Assessing the skill of satellite-based precipitation estimates in hydrologic applications

Ming Pan,1 Haibin Li,1 and Eric Wood1

Received 12 June 2009; revised 9 February 2010; accepted 3 May 2010; published 28 September 2010.

[1] An important application of global precipitation measurement rainfall products is providing forcing inputs for hydrologic applications, and the goal of this study is to assess how skillful they are for such applications. To do that, we force a land surface model with both satellite estimates and ground-based measurements and test how well they can predict hydrologic states and fluxes useful for water resource applications, i.e., soil moisture, evapotranspiration, and river streamflow. A number of satellite-based precipitation products ranging from retrievals based only on microwave measurements, combined microwave + infrared estimates, to gauge-corrected products are tested over the entire continental United States region. As a reference to the satellite retrievals, estimates from global and regional weather model reanalyses (the analysis fields from these models) are tested as well. It is found that the microwave + infrared combined estimates can match the skill of the coarse resolution European Center for Medium-Range Weather Forecasts global reanalysis but not the regional National Center for Environmental Predictions reanalysis. Gauge corrections to satellite products significantly enhance their skill by greatly reducing the bias in hydrologic predictions, especially over mountainous areas. Rainfall errors are shown to have strong impact on river streamflow predictions and column total soil moisture and relatively weak impact on near surface moisture and evapotranspiration. River streamflow experiments also suggest that satellite rainfall errors are highly correlated in space within the range of one storm system and thus do not reduce in magnitude with spatial scale (basin size).


1. Introduction

[2] Precipitation (P) is probably the most crucial variable in land surface hydrology. In warm areas, liquid precipitation, i.e., rainfall, provides most of the moisture input for hydrologic processes over land. Unless other moisture sources (e.g., snowpack) contribute significantly to the hydrologic system, rainfall serves as the major forcing to the surface dynamics. The estimation of land surface moisture states and fluxes therefore relies on accurate rainfall inputs, especially in the warm season.

[3] However, due to its highly intermittent occurrence and high variability in space and time, measuring precipitation requires dense sampling to achieve high accuracy [Huffman et al., 2007]. While large-scale gauge/radar networks exist in developed regions, ground measurements are sparse over areas like Africa or high latitudes. Therefore, rainfall estimates from spaceborne sensors offer a very valuable (and often the only) source of information for studying weather, climate, and hydrology over sparsely gauged areas. This is one of the most appealing promises that the global precipitation measurement (GPM) mission [Hou et al., 2008] is trying to fulfill.

[4] Nevertheless, it has been reported that satellite-based rainfall estimations have large uncertainties [Gebremichael et al., 2003; Hossain and Huffman, 2008], which is directly related to how the estimates are derived. Rain rates are usually estimated from measurements within two frequency ranges: microwave (MW) and infrared (IR) [Anagnostou, 2004]. Different storm cloud profiles result in different upwelling radiances at MW frequencies, and this makes it possible to relate passive MW measurements to rain rates. Such a MW to rain rate retrieval can be performed through physically based approaches like the widely used Goddard profiling algorithm (GPROF) [Kummerow et al., 1996; Olson et al., 1999]. MW retrievals generally offer reasonable accuracy, but MW sensors are limited to low-orbit satellites and one sensor usually completes no more than two to three full-Earth scans a day. To obtain higher-temporal sampling for more continuous rainfall estimates, IR measurements, available from geostationary satellites, are used to complement MW. Usually, an IR to rain rate relationship is created between IR temperatures and MW-derived rain rates using statistical approaches such as histogram matching [Huffman et al., 2007] or artificial neural networks [Hsu et al., 1997, 1999]. However, the IR to rain rate relationship is much more elusive than the MW to rain rate, because IR emissions do not penetrate the cloud and only the cloud top
brightness is measured by the IR sensor. Estimating rain rates without knowledge of low-level clouds leads to large uncertainties.

[5] In general, the errors in rainfall have a very complicated space-time structure [North and Nakamoto, 1989]. There is a long heritage of studies characterizing and quantifying errors in spatial fields of rainfall measurement, which dates back to the start of weather radar as reviewed in the work of Krajewski et al. [2010]. Traditional metrics like root mean square error, scaling (multiplicative) error, correlation coefficient, hit frequency bias, and false alarm ratio are often insufficient to describe the space-time structure, and many sophisticated models have therefore been proposed. Examples include the error variance separation [Ciach and Krajewski, 1999], the two-dimensional satellite rainfall error model [Hossain and Anagnostou, 2006], statistical validation [Villarini et al., 2009], fuzzy validation [Ebert, 2008], and the works by Astin [1997], Gebremichael and Krajewski [2004], and Steiner et al. [2003], among others. At the same time, validation of individual satellite products have been well performed [e.g., Huffman, 1997; Sorooshian et al., 2000; Kubota et al., 2007; Villarini and Krajewski, 2007; Villarini et al., 2009]. Collectively, the program for evaluation of high-resolution precipitation products (PEHRPP) [Arkin and Turk, 2006] tries to characterize errors in multiple satellite products at different spatial and temporal scales. These error models and validation studies provide very valuable knowledge for hydrology, because the space-time structure of rainfall errors will have a large impact on how the errors would propagate through the hydrologic processes and translate into errors in surface states/fluxes [Lee and Anagnostou, 2004]. Some recent studies [Hossain and Huffman, 2008] have already focused on scales useful for hydrology.

[6] Most of the aforementioned studies focus on rainfall errors by making direct comparisons between satellite and ground gauge/radar. As terrestrial hydrologists are more interested in using satellite rainfall in hydrologic modeling, it is also important to investigate errors in the simulated surface variables caused by noise in the rainfall input. This has been a focus of some recent studies, for example, the work of Harris and Hossain [2008], over a single basin, in the work of Hong et al. [2007] over a set of large river basins around the world and in the work of Su et al. [2008] over several subbasins of La Plata. In this paper, we force a land surface model (LSM) with both satellite and ground reference data over the continental United States (CONUS) region. We then evaluate the skill of the satellite retrievals by how well they can be used to predict a number of hydrologic variables of interest. The major goal is to determine how uncertainties in retrieved rainfall products will affect their use in hydrologic applications. Compared to similar previous studies over a single basin or small region, our work is of a large-scale nature (over CONUS) and encompasses almost all major high-resolution satellite rainfall products that can be considered as forerunners of the GPM-era products. The hydrologic variables to analyze include soil moisture (SM, surface and column total), evapotranspiration (ET), and river streamflow (Q). SM and ET are predicted by the LSM, and Q is calculated from runoff predictions using a river routing model.

[7] Section 2 introduces the LSM and error metrics, and section 3 describes the rainfall products. Sections 4 and 5 present the skill evaluation experiments with discussions and conclusions.

2. Land Surface Modeling and Error Metrics for Skill Assessment

[8] The basic premise of this study is to translate errors in precipitation (P) into errors in hydrologic states/fluxes (SM, ET, and Q) using a LSM. The reason for using a LSM, instead of a simple hydrologic model as in the work of Hong et al. [2007] or Harris and Hossain [2008], is that the scale mismatch between hydrologic processes (<1 km) and satellite rainfall data (>5 km) needs to be addressed. LSMs are developed to provide a dynamical lower boundary for general circulation models (GCMs) by resolving the related moisture and energy exchanges between the atmosphere and land [Wood et al., 1992]. Since the earliest GCMs ran at very coarse resolution, LSMs are designed to address scale issues (e.g., subpixel variability) caused by coarse forcing inputs. Here the variable infiltration capacity (VIC) hydrologic model [Liang et al., 1994, 1999] serves as the LSM to derive the hydrologic states and fluxes. A number of parameterizations are built into VIC to resolve the scale issue, for example, statistically distributed soil water holding capacity and moisture status [Wood et al., 1992], fractional storm area and precipitation redistribution within a grid pixel [Liang et al., 1996], internal temporal downscaling, subpixel vegetation tiles and elevation bands, subpixel forcing adjustment for elevation effects [Liang et al., 1994, etc. These techniques make VIC ideal for working with satellite rainfall. The VIC model has been implemented and validated in a large number of applications at regional [Abdulla et al., 1996], continental [Maurer et al., 2001; Mitchell et al., 2004], and global [Nijssen et al., 2001; Sheffield and Wood, 2007] scales. For this study, VIC calculates surface states and fluxes given daily meteorological inputs (precipitation, daily minimum and maximum air temperatures, and wind speed). VIC is a distributed model in that the predictions are made at each pixel independently. To obtain river streamflow, computed runoff at the grid pixels are routed to locations of interest (usually river gauge stations) using a linear routing model [Lohmann et al., 1996, 1998]. The model parameters used here are taken from the calibration effort in the work of Maurer et al. [2002], whose study domain (CONUS) and resolution (0.125°) are identical to this study. Most parameters like topography, land covers, vegetation properties, and soil properties follow the NLDAS [Mitchell et al., 2004] standard values and parameters describing soil depth, base flow drainage, and infiltration capacity of the soil layers are calibrated against gauge streamflow observations at the outlets of major river basins.

[9] To assess rainfall skill, we first collect the satellite rainfall estimates (PSat) and ground reference data (PGrnd). Then two sets of LSM simulations are performed forced by PSat and PGrnd to obtain two sets of estimates for the hydrologic variables:

\[
\text{LSM}(P_{\text{Sat}}) \rightarrow \{\text{SM, ET, } Q\}_{\text{Sat}}
\]

\[
\text{LSM}(P_{\text{Grnd}}) \rightarrow \{\text{SM, ET, } Q\}_{\text{Grnd}}.
\]
Errors are computed between predictions driven by \( P_{\text{Sat}}^{\text{Grnd}} \) and \( P_{\text{Grnd}}^{\text{Sat}} \). Here we use two error metrics: the mean of the errors (i.e., BIAS) and standard deviation of the errors (STDE), for example,

\[
\text{BIAS}_{\text{Sat}}^{\text{Grnd}} = \text{Mean}(\text{SM}_{\text{Sat}} - \text{SM}_{\text{Grnd}})
\]
\[
\text{STDE}_{\text{Sat}}^{\text{Grnd}} = \text{STD}(\text{SM}_{\text{Sat}} - \text{SM}_{\text{Grnd}}).
\]

The two metrics decompose the errors into two parts, a slow–varying part captured by BIAS and a fast-varying part captured by STDE. These two metrics are computed for all hydrologic variables studied. For river streamflow \( Q \), an additional error metric, the Nash-Sutcliffe coefficient, is computed. The Nash-Sutcliffe coefficient is added to provide insights on the temporal behavior of errors, which is important for river streamflow predictions, but is not revealed by BIAS or STDE.

### 3. Data Sources

[11] The ground rainfall data set \( P_{\text{Grnd}}^{\text{Clim}} \) is based on the gridded precipitation forcing data set [Cosgrove et al., 2003] for the North American Land Data Assimilation System (NLDAS) project [Mitchell et al., 2004]. NLDAS rainfall combines hourly WSR–88D radar analyses from the National Weather Service (NWS) and daily gauge reports (∼13,000 per day) from the Climate Prediction Center (CPC). The NLDAS project aimed to compile the best possible high-resolution (0.125°) hourly observational precipitation data set for CONUS, and its accuracy has been well validated [Luo et al., 2003].

[12] For \( P_{\text{Sat}}^{\text{Grnd}} \), six different satellite-based rainfall products are tested:

[13] 1. Tropical Rainfall Measurement Mission (TRMM) rescaled multisatellite rainfall product version 3B42V6 [Huffman et al., 2007]. This 3 hourly, 0.25° product is based on multisatellite retrievals that combine MW and IR estimates and are rescaled to match monthly gauge observations. The retrieval applies an IR to rain rate relationship using histogram matching.

[14] 2. TRMM real-time multisatellite rainfall product version 3B42RT [Huffman et al., 2003]. This is the real-time version of TRMM-3B42 and as such is not rescaled using gauge data.

[15] 3. CPC morphing (CMORPH) data [Joyce et al., 2004]. CMORPH uses multiple MW (SSM/I, AMSU-B, AMSR-E, and TMI) and IR sensors. CMORPH does not apply an IR to rain rate relationship. Instead, rain rates are derived from MW measurements and geostationary IR images are used to infer motion fields, which are then used to propagate the MW rain fields in space and time.

[16] 4. Global Precipitation Climatology Project (GPCP) One Degree Daily (1DD) product [Huffman et al., 2001]. GPCP-1DD applies the threshold-matched precipitation index (TMPI) to estimate rain rate from IR measurements in low latitudes (40°S–40°N). The threshold and conditional rain rate used in the TMPI are set based on SSM/I derived precipitation frequencies and monthly accumulations from the GPCP satellite–gauge product. GPCP-1DD offers rescaled daily television and infrared observation satellite operational vertical sounder (TOVS) precipitation for higher latitudes.

[17] 5. Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [Hsu et al., 1997; Hsu et al., 1999; Sorooshian et al., 2000]. PERSIANN establishes the IR to rain rate relationship using a neural network function with network parameters updated regularly using TMI 2A12 data.

[18] 6. Rain rates derived from TRMM microwave imager (TMI) only [Kummerow et al., 1996, 2001]. This is a single-sensor retrieval so its temporal sampling properties are sparser than the other products above and is included for comparison purposes.

[19] To form a reference for assessing skill, rainfall data from two reanalysis products are also included:


[21] 2. North American Regional Reanalysis (NARR) [Mesinger et al., 2006] produced at the National Center for Environmental Predictions (NCEP). The NARR assimilates measured (gauge) rain rates, so its rainfall is expected to be very close to \( P_{\text{Grnd}}^{\text{Clim}} \).

[22] Details about the data sources, e.g., the spatial and temporal coverage and resolution, can be found in Table 1. Also see the work of Ebert et al. [2007] for a comprehensive comparison study between satellite and reanalysis products.

[23] To provide a reference of the lower bound of skill, we introduce one more data source called the “climatological” rainfall \( P_{\text{Clim}}^{\text{Clim}} \). \( P_{\text{Clim}}^{\text{Clim}} \) is created by randomly sampling from a long-term rainfall record on a month–by-month basis. When no rainfall “observations” exist (e.g., from either a satellite or ground-based), forecasters often take random samples from the climatological record as a substitute, the
only information available. Thus, $P^{\text{Clim}}$ forms a minimum information baseline and provides a lower bound of skill, if we know the rainfall climatology at that location. The climatological record used here is the ECMWF long-term reanalysis data set ERA-40 [Uppala et al., 2005]. ERA-40 is a global product with multidecadal coverage: September 1957 to August 2002 (more details are given in Table 1). It is an ideal choice for the rainfall climatology because (1) it is long enough to allow stable random sampling and (2) the global coverage makes it easier to extend the study/findings to beyond the CONUS. The month-by-month sampling is done in such a way that, for example, when generating data for January 2000, the January data from another random year in ERA-40 is selected and used. This essentially ensures that the spatial and temporal structures of storms in the $P^{\text{Clim}}$ data set are reasonable.

4. Experiments and Results

4.1. VIC LSM Simulations

[24] Following the procedure defined in section 2, the skill analysis is carried out for all the specified precipitation data sets: ground based, climatological, satellite based, and reanalyses. The VIC model is set up to run at 0.125° resolution for the 9 year period from 1 January 2000 to 31 December 2008. The four required meteorological forcing fields are obtained from the NLDAS hourly forcing data base and aggregated to daily. This forms the forcing

Figure 1. (a) STDE and (b) BIAS in top layer soil moisture in the VIC simulations forced with different rainfall data. Domain averaged STDE and absolute BIAS are given in the boxes.
data set for the ground reference simulation. All other precipitation data products are also aggregated to daily and subsampled to 0.125°. For each VIC run other than using the ground-based forcing data set, we replace the precipitation field in the ground-based forcing data set with the corresponding satellite/reanalysis or climatological rainfall field. Some data products do not cover the first couple of years; therefore, the satellite/reanalysis rainfall data are inserted into the forcing data set only when/where they are available (Table 1). Missing values in the satellite data are padded with zeros (very little missing data exists in all products except TMI). The VIC model is initialized with its land surface states from the NLDAS-based nowcast system established in the work of Luo et al. [2007]. All the error metrics are calculated only over the period and region where the corresponding satellite/reanalysis data are available. Also to avoid skewed sampling for \( P_{\text{clim}} \), random sampling from the 45 year long ERA-40 is performed 5 times to create five 9 year long \( P_{\text{clim}} \) data sets. The five data sets have mutually exclusive samples and are all used to force the VIC model. Error metrics for the climatology (shown later) are the average of the five VIC simulations.

[25] Three gridded output hydrologic variables are analyzed: top layer soil moisture (SM\(_1\), top 100 mm), total soil moisture (SM\(_{\text{tot}}\)) (soil column depth varies in space), and evapotranspiration (ET). Figures 1a and 1b show the STDE and BIAS in the predicted SM\(_1\) using the different rainfall products. The NARR rainfall produces the smallest errors in terms of both STDE and BIAS, and this is anticipated because NARR assimilates ground observations and therefore should

**Figure 2.** (a) STDE and (b) BIAS in column total soil moisture (mm) in the VIC simulations forced with different rainfall data.
be close to the NLDAS rainfall. All the gauge-adjusted multisensor satellite products (TRMM-3B42V6 and GPCP-1DD) have STDE values similar to those using the rainfall estimates from the ERA-Interim global reanalysis with the domain averaged STDE below 3 mm. This result is interesting since it implies that over CONUS, where the upper atmosphere observations are plentiful, global Four-Dimensional Data Assimilation (4DDA) weather model rainfall fields are of a quality equivalent to gauge adjusted satellite retrieved rainfall. Ebert et al. [2007] compared rainfall estimates from weather models and satellites and found that the weather models examined performed better than the satellite estimates during the cool season when nonconvective precipitation is dominant.

Furthermore, in terms of the STDE of SM$_1$ (Figure 1a), multisensor estimates with no gauge adjustments (CMORPH, PERSIANN, and TRMM-3B42RT) perform only slightly worse than the gauge-adjusted estimates, especially in regions without significant topography. However, both the STDE and BIAS of these products degrade sharply over mountainous regions (e.g., the Rocky Mountains). This is especially obvious between the adjusted TRMM-3B42V6 and real-time TRMM-3B42RT (the second and third frames in the first row of Figures 1a and 1b). This indicates a larger uncertainty in the satellite retrievals over mountainous areas. It may also suggest that the storm structures are more complicated over complex terrain (e.g., orographic effects) and thus more difficult to resolve from the low-resolution passive MW.

Figure 3. (a) STDE and (b) BIAS in evapotranspiration (mm) in the VIC simulations forced with different rainfall data.
or low-penetration IR observations. Among the three unadjusted multisensor products, CMORPH and PERSIANN work slightly better than TRMM-3B42RT, but different products have advantages/disadvantages in different regions. This suggests that we should encourage diversity in GPM products.

[27] The single-sensor TMI (bottom right frame in Figures 1a and 1b) performs the poorest in predicting SM$_1$; the STDE is close to and often worse than the climatology-driven simulation, and the negative BIAS is larger than any other products. TMI consists of only about two instantaneous measurements per day, thus misses most of the rainfall events. This suggests that a single sensor is insufficient even to provide a climatological description of rainfall at time scales relevant to soil moisture, say 2 weeks to a month.

[28] Figure 2 shows the STDE and BIAS in predicting SM$_{tot}$. The results for SM$_{tot}$ are very similar to those for SM$_1$ and almost all the observations regarding SM$_1$ are repeated. One striking difference, however, is that the magnitude of BIAS in SM$_{tot}$ is close to its STDE, and the BIAS in SM$_1$ is much smaller than its STDE (see the gray scale legends). This suggests that the column total moisture is more sensitive to rainfall errors than the surface moisture. Here biases in rainfall (usually multiplicative) are translated into biases in soil moisture, though such biases are no longer multiplicative because soil moisture is bounded by its holding capacity. However, as the surface drains very quickly, its memory of rainfall errors is short while large biases can persist in the whole soil column. Furthermore, surface moisture is no longer sensitive to rainfall errors once it reaches saturation, which happens quite often, but deep layers will not be at saturation most of the time.

[29] Figure 3 shows the STDE and BIAS for the ET predictions. The relative performance in terms of both STDE and BIAS among products generally follows that of SM$_1$, but the difference among them is much less obvious. As soil moisture is bounded by holding capacity, evapotranspiration is further bounded by soil moisture availability and also by the potential evapotranspiration (PET), which is a function of radiation and atmospheric conditions. It is therefore even less sensitive to rainfall errors. Similar to the case for soil moisture, gauge adjustments can still significantly reduce the BIAS in ET predictions (comparing the TRMM-3B42V6 and TRMM-3B42RT, second and third in the first row of Figure 3b).

[30] Figure 4 summarizes the CONUS averaged STDE and absolute BIAS. Most of the observations made from Figures 1 through 3 are reconfirmed here. For example, the overall skill among products (judged from both STDE and BIAS) in predicting SM$_{tot}$ and ET follows a similar order of performance to SM$_1$: NARR > gauge-adjusted multisensor estimates and ERA-Interim global reanalysis > unadjusted multisensor estimates > single-sensor retrievals. The comparison between the magnitude of STDE and BIAS (blank and shaded bars in Figure 4) reveals that the BIAS is very significant for SM$_{tot}$, but comparatively less important for SM$_1$ and ET. Also, gauge adjustments contribute to the reduction of both STDE and BIAS. Other than the single-sensor TMI, all satellite products lead to predictions significantly better than the climatological rainfall for all hydrologic

![Figure 4. Domain averaged STDE and absolute BIAS in all experiments.](image)

or low-penetration IR observations. Among the three unadjusted multisensor products, CMORPH and PERSIANN work slightly better than TRMM-3B42RT, but different products have advantages/disadvantages in different regions. This suggests that we should encourage diversity in GPM products.

![Figure 5. The Ohio River basin (highlighted area), flow network, and gauging locations selected for streamflow simulation in the river routing experiment. Gauging station symbols are drawn relative to their corresponding drainage area.](image)
variables investigated here. Further, if we assume the BIAS does persist in time (to be tested yet) and thus correctable through calibration, then large BIAS, especially in column total soil moisture, does not pose a significant problem, and STDE could be a good skill metric by itself.

4.2. River Routing Experiments

Streamflow is another important variable in hydrologic applications. In climatic regions that are relatively wet and warm, streamflow is basically dominated by rainfall. Also, the streamflow value at a specific location integrates the runoff and base flow input over the corresponding drainage basin. Thus, river streamflow and its absolute errors are scale dependent, so an analysis of the errors in discharge should provide insights into the propagation of precipitation errors across spatial scales.

The grided runoff values from the previous VIC simulations are fed into a linear routing model [Lohmann et al., 1998] to calculate the streamflow values at 61 U. S. Geological Survey (USGS) gauging stations in the Ohio River basin (Figure 5). The routing model has constant flow velocity and diffusivity. Selected gauging locations have drainage areas ranging from $10^3$ to $10^5$ km$^2$.

In Figure 6, the STDE, BIAS, and Nash-Sutcliffe coefficient of daily streamflow are plotted against the drainage area at the gauging stations. In the log-log plot of daily streamflow STDE versus drainage area (Figure 6, top), we observe a linear growth of errors with drainage basin size for all rainfall sources. This indicates a linear growth in the integrated rainfall errors with integration area, which implies that the per unit area rainfall errors remain fairly constant.

The same trend is observed in the absolute values of BIAS (Figure 6, middle). This further suggests that the rainfall errors are highly correlated in space such that spatial averaging does not cancel out errors or that the errors accumulate as the runoff is routed through the river network. This behavior is different from variables like soil moisture, whose errors reduce with spatial averaging since the errors are less correlated in space. Note that in order to test the impact of rainfall error correlation, the basins used here are all within the large Ohio River basin (Figure 5) and mostly nested within each other, i.e., sharing some geographical similarities. Hong et al. [2007] performed a similar study over selected nonoverlapping large basins ($>10^5$ km$^2$) over the globe, and they found a decreasing trend of per area streamflow errors against basin size. Our findings are not in conflict with Hong et al. [2007] because the small basins tested here are mostly dominated by single-storm systems and rainfall errors tend to accumulate as basin size grows, while basins tested in the work of Hong et al. [2007] are generally larger than single-storm systems and the storm contributes less to the streamflow errors as basin size grows. If the same analysis is conducted for disjoint large basins in CONUS, we expect findings similar to Hong et al. [2007].

The Nash-Sutcliffe coefficient reflects errors in both the scalar offset and temporal mismatch between two data sets, and it serves as an overall performance metric for streamflow. The value for the coefficient can range from $-\infty$ to 1 with one indicating a perfect match between the model time series and the reference. The coefficient values (Figure 6, bottom) indicate the same general order of performance among products as in SM$_4$, and other variables. The coefficient value varies among small basins for the same product.
but as basin size grows, the coefficient stabilizes, with a very slight improvement over small basins.

5. Conclusions

[35] A series of LSM simulation and routing experiments are carried out using a set of satellite retrieved and reanalysis rainfall products. The resulting predicted hydrologic variables are analyzed to determine the skill of the products. Effort has been made to include a set of diversified rainfall products, but the chosen ones have commonality of the satellite sensors. With respect to the overall skill among the rainfall products, the fine resolution NARR regional reanalysis performs the best, to a great extent because it assimilates rainfall rates. More interestingly, the gauge-adjusted multisensor satellite products and the ERA-Interim coarse resolution global reanalysis rainfall product perform equally. Finally, the unadjusted multisensor products perform less well. Gauge adjustments result in a significant improvement in rainfall skill, especially over mountainous regions. Estimates from a single MW sensor are insufficient to provide skillful predictions in hydrologic applications. With respect to the difference among hydrologic variables, rainfall quality has a relatively weak impact on the upper 100 mm layer soil moisture and evapotranspiration but exerts strong impact on total soil moisture. Streamflow errors are dominated by rainfall quality. The routing experiments suggest that the rainfall errors are highly correlated in space and that errors in the runoff accumulate as it is routed through the river network. This implies that assimilating river observations from the proposed NASA Surface Water Ocean Topography (SWOT) mission would contribute to improved satellite-based water budget studies [Andreadis et al., 2007].

[36] All multisensor products tested here perform significantly better than the climatological rainfall. This confirms that global precipitation measurement from space has already offered great value for hydrology and water resource applications, especially for areas that lack ground measurements. However, depending on the specific purpose of the application, such as drought monitoring or flooding forecasting, continued research is required before satellite rainfall products are skillful enough for operational use. Our work confirms that a thorough understanding of the error in satellite rainfall is critical to any analysis of its skill in hydrologic predictions. Nonetheless, the work offers insights as to the pathway forward, which includes the following: (1) real-time adjustment of the retrieved rainfall to remove bias, which was implemented into the TRMM real-time product as of February 2009 (Adler, personal communication); (2) continued field campaigns are needed to better understand precipitation microphysics and improve the retrieval algorithms, as well as the development and testing of calibration methods for the algorithms. Given that the ERA-Interim global reanalysis rainfall product has equivalent skill to the satellite/gauge adjusted rainfall, the challenge to the satellite rainfall community is clear, yet perhaps merging the products is the best path forward.

References


Arkin, P., and J. Turk (2006), Program to Evaluate High Resolution Precipitation Products (PEHRPP): Contribution to GPM Planning, in 6th GPM International Planning Workshop, Annapolis, Maryland, USA.

Astin, I. (1997), A survey of studies into errors in large scale space-time averages of rainfall, cloud cover, sea surface processes and the Earth’s radiation budget as derived from low Earth orbit satellite instruments because of their incomplete temporal and spatial coverage, Surv. Geophys., 18(4), 385–403.


Steiner, M., T. L. Bell, Y. Zhang, and E. F. Wood (2003), Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar data set, *J. Clim.*, 16(22), 3759–3778.


H. Li, M. Pan, and E. Wood, Department of Civil and Environmental Engineering, Princeton University, Rm. E208, EQuad, Olden St., Princeton, NJ 08544, USA. (mpan@princeton.edu)