Contribution of Data Set Construction Methodology to Data Set Uncertainties
Deborah K. Smith (smith@remss.com), Carl A. Mears, Kyle A. Hilburn, Lucrezia Ricciardulli, Frank J. Wentz, Remote Sensing Systems
Rick Schutte, student at Santa Clara University (summer intern at Remote Sensing Systems)

Summary
This poster summarizes the analysis of similar, yet distinctly different, water vapor products created using DISCOVER microwave radiometer data. The purpose of the analysis was to construct a 1-deg, monthly, merged water vapor data set best suited for climate study. Different approaches were applied and the impact of each approach studied. The new water vapor data product is now available to Users in netCDF format and contains water vapor values, a climatology, vapor anomalies, and a trend map. The construction process is described.

Background
Remote Sensing Systems (RSS) processes satellite microwave radiometer data with the help of NASA MODIS/LIICS funding. We carefully internalize the data at the brightness temperature level and apply a consistent processing methodology to all sensors to produce Version-7 ocean products which include wind speeds, water vapor, cloud liquid water, rain rates and sea surface temperatures (for some instruments). A radiative transfer model is used to relate the brightness temperatures to ocean parameters. Water vapor is the most robust of the ocean products we retrieve due to the strong spectral signature in the radiometer frequency measurements, so it is the first merged product we chose to create. Individual satellite ocean products are available, however, we determine the best approach for merging the satellite data to create a water vapor product best suited to climate research.

Four scientists at RSS, each very familiar with the data we produce, independently created a 1-deg, monthly water vapor data set (see Table 1, below). We expected the outcomes of the independent approaches to be very similar. Indeed, it was found that initial differences varied over by half a millimeter, showed regional differences, and had different trends despite using the same high quality input data. The differences are in Table 1. We next explored how each product construction decision affected the outcomes. The decisions are summarized to the right. Most decisions were then automated except in a few cases for which data were excluded/included by the processing scientist.

Data Product Construction Decisions

Major Impacts

Geographic sampling - An important contribution to the differences between the four approaches involved construction decisions that altered what data were used or omitted. Affected data included coastal areas, areas of ice and regions of rain.

Minimal Number of Data per Grid Cell:
Radiometers suffer from side lobe interference that prevents obtaining vapor values near land. Due to mixed pixel detection, observation footprints, same grid cells have more observations than others. This can result in poorer quality vapor averages. Each approach listed in Table 1 used a different minimal number of data requirement. See Figure 3 to visualize the amount of data in the grid cells examined.

Ice Flagging: Ice is likely to exist more at one end of a month than another (with the exception of freezing shoulder). Three approaches removed grid cells with ice present during the month, using thresholds varied. The mean vapor was lower for Scientist 3 who did not use ice as a removal scheme, despite including data along the growing and melting ice edge. We determined that ice removal was necessary, and developed a mean-day of-month quality calculation to remove ice edge grid cells. To handle ice, we used the number of ice observations within a grid cell during the month to exclude when too much ice existed. Figure 3 top right shows an example of ice counts for a southern ocean region.

Rain Flagging: Scientist 2 included a 1/3 cell extended rain flag and his vapor mean and trend were different from the other approaches (see Figure 4 difference map). The inclusion of a rain flag creates a geometric area mapping problem as it removes data from primarily high, rainy, tropical areas resulting in lower vapor and an under-represented trend. Figure 6 shows a comparison of SSM/I/TDD vapor to GPS vapor in rain-free and rainy conditions. Since little difference can be attributed to rain, we determined that an extended rain flag was not needed.

Minor Impacts (if any)

Averaging Stacking: Three scientists used a drop-in-bucket averaging method. The fourth first determined monthly vapor averages and then stacked the reported vapor data set. There was virtually no difference between the methods. However, by calculating individual instrument monthly vapor values and then merging, we had better quality control over what data were added. We determined this approach to be best for creating the public product.

Remaining Instrument Offsets:
We used only RSS Version 7 water vapor data.

At the time of this study, that included all SSM/I, F17 SSM/I, AMSR-E and WindSat. As shown in Table 1, some approaches used only SSM/I while others included AMSR-E and WindSat. Figure 2 shows increased differences between the four data sets after 2002 when AMSR-E data is used. Some of these differences may be due to the later sampling time of AMSR-E (1am) as opposed to 0.5 am of other radiometers. AMSR-E observations may be due to the fact that these instruments were calibrated with rain-free data as opposed to using all data (rain-free and clear). We believe these factors cause a slight difference requiring correction. We calculated offsets of AMSR-E - (0.2 mm) and WindSat (0.05 mm) and applied these to the data.

Type of Data Reporting - As listed above, it was agreed that a correction was needed for both AMSR-E and WindSat data. We examined the differences created by using a simple additive offset or a more complicated multiplicative correction. The difference was negligible so we used the simpler additive correction.

Table 1: The four approaches used different microwave radiometers, set different boundaries on ice and sea ice, used different algorithms and detection schemes. Despite high quality consistently processed, internetworked ocean data, the water vapor and global trends used were not identical.

New Vapor Product at www.remss.com
From this exercise, we learned the lessons summarized in the Guidelines listed below. We took these lessons learned and applied them to create a new 1-deg, merged water vapor product for release to the public. We first created 312-deg maps of water vapor from individual satellites, keeping track of number of obs, number of ice obs and the mean day of month (as shown in Fig. 5). A mean water vapor value is only reported if the number of obs is greater than 160, number of ice obs is less than 30, and the [mean day of month - center of month] is less than 6 days. 3 satellite-months did not meet these requirements: F08 in Jan 1998, F08 in Oct 1999 and F10 in Dec 1995. In each of these cases, these mean values are needed for consistency of the time series so an exception was made and the data were included. Quality checks and post hoc corrections were applied in the second stage of processing. A set of 12 month climatology maps and monthly anomaly maps are then calculated from the monthly vapor values and added to the data set. In addition, we include a water vapor trend map and a homesticker plot.

The RSS DISCOVER Water Vapor 1-deg Monthly Volume Data Set will be available at ftp.remss.com/vapor and www.discover-earth.org/vapor in mid December 2012. The file contains monthly mean water vapor maps for Jan 1988 to Nov 2012, 12 month climatology vapor maps (Jan-Dec), associated monthly anomaly maps (monthly water vapor minus climatology), vapor trend maps and a time-latitude [homesticker] plot. The netCDF 4 file contains 24-compliant metadata. This water vapor product will be updated each month (for example, for Feb 2013).

Guidelines for Merging Data from Individual Satellites into a Climate-Quality Product

1. Use interpolated data that is consistently processed throughout instrument operation and different instruments. Do not apply for the great basin project (monthly reflectance) to ocean applications.
2. Use monthly maps to correct for differences since differences exist for the same month.
3. Keep track of all ice and water present for best product use, however, only the non-ice data are used.
4. If any adjustments are needed, use a simple correction.
5. Create an end product that uses the individual satellite products for the best possible data product. This trend map will be updated monthly along with other data.

Figure 1: The four global mean of water vapor mean from the four original data sets. The difference is expressed as a % error. Three of the differences decreased (see Figure 5) as we defined the approaches, giving a better understanding of the unique features of the product. In the final product similarity, we use the textual descriptions for construction purposes.

Figure 2: The series of the same four data sets shown in Figure 1, only that data for a grid cell was included in the calculations when it was present at all four data sets. This proves virtually all differences due to different algorithms of reading products. In Figure 3, the left side shows the difference in each approach. The right side shows the percent difference in each approach.

Figure 3: This comparison of SSM/I and WindSat vapor products both are on final and mining calculations. The mean difference shows no greater difference than before.

Figure 4: This comparison of SSM/I vapor products both on final and mining calculations. The mean difference shows no greater difference than before.