

A PROCESSING SYSTEM TO PRODUCE BLENDED TOTAL PRECIPITABLE WATER PRODUCTS

Stanley Q. Kidder and Andrew S. Jones*
Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, Colorado

1. INTRODUCTION

The number of environmental satellites is growing exponentially. Many of these satellites measure the same variables, though in different ways. To understand the weather and climate, it is essential to blend data from many different sensors in real time into a unified product for each variable of interest. Such blending requires that new thought be given to the systems employed to do this processing.

In this paper we detail the Cooperative Institute for Research in the Atmosphere (CIRA) processing system for blending total precipitable water (TPW) data from the Advanced Microwave Sounding Unit (AMSU) instruments on three NOAA satellites with data from the Special Sensor Microwave/Imager (SSM/I) instruments on three Defense Meteorological Satellite Program (DMSP) satellites to form a unified product. The discussion includes data acquisition, hardware, software (including grid computing techniques and dynamic cross-sensor intercalibration), and product distribution and archiving. Nearly global, Mercator TPW composites are constructed hourly and made available in real time to forecasters using this system.

2. DATA ACQUISITION

Increasingly satellite data are acquired over the Internet, where the name and location of the data can change rapidly. We have developed a simple scripted system for File Transfer Protocol (FTP) access to data

- which is flexible and easily updateable, to keep up with changing addresses,
- which can be scheduled to run as often as needed,
- which never gets the same data twice, to minimize the burden on the supplying machine, and
- which automatically puts the newly acquired data in the proper place in a directory structure for ease of access by processing programs,

This data acquisition system, like all of the system described in this paper, runs unattended around the clock. It can be replicated as necessary for each new data source. Current sources include AMSU and SSM/I data from NESDIS, and AGRMET data from the Air Force Weather Agency (AFWA). While simple, the system is quite robust with a fault-tolerant monitoring system that can automate failover to other functional backup machines during adverse machine hardware conditions (e.g., hard drive failures and the like).

*Author addresses: CIRA/Colorado State University, Fort Collins CO 80523-1375, Stanley.Kidder@ColoState.EDU, Andrew.S.Jones@ColoState.EDU.

3. HARDWARE

The system hardware is composed of a cluster of PCs running the Windows Server 2003 operating system. One computer acts as the supervisor of the other members of the cluster. A second computer acts as the assistant supervisor; it automatically takes over when the supervisor becomes incapacitated. The other members of the cluster act as workers, performing tasks as directed by the supervisor. Currently there are seven PCs in the cluster.

The cluster is easy to modify. Members of the cluster are identified by a text file. Removing a member from the cluster or adding a member to it is as easy as adding or removing a line from a text file. Computers can be removed for repair or upgrades, and new computers can be added to provide additional computing power. Indeed, some of the members of the cluster can be desktop computers which are in current use for office applications. Spare computing cycles are thus utilized (Guch et al. 2003).

Both input and output files are mirrored on the disk drives in the cluster so that data are not lost if one disk drive fails. Total disk space on the cluster is approximately two terabytes.

4. SOFTWARE

The primary software used in this system is the Data Processing Error and Analysis System (DPEAS, Jones and Vonder Haar 2002). DPEAS is a dynamic, parallel data processing system for the merger and analysis of data from multiple satellite sensors. DPEAS was created to overcome the inherent difficulties of working with large volumes of multiple data formats. Among these difficulties are:

- Data from different satellite sensors come in different formats. Thus, different code must be written for each combination of sensors that is desired.
- Satellite data are voluminous both in total number of bytes and in the number of files that must be processed, including backup and archival.
- The computational burden is not uniform. Parallel processing of the data to avoid processing bottle necks is highly desirable.
- Recoding the system for each new application is far too costly and time consuming.

DPEAS has five main aspects designed to overcome these difficulties:

- A. The memory-resident data structure is HDF-EOS (currently HDF-EOS version 2.5). All data are translated on input into the HDF-EOS structure, and then processing continues. On output, a simple subroutine call writes the output data in HDF-EOS format.

Other output formats are accomplished with format translators. Therefore, processing code is independent of input or output data format.

- B. A large number of utilities are included in DPEAS for the handling of satellite data. Due to the common data structures, most of these routines are generic and can operate on many different satellite data types. This improves the reusability of the advanced satellite processing codes.
- C. DPEAS automatically assigns computational tasks to free nodes on a cluster of computers to parallelize the data processing.
- D. DPEAS has a number of fault-tolerant features to enable the parallel computing system to reroute data flows dynamically in the event of a hardware failure.
- E. DPEAS is run using a scripting language, which is a subset of Fortran 90 (F90). All operations are accomplished through subroutine or function calls. Thus the operational data processing is easy to monitor and change.

DPEAS itself is composed of about 250,000 lines of Fortran 90 and C/C++ code.

The software used to operate the “system” discussed in this paper, then, consists of the DPEAS executable, several DPEAS script files, which tell DPEAS which functions to apply to which data and where to place the output, and a scheduling program which runs the system according to a schedule.

5. COMPOSITING ALGORITHMS

Integral parts of DPEAS are routines which (1) “correct” data so that retrievals from many instruments appear to have come from a single instrument, and (2) form composite images from individual orbits of data. Details of these routines for making TPW composites have been discussed by Kidder and Jones (2004).

The first step in the correction process is constructing histograms of TPW values for a five-day period. A histogram is constructed for each satellite instrument at each scan angle. The assumption is that in a five-day period, each scan position of each instrument will observe the global distribution of TPW. We choose to use the average TPW PDF of the three AMSU instruments (scan positions 6–25 only) as the “correct” PDF, and we call this PDF “Truth.” Next, a polynomial correction is developed from the histograms which maps the observed PDF onto the truth PDF [see Kidder and Jones (2004) for details]. Figure 1 shows the results of applying the polynomial correction to one scan position of one SSM/I instrument.

After the correction process is developed, each orbit of each instrument to be composited is corrected and mapped to the appropriate map projection. Then the compositing routine is called.

Satellite data may be composited or blended in a variety of ways depending on the use to which the blended product is to be put. Perhaps the most common way to blend data is to average them over a specified time period. Another way to composite data is to overlay newer data on top of older data; only the latest data are

displayed. This method of compositing is favored by forecasters because it is the most up-to-date image possible. When the overlaid composite is constructed, we can optionally map the time of the most recent observation and the satellite which made it. These data are useful for analyzing the resultant TPW field.

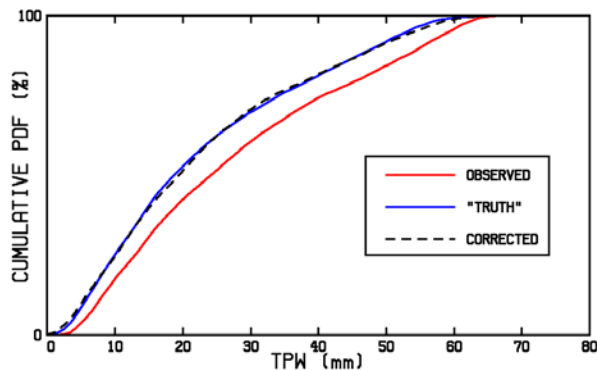


Figure 1. The dashed line shows the results of correcting the observed TPW data with the polynomial correction.

6. EXAMPLES

Figure 2 shows TPW values from a single orbit of DMSP F13. The TPW Values have been corrected by the polynomial correction procedure and mapped onto a Mercator map. These single-orbit map files are saved for several hours so that the compositing procedure can access them. Each orbit needs to be corrected and mapped only once.

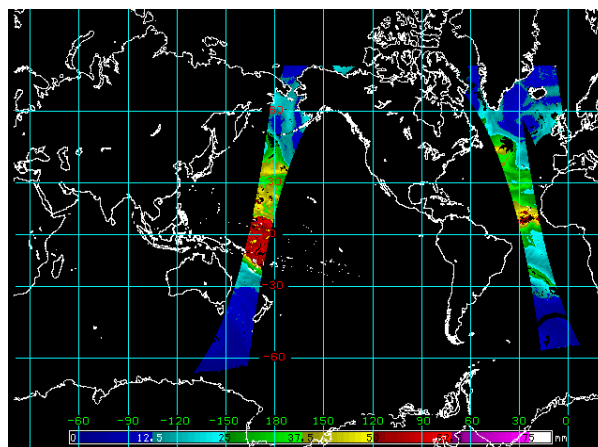


Figure 2. Corrected and mapped TPW values during one orbit of the DMSP F13 satellite.

Figure 3 shows the TPW from three AMSU instruments and 3 SSM/I instruments averaged for a 12-h period ending at 2030 UTC on 21 July 2004. Because data from six satellites are used in the composite, there are few places which are unobserved, which is the goal of compositing—one wants to know the water vapor field for the entire globe, not simply the field as observed by a single satellite in one orbit, as in Fig. 2. Thirty orbits went into this composite.

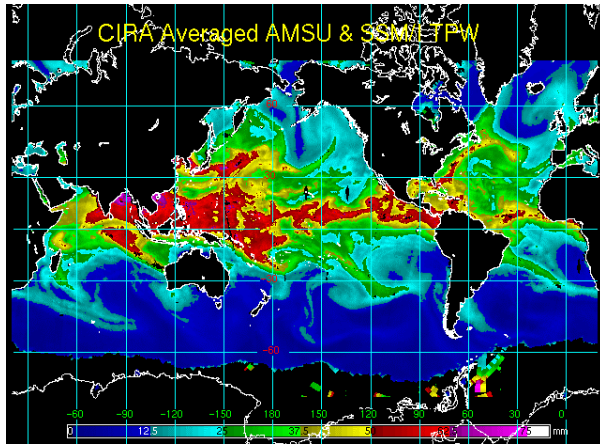


Figure 3. Average TPW for the 12-h period ending at 2030 UTC 21 July 2004.

Figure 4 shows the overlaid TPW composite for the 12-h period ending at 2030 UTC 21 July 2004 (In the overlaid composite, only the most recent datum at a point is shown.) The same 30 orbits went into this composite as into the averaged composite shown in Fig. 3. Note that the polynomial correction described in Section 5 has removed the “seams” in the data caused by different instruments observing adjacent locations. In other words, it is difficult to identify the individual orbits in the composite.

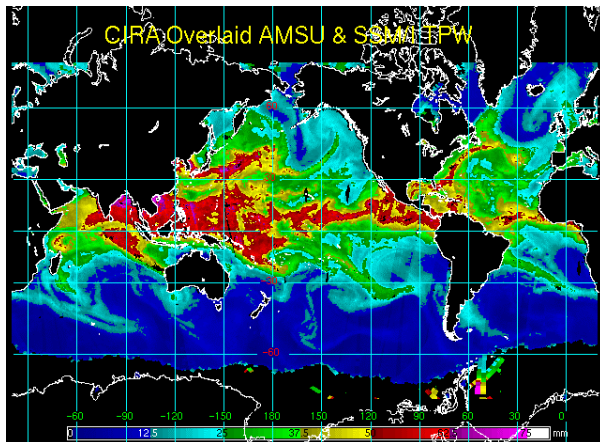


Figure 4. Overlaid TPW for the 12-h period ending at 2030 UTC 21 July 2004.

When constructing the overlaid composite (but not the averaged composite) it makes sense to map the time of the observation and the satellite which made it at the same time that one is constructing the composite itself. Fig. 5 shows the time of the observation in Fig. 4. The times are UTC to the nearest 10 minutes. Using the color bar (left is 0000 UTC, right is 2350 UTC), one can get an approximate idea of the time of observation. We use McIDAS to display the data, and with the IMGPROBE command, one can get the precise time (within 10 min) of each point. Figure 6 shows the satellite which made the observations in Fig. 4. Green points were observed with AMSU, purple points were observed

with SSM/I. Using the IMGPROBE command in McIDAS, one can discover which NOAA satellite (NOAA 15, 16, or 17) made the AMSU observations and which DMSP satellite (F13, F14, F15) made the SSM/I observations.

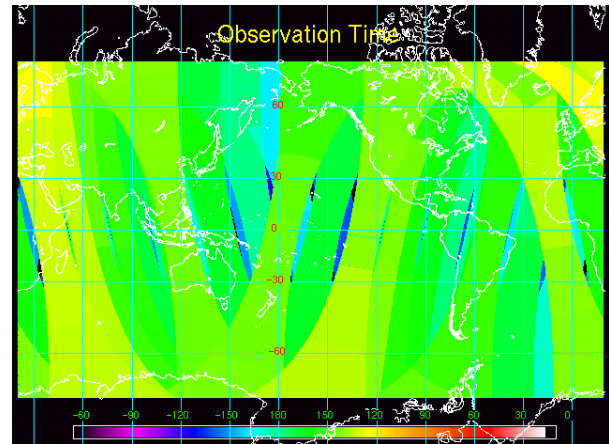


Figure 5. Time of latest observation for the composite shown in Fig. 4.

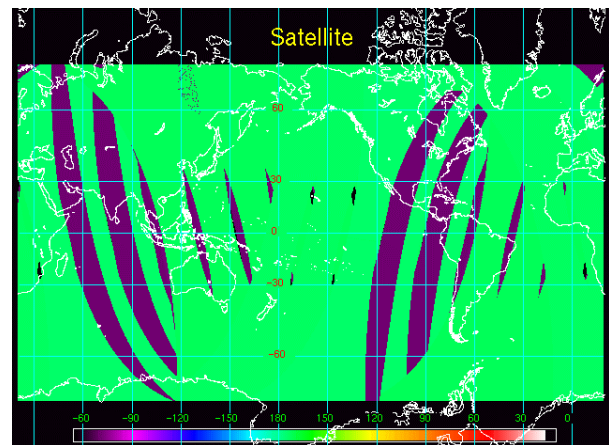


Figure 6. The satellite which made the observations plotted in Fig. 4.

7. PRODUCT DISTRIBUTION

Products are distributed by placing them in a directory where they are available by FTP or other file exchange protocol. Products may optionally be archived to DVD simply by placing them in an archive directory. As of this writing, these products are available as McIDAS Areas at <ftp://amsu.cira.colostate.edu/composites>. AREA0025 is the averaged composite; AREA0026 is the overlaid composite. AREA0026 contains the time and satellite maps as well as the TPW map.

8. CONCLUSIONS

Satellites offer the only way to observe the global distribution of some meteorologically important parameters. Forecasters need these parameters to make in-

formed forecasts. When constructing these products, it is important to remember that forecasters need products that are accurate, reliably produced, readily available, and are free of distracting artifacts. This paper has shown how we blend TPW observations from six satellites in real time to produce nearly global, hourly TPW analyses for use by forecasters at SSD and elsewhere.

Acknowledgements

We thank the NOAA/NESDIS Office of Systems Development (OSD) Product Systems Development and Implementation (PSDI) program and the NOAA High Performance Computing and Communications (HPCC) program for supporting this research. The AMSU and SSM/I satellite data sets for this work were obtained via the NESDIS CEMSCS.

References

- Guch, I. C., A. S. Jones, R. R. Ferraro, M. J. Kane, C. W. Karlburg, and S. Q. Kidder, 2003: Harnessing the spare computing power of desktop personal computers for improved satellite data processing and technology transition. *19th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Long Beach, California, 9–13 February.
- Jones, A. S., and T. H. Vonder Haar, 2002: A dynamic parallel data-computing environment for cross-sensor satellite data merger and scientific analysis. *J. Atmos. Oceanic Technol.*, **19**, 1307–1317.
- Kidder, S. Q., and A. S. Jones, 2004: Blended satellite products for operational forecasting. Extended Abstract CD: *13th Conference on Satellite Meteorology and Oceanography*, Norfolk, Virginia, 20–24 September.