Evaluating Satellite Precipitation Error Propagation in Runoff Simulations of Mountainous Basins

YIWEN MEI
Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

EFTHYMIOS I. NIKOLOPOULOS
Innovation Technologies Center, and Department of Physics, University of Athens, Athens, Greece

EMMANOUIL N. ANAGNOSTOU
Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

MARCO BORGA
Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, Padua, Italy

(Manuscript received 7 May 2015, in final form 1 November 2015)

ABSTRACT

This study investigates the error characteristics of six quasi-global satellite precipitation products and their error propagation in flow simulations for a range of mountainous basin scales (255–6967 km²) and two different periods (May–August and September–November) in northeast Italy. Statistics describing the systematic and random error, the temporal similarity, and error ratios between precipitation and runoff are presented. Overall, strong over-/underestimation associated with the near-real-time 3B42/Climate Prediction Center morphing technique (CMORPH) products is shown. Results suggest positive correlation between the systematic error and basin elevation. Performance evaluation of flow simulations yields a higher degree of consistency for the moderate to large basin scales and the May–August period. Gauge adjustment for the different satellite products is shown to moderate their error magnitude and increase their correlation with reference precipitation and streamflow simulations. Moreover, ratios of precipitation to streamflow simulation error metrics show dependencies in terms of magnitude and variability. Random error and temporal dissimilarity are shown to reduce from basin-average rainfall to the streamflow simulations, while the systematic error exhibits no clear pattern in the rainfall–runoff transformation.

1. Introduction

Integration of satellite precipitation products with hydrologic models constitutes a potential solution for simulating hydrological processes at global scale. Such an approach is particularly important for mountainous areas where spatial coverage of precipitation observations is limited because of the generally low number of in situ sensors and the blockage of ground remote sensors due to complex terrain. In this context, satellite sensors offer unique advantages relative to ground sensors since they can provide fine-resolution observations of precipitation at quasi-global scale, uninhibited by mountains or spatial inconsistencies (Arkin and Ardanuy 1989; Kidd et al. 2003; Anagnostou et al. 2010). The significance of these advantages has been recognized by the hydrologic community, and numerous studies have focused on the use of satellite precipitation retrievals in hydrologic applications in the past two decades (e.g., Guetter et al. 1996; Tsintikidis et al. 1999; Grimes and Diop 2003; Yilmaz et al. 2005; Su et al. 2008; Bitew and Gebremichael 2011; Nikolopoulos et al. 2013). Focusing specifically on the satellite-based hydrologic applications over complex terrain, Table 1 provides a representative list of past studies focused on
evaluating the error propagation of satellite products over mountainous areas. It can be seen that before 2010, few studies exist on the topic, while the majority of these studies were designated for coarse temporal resolution (mostly daily or monthly) over medium- to large-scale basins (sizes greater than 1000 km²).

Some of these studies have examined the hydrologic response of heavy precipitation in small-scale complex terrain basins (Hossain and Anagnostou 2004; Nikolopoulos et al. 2010; Bitew and Gebremichael 2011; Bitew et al. 2012; Nikolopoulos et al. 2013). Bitew et al. (2012) and Bitew and Gebremichael (2011) conducted studies for basins in the Ethiopia highlands; Hossain and Anagnostou (2004) and Nikolopoulos et al. (2010; 2013) examined storm events over the northeast Italian Alps that had caused flash floods. Two common observations have resulted from the above studies. First, the accuracy of the satellite precipitation-driven simulations is affected by the storm event severity, the precipitation product resolution, and the basin scale (Gourley et al. 2011; Maggioni et al. 2013; Vergara et al. 2014). Specifically, moderate precipitation magnitudes, finer product resolutions, and larger basin scales are associated with the most accurate hydrological applications. Second, notable improvements in satellite-based streamflow simulations can be obtained after recalibration of the hydrologic model with corresponding satellite precipitation estimates. However, values of the recalibrated model parameters typically lie outside the physical range, indicating the lack of hydrologic representativeness of the parameters themselves (Tobin and Bennett 2009; Yong et al. 2012; Nikolopoulos et al. 2013).

Two questions specific to satellite precipitation applications in mountainous basins arise from the above studies: Does gauge adjustment of satellite products yield consistently lower uncertainty in hydrologic simulations? And, how does the precipitation error propagate through the hydrologic model simulation? For the first question, Hossain and Anagnostou (2004), Nikolopoulos et al. (2013), and most of the recent literature argues that the gauge-adjusted satellite products are more promising than the unadjusted counterparts. Yet, the investigation on the Ethiopia highlands done by Bitew et al. (2012) and Bitew and Gebremichael (2011) showed that this is not always the case; critical factors to the improvements resulting from gauge adjustment are the number and representativeness of included gauges (Wilk et al. 2006; Gourley et al. 2011). Gauge adjustment can be considered as qualified over densely gauged areas (e.g., continental United States or parts of western Europe). On the other hand, this adjustment may introduce unrealistic features in the precipitation products, particularly in areas exhibiting strong precipitation gradients. Regarding the patterns of error propagation, a nonlinear error transformation process is prevailing among the hydrologic studies; that is, the hydrologic models can tolerate a relatively small amount of error by the integrated basin processes, but may amplify this error in high precipitation magnitudes (Guetter et al. 1996; Artan et al. 2007; Yong et al. 2010, 2012). It has also been shown that the properties of error propagation (magnification/dampening and linear/nonlinear) depend on several factors such as antecedent moisture conditions (Nikolopoulos et al. 2011, 2013; Bitew et al. 2012), basin scale (Nikolopoulos et al. 2010; Cunha et al. 2012), and the choice of hydrologic model or modeling complexity (Carpenter and Georgakakos 2006; Zhu et al. 2013).

This study builds upon the above works, providing a comprehensive evaluation of three different satellite products and their gauge-adjusted counterparts and comparing them

<table>
<thead>
<tr>
<th>Study</th>
<th>Watersheds characteristics</th>
<th>Hydrologic simulation</th>
<th>Resolution</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area (km²)</td>
<td>Elev (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li et al. (2009)</td>
<td>12696</td>
<td>1134–2700</td>
<td>Daily</td>
<td>6 yr</td>
</tr>
<tr>
<td>Yong et al. (2010)</td>
<td>18 112</td>
<td>427–2017</td>
<td>Daily and monthly</td>
<td>6 yr</td>
</tr>
<tr>
<td>Nikolopoulos et al. (2010)</td>
<td>108–1200</td>
<td>100–2000</td>
<td>Hourly</td>
<td>1 HPE</td>
</tr>
<tr>
<td>Bitew and Gebremichael (2011)</td>
<td>299 and 1656</td>
<td>1880–3530</td>
<td>Daily</td>
<td>2 yr</td>
</tr>
<tr>
<td>Bitew et al. (2012)</td>
<td>299</td>
<td>1890–3130</td>
<td>Daily and monthly</td>
<td>5 yr</td>
</tr>
<tr>
<td>Li et al. (2012)</td>
<td>15 500</td>
<td>50–2138</td>
<td>Daily and monthly</td>
<td>7 yr</td>
</tr>
<tr>
<td>Jiang et al. (2012)</td>
<td>9972</td>
<td>49–2093</td>
<td>Daily and monthly</td>
<td>6 yr</td>
</tr>
<tr>
<td>Yong et al. (2012)</td>
<td>18 112</td>
<td>427–2017</td>
<td>Daily and monthly</td>
<td>9 yr</td>
</tr>
<tr>
<td>Thiemig et al. (2013)</td>
<td>53 000–130 000</td>
<td>40–287</td>
<td>Daily</td>
<td>6 yr</td>
</tr>
<tr>
<td></td>
<td>76 000</td>
<td>400–3100</td>
<td>Hourly</td>
<td>1 HPE</td>
</tr>
<tr>
<td>Nikolopoulos et al. (2013)</td>
<td>623</td>
<td>1140 (mean)</td>
<td>Daily and monthly</td>
<td>14 yr</td>
</tr>
<tr>
<td>Tong et al. (2014)</td>
<td>121 972</td>
<td>2728–5969</td>
<td>Daily and monthly</td>
<td>7 yr</td>
</tr>
<tr>
<td></td>
<td>137 704</td>
<td>3804–5959</td>
<td>Hourly</td>
<td>12 yr</td>
</tr>
<tr>
<td>Zulkafli et al. (2014)</td>
<td>4640–363 848</td>
<td>1300</td>
<td>Daily and monthly</td>
<td>12 yr</td>
</tr>
</tbody>
</table>
against a reference precipitation dataset derived from a dense gauge network over the upper Adige River basin of the eastern Italian Alps. Given the strong mountainous relief (200–3900 m MSL) of the study area, the representativeness of gauge measurements to spatial precipitation variability introduces error in area-average estimates and should be noted as demonstrated in Nikolopoulos et al. (2015). To evaluate the error propagation in flood simulations, satellite precipitation datasets were used to force a gauge-calibrated hydrologic model to simulate runoff for 16 cascade basins (areas ranging from 255 to 6967 km²) and comparing them to the gauge-driven simulated hydrographs for a range of moderate to high flood events spanning a 9-yr period. Our study brings a more holistic investigation relative to previous studies of the questions posed above by capturing the dependency of satellite-driven flow simulation error on basin scale, basin altitude, seasonality, and product type for different event severities.

In the next section we introduce the study area and precipitation datasets used, while section 3 describes the data processing procedure and details on the hydrologic model. Section 4 presents the categorization method for the time series and the error metrics applied in the following analysis. Results are reviewed in section 5, and conclusions are drawn in section 6.

2. Study area and data

a. Study area

The study area is the upper Adige River basin closed at Bronzolo (~7000 km²), a mountainous region covered by broadleaf and conifer forests located in the eastern Italian Alps (Fig. 1). This region is characterized by steep topographic gradients with elevation ranging from 200 to approximately 3900 m MSL, with mean elevation at about 1800 m MSL. There are 16 cascade basins.
involved in our study with areas ranging from 255 to 6967 km² and mean elevation above almost 1700 m MSL (see Table 2 for a summary). Precipitation in the region is primarily attributed to mesoscale convective systems during summer to early fall and frontal or organized convective systems during fall and early winter (Frei and Schär 1998; Norbiato et al. 2009b). The mean annual precipitation for the period of this study (2002–10) over the region is 788 mm with minimum and maximum annual values of 692 and 912 mm, respectively. The corresponding mean annual and maximum/minimum normalized runoff values measured at the outlet of Adige at Bronzolo (representing the entire drainage area of the study basin) for the same period are 603 mm and 747/482 mm, respectively.

b. Precipitation data

Three quasi-global satellite products and their gauge-adjusted counterparts are used in this study. The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) is a combined IR and microwave (MW) product from the National Aeronautics and Space Administration (NASA). The TMPA (Huffman et al. 2007) is available with a near-real-time version adjusted according to a climatological correction algorithm (CCA; 3B42-CCA, hereafter named TR; Huffman et al. 2010). In addition to the near-real-time product, the postprocessing gauge-adjusted equivalent product [3B42, version 7 (3B42-V7); hereafter named aTR] is used. The second product is the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center morphing technique (CMORPH, hereafter abbreviated as CM), which utilizes multisatellite-based MW rain estimates propagated spatiotemporally by IR-derived motion vectors (Joyce et al. 2004). Recently, a bias adjustment procedure was developed based on daily gauge estimates (30 000 gauges worldwide; Xie et al. 2011) and was applied on the entire CMORPH record to provide the gauge-adjusted equivalent of the CMORPH product (hereafter referred to as aCM). The third product evaluated in this work is the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN, hereafter named PE). This algorithm (Sorooshian et al. 2000) uses a neural network approach calibrated by MW data to derive relationships between IR data and rainfall estimates (Sorooshian et al. 2000). The bias-adjusted version of PERSIANN (hereafter named aPE) is computed based on a correction factor that represents the ratio of Global Precipitation Climatology Project (GPCP) product and PERSIANN estimates at 2.5° and monthly space–time windows (Adler et al. 2003; Huffman et al. 2009). Spatial and temporal resolutions of the satellite rainfall products evaluated in this study are 0.25° and 3-hourly time intervals covering 2002–10.

A total of 104 rain gauges are distributed over the study area (Fig. 1), providing a gauge density of approximately 1/67 [one gauge station per area (km²)] over the whole area. Note that gauge density over the 16 selected cascade basins varied between 1/25 (M1) and 1/67 (the entire basin). Table 2 reports the number of contributing gauges for each basin. The rain gauge rainfall record used in this study covers a 9-yr period (2002–10) at hourly temporal resolution. These rain gauge data have gone through a

<table>
<thead>
<tr>
<th>Basin ID</th>
<th>Area (km²)</th>
<th>Elev (m MSL)</th>
<th>Height (m MSL)</th>
<th>No.</th>
<th>Rainfall May–August</th>
<th>Simulated flow May–August</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>255</td>
<td>1858</td>
<td>10</td>
<td>1302</td>
<td>449</td>
<td>239</td>
</tr>
<tr>
<td>M2</td>
<td>345</td>
<td>1952</td>
<td>11</td>
<td>1402</td>
<td>469</td>
<td>235</td>
</tr>
<tr>
<td>M3</td>
<td>391</td>
<td>1846</td>
<td>9</td>
<td>1287</td>
<td>421</td>
<td>232</td>
</tr>
<tr>
<td>M4</td>
<td>417</td>
<td>1684</td>
<td>12</td>
<td>926</td>
<td>414</td>
<td>279</td>
</tr>
<tr>
<td>M5</td>
<td>417</td>
<td>2113</td>
<td>8</td>
<td>1142</td>
<td>459</td>
<td>689</td>
</tr>
<tr>
<td>M6</td>
<td>505</td>
<td>1910</td>
<td>18</td>
<td>1295</td>
<td>476</td>
<td>522</td>
</tr>
<tr>
<td>M7</td>
<td>613</td>
<td>2035</td>
<td>12</td>
<td>1108</td>
<td>456</td>
<td>575</td>
</tr>
<tr>
<td>M8</td>
<td>653</td>
<td>2181</td>
<td>13</td>
<td>1600</td>
<td>307</td>
<td>261</td>
</tr>
<tr>
<td>M9</td>
<td>892</td>
<td>2162</td>
<td>17</td>
<td>1546</td>
<td>303</td>
<td>260</td>
</tr>
<tr>
<td>L1</td>
<td>1262</td>
<td>1908</td>
<td>23</td>
<td>1180</td>
<td>461</td>
<td>416</td>
</tr>
<tr>
<td>L2</td>
<td>1673</td>
<td>2109</td>
<td>29</td>
<td>1384</td>
<td>309</td>
<td>257</td>
</tr>
<tr>
<td>L3</td>
<td>1906</td>
<td>1857</td>
<td>32</td>
<td>1221</td>
<td>449</td>
<td>277</td>
</tr>
<tr>
<td>L4</td>
<td>2712</td>
<td>1900</td>
<td>44</td>
<td>1296</td>
<td>343</td>
<td>303</td>
</tr>
<tr>
<td>L5</td>
<td>2863</td>
<td>1830</td>
<td>53</td>
<td>1198</td>
<td>447</td>
<td>355</td>
</tr>
<tr>
<td>L6</td>
<td>4166</td>
<td>1743</td>
<td>71</td>
<td>1155</td>
<td>436</td>
<td>309</td>
</tr>
<tr>
<td>L7</td>
<td>6967</td>
<td>1793</td>
<td>104</td>
<td>1216</td>
<td>398</td>
<td>305</td>
</tr>
</tbody>
</table>

Table 2. Summary of evaluated basins and hydrologic characteristics. M and L stand for medium- and large-scale basins; L7 is the entire basin.
quality-control (QC) process according to the guidelines of the World Meteorological Organization (Zahumensky 2010). Values that did not pass the QC were discarded. An average of 14.5% of the rain gauge rainfall records have been discarded as a result of the QC. Also, hourly temperature records are provided by a dense network of stations (143 stations) over the study area (Fig. 1).

3. Data processing and hydrologic simulations

a. Precipitation data processing

All satellite products were spatiotemporally interpolated, using the nearest neighbor method, to derive hourly areal weighted basin-average precipitation for each of the cascade basins analyzed. Additionally, since the TMPA (aTR) represents MW and IR rainfall estimates within ±1.5 h of the synoptic hours, a temporal matching was applied for this product. Specifically, the 3B42 time series were interpolated to half-hourly by the nearest neighbor method, then each two consecutive time steps were aggregated so as to get the hourly records. Hourly gauge precipitation and temperature time series averaged over the 16 cascade study basins were also generated using the nearest neighbor interpolation technique.

b. Hydrologic model

The Integrated Catchment Hydrological Model (ICHYMOD) is used in this study. This is an offline version of the modeling scheme run operationally by the Hydrologic Office of the Autonomous Province of Bolzano as part of the Adige River Flood Forecasting System. ICHYMOD involves a semidistributed conceptual rainfall–runoff model that consists of a snow routine, a soil moisture routine, and a flow routine. This model has been successfully applied in several studies in the greater area of northern Italy (Norbiato et al. 2008, 2009a). A summary of the modeling framework is provided below, while for a detailed description of the modeling structure, the interested reader is referred to Norbiato et al. (2008). Snow accumulation and melting is calculated using a distribution function approach based on a combined radiation index degree-day concept (Cazorzi and Dalla Fontana 1996). Potential evapotranspiration is estimated with the Hargreaves method (Hargreaves and Samani 1982). A probability distribution function (Moore 1985) is used to describe the spatial variation of water storage capacity across the basin. Surface runoff is generated via saturation excess at any point in the basin and is integrated over the basin to derive the total direct runoff entering the fast response pathways to the basin outlet. Drainage from the soil enters slow response pathways and is represented by a function of basin moisture storage (Moore 2007). Total flow at the outlet of the basin results from the summation of storage representations of the fast and slow response pathways. Direct runoff routing is based on the Muskingum–Cunge method (Cunge 1969) while slow or baseflow components of the total runoff are routed through an exponential store.

The model runs at hourly time steps using hourly input of temperature and precipitation. The semidistributed structure of the model allows dividing the modeling area into different subareas (e.g., areas with different hydrological properties), which comprise the computational elements of the model. This permits spatial variability in the model parameterization and allows simulating hydrologic response at several points within a greater basin. Application of the model requires specification of 14 parameters: three for the snow routine, eight for the runoff generation module, and three for the runoff propagation module (Norbiato et al. 2008). The operational version of ICHYMOD, which is an already calibrated version of the model, includes the area of study. The calibration procedure for determining the model parameters is detailed in Borga et al. (2014) and, in general, involved the minimization of error in runoff simulations. In the current work, we adopted the same modeling parameters for our simulations to allow results to be directly related to an existing flood forecasting application. We avoided further calibration of the model because improving the model performance (relative to its current status) is not within the scope of this study.

c. Setup of hydrologic simulations

Within ICHYMOD, our study area was discretized into 51 subbasins (i.e., computational elements) that range in area between 30 and 500 km². Discretization of the area into corresponding subbasins was based on a combination of criteria involving mainly data availability (most outlets coincide with the existence of stream gauge) and hydrologic interest (i.e., areas that are highly prone to floods/flash floods were discretized at a higher detail, that is, the number of subbasins to allow modeling of the hydrologic response at finer spatial scales). For each of these subbasins, precipitation was provided as subbasin-average hourly time series following the procedure described in section 3a. Model simulations were then carried out and the simulated discharge was extracted from model output for the 16 study cascade basins selected for analysis (Fig. 1). This procedure was repeated for all precipitation products (gauge and satellite based) to obtain simulated discharge for each precipitation forcing. An important note is that because precipitation in the region during winter and early spring (i.e., December–April) is dominated by snow, we restricted the simulation period between May and November to avoid, to a large extent, mixing in our analysis satellite estimation errors for both rainfall and snowfall. However, recognizing that snowfall plays an important role in the hydrologic response of the region during the melting period, we were
initializing each simulation cycle (i.e., May–November per year) using the state variables (snow and soil moisture storage) obtained from continuous gauge-based simulations. Note that the continuous gauge-based simulations were performed for the period 2000–10 (i.e., starting 2 years earlier than the analysis period), allowing the first 2 years of simulation to be used for model spinup, which, according to our previous experience with ICHYMOD, has shown to be adequate. This was important for two main reasons. First, runoff generated from the snow-melting process was included (because of available snow storage), thus making simulations realistic for the region. Second, simulations for all precipitation products were performed using the same initial conditions, allowing decoupling observed differences (among products) from the effect of initial conditions and focusing only on rainfall–runoff dynamics.

Throughout the analysis the gauge-based simulations are used as reference for comparison with the satellite-based simulations. Given that gauge-based precipitation is also used as reference for evaluating satellite-rainfall error metrics, using the gauge-based runoff simulations as reference (instead of observed discharge) permit us to directly analyze the characteristics of error propagation from rainfall to runoff. Although, as previously stated, a rigorous evaluation of the efficiency of the model is not within the scope of this study, a comparison with observed discharge was carried out to gain a basic understanding of the efficiency of gauge-based simulations (used as our reference) to represent actual streamflow. The Nash–Sutcliffe (NS) model efficiency coefficient (Nash and Sutcliffe 1970) between reference flow simulations and streamflow observations for the entire basin is 0.80, while the mean NS values for all the study basins is 0.60 (Borga et al. 2014).

While these results indicate considerable variability in model performance for the different cascade basins analyzed, they show that representation of hydrologic response by the reference simulations holds a certain degree of realism. An example of the various flow simulations is shown in Fig. 2, where hydrographs from satellite and reference simulations are shown for the largest and smallest basin (denoted as L7 and M1, see section 4a for the implications) examined and for their corresponding wettest and driest years. Variability in the performance of different products is clearly shown in the figure, and one can immediately recognize distinct product-related behaviors. For example, TR is overestimating in all four cases, particularly for the dry year, while PE and CM are underestimating. A detailed analysis on this is provided in section 5.

4. Methodology

The methodology in this study devises error metrics that evaluate the error structure of satellite products and their driven runoff simulations according to basin scale, basin elevation, seasonality, and event severity. Description of the metrics is provided in section 4b. Next, we describe the classification approach followed to separate the simulated flow data to different cases.

a. Classification

The 16 study basins considered in this study were grouped into two classes according to their area (see summary in Table 2), where basins with area greater (less) than 1000 km² were considered as large-scale (medium scale) basins indicated by the letter L (M). Nine years (2002–10) of precipitation time series and the corresponding simulated discharges from the six satellite precipitation products and the rain gauge network measurements were analyzed for each of the selected basins. Analysis was carried out for two different periods corresponding to warm (May–August) and cold (September–November) season months according to the regional climatologic patterns (note that the terms warm season months and May–August, as well as cold season months and September–November, are used interchangeably in the text). Table 2 lists the mean annual precipitation and runoff accumulations for the selected basins in the two seasons. It is seen that the warm season months’ precipitation/runoff accumulations are almost twice as much as those in the cold season months. A point to note from the table is that in some basins the runoff accumulation is higher than the corresponding basin-average precipitation because of the significant contribution from snow storage.

The hourly runoff values were also grouped according to different percentiles to investigate the error structure in flow simulation for different magnitude levels. The thresholds were defined based on the 90th percentile values determined from reference (i.e., gauge-based streamflow simulations) runoff time series. The determined threshold values for different basin scales and seasons are summarized in Table 3. It is noted that the runoff threshold values are greater in the May–August months than those in the September–November months for most of the study basins, indicating (as expected) that the high runoff regime in this region is higher in the warm than the cold period considered. This is expected given that most of the flash floods (i.e., high peak runoff events) in the region occur during the summer (Marchi et al. 2010).

b. Error metrics

Three evaluation metrics are used to describe the properties of error in basin-average precipitation and simulated runoff. They are the mean relative error (MRE), centered relative root-mean-square error (CRMSE), and correlation coefficient (CC), with the listed forms
MRE = \frac{\sum [S(G \in T_G) - G(G \in T_G)]}{\sum G(G \in T_G)}, \hspace{1cm} (1)

\text{CRMSE} = \left( \frac{\frac{1}{M} \sum \left[ S(G \in T_G) - G(G \in T_G) \right] - \frac{1}{M} \sum \left[ S(G \in T_G) - G(G \in T_G) \right]}{\frac{1}{M} \sum G(G \in T_G)} \right)^2, \hspace{1cm} (2)

CC = \frac{\sum \left[ S(G \in T_G) - \frac{1}{M} \sum S(G \in T_G) \right] \left[ G(G \in T_G) - \frac{1}{M} \sum G(G \in T_G) \right]}{\sqrt{\sum \left[ S(G \in T_G) - \frac{1}{M} \sum S(G \in T_G) \right]^2 \sum \left[ G(G \in T_G) - \frac{1}{M} \sum G(G \in T_G) \right]^2}}, \hspace{1cm} (3)

Variables G and S represent gauge and satellite precipitation time series, respectively. T_G defines the space of values satisfying a given condition (i.e., above or below the threshold value defined from the reference data); and M is the total number of values belonging to T_G from the gauge time series. Note that the corresponding error metrics for runoff can be obtained by replacing the precipitation series (G and S) and the thresholds (T_G) with the runoff series and thresholds calculated from the gauge-based simulated runoff time series. MRE is an error metric measuring the systematic error component with values greater or smaller than zero indicating over- or underestimation, respectively. This is complementary to the CRMSE, which is a metric measuring the random component of error, as bias has been removed. CC is an indicator of the temporal similarity between reference and satellite-derived basin-average precipitation and simulated runoff.

Aiming to demonstrate how error translates from basin-average precipitation to simulated runoff, the error metric ratio (denoted as γ) is used as follows:

\[ γ = \frac{EM_p}{EM_r}, \hspace{1cm} (4) \]

where EM_p and EM_r are the error metrics (i.e., MRE, CRMSE, and CC) determined in our analysis for basin-average precipitation and runoff simulations, respectively, for the different basins and seasons. The values of γ for different error metrics will have different implications. For example, γ for MRE (i.e., γMRE) or CRMSE (i.e., γCRMSE) that is smaller (greater) than one indicates dampening (amplification) of error through its transformation from rainfall to runoff simulations. By taking the absolute value, we neglect the direction of the systematic error so as to focus purely on the error magnitude propagation from precipitation to runoff. The indication of γ values reverses for CC (i.e., γCC), that is, γ above (below) one stands for increase (deterioration) of temporal covariation.

5. Results

An overview of the precipitation and simulated flow accumulations for different products and seasons over the entire study region is shown in Fig. 3. The most striking features from these results are the strong overestimation from TR and the relatively significant underestimation from CM; TR is 90% (for warm period) and 143% (for cold period) higher than the reference precipitation, while CM underestimates by 34% (warm) and 62% (cold). Similarly, TR-forced hydrologic simulations exhibit 106% and 125% overestimation of the reference flow, while CM-driven simulations underestimated the reference flows by 27% and 37% for the warm and cold periods, respectively. The magnitude of error for aTR and aCM is significantly lower than their unadjusted (or climatologically adjusted) counterparts. Specifically, the overestimation in the aTR product reduced to 22% and 18% (18% and 26%) for basin-average precipitation (simulated runoff) for the warm and cold periods, respectively. Performance of aCM was also greatly improved relative to the unadjusted product with 2% and 28% underestimation for basin-average precipitation in the warm and cold periods, respectively, and nearly unbiased results (0.4% and 3%) for the flow simulations. For PERSIANN, the gauge adjustment shifts both precipitation and runoff simulations from underestimation to overestimation. Specifically, PE exhibits 3% and 32% (4% and 14%) underestimation of the warm and cold period basin-average precipitation (simulated flow); aPE gives 19% and 17% (15% and 27%) overestimation for basin-average precipitation (simulated flow) over the two periods. Figure 3 also suggests that the mean annual accumulations of precipitation and runoff simulation for aTR and aPE are almost identical over the study area.
This is expected given that both products are adjusted to the global precipitation-gauge analyses from the GPCP (Huffman et al. 2009).

The following subsections provide more in-depth error analysis of satellite precipitation retrievals, the satellite-based flow simulations, and the characteristics of error propagation from basin-average precipitation to simulated runoff. The analysis was carried out with respect to three main aspects, including 1) dependence of systematic error on elevation; 2) systematic, random error and temporal correlation of the flow simulations; and 3) dependence of error metrics ratios on basin elevation.

a. Role of elevation on systematic error

Given the focus of this study on complex terrain, this section evaluates the role of elevation on the precipitation and runoff simulation error magnitudes. Figures 4 and 5 show scatterplots of biases in precipitation and the corresponding runoff simulations against mean basin elevation values for the two seasons. A positive linear relationship can be seen from most of the scatterplots, pointing to the fact that the systematic error in satellite-derived basin-average precipitation and the corresponding runoff simulations change from under- to overestimation with an increase in basin altitude. The linearity is stronger for the TR precipitation product and the aCM-driven runoff simulations (the coefficient of determination $r^2$ values for the two scenarios are 0.71 and 0.72, respectively). This correlation is not as strong for PE ($r^2$ is generally below 0.3), which may be explained by the fact that this is an IR-only satellite product. It is noted that the increase shown in the 3B42
overestimation with elevation is consistent to results presented in Yong et al. (2010).

From the perspective of error magnitude, we observe overestimation in the two 3B42 products and aPE and underestimation in both CMORPH products and PE (for most of the cases). The strongest over- and underestimations in basin-average precipitation and flow simulations are from TR and CM, which represent combined IR–MW and sole MW retrievals, respectively. Besides, improvements from unadjusted (or climatologically adjusted) to adjusted products are noted as the adjusted products have bias values generally closer to zero (not clear for the PERSIANN product, which is an IR alone–based technique), indicating the importance of improving accuracy by incorporating gauge adjustment to the satellite rainfall estimates (Wilk et al. 2006; Gourley et al. 2011).

From the seasonal perspective, we found a higher degree of consistency in warm season precipitation estimations and flow simulations than in the cold season counterparts, especially for the unadjusted precipitation products (with the exception of the PERSIANN product, where biases of the warm and cold season precipitation and flow simulation are similar, which again could attribute to the fact that PERSIANN is an IR-alone technique). A possible explanation for this seasonal difference is the effect of snow screening applied in the passive microwave satellite retrievals used by CMORPH and 3B42 techniques (Mei et al. 2014). In addition, the study reflects the effectiveness of gauge adjustment in eliminating the high bias in cold season precipitation retrievals.

Based on the trend of the scatterplots in Figs. 4 and 5, it is noted that slopes of the 3B42 products drop

![Fig. 3](image_url)
considerably from the CCA to the adjusted versions for both the precipitation and simulated flow over the two seasons (3.7 and 4.5 times as CCA overadjusted for warm and cold season basin-average precipitation values, while 6.7 and 3.1 times for the simulated flows). This decreasing trend in slopes yields lower magnitudes of systematic error in the basin-average precipitation estimation and corresponding flow simulations for higher-altitude basins. The slope values remain similar for the unadjusted and adjusted versions of the other products. Therefore, gauge adjustment in 3B42 greatly mitigates the strong systematic error over high-altitude basins.

b. Effects of basin scale, seasonality, and flow severity

As mentioned above, goal of this study is to quantify the performance of satellite precipitation-product-driven hydrologic simulations by rendering the MRE (Fig. 6), CRMSE (Fig. 7), and CC (Fig. 8) as a function of basin scale, streamflow magnitude, and seasonal period. Figure 6 demonstrates that there exists strong overestimation from TR and relatively strong underestimation from CM, which is consistent with the results from Fig. 5. It is seen that the flow simulations tend to be more underestimated in higher flow threshold (above the 90th percentile) with wider MRE variability (longer value range). On average MRE values are relatively stable among basin scales examined and for a given product; however, the medium-scale basins exhibit considerably higher variability in MRE values than the larger-size basins of this study. These aspects are especially noted for the extreme flow values (90th percentile). In terms of seasonality effects, results show stronger underestimation for the two PERSIANN and CMORPH precipitation products and a higher degree of variability in MRE values during cold period months for streamflows above the 90th percentile threshold; TR is an exception.

The random component of error in the flow simulation is quantified by the CRMSE (shown in Fig. 7). As a first glance, TR and PE (below the 90th percentile group and warm season) products are characterized with high CRMSE, while those for CMORPH products are mostly below 0.4. High CRMSE values for TR are expected since this product is corrected by the CCA, which has been found to introduce random error on the precipitation estimates (Yong et al. 2013; Mei et al. 2015). In addition, the random component of error is slightly higher for the high

![Fig. 4. Bias in basin-average satellite precipitation products vs mean basin elevation for the different basins and two periods.](image-url)
flows (above the 90th percentile) relative to those below the 90th percentile; TR and PE are two exceptions. Lower CRMSE values with narrower value ranges are found for the larger basin scale for all the adjusted products (and most of the unadjusted ones) regardless of flow threshold pointing to the more significant smoothing effect on random error from larger basin scales. Similar findings were revealed in the Maggioni et al. (2013) and Vergara et al. (2014) studies. Besides, cold season is characterized with higher CRMSE values as anticipated ascribing to the issue of satellite detection of solid- and mixed-phase precipitation during the cold season. A comparison between adjusted and unadjusted products shows that the gauge adjustments are able to reduce the random error in precipitation estimation in most of the cases (this improvement is not apparent for CMORPH), particularly pronounced for the flow simulations below the 90th percentile threshold.

The correlation coefficient was used to quantify the temporal similarity of the hydrologic simulations; the mean and range for CC values are shown in Fig. 8. The figure reveals different behaviors of CC for the two flow threshold groups where lower CC values are observed for the above 90th percentile group. Additionally, cold season is characterized by lower CC values and longer value ranges, implying the influence on temporal dynamics of the solid- and mixed-phase precipitation effect on passive microwave retrieval. Investigation on the scale dependency of the CC value demonstrates that there is a convergence trend in value variability from medium- to larger-scale basins. This means that the performance of satellite-based flow simulation is more consistent, in terms of correlation, for the larger (>1000 km²) basins. Moreover, similar to the improvements found over MRE and CRMSE statistics, the adjusted products give higher CC values in flow simulations. A product-wise comparison indicates that aTR outperformed the other products over the various error metrics determined for the flow simulations. The two CMORPH products gave acceptable temporal correlations during the warm season, but the CC value ranges dropped below zero during the September–November period for the extreme flow threshold group, which indicated the limitation of this technique for cold season precipitation events.

c. Error propagation

Investigation of the basin-scale precipitation-to-runoff simulation error propagation is conducted by
rendering the ratio between the three error metrics following Eq. (4). Figure 9 (top) shows the ratio of MRE indicating how the basin-average precipitation error translates to systematic error through the flow simulations (note that this ratio is defined as the absolute value of the error metrics). As noted from the figure, values of the $\gamma_{\text{MRE}}$ for the two 3B42 products are distributed between 0.5 and 2, which means there is

Fig. 6. Mean (circles) and max/min values (vertical bars) of the MRE of stream flows simulated based on the different satellite precipitation products. Results are shown for two value ranges, two basin-scale categories, and the two periods.

Fig. 7. As in Fig. 6, but for the CRMSE.
no good definition of the direction to either dampening or magnification of the systematic error of precipitation by the model. On the other hand, as most of the ratios for the PERSIANN and CMORPH products are less than one, this suggests that systematic error is mostly dampened through the rainfall–runoff transformation; this is especially noted for aCM. In addition, the values of precipitation-to-simulated-runoff error metric ratios show dependency on basin scale and seasons. Specifically, the midsize basins and cold period month events are characterized with wider $\gamma_{\text{MRE}}$ value ranges. This could be expected since storm systems occurring in the cold period over this area are characterized with a higher degree of heterogeneity (mix of frontal and mesoscale convective systems). Furthermore, the gauge adjustment for CMORPH is shown to give a higher degree of error dampening compared to the other two algorithms.

The error propagation of the random component is quantified by $\gamma_{\text{CRMSE}}$ with results demonstrated in Fig. 9 (middle). Compared to the pattern of $\gamma_{\text{MRE}}$, which shows no obvious tendency in terms of either under- or overestimation, the $\gamma_{\text{CRMSE}}$ values are all below one, highlighting the dampening effect on the random error of basin-average precipitation in simulated flows. Based on the magnitude of $\gamma_{\text{CRMSE}}$, the random component of error in basin-average precipitation is at least 3 times that in simulated runoff values. Also, we observed similar seasonal and basin-scale dependencies in the value ranges of $\gamma_{\text{CRMSE}}$ compared to the behavior of $\gamma_{\text{MRE}}$ (longer value range found over the medium-scale and cold season cases), attributed again to the more heterogeneous storm system characteristics over the cold period for this region. No clear basin-scale or seasonal dependency is revealed for the mean values of the CRMSE ratio. Besides, Fig. 9 shows that gauge-adjusted products have lower $\gamma_{\text{CRMSE}}$ with similar magnitudes compared to the unadjusted products (not very clear for CM). Specifically, for the unadjusted (and TR) products, the most significant (insignificant) error dampening effect is observed for the CM (TR) over the two periods.

The error propagation is also quantified in terms of the ratio between temporal correlations of simulated runoff over those of basin-average precipitation [Fig. 9 (bottom)]. A first observation is that most of the $\gamma_{\text{CC}}$ values are greater than one, pointing to improvement from rainfall to runoff. Besides, we can see that the values and value ranges of $\gamma_{\text{CC}}$ are dropping from the medium- to the large-scale basins, ascribing to possibly the increase in runoff routing timing error as basin scale increases. A season-based comparison reveals that the cold season $\gamma_{\text{CC}}$ values are typically with smaller mean values and longer variability. On the one hand, this observation reflects that the degree of model error dampening effect for the cold period months is not as effective as that for the warm months; on the other hand, error patterns of the winter storm estimation are more complex, introducing higher variability in streamflow simulation error. Moreover, adjusted products are
characterized with higher $\gamma_{CC}$ values (always greater than one), which demonstrates the positive aspect of the correction algorithm. A product-wise comparison indicates that warm season aCM and cold season aPE have more significant error-dampening effects in terms of the temporal similarity contrasting to the other cases.

To summarize, a higher degree of variability in terms of the three error metric ratios is observed for the smaller-sized basins. The cold period months’ ratio values for the random error and temporal similarity are characterized with wider value ranges. The ratio values of temporal correlation are notably higher for the warm period months in the case of midsize basins. It is noted that the random error is dampened through the rainfall–runoff transformation in all cases, while the systematic error showed no significant sensitivity.

### 6. Conclusions

This paper provided a rigorous hydrologic assessment of the error propagation of three widely used satellite precipitation products and their corresponding gauge-adjusted versions over a cascade of mountainous basins in the upper Adige basin of the northeast Italian Alps. Results reported in this study evaluate the error properties of both basin-scale precipitation and the simulated flows derived from a semidistributed hydrologic model, and the links between them, for different basin characteristics (size and elevation) and storm types. Error characteristics were defined in terms of error statistics (MRE, CRMSE, and CC) for precipitation and simulated runoff and ratios of precipitation to runoff error metrics (i.e., $\gamma$). The principal conclusions of this study are summarized below.

The systematic error of products and their corresponding simulations ranged from under- to overestimation as the mean basin elevation increased. Values of the systematic error are closer to zero after the gauge adjustments highlighting the necessity of incorporating the ground-based gauge measurements in satellite precipitation retrievals. This is of particular importance for the cold season precipitation retrievals and for the flow simulations as the performance of all products in the cold month period are shown to be worse than in the warm month period. The two 3B42 products (CCA-adjusted and V7) are characterized with overestimation, while CM is consistently underestimating both basin-average precipitation and flow simulations.
The magnitude of error metrics for flow simulations show that the low to moderate flow rates (below the 90th percentile threshold) are predicted with lower systematic and random errors and a higher degree of temporal similarity compared to the extreme flow rates. Also, the random errors are reducing with converging trends as basin scale increases and from cold to warm season months. Gauge adjustment was able to moderate the random error and increase the degree of temporal similarity in the simulated flows. TR is characterized with the highest random error, while the estimates from the two CMORPH products exhibited the lowest random error; the aTR-driven flow simulations exhibited the highest degree of temporal similarity with the gauge-driven flow simulations.

We evaluated the basin-scale precipitation-to-runoff error propagation by taking the error metric ratios of simulated flow to basin-average precipitation. Overall, the ratios show dependencies on basin scale and seasonality. Ratios from larger basins are characterized with a lower degree of variability. The cold period cases are characterized with a higher degree of heterogeneity exhibited by the wider value ranges of the error metric ratios of the random error and temporal similarity metrics. Furthermore, we showed that the flow simulation’s dampening effect in terms of temporal similarity is more significant for the cold period that is associated with more widespread precipitation systems. Finally, it was shown that gauge adjustment is meaningful in terms of error dampening for nearly all of the cases (seasons and basin scales).

We recognize that our analysis gives particular focus on a limited hydroclimatic and geomorphologic regime; therefore, results are directly relevant only for mountainous regions where precipitation processes are driven by orographic enhancement. However, the study rendered error propagation based on long-term data, accounting for both warm and cold season weather patterns, which represents a wide variety of precipitation events and basin responses (i.e., runoff generation). The potential effect of the physical mechanism of runoff processes (antecedent basin conditions, spatial and dynamic aspects of runoff generation) in the precipitation-to-runoff error propagation is not addressed in this study. This error analysis would require detailed event classification and distributed modeling of the rainfall–runoff transformation. Furthermore, use of satellite precipitation products with higher spatiotemporal resolution [e.g., PERSIANN–Cloud Classification System (PERSIANN-CCS) available at 0.04°/hourly, CMORPH available at 0.1°/half-hourly, and Integrated Multisatellite Retrievals for GPM (IMERG) available at 0.1°/hourly] should be examined to explore the trade-off between increased resolution and retrieval uncertainty in flow simulations.

Acknowledgments. This work was supported by a NASA Precipitation Measurement Mission award (NNX07AE31G). Efthymios Nikolopoulos was supported by EU FP7 Marie Curie Actions IEF (project PIEF-GA-2011-302720). We acknowledge and appreciate Roberto Dinale from the Province of Bolzano for making the gauge data available in this study. Datasets are available online for CCA-adjusted 3B42 (ftp://trmmopen.gsfc.nasa.gov/pub/merged/3B42RT/), gauge-adjusted CMORPH (ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/RAW/0.25deg-3HLY/), gauge-adjusted 3B42 (ftp://disc2.nascom.nasa.gov/ftp/data/s4pa/TRMM_L3/TRMM_3B42/), gauge-adjusted PERSIANN (http://fire.eng.uci.edu/PERSIANN/data/3hrly_adj_cact_tars/), near-real-time CMORPH (ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/0.25deg-3HLY/), and near-real-time PERSIANN (ftp://persiann.eng.uci.edu/pub/PERSIANN/tar_3hr/).

REFERENCES


