Estimation of Satellite Rainfall Error Variance Using Readily Available Geophysical Features

Abebe S. Gebregiorgis and Faisal Hossain

Abstract—The present study addresses the estimation of error variance (mean square error, MSE) of three satellite rainfall products: i) Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) product of 3B42RT; ii) Climate Prediction Center (CPC) Morph (CMORPH); and iii) Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS). Nonlinear regression model is used to fit the response variable (satellite rainfall error variance) with explanatory variable (satellite rainfall rate) by grouping them as function of three key geophysical features: topography, climate, and season. The results of the study suggest that the error variance of a rainfall product is strongly correlated with rainfall rate and can be expressed as a power-law function. The geophysical feature based error classification analysis helps in achieving superior accuracy for prognostic error variance quantification in the absence of ground truth data. The multiple correlation coefficients between the estimated and observed error variance over an independent validation region (Upper Mississippi River basin) and time period (2007–2010) are found to be 0.75, 0.86, and 0.87 for 3B42RT, CMORPH, and PERSIANN-CCS products, respectively. In another validation region (Arkansas-Red River basin), the correlation coefficients are 0.59, 0.89, and 0.92 for the same products, respectively. Results of the assessment of error variance models reveal that the type of error component present in a satellite rainfall product directly impacts the accuracy of estimated error variance. The model estimates the error variance more accurately when the precipitation error components are mostly hit bias or false precipitation, while for a product with extensive missed precipitation, the accuracy of estimated error variance is significantly compromised. The study clearly demonstrates the feasibility of quantifying the error variance of satellite rainfall products in a spatially and temporally varying manner using readily available geophysical features and rainfall rate. The study is a path finder to a globally applicable and operationally feasible methodology for error variance estimation at high spatial and temporal scales for advancing satellite rainfall applications in ungauged basins.

Index Terms—Climate, error variance, geophysical features, rainfall rate, regression model, satellite rainfall, season, topography.

I. INTRODUCTION

Satellite precipitation estimation has made considerable progress over the last few decades in terms of accuracy, resolution, and global coverage (hereafter the word “precipitation” will be used interchangeably with “rainfall”). It has developed from the archaic Global Precipitation Climatology Project (GPCP) [1]–[4] to the present variety of higher resolution rainfall products such as the TRMM Multi-satellite Precipitation Analysis (TMPA) [5]; CMORPH [6], [7]; PERSIANN-CCS [8]; PERSIANN [9]; Rain Estimation Using Forward-Adjusted Advection of Microwave Estimates (REFAME) [10]; and Global Satellite Mapping of Precipitation (GSMaP) [11].

The resolution of satellite rainfall data has also improved from one degree spatial and monthly time scale available in the 1990s to the 0.25° and 3 hourly or higher spatial and temporal scales, respectively which is available today. Through this evolution, satellite rainfall data has made a significant contribution to understanding of the dynamics of earth’s climatic and hydrologic system. Global rainfall data is nowadays routinely produced using a range of variety of satellite retrieval algorithms and techniques [5]–[8], [10], [11] and available for the users and scientific communities. Despite the weakness it may have, satellite rainfall has become indispensable for hydrologic simulation and climate prediction especially where there is no ground observation data.

Uncertainty in satellite rainfall products however remains inherent because of the fundamental constraint posed by the indirect approach of remote sensing. To tackle this inherent shortcoming, numerous approaches have been developed to reduce uncertainty of satellite rainfall estimation. Among the many, some common approaches are: combining multisensors infrared (IR) and microwave (MW) data [5], [12]; merging multi-satellite products with gauge observation [4], [13]; implementing different rainfall screening and retrieval techniques [14], [6], [8]; blending (or merging) different satellite rainfall products based on a priori (diagnostic) hydrologic predictability [15]; fusion of multiscale multisensor precipitation using Gaussian-scale mixtures in the wavelet domain [16]. Regardless of such efforts, non-negligible errors associated with the satellite rainfall products still remains a challenge. There always seems to be room to improve the quality of satellite rainfall data sets [9], [17], [18].

One approach, to aid the application of satellite rainfall data for hydrologic predictions, is to understand the characteristics of errors and their outcomes in hydrologic modeling under various possible scenarios [19]–[21]. As the usefulness of these rainfall data sets relies on users’ knowledge of uncertainty about the product, operational quantification of satellite rainfall uncertainty is a pressing need among data producers and users [22]. The issue of quantitative analysis of uncertainty can easily be addressed in a location where quality-controlled ground observation data is adequately available. However, most parts
of the globe are sparse regions that are not well-covered by gauges or ground radars and others cannot be observed by ground networks (e.g., large water bodies, mountainous and remote desert areas). Therefore, the question is how one can estimate the uncertainty of a satellite rainfall product at any location and time in the absence of ground validation data?

In the past few years, several studies have been reported on satellite rainfall uncertainty [19], [20], [22]–[37]. The main focus of these studies were on investigation of error characteristics, quantification of errors, and propagation and impacts of uncertainty on hydrologic model simulations. Huffman [22] was perhaps the first in formulating a functional relationship for RMS random error using the average rainfall rate and probability distribution parameters associated with the precipitation estimates. The RMS random error investigated in the study was a combination of both sampling and algorithmic error, and it was directly proportional to the rainfall rate as shown in

$$\sigma = \frac{\tau}{N_1} \left(\frac{H - p}{p}\right)^{0.5}$$  \hspace{1cm} (1)

where $\tau$ is the space–time average of precipitation over set of E, $H$ is a function of the shape of probability distribution of precipitation (approximately 1.5 for most of global), $p$ is the frequency of nonzero precipitation in set E, and $N_1$ is the number of independent samples in E. The estimated RMS was found to be reasonably comparable to the observed RMS error [38]. The study considered 2.5° spatial and monthly temporal scales to compute the RMS error [22].

Steiner [34] developed a framework to express the sampling error variance of radar rainfall estimate as function of rainfall rate $R$, domain size $A$, time $T$, and sampling interval $\Delta t$ per

$$\sigma = f \left(\frac{1}{R}, \frac{1}{A}, \frac{\Delta t}{T}\right).$$  \hspace{1cm} (2)

According to this study, the random error due to sampling was assumed directly proportional to sampling time interval and inversely to size of space and time domain and rainfall rate. In the case of Huffman [22], the direct proportional relation between the total random error (which was a combination of both sampling and measurement-algorithmic errors) and rainfall rate was dominated by the existence of measurement-algorithmic error. This suggested that measurement-algorithmic error was the major component of the total uncertainty and it was directly proportional to the rainfall rate. In the current era of significantly improved temporal sampling by the constellation of passive microwave (PMW) sensors, it is fair to claim that sampling error is now a relatively negligible source of uncertainty at the daily or higher timescales.

To explore the algorithmic uncertainty in detail, Tian and Peters-Lidard [19] developed a global map of satellite rainfall uncertainty (which reflected both systematic and random errors) by computing ensemble mean of six different satellite rainfall products. The standard deviation was computed from the mean (i.e., anomalies) as a measurement of uncertainty. The finding revealed that the uncertainty over the ocean was relatively smaller, as expected, when compared over land. Besides, large amount of uncertainty was observed over high latitude during the cold season. It is clear from past studies that the knowledge of uncertainly inherent in satellite rainfall estimates is important for data users and producers. For instance, Huffman [22] indicated that spatially and temporally varying uncertainty is more important than single data set-averaged estimate. The former helps data producers to evaluate the performance of their algorithms and make the necessary adjustment as a function of location, storm systems and seasons. It also assists data users to assess models’ simulation outputs and make more reliable prediction.

However, the nature and magnitude of rainfall errors associated with different satellite rainfall products are not thoroughly investigated and fully addressed to broaden their application at relevant spatial and temporal scales. Unless ground truth data is available, there is no way for the users to know error information associated with 3B42RT, CMORPH, or PERSIANN-CCS products at different part of the globe. Without the fundamental knowledge and guideline on selection of publicly available satellite rainfall products, users rightfully can ask “which product should one use for a specific hydrological investigation at a particular location? Is there a mechanism for users to know the uncertainty of these products without having access to ground truth data at a specific location?” This study seeks an answer to the latter question with a view to further promote satellite rainfall products for hydrological application. The study aims at quantifying the error variance of three satellite rainfall products using regression models classified according to easily available geophysical features of the basin and satellite rainfall estimates. Previous studies on estimation of error variance have not leveraged the role played by readily available geophysical features and hence, this study represents a new contribution to the body of knowledge.

The finding of the current study (estimation of spatially and temporally varying error variance using regression model) is also expected to contribute to the merging of various satellite products to a unified state in ungauged basins. Gebregiorgis and Hossain [15], [39] earlier proposed a merging scheme for different satellite rainfall products based on individual performance. The developed merging scheme used error variance of hydrologic predictability to generate weight factors of individual products in the merging process. The finding suggested that leveraging runoff error variance yielded a more accurate merged product [15]. In a follow-up paper, Gebregiorgis and Hossain [39] reported that the use of both spatial and temporal varying error signatures of hydrologic predictability was more useful in merging products for better hydrologic prediction. Therefore, any space–time estimation of error variance over ungauged regions will be valuable in the merging of rainfall products.

In another work, Gebregiorgis and Hossain [27] showed that investigating rainfall uncertainty based on topography, climate regions and seasons is a systematic approach to understanding the nature and magnitude of errors in satellite rainfall products. Topography has a major effect on climate and formation of precipitation. For instance, mountains can affect climate by changing the patterns of temperature, precipitation, and wind circulation. Based on these fundamental drivers, Gebregiorgis...
and Hossain [27] studied the dependency of satellite rainfall uncertainty on topography and climate. Their study reported that the low land regions of a basin are mostly characterized by missed precipitation while the highland regions are dominated by hit bias (deviation of satellite prediction from ground observation during rainfall detection) and false precipitation. The climate of a region can also be used in segregating the nature of rainfall error attributed to a particular topographic region. For instance, in mountainous region where orographic rainfall is dominant, the rainfall rate falling within the same topographic region could significantly be different in the leeward and windward sides of the mountain. In such situations, the climate type of the region can control rainfall characteristics, which has an impact on the uncertainty of estimated satellite rainfall. Exploring the nature of satellite rainfall errors based on seasons is also essential to understand the effect of other seasonally varying meteorological and geophysical processes (such as temperature, snow cover, land use and land cover) that take place on the land surface.

Based on the above fundamental premise, a similar procedure (use of geophysical features) is adopted in the current study to estimate error variance using mathematical models. In the following sections of this paper, detail description of the study area, data and methodology are presented. Next, the results of the study are discussed. In addition, the performance of the regression model in estimating the error variance is evaluated. Finally, the finding and limitation of this work are summarized together with the future extension of the study.

II. STUDY AREA, DATA, AND METHODOLOGY

The Mississippi River Basin (MRB) is chosen as the study region due to its diverse geophysical features and existence of ground truth data for validation purpose (Fig. 1). MRB has five major sub-basins: Missouri, Ohio, Lower Mississippi, Upper Mississippi, and Arkansas-Red basins. The first three basins are considered for validation purpose and the remaining two are selected for model calibration [Fig. 1(a)]. The basin is delineated into five regions based on topography features [Fig. 1(b)] to develop the regression model framework. Moreover, each region is classified according to the dominant Köppen climate type as shown on Figs. 1(c) and 2. The major land use land cover (LULC) types that inform the user about the geophysical nature of the regions are also shown on Fig. 1(d). Interestingly, the topographic regions, Köppen climate classes, and LULC somehow similar spatial patterns as shown on Fig. 1(b)–(d). Detailed description of regions is presented in Table I.

Three satellite rainfall products, namely 3B42RT, CMORPH, and PERSIANN-CCS, are used to develop error variance regression model over MRB. These satellite rainfall products are widely used, available on near-real time, and are considered fairly high resolution products for satellite-based hydrologic
application. Both 3B42RT and CMORPH, data is available at
0.25° spatial and 3 hourly temporal resolution. For the purpose
of this study, the spatial resolution is downsampled to 0.125° using
local scaling method [40] and the 3 hourly time scale
is aggregated to a daily time step. Likewise, the PERSIANN-
CCS product is aggregated to 0.125° and daily resolution from
a from a 0.04° spatial and hourly resolution.

Although the focus of this study is exclusively on assessing
how well error variance can be modeled mathematically
regardless of scale, the choice of 0.125° as the scale of study
is governed by the following factors. First, the ground truth
data on rainfall is available to us at the 0.125° resolution for the
CONUS region (see the detail in the next paragraph). Second,
0.125° offers a standard compromise between the PERSIANN-
CCS scale of 0.04° and the typical scale of 0.25° for CMORPH
and 3B42RT. Third, the aim of this study is not to compare
products for their ability to be “modeled” of error variance
per se, but rather, to assess the underlying factors of regression
models and geophysical features that can assist in estimating
error variance of satellite rainfall. For detail algorithm and
retrieval technique of each product, readers are referred to
Huffman et al. [5], Joyce et al. [6], and Hsu et al. [8] for
3B42RT, CMORPH, and PERSIANN-CCS, respectively.

To calibrate and validate the regression model on error
variance, the observed error variance is computed using gridded
ground observation rainfall data available from the Washington
University’s Surface Hydrology Group at the 0.125° daily scale.
This data pertains to the contiguous United States (CONUS)
and is derived from more than 7000 stations collected from the
National Oceanic and Atmospheric Administration (NOAA)
at an average density of one station per 700 km². The point
data is gridded using synergraphic mapping system (SYMAP)
interpolation algorithm [41].

A nonlinear regression model framework is developed based
on the following procedure. In the first step, the MRB is
grouped into five regions based on topography. Each topo-
gerographic region is classified into three dominant Köppen climate
classes except for region 1, which is dominated by only one type
of climate class [Fig. 2(a) and (c)]. Thus, the regression model
validation is performed for 13 regions. Each of these scenarios
is further broken down per season leading to
52(= 4 × 13) scenarios. Thus, to estimate the error variance over a pixel and
a given day, the user first needs to have the information on
topography (to classify the region it falls under), climate type
and season of the day and is then guided to the appropriately
calibrated regression model. In the next step, pixel’s observed
error variance and satellite rainfall rate are extracted for each
region during the period of 2003–2006 (validation period). The
satellite rainfall rate is considered as independent (explanatory)
variable; whereas the observed error variance is dependent
TABLE I

<table>
<thead>
<tr>
<th>Region</th>
<th>Area, Km²</th>
<th>Climate</th>
<th>%age area</th>
<th>Remark</th>
<th>Percentage area of land use land cover</th>
<th>Barren</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19068</td>
<td>Cfa</td>
<td>98</td>
<td>2% BS, Dfa</td>
<td>2.7</td>
<td>32.0</td>
</tr>
<tr>
<td>2</td>
<td>155670</td>
<td>Dfa</td>
<td>60</td>
<td></td>
<td>0.2</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cfa</td>
<td>28</td>
<td>1% Dwa</td>
<td>0.6</td>
<td>29.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dfa</td>
<td>11</td>
<td></td>
<td>1.5</td>
<td>42.4</td>
</tr>
<tr>
<td>3</td>
<td>1158034</td>
<td>BSk</td>
<td>50</td>
<td>5% Cfa, Dwa, Dwb, Cfa, Dfa</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dfa</td>
<td>24</td>
<td></td>
<td>0.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dwb</td>
<td>21</td>
<td></td>
<td>0.4</td>
<td>35.1</td>
</tr>
<tr>
<td>4</td>
<td>262140</td>
<td>BSk</td>
<td>39</td>
<td></td>
<td>0.0</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dfa</td>
<td>38</td>
<td>5% Dfa, Dwb</td>
<td>0.1</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dwb</td>
<td>18</td>
<td></td>
<td>0.6</td>
<td>35.8</td>
</tr>
<tr>
<td>5</td>
<td>57103</td>
<td>Dic</td>
<td>61</td>
<td></td>
<td>0.3</td>
<td>42.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dfb</td>
<td>22</td>
<td>6% BS, Dsc</td>
<td>0.0</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>11</td>
<td></td>
<td>0.0</td>
<td>30.0</td>
</tr>
</tbody>
</table>

Topography description

Region 1: 0 – 100 m above sea level (a.s.l)
Region 2: 100 – 500 m
Region 3: 500 – 1500 m
Region 4: 1500 – 2500 m
Region 5: 2500 – 4500 m

Land use land cover description

Forest: evergreen needle leaf, evergreen broad leaf, deciduous needle leaf, deciduous broad leaf, mixed forests
Shrub land: closed and open shrub lands
Savanna-grassland: savannas, grasslands, woody savannas
Croplands: cultivated croplands (irrigated land)
Developed: Urban and built-up
Barren: Barren or sparsely vegetated

Köppen climate description

BSk: Mid-latitude steppe - Mid-latitude dry
Cfa: Humid subtropical - Mild with no dry season, hot summer
Dfa: Humid continental - Mild with severe winter, no dry season, hot summer
Dfb: Humid continental - Mild with severe winter, no dry season, warm summer
Dfc: Subarctic - Severe winter, no dry season, cool summer
Dwa: Humid continental - Humid with severe, dry winter, hot summer
Dwb: Humid continental - Humid with severe, dry winter, warm summer
Dsc: Continental subarctic or Boreal climate
H: Highland climate

III. RESULTS AND DISCUSSION

As seen on Fig. 2(b), the error variance is expressed as a function of satellite rainfall rate. It is, therefore, important to identify the spatial distribution of error component of satellite rainfall to understand its impact on error variance estimation using the regression model. On Fig. 3, 3B42RT and CMORPH products have significant missed precipitation during the winter and summer seasons in the eastern and southern part of Mississippi basin (region 1, 2 and part of region 3). Thus, the regression model expected to underestimate or produce nil error variance in these regions.

The error variance of three satellite rainfall products is quantified at the 0.125° spatial and daily temporal scale for the period of 2003–2010. Fig. 4 presents the observed and estimated error variance spatially and seasonally averaged over topography-climate regions during the validation period (2007–2010) for 3B42RT. Generally, the magnitude of the error variance is small in region 4 and 5. This is directly related to the considerably lower rainfall intensity in these regions (Fig. 3). The observed and estimated error variances have also good agreement in these regions. However, the model underestimates the error variance in region 1-Cfa, 2-Cfa, and 2-Dfa due to the presence of missed precipitation (Fig. 3) in this region [26]. These regions are mainly dominated by cropland, forest and savanna-grassland systems.
Fig. 3. Seasonal average satellite rainfall rate and the total error components (total bias, missed-rain bias, hit bias, false-rain bias) for the winter and summer seasons.

Fig. 4. Estimated and observed error variance for 3B42RT satellite rainfall product, spatially and seasonally averaged over topographic and Köppen climate regions of MRB for the validation period (2007–2010). Region code is based on Table I. (Wn: winter; Sp: spring; Sm: summer; Fl: fall seasons).
Fig. 5. Same as Fig. 4, except for CMORPH rainfall product.

As seen in Fig. 5, the estimated error variance from CMORPH is also underestimated in regions 1-Cfa, 2-Dfb, and 3-Dfa, mostly as a result of missed precipitation (Fig. 3). These regions are mostly characterized by forest, cropland and savanna-grassland system. But the model slightly overestimates in regions 3-BSk and 3-Dfa (both regions are dominated by savanna-grassland system) due to the existence of positive hit bias for CMORPH product, reported earlier by Gebregiorgis et al. [26] or refer to Fig. 3. On Fig. 6, the error variance estimated from PERSIANN-CCS is relatively more accurate for regions 1 and 2. But in regions 3-BSk, 3-Dfa, and 4-BSk, this product over predicts the error variance especially during the winter and spring seasons. In regions 4-Dfb, 4-Dfc, and 5-H, it overestimates the error variance during winter season. These regions are also mostly covered by savanna-grassland systems. This is mainly caused by false precipitation and positive hit bias during the cold seasons (Fig. 3). In general, the above results reveal that the regression model performs well and captures the trend of observed error variance in the region where hit bias and false precipitation are dominant components of the error. However, if the region is dominated by missed precipitation, the model’s predictability is compromised because of the satellite data reporting extensive zero rainfall rates. This is one of the main limitations of the power-law type multiplicative type error model used in the study (discussed later).

Fig. 7–9 present the time series of error variance spatially averaged over the major Mississippi sub-basins (calibration and validation regions) for the entire period of simulation (2003–2010). A 31-day moving average is applied to the time series data (observed and estimated error variances) to reduce visual cluttering. In case of 3B42RT product, the trends of the observed and estimated error variance are closely similar in Missouri basin during the entire period (Fig. 7). This is not true, however, for Lower Mississippi, Ohio, Upper Mississippi, and Arkansas-Red basins. The model fails to capture the peaks of observed error variance, particularly during the cold seasons. On the other hand, in case of CMORPH product (Fig. 8), the model displays outstanding performance for all sub-basins during the entire period of validation. For PERSIANN-CCS, the drift of observed and estimated error variances in all sub-basins is qualitatively similar (Fig. 9). However, the model overestimates the error variance in Missouri basin almost during the entire period due to the inherent problem of false precipitation in PERSIANN-CCS as reported earlier by Gebregiorgis et al. [26] (also see Fig. 3). In general, this reinforces that the main type of error component (hit, miss or false precipitation) that is associated with a particular product directly affects the performance of the regression model at a given location.

The residual error variance (which is defined in this case, as the difference between estimated and observer error variance), computed over the entire basin, is shown on Fig. 10. Residuals
Fig. 6. Same as Fig. 4, except for PERSIANN-CCS rainfall product.

Fig. 7. Estimated and observed daily time series of error variance for 3B42RT rainfall product, spatially averaged over the selected calibration (top two panels) and validation sub-basins (bottom two panels) during the period of 2003–2010. A 31-day moving average is applied to reduce visual cluttering.
provide a general idea of the required unbiased nature of a predictive model. In general, the modeling approach seems to be relatively unbiased for all satellite rainfall products during the independent validation period (2007–2010). It captures the spatial pattern of the observed error variance. The unbiased regions mostly show non-zero actual rainfall records (ratio of error variance to ground rainfall (EV/GR) is greater than zero). The estimated error variance shows quantitative offset from the observed in range of $-400$ to $400$ (mm/day)$^2$ (or $-20$ to $20$ mm/day in standard deviation). As expected, 3B42RT is a little skewed toward negative residual error variance because of more missed rain.

Fig. 11 shows the comparison of spatially and seasonally averaged observed and estimated error variance over the developed topography-climate regions and seasons. In case of all products, the estimated and observed error variances are reasonably comparable. In case of 3B42RT, the computed mean error variance in region 1-Cfa is underestimated during winter.
Fig. 10. Estimated, observed, and residual error variance and ratio of computed error variance to ground rainfall value for three satellite rainfall products over the MRB for four randomly selected days during the validation period (2007–2010). (EV: error variance; GR: ground rainfall).

Fig. 11. Mean observed and estimated error variance of satellite rainfall products during the calibration period (2003–2006) for the 13 topographic and climate regions and 4 different seasons. The x-axis on the lower panel shows the topographic region and climate types, and on the middle panel it displays the seasons for each region.

and fall seasons; for CMORPH, the estimated error variance is larger than the observed in region 2-Dfa and 2-Cfa during the summer season; and for PERSIANN-CCS, the predicted mean variance is underestimated only in region 1-Cfa during fall season. In general, the model shows good performance for CMORPH and PERSIANN-CCS products.

To understand the proportion of variance of the dependent variable (error variance) explained by the independent variable (rainfall rate), the following parameters are computed: the total variance of the dependent variable (total sum of squares, TSS), the proportion of variance due to the residuals (error sum of squares, SSE), and the proportion of variance due to the
The mean relative bias (MBias) also measures the agreement between observed and estimated error variance. The allowable limit for MBias is between $-40\%$ and $40\%$ with zero as the ideal value [42]. Table II illustrates the correlation coefficient (SSR/TSS) and mean relative bias (MBias) based on sub basins. Accordingly, the model predicts the error variance very well over Missouri, Lower Mississippi and Ohio basins using CMORPH and PERSIANN-CCS products; whereas the predic-

<table>
<thead>
<tr>
<th>Sub-basin</th>
<th>3B42RT SSR/TSS</th>
<th>CMORPH MBias</th>
<th>PERSIANN-CCS MBias</th>
<th>MBias</th>
<th>SSR/TSS</th>
<th>MBias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missouri</td>
<td>0.63</td>
<td>46.0</td>
<td>0.83</td>
<td>14.4</td>
<td>0.80</td>
<td>-10.0</td>
</tr>
<tr>
<td>Lower Mississippi</td>
<td>0.80</td>
<td>36.3</td>
<td>0.74</td>
<td>12.0</td>
<td>0.63</td>
<td>-15.4</td>
</tr>
<tr>
<td>Upper Mississippi</td>
<td>0.75</td>
<td>15.9</td>
<td>0.86</td>
<td>-3.8</td>
<td>0.87</td>
<td>-46.8</td>
</tr>
<tr>
<td>Arkansas-Red basins</td>
<td>0.59</td>
<td>45.6</td>
<td>0.89</td>
<td>35.9</td>
<td>0.92</td>
<td>-32.1</td>
</tr>
<tr>
<td>Northwest basin</td>
<td>0.85</td>
<td>-34.0</td>
<td>0.48</td>
<td>17.0</td>
<td>0.60</td>
<td>-37.0</td>
</tr>
<tr>
<td>Northwest basin **</td>
<td>0.63</td>
<td>66.0</td>
<td>0.44</td>
<td>-53.0</td>
<td>0.51</td>
<td>83.0</td>
</tr>
</tbody>
</table>

* With geophysical feature based classification
** Without geophysical feature based classification

To look further into the performance of the regression model, the error variance has been estimated for Northwest basin. Fig. 13 presents the spatial error variance distribution of the three satellite rainfall products for the four seasons over the Northwest basin. In this case, the regression model that considers geophysical features is implemented. For 3B42RT (Fig. 13 top three panels), the model captured the spatial distribution pattern adequately over entire region despite quantitative accuracy of prediction. The model performance for CMORPH (Fig. 13 middle three panels) is similarly good except it has underestimated the error variance during the winter season. As
Fig. 13. Computed, observed, and residual seasonal average error variance for three satellite rainfall products over the NWB for the period of 2004–2005. The regression model is developed based on geophysical regions (topography and Koppen climate).

Fig. 14. Same as Fig. 13, except the regression model is developed without the consideration of geophysical features.

During the summer season, Table II also presents the SSR/TSS and MBias of the estimated and observed error variance for the Northwest basin. The regression model based on classification of geophysical feature demonstrates good performance as compared to without region classification. The former scenario shows that the SSR/TSS is 0.85, 0.48, and 0.60 for 3B42RT, CMORPH, and PERSIANN-CCS, respectively. Without the classification of geophysical feature, the SSR/TSS is 0.63, 0.44, and 0.51 for the respective rainfall products.

One of the limitations for independently testing the regression model is that some of the climate types in the Northwest basin are not available in MRB. Therefore, matching of closely similar climate type was made in error variance estimation procedure. The Koppen climate classes specific to the Northwest basin (NWB) includes BWh (dry subtropical desert), BSh (dry subtropical steppe), Csa (Mild mid-latitude with dry, hot summer), Csb (Mild mid-latitude with dry, warm summer), and Dsb (Sever mid-latitude continental with warm summer). For instance, the dominant climate class in region 2 of NWB Csb
Fig. 15. Estimated and observed daily time series of error variance spatially averaged over NWB for the period of 2004–2005 using the regression model developed with and without geophysical region classification (upper and lower panels, respectively). A 31-day moving average is applied.

is matched with Cfa of MRB; Dsa in region 4 of NWB with Dfb of MRB and so on. Such a “mapping” could affect the estimation of error variance. Therefore, it is recommended to develop the regression model on a large-scale basin where the entire spectrum of the diverse geophysical features is existent.

IV. CONCLUSION AND RECOMMENDATION

As satellite rainfall estimates become more important for hydrologic and atmospheric applications, users’ knowledge on uncertainty associated with the satellite rainfall product is a
necessary step to advance its application. A simple nonlinear regression model has been developed for 3B42T, CMORPH, and PERSIANN-CCS products to estimate the error variance at the 0.125° spatial and daily temporal resolution. Topography, climate and seasons are considered as readily available geophysical features for enhancing the predictive ability of

Fig. 16. Computed value of scaling factor ($\alpha$) for different regions of MRB for the three satellite rainfall products.

Fig. 17. Computed value of power or exponent ($\beta$) for different regions of MRB for the three satellite rainfall products.
the model at any location. In general, topography plays a direct role on the pattern of climate of the region due to its forcing on the formation of clouds, temperature and albedo. These climatic and weather-scale processes strongly influence the effectiveness of indirect approach of the remote sensing measurement technique. Therefore, use of topography, climate and season as major governing factors in the development of regression framework is logical to identify the uncertainty type associated with satellite rainfall estimates.

The findings of this study can be summarized into the following three major points.

1) The error variance (EV) has a strong correlation and is directly proportional to the rainfall rate (RR). The relation between the two variables can be adequately expressed by a power function \( EV = \alpha (RR)^{\beta} \), where \( \alpha \) and \( \beta \) are constant real numbers. The parameter \( \alpha \) serves as simple scaling factor, moving the values of \( (RR)^{\beta} \) up or down as the value of \( \alpha \) increases or decreases, respectively. The parameter \( \beta \) is called the exponent or power that determines the rate of growth or decay and also shape and behavior of the function.

2) In general, the parameter \( \alpha \) is high for 3B42RT, moderate for CMORPH, and small for PERSIANN-CCS products (see Fig. 16). For all products, \( \alpha \) gradually decreases from lowland to highland regions. The value of \( \beta \) is found to be high for PERSIANN-CCS product and there is no significant variation among different regions. For 3B42RT and CMORPH, \( \beta \) shows considerable variation from season to season without obvious systematic trend across the regions (see Fig. 17).

3) The type of error components (missed rain, hit, and false-rain biases) that is present in the satellite rainfall estimates has a direct impact on the performance of regression model. The model estimates the error variance more accurately when hit or false precipitation is dominant in the product. On the other hand, the presence of large missed precipitation, makes a product less amenable for error variance estimation.

The key limitation of this study is the model’s inability to predict the error variance accurately in a region where missed precipitation is dominant. If satellite product predicts zero rainfall, then the estimated error variance will be zero by design. Therefore, further exploration is needed to know the location where missed rain is more likely to occur and include other independent (additive) variables in the regression model to estimate the error variance in such conditions. Moreover, the proposed method has a conceptual limitation when the precipitation errors depend heavily on other factors such as physical properties of rain systems. Further exploration on these issues will help improve the concept of this study for practical applications.

On the basis of the promising results reported herein, further investigation into the impact of diverse geophysical features on the performance of regression model by extending the study region to a global scale is now appropriate. Future investigation should also target the quantification of the probabilistic behavior of missed rain as a function of terrain, climate and satellite rain rate. Such an assessment may allow proxy adjustments to avoid the aforementioned limitation of zero error variance prediction.

In summary, high resolution and multisensor satellite-based precipitation estimates, such as those analyzed in this study and those anticipated from the Global Precipitation Measurement (GPM; http://gpm.gsfc.nasa.gov) satellites, now hold great promise for hydrologic applications, especially over parts of the world where surface observation networks are sparse, declining or non-existent. However, the usefulness of such precipitation products for hydrological applications will depend on their error characteristics and how successful we are in intelligently harnessing the implications of uncertainty for surface hydrology. The decline of the few existing global ground based measurement networks for rain and stream flow and the absence of in-situ measurement in most parts of the world represent a “paradoxical” situation for evaluating satellite rainfall estimation uncertainty. By developing simple models for estimation of error variance for satellite data that a user can use anywhere and anytime using only readily available geophysical features, our study represents a first comprehensive step at resolving the paradox for the GPM era.

APPENDIX A

MATHEMATICAL FORMULATION OF REGRESSION MODEL

Regression is a highly useful statistical method for developing a quantitative relationship between dependent variable and one or more independent variables. The dependent variable often is called response or predicted variable. The independent variables that explain the response variable are called explanatory or predictor variables. In this paper, the dependent variable is defined as error variance, EV and the independent variable is satellite rainfall rate, RR.

The proposed nonlinear functional relation between the two variables is expressed as

\[ EV = \alpha (RR)^{\beta} \]  

(A1)

where, \( \alpha \) and \( \beta \) are constant real number and are called scaling factor and power, respectively. The nonlinear equation can be converted to linear equation by applying logarithmic function in (A1)

\[ \log(EV) = \log(\alpha) + \beta \log(RR). \]  

(A2)

Let \( y = \log(EV), \alpha_o = \log(\alpha), \alpha_1 = \beta, \) and \( x = \log(RR), \) then (A2) can be written as form of linear regression form.

\[ y = \alpha_o + \alpha_1 x. \]  

(A3)

For \( i^{th} \) observation, the predicted value can be written as:

\[ Y_i = \alpha_o + \alpha_1 x_i. \]

If \( EV_o \) is the observed error variance then the logarithmic form of the observed error variance for the \( i^{th} \) observation, \( y_{oi} \)

\[ y_{oi} = \log(EV_o). \]
Applying least square method (LSM) to minimize the error sum of the squares (SSE) 
\[ SSE = \sum_{i=1}^{n} (y_{oi} - \alpha_0 - \alpha_1 x_i)^2. \] 

(A4)

To minimize the sum of error square differential (A4) and equate to zero

\[ \frac{\partial (SSE)}{\partial \alpha_0} = \sum_{i=1}^{n} (-2(y_{oi} - \alpha_0 - \alpha_1 x_i)), \]
\[ \frac{\partial (SSE)}{\partial \alpha_1} = \sum_{i=1}^{n} (-2x_i(y_{oi} - \alpha_0 - \alpha_1 x_i)). \]

Equations (A5) and (A6) are called normal equations. Carrying out the differentiation, we obtain

\[ n\alpha_0 + \alpha_1 \sum x_i = \sum y_{oi}, \]
\[ \alpha_0 \sum x_i + \alpha_1 \sum x_i^2 = \sum y_{oi}x_i. \]

where all the summation go from \( i = 1 \) to \( i = n \). The solution to these normal equations can be given as

\[ \alpha_0 = \overline{y} - \alpha_1 \overline{x} \quad \text{and} \]
\[ \alpha_1 = \frac{n \sum x_i - \sum x_i \overline{x}}{n \sum x_i^2 - \sum x_i^2 \overline{x}}. \]

Finally, convert the logarithmic scale back to normal

\[ \alpha = \text{anti} - \log(\alpha_0) \]
\[ \beta = \alpha_1. \]

For 3B42RT, CMORPH, and PERSIANN-CCS products, the value of \( \alpha \) and \( \beta \) are computed for the developed topography-climate regions as shown in Figs. 16 and 17, respectively.

**ACKNOWLEDGMENT**

The authors are grateful to the hydrology research group at the University of Washington for providing ground rainfall data. The authors also acknowledge the technical support received from M. Renfro of the Computer-Aided Laboratory at the Center for Manufacturing Research, Tennessee Technological University.

**REFERENCES**


