A Blended Satellite Total Precipitable Water Product for Operational Forecasting

STANLEY Q. KIDDER AND ANDREW S. JONES

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado

(Manuscript received 30 November 2005, in final form 17 May 2006)

ABSTRACT

Total precipitable water (TPW), the amount of water vapor in a column from the surface of the earth to space, is used by forecasters to predict heavy precipitation. In this paper, a process for blending TPW values retrieved from two satellite sources is detailed: the Advanced Microwave Sounding Unit (AMSU) instruments on three NOAA satellites, and the Special Sensor Microwave Imager (SSM/I) instruments on three Defense Meteorological Satellite Program (DMSP) satellites. The process starts with a blending algorithm, which matches the cumulative probability distribution functions of TPW retrievals from the two instruments to lessen their differences. The data are then mapped to a map projection useful to forecasters and composited for 12 h to make a global map. These maps are produced hourly using Data Processing and Error Analysis System (DPEAS) software and made available to forecasters online.

1. Introduction

Forecasters today are faced with many sources of data. What they need is meteorologically significant data fields blended from all available data sources, not numerous maps of observations from individual sources. In this paper we detail our process for blending data for one such meteorological parameter, the total precipitable water (TPW), which is the amount of water vapor in a column from the surface of the earth to space (in kilograms per square meter or, equivalently, in millimeters of condensate). TPW is used by forecasters to predict heavy precipitation.

Two satellite sources are used in this TPW blending process, one from the Advanced Microwave Sounding Unit (AMSU) instruments on three National Oceanic and Atmospheric Administration (NOAA) satellites, and the other from the Special Sensor Microwave Imager (SSM/I) instruments on three Defense Meteorological Satellite Program (DMSP) satellites.

One might think that the blending process could be simply plotting TPW values from the two satellite sources on a map for the length of time desired (several hours, during which the earth is nearly completely sampled by the satellites). However, there are real differences between the instruments. AMSU is a cross-track scanner; it scans perpendicular to the satellite's velocity vector through nadir. The AMSU-A instrument scans in 30 steps (or scan positions), which results in 48 km $\times$ 48 km resolution at nadir and about 79 km $\times$ 148 km at the ends of the scan. TPW is retrieved (only over water) from the 23.8- and 31.4-GHz channels (Weng et al. 2003; Ferraro et al. 2005). SSM/I scans in a cone with the vertex at the satellite and the axis passing through nadir. Each observation is made at nearly the same earth-viewing angle, and each has the same size, effectively 25 km $\times$ 25 km. There are 64 scan positions on each SSM/I scan line for the frequencies used. TPW is retrieved (only over water) from the 19-, 22-, and 37-GHz vertically polarized channels in an algorithm that is quite different from that used for AMSU (Ferraro et al. 1996). These differences result in differences in the statistical properties of the retrieved TPW values (see below), and if one simply plots the retrieved values on a map, artifacts appear in the data that have nothing to do with TPW, but are distracting for forecasters. They wonder whether a feature in the map is a real TPW feature or was caused by the fact that it was observed by different satellites. An example of these artifacts is shown in section 4.

We developed a blending process for these two sources of TPW data, which makes them appear to be made from the same instrument and thus removes the

Corresponding author address: Stanley Q. Kidder, CIRA, Colorado State University, Fort Collins, CO 80523-1375.
E-mail: stanley.kidder@colostate.edu

DOI: 10.1175/JTECH1960.1

distracting artifacts. The process involves applying adjustments to ensure that the data from different satellites are compatible, mapping the data on an orbit-by-orbit basis to a convenient projection, and compositing the mapped data into a combined product in a suitable format for use in operations. In addition, the computer-processing environment in which the products are produced is discussed. Nearly global, Mercator TPW composites are constructed hourly and made available in real time to forecasters at the Satellite Analysis Branch (SAB) of the Satellite Services Division of the NOAA/National Environmental Satellite, Data, and Information Service (NESDIS) Office of Satellite Data Processing and Distribution (OSDPD) and online (see section 5), to forecasters everywhere, including those with the National Weather Service.

2. Blending algorithm

When one thinks of making corrections to data, one usually thinks of removing biases and, perhaps, adjusting the standard deviations. This works well for data that are normally distributed and for which there is a standard, that is, for which "truth" is known. TPW data from different satellite instruments do not fit this description. First, a truth dataset is not readily available; and, second, TPW retrievals from the two instruments differ in a statistically nonnormal way from each other. To solve this problem, we developed a technique to make the probability distribution function (PDF) of the SSM/I TPW data look like the PDF of the AMSU TPW data. We call this a blending algorithm.

The first step in the blending algorithm is the construction of histograms of TPW values for a 5-day period. A histogram is constructed for each satellite instrument at each scan position. The assumption is that in a 5-day period, each scan position of each instrument will sample the global distribution of TPW. Both NOAA satellites and DMSP satellites are in sun-synchronous, near-polar orbits, which means that they observe the whole earth. Each satellite makes about 14 orbits per day sampling the moist Tropics and dry extratropics on each orbit. At each scan position, AMSU makes about 30 000 ocean observations, and SSM/I makes about 50 000 observations in 5 days.

Let $n(i_{\mathrm{TPW}}, i_{\mathrm{SCAN}}, i_{\mathrm{SAT}})$ be the 5-day histogram, where i_{TPW} is the integral value of the retrieved TPW value in millimeters (range of 0–100), i_{SCAN} is the scan position (range of 1–30 for AMSU and 1–64 for SSM/I), and i_{SAT} is the index for the satellite (range of 1–6 for the six satellites used). The cumulative probability distribution function is defined as

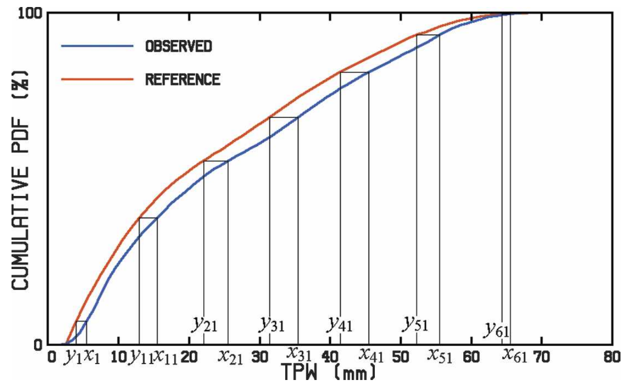

FIG. 1. Illustration of the blending algorithm.

$$\mathrm{PDF}(i_{\mathrm{TPW}}, i_{\mathrm{SCAN}}, i_{\mathrm{SAT}}) \equiv \frac{\sum_{j_{\mathrm{TPW}}=0}^{i_{\mathrm{TPW}}} n(j_{\mathrm{TPW}}, i_{\mathrm{SCAN}}, i_{\mathrm{SAT}})}{\sum_{j_{\mathrm{TPW}}=0}^{100} n(j_{\mathrm{TPW}}, i_{\mathrm{SCAN}}, i_{\mathrm{SAT}})}, \tag{1}$$

where, of course, PDF ranges from 0 to 1.

The second step in the process is constructing a reference PDF. While we do not know what the true TPW distribution is, we can choose one set of observations to be the reference and calculate an adjustment to be applied to all observations that will make the distribution approximate the reference distribution. We chose to use the average TPW PDF (scan positions 6–25 only) of the *NOAA-17* AMSU instrument as the reference PDF. We considered using the average of all AMSU instruments as the reference, but we decided that using a single instrument (then the newest of the three AMSUs) was cleaner than using the average because it will not change as the older AMSUs fail. We also considered, but decided against, using SSM/I as the reference in part because our grant was to study AMSU data and in part because we had no data to indicate that SSM/I TPW retrievals are more accurate than AMSU retrievals. This choice is more than adequate to test the concept of a blending algorithm, to see if it is useful to forecasters, which is our purpose in this study. If the blended product were to be used for other purposes, such as climate applications, however, the choice of a reference should be carefully reconsidered. The reference PDF for the 5-day period ending at 2215 UTC 29 March 2006 is shown in Fig. 1 by the red line.

The third step in the process is the construction of an adjustment for each scan position of each instrument (3 SSM/I instruments × 64 scan positions + 3 AMSU instruments × 30 scan positions = 282 adjustments). The blue line in Fig. 1 shows the cumulative PDF for scan

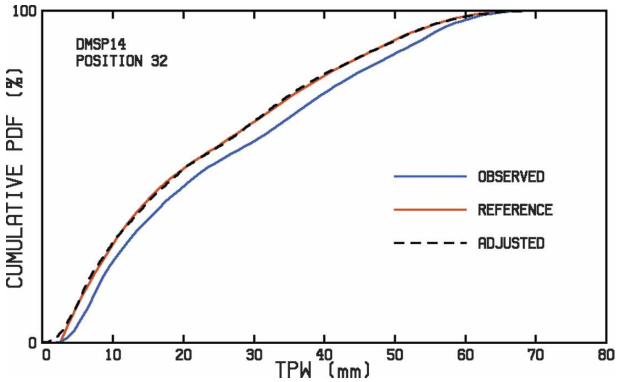

FIG. 2. The dashed line shows the results of adjusting the observed TPW data with the blending algorithm.

position 32 on the *DMSP F14* SSM/I for the time period above. (The reference PDF is the same for each of the 282 adjustments.) The TPW histograms are tabulated with 1-mm-width bins centered at 0.5, 1.5 mm, etc. For each bin from 5.5 to 68.5 mm (the 64 x_i values) a y_i is interpolated such that the observed cumulative PDF has the same value as the cumulative reference PDF. This step is illustrated in Fig. 1 for a subset of the x_i and y_i values.

Note that the SSM/I TPW values are generally higher than the AMSU TPW values, so the SSM/I values need to be adjusted downward to match the AMSU reference; however, the adjustment is not uniform. A larger correction is required in the middle range of TPW (where the "bars" in Fig. 1 are wider) than at either high or low TPW values. This adjustment cannot be accomplished accurately with a simple bias adjustment, which would simply shift the blue curve to the left, nor with a linear adjustment, which would shift the blue curve to the left and change its slope. We chose to fit a cubic polynomial to the (x_i, y_i) values. To form the adjustment, the 282 sets of four polynomial coefficients are calculated and stored for use during the adjustment procedure.

Applying the adjustment is a simple process of selecting the coefficients (as a function of satellite and scan position), using the observed TPW as x, and calculating y:

$$\text{adjusted TPW} = a_0 + a_1\,\text{TPW} + a_2\,\text{TPW}^2 + a_3\,\text{TPW}^3. \qquad (2)$$

Figure 2 shows the adjusted cumulative PDF (dashed line) for the data shown in Fig. 1.

This blending algorithm is quite robust; it will work as long as the retrieved TPW values are monotonic with the true TPW values. It does not produce an absolutely calibrated value, but it does bring one set of observations into agreement with another set. In effect, it makes one set of observations (i.e., *NOAA-17* AMSU TPW retrievals) a reference with which other observations must agree. The advantage to the forecaster of doing this is illustrated in section 4.

It must be noted that although we calculate the polynomial coefficients using TPW values between 5 and 69 mm, we apply them regardless of magnitude, and we clip the result to be within the AMSU TPW range, 0–75 mm. The high and low values of TPW, therefore, are less accurate than the center values. This is another area that deserves further study.

This blending algorithm is a dynamic algorithm. Suppose, for example, that the retrieval algorithm for one instrument is changed. In 5 days the blending algorithm will automatically adjust to the new reality.

We note that this blending algorithm is fundamentally different than a retrieval algorithm. The duty of the retrieval algorithm is to produce the desired parameter (i.e., TPW) as accurately as possible from the measured radiances. The job of the blending algorithm is to lessen the nearly inevitable differences between parameters retrieved with different algorithms from measurements made by different instruments. One would like the blending algorithm to combine the observations in a way that is closer to the truth. However, simply removing artifacts from the retrievals to aid the forecaster is useful even if increased accuracy cannot be claimed.

Since all of the satellites used in this study are sun synchronous and, therefore, make observations at the same local time each day, the diurnal variation of TPW is possibly an issue. We believe that the diurnal variation is small, and our blending algorithm assumes that it is zero, that is, each satellite views that same distribution of TPW regardless of the time of observation. Clearly, this deserves further study, as does the effect of varying field of view in the AMSU data on TPW.

We are grateful to a reviewer for pointing out that this blending algorithm is a type of histogram matching algorithm with roots in image processing (e.g., Richards and Jia 1999). Histogram matching has also been used to calibrate IR rainfall algorithms with passive microwave retrievals (Kidd et al. 2003) and to calibrate brightness temperatures between two different satellite instruments (Jones and Cecil 2006).

3. Mapping

Once the data have been adjusted, they need to be mapped for analysis. The base map that we use was chosen to be compatible with a map used by our colleagues at SAB. It is a Mercator projection with 16-km resolution at the equator. The map is centered at the

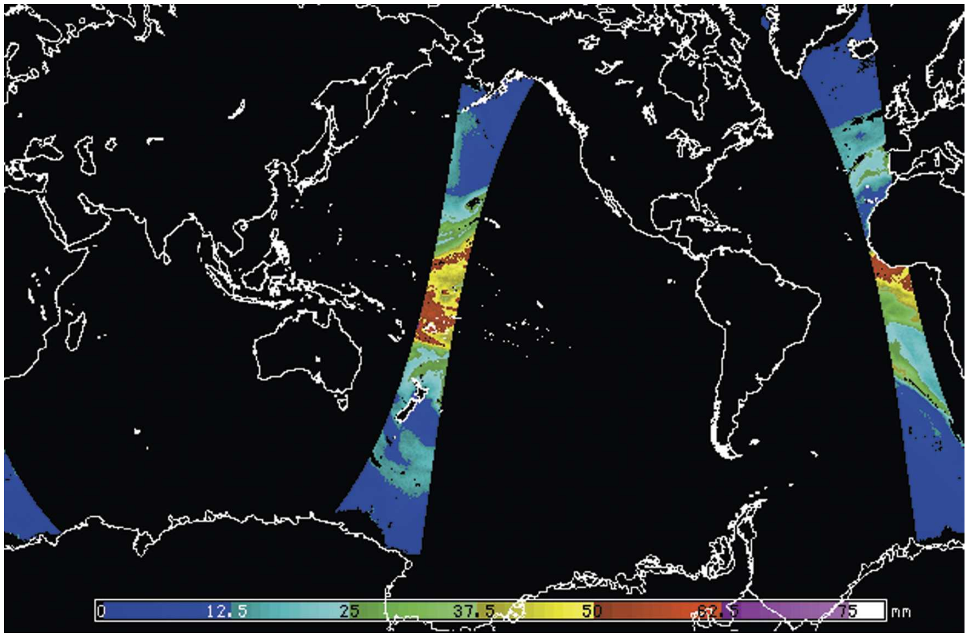

FIG. 3. Adjusted and mapped TPW values during one orbit of *DMSP F14*.

equator and 160°W. It has 1437 lines and 2500 elements, which covers the earth from about 71°N to 71°S. The cut line is at 20°E, which was chosen to emphasize ocean areas.

The TPW data that we receive from NESDIS are in files that represent approximately one orbit. Before mapping, the TPW value for each scan spot is adjusted with the blending algorithm as described in section 2. Then each scan spot is mapped by filling a quadrilateral that represents the scan spot. (Actually, the scan spots overlap somewhat, but we have ignored the overlap in this study.) The SSM/I quadrilaterals are 25 km $\times$ 25 km, and the AMSU quadrilaterals are 48 km $\times$ 48 km at nadir and approximately 79 km $\times$ 148 km at the edges of the scan. The quadrilaterals are contiguous both along the scan lines and from scan line to scan line. Thus, the resulting map has no holes within the scanned area. (However, because the microwave swaths do not overlap near the equator, there are gaps between adjacent swaths. Also, there are holes in the map over land and where it is raining.) Figure 3 shows TPW values from a single orbit of *DMSP F14*.

Each orbit of data is adjusted and mapped once, then used without reprocessing in as many composites as desired. We make a 12-h composite each hour, which includes about 30 orbits. Only up to six orbits (one for each satellite) are "new"; the remaining 24 (approximately) are accessed from disk files and composited again, which results in a considerable savings in computer processing.

In addition to mapping the TPW value, we map the time that the scan spot was observed and the satellite that observed it. When the single orbits are composited (see next section) these additional mapped values help interpret the data.

4. Compositing

Satellite data may be composited or blended in a variety of ways depending on the intended use of the blended product. Perhaps the most common way to blend data is to average them over a specified time period. Figure 4 shows the TPW from three AMSU instruments and three SSM/I instruments averaged for a 12-h period ending at 1929 UTC 29 March 2006. Because data from six satellites are used in the composite, there are few places that 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. 3.

Another way to composite data is to overlay newer data on top of older data, such that only the newest data are displayed. This method of compositing is favored by forecasters because it is the most up-to-date image possible. Figure 5 shows an overlaid composite for the same time period as the averaged composite in Fig. 4. A disadvantage of the averaged composite, from the forecaster's point of view, is that averaging "slows down" the weather systems; that is, a moving weather system,

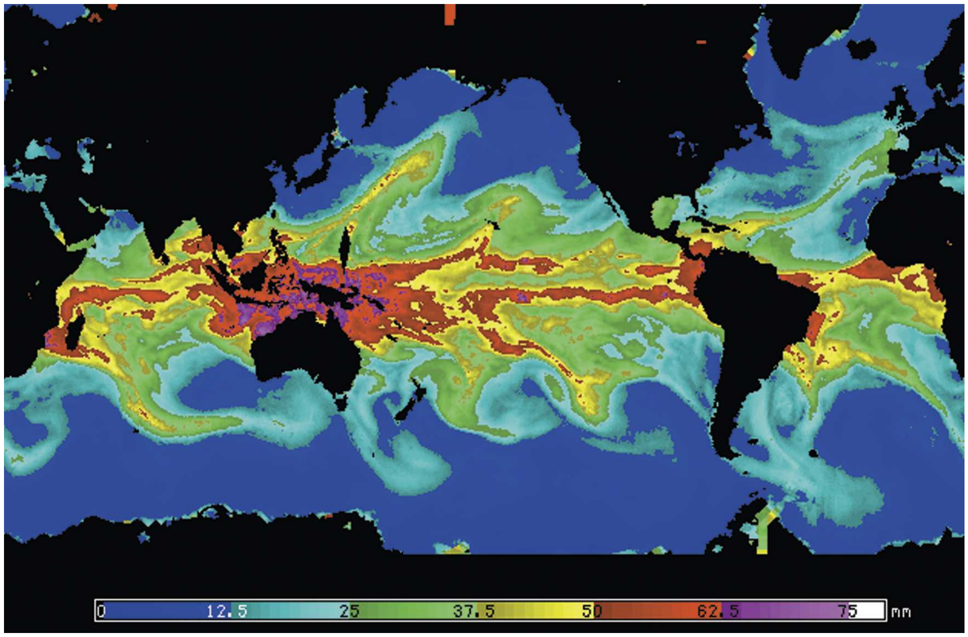

FIG. 4. Average TPW for the 12-h period ending at 1929 UTC 29 Mar 2006. Approximately 30 orbits went into this composite.

if observed more than once in a 12-h period, will appear to be "behind" its position in the overlaid composite. An advantage of the averaged composite is that it is smoother than the overlaid composite. The Data Processing and Error Analysis System (DPEAS) software (see section 5), which we used to construct these composites, is also capable of doing a weighted average of the observations, with older observations being weighted less than newer observations. This method is "between" the averaged product, which has uniform

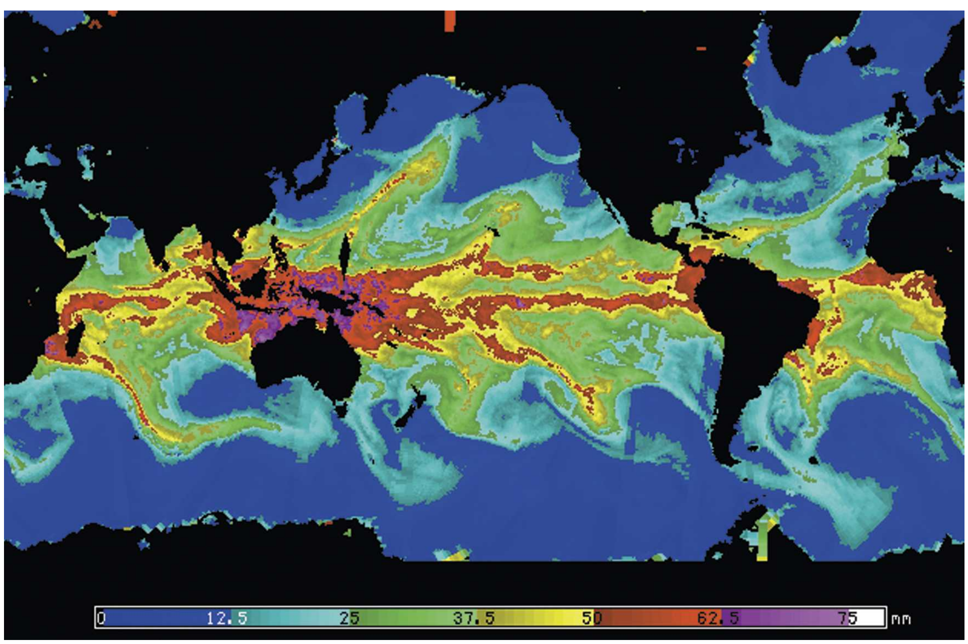

FIG. 5. Overlaid TPW for the 12-h period ending at 1929 UTC 29 Mar 2006 (only the most recent datum at a point is shown). The same orbits were used in this composite as in Fig. 4.

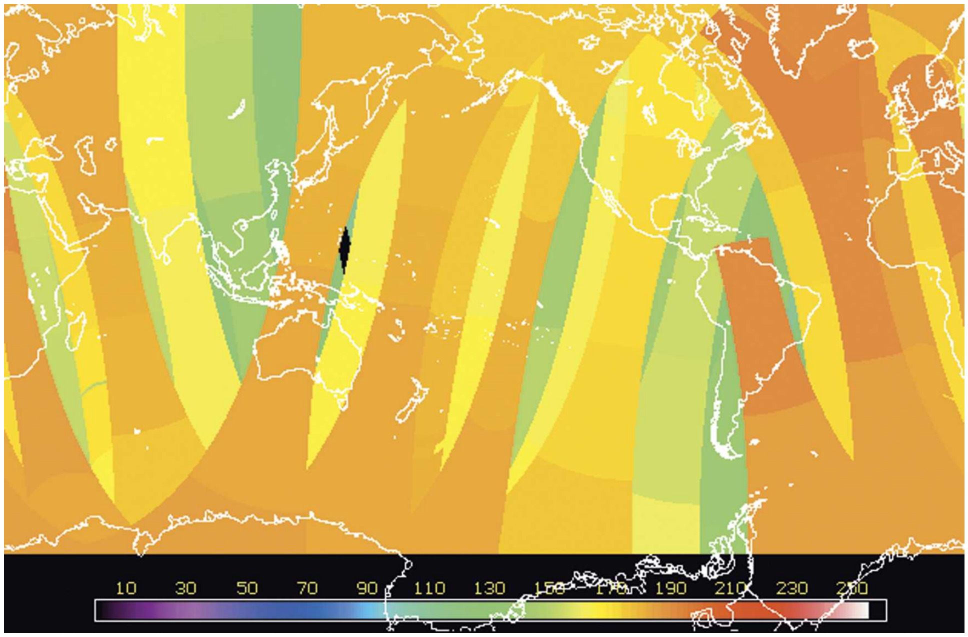

FIG. 6. Time of latest observation for the composite shown in Fig. 5. The times are UTC to the nearest 10 min. Using the color bar [(left) 0000 and (right) 2350 UTC], one can get an approximate idea of the time of observation. We use McIDAS to display the data, and with the IMGPROBE command, one can get the precise time (within 10 min) of each point.

weights for every data point, and the overlaid product, which weights the newest observation as 1 and all older observations as 0.

When the overlaid composite is constructed, we can optionally map the time of the most recent observation and the satellite that made it. These data are useful for analyzing the resultant TPW field. Figure 6 shows the mapped times. The individual orbits are clearly shown,

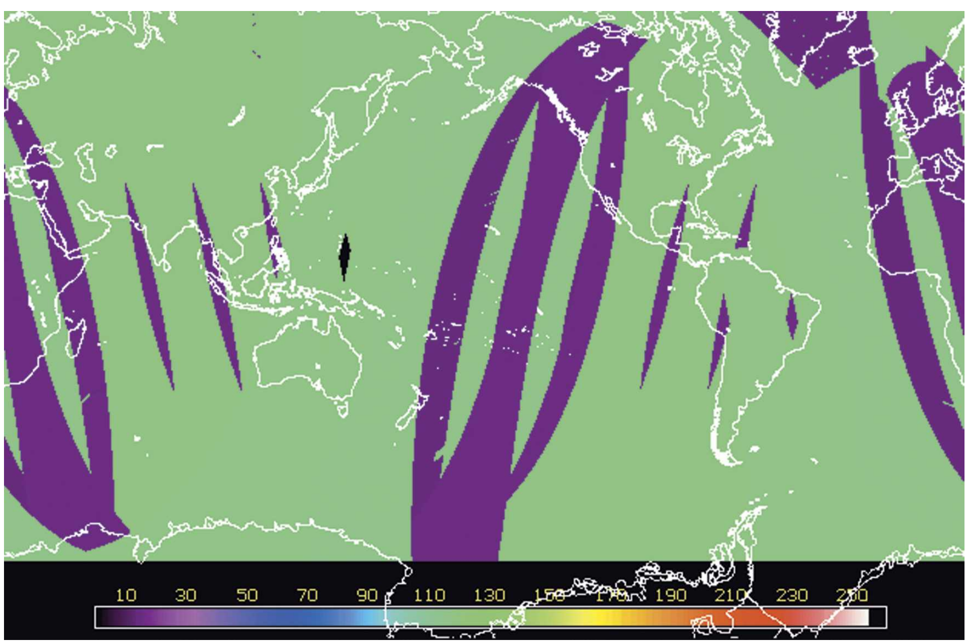

FIG. 7. The satellite that made the observations plotted in Fig. 5. Green points were observed with AMSU, purple points were observed with SSM/I. Using the IMGPROBE command in McIDAS, one can discover which NOAA satellite (*NOAA-15, -16,* or *-17*) made the AMSU observations and which a DMSP satellite (*F13, F14,* or *F15*) made the SSM/I observations.

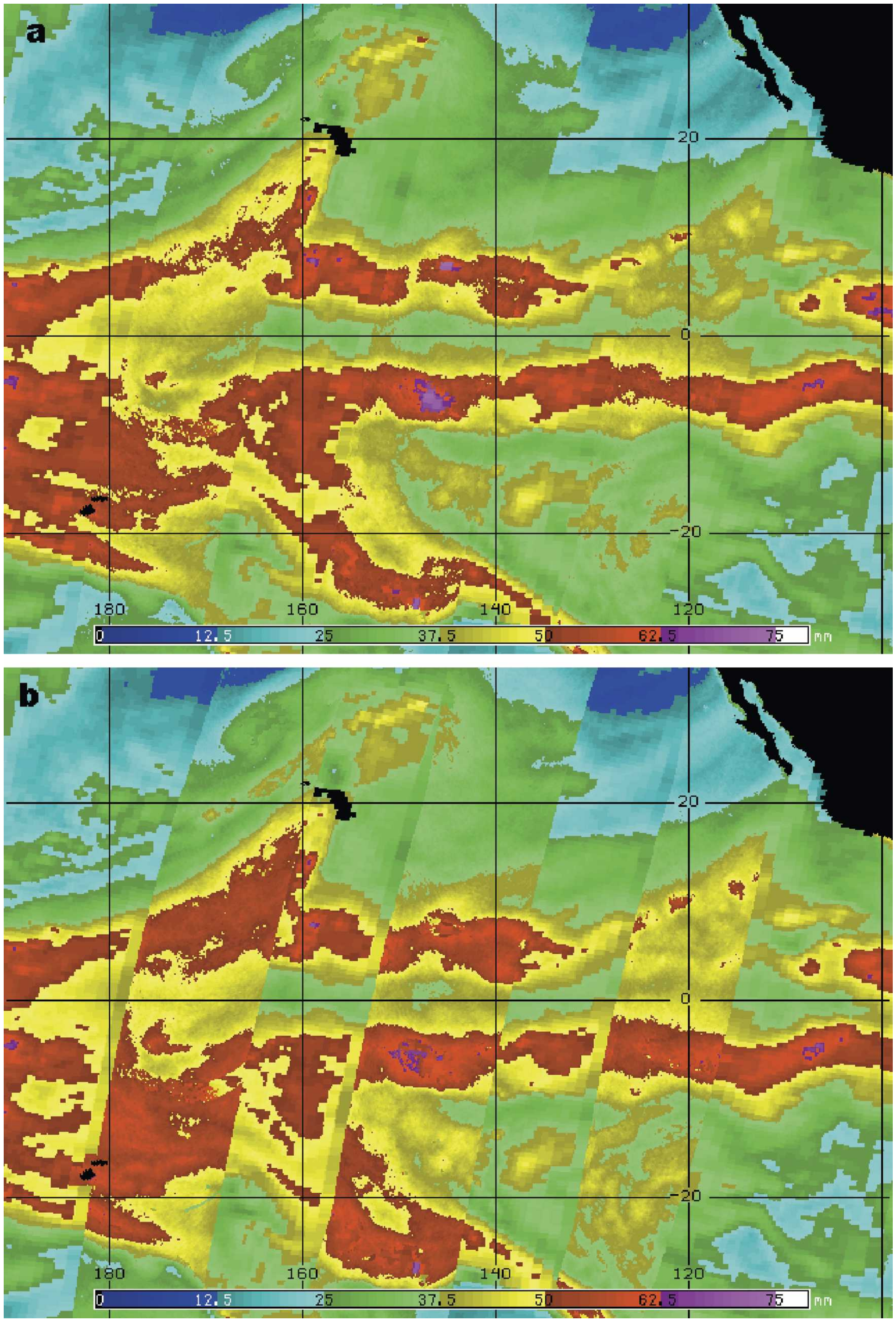

FIG. 8. Illustration of the removal of artifacts from composite images by use of the blending algorithm: (a) composite with blending algorithm and (b) composite without blending algorithm.

as are the older data. Figure 7 shows the satellite that made each observation.

To illustrate the primary reason for using a blending algorithm, we constructed Fig. 8, which shows a close-up view of the equatorial east Pacific with (Fig. 8a) and without (Fig. 8b) the blending algorithm. In Fig. 8b, individual swaths can be clearly seen (and can be identified with the aid of Figs. 6 and 7). These individual swaths are what we referred to (above) as artifacts, which are distracting to forecasters. The blending algorithm removes most of these artifacts and results in a meteorologically much more pleasing image, particularly when the images are animated.

5. Processing environment

The TPW composites are produced in real time using DPEAS (Jones and Vonder Haar 2002). The system runs on a cluster of Windows computers. Each hour, new AMSU and SSM/I data are acquired from NESDIS, the adjustments are constructed, the data are mapped, and the composites are formed. The system handles about 200 GB of data per day. The DPEAS reliability exceeds 99.97% due to its fault-resilient, grid-computing capabilities. Internet reliability and other non-DPEAS issues reduce the total aggregate system reliability to approximately 98%. The TPW composite algorithm was installed at OSDPD in less than 8 days, and upgrades have been performed in a matter of hours. A cost/benefit analysis of the system has shown large returns on investment (Jones et al. 2005). The cost savings are primarily due to a portable software framework approach used by DPEAS. The system is capable of compositing numerous additional satellite data flows in near–real time in a manner suitable to the OSDPD operational environment. The DPEAS software has been installed at OSDPD, and is being evaluated for operational use. (The hourly output maps in McIDAS format are available online at http://amsu.cira.colostate.edu/TPW/ftp.htm, and 5-day animations of TPW images are available online at http://amsu.cira.colostate.edu/TPW.)

6. Summary and conclusions

Satellites offer the only way to observe the global distribution of meteorologically important parameters. Forecasters need these parameters to make informed 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 SAB and elsewhere.

Acknowledgments. The research was supported by 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 under Cooperative Agreement NA17RJ1228. Additional support was provided by the Department of Defense Center for Geosciences/Atmospheric Research (CG/AR) at Colorado State University under Cooperative Agreement W911NF-06-2-0015 with the Army Research Laboratory.

REFERENCES

Ferraro, R. R., F. Weng, N. C. Grody, and A. Basist, 1996: An eight year (1987–94) climatology of rainfall, clouds, water vapor, snowcover, and sea-ice derived from SSM/I measurements. *Bull. Amer. Meteor. Soc.,* **77,** 894–905.

——, and Coauthors, 2005: NOAA operational hydrological products derived from the Advanced Microwave Sounding Unit. *IEEE Trans. Geosci. Remote Sens.,* **43,** 1036–1049.

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.

——, P. C. Shott, L. Zhao, I. Tcherednitchenko, R. R. Ferraro, C. Karlburg, and S. Q. Kidder, 2005: Maximizing the usefulness of grid technology on NOAA Office PCs: Cost/benefit analysis. Report to the NOAA High Performance Computing and Communications Program and NOAA/NESDIS/OSDPD, 39 pp.

Jones, T. A., and D. J. Cecil, 2006: Histogram matching of AMSR-E and TMI brightness temperatures. Preprints, *14th Conf. on Satellite Meteorology and Oceanography,* Atlanta, GA, Amer. Meteor. Soc., CD-ROM, P1.23.

Kidd, C., D. R. Kniveton, M. C. Todd, and T. J. Bellerby, 2003: Satellite rainfall estimation using combined passive microwave and infrared algorithms. *J. Hydrometeor.,* **4,** 1088–1104.

Richards, J. A., and X. Jia, 1999: *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 363 pp.

Weng, F., L. Zhao, G. Poe, R. Ferraro, X. Li, and N. Grody, 2003: AMSU cloud and precipitation algorithms. *Radio Sci.,* **38,** 8068–8079.