A-Train Data Depot: Integrating and Exploring Data Along the A-Train Tracks

G. Leptoukh, S. Kempler, P. Smith
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771
Gregory.Leptoukh@nasa.gov

A. Savtchenko, R. Kummerer, A. Gopalan, J. Farley
RSIS / NASA GSFC
Greenbelt, Maryland 20771

A. Chen
George Mason University /
NASA/GSFC
Greenbelt, Maryland 20771

Abstract—The immense potential for new science findings as a result of inter-instrument data analysis has led to the development of a new data portal at GSFC: the A-train Data Depot. The power and utility of this new service to the general public is amplified immensely when the archived data are used in conjunction with online data analysis services like Giovanni. This presentation details some of the challenges of data usage from multiple distinct missions and how the tool sets we have developed can help to overcome these challenges, considerably cut down on analysis overhead and promote science exploration in an otherwise very challenging arena.

Keywords-component; Remote sensing, Data tools, A-Train, ATDD, Value-added services, CloudSat, CALIPSO, MODIS, MISR, Terra, Aqua, Giovanni

I. INTRODUCTION

The succession of US and international satellites that follow each other, seconds to minutes apart, across the local afternoon equator crossing is called the A-Train [1]. In order of equator crossing, these satellites are: OCO, Aqua, CloudSat, CALIPSO, PARASOL, and Aura all in the same sun-synchronous orbit (Fig.1). Flying in such formation increases the number of observations, validates observations, and enables coordination between observations, resulting in a more complete "virtual science platform". However, to get the full benefit of data from various A-Train missions, it is necessary to co-register (vertical/horizontal) or regrid datasets. This can be a daunting task with an extensive overhead in time even for an experienced user. In the absence of a viable and readily available tool, scientists must individually allocate much of their time and resources acquiring A-Train related datasets

residing at various locations, developing algorithms to match up and graph datasets along the A-Train track, and search through large amounts of data for areas and/or phenomena of interest. The aggregate effort expended on performing and repeating these tasks could climb into the tens of millions of dollars; this entire overhead is necessary before any real science can usually be performed. Our solution at GSFC to alleviate this overhead is the A-Train Data Depot (ATDD).

The goal of the ATDD project is to create the first ever A-Train virtual data portal/center, to process, archive, access, visualize, analyze and correlate distributed atmosphere measurements from various A-Train instruments. ATDD brings together data hosted at different and remote centers so that they can be combined to create a consolidated vertical view of the Earth's Atmosphere along the A-Train tracks. Whereas access to data from instruments with relatively narrow fields of view is easily facilitated, the ATDD provides subsets of data from wider field of view (FOV) instruments that only coincide with data from the narrower FOV instruments. Narrowing the data field is generally an essential first step for cross instrument data analysis. Making use of the online analysis tools available at ATDD portal, the need to transfer data to the scientist's site is also minimized. The innovative approach of analyzing and visualizing atmospheric profiles
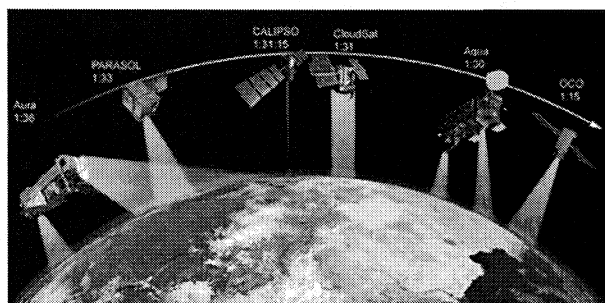

Figure 1. An approximate schematic reflecting the succession of the platforms in the A-Train formation. Numbers indicate approximate local Equatorial crossing time.

TABLE I. AVAILABLE COLLOCATED MODIS SUBSETS IN A-TRAIN DATA DEPOT

http://disc.gsfc.nasa.gov/atdd http://disc.gsfc.nasa.gov/atdd/data_products.shtml	
MODIS-CloudSat collocated subsets	*MODIS-MLS collocated subsets*
±5, ±100-km swath	±100-km swath
MODIS geolocation; 1- and 0.25-km calibrated radiances; and all Atmospheres products: aerosol, total column water vapor, clouds physical and optical properties, atmospheric profiles, and cloud mask.	MODIS geolocation; and all Atmospheres products: aerosol, total column water vapor, clouds physical and optical properties, atmospheric profiles, and cloud mask. Radiances are not subset

along the platforms track (i.e., time) is accomplished by bringing together data from Aqua (MODIS, AIRS, AMSR-E), CloudSat and CALIPSO (CALIOP, IIR), as well as the Aura (OMI, MLS, HIRDLS, TES) to create a consolidated vertical view of the Earth's Atmosphere along the A-Train tracks (Table 1).

II. ATDD CAPABILITIES

Data residing at the ATDD is archived on-line for fast data access, using the Simple, Scalable, Script-based Science Processor for Archive (S4PA) data management system. S4PA has proven to be an efficient tool for quick data access in its previous uses. In addition, the Mirador data search and access tool will enable users to find specific data of interest. The ATDD will continue to evolve as A-Train datasets are added and services are implemented in response to science needs. As a data service portal oriented toward providing Atmospheric Scientists with a central interface for the retrieval, analysis, and visualization of A-Train data, new enhancements and capabilities will be added to make this service user friendly with access to tools that facilitate research and exploration. ATDD provides the following capabilities to users:

- Provide a portal with services to facilitate the effortless access to ATDD data using simple graphical user interfaces (GUI) as in Fig 2. ATDD thereby intends to provide the community with a "one-stop shopping" location for all A-Train related information, even for sensor data stored at locations other than GES DISC.

- Provide tools to subset, process and visualize A-Train sensor data along with the basic temporal and spatial correlations.

- Provide "subscriptions" to operational users - users will receive notification when new data files are available.

- Promote collaboration among scientists using the A-Train sensors and be responsive to the Atmospheric community needs.

In atmospheric retrieval algorithms, common assumptions are made for the lack of observational data for a particular retrieval area. Lacking observations on clouds, humidity, temperature, etc., the actual operational algorithms rely either on climatologic or model data. In some cases, these can be far off from real conditions, making retrieval results quite
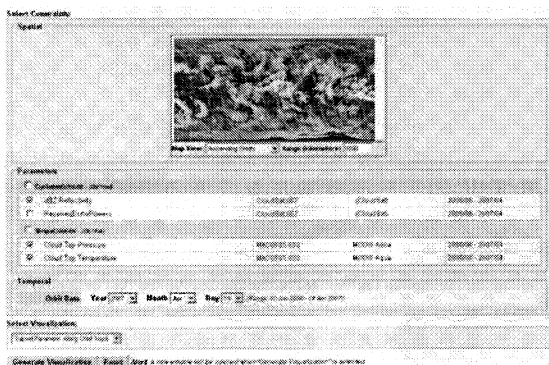
inaccurate. On the other hand, if the actual information obtained by the A-Train sensors could be used, the results will be much more accurate. For example, in areas close to the A-Train tracks, it is possible to use the CloudSat information on cloud properties (cloud top pressure, cloud phase, number and position of layers, etc.), or aerosol vertical position from CALIPSO (where it is critical to know location of aerosol layer relative to cloud layer), or humidity and temperature from AIRS. Even in between tracks using interpolation, better understanding of the actual status of the atmosphere can be achieved using A-Train data directly or in conjunction with models. ATDD therefore attempts to fill these gaps by providing data and services to promote such research.

III. DATA VISUALIZATION WITH GIOVANNI

One of the most important analysis capabilites of the ATDD is fullfilled using the GES DISC online visualization tool Giovanni [2]. Giovanni is a Web-based graphics utility that can display co-located ATDD data sets. A user is able to select a spatial area "box" for a desired region via a Java image map applet or manually enter coordinates defining the bounding box for data of interest, the temporal range for the data, or the parameters from a data set to plot along with desired output type.

For ATDD, the latest version of Giovanni (G3) is used to select various A-Train swath data (MODIS, AIRS, and OMI) collocated with CloudSat, CALIPSO and MLS/Aura ground tracks. Where available, vertical profile data from MODIS, AIRS, CloudSat, CALIPSO, and MLS are previewed as curtain plots. As an illustration of ATDD comparison obtained using Giovanni, the MLS humidity relative to ice (RHI), and CloudSat dBZ reflectivity produced in ATDD using Giovanni are shown in Fig.3. The CloudSat data shows what appears to be a polar nimbostratus with anvil-like wing aloft (the MLS cutoff is at 316 mb for data quality). The RHI shows oversaturation above that cutoff line, likely resulting from the

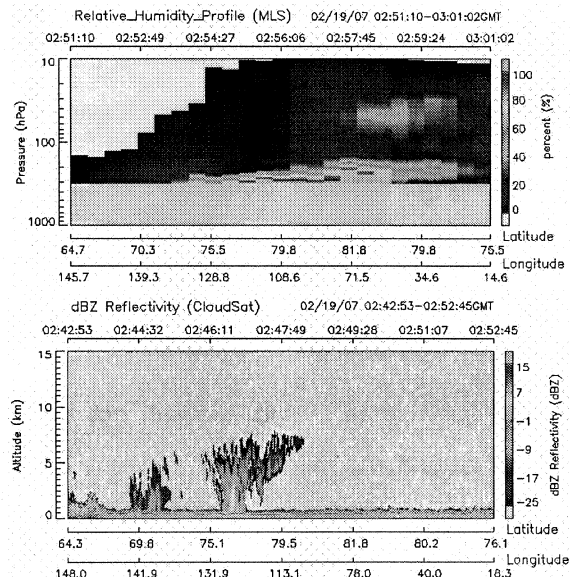Figure 2. User interface to the new Giovanni tool.

Figure 3. Example of MLS humidity relative to ice (RHI), and CloudSat dBZ reflectivity plots produced by Giovanni.
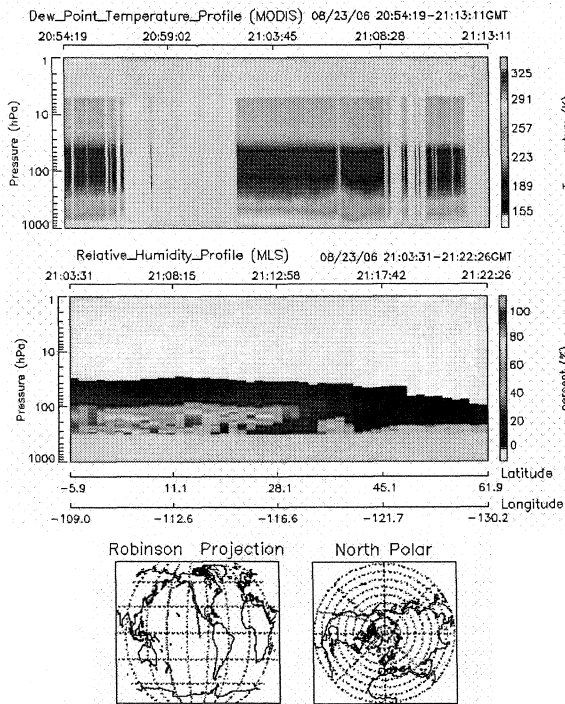
Figure 4. MODIS/Aqua and MLS transect of Hurricane Ileana, and maps showing the coverage of the particular portions of the ground tracks.

moisture due to the same source, which would appear as thin polar cirrus-ice crystals). While CloudSat penetrates through these clouds without getting distinguishable response, MLS on the other hand, wouldn't be able to resolve the horizontal and vertical structure of the cumulonimbus and MODIS does not retrieve atmospheric profiles in cloudy regions. Fig. 4 and 5
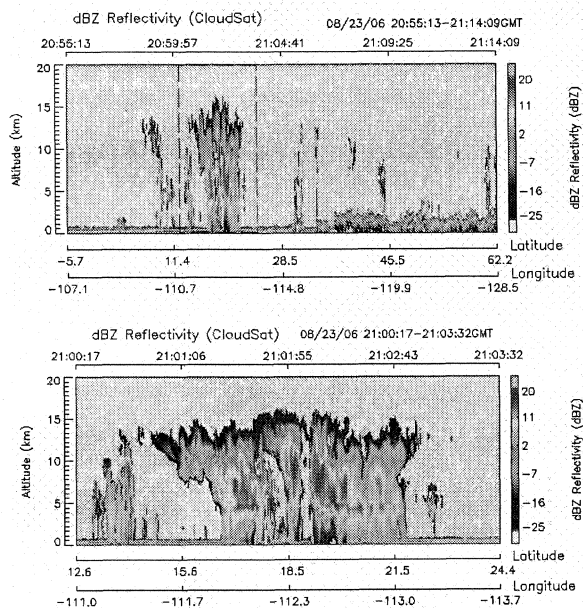

Figure 5. CloudSat transect of Hurricane Ileana (top). Bottom panel corresponds to the zoomed portion (indicated by the dashes) in top image.

show additional ATDD visualization pertaining to Hurricane Ileana (Aug 23, 2006).

IV.   ATDD GOOGLE EARTH

Google Earth has drastically increased the level of popular interest in remote sensing data with added expectation and demand on the types of data available. Google Earth (GE) provides a unique way to interactively view "curtains" of vertical profiles from CloudSat and other A-Train sensors. In addition, it also provides an option to overlay curtains from different sensors.

With the highest resolution of CloudSat orbit -- 5-second interval orbit data, the final KMZ (the zip format of KML) file for one hour CloudSat data is less than 1 MB.

The operational processing of CloudSat into Google Earth involves processing the vertical CloudSat data at 45-second intervals through Giovanni, and then chopping into 15-second vertical data images. The actual location of curtains in GE is precisely calculated based on the CloudSat data geolocation coordinates. Fig 6 illustrated how the hurricane Ileana (shown via Giovanni in Figs 4 and 5) can be seen in Google Earth.

V.   GEOMETRICAL CONSIDERATIONS

To understand the impact of spatially distinct measurements taken from different orbits of the various A-Train platforms, user requires sophisticated tools to match up the datasets to extract science results properly - more detailed discussions of the type of problems encountered can be found in [3]. To illustrate the spatial differences between A-Train satellite tracks, and considerations that must be accounted for differing horizontal resolutions consider Fig. 7 showing a MODIS, AIRS, and CloudSat collocation with their respective footprints. ATDD relieves the users from performing the time intensive task of collocating data by, for example, providing MODIS subsets that are already collocated with CloudSat and
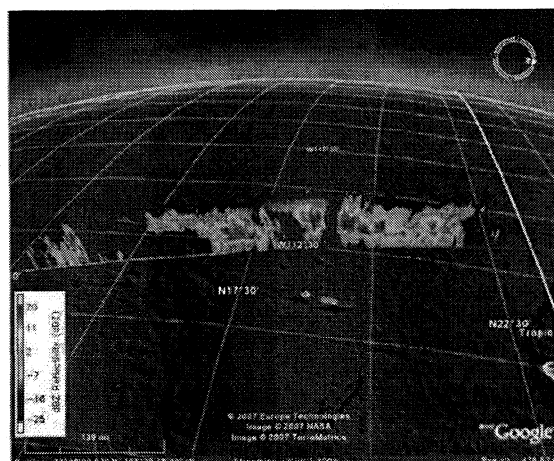

Figure 6 Google Earth rendering of Hurricane Ileana from the same CloudSat vertical profile shown earlier.
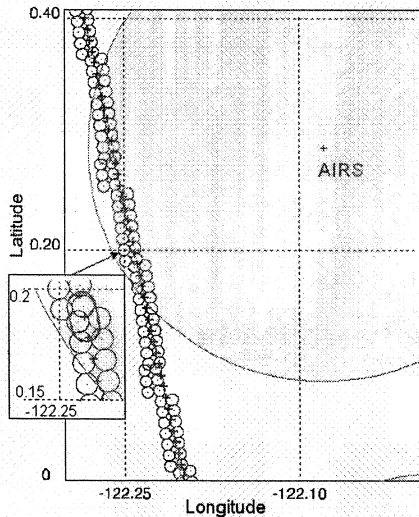
Figure 7. Comparative example of MODIS, AIRS, and CloudSat collocation, and relative sizes of the footprints of the corresponding retrievals. All proportions are preserved. Crosses show locations of CloudSat time stamps of retrieved profiles. The bean-shaped form is the resulting CloudSat surface footprint. Thus, consecutive profiles have roughly 50% overlap. The circles show the closest pair of pixels from the collocated 1-km radiances.
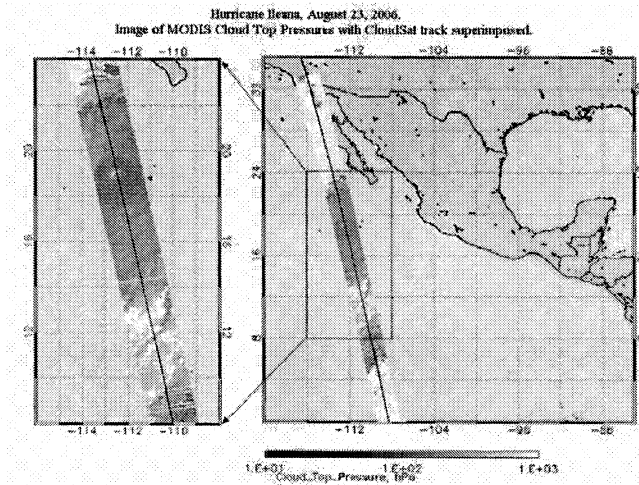

Figure 8. An example of the spatial coverage of MODIS/Aqua swath subsets collocated with CloudSat (center dark line). Available subsets are collocated with CloudSat and MLS (Table 1). CloudSat flew exactly over hurricane Ileana, August 23, 2006.

MLS. Fig. 8 shows the MODIS cloud top pressure subset, with overlaid CloudSat track for Hurricane Ileana. One can easily see the hurricane eye from this horizontal MODIS strip, which can be then associated with the gap in the CloudSat profile.

VI.   FUTURE ATDD DEVELOPMENT

Based on multiple suggestions from the atmospheric community, ATDD in the future will try to include additional datasets from other A-Train sensors: profile data from the Tropospheric Emission Spectrometer (TES) onboard Aura; aerosol data from the Polarization and Directionality of the Earth's Reflectances (POLDER) onboard PARASOL; and later extend to data from the Orbiting Carbon Observatory (OCO).

As a result, the atmospheric community will be able to get access to collocated and mutually compatible data from the whole suite of A-Train sensors, thus allowing the most comprehensive studies of the atmospheric properties.

ACKNOWLEDGMENT

REFERENCES

[1]   Graeme Stephens et al, 2002. "THE CLOUDSAT MISSION AND THE A-TRAIN: A New Dimension of Space-Based Observations of Clouds and Precipitation", BAMS, 83, pp 1771–1790

[2]   J.G. Acker and G. Leptoukh, 2007. Online Analysis Enhances Use of NASA Earth Science Data, EOS, Transactions of American Geophysical Union, 88, 14

[3]   Andrey Savtchenko, Robert Kummerer, Peter Smith, Steve Kempler, Greg Leptoukh, 2007. "A-Train Data Depot- Bringing Atmospheric Measurements Together," these proceedings