WEB SERVICES FOR CUSTOM LEVEL 2 AND LEVEL 3 DATA SUMMARIZATION OF NEWS MERGED A-TRAIN DATA

Hook Hua, Eric Fetzer, Steven Lewis, Mathew Henderson, Alexandre Guillaume, Seungwon Lee, Manuel de la Torre Juarez, Van Dang, and Amy Braverman
Jet Propulsion Laboratory/California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109

1. INTRODUCTION

Scientists routinely perform conditional subsetting and summarization on Earth science Level 2 data (geophysical state derived from directly observed radiances) in order to capture the trends and correlations hidden in the voluminous data and to feed the data into their global models. A NASA Energy and Water cycle Study (NEWS) project [1] is creating a merged Level 2 data set containing information from multiple instruments of NASA's A-Train satellite constellation in a common spatial and temporal resolution. However, the immense volume and inhomogeneity of Level 2 data from a large number of instruments restrict the analyses to small-scale and short-term scopes. Level 3 data, where geophysical parameters that have been spatially and/or temporally summarized from Level 2 data, are currently produced in advance and have limited selections so that a majority of scientists develop their own tool to customize Level 3 data from Level 2 data. This results in redundancy of work in the community and the dwindling use of Level 3 data available in NASA's distributed active archive centers.

Existing data access approaches require manual FTP-like or web portal interaction and therefore force the scientist to break away from their familiar data analysis environments. To simplify access to large and complex satellite data sets for climate analysis and model verification, we developed service-oriented tools that are used to study long-term and global-scale trends in climate, water and energy cycle, and weather variability. The services and tools leverage the NEWS Level 2 data to enable creation of climatologies that include correlation between observed temperature, water vapor and cloud properties from the A-Train sensors. Our services enable scientists to customize the conditional subsetting, averaging, and masking of the voluminous NEWS Level 2 data and the production of Level 3 data using a customizable Level 3 data reduction technique. An OpenSearch service was also developed to enable spatial, temporal, and full-text search of the data. We are also collaborating with GSFC's "Mirador Earth Science Data Search Tool" [2] at the GES DISC to provide a federated search capability of our data sets from Mirador via our OpenSearch service interface.

1.1 NEWS Level 2 Data

The NEWS A-Train project led by Dr. Eric Fetzer (also a Co-Investigator of the work presented here) has been generating a merged Level 2 product containing information from the A-Train instruments in a common spatial and temporal resolution. The A-Train consists of a series of satellites flying in formation crossing the equator

within a few minutes of one another in the "Afternoon" around 1:30pm local time. The Level 2 data from the merged observations of A-Train constellation of satellites is very large and we are not aware of other reasonable approaches to formally summarize the data as the one presented here. The NEWS Level 2 data contains temperature, atmospheric water vapor, cloud fraction, cloud-top temperature and pressure, cloud liquid water, and cloud water ice data from several A-Train instruments and CloudSat. Temperature field data are taken from Atmospheric Infrared Sounder (AIRS) and Microwave Limb Sound (MLS) Level 2 data. Atmospheric water vapor data are also obtained from AIRS and MLS. Cloud fraction and cloud-top temperature and pressure data are taken from AIRS and Moderate Resolution Infrared Sounder (MODIS). Cloud liquid water data are collected from Advanced Microwave Sounding Radiometer for Earth Observing System (AMSR-E). The sources of cloud water ice data are Advanced Microwave Sounder (AMSU), AMSR-E, MLS, and AMSU-B. Cloud ice and liquid water content data are also obtained from CloudSat. The physical parameters from A-Train instruments and CloudSat are merged and placed on a common, regular, nested spatial grid through appropriate interpolation schemes depending on the quantity. The errors in each data source and the mapping are estimated. We leveraged the NEWS merged data of the A-Train sensors to enable the science objective of creating climatologies that include correlations between observed quantities.

2. APPROACH

Instead of imposing on the user an often rigid and limiting web-based analysis environment, we recognize the need for well-designed distributed services so that users can perform analysis in their own familiar computing environments. We developed server-side processing capabilities necessary to reduce the voluminous NEWS A-Train data to be transferred to the client. Correspondingly, we developed client-side service APIs and utilities to enable scientists to perform analysis of voluminous server-side data from within their own familiar computing environment (command-line, Matlab, IDL, Python, Java, and C/C++). As shown in Figure 1, these targeted services aim to bring seamless data interconnectedness back into their multiplatform working environments. Developing distributed components in a multitude of programming environments commonly used by scientists enables the streamlining of scientific data processing and analysis by minimizing the manual intervention previously required and bringing the new capabilities into the scientist's familiar working environments. These

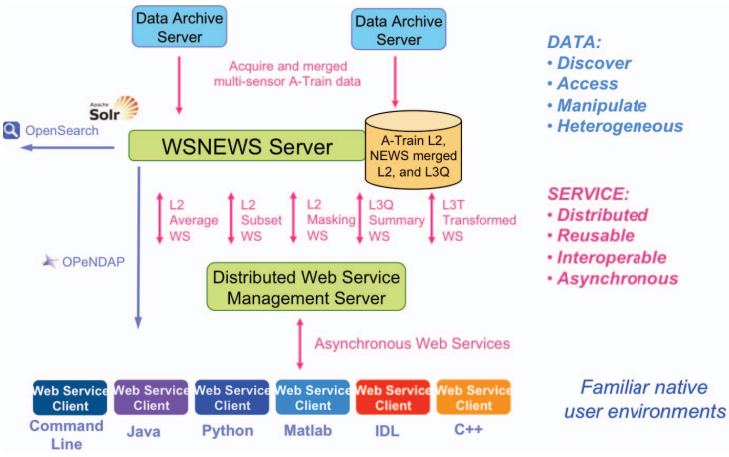

Figure 1. Overview of the distributed system for accessing and manipulating the NEWS merged A-Train Level 2 data. Multiple web service clients enable users to perform analysis directly within their familiar computing environments.

services can be used to analyze A-Train sensor data in order to investigate long-term and global-scale trends in climate, water and energy cycle and weather variability. Previously there did not exist a capability to easily discover and access data from the A-Train's multiple instruments as merged multi-parameter data sets.

The set of data products are being generated either offline or by on-demand services and delivered seamlessly to users. Our server hosts the services and has direct access to a large volume of staged data. Processing and summarization of this long-term and global-scale data on the server avoids needless and repeated downloading of massive amounts of data by various clients to perform potentially the same processing.

The A-Train set of Level 2 instrument data are being merged offline to create a merged product that preserves the relationship of observed atmospheric water properties. An on-demand averaging service is offered to create custom global-scaled averages as NetCDF products. The customization enables creation by temporal constraints (yearly, monthly, weekly, daily, and hourly averages) as well as a user-selected set of parameter measurements. An on-demand subsetting service was also developed to create custom subsets of the merged Level 2 data matching the temporal, spatial, and parameter constraints of the user. Similar to the subsetting service, we also developed a unique masking service that enables geographical and parameter subsets that match frequently used geographical regions such as the North American continent or northern/southern Pacific ocean.

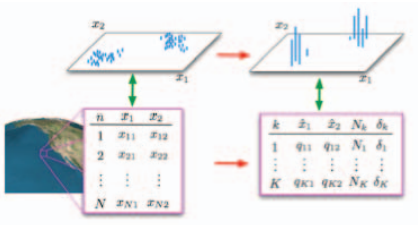

Figure 2. Raw and summarized data for one grid cell from applying statistical clustering methods.

The NEWS A-Train data record created was specifically designed to preserve the critical instantaneous relationship between highly variable atmospheric water components. However, no methodology has been previously developed to comprehensively characterize the variability, and covariability, of global satellite observations from any source. We addressed this shortcoming by applying a well-established statistical summarization technique to this data set, and describe multi-year behavior of the A-Train observations. We applied statistical clustering methods to a multiple-parameter set of observations from the A-Train instruments over the multi-year record (Figure 2). This data reduction replaces a large number of individual data points with a smaller number of representative data points and associated weights and quality measures. The resulting Level 3 quantitative summaries are made accessible through our service tool.

3. CONCLUSIONS

To achieve the science research goal of investigating long-term and global-scale trends in climate, water and energy cycle, and weather variability, we developed services and tools for discovering, accessing, and manipulating of NEWS merged A-Train data sets. The data product is a formal model containing data from the AIRS, AMSR-E, MLS, MODIS, and CloudSat instruments. Previously, scientists wanting to perform long-term and global-scale studies encompassing simultaneous measured quantities would quickly face a data access hurdle of first finding the data, then manually downloading them, and finally merging the data into a cohesive model—

all before starting their analysis. Additionally the voluminous nature of the data results in each scientist potentially downloading the same data resulting in redundancy of reprocessing on the client sides. Our paradigm pushes more of the commonly repeated processing onto the server side. Moreover, this avoids repeated downloading of the same data among the science users. We can deliver customized averaged, subsetted, masked, and summarized data of the NEWS merged A-Train observations directly into the scientist's working environment to do their analysis work. Our data reduction service enables analysis of long-term and global-scale climate where information about the covariability of multiple instrument observations, such as those from the A-Train, is preserved.

4. EXAMPLE RESULTS

Our services and tools have been used to create time-averaged data sets (Figures 3 and 4). Longer-term monthly averaged comparisons from different years have also been used to compare with standard indices of El Nino-Southern Oscillation and the Quasi-biennial Oscillation in the equatorial band.

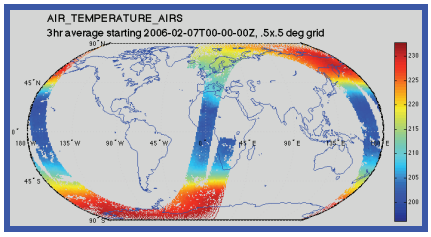

Figure 3. 3-hour average of AIRS temperature at NNN hPa shows orbit sampling and local variability.

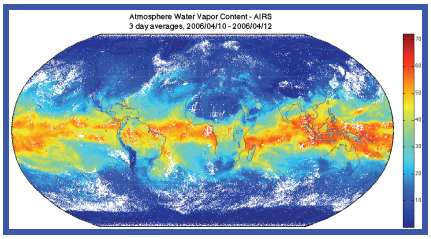

Figure 4. 3-day average of AIRS total water vapor shows global variability and local structure.

5. REFERENCES

[1] NASA and Energy and Water cycle Study (NEWS), website: http://www.nasa-news.org/
[2] Mirador Earth Science Data Search Tool, website: http://mirador.gsfc.nasa.gov/

6. BIBLIOGRAPHY

Hook Hua *is a member in the Instrument Software and Science Data Systems section at the Jet Propulsion Laboratory. He is the Principle Investigator of this project to develop services for use in exploratory analysis of Level 2 and Level 3 NEWS merged A-Train data, which is used to study long-term and global-scale atmospheric trends. He is also the Principle Investigator of a project to develop provenance capabilities to be used in the creation of a merged and multi-decadal climate data records. Hook was also the lead developer of an Observing System Simulation Experiments (OSSE) system that utilized distributed Web Services-based workflows of heterogeneous models to assess the HyspIRI hyperspectral remote sensing instrument model. Hook was also the lead in the development of an ontology knowledge base and expert system with reasoning to represent the various processing and data aspects of Interferometric Synthetic Aperture Radar processing. He has also been involved with Web Services and dynamic language enhancements for the Satellite Orbit Analysis Program (SOAP) tool. Hook received a B.S. in Computer Science and a B.S. in Applied Mathematics from the University of California, Los Angeles.*