6B.3 A NEW GLOBAL ANALYSIS OF PRECIPITATION

Mathew R. P. Sapiano[1]*, Thomas Smith[2] and Philip A. Arkin[1]

[1]The Cooperative Institute for Climate Studies (CICS), University of Maryland, College Park, MD
[2]NOAA/NESDIS Center for Satellite Applications and Research, College Park, MD

1. INTRODUCTION

The effect of climate change on the hydrological cycle is still largely unknown due to the complex mechanisms related to water in the atmosphere (amongst other things). Paradoxically, changes in the hydrological cycle have the highest societal relevance since phenomena such as precipitation are crucial to life on Earth and the hydrological cycle is related to some of the most damaging extreme weather. Crucial to the understanding of climate change effects is knowledge of recent global trends in hydrological parameters, which are more challenging to quantify than temperature trends. There is a lack of agreement between the precipitation changes obtained from models and observed precipitation from satellites (Wentz et al. 2007) and from theoretical considerations (Allen and Ingram 2002; Pall et al. 2007). Global observations of precipitation, which are necessarily reliant on satellite estimates over the ocean, are often considered to be inadequate for climate studies due to issues with the length and homogeneity of such datasets. Furthermore, the size of the trends are often too small to detect against the noisy background (which is only made worse by the additional noise in satellite estimates).

Many merged multi-source global analyses of precipitation exist, including the well-known and commonly used Global Precipitation Climatology Project (GPCP; Adler et al. 2003) analysis and the CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997). Such datasets arose in response to the need for longer global time-series of precipitation and were designed to exploit all available information: gauges are superior over land but satellite estimates are required over ocean and sparsely sampled land. The basic ethos of both GPCP and CMAP was to merge geostationary infra-red estimates (which start in 1979) with higher quality polar-orbiting passive microwave estimates (which start in 1987) and to apply a bias correction over land to make this estimate agree with the gauges. This approach allows such datasets to use the best data available to produce estimates of precipitation which are long, but can also lead to discontinuities, artifacts and inhomogeneities in the longer record. Such issues are caused by changes in the number and type of available sensors and can impact their suitability for investigation of long-term changes in global precipitation.

An additional problem is that high latitude precipitation is poorly represented in these datasets since the available estimates originate from either gauges (which suffer from problems related to under-catch of solid precipitation) or satellite data (which exhibit large errors at high latitudes due to ice-contamination issues). A solution to this problem is to use reanalysis data with the satellite/gauge estimates. This approach was tried in CMAP with the incorporation of NCEP/NCAR reanalysis data (Kalnay et al. 1996) but has been underused – probably due to the serious problems associated with the rainfall estimates from the NCEP/NCAR reanalysis. Estimates from the European Centre for Medium Range Weather Forecasts (ECMWF) ERA-40 (Uppala et al. 2005) reanalysis have been shown to be superior to satellite estimates in the arctic due to the limitation of satellite estimates over snow/ice (Serreze et al. 2005; Su et al. 2006). With the continued development of reanalysis data from multiple agencies, it is expected that such estimates will only improve and that their incorporation into merged precipitation products will become more commonplace.

We have produced a new global analysis of precipitation which uses Optimum Interpolation (OI) to blend high quality passive microwave satellite data with ERA-40 reanalysis data. The OI technique allows the satellite data to have greater weight in low latitudes and the reanalysis data to have higher weight in higher latitudes with a mix in mid-latitudes. Thus, the new dataset exploits the strengths of the components. The OI methodology also allows for the calculation of errors associated with the analysis, which are needed for many applications and can be used in climate studies. One such use of the new analysis is as a base-period for a new reconstruction of historical precipitation (Smith et al. 2008). This paper outlines the data, methodology and preliminary results obtained with this approach and the prospects for future enhancements.

2. DATA

As stated above, the inputs for this analysis are precipitation estimates derived from polar-orbiting passive microwave instruments and reanalysis

* *Corresponding author address:* Mathew R. P. Sapiano, Univ. of Maryland, CICS/ESSIC, 4114 Computer and Space Sciences Building, College Park, MD, 20742; email: msapiano@essic.umd.edu.

precipitation estimates. There are many different passive microwave estimates available but the longest record is that from the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I). The SSM/I record starts in 1987 with a single sensor and has at least dual sensor coverage from 1992 until present. The local measurement times are fixed for each sensor although satellite drift leads to changes of up to one to two hours over several years. The goal of the DMSP has been to maintain a sensor in orbit with a local equator crossing time of around 6am/pm (depending on whether the satellite is in its ascending/descending phase) with a second sensor with local crossing time around 8am/pm. It takes one day to get almost complete global coverage with a single sensor so an additional sensor yields the advantage of filling in the few gaps remaining from the first sensor and allows for many locations to be sampled twice. The earliest SSM/I sensor was launched in 1987 but suffered a system failure in 1990 so that one of the key channels for precipitation estimation was lost completely. Estimates of precipitation from the SSM/I are available over this time, although they are of lower quality.

Other passive microwave instruments are available such as the Tropical Rainfall Monitoring Mission (TRMM) Microwave Imager (TMI) and the Advanced Microwave Scanning Radiometer (AMSR). Whilst similar to the SSM/I, these instruments have different foot-prints, channels and orbits to the SSM/I data which would surely introduce discontinuities into the data which would present problems for its use in climate studies. We therefore use only precipitation estimates from the SSM/I data although we do use all available sensors. A consequence of this is that the sampling is increased during certain periods and so the accuracy of the data might be considered to be higher (this can be incorporated into the error estimates, but is not at present).

There are several different methods for estimating precipitation from SSM/I including the Goddard Profiling algorithm (GPROF; Kummerow and Giglio 1994), the Unified Microwave Ocean Retrieval Algorithm (UMORA; Wentz 1997, Wentz and Spencer 1998, Hilburn and Wentz 2007) and the NOAA/NESDIS algorithm (Grody 1991; Ferraro 1997). As the name suggests, UMORA is restricted to oceans only whilst the other two are land and ocean. All of these datasets utilize the same two techniques based on radiative transfer: emission and scattering estimates. The emission estimation technique directly measures the emission from hydrometeors in the atmosphere, which requires a homogeneous background and is therefore unsuitable over land and coast. The indirect scattering technique is based in the scattered radiation from ice particles in clouds and works over any background (ocean or land) but is inferior to the emission technique. Careful analysis of the available datasets showed that the recently reprocessed UMORA V6 was superior over the ocean and has the added benefit of careful inter-calibration between sensors. Over land, GPROF (Version 6.5 for SSM/I) was found to be superior to the NOAA/NESDIS algorithm although it should be noted that this is not the most recent version of the GPROF algorithm (the most recent version has currently only been re-processed to 1997). For the current analysis, we have used GPROF over land and a combination of Wentz and GPROF over the ocean. The combination was done using the OI technique and was designed to minimize discontinuities between oceans and land. It is expected that the choice of data will be re-visited in the future as improvements are made to algorithms, although the SSM/I record will remain the primary satellite input.

The GPROF data were obtained from the NASA Goddard Space Flight Center website on a 0.5° grid at the original sub-daily timings. Ambiguous flags were developed for these data (George Huffman, *personal communication*) which are applied to the original data to remove pixels contaminated with snow/ice. Once this mask is applied, the data are accumulated to produce monthly means on a 2.5° grid (in *mm day*$^{-1}$). The UMORA data were downloaded from the Remote Sensing Systems (RSS) website as daily, 0.5° averages. A simple mask is applied to remove any data that is unrealistically high (presumably due to contamination by snow/ice) before the data are accumulated to monthly means on a 2.5° grid (in *mm day*$^{-1}$).

There are several different reanalysis products available, although we have started by using only the ERA-40 analysis which runs from 1958 to 2002. This decision was largely motivated by the evidence that the precipitation estimates from this dataset possess skill at high latitudes although we expect to be able to use the same technique with other data. One obvious drawback to using this data is that it limits the dataset to 1987-2002 despite the existence of SSM/I data from 1987-2007. The ECMWF is currently processing an “interim” version of their reanalysis which is based on ERA-40, but has small improvements which are expected to improve the precipitation. Other datasets are becoming available such as the Japanese Aerospace Exploration Agency’s (JAXA) JRA-25 (Onogi et al. 2007), the NASA Modern Era Reanalysis for Research and Applications (MERRA; Mike Bosilovich, NASA/GMAO, *personal communication*) and NOAA expects to initiate a CFSRR (Coupled Forecast System Reanalysis/Reforecast. personal communications, Hua-Lu Pan, NCEP), both for the period 1979 to present.

The ERA-40 reanalysis data are obtained directly from the ECMWF as six-hourly total forecast precipitation accumulations which are totaled to yield a monthly precipitation value. The overlapping period of SSM/I and ERA-40 thus constrains the length of these data currently to 1987-2002, but is expected to be extended with the use of other products in the near-future.

3. Methodology

The input data are combined using an OI (e.g. see Reynolds and Smith 1994) which estimates a value at a point from a weighted combination of the surrounding

points. The weights for the surrounding points are calculated by considering their spatial correlation with each other and the point to be estimated, as well as their noise-to-signal ratio. The spatial correlation is estimated using Gaussian functions and is here set to have a fixed e-folding scale of 2.5° to minimize spatial smoothing in the OI. The relative noise-to-signal ratio is calculated through comparison with GPCP and is given by

$$\varepsilon^2 = \frac{1 - r^2}{r^2} \quad (1)$$

where r^2 is the correlation between the input data and GPCP. Since GPCP cannot be considered *the truth*, this method essentially relies on the assumption that differences (and, therefore, low correlations) between GPCP and the input data occur in areas where noise is higher in the estimates and vice-versa. For this analysis, we used a temporally fixed, smoothed zonal mean of the relative noise-to-signal of the inputs as shown in Figure1. Using this technique, the zonal mean noise-to-signal ratio for ERA-40 is approximately constant, whilst the noise-to-signal ratio for SSM/I is higher in the high-latitudes and lower at the equator. This pattern means that ERA-40 is given higher weight at high latitudes and the SSM/I estimates are given higher weight at the equator, which agrees with the hypotheses on the relative merits of the inputs at different locations. Since the input data are all on the same regularly spaced grid, and the spatial correlation scale is fixed, the weights are dominated by the relative signal-to-noise ratio and the SSM/I data are favored in the tropics, the ERA-40 data are favored at higher latitudes and a gradual mix of the two sources occurs in between.

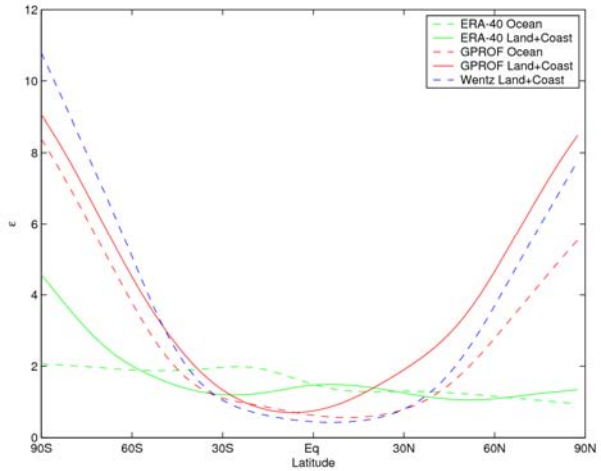

Figure 1: Noise-to-signal ratios obtained by comparison with GPCP.

The usual methodology for an OI is that a search radius is chosen around the point to be estimated and if the initial radius does not contain enough points, then the search radius is extended. However, in this case ERA-40 data exists in every grid-point so that the initial three-by-three search area never needs to be extended. It is unusual for missing data in the SSM/I record except in over very dry areas where it is assumed that there is no rain (although these rare points would be estimated from ERA-40 data) and over snow/ice. Since ERA-40 is preferred at high latitudes, the existence of missing data over snow/ice is not problematic. The other important period which contains missing SSM/I data is during the period of technical problems with the first SSM/I sensor which affected the land retrievals; these data are estimated using ERA-40 only.

Before the OI is carried out, the two SSM/I estimates are merged using Optimal Averaging with the noise-to-signal ratios shown in fig. 1. Once this combined SSM/I is produced, new noise-to-signal ratios are calculated (which are almost identical to those in fig.1) and the procedure is repeated to merge ERA-40 and the combined SSM/I. The OI is performed on the raw monthly means of the inputs and the weights are normalized to ensure their addition to unity. This initial product is available as a monthly mean on a 2.5° grid from July 1987 until August 2002.

4. RESULTS

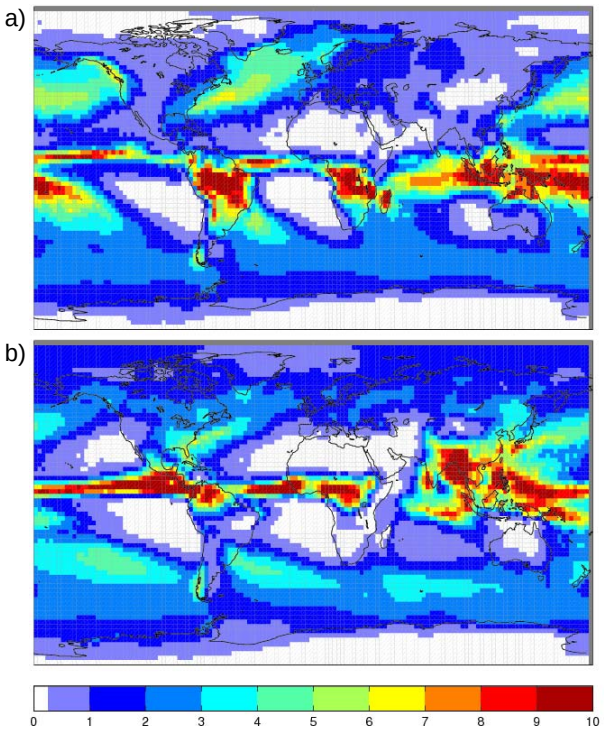

Figure 2: Mean precipitation from the OI for (a) December-February and (b) June-August in *mm day*$^{-1}$.

Figure 2 shows the December-February and June-August mean precipitation from the OI. The pattern and magnitude of global precipitation matches expectations with the main features present, including the ITCZ, the storm tracks, the dry subtropical high pressure zones and precipitation associated with monsoonal circulations. Note that the OI produces a value for every grid point, although the quality of the

estimates might vary. As already stated, an advantage of the OI is that an estimate of the error in the output datasets is calculated as part of the procedure. Here, errors from systematic data biases are not considered. The error estimate is normalized by the total variability in the data, hence, this formulation of the error is a proportion where zero denotes high skill and unity denotes no skill. The mean for December-February and June-August of the normalized error is shown in Figure 3. In this case, the error estimates are highly reflective of the input noise-to-signal ratios used to calculate the weights since most other parameters where more-or-less constant over the regular grid. Fig. 3 shows the highest skill lies over the tropical oceans where the SSM/I data is of the highest quality. At higher latitudes, the data is of lower quality and there is almost no skill in the Antarctic. Tropical land masses also have surprisingly low skill, due to the low variance over dry desert areas (see fig. 1).

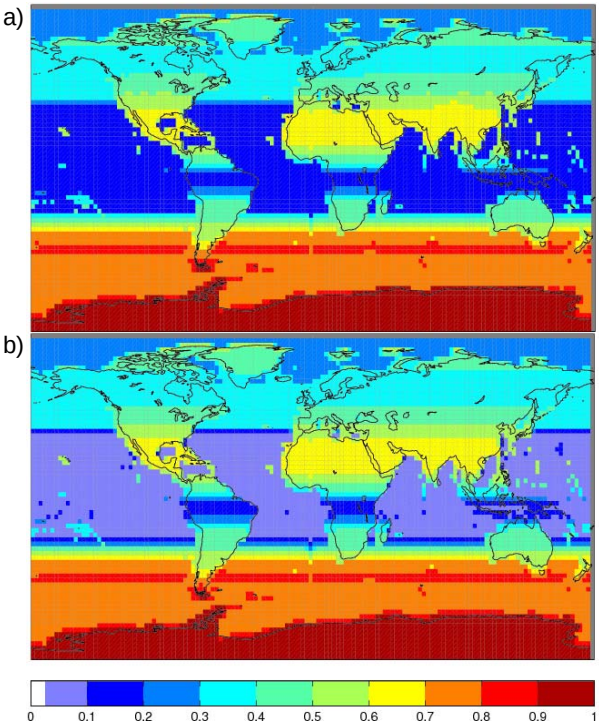

Figure 3: Mean normalized error estimates from the OI for (a) December-February and (b) June-August.

Assessment of the skill of this dataset is somewhat challenging since the true state of oceanic global precipitation is only known from satellite sources which all use the SSM/I data. One way of assessing the skill is to compare to gauge data over land such as those used in GPCP and the independent Climate Prediction Center (CPC) PREC-L (Chen et al. 2002). Anomaly correlations for the OI compared with these two datasets is shown in Figure 4; note that the GPCP analysis is very similar to the gauges over land but is highly reliant on the satellite estimates over the ocean. Fig. 4a and 4b show very similar correlations, with poor performance over the Sahara and the Antarctic (both of these deficiencies are manifested in the error estimates of fig. 3) although the quality of the gauges is questionable in these areas. In general, correlations between the OI and both gauge analyses are around 0.8-0.9 and higher over parts of India, China and Russia, but lower over significant orography and some coastal areas such as over Indonesia.

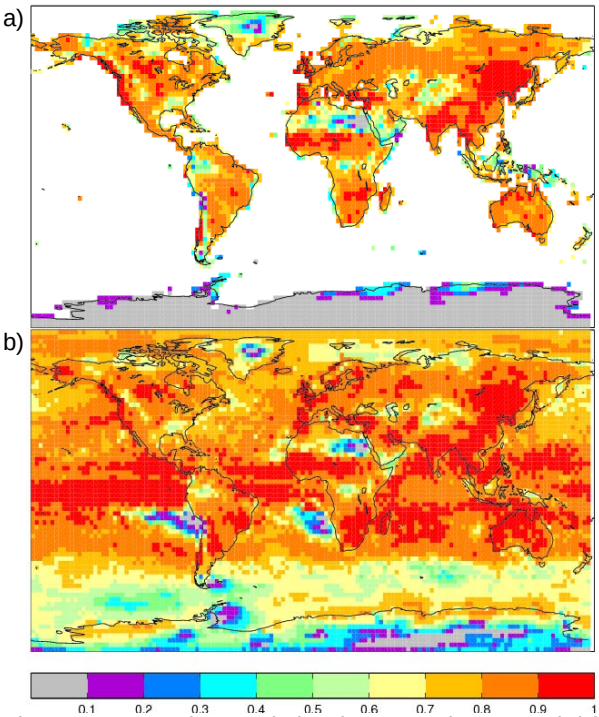

Figure 4: Anomaly correlation between the OI and (a) CPC PREC-L and (b) GPCP.

One of the useful outcomes from this work may be the adoption of these methodologies and perhaps these inputs by the GPCP in the next generation of their monthly precipitation product. Figure 5 shows the December-February correlation (on the raw data, not the anomalies as was shown in fig. 4) between the GPCC gauge data (which is used in GPCP) with the OI and the GPCP multi-satellite (GPCP MS) product. The GPCP MS is an interim GPCP product which is the combination of all the satellite data; this product is then bias-corrected to be consistent with the GPCC gauges to obtain the final GPCP product. Differences between the GPCP MS should occur either because of the use of newer input precipitation estimation algorithms or because of the additional reanalysis data, with the latter being more likely to make a large impact. Fig. 5 shows that the OI has higher correlations with the GPCC gauge data particularly in the Northern Hemisphere high latitudes. Such skill at high latitudes must be due to the inclusion of the reanalysis data and shows the clear value of including such data. Whilst the OI has higher than GPCP MS correlations virtually everywhere (probably due to the use of newer inputs), low correlations against the raw DJF values are still observed over large parts of Africa and South America,

both of which experience very low rainfall at this time of the year suggesting that the SSM/I data poorly represents very low rainfall amounts probably due virga.

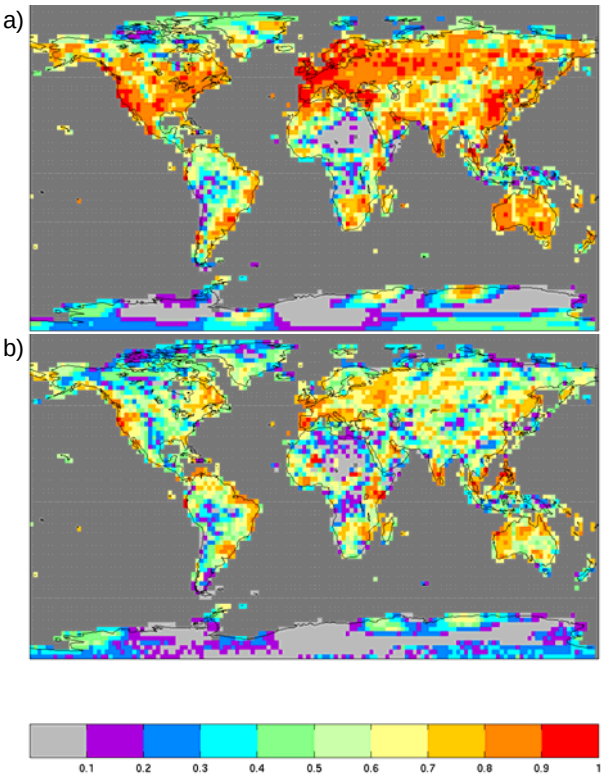

Figure 5: Anomaly correlation between the GPCC gauge analysis used in GPCP and (a) the OI and (b) the GPCP multi-satellite interim product.

5. SUMMARY AND FUTURE PLANS

A new analysis of global precipitation has been developed using Optimal Interpolation to blend passive microwave and reanalysis forecasts in such a way as to reduce the deficiencies of these inputs. These early results are very promising with high correlations obtained when compared to existing gauge analyses. In particular, the results show that reanalysis techniques have improved to the level where precipitation forecasts might be the most reliable source of information at high latitudes. A further benefit of the OI procedure is the production of error estimates as part of the blending technique. The error estimates reflect the expectation that the analysis would have highest skill over the tropics (particularly the tropical oceans where convective processes are more dominant).

There are still some deficiencies that need to be addressed in the analysis and the technique. Currently, the noise-to-signal ratio is calculated using GPCP, a similar product. Whilst this approach does give a meaningful estimate of the relative error, it would be more satisfactory to use an independent global precipitation estimate (one that excludes SSM/I). Additionally, there are still some clear issues over land (see fig. 5) which would be solved through the incorporation of gauges. The inclusion of other data such as the IR data is also possible.

As was mentioned earlier, it is expected that newer, improved versions of the input datasets will become available over the coming months and years. It is expected that a new version of GPROF will be available sometime in 2008-2009 which might be superior to the latest UMORA algorithm. Perhaps more importantly, new reanalysis datasets are expected to become available with the ERA-Interim data expected in mid-2008 and the NSAS MERRA analysis expected sometime in 2008-2009. We plan to further update this dataset based on the availability of these datasets, with the first "full" release of the data utilizing the ERA-interim in 2008.

An important area for further work is the validation of these data in areas where we lack good data such as over the oceans and at high latitudes. We plan to conduct further evaluation of the data by comparing it with the TAO/TRITON buoy array in the Tropical Pacific as well as the against some Arctic gauge data. Whilst both of these datasets are limited, it is hoped that such results would inform the choice of input weights. Finally, validation of climate phenomena is required to evaluate the representation ENSO.

ACKNOWLEDGEMENTS

Some of the data used here are produced by Remote Sensing Systems and sponsored by the NASA Earth Science REASoN DISCOVER Project. Data are available at www.remss.com. The GPROF data were produced by E. J. Nelkin at Goddard Space Flight Center as part of the TRMM project, and the ERA-40 data supplied by the European Centre for medium Range Weather Forecasts.

REFERENCES

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.*, **4**, 1147-1167.

Allen, M. R. and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419**, 224-232.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations. *J. Hydrometeor.*, **3**, 249-266.

Ferraro, R.R., 1997: SSM/I derived global rainfall estimates for climatological applications. *J. Geophys. Res.*, **102**, 16,715-16,735.

Grody, N. C., 1991: Classification of Snow Cover and Precipitation using the Special Sensor Microwave Imager. *J. Geophys. Res.*, **96** (D4), 7423-7435.

Hilburn, K. A. and F. J. Wentz, 2007: Intercalibrated Passive Microwave Rain Products from the Unified Microwave Ocean Retrieval Algorithm (UMORA). *J. Appl. Meteorol*, in press.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-472.

Kummerow C. and L. Giglio, 1994: A Passive Microwave Technique for Estimating Rainfall and Vertical Structure Information from Space .1. Algorithm Description. *J. Appl. Meteorol.*, **33**(1), 3-18.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi, S. Kadokura, K. Wada, K. Kato, R. Oyama, T. Ose, N. Mannoji and R. Taira, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369-432.

Pall, P., M. R. Allen and D. A. Stone, 2007: Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO2 warming. *Climate Dyn.*, **28**, 351–363.

Reynolds, R. W. and T. M. Smith, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *J. Clim.*, **7**(6), 929–948.

Serreze, M.C., A. Barrett, and F. Lo, 2005: Northern high latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Mon. Wea. Rev.*, **133**, 3407-3430.

Smith, T. M., M. R. P. Sapiano and P. A. Arkin, 2008: Historical Reconstruction of Monthly Oceanic Precipitation: Methods and Preliminary Results. *In preparation*.

Su, F., J. C. Adam, K. E. Trenberth, and D. P. Lettenmaier, 2006: Evaluation of surface water fluxes of the pan-Arctic land region with a land surface model and ERA-40 reanalysis. *J. Geophys. Res.*, **111**, D05110, doi:10.1029/2005JD006387.

Wentz, F. J., 1997: A well-calibrated ocean algorithm for special sensor microwave/imager. *J. Geophys. Res.*, **102**(C4), 8703-8718, doi:10.1029/96JC01751.

Wentz, F. J. and R. W. Spencer, 1998: SSM/I rain retrievals within a unified all-weather ocean algorithm. *J. Atmos. Sci.*, **55**(9), 1613-1627.

Wentz, F. J., L. Ricciardulli, K. Hilburn and C. Mears, 2007: How Much More Rain Will Global Warming Bring? *Science*, **317**(5835), 233-235.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.