Transition from TSDIS to the Precipitation Processing System

John Kwiatkowski, Erich F. Stocker, Yimin Ji, John Stout, Joyce Chou, Patty McCaughey, Owen Kelley, Tony Stocker, Michael Hensley, and Steve Bilanow
Precipitation Processing System (PPS), NASA Goddard. *Contact: helpdesk@pps-mail.gsfc.nasa.gov*

Introduction

The Precipitation Processing System (PPS) will take over the operational production of TRMM Version 6 data. PPS will be used to process future versions of TRMM data as well as data from the Global Precipitation Measurement (GPM) mission at NASA.

The TRMM Science Data Information System (TSDIS) has generated the NASA TRMM products since launch in 1997. While automated, TSDIS is a single satellite system with little flexibility to incorporate new data sources. As the TRMM mission progressed the Precipitation Measuring Missions (PMM) Science team became able to incorporate new data sources and more complex precipitation retrieval schemes. In addition, NASA's Global Precipitation Measurement (GPM) mission will require the ability of the data system to incorporate data from multiple satellites at different rates, spatial resolutions and sources. As a response to these requirements, the Precipitation Processing System has been created to take on Version 6 and future TRMM processing and will be used as the processing system at NASA for GPM data.

PPS for TRMM Version 6 Processing

An initial build of the PPS system with similar functionality to TSDIS will be used for processing TRMM Version 6 data. This system is designated **PPS--** and is a subset of the planned functionality of the full PPS for GPM. **PPS--** runs on clusters of commodity PC's running Linux and legacy SGI systems.

Production System Hardware (Creation of retrieval products)
- Dual Host nodes (for redundancy)
  - two dual core AMD 2216 Opteron CPU's, 8 GB of RAM.
- 96 Compute nodes
  - single dual core AMD 2216 Opteron CPU, 6 GB of RAM
  - diskless nodes running in RAM only.
- Scyld ClusterWorks 4.1 Beowulf software on top of CentOS 4.6 Operating System.
- 34 TB of Panasas high-speed clustered storage using DirectFlow 3.0.6 client.

Storage management (Access and Archiving Across System)
- Host node, with two dual core AMD Opteron CPU's, 8 GB of RAM.
- 8 Compute nodes
  - single dual core AMD Opteron CPU, 8 GB of RAM.
  - additional Gigabit ethernet connections.
- Approximately 800 TB of RAID 6 SATA storage.
- Front-end servers
  - Four Dell 2950 servers running RHEL 4.6.
  - using PolyServer for load-balanced request servicing.

Integration and Test Environment (ITE)

Experience with TRMM retrieval algorithms has shown that a test environment at the data system is an invaluable tool when evaluating new retrieval methods and performing sensitivity studies during algorithm checkout.

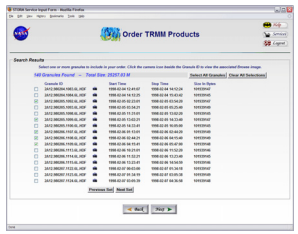
ITE Cluster Hardware
- Host node, with two dual core AMD 2216 Opteron CPU's, 8 GB of RAM.
- 32 Compute nodes
  - single dual core AMD 2216 Opteron CPU, 6 GB of RAM.
  - diskless nodes running in RAM only.
- Scyld ClusterWorks 4.1 Beowulf software on top of CentOS 4.6 OS.
- 17 TB of Panasas high-speed clustered storage using DirectFlow 3.0.6

The clustered systems replace the large, expensive SGI systems of TSDIS. The PPS-- cluster is based on little-endian architecture and results in code porting issues that had to be addressed (see section at right). A legacy SGI system will remain in use for all Real-Time and L1 PR algorithm codes.

New User Interfaces for Searching and Ordering TRMM Data

STORM: Science Team On-Line Request Module.

Access to all TRMM data with full search capabilities for members of the Precipitation Measuring Missions (PMM) Science Team and the public.

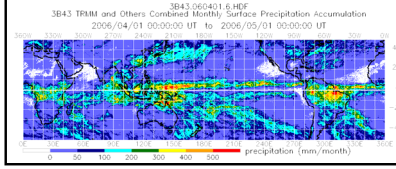


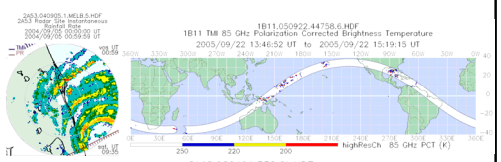
- Search and order TRMM data products.
- Create standing orders.
- Examine browse images.
- Select geographic areas of interest.
- Interface to TRMM On-Line Archive.
- All data delivered by network.
- Direct FTP interface to archive available for automated access.
- Trending data access and plots.

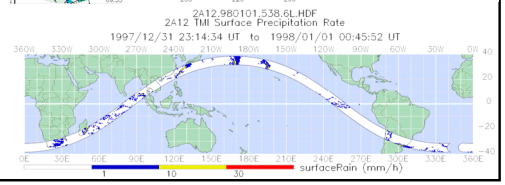
New TRMM Browse Images

TRMM browse images in TSDIS exist as static images within HDF files which allow some useful metadata attributes. However, HDF files themselves have little support in non-scientific software applications.

New browse images are
- Simple, static, raster browse images.
- Portable Network Graphics (PNG) format.
- Supported in all modern web browsers.
- Accessible to all data users, scientists and non-scientists.
- Images linked directly to the Search/Ordering process in STORM.

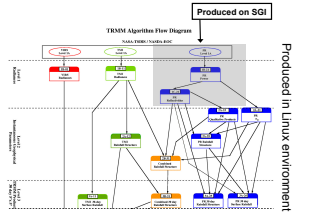






TRMM Version 6 Precipitation Retrieval Algorithms in the PPS-- System

PPS-- will use a hybrid processing system combining SGI (big-endian) and PC-Linux (little-endian) computers. TRMM version 6 algorithms were intended for a big-endian platform and took advantage of some non-portable language/compiler features. This has presented porting challenges. In addition, Linux machines in PPS-- are 64-bit environments whereas the SGI is a 32 bit platform. The TRMM Version 6 algorithms have been ported to Linux with the help of the TRMM algorithm developers. In order to preserve the integrity and continuity of the TRMM Version 6 retrieval results, only code changes critical to porting were made. Several months of test data have been generated in the PPS-- system to determine that no biases or errors are introduced into the TRMM retrievals. The results of comparisons between the standard TRMM products produced in TSDIS and the test products produced in PPS-- are presented below.



3B42, 3B43 and 3G68 will be produced on Linux cluster.

*The Level-1 PR algorithms will be ported to Linux by JAXA. When this occurs, all algorithms will be run in the **PPS--** Linux cluster environment. As of now the 1A21, 1B21 and 1C21 products in **PPS--** are identical to the ones produced in TSDIS and there is no direct impact to algorithms reading these products.*

Floating point calculations

The TRMM algorithms perform floating point calculations which have inherent porting issues. Rounding can be machine dependent and/or compiler dependent as in the case of FORTRAN. This has implications for any decision constructs based on floating point values such as thresholding, which then impact the final values of retrieved quantities, including rainfall rate.
- Gridding with floating point latitude and longitude.
- Scaling of values written into HDF products.
- Algorithm decision trees:
  - Rain/no-rain detection.
  - Surface type detection (land/ocean/coast).

Geolocation

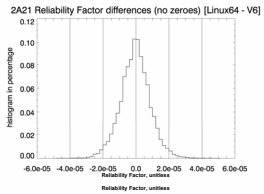
The IFOVs of each instrument are geolocated onto the Earth ellipsoid at Level-1. Latitude and longitude are stored as single precision floating point values in the HDF products. While the changes are small and at the limit of single precision floating point accuracy, these differences can have an impact on the final retrievals.

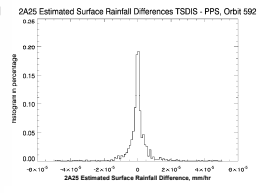
Orbital maximum difference Latitude: ± *3.0x10^{-6} degrees*
Orbital maximum difference Longitude: ± *1.5x10^{-5} degrees*

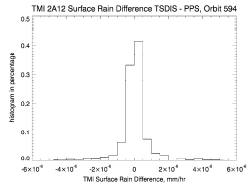
Orbit Boundaries are based on a floating point representation of time. Over the period of 1511 orbits, orbit boundary times were different by 1 msec in 128 orbits. 20 orbits had orbit boundary times different by 2 msec. This results in some scans being placed in different orbits between the two systems. No scans are lost.

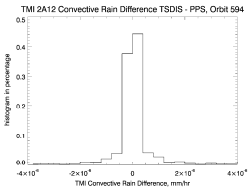
Instantaneous Pixel by Pixel Comparisons

Examples of some level-2 instantaneous comparisons. Cases where the TSDIS and PPS-- data agree are not shown in the figures.









DIFFERENCE	2A21.070901.55799.6.HDF - 2A21.070901.55799.6L.HDF							
	Frac	Data1 at	Data2 at	Frac	Frac	
Name	MaxDiff	MaxDiff	MaxDiff	MaxDiff	MeanDif	#NonZero	%NonZero
geolocation	0.00	0.	-36.2	-36.2	0.00e+00	0	0.e+00
sigmaZero	1.00	0.02	65.0	66.0	8.42e-07	140	3.e-04
pathAtten	1.00	0.02	61.0	62.0	1.12e-05	12	3.e-05
reliabFlag	0.00	0.	1.99e+04	1.99e+04	0.00e+00	0	0.e+00
reliabFactor	0.000248	6.e-06	44.3	44.3	2.55e-10	15836	3.e-02
incAngle	0.00	0.	182.	182.	0.00e+00	0	0.e+00
rainFlag	0.00	0.	0.00	0.00	0.00e+00	0	0.e+00
TOTAL						15988	

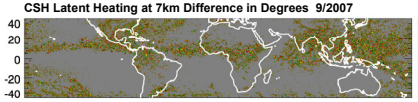
Example of an orbit comparison of two 2A21 (PIA, surface reference technique) products between the two systems. Note the surface cross section and path integrated attenuation are scaled to integers in the HDF products. A difference of 1.0 indicated rounding in the last significant digit for those fields.

Some algorithms are more sensitive to rounding effects. Combined algorithms that match IFOVs from the different instruments and those that depend heavily on surface type can produce outliers. In some cases the passive retrieval will determine rain over ocean but make no retrieval if the surface is changed to coast when the geolocation pushes the IFOV into another grid box in the surface type reference data.

Maps

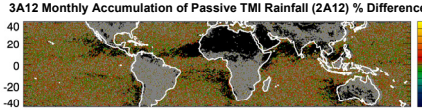
Maps of differences or % differences (TSDIS-PPS)/TSDIS*100.0 for various monthly, gridded averages. Grey indicates a value close to zero in the middle of the color scale, while **black** indicates no data (no rain). Positive differences are **red**/**yellow** while negative differences are **green**/**blue**.

CSH Latent Heating at 7km Difference in Degrees 9/2007



Max Difference: 5·10^{-5} degrees
Mean Difference: 7·10^{-10}degrees
Max % Difference: 0.1%
Mean % Difference: 1·10^{-6}%

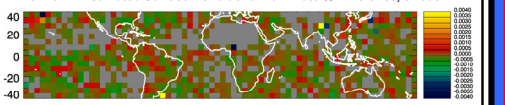
3A12 Monthly Accumulation of Passive TMI Rainfall (2A12) % Differences, 2/1998



Max Difference: 0.0015 mm/hr
Mean Difference: 3·10^{-4} mm/hr
Max % Difference: 0.04%
Mean % Difference: 4·10^{-6} %

Max Difference: 0.007 mm/hr
Mean Difference: 1.5·10^{-3} mm/hr
Max % Difference: 6%
Mean % Difference: 5·10^{-4} %

3A25 PR Estimated Surface Conditional Rain Rate % Difference, 3/1998



Monthly Difference Maps indicate very little impact from the new system. Grid boxes with larger differences result from a combination of thresholding and low measurement counts.

Monthly zonal means of unconditional rainfall rate from Level-3 products. Zonal Means indicate very little impact from the new system as well. Max % difference = 0.03%

mm/hr	2A25 (PR)		2A12 (TMI)		2B31 (Combined)		3A11 (TMI)	
YYMM	TSDIS	PPS--	TSDIS	PPS--	TSDIS	PPS--	TSDIS	PPS--
9801	0.118140	0.118140	0.123861	0.123861	0.125406	0.125370	0.122688	0.122687
9803	0.110881	0.110881	0.121790	0.121790	0.120065	0.120065	0.122172	0.122175
9804	0.113370	0.113370	0.118595	0.118595	0.122054	0.122049	0.115658	0.115658
0709	0.105343	0.105343	0.109662	0.109662	0.120338	0.120347	0.104491	0.104490
0710	0.103299	0.103299	0.109844	0.109844	0.118124	0.118099	0.108366	0.108366

Summary

The TRMM algorithms are performing well in the PPS-- system with no indications of errors or biases. We will continue to process and analyze test data in preparation of the PPS-- system assuming TRMM Version 6 processing duties. The exercise of porting the TRMM algorithms to Linux has provided valuable experience for both the algorithm developers and data system. This experience is beneficial for the continued development of the TRMM and GPM retrieval algorithms.

Authors acknowledge the contributions of many individuals involved in PPS from NASA Goddard, George Mason University, RSIS and as always, the NASA Precipitation Measurement Missions Science Team and support from NASA HQ.