Journal: HESS
Title: Revisiting the Hydrological Basis of the Budyko Framework with the Hydrologically Similar Groups Principle
MS No.: hess-2022-290
MS Type: Research article

Response to Reviewer#2

General Comments:
This manuscript proposes a framework to estimate the parameter of a parametric Budyko-type equation. The originality compared to other studies on the same issue is the preliminary classification of the catchments. In general, I enjoyed reading the paper and some results are very interesting, e.g., the ambiguous role of soil moisture on the evaporative ratio. There are two major comments that I think the authors should respond.
Response:

Thank you for your positive comments. Your suggestions are very useful for us to improve our research. We revised our manuscript according to your comments. The changes in our manuscript are underlined with red. We believe our manuscript improved a lot after the modification. Please see the response below.

Major Comments of Reviewer 2#:
Comment 1:
Some methodological choices are not presented / enough discussed. See the exhaustive list in the minor comments below. Some key information is missing, e.g., the time step used for establishing the equations between m and vegetation fractions, and the settings of the classifier are not presented, as the output of the classification performance.
Response:
(1) Reply on the time step

The time step used in our study is annual. For the collected datasets, the times of observation are discrete and discontinuous (listed in Supplement 1). The time range of verification datasets, including the data used for reconstructing runoff and the observed runoff data from GRDC, is from 2000 to 2016.

(2) Reply on the classifier

In the revised manuscript, we further describe the setup and performance of classifier, as follows,

"Three watershed characteristic variables — surface soil moisture (SM), rainfall seasonality index (SI), and fractional vegetation cover (FVC) — were selected for classification. For SM and FVC, the bounded intervals of the variables were given by the Decision Tree Regressor (DTR). The locations of splits in DTR were used as dividing intervals. The Scikit-learn library (Pedregosa et al., 2011) in Python provides the DTR used in this study. The criterion for measuring the quality of the split was set to "poisson" which uses reduction in Poisson deviance to find splits. The "random" strategy was used to choose local optimal splitting at each node. The results and performances of DTR are shown in Supplement 2. Based on the criteria using by Walsh and Lawler (1981), we divided the SI into three parts (SI≤0.4, 0.4<SI≤0.8, SI>0.8) to represent three hydroclimatic seasonality (precipitation spread throughout the year, marked seasonality with a short drier season, extreme seasonality with a long drier season). Finally, six hydrologically similar groups were classified (Table 3).

Table 3. Classification of watersheds

Soil moisture classifier	Water soil regime	Seasonality index classifier	Seasonality precipitation regime	Fractional vegetation cover classifier	vegetation cover regime	Name of the group
SM≤20	Dry soil	——	——	——	——	IN_D
SM>20	Wet soil	SI ≤ 0.4	Seasonless	——	——	IN_{WP}
		0.4 < SI ≤ 0.8	Marked seasonality	FVC ≤ 0.2 0.2 < FVC ≤ 0.5 FVC > 0.5	Low density Middle density High density	IN_{WMS} IN_{WMM} IN_{WML}
		SI > 0.8	Extreme seasonality	——		IN_{WE}

”

Comment 2:

The added value of the classification step is not demonstrated. I suggest the authors compare the performance of the model with relationships for each group with the performance of the model when a single relationship is used for the whole catchment set. At this stage, the classification provides insights in terms of the physical processes but we cannot measure the added value of this refined description in terms of predicted runoff.

Response:

Good idea. In our revised manuscript, we have setup a prediction model without the hydrologically similar groups (non_PwM) to show the effect of grouping on the PwM. The Cross-validations result of the PwM and non_PwM show that the performance of the PwM with the hydrologically similar groups is better and more stable than that of the non_PwM. In the revised manuscript, we have added the description and analysis of this part.

“**4.2.2 PwM without the hydrologically similar groups**

For comparison, we estimated Pw without the hydrologically similar groups, defined as non_PwM. The non_PwM is as follows:

$$non\_Pw = a_1 \times SM^2 + a_2 \times SM + b_1 \times FVC^2 + b_2 \times FVC \quad (3)$$

where non_Pw is the annual value of Pw simulated by non_PwM; *SM* is annual average value of surface soil moisture (0-10cm underground); *FVC* is annual average value of fractional vegetation cover; a_1, a_2, b_1 and b_2 represent the empirical coefficient fitted by least square method.”

“**5.2 Cross-validations based on data collected from globally published literatures**

The performance of the PwM and non_PwM were cross-validated based on the data collected from globally published literatures using the bootstrap sampling method (Fig. 3). On average, the ensemble RelBIAS of the Pw simulated by the PwM is slightly negative (Fig. 3a), indicating a weak tendency to underestimate the values of Pw, but the maximum relative bias is less than 0.1. The interquartile range of R^2 for the PwM is from 0.35 to 0.40, with a median of 0.37. The scores of R^2 are higher than 0.3 in more than 95% of the bootstrap sampling events. The NSE skill scores show that in most bootstrap samplings, the estimation error estimated variance for the PwM is less than the variance of the observations (NSE > 0), with the interquartile range from 0.33 to 0.39. In comparison, the maximum relative bias of the Pw simulated by the non_PwM is 0.12, the median of R^2 is 0.13, and the median of NSE is 0.13. Overall, cross validations show that the performance of the PwM with the hydrologically similar groups is better and more stable than that of the non_PwM.

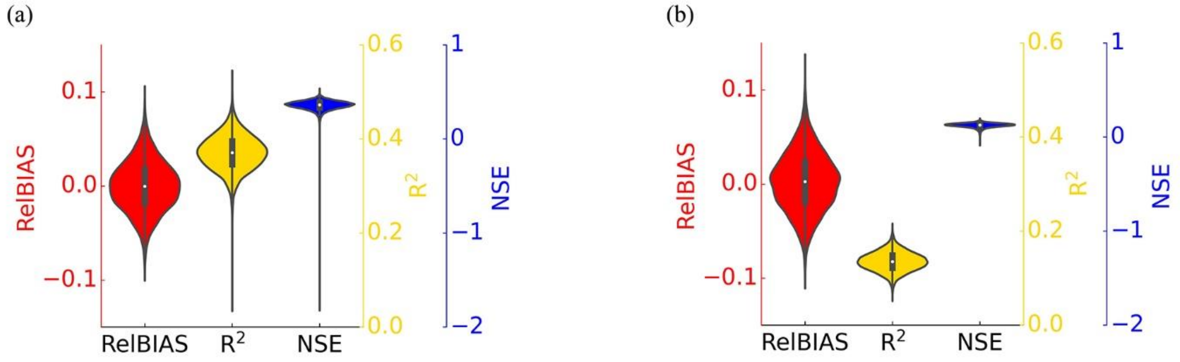

Figure 3. Cross-validation results of (a) PwM and (b) non_PwM. A violin represents the distribution of the considered skill scores. The white dot on the violin plot represents the median. The black bar in the center of the violin represents the interquartile range. Colors distinguish three performance metrics: Red (RelBIAS), yellow (R2) and blue (NSE)."

Minor comments:

Comment 3:

l.48: Note that the climate seasonality is not taken into account in basic Budyko-type equations so the sentence needs modification, maybe change "climatic conditions" by "mean annual climatic conditions".

Response:

Thank you, according to your comments, we have changed the "climatic conditions" to "mean annual climatic conditions".

Comment 4:

Table 1: please add a column with the parameter to be calibrated, the analytical role of the parameter in the equation (increase/decrease of evaporative ratio with increasing parameter value) and it would be highly beneficial to the reader if some information on previous estimation/calibration of these parameters could be given in this table.

Response:

Thank you. We have added two columns in table 1. One column is used to list the symbols for the watershed characteristic parameter (Pw) and their theoretical range, and the other column is used to list the reference values of Pw in the previous research. Table 1 in the revised manuscript was modified as follows,

Table 1. Parametric Budyko-type formulations (Pw - watershed characteristic parameter; ET - actual evaporation, R - runoff, P - precipitation, PET - potential evapotranspiration, all in mm yr^{-1}).

Reference	Formulation	Pw (Theoretical range)	Reference values of Pw
Budyko (1974)	$\frac{ET}{P} = \left[\frac{PET}{P}\tanh\left(\frac{PET}{P}\right)^{-1}(1-exp(-\frac{PET}{P}))\right]^{0.5}$	0.5	0.5
Zhang et al. (2001)	$\frac{ET}{P} = \frac{1+w\frac{PET}{P}}{1+w\frac{PET}{P}+(\frac{PET}{P})^{-1}}$	w $(0, \infty)$	Trees – 2.0, Plants – 0.5

Turc (1954), Mezentsev (1955), Choudhury (1999), Yang et al. (2008)	$\frac{ET}{P}=\frac{1}{\left[1+(\frac{P}{PET})^n\right]^{\frac{1}{n}}}$	n (0, ∞)	Field – 2.6, River basins – 1.8
Wang and Tang (2014)	$\frac{ET}{P}=\frac{1+\frac{PET}{P}-\sqrt{(1+\frac{PET}{P})^2-4\varepsilon(2-\varepsilon)\frac{PET}{P}}}{2\varepsilon(2-\varepsilon)}$	ε (0,1)	0.55 - 0.58
Tixeront (1964), Fu (1981), Zhou et al. (2015)	$\frac{R}{P}=\left[1+\left(\frac{P}{PET}\right)^{-m}\right]^{\frac{1}{m}}-(\frac{P}{PET})^{-1}$	m (1, ∞)	Forest – 2.83, Shrub – 2.33, Grassland or cropland – 2.28, Mixed land – 2.12

Comment 5:

l.58-63: At this stage of the manuscript, it is unclear what Pw stands for. Is it an a priori estimation of the parameters that are apparent in the equations of Table 1? To make things clearer the column of table 1 indicating the free parameter could be headed Pw.

Response:

Thanks for your question. The Pw is the w in Zhang equations, n in Yang equations, ε in Wang and Tang equations, and m in Fu equations. As you suggested, we have added a column for listing the symbol of Pw and their theoretical range respectively. The modifications are as follows,

"As a result, hydrologists have invested considerable efforts to improve model performance by introducing parameters related to watershed characteristics (watershed characteristic parameter, Pw) into the original Budyko equation. The popular parametric equations are presented in Table 1.

Table 1. Parametric Budyko-type formulations (Pw - watershed characteristic parameter; ET - actual evaporation, R - runoff, P - precipitation, PET - potential evapotranspiration, all in mm yr^{-1}).

Reference	Formulation	Pw (Theoretical range)	Reference values of Pw
Budyko (1974)	$\frac{ET}{P}=\left[\frac{PET}{P}\tanh\left(\frac{PET}{P}\right)^{-1}(1-exp(-\frac{PET}{P}))\right]^{0.5}$	0.5	0.5
Zhang et al. (2001)	$\frac{ET}{P}=\frac{1+w\frac{PET}{P}}{1+w\frac{PET}{P}+(\frac{PET}{P})^{-1}}$	w (0, ∞)	Trees – 2.0, Plants – 0.5
Turc (1954), Mezentsev (1955), Choudhury (1999), Yang et al. (2008)	$\frac{ET}{P}=\frac{1}{\left[1+(\frac{P}{PET})^n\right]^{\frac{1}{n}}}$	n (0, ∞)	Field – 2.6, River basins – 1.8
Wang and Tang (2014)	$\frac{ET}{P}=\frac{1+\frac{PET}{P}-\sqrt{(1+\frac{PET}{P})^2-4\varepsilon(2-\varepsilon)\frac{PET}{P}}}{2\varepsilon(2-\varepsilon)}$	ε (0,1)	0.55 - 0.58

Tixeront (1964), Fu (1981), Zhou et al. (2015)	$\frac{R}{P}=\left[1+\left(\frac{P}{PET}\right)^{-m}\right]^{\frac{1}{m}}-(\frac{P}{PET})^{-1}$	m $(1,\infty)$	Forest – 2.83, Shrub – 2.33, Grassland or cropland – 2.28, Mixed land – 2.12

Comment 6:

l.67-68: Does it depend on the equation? Since the statement is general and not specific to a given equation, this needs more details.

Response:

Thanks for your question. These results are based on the parametric Budyko-type formulations. We have rewritten this fragment in the revised manuscript, detailing the relationship between the Pw and the watershed characteristic factors in previous studies.

"During the past decades, researchers have done lots of work to quantify the Pw for the accurate simulation of evapotranspiration or runoff using the Budyko framework (Wang et al., 2022; Yao et al., 2017; Guo et al., 2019; Yu et al., 2021) and made considerable contributions for improving the estimation of Pw by taking into account the influences from watershed characteristics (Fu, 1981; Liu and Liang, 2015; Guan et al., 2022; Yang et al., 2008). Although there is agreement that the Pw represents the integrated effects of various environmental factors (Wang et al., 2022; Liu et al., 2022b; Yu et al., 2021; Gan et al., 2021), studies still differed greatly as to what factors and effects should relate to the Pw and failed to give a general framework for quantifying the Pw. For instance, whether the Pw in the Budyko framework is controlled by vegetation or not has been much debated. Ning et al. (2017) found that the Pw generally had a positive correlation with vegetation coverage. Zhang et al. (2018) obtained the sensitivity of the Pw to changes in LAI by taking a derivative of the Pw function with respect to LAI, implying a crucial role of vegetation cover in impacting the Pw. However, some other studies indicated that most regions or watersheds show no significant influences of vegetation indices or coverage on Pw (Li et al., 2013; Liu et al., 2021). For example, Li et al. (2013) pointed out the variations in the Pw values are not entirely controlled by vegetation coverage in the small catchments. Another study from Liu et al. (2021) also found a weak correlation between vegetation leaf area index and the Pw. Therefore, more in-depth studies are in need for revisiting the hydrological Basis of Pw in the Budyko Framework."

Comment 7:

l.72: The term "contradictory" is not appropriate. There is no clear consensus but some results are relatively consensual (e.g., positive relationship between Pw and vegetation cover).

Response:

Thank you, according to your comments, we have replaced the simple word "contradiction" with the description of the results of the previous studies. The details are as follows,

"During the past decades, researchers have done lots of work to quantify the Pw for the accurate simulation of evapotranspiration or runoff using the Budyko framework (Wang et al., 2022; Yao et al., 2017; Guo et al., 2019; Yu et al., 2021) and made considerable contributions for improving the estimation of Pw by taking into account the influences from watershed characteristics (Fu, 1981; Liu and Liang, 2015; Guan et al., 2022; Yang et al., 2008). Although there is agreement that the Pw represents the integrated effects of various environmental factors (Wang et al., 2022; Liu et al., 2022b; Yu et al., 2021; Gan et al., 2021), studies still differed greatly as to what factors and effects should relate to the Pw and failed to give a general framework for quantifying the Pw. For instance, whether the Pw in the Budyko framework is controlled by vegetation or not has been much debated. Ning et al. (2017) found that the Pw generally had a positive correlation with vegetation coverage. Zhang et al. (2018) obtained the sensitivity of the Pw to changes in LAI by taking a derivative of the Pw function with respect to LAI, implying a crucial role of

vegetation cover in impacting the Pw. However, some other studies indicated that most regions or watersheds show no significant influences of vegetation indices or coverage on Pw (Li et al., 2013; Liu et al., 2021). For example, Li et al. (2013) pointed out the variations in the Pw values are not entirely controlled by vegetation coverage in the small catchments. Another study from Liu et al. (2021) also found a weak correlation between vegetation leaf area index and the Pw. Therefore, more in-depth studies are in need for revisiting the hydrological Basis of Pw in the Budyko Framework."

Comment 8:

l.78: the term essential is debatable. Splitting into groups leads to non-universal laws. I agree this could lead to better performance and it is, therefore, to be tested but the motivations in terms of the physical process are not clear at this stage of the manuscript. So the term essential is in my opinion too strong and I suggest changing it to "useful".

Response:

Thank you, according to your comments, we have changed the "essential" to "useful".

Comment 9:

Data section: it is unclear why data are not taken homogeneously among published datasets and GRDC. The main caveat lies in the differences in the climatic forcing data (P and PET). Why not merge the data and use a single product to derive precipitation? Also, this would allow the authors to homogenize the calibration and validation datasets that appear largely different in terms of geographic locations (and climate settings). Also, is there a criterion on the number of years of data for including a catchment in the dataset? Last, it is not clear if climatic data are aggregated over catchment areas. Do the authors delineate catchment boundaries?

Response:

(1) Reasons for using published datasets and GRDC datasets

We used the published datasets for modeling and the GRDC data for verification of runoff reconstruction. All of the published data we collected in this study came from the conservative watersheds (i.e., precipitation is the sum of streamflow and evapotranspiration). We need to use such conservative watershed data for modeling. However, the time of collected datasets are discrete and discontinuous. To verify the performance of the model in time series, we used the GRDC data to verify reconstructed time-series runoff.

(2) Reasons for using GPCC precipitation data

Compared to the CRU TS precipitation dataset, the Global Precipitation Climatology Centre (GPCC) precipitation data was found to be more agreeable with the observation in the previous researches (Ahmed et al., 2019; Degefu et al., 2022; Fiedler and Döll, 2007; Hu et al., 2018; Salaudeen et al., 2021). Therefore, the P values for runoff reconstruction were extracted from GPCC Precipitation data. In the revised manuscript, we have added the explanation on the reasons for using GPCC precipitation data, as follows,

"The P values for runoff reconstruction were extracted from Global Precipitation Climatology Centre (GPCC) Precipitation Total Full V2018 (0.5×0.5) data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA. It is because that the Global Precipitation Climatology Centre (GPCC) precipitation data was found to be more agreeable with the observation in the previous researches compared to the CRU TS precipitation dataset(Ahmed et al., 2019; Degefu et al., 2022; Fiedler and Döll, 2007; Hu et al., 2018; Salaudeen et al., 2021)"

(3) The time range of datasets

For the collected datasets, the times of observation are discrete and discontinuous (listed in Supplement 1). The time range of verification datasets, including the data used for reconstructing runoff and the observed runoff data from GRDC, is from 2000 to 2016.

(4) The Climate data extraction method

For the GRDC watersheds, the climate data (including P and PET data) were extracted from grid data based on the boundary files provided by GRDC Watershed Boundaries (2011). For the collected watersheds from published literatures without boundary files, the PET data were extracted from grid data according to the coordinate points of these watersheds. We have added that description in the revised manuscript,

"Potential evapotranspiration (PET, mm yr^{-1}) data were downloaded from version 4.05 of the CRU TS (Climatic Research Unit gridded Time Series) climate dataset (https://doi.org/10.6084/m9.figshare.11980500), which is produced by the CRU at the University of East Anglia. For consistency, we used PET values extracted from the CRU TS dataset of all watersheds listed in Supplement 1, even for studies with PET values reported. The PET values were extracted based on the coordinate points of watersheds."

"We used the boundary of watersheds provided by GRDC Watershed Boundaries (2011) to extract the average values of PET and P from grid datasets for each watershed. The PET values were extracted from the CRU TS dataset. The P values for runoff reconstruction were extracted from Global Precipitation Climatology Centre (GPCC) Precipitation Total Full V2018 (0.5×0.5) data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA."

Comment 10:

l.95-96: Please indicate the formulation used for potential evaporation.

Response:

The potential evapotranspiration (PET, mm yr-1) data were downloaded from version 4.05 of the CRU TS (Climatic Research Unit gridded Time Series) climate dataset (https://doi.org/10.6084/m9.figshare.11980500), which is produced by the CRU at the University of East Anglia. The formulation used for CRU TS potential evapotranspiration is as follows,

$$PET = \frac{0.408 \triangle (R_n - G) + \gamma \frac{900}{T + 273.16} U_2 (e_a - e_d)}{\triangle + \gamma (1 + 0.34 U_2)}$$

We have added this formula to Supplement 1.

Comment 11:

l.98-100: why not do the same for precipitation data?

Response:

Thank you for your question. Generally, the monitoring data of watershed stations are more accurate than the data extracted by remote sensing data. Modelling may be more beneficial by using climate data monitored from watersheds stations. However, many sites in the collected data set did not provide corresponding values for potential evapotranspiration. Therefore, we used potential evapotranspiration values extracted from the CRU TS dataset of all watersheds in the collected data set, and used precipitation data collected from published data sets.

Comment 12:

l.132-133: why these three watershed characteristics? why only three? why not topographic attributes? These watershed characteristics are not stationary, do the authors change the value of these characteristics each year, or do they use aggregated statistics?

Response:

Thank you. In natural watershed, there are many factors affecting Pw, including soil moisture, vegetation coverage, seasonality, topography and so on. However, the topographic factor has little influence by other factors and remains stable for a long time. Therefore, topographic features are not considered in this study. Additional watershed characteristic factors may be considered in future studies. In the discussion section of the revised manuscript, we put forward the direction of future research, as follows, "Although the overall performance of PwM was satisfactory, we noted that the accuracy of the runoff simulated by the Budyko framework in some regions show either an overestimation or an underestimation. It is because the Pw in our study is only forced with soil moisture, seasonality index and fractional vegetation cover, and thus the estimated runoff could not clearly account for impacts from other drivers, like the effects of temperature anomalies and glacial meltwater on the hydrological regimes (Liu et al., 2022a). This is probably one of the main reasons for the severe underestimation of runoff in western North America and southern Europe (Fig. 7a, d). Future in-depth researches are in need to examine influences from other impact factors to improve the accuracy of Pw estimation in the Budyko framework."

Comment 13:

l.133-139: it is not clear how the regression tree is parametrized and optimized. Is it a supervised or unsupervised classification? As stated in lines 130-131, it seems that the authors want a supervised classification, but this would require a preliminary calibration of m. Is the number of groups imposed by the authors or it is the result of a cross-calibration experiment?

Response:

Thanks for your consideration. Since watershed feature factor data are continuous data, we used Decision Tree Regressor (DTR) instead of Decision Tree Classification (DTC) to find the turning point of the relationship between Pw and SM, and Pw and FVC in the study. The DTR does not involve supervised classification or unsupervised classification. We used the locations of splits in DTR as dividing intervals. We further describe the setup and performance of classifier in the revised manuscript,

"Three watershed characteristic variables — surface soil moisture (SM), rainfall seasonality index (SI), and fractional vegetation cover (FVC) — were selected for classification. For SM and FVC, the bounded intervals of the variables were given by the Decision Tree Regressor (DTR). The locations of splits in DTR were used as dividing intervals. The Scikit-learn library (Pedregosa et al., 2011) in Python provides the DTR used in this study. The criterion for measuring the quality of the split was set to "poisson" which uses reduction in Poisson deviance to find splits. The "random" strategy was used to choose local optimal splitting at each node. The results and performances of DTR are shown in Supplement 2. Based on the criteria using by Walsh and Lawler (1981), we divided the SI into three parts (SI≤0.4, 0.4<SI≤0.8, SI>0.8) to represent three hydroclimatic seasonality (precipitation spread throughout the year, marked seasonality with a short drier season, extreme seasonality with a long drier season). Finally, six hydrologically similar groups were classified (Table 3).

Table 3. Classification of watersheds

Soil moisture classifier	Water soil regime	Seasonality index classifier	Seasonality precipitation regime	Fractional vegetation cover classifier	vegetation cover regime	Name of the group
SM≤20	Dry soil	——	——	——	——	IN_D
SM>20	Wet soil	SI ≤ 0.4	Seasonless	——	——	IN_{WP}
		0.4 < SI ≤ 0.8	Marked seasonality	FVC ≤ 0.2 0.2 < FVC ≤ 0.5 FVC > 0.5	Low density Middle density High density	IN_{WMS} IN_{WMM} IN_{WML}
		SI > 0.8	Extreme seasonality	——		IN_{WE}

"

Comment 14:

l.153-155: Not clear at this stage whether the time step is annual. Numerous studies pointed out the problems of using the Budyko-type equations on an annual time step. This should be taken into account by the authors.

Response:

Thanks for your consideration. The time step used in our study is annual.

Comment 15:

l.159-168: Are the metrics computed on each catchment or all catchment runoff values? What is the minimum number of years for considering a catchment? If the record periods are too short, the resulting performance metrics might be meaningless.

Response:

Thank you. In the process of cross-validation (using the bootstrap sampling method), for the performance metrics, the term N is the number of test sets, i is the i^{th} value to be simulated by the trained PwM, and y_s and y_o are the simulated and observed series of test sets, respectively. In the process of runoff reconstruction verification, for the performance metrics, the terms y_s and y_o are the simulated and observed 17 years of each watershed, respectively. n the revised manuscript, we have added explanations on these terms,

"**4.3.2 Cross-validations using the bootstrap sampling method**

We used cross-validation to test the stability of the proposed PwM using the bootstrap sampling method. The collected public data were split into two parts, one for model training and the other for model validation. A subset of 60% of the data was randomly selected using the bootstrap sampling method for training PwM. The remaining 40% of the runoff data were used to evaluate the model performance using the validation metrics in section 4.3.1. For each metric, the term N is the number of test sets, i is the i^{th} value to be simulated by the trained PwM, and y_s and y_o are the simulated and observed series of test sets, respectively. The process was repeated randomly 10000 times. We documented the cross-validation result of each bootstrapping and showed them in the violin plot (Fig. 3).

4.3.3 Validations of GRDC time-series runoff reconstruction results

To further assess the model performance, we applied the proposed PwM into Fu's model to reconstruct the time-series runoff data of GRDC from 2000 to 2016. Finally, the time-series runoff data from 545 GRDC stations, which were selected by Sect. 3.1, were used to evaluate the model performance using the validation metrics in section 4.3.1. For each metric, the terms y_s and y_o represent the simulated and observed time-series runoff data, respectively."

Comment 16:

l.170: In the data section, the authors present a calibration and a validation dataset, now, they say they perform bootstrapping... Is it a bootstrapping on the calibration dataset?

Response:

Thank you. For cross-validation (using the bootstrap sampling method), we split the collected data set into two parts, one for model training and the other for model validation. A subset of 60% of the data was randomly selected using the bootstrap sampling method for training PwM. The remaining 40% of the runoff data were used to evaluate the model performance using the validation metrics in section 4.3.1. The process was repeated randomly 10000 times. For the use of data, we have made the detailed descriptions in the revised manuscript,

"**4.3.2 Cross-validations using the bootstrap sampling method**

We used cross-validation to test the stability of the proposed PwM using the bootstrap sampling method. The collected public data were split into two parts, one for model training and the other for model validation. A subset of 60% of the data was randomly selected using the bootstrap sampling method for training PwM. The remaining 40% of the runoff data were used to evaluate the model performance using the validation metrics in section 4.3.1. For each metric, the term N is the number of test sets, i is the i^{th} value to be simulated by the trained PwM, and y_s and y_o are the simulated and observed series of test sets, respectively. The process was repeated randomly 10000 times. We documented the cross-validation result of each bootstrapping and showed them in the violin plot (Fig. 3).

4.3.3 Validations of GRDC time-series runoff reconstruction results

To further assess the model performance, we applied the proposed PwM into Fu's model to reconstruct the time-series runoff data of GRDC from 2000 to 2016. Finally, the time-series runoff data from 545 GRDC stations, which were selected by Sect. 3.1, were used to evaluate the model performance using the validation metrics in section 4.3.1. For each metric, the terms y_s and y_o represent the simulated and observed time-series runoff data, respectively."

Comment 17:

l.180-201: I think this should be placed in the Data section.

Response:

Good idea. We have placed this to the data section, as follows,

"Observed river discharge data for validation were obtained from the Global Runoff Data Centre (GRDC, https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html). Only the GRDC stations meeting the following criteria were selected for further analysis: (1) The sites with continuous time-series runoff observations during the period 2000–2016 and corresponding SM, FVC, and SI were also available during such a period; (2) The drainage area reports can be found in the original data to provide area parameters for converting original flow volumes to runoff rates; (3) The geographical coordinates reports can be found in the original data and the shape of the drainage can be found in the GRDC Watershed Boundaries (2011); (4) The watersheds of "non-conservative" ($m>10$) and unrealistic runoff rates ($m<1$) are removed. Based on these criteria, 545 GRDC stations were selected for validation (Fig. 1). Then, the flow volumes of selected sites were converted to runoff rates (Ghiggi et al., 2019)."

Comment 18:

l.198-201: I think we can assume that the reader knows how to convert volumetric discharge to runoff depth.

Response:

Thank you. We have deleted this part as suggestion.

Comment 19:

Figure 2: not clear at all what is represented. FVC changes each year. Do the authors plot the aggregated FVC over the temporal range of measured streamflow? Do the calibrated m for each year or globally over the entire record period? Figure caption should detail each panel explicitly.

Response:

Thank you for your question. Figure 2 shows the regressions between Pw in Fu's formula and watershed characteristic variables collected from globally published datasets. Here, the FVC were the annual average values corresponding to the runoff observation period, and extracted by the coordinates of site from grid data. We performed regression analysis between the Pw and watershed characteristic variables to determine the input variables of the PwM. The variables whose R^2 of the regression model was greater than 0.1 were selected as input variables. We used a polynomial as the basic model form. Each term of the polynomial depends on the regression model of the corresponding variable and the Pw. After the model was determined, we extracted the annual average values of FVC from 2000 to 2016 to reconstruct the time-series runoff in 545 GRDC stations. Therefore, the proposed PwM not only represented the runoff observations in 366 watersheds from global published literatures, but can also reconstruct the time-series runoff in 545 GRDC stations.

We have redrawn Fig. 2 and modified the related description and analysis, as follows,

"The regressions between Pw in Fu's formula and watershed characteristic variables collected from globally published datasets are shown in Fig. 2. Analyses show that soil moisture (SM) and fractional vegetation cover (FVC) are strongly correlated to Pw in each group. The Pw values in dry watersheds with SM≤20mm monotonically increases with SM following a power function (Fig. 2a). However, in humid watersheds with SM>20mm, the Pw values converts to monotonically decrease with SM, which is also in a power function (Fig. 2b). And the fractional vegetation cover (FVC) shows linearly correlated with the Pw values of watersheds in most hydrologically similar groups but differ greatly between different groups (Fig. 2c-h). There is positive linear correlation between Pw and FVC in the IN_D, IN_{WSS} and IN_{WE} groups; while the relationship turns to be a negative linear equation in the IN_{WMM}

and IN_{WML} groups. However, in the IN_{WP} group, the relationship between Pw and FVC is not significant. Therefore, in the proposed PwM, SM and FVC were selected as input variables (i.e., *Var_n*) for all the groups, except that FVC was rejected in the IN_{WP} group. The formula in PwM for calculating the Pw is modeled as sum of a power function of SM and a linear function of FVC, given by Equation 7."

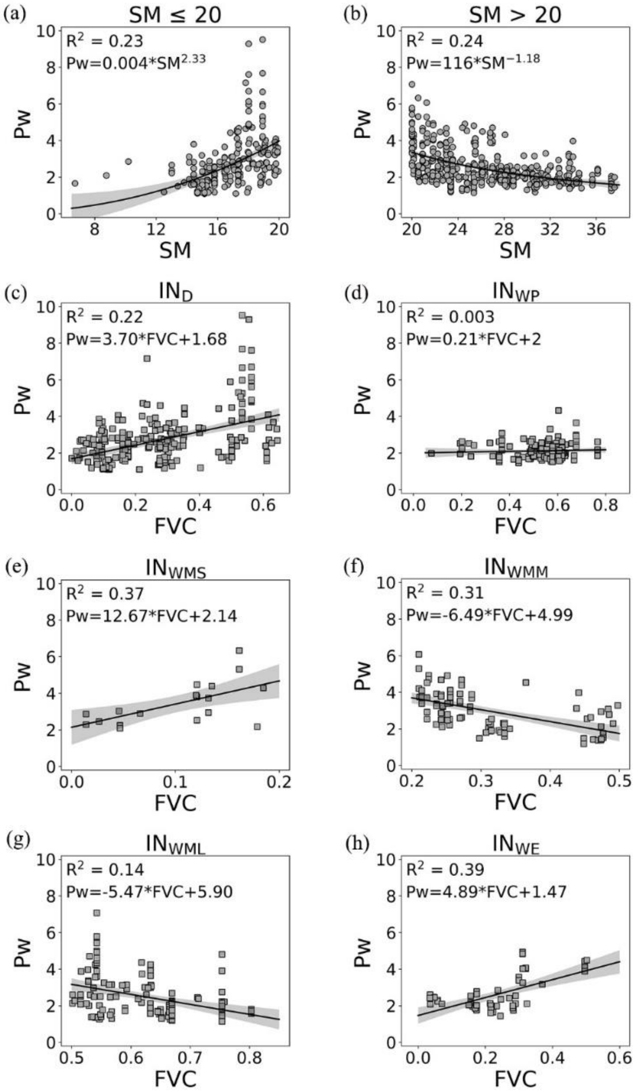

Figure 2. Regression between Pw in Fu's formula and (a) SM (SM≤20mm), (b)SM (SM>20mm), (c)FVC (IN_D), (d)FVC (IN_{WP}), (e)FVC (IN_{WMS}), (f)FVC (IN_{WMM}), (g)FVC (IN_{WML}), and (h)FVC (IN_{WE}). Symbol shapes indicate SM (dot) and FVC (square).

References

Ahmed, K., Shahid, S., Wang, X., Nawaz, N., and Khan, N.: Evaluation of gridded precipitation datasets over arid regions of Pakistan, Water, 11, 210, 2019.

Budyko, M. I.: Climate and life, Academic press1974.

Choudhury, B.: Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model, Journal of Hydrology, 216, 99-110, 1999.

Degefu, M. A., Bewket, W., and Amha, Y.: Evaluating performance of 20 global and quasi-global precipitation products in representing drought events in Ethiopia I: Visual and correlation analysis, Weather and Climate Extremes, 35, 100416, 2022.

Fiedler, K. and Döll, P.: Global modelling of continental water storage changes–sensitivity to different climate data sets, Advances in Geosciences, 11, 63-68, 2007.

Fu, B.: On the calculation of the evaporation from land surface, Chinese Journal of Atmospheric Sciences, 5, 23-31, 1981.

Gan, G., Liu, Y., and Sun, G.: Understanding interactions among climate, water, and vegetation with the Budyko framework, Earth-Science Reviews, 212, 103451, 2021.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth System Science Data, 11, 1655-1674, 2019.

GRDC: Watershed Boundaries of GRDC Stations / Global Runoff Data Centre, Federal Institute of Hydrology (BfG) [dataset], 2011.

Guan, X., Zhang, J., Yang, Q., and Wang, G.: Quantifying the effects of climate and watershed structure changes on runoff variations in the Tao River basin by using three different methods under the Budyko framework, Theoretical and Applied Climatology, 1-14, 2022.

Guo, A., Chang, J., Wang, Y., Huang, Q., Guo, Z., and Li, Y.: Uncertainty analysis of water availability assessment through the Budyko framework, Journal of Hydrology, 576, 396-407, 2019.

Hu, Z., Zhou, Q., Chen, X., Li, J., Li, Q., Chen, D., Liu, W., and Yin, G.: Evaluation of three global gridded precipitation data sets in central Asia based on rain gauge observations, International Journal of Climatology, 38, 3475-3493, 2018.

Li, D., Pan, M., Cong, Z., Zhang, L., and Wood, E.: Vegetation control on water and energy balance within the Budyko framework, Water Resources Research, 49, 969-976, 2013.

Liu, J., You, Y., Zhang, Q., and Gu, X.: Attribution of streamflow changes across the globe based on the Budyko framework, Science of The Total Environment, 794, 148662, 2021.

Liu, Q. and Liang, L.: Impacts of climate change on the water balance of a large nonhumid natural basin in China, Theoretical and Applied Climatology, 121, 489-497, 2015.

Liu, S.-W., Wang, X., Zhang, L., Kong, W.-M., Gao, H.-K., and Xiao, C.-D.: Effect of glaciers on the annual catchment water balance within Budyko framework, Advances in Climate Change Research, 13, 51-62, 2022a.

Liu, S., Wang, X., Zhang, L., Kong, W., Gao, H., and Xiao, C.: Effect of glaciers on the annual catchment water balance within Budyko framework, Advances in Climate Change Research, 13, 51-62, 2022b.

Mezentsev, V.: Back to the computation of total evaporation, Meteorologia i Gidrologia, 5, 24-26, 1955.

Ning, T., Li, Z., and Liu, W.: Vegetation dynamics and climate seasonality jointly control the interannual catchment water balance in the Loess Plateau under the Budyko framework, Hydrology and Earth System Sciences, 21, 1515-1526, 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V.: Scikit-learn: Machine learning in Python, the Journal of machine Learning research, 12, 2825-2830, 2011.

Salaudeen, A., Ismail, A., Adeogun, B. K., Ajibike, M. A., and Zubairu, I.: Evaluation of ground-based, daily, gridded precipitation products for Upper Benue River basin, Nigeria, Engineering and Applied Science Research, 48, 397-405, 2021.

Tixeront, J.: Prediction of streamflow, IAHS publication, 118-126, 1964.

Turc, L.: The water balance of soils. Relation between precipitation, evaporation and flow, Ann. Agron, 5, 491-569, 1954.

Walsh, R. and Lawler, D.: Rainfall seasonality: description, spatial patterns and change through time, Weather, 36, 201-208, 1981.

Wang, D. and Tang, Y.: A one-parameter Budyko model for water balance captures emergent behavior in Darwinian hydrologic models, Geophysical Research Letters, 41, 4569-4577, 2014.

Wang, F., Xia, J., Zou, L., Zhan, C., and Liang, W.: Estimation of time-varying parameter in Budyko framework using long short-term memory network over the Loess Plateau, China, Journal of Hydrology, 607, 127571, 2022.

Yang, H., Yang, D., Lei, Z., and Sun, F.: New analytical derivation of the mean annual water-energy balance equation, Water resources research, 44, 2008.

Yao, J., Mao, W., Yang, Q., Xu, X., and Liu, Z.: Annual actual evapotranspiration in inland river catchments of China based on the Budyko framework, Stochastic Environmental Research and Risk Assessment, 31, 1409-1421, 2017.

Yu, K.-x., Zhang, X., Xu, B., Li, P., Zhang, X., Li, Z., and Zhao, Y.: Evaluating the impact of ecological construction measures on water balance in the Loess Plateau region of China within the Budyko framework, Journal of Hydrology, 601, 126596, 2021.

Zhang, L., Dawes, W., and Walker, G.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water resources research, 37, 701-708, 2001.

Zhang, S., Yang, Y., McVicar, T. R., and Yang, D.: An analytical solution for the impact of vegetation changes on hydrological partitioning within the Budyko framework, Water Resources Research, 54, 519-537, 2018.

Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., Sun, G., Scott, D. F., Zhou, S., and Han, L.: Global pattern for the effect of climate and land cover on water yield, Nature communications, 6, 1-9, 2015.

Supplement 2: The results and performances of Decision Tree Regressor

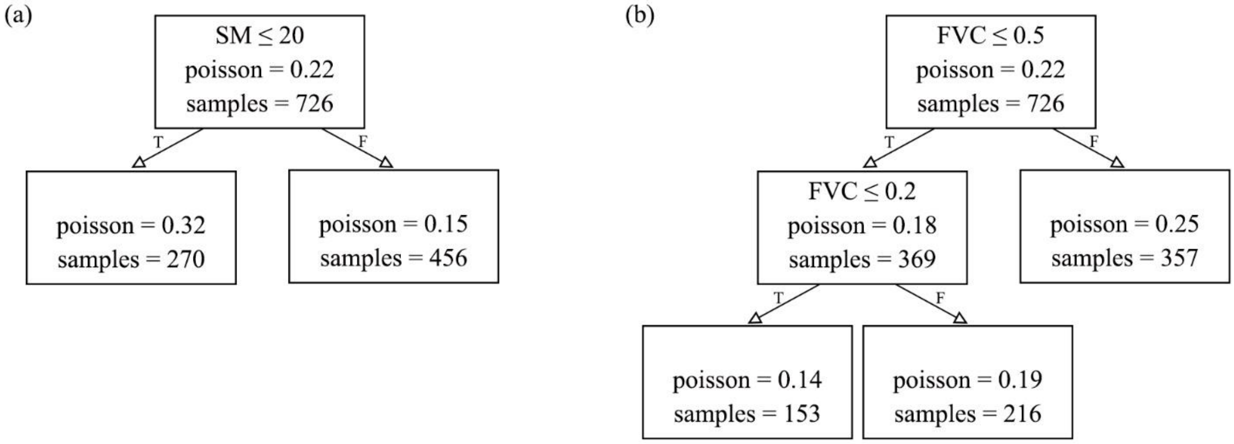

Figure S.1 The results and performances of Decision Tree Regressor for (a) surface soil moisture (SM) and (b) fractional vegetation cover (FVC). The "poisson" indicates the value of Poisson deviance, "samples" indicates the number of samples, "T" means True, and "F" means Fales.