetzer, Robert B. Oliphant, Amy Braverman, Sung-Yung Lee, Bjorn H. Lambrigtsen

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Abstract— The Atmospheric Infrared Sounder (AIRS) sounding system is a suite of infrared and microwave instruments flown as part of NASA's Earth Observing System (EOS) onboard the Aqua platform. The AIRS dataset provides a daily, global view of Earth processes at a finer vertical resolution than ever before. However, analysis of the AIRS data is a daunting task given the sheer volume and complexity of the data. The volume of data produced by the EOS project is unprecedented; the AIRS project alone will produce many terabytes of data over the lifetime of the mission. This paper describes development of AIRS Level 3 data products that will help to alleviate problems of access and usability.

Keywords; *global analysis, climate studies, gridded data, data reduction*

I. INTRODUCTION

Level 3 products are composed of geophysical parameters that have been spatially and temporally re-sampled from Level 2 data products. Due to re-sampling and selecting a reduced set of reporting parameters, Level 3 datasets are substantially smaller than the lower level source products from which they are derived. Thus, Level 3 products can be used without a great deal of overhead in terms of data handling [1].

AIRS Level 3 product development is driven by the need for global analysis of AIRS data for weather and climate studies and to promote the use of AIRS data sets by providing easy-to-use, quantitative gridded data products for interdisciplinary science. The overarching objective of the AIRS Level 3 product development effort is to provide a consistent, standard global AIRS data set of relatively low volume to aid in research and data discovery.

The Atmospheric Infrared Sounder (AIRS) is one of six instruments launched on board the Aqua spacecraft on May 4, 2002, from Vandenberg Air Force Base in California. The AIRS sounding system is made up of a suite of three instruments designed to measure constituents of the earth's atmosphere: AIRS, the Advanced Microwave Sounding Unit (AMSU-A) and the Brazilian Humidity Sounder (HSB). The AIRS suite of instruments will observe global water and energy cycles, climate variation and trends, and the climate system response to increased greenhouse gases by making highly accurate measurements of air temperature and humidity, clouds, and surface temperature.

AIRS generates data products at several processing levels. The data volume of AIRS geophysical or retrieved products can range 4.7 MB to 18 MB per file with a daily data volume including metadata of about 5.5 GB. For a climate researcher the volume of data can be unwieldy.

Researchers interested in atmospheric dynamics, climate variability and change and the hydrologic cycle are interested in daily, multi-day and monthly global, gridded datasets [2]. A member of the research community would have to download, read in and process at least 240 Level 2 standard products to produce a single daily gridded file to satisfy their research needs. For communities interested in variability or trends, the data requirements expand proportionately. This raises many issues for the independent researcher including bandwidth limitations, lack of local storage, program development, and processing time.

II. CHARACTERISTICS OF THE INSTRUMENTS

The EOS Aqua platform is in a polar, sun-synchronous orbit with a nominal altitude of 705 km, an inclination of 98.2° and an orbital period of 98.8 minutes. The repeat cycle is 233 orbits (16 days) with a ground track repeatability of +/- 20 km. The platform will have equatorial crossing times of 1:30 AM and 1:30 PM.

A. AIRS

AIRS is a continuously operating cross-track scanning sounder, consisting of a telescope that feeds an echelle spectrometer [3]. The AIRS high spectral resolution infrared channels cover 3.4 to 15 μm region. The spatial footprint of the infrared channels is 1.1° in diameter, which corresponds to about 15x15 km in the nadir.

During each scan, the rotating external mirror scans the underlying Earth from 49° on one side of the nadir to 49° on the other side, in 90 integration periods, and provides two views of dark space, one view of an internal radiometric calibration target, and one view of an internal spectral calibration target. Thus each scan produces 94 sets of measurements (90 earth scenes and 4 calibrations). The scan is repeated every 8/3 seconds. The downlink data rate from the AIRS instrument is 1.2 Mbit/sec.

B. AMSU-A

AMSU-A is a 15-channel microwave temperature sounder implemented as two independently operated modules. Module 1 (AMSU-A1) has 12 channels in the 50-58 GHz oxygen absorption band which provide the primary temperature sounding capabilities and 1 channel at 89 GHz which provides surface and moisture information. Module 2 (AMSU-A2) has 2 channels – one at 23.8 GHz and one at 31.4 GHz which provide surface and moisture information (total precipitable water and cloud liquid water). Like AIRS, AMSU-A is a cross-track scanner. The three receiving antennas – two for AMSU-A1 and one for AMSU-A2 – are parabolic focusing reflectors that are mounted on a scan axis at a 45° tilt angle, so that radiation is reflected from a direction along the scan axis (a 90° reflection). AMSU-A scans approximately three times as slowly as AIRS (once per 8 seconds) and its footprints are approximately three times as large as those of AIRS (45 km at nadir). This results in three AIRS scans per AMSU-A scans and nine AIRS footprints per AMSU-A footprint.

C. HSB

The Humidity Sounder Brazil (HSB) is a 4-channel microwave moisture sounder implemented as a single module. Three channels are located near 183 GHz, while the fourth is a window channel at 150 GHz. Physically HSB is identical to AMSU-B, which is operated by NOAA on its most recent POES satellites, but HSB lacks the fifth channel (89 GHz) of AMSU-B. Like AIRS, HSB is a cross-track scanner, and it has only one parabolic scan mirror. Its scan speed as well as its footprints is similar to AIRS (three scans per 8 seconds and about 15 km at nadir, respectively). There is therefore one HSB footprint per AIRS footprint. The HSB instrument has been inoperative since February 2003.

III. AIRS LEVEL 2 DATA PRODUCTS

The ability of AIRS instruments to provide simultaneous observations of the earth's atmospheric temperature, ocean surface temperature, and land surface temperature, as well as humidity, clouds, albedo and the distribution of greenhouse gases, makes AIRS the primary EOS instrument for investigating several interdisciplinary issues to be addressed in earth science [3].

The AIRS data are organized in terms of data levels as defined by the Committee on Data Management, Archiving and Computing (CODMAC) as shown in the following table.

| CODMAC Data Levels | |
|--------------------|---|
| Level | Description |
| 0 | Reconstructed unprocessed instrument/payload data at full resolution; raw engineering measurements. |
| 1 ¹ | Reconstructed unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, computed and appended, but not applied, to the Level 0: processed tracking data. |
| 2 | Derived geophysical variables at the same resolution and location as the Level 1 source data. |
| 3 | Variables mapped on uniform space-time grid scales, usually with some completeness and consistency. |

The AIRS derived geophysical data (Level 2) consist of:

- Atmospheric temperature profiles with an average layer accuracy of 1K in 1 km thick layers in the troposphere and 1K in 4 km layers in the stratosphere
- Land surface temperature, sea surface temperature and infrared spectral surface emissivity
- Humidity profiles of 2 km resolution and total precipitable water vapor
- Fractional cloud cover, cloud spectral infrared emissivity, and cloud-top pressure and temperature
- Total ozone column density and column density in several layers of the atmosphere
- Trace gas column densities (e.g., CH₄, and CO and where possible in various layers with the atmosphere.)

All Level 2 products with the exception of cloud fraction are produced at the AMSU-A footprint resolution (45 km at nadir). The spatial resolution of AIRS cloud fraction is on the AIRS IFOV, or 15 km at nadir.

The AIRS Level 2 files range in size from 4.7 to 20 MB, with each file containing approximately 6 minutes of satellite coverage data, for a total of 240 files per day per instrument, per processing level. The Level 2 files are further broken down into standard and support files, the size of a single 6 minute Level 2 standard file is 4.7 MB and for a single Level 2 support file is 18 MB. The total daily volume for Level 2 standard products is ~ 1.1 GB per day, for support products the total is ~ 4.3 GB per day. Additionally, each standard and support file has a separate associated 122 KB metadata file that adds another ~ 29.3 MB per day per product type. Data volumes for different periods of time are given in the table below.

¹ AIRS and many other earth observing instruments break down the definition into Level 1A and Level 1B data. Level 1A refers to uncalibrated instrument data and level 1B refers to calibrated instrument data.

| Level 2 Data Volume | Products | |
|---------------------|-----------------------------|------------------------------|
| | Level 2 Standard | Level 2 Support |
| Daily | 4.7 MB * 240 files = 1.1 GB | 18 MB * 240 files = 4.3 GB |
| Weekly | 1.1 GB * 7 days = 7.7 GB | 4.3 GB * 7 days = |
| Monthly | 1.1 GB * ~30 days = 33 GB | 4.3 GB * 30 days = 129 GB |
| Yearly | 33 GB * 12 months = 396 GB | 129 GB * 12 months = 1.25 TB |

IV. AIRS LEVEL 3 PRODUCTS

Much of the analysis that will be undertaken using AIRS data requires the higher volume Level 1B and Level 2 data products given the high resolution and accuracy of these products. However, these lower level products are difficult to use given the size of the files and the data handling issues that can arise. Level 3 data sets are lower in volume than Level 1B and Level 2 data sets and are thus easier to use.

The AIRS project is developing Level 3 gridded data products to address issues of data volume and ease of use and also to facilitate inter-comparisons of data from other instruments as well as climate and trending studies. The AIRS Level 3 data products are based on Level 2 geophysical products and are created directly from the geophysical parameters, no additional parameters are derived.

We are planning the development of two types of Level 3 products: standard and quantization. The quantization product is a future release and is discussed in Section VI. For the initial release, we will produce global, gridded mean products with a spatial resolution of $1^\circ \times 1^\circ$ [4]. The temporal resolution of the products is daily, 8-day (half of the 16 day Aqua orbit repeat cycle) and monthly, based on the needs of different user communities [2].

The communities interested in the standard daily Level 3 product are the Numerical Weather Prediction (NWP) and numerical modeling community. They will be interested in temperature, specific humidity, and geopotential height profiles, cloud thickness, height and fraction and surface moisture/emissivity. The community interested in the 8-day product is researchers of climate quasi-oscillations and assorted phenomena, such as the Madden-Julian Oscillation, the annular modes, etc. The monthly products will interest those involved in climate trend analysis. They are typically interested in monthly means over long timescales and prefer data products with the lowest possible systematic errors. The characteristics of these three data types are summarized in the table below and example images are shown in Fig. 1.

| L3 Standard Product Characteristics | | |
|--|---|---|
| Daily | 8-Day | Monthly |
| "complex" data, leaves in gores between satellite tracks (missing) | "moderate" data, no gores, some data dropouts | "simple" data, no gores, complete coverage |
| $1^\circ \times 1^\circ$ spatial resolution | $1^\circ \times 1^\circ$ spatial resolution | $1^\circ \times 1^\circ$ spatial resolution |
| 1-day temporal resolution | 8-day temporal resolution | Monthly (calendar) |

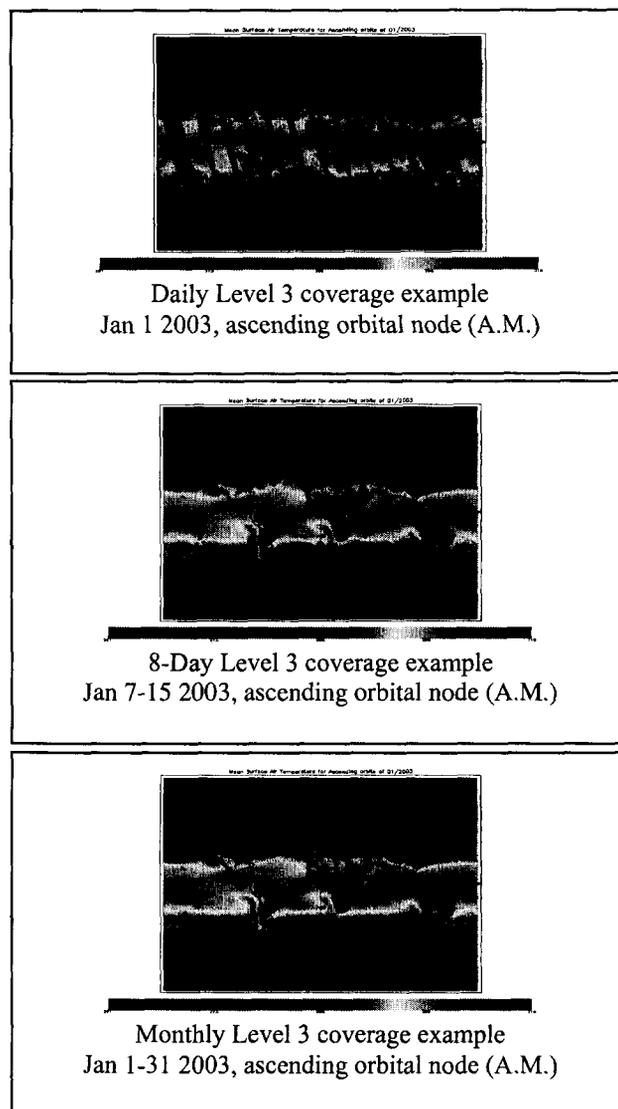


Figure 1. Examples of Level 3 coverage using images of atmospheric surface temperature (K)

V. LEVEL 3 DATA PRODUCT ISSUES

The AIRS project is encouraged by the EOS Project to produce output files in the Hierarchical Data Format (HDF). The Level 3 files are written using the HDF-EOS Grid API [6]. The HDF-EOS Grid API provides a convenient method of storing gridded data in binary format.

The Level 3 products include many of the geophysical parameters separated by ascending and descending orbits, however, the files have quickly become quite large as additional data parameters are added. We have explored techniques to reduce the size of the files including compression techniques and precision reduction.

A. Compression

Uncompressed AIRS Level 3 data are on the order of 153 MB daily. Several techniques have been tested for compression of AIRS Level 3 files. HDF-EOS offers three internal

compression techniques for the Grid API that are transparent to the user and have shown good results (Table 3). The three techniques are Run Length Encoding (RLE), Skipping-Huffman, and Deflate. Deflate is the same as Gzip and the level of compression can be adjusted through use of an integer flag with values from 1-9. Test runs using Run-Length Encoding have shown no significant results. Skipping-Huffman and Deflate result in significant reduction in file size for the daily Level 3 files. Skipping-Huffman compression reduces the file size from 153 MB to 50 MB, but I/O speed is negatively impacted. The results of the Deflate compression technique using a compression parameter of 5 give the most significant amount of compression, from 153 MB to 36 MB with little or no impact on I/O throughput.

B. Precision reduction

Early attempts to reduce the size of AIRS Level 3 products using standard lossless compression methods are promising. Further reduction can be realized by reducing the precision of the data values [7]. The compression methods described above work best when they can find patterns in the bytes, but floating-point representation will make two numbers that are not significantly different have very different bit patterns (Table 4).

| Value | Binary Representation |
|-----------|-----------------------|
| 1.998000 | 3F FF BE 77 |
| 2.002000 | 40 00 20 CC |
| 52.006000 | 40 00 62 4E |

The difference in binary representations between 1.998 and 2.002 is dramatic, but even the more typical case of the change from 2.002 to 2.006 changes 2 of 4 bytes for a 0.2% change.

For many AIRS product fields, the significance of the result is limited to ~1%, so the lower bits are filled with uncorrelated instrument and computational noise. In order to make AIRS data more compressible while maintaining the ease-of-use of floating-point notation, we determine how many bits are significant and round the number so that all lower bits will be zero. Use of rounding instead of truncation assures that the result will be unbiased.

Tests using a combination of a simple rounding technique and Deflate compression have shown a further reduction of Level 3 file size to 30 MB/daily file.

VI. FUTURE DIRECTIONS

We are planning development of a Level 3 quantization product as a future release [4]. This approach summarizes lower level data with a set of statistical summaries. Lower level data are partitioned into grid cells of $N^\circ \times N^\circ$ and for a given temporal period (e.g., one month). Each grid cell is separately summarized by a number of representative values and their associated weights. The collection of lower level data points with the same representative is called a cluster, and its representative is the cluster mean. Cluster membership is determined by assigning lower level data points to clusters in such a way that the errors between original values and their representatives are minimized. This constraint prevents there being one cluster for each data point.

The resulting statistical summaries can be thought of as high-dimensional histograms, with the following properties: 1) the bin representatives are means, not mid-points; 2) the bin shapes are allowed to adapt to the shape of the data; and 3) the histograms representing different grid cells are comparable in their quality as representations of the underlying data. The representatives and their weights can be used as proxies for the lower level (Level 2) data they summarize with the understanding that some information loss is suffered. That loss is quantified by the mean squared error between the summary and the original data, and is reported as part of the Level 3 product.

VII. CONCLUSIONS

AIRS Level 3 products will aid independent researchers and the climate community interested in AIRS data. They are easier to use and lower in volume than the Level 2 products. Several issues have arisen in terms of file size and organization. Compression techniques provided by NCSA within the HDF API's have shown good results in reducing the size of the Level 3 output products. In addition, precision reduction using the algorithm developed by Manning [7] for the AIRS project will reduce the data volume further and help to avoid propagating noise through the AIRS Level 3 products.

ACKNOWLEDGMENT

The authors wish to express sincere appreciation to our colleagues on the EOS AIRS team for providing support and comments for this work. In particular, we would like to thank Mike Gunson, and George Aumann of the Jet Propulsion Laboratory and Joel Susskind and Lena Iredell of the Goddard Spaceflight Center. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration, through the office of the Earth Systems Enterprise.

REFERENCES

- [1] R. Kahn and A. Braverman, "What shall we do with the data we are expecting from upcoming earth observation satellites?", American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America, *Journal of Computational and Graphical Statistics*, Vol. 8, Number 3, pp. 575-588, 1999
- [2] S. Leroy and S. Granger, "AIRS level 3 science requirements, version 1.0", AIRS Design File Memo, ADFM-660, April 2001
- [3] M. T. Chahine, "AIRS level 2 algorithm theoretical basis document, version 2.2", AIRS Project Office, JPL D-17006, Version 2.2, April 2001
- [4] A. Braverman, "Compressing massive datasets using quantization: an application to geophysical data.", *Journal of Computational and Graphical Statistics*, 2002
- [5] S. Granger et al, "AIRS level 3 software requirements, version 1.0", AIRS Project Office, JPL D-28558, May 2004-06-25
- [6] "HDF-EOS User's Guide for the ECS Project", 170-TP-005-001, 2003
- [7] E. Manning, "Precision reduction for increased compression of AIRS products", unpublished

Development of Level 3 (gridded) Products for the Atmospheric Infrared Sounder (AIRS)

Stephanie L. Granger, Stephen S. Leroy, Evan M. Manning, Eric J. Fetzer, Robert B. Oliphant, Amy Braverman,
Sung-Yung Lee, Bjorn H. Lambrigtsen
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Abstract— The Atmospheric Infrared Sounder (AIRS) sounding system is a suite of infrared and microwave instruments flown as part of NASA's Earth Observing System (EOS) onboard the Aqua platform. The AIRS dataset provides a daily, global view of Earth processes at a finer vertical resolution than ever before. However, analysis of the AIRS data is a daunting task given the sheer volume and complexity of the data. The volume of data produced by the EOS project is unprecedented; the AIRS project alone will produce many terabytes of data over the lifetime of the mission. This paper describes development of AIRS Level 3 data products that will help to alleviate problems of access and usability.

Keywords; *global analysis, climate studies, gridded data, data reduction*

I. INTRODUCTION

Level 3 products are composed of geophysical parameters that have been spatially and temporally re-sampled from Level 2 data products. Due to re-sampling and selecting a reduced set of reporting parameters, Level 3 datasets are substantially smaller than the lower level source products from which they are derived. Thus, Level 3 products can be used without a great deal of overhead in terms of data handling [1].

AIRS Level 3 product development is driven by the need for global analysis of AIRS data for weather and climate studies and to promote the use of AIRS data sets by providing easy-to-use, quantitative gridded data products for interdisciplinary science. The overarching objective of the AIRS Level 3 product development effort is to provide a consistent, standard global AIRS data set of relatively low volume to aid in research and data discovery.

The Atmospheric Infrared Sounder (AIRS) is one of six instruments launched on board the Aqua spacecraft on May 4, 2002, from Vandenberg Air Force Base in California. The AIRS sounding system is made up of a suite of three instruments designed to measure constituents of the earth's atmosphere: AIRS, the Advanced Microwave Sounding Unit (AMSU-A) and the Brazilian Humidity Sounder (HSB). The AIRS suite of instruments will observe global water and energy cycles, climate variation and trends, and the climate system response to increased greenhouse gases by making highly accurate measurements of air temperature and humidity, clouds, and surface temperature.

AIRS generates data products at several processing levels. The data volume of AIRS geophysical or retrieved products can range 4.7 MB to 18 MB per file with a daily data volume including metadata of about 5.5 GB. For a climate researcher the volume of data can be unwieldy.

Researchers interested in atmospheric dynamics, climate variability and change and the hydrologic cycle are interested in daily, multi-day and monthly global, gridded datasets [2]. A member of the research community would have to download, read in and process at least 240 Level 2 standard products to produce a single daily gridded file to satisfy their research needs. For communities interested in variability or trends, the data requirements expand proportionately. This raises many issues for the independent researcher including bandwidth limitations, lack of local storage, program development, and processing time.

II. CHARACTERISTICS OF THE INSTRUMENTS

The EOS Aqua platform is in a polar, sun-synchronous orbit with a nominal altitude of 705 km, an inclination of 98.2° and an orbital period of 98.8 minutes. The repeat cycle is 233 orbits (16 days) with a ground track repeatability of +/- 20 km. The platform will have equatorial crossing times of 1:30 AM and 1:30 PM.

A. AIRS

AIRS is a continuously operating cross-track scanning sounder, consisting of a telescope that feeds an echelle spectrometer [3]. The AIRS high spectral resolution infrared channels cover 3.4 to 15 μm region. The spatial footprint of the infrared channels is 1.1° in diameter, which corresponds to about 15x15 km in the nadir.

During each scan, the rotating external mirror scans the underlying Earth from 49° on one side of the nadir to 49° on the other side, in 90 integration periods, and provides two views of dark space, one view of an internal radiometric calibration target, and one view of an internal spectral calibration target. Thus each scan produces 94 sets of measurements (90 earth scenes and 4 calibrations). The scan is repeated every 8/3 seconds. The downlink data rate from the AIRS instrument is 1.2 Mbit/sec.

B. AMSU-A

AMSU-A is a 15-channel microwave temperature sounder implemented as two independently operated modules. Module 1 (AMSU-A1) has 12 channels in the 50-58 GHz oxygen absorption band which provide the primary temperature sounding capabilities and 1 channel at 89 GHz which provides surface and moisture information. Module 2 (AMSU-A2) has 2 channels – one at 23.8 GHz and one at 31.4 GHz which provide surface and moisture information (total precipitable water and cloud liquid water). Like AIRS, AMSU-A is a cross-track scanner. The three receiving antennas – two for AMSU-A1 and one for AMSU-A2 – are parabolic focusing reflectors that are mounted on a scan axis at a 45° tilt angle, so that radiation is reflected from a direction along the scan axis (a 90° reflection). AMSU-A scans approximately three times as slowly as AIRS (once per 8 seconds) and its footprints are approximately three times as large as those of AIRS (45 km at nadir). This results in three AIRS scans per AMSU-A scans and nine AIRS footprints per AMSU-A footprint.

C. HSB

The Humidity Sounder Brazil (HSB) is a 4-channel microwave moisture sounder implemented as a single module. Three channels are located near 183 GHz, while the fourth is a window channel at 150 GHz. Physically HSB is identical to AMSU-B, which is operated by NOAA on its most recent POES satellites, but HSB lacks the fifth channel (89 GHz) of AMSU-B. Like AIRS, HSB is a cross-track scanner, and it has only one parabolic scan mirror. Its scan speed as well as its footprints is similar to AIRS (three scans per 8 seconds and about 15 km at nadir, respectively). There is therefore one HSB footprint per AIRS footprint. The HSB instrument has been inoperative since February 2003.

III. AIRS LEVEL 2 DATA PRODUCTS

The ability of AIRS instruments to provide simultaneous observations of the earth's atmospheric temperature, ocean surface temperature, and land surface temperature, as well as humidity, clouds, albedo and the distribution of greenhouse gases, makes AIRS the primary EOS instrument for investigating several interdisciplinary issues to be addressed in earth science [3].

The AIRS data are organized in terms of data levels as defined by the Committee on Data Management, Archiving and Computing (CODMAC) as shown in the following table.

| CODMAC Data Levels | |
|--------------------|---|
| Level | Description |
| 0 | Reconstructed unprocessed instrument/payload data at full resolution; raw engineering measurements. |
| 1 ¹ | Reconstructed unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, computed and appended, but not applied, to the Level 0: processed tracking data. |
| 2 | Derived geophysical variables at the same resolution and location as the Level 1 source data. |
| 3 | Variables mapped on uniform space-time grid scales, usually with some completeness and consistency. |

The AIRS derived geophysical data (Level 2) consist of:

- Atmospheric temperature profiles with an average layer accuracy of 1K in 1 km thick layers in the troposphere and 1K in 4 km layers in the stratosphere
- Land surface temperature, sea surface temperature and infrared spectral surface emissivity
- Humidity profiles of 2 km resolution and total precipitable water vapor
- Fractional cloud cover, cloud spectral infrared emissivity, and cloud-top pressure and temperature
- Total ozone column density and column density in several layers of the atmosphere
- Trace gas column densities (e.g., CH₄, and CO and where possible in various layers with the atmosphere.)

All Level 2 products with the exception of cloud fraction are produced at the AMSU-A footprint resolution (45 km at nadir). The spatial resolution of AIRS cloud fraction is on the AIRS IFOV, or 15 km at nadir.

The AIRS Level 2 files range in size from 4.7 to 20 MB, with each file containing approximately 6 minutes of satellite coverage data, for a total of 240 files per day per instrument, per processing level. The Level 2 files are further broken down into standard and support files, the size of a single 6 minute Level 2 standard file is 4.7 MB and for a single Level 2 support file is 18 MB. The total daily volume for Level 2 standard products is ~ 1.1 GB per day, for support products the total is ~ 4.3 GB per day. Additionally, each standard and support file has a separate associated 122 KB metadata file that adds another ~ 29.3 MB per day per product type. Data volumes for different periods of time are given in the table below.

¹ AIRS and many other earth observing instruments break down the definition into Level 1A and Level 1B data. Level 1A refers to uncalibrated instrument data and level 1B refers to calibrated instrument data.

| Level 2 Data Volume | Products | |
|---------------------|-----------------------------|------------------------------|
| | Level 2 Standard | Level 2 Support |
| Daily | 4.7 MB * 240 files = 1.1 GB | 18 MB * 240 files = 4.3 GB |
| Weekly | 1.1 GB * 7 days = 7.7 GB | 4.3 GB * 7 days = |
| Monthly | 1.1 GB * ~30 days = 33 GB | 4.3 GB * 30 days = 129 GB |
| Yearly | 33 GB * 12 months = 396 GB | 129 GB * 12 months = 1.25 TB |

IV. AIRS LEVEL 3 PRODUCTS

Much of the analysis that will be undertaken using AIRS data requires the higher volume Level 1B and Level 2 data products given the high resolution and accuracy of these products. However, these lower level products are difficult to use given the size of the files and the data handling issues that can arise. Level 3 data sets are lower in volume than Level 1B and Level 2 data sets and are thus easier to use.

The AIRS project is developing Level 3 gridded data products to address issues of data volume and ease of use and also to facilitate inter-comparisons of data from other instruments as well as climate and trending studies. The AIRS Level 3 data products are based on Level 2 geophysical products and are created directly from the geophysical parameters, no additional parameters are derived.

We are planning the development of two types of Level 3 products: standard and quantization. The quantization product is a future release and is discussed in Section VI. For the initial release, we will produce global, gridded mean products with a spatial resolution of $1^\circ \times 1^\circ$ [4]. The temporal resolution of the products is daily, 8-day (half of the 16 day Aqua orbit repeat cycle) and monthly, based on the needs of different user communities [2].

The communities interested in the standard daily Level 3 product are the Numerical Weather Prediction (NWP) and numerical modeling community. They will be interested in temperature, specific humidity, and geopotential height profiles, cloud thickness, height and fraction and surface moisture/emissivity. The community interested in the 8-day product is researchers of climate quasi-oscillations and assorted phenomena, such as the Madden-Julian Oscillation, the annular modes, etc. The monthly products will interest those involved in climate trend analysis. They are typically interested in monthly means over long timescales and prefer data products with the lowest possible systematic errors. The characteristics of these three data types are summarized in the table below and example images are shown in Fig. 1.

| L3 Standard Product Characteristics | | |
|--|---|---|
| Daily | 8-Day | Monthly |
| "complex" data, leaves in gores between satellite tracks (missing) | "moderate" data, no gores, some data dropouts | "simple" data, no gores, complete coverage |
| $1^\circ \times 1^\circ$ spatial resolution | $1^\circ \times 1^\circ$ spatial resolution | $1^\circ \times 1^\circ$ spatial resolution |
| 1-day temporal resolution | 8-day temporal resolution | Monthly (calendar) |

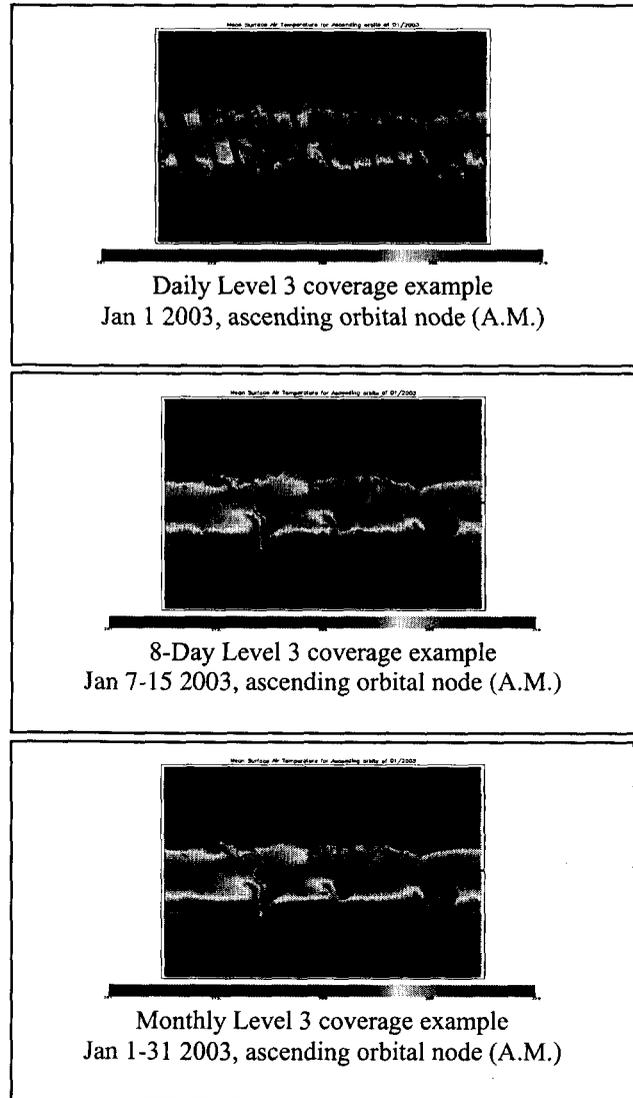


Figure 1. Examples of Level 3 coverage using images of atmospheric surface temperature (K)

V. LEVEL 3 DATA PRODUCT ISSUES

The AIRS project is encouraged by the EOS Project to produce output files in the Hierarchical Data Format (HDF). The Level 3 files are written using the HDF-EOS Grid API [6]. The HDF-EOS Grid API provides a convenient method of storing gridded data in binary format.

The Level 3 products include many of the geophysical parameters separated by ascending and descending orbits, however, the files have quickly become quite large as additional data parameters are added. We have explored techniques to reduce the size of the files including compression techniques and precision reduction.

A. Compression

Uncompressed AIRS Level 3 data are on the order of 153 MB daily. Several techniques have been tested for compression of AIRS Level 3 files. HDF-EOS offers three internal

compression techniques for the Grid API that are transparent to the user and have shown good results (Table 3). The three techniques are Run Length Encoding (RLE), Skipping-Huffman, and Deflate. Deflate is the same as Gzip and the level of compression can be adjusted through use of an integer flag with values from 1-9. Test runs using Run-Length Encoding have shown no significant results. Skipping-Huffman and Deflate result in significant reduction in file size for the daily Level 3 files. Skipping-Huffman compression reduces the file size from 153 MB to 50 MB, but I/O speed is negatively impacted. The results of the Deflate compression technique using a compression parameter of 5 give the most significant amount of compression, from 153 MB to 36 MB with little or no impact on I/O throughput.

B. Precision reduction

Early attempts to reduce the size of AIRS Level 3 products using standard lossless compression methods are promising. Further reduction can be realized by reducing the precision of the data values [7]. The compression methods described above work best when they can find patterns in the bytes, but floating-point representation will make two numbers that are not significantly different have very different bit patterns (Table 4).

| Value | Binary Representation |
|-----------|-----------------------|
| 1.998000 | 3F FF BE 77 |
| 2.002000 | 40 00 20 CC |
| 52.006000 | 40 00 62 4E |

The difference in binary representations between 1.998 and 2.002 is dramatic, but even the more typical case of the change from 2.002 to 2.006 changes 2 of 4 bytes for a 0.2% change.

For many AIRS product fields, the significance of the result is limited to ~1%, so the lower bits are filled with uncorrelated instrument and computational noise. In order to make AIRS data more compressible while maintaining the ease-of-use of floating-point notation, we determine how many bits are significant and round the number so that all lower bits will be zero. Use of rounding instead of truncation assures that the result will be unbiased.

Tests using a combination of a simple rounding technique and Deflate compression have shown a further reduction of Level 3 file size to 30 MB/daily file.

VI. FUTURE DIRECTIONS

We are planning development of a Level 3 quantization product as a future release [4]. This approach summarizes lower level data with a set of statistical summaries. Lower level data are partitioned into grid cells of $N^{\circ} \times N^{\circ}$ and for a given temporal period (e.g., one month). Each grid cell is separately summarized by a number of representative values and their associated weights. The collection of lower level data points with the same representative is called a cluster, and its representative is the cluster mean. Cluster membership is determined by assigning lower level data points to clusters in such a way that the errors between original values and their representatives are minimized. This constraint prevents there being one cluster for each data point.

The resulting statistical summaries can be thought of as high-dimensional histograms, with the following properties: 1) the bin representatives are means, not mid-points; 2) the bin shapes are allowed to adapt to the shape of the data; and 3) the histograms representing different grid cells are comparable in their quality as representations of the underlying data. The representatives and their weights can be used as proxies for the lower level (Level 2) data they summarize with the understanding that some information loss is suffered. That loss is quantified by the mean squared error between the summary and the original data, and is reported as part of the Level 3 product.

VII. CONCLUSIONS

AIRS Level 3 products will aid independent researchers and the climate community interested in AIRS data. They are easier to use and lower in volume than the Level 2 products. Several issues have arisen in terms of file size and organization. Compression techniques provided by NCSA within the HDF API's have shown good results in reducing the size of the Level 3 output products. In addition, precision reduction using the algorithm developed by Manning [7] for the AIRS project will reduce the data volume further and help to avoid propagating noise through the AIRS Level 3 products.

ACKNOWLEDGMENT

The authors wish to express sincere appreciation to our colleagues on the EOS AIRS team for providing support and comments for this work. In particular, we would like to thank Mike Gunson, and George Aumann of the Jet Propulsion Laboratory and Joel Susskind and Lena Iredell of the Goddard Spaceflight Center. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration, through the office of the Earth Systems Enterprise.

REFERENCES

- [1] R. Kahn and A. Braverman, "What shall we do with the data we are expecting from upcoming earth observation satellites?", American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America, *Journal of Computational and Graphical Statistics*, Vol. 8, Number 3, pp. 575-588, 1999
- [2] S. Leroy and S. Granger, "AIRS level 3 science requirements, version 1.0", AIRS Design File Memo, ADFM-660, April 2001
- [3] M. T. Chahine, "AIRS level 2 algorithm theoretical basis document, version 2.2", AIRS Project Office, JPL D-17006, Version 2.2, April 2001
- [4] A. Braverman, "Compressing massive datasets using quantization: an application to geophysical data:", *Journal of Computational and Graphical Statistics*, 2002
- [5] S. Granger et al, "AIRS level 3 software requirements, version 1.0", AIRS Project Office, JPL D-28558, May 2004-06-25
- [6] "HDF-EOS User's Guide for the ECS Project", 170-TP-005-001, 2003
- [7] E. Manning, "Precision reduction for increased compression of AIRS products", unpublished