AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR020593

Key Points:
- A novel validation framework for remote sensing data is presented
- Proposed framework compares well with traditional validation methods
- Application of the framework to a data-scarce catchment is presented

Correspondence to:
A. Koppa,
akashkoppa@ucla.edu

Citation:
Koppa, A., & Gebremichael, M. (2017). A framework for validation of remotely sensed precipitation and evapotranspiration based on the Budyko hypothesis. *Water Resources Research*, *53*, 8487–8499. https://doi.org/10.1002/2017WR020593

Received 15 FEB 2017
Accepted 20 SEP 2017
Accepted article online 25 SEP 2017
Published online 25 OCT 2017

A Framework for Validation of Remotely Sensed Precipitation and Evapotranspiration Based on the Budyko Hypothesis

Akash Koppa[1] and Mekonnen Gebremichael[1]

[1]Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

Abstract Despite offering spatially and temporally continuous measurements, the use of remotely sensed P and E in data-scarce catchments is hindered by the lack of ground-based measurements that enable comprehensive validation. This study proposes a novel validation framework that characterizes the combined error in the long-term average estimates of remotely sensed P and E by making use of the Budyko hypothesis, specifically Fu's equation. A Root Mean Square Error (RMSE)-based error metric that is capable of translating individual biases in P and E estimates onto the Budyko space is developed. A controlled sensitivity experiment using data from Model Parameter Estimation Experiment (MOPEX) catchments in the United States showed that the developed error metric is more sensitive to biases in P compared to biases in estimates of E. Validating the framework using combinations of different satellite-based estimates of P and E revealed that the framework succeeds in arriving at the same conclusions as a traditional validation method with regards to the quality of P and E data sets. The framework offers a physically consistent, parametrically efficient basis for the selection of remotely sensed P and E data sets for hydrologic studies. Due to lack of consideration for catchment storage in the formulation of Fu's equation, the developed error metric is limited to long temporal time scales. As a result, the error metric is capable of characterizing the bias in P and E data sets and not the variance.

1. Introduction

Advances in remote sensing have enabled continuous monitoring of water and energy fluxes at spatial and temporal scales appropriate for a wide range of hydrologic applications (Lettenmaier et al., 2015), especially in ungauged basins (Lakshmi, 2004). However, these estimates are subject to large uncertainties arising, primarily, from differences in retrieval algorithms and sensor types (Gebregiorgis & Hossain, 2014; Kidd & Huffman, 2011). For example, the Precipitation Intercomparison Project-3 (PIP-3), which evaluated a number of satellite-based precipitation products concluded that, in general, Passive Microwave (PMW)-based products are better than Infrared (IR)-based products at shorter time scales (Adler et al., 2001). It is also seen that the performance of data products vary considerably in space, over different terrains. In a study by Hirpa et al. (2010), it was found that the Climate Prediction Center Morphing Technique (CMORPH) data set performed satisfactorily over a complex terrain in Ethiopia whereas the same data set failed to capture the temporal and spatial variation of rainfall over an urban region in China (Chen et al., 2014). Therefore, the use of remotely sensed data must be subject to comprehensive validation in the catchments of interest.

But, validation efforts are often hindered by the lack of ground-based observations in ungauged and data-scarce catchments. With substantial decline observed in in-situ precipitation and runoff measurement networks across the globe (Lorenz & Kunstmann, 2012; Stokstad, 1999), robust methodologies that can test the physical consistency of observational data using hydrologic principles can offer insight into their quality. Several studies make use of physically based hydrologic models to validate remote sensing data. These studies are based on the hypothesis that any error in the input precipitation will result in a commensurate error in the closure of water balance. An effort to carry out the water balance of the Mississippi river basin using data exclusively from satellite remote sensing revealed that biases in precipitation resulted in overestimation of runoff (Sheffield et al., 2009). Stisen and Sandholt (2010) used a distributed hydrological model, calibrated using observed discharge, to evaluate precipitation data sets derived from satellite-borne sensors over a data-scarce catchment in Senegal. In addition to requiring streamflow observations, the main drawback of such validation studies is the use of calibrated and over-parameterized models which prohibit drawing meaningful conclusions regarding the quality of the data set (Bitew & Gebremichael, 2011).

In this study, we address the issues of requiring concurrent streamflow observations, calibration, and over-parameterization of models by invoking the Budyko hypothesis (Budyko, 1974) for validating remotely sensed water and energy balance components, specifically precipitation and evapotranspiration. The Budyko hypothesis is a semiempirical model that describes the long-term combined water and energy balance of catchments. In recent years, the hypothesis has witnessed widespread application in climate change (Greve et al., 2014), land use change (Zhou et al., 2015), and ecohydrological (Gentine et al., 2012) studies. Through this study we intend to rigorously test the applicability of the Budyko hypothesis to validation of observational data derived from remote sensing. Specifically, we intend to answer the following questions: (1) Can the Budyko hypothesis be used to develop a data validation framework that does not require concurrent ground-based measurements? (2) How does the validation framework compare with traditional methods of evaluation that make use of in-situ measurements?

2. Study Area and Observational Data

For the development and validation of the proposed framework, the Model Parameter Estimation Experiment (MOPEX) catchments, consisting of 438 small to medium size basins spread across the United States are selected (Figure 1a). For these catchments, MOPEX provides ground-based daily observations of precipitation (P), runoff (Q), and climatological potential evapotranspiration (E_p) spanning 56 years (1948–2003) (downloaded from ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/). The size of the catchments vary from 70 to 10,000 km^2. Based on continuous availability of data for at least 30 years, 393 out of 438 catchments were selected for the development of the framework. Evapotranspiration (E) is calculated as (P-Q) at annual time scales based on previous studies on MOPEX catchments (Greve et al., 2015; Li et al., 2013).

3. Development of the Validation Framework

3.1. Budyko Space

The Budyko space is a two-dimensional space in which every point is described by two dimensionless indices; evaporative index (E/P, abbreviated as EI) and Aridity Index (E_p/P, abbreviated as AI) (Figure 1). Physically, EI represents the partitioning of P into E in the long term whereas AI can be interpreted as a measure of the mean climate of the catchment (Carmona et al., 2016). The original Budyko formulation (Budyko, 1974) relates AI to EI through a nonparametric, nonlinear function that describes the long-term water-energy balance of catchments. But, owing to observed deviations from the original Budyko curve (Donohue et al., 2012; Istanbulluoglu et al., 2012; Porporato et al., 2004), several parametric variations of the Budyko function have been formulated (Choudhury, 1999; Fu, 1981; Yang et al., 2008; Zhang et al., 2004). In this study, Fu's equation (Fu, 1981), a single parameter Budyko function is made use of Figure 1. It is expressed as follows:

$$\frac{E}{P}=1+\frac{E_p}{P}-\left(1+\left(\frac{E_p}{P}\right)^{\omega}\right)^{\frac{1}{\omega}} \tag{1}$$

In equation (1), the parameter ω has no analytic solution but various parameterizations have been proposed that relate ω to catchment characteristics such as vegetation, soil, topography, and location of catchments (Li et al., 2013; Xu et al., 2013).

The Budyko curve (equation (1)) is constrained by the following water and energy limits:

$$\frac{E}{P}=1, \quad \frac{E_p}{P}>1 \quad (water\ \ limit) \tag{2}$$

$$\frac{E}{P}=\frac{Ep}{P}, \quad \frac{E_p}{P}<1 \quad (energy\ \ limit) \tag{3}$$

Equation (2) implies that the water available for E is limited by P. Thus, the hypothesis assumes that the contribution of catchment storage to water availability is negligible over the long term. If the supply of water in the catchment is unlimited, E is constrained by available energy in the form of E_p (equation (3)). In summary, the Budyko space consists of Budyko curve, based on catchment specific parameter ω, and the limits (equations (2) and (3)).

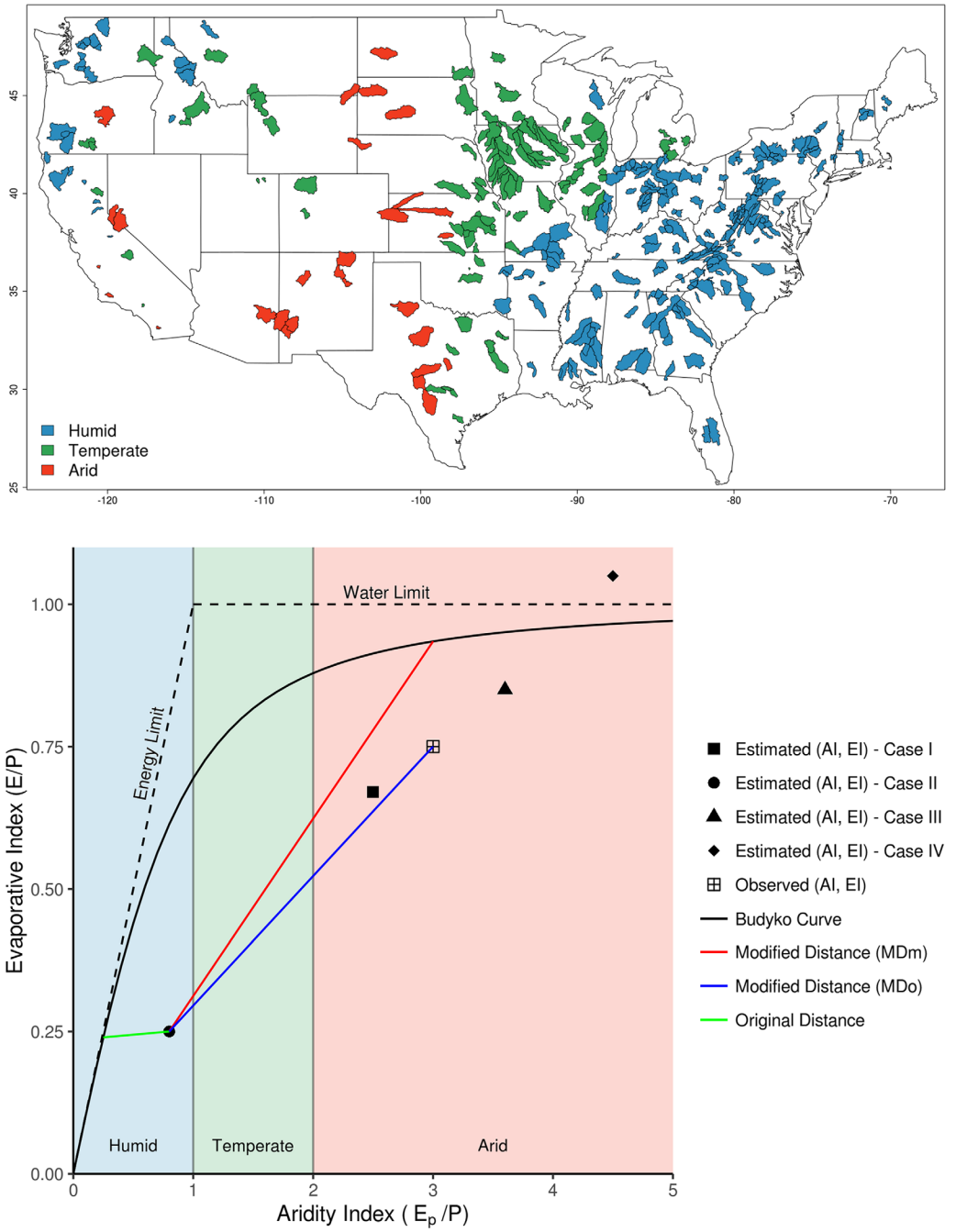

Figure 1. (a) MOPEX catchments in the United States classified according to aridity and (b) a representation of the Budyko space consisting of Fu's curve with $\omega = 2.6$ (Fu, 1981), water and energy limits. The original definition of distance following Greve et al. (2014) and the modified distances (MDm and MDo) are represented. Four cases of estimated (AI, EI) points representing different precipitation biases are shown: Case 1: low positive bias, Case 2: high positive bias, Case 3: low negative bias, and Case 4: high negative bias (refer to section 3.4).

3.2. Defining the Error Metric

In traditional data validation approaches, observational data sets are directly evaluated with the help of statistical error metrics such as Mean Bias and Root Mean Square Error (RMSE). Even when models are used to evaluate remote sensing data sets, the error in the data set is assessed by comparing model outputs, like runoff, with ground-based measurements using similar error metrics. But using the Budyko function (equation (1)) for validation requires the projection of the three-dimensional space described by (P, E, E_p) onto

the two-dimensional Budyko space represented by (AI, EI). Therefore, the objective of this study is to develop and test an error metric that can translate the individual errors in P, E, and E_p to an equivalent error in the Budyko space. For this, we use the Root Mean Square weighted Error (RMSwE) metric developed by Greve et al. (2014) with suitable modifications. This approach hypothesizes that the error in the (P, E, E_p) space is analogous to the Euclidean distance between the estimated (AI, EI) point (Figure 1b), determined from satellite-based estimates of (P, E and E_p), and the catchment specific Budyko curve (original distance in Figure 1b). Physically, original distance (OD), as defined by Greve et al. (2014), represents the combined error of precipitation and evapotranspiration data sets in representing the long-term water and energy balance of the catchments. In addition, a weight that penalizes the data sets that exceed the water and energy limits (equations (2) and (3)) is introduced.

In this study, we set the value of the weight as unity based on the following arguments. One of the main objectives of this study is to compare the proposed framework with traditional data validation methods which do not contain any additional weight or penalty terms for exceeding water or energy limits. Therefore, comparing such an error metric with a metric that is inflated by an extraneous penalty would not be logical. The developed error metric may imply higher error in the P, E, or E_p data than that would be implied by traditional error metrics which may lead to erroneous conclusions regarding the quality of the data set. Hence, $RMSE_{OD}$ in the Budyko space, following Greve et al. (2014) is expressed as

$$RMSE_{OD}=\sqrt{\frac{\sum_{i=1}^{n}(w_i.D_i)^2}{n}} \tag{4}$$

where the weight, $w_i = 1$, n is the number of points in the Budyko space and D_i or OD is the Euclidean distance (minimum) between the point i and the catchment specific Budyko curve, $RMSE_{OD}$ is the RMSE metric as defined by Greve et al. (2014) in which the subscript OD refers to the original definition of distance, D (refer to Figure 1).

3.3. Estimating the Error Metric

Estimation of the $RMSE_{OD}$ metric (equation (4)) for the MOPEX catchments involves the calculation of the Euclidean or the minimum distance, OD, between the estimated (AI, EI) point determined using remote sensing data (Figure 1) and the Budyko curve defined by Fu's equation (equation (1)). Thus, OD, according to Greve et al. (2014) is given by

$$\min_{AI} OD=\sqrt{(AI_{est}-AI)^2+(EI_{est}-EI_{mod})^2} \tag{5}$$

where AI_{est} and EI_{est} are the long-term average AI and EI estimated from remote sensing data, EI_{mod} is modeled EI from Fu's equation given by $EI_{mod}=1+AI-(1+(AI)^{\omega})^{1/\omega}$.

The following procedure is adopted to estimate the $RMSE_{OD}$ metric for the MOPEX catchments. First, the parameter ω in Fu's equation is estimated for each of the 393 MOPEX catchments under consideration following the method prescribed by Li et al. (2013). In this method, ω is determined by minimizing the squared error between observed annual EI and EI inferred from Fu's equation (equation (1)). Thus, the objective function is given by

$$obj=\min\sum_{i}\{EI^i_{act}-(1+AI^i_{act}-(1+(AI^i_{act})^{\omega})^{1/\omega})\}^2 \tag{6}$$

where EI^i_{act} and AI^i_{act} are the average EI and AI estimated for the year i using observed data from MOPEX catchments.

Next, the original distance, OD, is determined for all the 393 catchments using equation (5). Finally, $RMSE_{OD}$ is estimated using equation (4).

3.4. Sensitivity Analysis of the Error Metric

We test the hypothesis that $RMSE_{OD}$ in the Budyko space is sensitive to individual biases in the P and E estimates. For this, we setup a controlled sensitivity analysis experiment using the 393 MOPEX catchments. We first divide the Budyko space into three hydroclimatic regimes based on AI; humid ($0 < AI < 1$), temperate ($1 < AI < 2$) and arid ($AI > 2$) (Sankarasubramanian & Vogel, 2003). Out of 393 catchments, 253 are humid,

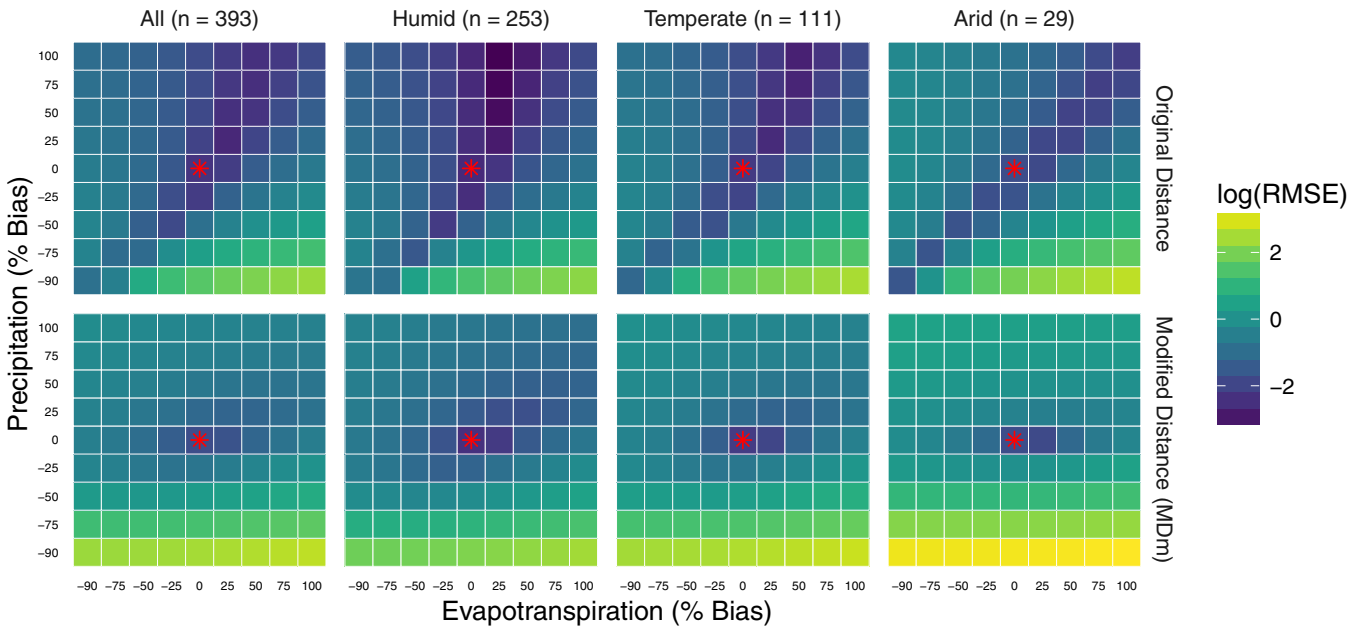

Figure 2. Sensitivity maps of the logarithm of (1) $RMSE_{OD}$ (original distance) and (2) $RMSE_{MDm}$ (modified distance). The red star represents the RMSE value for no bias in either precipitation or evapotranspiration. *n* is the number of catchments in each hydroclimatic regime.

111 are temperate whereas 29 are arid (Figure 1a). The observed P and E data sets provided by MOPEX are artificially biased by fixed percentages (0, 25, 50, etc.) and for each combination of the biased P and E data, $RMSE_{OD}$ (equation (4)) is determined for the defined aridity classes. Then, the logarithm of the $RMSE_{OD}$ value for each combination of artificially biased P and E data is plotted as a single pixel on the sensitivity map (Figure 2). As the focus of the study is on validation of P and E data sets, the sensitivity of the error metric to changes in E_p was not considered.

The sensitivity analysis results are interpreted using four cases: (1) Case I: low positive precipitation bias, (2) Case II: high positive precipitation bias, (3) Case III: low negative precipitation bias, and (4) Case IV: high negative precipitation bias. These cases are represented in Figure 1. If the $RMSE_{OD}$ metric is sensitive to individual biases in P and E values, then, ideally, the logarithm of $RMSE_{OD}$ values must be minimum at the center of the sensitivity map (darker colors in Figure 2) and then gradually increase toward the edges (lighter colors in Figure 2). It is evident from the sensitivity map of $RMSE_{OD}$ (Figure 2) that when the precipitation is positively biased to a large degree (>50%) the RMSE value is very low, irrespective of the hydroclimatic regime (Case II in Figure 1). This is contrary to expectation as RMSE value should increase with increasing bias in either P or E. The reason for such a behavior is that the Euclidean distance, *D*, in equation (4) is the minimum distance between the Budyko curve and the estimated (AI, EI) point (original distance in Figure 1b). As a result, if the precipitation has a large positive bias, the estimated (AI, EI) point moves toward the origin, or the humid region, of the Budyko space. In this region, distance to the Budyko curve is observed to be considerably lower than in either temperate or arid regions, thereby resulting in a very low $RMSE_{OD}$ value. This is seen in cases I, III, and IV as well in which the $RMSE_{OD}$ metric behaves as expected when the estimated (AI, EI) point is not very close to the Budyko curve.

Figure 2 also reveals that the $RMSE_{OD}$ metric is more sensitive to biases in P rather than E. This is expected as both AI and EI feature P in their definitions, whereas changes in E only affect the evaporative index (EI).

3.5. Modification of the Error Metric

We address the issue of high precipitation values leading to low RMSE by modifying the definition of the distance (*D*) in equation (4). Instead of determining the minimum distance from the estimated (AI, EI) point to the Budyko curve, *D* is calculated as the distance from the estimated (AI, EI) point to a point on the Budyko curve corresponding to the observed long-term aridity index of the specific catchment (modeled modified distance in Figure 1b). This modification preserves the aridity of the catchment which is violated

by the original definition of distance but it introduces an additional data requirement in the form of the actual aridity index. The RMSE calculated using the modified definition of distance is henceforth referred to as $RMSE_{MDm}$, where MDm refers to modeled modified distance calculated as

$$MDm = \sqrt{(AI_{est} - AI_{act})^2 + (EI_{est} - EI_{mod})^2} \quad (7)$$

where EI_{est} and AI_{est} are the long-term average evaporative and aridity indices estimated from remote sensing data, AI_{act} is the actual AI determined using observed E_p and P data available for the MOPEX catchments (described in section 2). EI_{mod} is modeled EI from Fu's equation (equation (1)) as $EI_{mod} = 1 + AI_{act} - (1 + (AI_{act})^{\omega})^{1/\omega}$.

Sensitivity analysis is performed on the $RMSE_{MDm}$ metric to test whether the new definition can overcome the shortcomings of $RMSE_{OD}$. In the sensitivity maps of $RMSE_{MDm}$ (Figure 2), very low values are found only around the pixel representing $RMSE_{MDm}$ for zero bias in P and E. The positive and negative bias regions in the heat maps exhibit increasing gradients from the center (point of no bias) toward the edges (points of maximum bias). This behavior of $RMSE_{MDm}$ is essential if any error metric in the Budyko space is to reflect the errors in the individual estimates of P and E. Comparing the sensitivity maps of $RMSE_{MDm}$ with that of the $RMSE_{OD}$, it is quite apparent that the magnitude of error is higher in all hydroclimatic regimes. As expected $RMSE_{MDm}$ is still more sensitive to biases in P rather than E. Across hydroclimatic regimes, $RMSE_{MDm}$ values in the arid region is higher than in temperate or humid regions for the same percentage of bias. The biggest improvement is seen in the humid regions, where insensitive regions in the $RMSE_{OD}$ heat map now exhibit higher sensitivity.

3.6. Sensitivity of Modified Distance to ω and AI

Determining the modeled modified distance (MDm in equation (7)), and thus the application of the developed $RMSE_{MDm}$ error metric, requires estimates of both the Budyko parameter, ω, and the long-term aridity index, AI, of the catchment. For data-scarce catchments where no ground-based observations of P and E are available, estimates of ω and AI (AI_{act} in equation (7)) have to be obtained from other sources. As stated earlier, several parameterizations are available for ω (Li et al., 2013; Xu et al., 2013) and for AI_{act}, long-term aridity index maps are available (Trabucco & Zomer, 2009). As these parameterizations and estimates are subject to uncertainty, it is important to understand the importance of accurately estimating ω and AI_{act} to the determination of the developed $RMSE_{MDm}$ metric. In this section, we answer the question: How sensitive is the modeled modified distance, MDm, in equation (7) to changes in (1) ω and (2) AI_{act}?

To answer the question, a theoretical experiment is setup in the Budyko space. First, three estimated (AI, EI) points (AI_{est} and EI_{est} in equation (7)) representing the remotely sensed data are selected, one each in the humid ($AI_{est} = 0.5$, $EI_{est} = 0.5$), temperate ($AI_{est} = 1.5$, $EI_{est} = 0.5$), and arid regions ($AI_{est} = 4$, $EI_{est} = 0.5$) of the Budyko space. To quantify the effects of changes in ω, the value of the actual AI_{act} is fixed at 0.5, 1.5, and 4 for humid, temperate, and arid regions of the Budyko space, respectively. To quantify the sensitivity of the modified distance (MDm) to changes in AI_{act}, the value of ω is fixed at 2.6. Then, the sensitivity analysis is carried out according to Saltelli et al. (2008). Specifically, we adopt the stratified single parameter sampling strategy suggested by the authors. Accordingly, ω is sampled at equal intervals in the range [1,6]. The range of ω is selected based on the estimated ω values for the MOPEX catchments. AI_{act} is sampled in a similar manner in the ranges [0,1] for humid, [1,2] for temperate, and [2,6] for arid regions in the Budyko space. About 2,000 samples of ω and AI_{act} are obtained to estimate the mean and variance of the resulting distribution of modified distance, with variance being a measure of sensitivity.

Table 1 summarizes the results of sensitivity analysis. It is seen that irrespective of the position of the estimated (AI_{est}, EI_{est}) point or the aridity region in which sensitivity analysis is carried out, the variance of modified distance with varying AI_{act} is orders of magnitude higher than the variance in modified distance with varying ω. Therefore, modified distance is more sensitive to changes in AI rather than ω. In absolute terms, changes in the Budyko parameter do not seem to significantly affect the estimates of modified distance. Even for AI_{act}, the magnitude of the variance of the modified distance is higher in arid region compared to the other regions. This is due to the fact that the range of aridity index in the arid region is higher than the other regions and the nonlinear nature of the Budyko curve. Therefore, when applying the framework to data-scarce regions it is important to have more reliable estimates of AI, especially if the catchment is in the

Table 1
Variance and Mean (in Brackets) of the Modified Distance (MDm) for (1) Fixed AI_{act} and Varying ω and (2) Fixed ω and Varying AI_{act}

Aridity	Humid	Temperate	Arid
(1) Estimated (AI_{est} = 0.5, EI_{est} = 0.5) point in humid region			
Fixed AI_{act} and varying ω	0.012 (0.09)	1.0E-03 (1.06)	1.2E-04 (3.52)
Fixed ω and varying AI_{act}	0.03 (0.32)	0.085 (1.05)	1.32 (3.53)
(2) Estimated (AI_{est} = 1.5, EI_{est} = 0.5) point in temperate region			
Fixed AI_{act} and varying ω	4.3E-04 (1.01)	0.02 (0.33)	2.4E-04 (2.54)
Fixed ω and varying AI_{act}	0.09 (1.02)	0.08 (0.42)	1.28 (2.55)
(3) Estimated (AI_{est} = 4, EI_{est} = 0.5) point in arid region			
Fixed AI_{act} and varying ω	3.8E-05 (3.50)	1.9E-04 (2.52)	0.015 (0.41)
Fixed ω and varying AI_{act}	0.09 (3.51)	0.08 (2.52)	0.24 (1.14)

Note. Depending on the values of estimated (AI_{est}, EI_{est}) and actual AI_{act}, 18 cases are presented. The fixed AI_{act} values for humid, temperate, and arid regions are 0.5, 1.5, and 4, respectively. The fixed ω value is 2.6.

arid region. Moreover, as the variance magnitudes are not very high, the uncertainty in $RMSE_{MDm}$ estimates is relatively low (except for arid catchments). It is to be noted that sensitivity analysis was carried out for a number of different estimated (AI_{est}, EI_{est}) points in the Budyko space but the results are similar to the results presented here.

3.7. Statistical Significance Test

The difference in $RMSE_{MDm}$ values between every combination of P and E data sets is tested for statistical significance using a two-sample Student's t-test (two-sided) (Snedecor & Cochran, 1989). To apply the t-test a sample of 100 $RMSE_{MD}$ values are generated for each combination of P and E data sets. The large sample size, $N = 100$, is to overcome the requirement of normality for applying t-tests. Each of the 100 $RMSE_{MDm}$ values are calculated by randomly selecting 150 catchments from the 393 MOPEX catchments. The null hypothesis in this case is $H0 : \mu_i = \mu_j$, where μ_i and μ_j are the means of $RMSE_{MDm}$ values of two different combinations of P and E data sets. For example, μ_1 could be mean of $RMSE_{MDm}$ values of (TRMM, AVHRR) and μ_2 could be mean for (TRMM, MODIS). The threshold selected for statistical significance is $\alpha = 0.05$.

4. Validating the Framework

For validating the framework, several satellite-based P and E data sets are used. For P estimates, three data sets are chosen; Climate Prediction Center Morphing Technique (CMORPH) (Joyce et al., 2004) (downloaded from http://rda.ucar.edu/datasets/ds502.0), Tropical Rainfall Measuring Mission (TRMM) 3B42RT (Huffman et al., 2007) (downloaded from http://mirador.gsfc.nasa.gov/), and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks—Climate Data Record (PERSIANN) (Ashouri et al., 2015) (downloaded from http://chrs.web.uci.edu/persiann/data.html). Three evapotranspiration products are used; Advanced Very High Resolution Radiometer (AVHRR) (Zhang et al., 2010) (from http://www.ntsg.umt.edu/project/et), MOD16 from the Moderate Resolution Imaging Spectrometer (MODIS) (Mu et al., 2007) (from http://www.ntsg.umt.edu/project/mod16), and the Global Land Evaporation Amsterdam Model (GLEAM) (Martens et al., 2016; Miralles et al., 2011) (from http://www.gleam.eu/). As the study focuses on validation of P and E data sets, E_p is sourced only from GLEAM. As most of the remote sensing data span the years 1998–2015, the MOPEX data set was also extended from 2003 to 2015 using precipitation from Parameter-elevation Relationships on Independent Slopes (PRISM) (PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu) and river discharge data from the United States Geological Survey.

4.1. Validation in the Budyko Space

To validate the framework, RMSE metric in the Budyko space modeled using Fu's equation ($RMSE_{MDm}$) is compared against RMSE metric in the Budyko space calculated using observed data ($RMSE_{MDo}$), to see whether both the metrics lead to similar conclusions regarding the quality of data sets under consideration. $RMSE_{MDo}$ is calculated using equation (4), but the distance, D, is the observed modified distance (MDo) between the estimated (AI_{est}, EI_{est}) point and the actual (AI_{act}, EI_{act}) point (Figure 1) given by

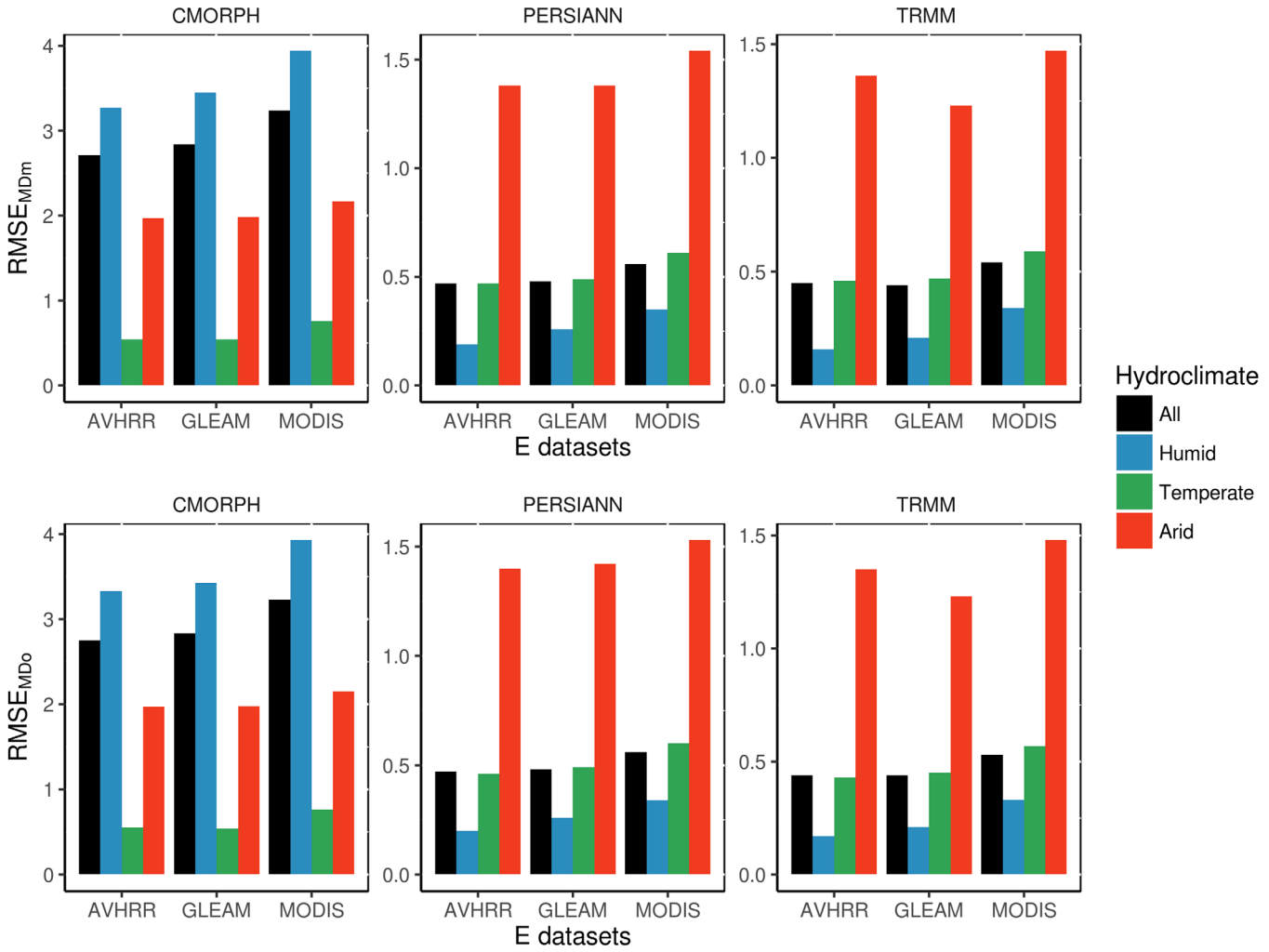

Figure 3. Bar plots of $RMSE_{MDm}$ and $RMSE_{MDo}$ values for different combinations of P and E data sets and for all hydroclimatic regimes. The three plots correspond to the three satellite-based precipitation products.

$$MDo = \sqrt{(AI_{est} - AI_{act})^2 + (EI_{est} - EI_{act})^2} \tag{8}$$

To determine the estimated (AI_{est}, EI_{est}) point, the long-term annual average estimates of P and E are determined from the aforementioned satellite-based products. To determine $RMSE_{MDm}$, the actual AI (AI_{act} in equation (7)) is calculated using E_p from the MOPEX data set and P from the MOPEX data set. To determine $RMSE_{MDo}$, the actual (AI_{act}, EI_{act}) point (equation (8)) is calculated using the extended MOPEX data set.

It is evident that the developed error metric using Fu's equation ($RMSE_{MDm}$) compares well with the observed $RMSE_{MDo}$ metric in characterizing the combined error in P and E data sets (Figure 3) in all hydroclimates. The agreement between $RMSE_{MDm}$ and $RMSE_{MDo}$ is quite apparent in the scatter plot of observed and modeled distances (Figure 4). Except for a few catchments in the arid and humid regions, the distances are well correlated. Analyzing $RMSE_{MDo}$ values determined using ground-based measurements of P and E (Figure 3) for all the hydroclimatic regimes combined, it is seen that the (P, E) combination that exhibits the least error is (TRMM, AVHRR). A similar conclusion is reached by the proposed Budyko hypothesis-based validation framework (Figure 3). As seen earlier, $RMSE_{MDm}$ exhibits different sensitivity in different hydroclimates. To test whether the conclusion regarding the best combination of (P, E) data set holds true for different hydroclimates, we determine the $RMSE_{MDm}$ and $RMSE_{MDo}$ statistics separately for humid, temperate, and arid catchments. The results show that both metrics point to (TRMM, AVHRR) as the best combination across all hydroclimates. As far as the poorest performing (P, E) data sets are concerned, $RMSE_{MDm}$ and

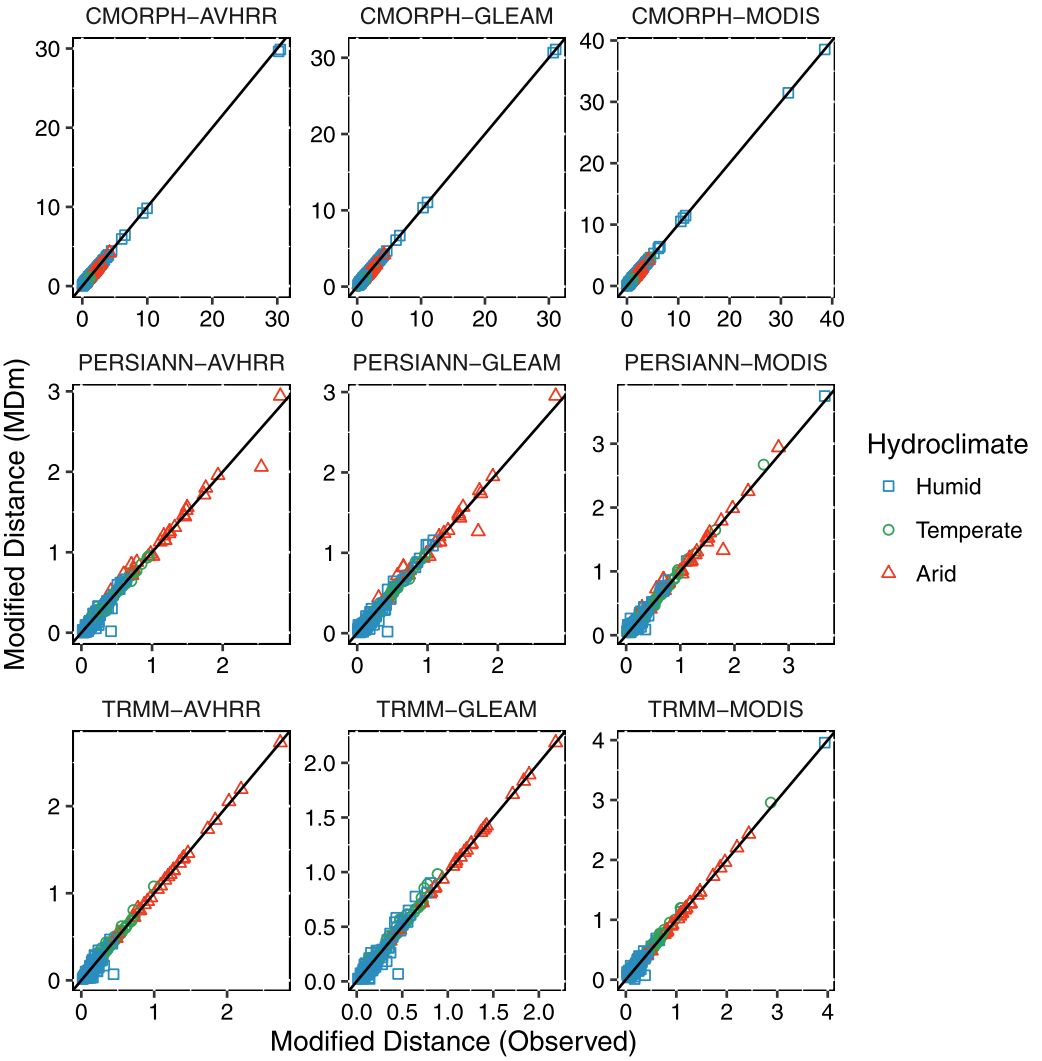

Figure 4. Scatter plots of observed modified distance (*X* axis) versus modeled modified distance (*Y* axis) for the MOPEX catchments and for all combinations of remotely sensed P and E data sets.

$RMSE_{MDo}$ metrics agree that (CMORPH, MODIS) combination fails to represent the combined water and energy balance of the study catchments. CMORPH seems to have high errors, especially in humid catchments (Figure 3). On closer examination of the distribution of (AI, EI) points in the Budyko space for CMORPH, it is seen that in a number of catchments, the water and energy limits are exceeded by a large margin, mainly due to severe underestimation of precipitation. MODIS combined with P data sets results in higher $RMSE_{MDm}$ values compared to either GLEAM or AVHRR.

4.2. Validation With Traditional RMSE

In general, the proposed error metric is seen to be capable of identifying the best performing combination of P and E data sets in the Budyko space. But it is also important to understand how the developed $RMSE_{MDm}$ metric compares with traditional $RMSE_T$ metric for several reasons. First, it is essential to analyze whether combining P and E data sets, as is done in the Budyko space, substantially differs from evaluating P and E separately, as is done traditionally. Second, $RMSE_{MDm}$ is seen to be less sensitive to changes in E compared to changes in P. To understand the effect of combining P and E data sets and the reduced sensitivity of the error metric to E, we compare $RMSE_{MDm}$ to traditional $RMSE_T$ values (Figure 5). The long-term average from the extended MOPEX data set and the satellite-based estimates of P and E are used to determine $RMSE_T$ of individual data sets across all the hydroclimatic regimes.

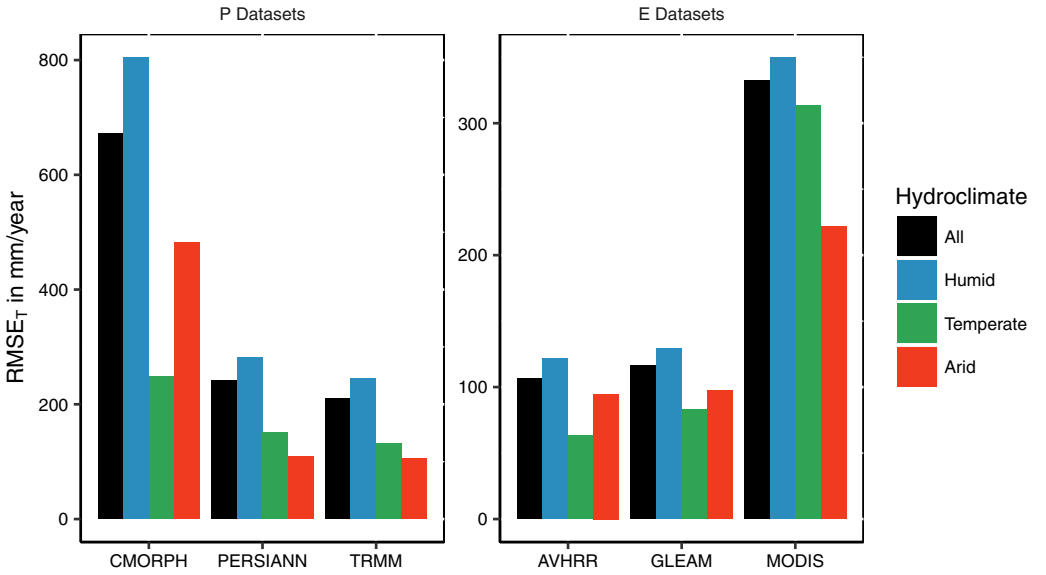

Figure 5. Bar plots of RMSE$_T$ (traditional RMSE) values for different P and E data sets and for all hydroclimatic regimes.

As far as the best (TRMM and AVHRR) and the worst (CMORPH and MODIS) performing P and E data sets are concerned, the RMSE$_T$ metric agrees with the developed framework. Therefore, combining P and E data sets in the Budyko space does not seem to have substantial effect on ranking the P and E data sets. But, it is seen that RMSE$_{MDm}$ does indeed suppress the biases in E data sets. For example, MODIS is seen to have a very high RMSE$_T$ error in the arid region, but combining MODIS and CMORPH in the Budyko space, the error is comparable to other combinations in magnitude. Next, disaggregating the catchments according to AI reveals certain differences between RMSE$_{MD}$ and RMSE$_T$ metrics (Figures 3 and 5). For example, if RMSE$_{MDm}$ is followed, the (TRMM, AVHRR) combination performs best in humid catchments. But RMSE$_T$ for TRMM is minimum in arid catchments and AVHRR is seen to be better in temperate regions than in either humid or arid catchments. This behavior is primarily due to the fact that the distance, D, in humid regions are relatively small for the same magnitude of error in (P, E, E$_p$) compared to distance in arid and temperate regions (Figure 2). Therefore, care must be taken in comparing RMSE$_{MDm}$ values calculated in different regions of the Budyko space.

4.3. Statistical Significance of the Results

Owing to a large combination of P and E data sets, only a summary of the results of the statistical significance tests is presented here. Except a few exceptions, the difference in RMSE$_{MDm}$ values between all combinations of P and E data sets is seem to be significant with a p-value less than 0.05. The exception involves the AVHRR and GLEAM evapotranspiration data sets where the difference in RMSE$_{MDm}$ values is statistically insignificant when AVHRR and GLEAM are combined with the same P data set. (For example, p-value is greater than 0.05 between TRMM-AVHRR and TRMM-GLEAM.) This is also seen in Figure 5, where the traditional RMSE$_T$ values are close to each other. The results of statistical significance tests show that despite being relatively more insensitive to biases in E, the RMSE$_{MDm}$ metric can represent these biases given that their magnitude is high. For example, the difference between TRMM-GLEAM and TRMM-MODIS combinations is significant.

5. Applying the Framework to a Data-Scarce Catchment

We demonstrate the application of the proposed framework to a catchment in which no ground-based measurement of P and E are available. The steps involved in determination of the Budyko parameter (ω), the modified distance, D, and the RMSE$_{MDm}$ error metric using publicly available data are detailed.

The study area selected for applying the framework is the Omo Gibe river basin in East Africa. As all the satellite-based P and E products are gridded, we divide the catchment into regular grids of

resolution 0.25° × 0.25°. Thus the Omo Gibe basin, with an area of 80,000 km^2, is divided into 592 grids and all the six satellite-based P and E products are interpolated onto the grids using nearest-neighbor interpolation.

Next step is to determine the Budyko parameter, ω, for each of the 592 grids. For this, we use the parameterization developed by Xu et al. (2013). Specifically, we make use of the following multiple linear regression (MLR) model developed for small catchments

$$\omega = 5.05722 - 0.09322lat + 0.13085CTI + 1.31697NDVI + 0.00003A - 0.00018elev \quad (9)$$

where *lat* is the latitude of each of the grid points in degrees, *A* is the grid resolution in km^2, *CTI* is the compound topographic index (Beven & Kirkby, 1979), *NDVI* is the Normalized Difference Vegetation Index, and *elev* is the elevation of the grids in meters. In equation (9), the topographic variables (*lat*, *elev*, and *CTI*) were determined using the Hydro-1K data set (Lehner et al., 2006) (downloaded from http://hydrosheds.cr.usgs.gov/dataavail.php), *NDVI* is derived from the Global Inventory Modeling and Mapping Studies (GIMMS) AVHRR NDVI3g data set (https://nex.nasa.gov/nex/projects/1349/).

Determination of $RMSE_{MD}$ requires the knowledge of long-term AI of the catchments. It is to be noted that only a climatological estimate of AI of the study catchment is required and not concurrent observations corresponding to the study period. AI of the 592 grids are derived from the global aridity index data set (spatial resolution of 1 km) developed by Trabucco and Zomer (2009) (from http://www.cgiar-csi.org/data/global-aridity-and-pet-database). Next the distance, *D*, in equation (4) is determined. For this, the satellite-based estimates of (P, E, E_p) are mapped onto the Budyko space. Then, the distance, MDm is calculated using equation (7) and $RMSE_{MDm}$ metric is determined using equation (4).

The $RMSE_{MDm}$ metric is determined for all combinations of P and E data sets (equation (4)): 0.135 (CMORPH, AVHRR), 0.149 (CMORPH, GLEAM), 0.876 (CMORPH, MODIS), 0.132 (PERSIANN, AVHRR), 0.147 (PERSIANN, GLEAM), 0.53 (PERSIANN, MODIS), 0.133 (TRMM, AVHRR), 0.139 (TRMM, GLEAM), and 0.462 (TRMM, MODIS). It is apparent that the MODIS E data set performs the poorest among the E data sets. Unlike in the MOPEX catchments, all the $RMSE_{MDm}$ values of all the three P data sets under consideration are close to each other, especially when combined with AVHRR and GLEAM E estimates. In summary, the best precipitation products according to the developed framework are TRMM and PERSIANN (PERSIANN-CDR). As far as the best E data set is concerned, AVHRR is seen to have the least bias. The combination that shows the largest deviation from the Budyko curve is (CMORPH, MODIS). The results from the developed framework are compared with past studies which validate satellite products for Ethiopian river basins. Romilly and Gebremichael (2011) and Hirpa et al. (2010) concluded that TRMM 3B42RT outperforms CMORPH in Ethiopian basins. As both the aforementioned studies evaluate the PERSIANN-CCS product and not the PERSIANN-CDR multisatellite product used in the present study, the result from the developed framework for PERSIANN could not be validated with past studies.

It is to be noted here that the methodology presented here is just one of the ways in which this framework can be applied. This application assumes that no ground-based measurements of P and E for the study period are available. But in data-scarce catchments, where some historical data are available, the Budyko parameter ω and aridity index, AI, can be estimated using the methods presented in the paper for MOPEX catchments.

6. Conclusions, Limitations, and Future Work

Hydrologic studies in data-scarce regions use remotely sensed precipitation as meteorological forcing and satellite-based estimates of evapotranspiration for data assimilation. But remotely sensed P and E data sets exhibit large uncertainty requiring comprehensive validation for the area of study. This study addresses this important issue by developing a validation framework that tests for the physical consistency of remotely sensed P and E data sets without the use of concurrent ground-based measurements. A RMSE-based error metric is developed and comprehensively tested to see whether the metric can translate individual biases in P and E data sets onto the Budyko space. Results show that the proposed validation framework is capable of arriving at the same conclusions as traditional validation methodologies regarding the quality of P and E data sets. The application of the developed framework to a data-scarce catchment using publicly available topographic, vegetation, and aridity information is also presented. In contrast to previous validation studies

that employ complex distributed hydrological models, the use of the single parameter Budyko function highlights the effectiveness of using simple water and energy balance principles in validation of observational data.

Owing to the limitations of the original Budyko formulation, the developed framework can only test whether the combination of P and E data sets can describe the long-term combined water and energy balance of catchments. This implies that the developed $RMSE_{MDm}$ metric characterizes the bias in P and E data sets and not the variance. Recent studies have focused on extending the Budyko hypothesis to subannual time scales (Greve et al., 2016; Zhang et al., 2008). Therefore, future work involves the extension of the framework to validation of P and E data sets at monthly and daily time scales, which is crucial for characterizing the variance and also for hydrologic applications such as streamflow forecasting and reservoir operations. In addition, it is assumed that in the long term, storage in the catchment is negligible. Therefore, care must be taken when applying the framework in catchments which have long-term storage such as snow, ice, or reservoirs and also in small catchments where storage influences water availability. It is seen that the $RMSE_{MDm}$ metric is more sensitive to biases in P rather than E. Therefore, care must be taken in interpreting the error metric when the focus of a study is solely on evaluating E data sets which are relatively close to each other. But E data sets can be still be effectively evaluated using the framework if accurate estimates of P are available and the focus of the study is to validate only E data sets.

It is to be noted here that the developed framework does not require concurrent observations of precipitation and evapotranspiration for validating remote sensing data. But the application of this framework to a data-scarce region requires reliable estimates of AI, which could be sourced from nonconcurrent ground-based measurements, as was done in this study. Although a large sample of catchments, representing a wide range of aridities, have been used in the study, we encourage researchers to validate the robustness of the developed framework in other geographies having different topographic, hydrologic, and climatic characteristics.

Acknowledgments
We acknowledge funding support from the NASA Applied Sciences Water Resource Application grant NNX15AC33G. We thank the three reviewers for comments that helped improve the paper. We thank Peter Greve for his valuable inputs which improved the paper. The sources of the data sets used in the study are mentioned in the text along with appropriate references.

References

Adler, R. F., Kidd, C., Petty, G., Morissey, M., & Goodman, H. M. (2001). Intercomparison of global precipitation products: The third precipitation intercomparison project (pip-3). *Bulletin of the American Meteorological Society*, *82*(7), 1377–1396. https://doi.org/10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., . . . Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Beven, K., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, *24*(1), 43–69.

Bitew, M., & Gebremichael, M. (2011). Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. *Hydrology and Earth System Sciences*, *15*(4), 1147–1155.

Budyko, M. (1974). *Climate and life* (508 pp.). New York, NY: Academic Press.

Carmona, A., Poveda, G., Sivapalan, M., Vallejo-Bernal, S., & Bustamante, E. (2016). A scaling approach to Budyko's framework and the complementary relationship of evapotranspiration in humid environments: Case study of the Amazon River basin. *Hydrology and Earth System Sciences*, *20*(2), 589–603.

Chen, S., Liu, H., You, Y., Mullens, E., Hu, J., Yuan, Y., . . . Hong, Y. (2014). Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations. *PLoS One*, *9*(4), e89681. https://doi.org/10.1371/journal.pone.0089681

Choudhury, B. (1999). Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *Journal of Hydrology*, *216*(1–2), 99–110. https://doi.org/https://doi.org/10.1016/S0022-1694(98)00293-5

Donohue, R. J., Roderick, M. L., & McVicar, T. R. (2012). Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko hydrological model. *Journal of Hydrology*, *436–437*, 35–50. https://doi.org/10.1016/j.jhydrol.2012.02.033

Fu, B. (1981). On the calculation of the evaporation from land surface. *Scientia Atmospherica Sinica*, *5*(1), 23–31.

Gebregiorgis, A., & Hossain, F. (2014). Making satellite precipitation data work for the developing world. *IEEE Geoscience and Remote Sensing Magazine*, *2*(2), 24–36. https://doi.org/10.1109/MGRS.2014.2317561

Gentine, P., D'odorico, P., Lintner, B. R., Sivandran, G., & Salvucci, G. (2012). Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. *Geophysical Research Letters*, *39*, L19404. https://doi.org/10.1029/2012GL053492

Greve, P., Gudmundsson, L., Orlowsky, B., & Seneviratne, S. I. (2015). Introducing a probabilistic Budyko framework. *Geophysical Research Letters*, *42*, 2261–2269. https://doi.org/10.1002/2015GL063449

Greve, P., Gudmundsson, L., Orlowsky, B., & Seneviratne, S. I. (2016). A two-parameter Budyko function to represent conditions under which evapotranspiration exceeds precipitation. *Hydrology and Earth System Sciences*, *20*(6), 2195–2205.

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, *7*(10), 716–721.

Hirpa, F. A., Gebremichael, M., & Hopson, T. (2010). Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology*, *49*(5), 1044–1051. https://doi.org/10.1175/2009JAMC2298.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., . . . Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Istanbulluoglu, E., Wang, T., Wright, O. M., & Lenters, J. D. (2012). Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. *Water Resources Research*, *48*, W00H16. https://doi.org/10.1029/2010WR010100

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Kidd, C., & Huffman, G. (2011). Global precipitation measurement. *Meteorological Applications*, *18*(3), 334–353. https://doi.org/10.1002/met.284

Lakshmi, V. (2004). The role of satellite remote sensing in the prediction of ungauged basins. *Hydrological Processes*, *18*(5), 1029–1034. https://doi.org/10.1002/hyp.5520

Lehner, B., Verdin, K., & Jarvis, A. (2006). *Hydrosheds technical documentation*. Washington DC: World Wildlife Fund US. Retrieved from http://hydrosheds.cr.usgs.gov

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, *51*, 7309–7342. https://doi.org/10.1002/2015WR017616

Li, D., Pan, M., Cong, Z., Zhang, L., & Wood, E. (2013). Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, *49*, 969–976. https://doi.org/10.1002/wrcr.20107

Lorenz, C., & Kunstmann, H. (2012). The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *Journal of Hydrometeorology*, *13*(5), 1397–1420. https://doi.org/10.1175/JHM-D-11-088.1

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., . . . Verhoest, N. E. C. (2016). Gleam v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development Discussions*, *2016*, 1–36. https://doi.org/10.5194/gmd-2016-162

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., & Dolman, A. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453.

Mu, Q., Heinsch, F. A., Zhao, M., & Running, S. W. (2007). Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, *111*(4), 519–536. https://doi.org/10.1016/j.rse.2007.04.015

Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. *The American Naturalist*, *164*(5), 625–632.

Romilly, T. G., & Gebremichael, M. (2011). Evaluation of satellite rainfall estimates over Ethiopian river basins. *Hydrology and Earth System Sciences*, *15*(5), 1505–1514.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). *Global sensitivity analysis: The primer*. West Sussex, UK: John Wiley.

Sankarasubramanian, A., & Vogel, R. M. (2003). Hydroclimatology of the continental United States. *Geophysical Research Letters*, *30*(7), 1363, https://doi.org/10.1029/2002GL015937

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, *36*, L07403, https://doi.org/10.1029/2009GL037338

Snedecor, G., & Cochran, W. (1989). *Statistical methods*. Ames, IA: Iowa University Press.

Stisen, S., & Sandholt, I. (2010). Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrological Processes*, *24*(7), 879–891. https://doi.org/10.1002/hyp.7529

Stokstad, E. (1999). Scarcity of rain, stream gages threatens forecasts. *Science*, *285*(5431), 1199–1200. https://doi.org/10.1126/science.285.5431.1199

Trabucco, A., & Zomer, R. (2009). Global aridity index (global-aridity) and global potential evapotranspiration (global-pet) geospatial database. In *CGIAR Consortium for Spatial Information*. Retrieved from http://csi.cgiar.org

Xu, X., Liu, W., Scanlon, B. R., Zhang, L., & Pan, M. (2013). Local and global factors controlling water-energy balances within the Budyko framework. *Geophysical Research Letters*, *40*, 6123–6129. https://doi.org/10.1002/2013GL058324

Yang, H., Yang, D., Lei, Z., & Sun, F. (2008). New analytical derivation of the mean annual water-energy balance equation. *Water Resources Research*, *44*, W03410. https://doi.org/10.1029/2007WR006135

Zhang, K., Kimball, J. S., Nemani, R. R., & Running, S. W. (2010). A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research*, *46*, W09522. https://doi.org/10.1029/2009WR008800

Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., & Briggs, P. R. (2004). A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, *40*, W02502. https://doi.org/10.1029/2003WR002710

Zhang, L., Potter, N., Hickel, K., Zhang, Y., & Shao, Q. (2008). Water balance modeling over variable time scales based on the Budyko framework-model development and testing. *Journal of Hydrology*, *360*(1), 117–131.

Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., . . . Su, Y. (2015). Global pattern for the effect of climate and land cover on water yield. *Nature Communications*, *6*, 5918.