Gridded Hourly Text Products: A TRMM Data Reduction Approach

Erich Franz Stocker,
NASA Goddard Space Flight Center, Greenbelt, Maryland

John Kwiatkowski and Owen Kelley
George Mason University, Fairfax, Virginia

Abstract—The quantity of satellite observations available for the study of global precipitation is both a blessing and a curse. The sheer volume and complexity of many of the data products makes it difficult for many researchers to use. This paper describes one approach that TRMM used to reduce the volume and complexity of data.

I. INTRODUCTION

The Tropical Rainfall Measuring Mission (TRMM) is a joint U.S./Japanese mission. The TRMM satellite was launched from Tanegashima Island, Japan, on November 27, 1997. Three-fourths of the energy driving atmospheric wind circulation derives from latent heat released by tropical precipitation [1]. TRMM was launched to help address some of the questions pertaining to tropical precipitation. It provides observations from space in an inclined orbit extending from 35° N to 35° S. TRMM carries three instruments that measure precipitation: the Visible and Infrared Scanner (VIRS), the TRMM Microwave Imager (TMI) (a radiometer), and the first operational spaceborne Precipitation Radar (PR) [2]. In addition, TRMM carries two Earth Observing System (EOS) sensors: the Lightening Imaging Sensor (LIS) and the Clouds and Earth Radiant Energy System (CERES).

Compared to some satellites such as the EOS Terra satellite, TRMM does not generate a huge volume of data daily. The three TRMM rain instruments transmit 1.8 GB of raw science and housekeeping data daily. However, this represents only the tip of the data volume with which potential users must deal. The TRMM Science and Data Information System (TSDIS) processes these initial satellite data through three transformation levels. Level 1 processing generates calibrated, geo-located products at the instrument instantaneous field of view (IFOV). Level 2 processing converts the previous level products to appropriate geophysical parameter values (e.g. rain rates) at the IFOV. Level 3 processing accumulates the geophysical parameter values within spatial grids (generally a 5° x 5° grid) and over pre-determined time periods (usually monthly). This science processing generates more than 17.4 GB of additional data daily.

While a single day of data might not seem an overwhelming retrieval and storage burden for some scientists, data for the entire mission would prove daunting for most. For scientists without sufficient network bandwidth or storage even retrieving a single day of TRMM data would be difficult. The study of diurnal, seasonal and annual variations in precipitation is an important research area and multi-year TRMM data would certainly be required. Retrieving only the appropriate level 3 data rather than the instantaneous data could mitigate the volume issue. This could be done if the standard TRMM level 3 products contained both the spatial and time sampling required. Standard TRMM level 3 products from TMI, PR and the combined algorithm would consume only 4 GB for three mission years. However, level 2 products remain the best tools for examination of diurnal precipitation variation. Also, studies requiring smaller temporal or finer spatial sampling than those contained in the standard level 3 products would still require the retrieval of level 2 products. As a result not every investigator would be able to take advantage of the reduced data volume of the TRMM level 3 products.

All TRMM data is produced, archived and distributed in the Hierarchical Data Format (HDF) developed by the National Center for Supercomputing Applications (NCSA) at the University of Illinois. Many public domain and commercial software tools are able to use this format directly. Additionally, TSDIS has made available a toolkit that simplifies reading and writing HDF TRMM data. Nevertheless, some users continue to prefer a simpler, more direct format with fewer parameters, less overhead and less time required to learn.

II. DATA REDUCTION APPROACH

In 1999, after discussions with TRMM science team members, the TRMM Project Scientist determined that TRMM should address the issues of data volume, parameter inclusion and simplicity of format. An effort began to establish the content and format of a simple rain product. Five goals were established for this data reduction effort. The product was to use the simplest and most easily understood format. A single product was to include all level 2 TRMM surface rain estimates. It was to be as small as possible to facilitate electronic retrieval even over narrow network bandwidth as well as to minimize storage. It would use the finest spatial grid appropriate for TRMM data and allow users to combine grid cells in a straightforward manner. Finally the product was to present the data in a fine temporal resolution that allows users to create longer temporal aggregations using only the data contained in the product. A decision was also made to make this product generally available to anyone that wished to retrieve it. The idea was to avoid use of special interfaces, registration mechanisms, and the overhead required for the management of the more voluminous standard TRMM.
The product implemented to meet the goals stated in the previous paragraph was given the TRMM identification 3G68 [3]. To address the goal of simplicity, 3G68 is produced as a line-oriented ASCII text file. Each line has a particular format and each line is terminated with the ASCII new-line character. The first five lines provide metadata that describe the product, the numeric grid information, numeric TRMM orbital information, words identifying the purpose of the numeric grid information and finally a description of the content of each data line. HDF was adopted for TRMM because a self-describing format was desired and the volume of the standard products would have made a text format impractical. With the inclusion of the metadata lines the 3G68 is also self-describing but using a much simpler format. However, the much simpler and more compact data presentation in 3G68 make text the most appropriate self-describing format. It also has the advantage of being read by standard text editors.

Among the TRMM standard level 2 products are three products that contain surface rain estimates. One of these is the rain estimate produced from TMI (TRMM ID 2A12). The second is produced from the PR (2A25). This product contains a rain estimate profile from 80 vertical levels. The last product is a combined product (2B31) that produces a rain estimate based on a combination of TMI and PR data. The combined product also accounts for the different swath widths of TMI and PR. Each of these level 2 products contains rain estimates at the IFOV level. Each product also contains additional and useful parameters. Some of these are parameters needed by the algorithm developer to ascertain that the algorithm is working properly. These contribute to making these products large. The data reduction objective for 3G68 led to the decision to include surface rain-rate information only. Additionally, surface rain rates from all three products would be included in single product rather than create a separate product for each of the estimates.

A number of different suggestions were made for the base temporal aggregation to use for 3G68. All could easily be accommodated by implementing 3G68 using a one-hour grid. Additionally, each of the 24 grids required for a day would be packaged together in a single daily product. This packaging would make it easy for users to aggregate as many hourly grids as necessary to create the desired temporal resolution. Each of the hourly grids is composed of 0.5° x 0.5° grid cells. This cell size was determined to be the smallest cell appropriate given the number of observations likely as a result of the TRMM revisit time. Once again if a coarser spatial resolution is desired to provide more samples, multiple grid cells can easily be combined into a larger cell. It was also decided to identify each cell with reference to a global grid rather than one based on the TRMM coverage area of 35° S to 35° N. This would facilitate combining rain data from instruments on other satellites (e.g., SSM/I on the DMSP satellites) that have orbital characteristics different from TRMM. Indeed, there is an ongoing effort to create a product like 3G68 for SSM/I rain data to make multisensor combination easier.

Two strategies were applied to reduce the size of the daily 3G68 products. Only surface rain rate data was included in the product and only grid cells that contained data would be written to the file. The format chosen for 3G68 data also enabled a further reduction strategy. Each cell’s data is written on a separate text line of the file. Data items are delimited by a single space. Each of the cell lines contains an integer identifying the cell by its row number followed by one identifying the column number of the cell. These are followed by integers identifying first the hour and then the minute of the first pixel in that cell. The rain data from the three rain sources follows. TMI rain data appears first. The PR data follows TMI, and combined data follows the PR. Each instrument group provides a count of the total pixels in that cell, a count of the rainy pixels in the cell, the mean rain rate in mm/hr and an integer indicating the percentage of the rain determined to be convective. Because of the difference in the swath width of TMI and PR there are cells that may have data from TMI but not from PR or the combined. The opposite could be true also. As part of the data reduction strategy, if a cell has TMI data but no PR and combined data, 3G68 writes out the TMI data and a –9 for the PR total pixel value. No other PR and combined data is written. Initial tests demonstrated that this approach would reduce the daily product size by as much as 50%. Yet the format remained easy to read and easy to manipulate. TSDIS also provides sample read routines in FORTRAN 77, C and IDL.

III. CONCLUSIONS

The 3G68 product contains the surface rain estimates from three different standard TRMM instantaneous field of view products. The product is created using hourly grids but packaged as daily files. A single day’s product consumes only 5.5 MB, uncompressed, but gzip compression reduces 3G68 by a factor of five to 1.1 MB. One year of mission data would require 1.9 GB of storage uncompressed or approximately 396 MB compressed. The entire mission to 1 April 2001 would require 6.6 GB uncompressed or 1.9 GB compressed. Comparing these volumes with the standard TRMM product volumes in Table 1 indicates the magnitude of the data reduction in 3G68.

TABLE 1

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Daily (GB)</th>
<th>Annual (GB)</th>
<th>Mission (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMI</td>
<td>1.6</td>
<td>579.6</td>
<td>1883.7</td>
</tr>
<tr>
<td>PR</td>
<td>4.0</td>
<td>1443.6</td>
<td>4691.7</td>
</tr>
<tr>
<td>Combined</td>
<td>2.5</td>
<td>900.0</td>
<td>2925.0</td>
</tr>
<tr>
<td>Total</td>
<td>8.1</td>
<td>2923.2</td>
<td>9500.4</td>
</tr>
</tbody>
</table>

It is clear that 6.6 GB for the entire TRMM mission represents a huge reduction over the 9.5 TB for the standard TRMM data products. Additionally, compression of the standard TRMM products yields only a ~50% reduction rather than the much higher compression rate of 3G68 products. The entire mission’s worth of rain estimates could be represented in 1.9 GB.
of compressed 3G68 products while the equivalent standard TRMM compressed products would still require 4.3 TB.

Of course, the main issue in any data reduction effort is not just achieving the reduction goal but creating a product useful for research. 3G68 appears to have succeeded in this area as well. Even with minimal publicity, an increasing number of users are downloading the 3G68 product. Additionally, many requests have been received to create a similar product for SSM/I data. Users have indicated that they are using the product to improve precipitation algorithms, to research tropical precipitation, to study temporal variations in precipitation, and to uncover spatial patterns in precipitation. The product is also being used in research for masters and doctorate degrees.

An important objective of the 3G68 implementation was to facilitate aggregating hourly grids into longer time aggregations. This is important to increase the sampling within each grid. Aggregation also allows precipitation to be monitored over longer time periods. The TSDIS OrbitViewer is a tool used for visualizing any TRMM data product [4]. The OrbitViewer, written in IDL, can dynamically create multi-hour images from 3G68 hourly grids. It can also create a daily image from a 3G68 file. Each image in Fig. 1 is a multi-hour aggregation of the 3G68 TMI precipitation data showing typhoon Saomai as it moves toward the southern coast of Japan. The images track the progress of the typhoon over two days. This required retrieving only two 3G68 files (2.2 MB compressed). The 3G68 data provides the rain rates (mm/hr) within the typhoon as well as showing the shape and size of Saomai. The red/orange colors in the images represent higher rain rates. The multi-hour aggregations shown over several days also provide some indication of the motion of the typhoon. These images clearly indicate that 3G68 provides the researcher with the ability and flexibility to look at patterns and variations using the temporal and spatial aggregations required by the study.

Unfortunately, TRMM revisit times provide poor sampling at very high spatial resolutions and low temporal resolutions. NASA is currently studying the possibility of a future mission that could address the sampling issues by producing an approximately three-hour global precipitation estimate. Such a mission might provide approximately three hour global precipitation data. This coverage would make a product such as 3G68 even more useful.

REFERENCES

Fig. 1. Two days of 3G68 data showing typhoon Saomai south of Japan.