# Replies to the comments from the editor and reviewers

We really appreciate the comments from the editor and reviewers that help improve the current paper. We reply the comments in this document with a brief introduction of the changes we made in the revised paper. The full revision as well as the marked-up revision can be found in the submission system.

Editor's comment: Reviewers provide positive comments on the paper. Authors are required to address these comments on the revision (i.e., some math description, novelty of the study). Particularly, it is recommended to further elaborate the new contribution of the study (especially in the cold region and lack of data), to make concise and smooth transition among sections (abstract and some parts are too long), and to include additional detailed analysis on the results.

Reply: Thanks for the suggestions. We have revised the manuscript according to the editor and reviewers' comments and suggestions. Particularly, we improved the structures of the paper and we simplified the abstract and conclusion that made these parts more concise and focused on the ORCHIDEE-Budyko framework, which is the innovative approach in this paper. The methodologies have been introduced in a better way with the improved structure and additions of an illustration and a few equations. Necessary explanations on the results are added in the revised manuscript.

Reply to reviewer #1

Please, find here my comments for the paper 'Understanding the water cycle over the upper Tarim basin: retrospect the estimated discharge bias to atmospheric variables and model structure.
First of all, in my opinion, this is a good paper that contributes substantially to the state of the art.
The authors show an in-depth and well-articulated understanding about the water cycle over the upper Tarim basin and provide a way to retrospect the bias from the discharge estimation. The authors found that the forcing variables contribute more to the bias in the discharge than the model structure. Therefore, the bias in the forcing variables should be first assessed in order to perform the analysis of water cycle. They provide guide then for futures studies.
In terms of the overall quality of the paper, I think it is good, and the novelty of the contribution is relevant. The results are very well discussed and articulated, and the references are up to date. The paper is also well structured, and the scientific results and conclusions are presented in a very clear, concise and sensitive way. This makes the paper an easy reading, and proves the researchers are expert in the field. The content and the quality of the paper fit into the HESS standards (novelty, substantial conclusions, validity of methods and assumptions). I also highlight the thorough interpretation of the results, which support the conclusions. Moreover, the detailed description of the experiments makes the traceability of the results very easy for

future researchers. The contribution is relevant for our community, and the authors give proper credit to related work. The title directly reflects the contents of the paper and the abstract provides a concise and complete summary.
In terms of the form, the paper is well written (as I indicate above). And the paper is well balanced.
However, in terms of mathematical formulation, I would like to see more details about how the method was implemented.
To wrap up, although, if possible, I would like to see some additional mathematical details (e.g. some technical specifications in the methodology), I think the paper is good to be published in its present form.
In case the Editor decides to ask for a reviewed manuscript, I am happy to review it again.

Reply: We appreciate the reviewers' high evaluation on our paper. We consider the mathematical formulations important in the model set-up and the expressions of them helpful for readers understand more easily the ways we took during our simulation. While, we did not emphasize the formulations in this paper mainly for the following two reasons. (1) The major novelty of this paper is the analysis of bias from different sources. The idea is to find the possible process which is missing (e.g. human interventions) or largely biased (e.g. precipitation or evapotranspiration) in the simulation. The goal is achieved by a set of comparisons between our model outputs and theoretical values (Budyko hypothesis) and other reference works. The bias in any single formulations or parameters may not exert such large variations in the final results. (2) We use references instead of giving a few generalized formulations for the model description because we think the general idea has been presented in the words and people can find the corresponding references for details. This way is better for readers with different interests.

Therefore, in the revised paper, we didn't add the formulations of modules in land surface model ORCHIDEE which are too complex. But, we added one illustration (Figure 3) and a few simple equations (attached in the end of this document) to quantify the bias that are helpful to simplify the understanding the ORCHIDEE-Budyko framework and the bias calculation (revised paper subsection 3.5).

Reply to reviewer #2

The manuscript addresses the different sources of precipitation and the effect on dis- charge, and also did the attribution analysis of discharge bias to forcing variables and model structure by using the Budyko assumption. The analysis of resources of precipitation and discharge is very important for the water cycle of the cold regions, and the attribution analysis contributes to the understanding and the simulation of water cycle. The research in this paper is very interesting and important. The content and the quality of the paper fit into the HESS standards. However, the innovation is not outstanding and some of the results is not well explained. Specific comments are as follows:

1) The logical structure, the abstract as well as the conclusions should be adjusted to highlight the innovation of the manuscript.

Reply: Thanks for the suggestions. The innovation of this paper is a ORCHIDEE-Budyko framework that help identify the discharge simulation bias to different sources which is difficult in regions lack of data. With considering all the comments together, we revised the paper in the following three ways.

a. We revised the abstract (attached in the end of this document) by clearly separating it into four parts (i.e. brief background, the proposed method, the general results and a short conclusion). The current abstract is also shortened and the innovation of the manuscript is easily to be found and emphasized in the abstract.

b. The most significant change in the structure is that we moved the previous subsection 4.3 (the introduction of biases and their classification as forcing or model structure) and part of subsection 4.4.1 (introduction of the Budyko analytical tool for bias analysis) forward to the end of the methodology part. These introductions are important for readers to understand the biases we considered in the paper and the way ORCHIDEE-Budyko solve the bias attribution task.

c. The conclusion part was revised accordingly. We shortened the conclusion but emphasized the framework we introduced as we did in the revised abstract.

2) More details on the methodology to identify the bias from the model structure and forcing inputs, as well as the evaluation index should be provided.

Reply: To better introduce the method, we firstly reorganized the paper as we replied to the question 1. We moved the previous subsections 4.3 and 4.4.1, where different biases and their classifications as forcing bias or model bias and the analytical approach, to the methodology part. So that the biases will be introduced in methodological part which is more logical and easier to find for readers. We then added a detailed illustration (Figure 3, also attached in the end of the document) which explains the ORCHIDEE-Budyko framework (e.g. the assumptions, the way to estimate the bias). The evaluation index (bias calculation) is also added in current section 3.5 (also attached in the end of the document). These modifications will help readers better understand the methodology.

3) For bias from the model structure, in section 4.1.2 the bias of the model structure is identified as the discharge difference between WFDEI-CCG and WFDEI-CCG-SF. While in section 4.4.1, the change caused by ET with unchanged P and PET is identified as the bias affected by the model structure. They are inconsonant, and what is the difference between these two?

Reply: We attribute the biases in the model simulation to either the forcing or the model structure. The forcing bias includes biases in precipitation and other atmospheric variables (e.g. radiation, wind etc.). Because the potential evapotranspiration (PET) is mainly determined by forcing variables, the bias in PET belongs to the forcing bias. Any other biases are involved in the bias from model structure.

The snow and soil freezing scheme (Sect 3.2.2, 'SF' as an abbreviation) is improved in the model used for WFDEI-CCG-SF simulation compared to the old version used for WFDEI-CCG simulation. The differences between the two simulations are induced by model structures because the forcing is the same. The new snow and soil freezing scheme determines the snow processes and the water movement in the soil, which will further affect the energy balance and evaporable water. Hence, the final response to the new scheme is the changes in ET and the discharge is changing as a result.

In the assumption 3 of subsection 4.4.1, the P and PET which are related to forcing conditions are considered right. The difference of estimated discharge from the observations is the bias of ET. The bias is caused due to discrepancies of ET processes in the models rather than the forcing. Therefore, the ET bias is also belonging to category of 'model structure'. It is consonant with the previous bias (SF module), as the SF module also changes the ET estimation at last. Bias in any processes that change the ET will be involved as the bias from model structure. The difference is not that ET in the previous subsection 4.4.1 is a cumulative result of all processes while the SF module is just one of the processes.

We added the necessary explanation of different bias as well as their bias sources in the revised paper as subsection 3.3.6 (Page 10, Line 14-19).

> "With the main biases listed as above, we consider bias in any processes that changes *P* or *PET* as bias from forcing and bias in any processes that directly changes *ET* as bias caused by model structure. Although the shifts in forcing variables will change the *ET* estimation, for example, the *P* restricts the available water for *ET*, this shift is still belonging to forcing category since the relation is indirect. The biases which directly affects *ET* includes biases in infiltration, soil water movement, snow processes, vegetation representation and many other model processes. And all these are considered as biases caused by model structures."

4) For the results analysis, more in-deep reason should be put forward instead of just describing the phenomenon. For example, why there exits deviation between WFD-CRU and WFDEI-CRU after 1990, as shown in Page 12? Why the discharge correlation decreases for the upper

Aksu, shown in Page 13.

Reply: a. The precipitation difference between WFD-CRU and WFDEI-CRU (Figure 4-b) is only determined by the CRU versions used for correcting the precipitation. As introduced in 3.1.2 Near-surface atmospheric conditions, the precipitation in WFD was corrected using CRU TS 2.10 and that in WFDEI was corrected using CRU TS 3.1(Weedon *et al.*, 2014). The two CRU versions differ in the time period, in the stations used and the methods employed (Harris *et al.*, 2014). The version difference induces the final difference shown in Figure 4-b and the difference enlarges after 1990.

We added in Page 12, Line27-28

> "The precipitation difference is mainly due to that two different versions of CRU (CRU TS 2.10 and CRU TS 3.1) are used for precipitation correction in WFD and WFDEI (See subsection 3.1.2 or Weedon et al., 2014)."

b. The correlation between the simulated discharge and observation at Aksu decreases in experiment WFDEI-CCG to that of WFDEI-CCG-SF (Figure 7-b). The decrease in the correlation can be understood by checking the annual cycle of discharge between the two experiments for Aksu (Figure 6-e) as "**an early discharge peak exists in May, while not enough runoff is generated in the summer period (Figure 6-e)" (Page 13, Line 30-31)**. The difference between the two experiments only attributes to the usage of new developed snow and soil freezing module (introduced in subsection 3.2.3). Therefore, we consider the change is due to better description of the new SF module because the snow-melting in spring is better captured in the WFDEI-CCG-SF simulation (March-May, Figure 5-e). Although the correlation decreases, it does not mean the model/simulation deteriorates because correlation only evaluates the similarity of temporal variation but ignores the fact that the discharge amount has been better estimated (Figure 6-e and Figure 7-a).

We added in Page 13, Line 31-33

> "Although the correlation decreases, it does not mean the model/simulation deteriorates because correlation only evaluates the similarity of temporal variation but ignores the fact that the discharge amount has been better estimated (Figure 6-e and Figure 7-a)."

5) How do you calculate the biases range in Table 5 since there is no observations for each of the forcing variables? The manuscript declares that the observations in the nearby regions or the regions with similar climatic and regional characteristic are used. But only the underestimate or overestimate can be concluded compared to the observations nearby regions instead of an exact value.

Reply: As presented in the caption of Table 5, the bias implies the bias of the values in the current variables compared to the values they should be. The current value is the value estimated by model and the ideal value is obtained from the Budyko curve. For instance, with assumption 1 for the upper Yarkand, we consider P as the only factor biased. To meet the

discharge observation, we got the 'corrected' P by Budyko approach as 453.4 mm/yr (Figure 9a, from point A to point B). Because the original P is 247.3 mm/yr, the bias is (247.3 – 453.4) / 453.4 * 100%= -43.2%. However, the value is obtained when we consider only P is incorrect, if there are other variables with bias, the possible bias of P could be smaller. So, the possible maximum range of P bias is -43.2% (Page 16, Line 21-23).

We do use the observations in nearby or similar regions, but these references are not used for obtaining the exact values in Table 5. They are used to analyze the possibility of the assumptions because we are not sure whether the assumptions exist (shown in Table 6). For instance, because the PET for Tibet Plateau ranges from 580-720 mm/yr (Chen, Liu and Axel, 2006), the PET (=1153.7 mm/yr) is therefore very likely overestimated for the Upper Hotan where the climate is not very different from the Tibet (Page 17, Line 10-13). On the other hand, it avoids over-correction. For example, if we consider only PET is biased in the Upper Yarkand, we have to correct it to 225 mm/yr to meet the observations, which is too low compared to the reference. The PET is thus not the only biased variable for this region (Page 17, Line 13-18). With the reference values, we can obtain the relative possibility of biases in variables P, ET and PET. Then, we can know the action we need to take first to improve our model simulation (Page 17, Line 19-21).

All the proposed changes are traced in the new version of manuscript.

Reference:

Chen, S., Liu, Y. and Axel, T. (2006) 'Climatic change on the Tibetan Plateau: Potential evapotranspiration trends from 1961-2000', *Climatic Change*, 76(3–4), pp. 291–319. doi: 10.1007/s10584-006-9080-z.

Harris, I. *et al.* (2014) 'Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset', *International Journal of Climatology*, 34(3), pp. 623–642. doi: 10.1002/joc.3711.

Weedon, G. P. *et al.* (2014) 'The WFDEI meteorological forcing data set: WATCH Forcing data methodology applied to ERA-Interim reanalysis data', *Water Resources Research*, 50(9), pp. 7505–7514. doi: 10.1002/2014WR015638.

**Abstract:**

The bias in atmospheric variables as well as that in model computation are two major causes of failures in discharge estimation. Attributing the bias in discharge estimation becomes difficult if the forcing bias cannot be evaluated and excluded in advance in places lack of qualified meteorological observations, especially in those cold and mountainous area (e.g. the Upper Tarim basin). In this study, we proposed an ORCHIDEE-Budyko framework which helps identify the bias range from the two sources (i.e. forcing and model structure) with a set of analytical approaches. A latest version of land surface model-ORCHIDEE was used to provide reliable discharge simulations based on the most improved forcing inputs. The Budyko approach was then introduced to attribute the discharge bias to two sources with prescribed assumptions. Results show that as the forcing biases, the water inputs (rainfall, snowfall or glacier melt) are very likely underestimated for the Tarim headwater catchments (-43.2%~-21.0%). Meanwhile, the potential evapotranspiration is unrealistically high over the upper Yarkand and the upper Hotan (1240.4 mm/yr and 1153.7mm/yr respectively). Determined by the model structure, the bias in actual evapotranspiration is possible but not the only contributor to the discharge underestimation (overestimated up to 105.8\% for the upper Aksu). Based on a simple scaling approach, we estimated the water consumption by human intervention ranging from $213.50\times 10^8$ $m^3$/yr to $300.58\times10^8$ $m^3$/yr up the Alar gauge station, which is another bias source in current version of ORCHIDEE. This study succeeded in retrospecting the bias from the discharge estimation to multiple bias sources of the atmospheric variables and the model structure. The framework provides a unique method for evaluating the regional water cycle and its biases with our current knowledge of observational uncertainties.

## Conclusion:

In this work, we proposed a ORCHIDEE-Budyko framework which is used to identify attribute the discharge bias in model simulation to different sources as the forcing and model structure. Bias in the precipitation (P) and any processes related to potential evapotranspiration (PET) are considered as bias from forcing and any processes affect the actual evapotranspiration estimation are considered as bias from model structures. The discharge simulation was provided by the land surface model-ORCHIDEE with latest improvements in its modules and driven by the most developed forcing inputs (WFDEI-CCG-SF). However, underestimation in the discharge still exists over the three headwater catchments, where the biases of P, PET and ET are analyzed with a Budyko analytical approach. With a set of assumptions, we isolated the biases in three variables and their possibility assessed with information from nearby and hydro-climatic similar regions. Results show that precipitation (here considered as the sum of rainfall, snowfall and glacier melt) underestimation is highly likely for the upper Yarkand and the upper Aksu, while the overestimation of PET is likely to affect the upper Yarkand and the upper Hotan. The overestimation in ET is possible but not likely cause for the discharge underestimation for all headwater catchments. In the lower oases, humans consume 83.1%-87.4% of the discharge for irrigation usage which is also a bias source in the current version of model. Thus, including detailed human modules is in need for any large-scale model.

In this attempt to analyze the performance of a complex land surface model over the Tarim basin, large biases are found in the discharge estimation. Our finding that the bias is most likely caused by the forcing variables rather than the model is probably the reason for the failures of other models as well. Our work provides more information about the Tarim basin's water cycle and guidance for future studies that the bias in forcing variables should firstly be assessed and reduced in order to perform pertinent analysis of the regional water cycle. Land surface models are a recommended tool for water cycle analysis because of their independence of calibration and good ability to simulate most variables of the water cycle and their interplay which facilitates the identification of bias sources. This kind of application along with the improvements of forcing data are also important for predicting water resources in the Tarim as well as other high-altitude basins in central Asia in a changing climate.

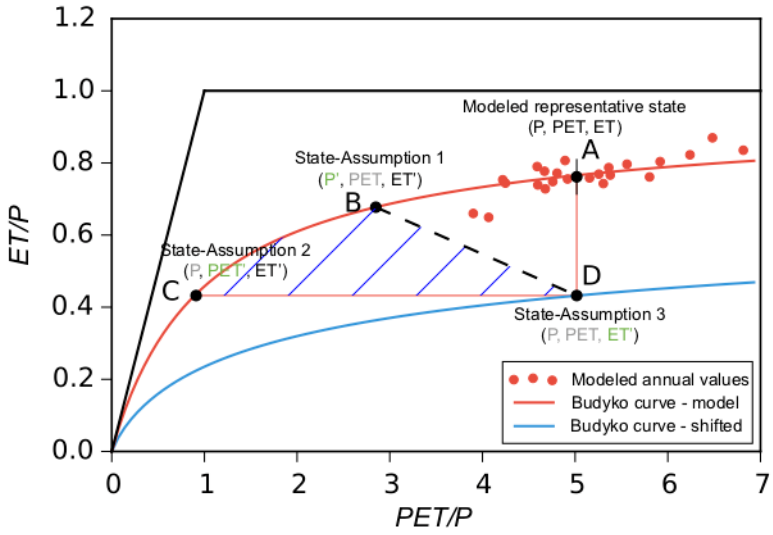


Figure 3. The illustration of the ORCHIDEE-Budyko framework. Point A represents the average state among the modeled annual values (with land surface model-ORCHIDEE), and the red curve is the simulated Budyko curve following the modeled state. Point B, C, D represent the representative state with shifting the P, PET and ET respectively with different assumptions to meet the discharge observations. A shifted Budyko curve (blue) is obtained crossing the point D which indicates a new state of model structure. The new points of B and C still stay on the original Budyko curve indicating that the model structure remain the same and the changes only relate to forcing variables. The shade area is the area among the three shifted states.

Assumption 1. ……

The possible maximum bias in P is calculated as

$$Bias(P)=(P-P')/P' \times 100\% \quad (1)$$

Assumption 2. ……

The PET is changed to a 'true' PET' and the possible maximum bias in PET is calculated as

$$Bias(PET)=(PET-PET')/PET' \times 100\% \quad (2)$$

Assumption 3. …...

The possible maximum bias in ET is calculated as

$$Bias(ET) = (ET - ET')/ET' \times 100\% \quad (3)$$

# Understanding the water cycle over the upper Tarim basin: retrospect the estimated discharge bias to atmospheric variables and model structure

Xudong Zhou[1,2], Jan Polcher[2], Tao Yang[1], Yukiko Hirabayashi[3], and Trung Nguyen-Quang[2]

[1]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, 210098, Nanjing, China
[2]Laboratoire de Météorologie Dynamique du CNRS, IPSL, École Polytechnique, 91128, Paris, France
[3]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

**Correspondence:** Xudong Zhou (xudong.zhou@lmd.polytechnique.fr)

**Abstract.**

The bias in atmospheric variables as well as that in model computation are two major causes of failures in discharge estimation. Attributing the bias in discharge estimation becomes difficult if ~~there is a~~ the forcing bias cannot be evaluated and excluded in advance in places lack of qualified meteorological observations~~. The problem is more complicated over the mountainous area where strong orographic effects exist and with severe heterogeneous geography~~ , especially in those cold and mountainous area (e.g. the Upper Tarim basin). In this study, we ~~in the first step improved the forcing inputs using local refined precipitation dataset and glacier simulations over the upper Tarim basin. The discharge bias was then obtained by comparing the estimated discharge from an advanced land surface model (ORCHIDEE) with *in-situ* discharge observations. We then introduced a framework with Budyko approach which succeeded in attributing the estimated bias to possible biases from major forcing variables~~ proposed an ORCHIDEE-Budyko framework which helps identify the bias range from the two sources (i.e. forcing and model structure) with a set of analytical approaches. A latest version of land surface model-ORCHIDEE was used to provide reliable discharge simulations based on the most improved forcing inputs. The Budyko approach was then introduced to attribute the discharge bias to two sources with prescribed assumptions. ~~The possibility of these biases was discussed by referring to many other studies with similar climatic or land surface characteristics.~~ Results show that ~~the~~ as the forcing biases, the water inputs (rainfall, snowfall or glacier melt) are very likely underestimated ~~especially for the headwater catchments of the upper Yarkand and the upper Aksu, with the largest range of 43.2%and 33.9%respectively~~for the Tarim headwater catchments (-43.2% -21.0%). Meanwhile, the potential evapotranspiration is unrealistically high over the upper Yarkand and the upper Hotan (1240.4 mm/yr and 1153.7mm/yr respectively). ~~The~~ Determined by the model structure, the bias in actual evapotranspiration ~~which is determined by the model structure is~~ is possible but not the only contributor to the discharge underestimation (overestimated ~~by 76.1%, 19.1% and~~ up to 105.8% for ~~its three headwater catchments~~the upper Aksu). Based on ~~the discharge simulation and bias analysis~~a simple scaling approach, we estimated the water consumption by human intervention ranging from $213.50\times10^8\mathrm{m}^3/\mathrm{yr}$ to $300.58\times10^8\mathrm{m}^3/\mathrm{yr}$ up the Alar gauge station, which is another bias source in current version of ORCHIDEE. This study succeeded in retrospecting the bias from the discharge estimation to multiple bias sources of the atmospheric variables and the model structure~~, although the framework needs further argumentations about its robustness, it~~ .

The framework provides a unique method for evaluating the regional water cycle and its biases with our current knowledge of observational uncertainties.

*Copyright statement.*

## 1 Introduction

A failure of discharge estimation can easily happen to a researcher especially when exploring a new region. It is often attributed to the model inapplicability to the region and tuning the model parameters is the common way to eliminate the discharge bias (Refsgaard, 1997; Westerberg et al., 2011). Although, a hidden assumption is often ignored that the atmospheric variables (or named here forcing) are essentially correct, while it may fail in some regions (Fekete et al., 2004; Adam et al., 2006). Without knowing the bias in forcing, the calibration becomes meaningless if the model parameters are tuned to values that are far from their physical meaning (Hernández and Francés, 2014). Thus, an important step before applying a model to a new region is to understand where the bias sources lie and their relative relations (Renard et al., 2010).

*In-situ* measurements are considered most reliable sources for atmospheric variables and thus can be used to evaluate or further correct the variables used to drive hydrological models. However, larger uncertainties are still found in mountainous and arid areas due to the poor representativity of *in-situ* observations (Adam et al., 2006; Harris et al., 2014; Yang et al., 2014). For instance, the precipitation over the mountainous area is mostly underestimated due to rare observations in high altitude and orographic effects (Harris et al., 2014). 20.2% of the precipitation is underestimated over global orographically affected regions according to Adam et al. (2006). Arid areas receive less water input but with larger relative uncertainties in precipitation (Fekete et al., 2004), which are crucial to regional runoff generation. Meanwhile, the energy flux over arid regions varies significantly. Thus the potential evapotranspiration ($PET$), as well as the actual evapotranspiration ($ET$), are quite uncertain over those areas (Federer et al., 1996; Weiß and Menzel, 2008) and in the meanwhile the $PET$ and $ET$ are variables unable to directly measure for a basin. Investigation of their errors and relations based on the model simulation becomes necessary.

Model efficiency needs to be verified firstly. The model performance is generally evaluated on the agreement of a single variable and discharge is the most common used as it is the result of all the water-energy processes. It reveals the accuracy of the whole system while it also accumulates all the errors from the forcing and the model. Therefore, a multivariable analysis based on the relation between variables is needed for overall evaluation (Kavetski et al., 2006). These relations represent typical climatic and regional characteristics, i.e., the aridity index ($PET/P$) reveals the energy and water input over a specific region (Zomer et al., 2007, 2008) and evapotranspiration ratio to precipitation ($ET/P$) is relevant to the land cover conditions (Liu et al., 2003; Yang et al., 2008). The Budyko hypothesis is a widely accepted empirical relation between $ET/P$ and $PET/P$ (Budyko, 1974). The shape of optimal Budyko curve reflects local climatic and underlying characteristics (Ponce et al., 2000; Yang et al., 2007). Hence with a Budyko curve derived from land surface model simulations, biases of the water-energy components ($P$, $ET$ or $PET$) can be assessed. For example, Adam et al. (2006) quantified the precipitation bias in

orographically effected areas using the Budyko hypothesis, although their work attributed all the bias in discharge simulation to the forcing with an incorrect assumption that their model was perfect.

Most of the hydrological models, with either lumped and distributed concept, are dependent on the calibration. Because of the assumption that water input $P$ is correct and a very crude description of energy processes, the $ET$ is the variables which can be adjusted to meet the water mass balance. Most of the bias is therefore assumed deriving from the $ET$. It may mislead the relations between $PET/P$ or $ET/P$ which represents the climatic and regional characteristics (Liu et al., 2003; Zomer et al., 2007; Yang et al., 2008; Zomer et al., 2008). Land Surface Models (LSMs), are almost independent of the calibration process with most of the model parameters obtained from multiple maps (e.g. land cover, land use, soil textures, etc.). Based on their substantial physical-based modules, the LSMs have been widely used to estimate most of the components in the continental water cycle (Yu et al., 1999; Yu, 2000; Pitman, 2003; Trenberth et al., 2006; Renard et al., 2010). Although the LSMs do not necessarily provide good estimates of discharge (Giorgi and Francisco, 2000; Fekete et al., 2004; Yu et al., 2006; Knutti et al., 2010), they prevent all the bias being revised by calibration. In return, the modelled discharge bias can reveal the biases related to model input or model structure, which have not attracted enough attention.

The Tarim basin, located in the Northwest of China, central Asia, surrounded by high mountains, oases in the center of deserts, is a region integrating the mountainous, arid and cold characteristics in different parts (Yang et al., 2015). The precipitation mainly distributes over the upper mountainous area around the boundaries, and the snow and glacier melt are major contributors to the local water resources (Gao et al., 2010; Pritchard, 2017). While the meteorologic observations on the water input components are sparse and the gauges are not representative because the surface conditions are heterogeneous especially in the mountainous area (Shen et al., 2010). In the lower oases, intensive irrigation is developed, which is hugely relied on the discharge flowing from the headwater catchments (Mamitimin et al., 2014), causing considerable changes in the natural river discharge downstream the area (Zhou et al., 2000; Tao et al., 2011). While, in reality, the human intervention is very difficult to model as it is policy related and because of the lack of efficient dataset. The anthropogenic effects on the water cycle, accompanying the climatic and topographic characteristics, make the Tarim one of the most challenging places to apply land surface models.

There are three major steps in this study, Firstly, we generated a best possible forcing dataset for the Tarim domain which reduces as far as possible the biases using refined data sets. The refined forcing then drove an improved land surface model (ORCHIDEE) to obtain the improved discharge estimations. Secondly, the estimated discharge was compared with *in-situ* discharge observations in section 4.1 and the evidence of their bias analyzed in section 4.2. In the third step the possible bias sources from the forcing and model structure were qualified with Budyko hypothesis in section 4.3 and their possibilities discussed in section 4.4. The model bias due to ignorance of human intervention is estimated based on the bias analysis over the headwater catchments in section 4.5.

## 2 Study area and hydro-meteorological characteristics

The Tarim basin locates in the Northwest China, surrounded by the Kunlun Mountains in the south, the Tienshan Mountain in the north and the Pamirs Plateau in the west (Figure 1). Its U-shaped terrain blocks the westerly atmospheric water vapor transport that leads to relatively low precipitation inside the basin (Wu et al., 2012). Simulated by Wu et al. (2012), 63% of the water vapor enters Tarim through the eastern passway, while it only happens in summer, contributing around 54% of the total annual precipitation, leading to a strong seasonality in precipitation (Tao et al., 2011). Combining with the glacier melt during warm summer, 70% of the annual discharge concentrates in the period from June to October (Liu et al., 2010). The high seasonality implies a high risk of water deficit in dry seasons and endangers ecosystem along the rivers (Döll et al., 2009) that requires human regulation to allow for efficient agriculture.

Despite the high inter-annual variability, the precipitation is heterogeneously distributed due to orographic effects (Wu et al., 2012). It ranges from 200 to 500 mm/yr in the mountainous area, while less than 50 mm in the central Tarim (Chen et al., 2007). The mountainous glacier/snow melt occurs in the same place where precipitation is generated. The mountainous area contributes almost all the river runoff of the Tarim basin, while the plains of Tarim contributes little to the water resource of the main Tarim (Yang et al., 2015; Shi et al., 2016). In the Upper Tarim, only three water systems (the Yarkand, the Hotan and the Aksu; see Figure 1) have natural hydraulic connections to the mainstream - the Tarim (Yang et al., 2015). The water originating from the mountains flows through the oases where people live and allows intensive agriculture (Mamitimin et al., 2014). As a consequence, a large proportion of the water is extracted for human utilization in the oases, so that only 8.7%, 43.4% and 30.6% of the river discharge from headwater region of Yarkand, Hotan and Aksu can finally reach Tarim mainstream (25.3% in overall in 1995; Zhou et al., 2000). The Kaxgar is another major tributary of the Tarim, while it has already dried up before water reaching the mainstream due to natural evaporation/leakage and human intervention. Of all the water consumption, the agriculture irrigation accounts for more than 95% in Tarim basin (Zhou et al., 2000), hence the dominated human influence in Tarim is considered to be the irrigation influence.

There are 11665 glaciers with a total area of 19878 $\mathrm{km}^2$ and a volume of 2313 $\mathrm{km}^3$ distributed over the Tarim (Liu et al., 2006). The glacier melt is a critical contributor to the local water resource. According to Zhou et al. (2000) and Shangguan et al. (2009), the estimated glacier melt accounts for around 40% of the total river runoff for the whole Tarim. While, due to the climate change, a large number of the glaciers were in retreat during last 40 years (1960s-2001). In the Upper Tarim, the Yarkand river has suffered the most significant glacier area changes (-205$\mathrm{km}^2$) with a relative proportion of -6.1%. The most significant retreat rate (-7.9%) in glacier area occurs in Pamirs Plateau in the west (Shangguan et al., 2009). All the changes in glaciers will result in the alteration in the river discharge and also the human interactions.

# 3 Data and Models

## 3.1 Data and simulation description

### 3.1.1 River discharge observations

River discharge is a very reliable and integrated observation of the continental water cycle which is always used as a validation variable. Over the headwater catchments, there are 13 hydrological gauges recorded in the Yearbook, though only six gauges are selected with consideration of their locations and data completeness. Two gauges (1,2) are in the upper Yarkand, two (3, 4) are in the upper Hotan, and two (5, 6) are in the upper Aksu, while no gauge is found on the Kaxgar river. For the gauges in headwater catchments, the discharge is considered free of human intervention, i.e., irrigation or reservoir regulation, so to a large extent they represent the natural environment. This facilitates model validation. Moreover, on the mainstream of Tarim, one gauge (7-Alar) was selected at the junction of three upstream rivers (Figure 1). Different from the headwater gauges, the river discharge at Alar has been significantly altered by human consumption after water flowing through the irrigation area (Mamitimin et al., 2014). Hence, the observations are no longer natural values but can be used to quantify the influence resulting from human activities.

### 3.1.2 Near-surface atmospheric conditions

Near-surface atmospheric conditions are crucial to hydrological responses (Adam et al., 2006). However, both the model simulation and gridded forcing generated from observations are proven to have large uncertainties where observations are sparse, and heterogeneity is strong (Harris et al., 2014; D'Orgeval et al., 2008), i.e., in the arid and mountainous area. So that, in practice, several forcing datasets are always used in parallel to generate an ensemble of climate conditions which hopefully tracks the uncertainties (Knutti et al., 2010; Tebaldi and Knutti, 2007). Alternatively, when possible, regional datasets which contain more information than global datasets are used to move the forcing closer to true values (Ines and Hansen, 2006). In this study, several sets of estimated forcing inputs based on WATCH reference (Harding et al., 2011; Weedon et al., 2014) are developed and then used to drive the land surface model (Figure 2). Among which the best simulations will be used to analyze the bias accompanying with its driving forcing.

A pair of reference forcing datasets are WFD (WATCH Forcing Data, 1958-2001) and WFDEI (WATCH Forcing Data methodology applied to ERA-Interim data, 1979-2012, Weedon et al., 2014). They use same methodologies but have slight differences in the basic data, processing and formatting (Weedon et al., 2014). In brief, WFDEI is an evolution of WFD where the underlying re-analysis is now ERA-Interim but using the same bias correction methodology. This product has been proven to be superior to WFD. CRU (Climate Research Unit) monthly total precipitation observations were used to bias correct the precipitation in WFD and WFDEI datasets. However, the WFD uses a previous version CRU TS2.10 before the CRU TS 3.1 used in WFDEI was released (Weedon et al., 2014). The two CRU datasets differ in the time period (CRU TS2.10:1901-2002, CRU TS3.1: 1979-2009), in the stations used and the methods employed, see (Harris et al., 2014) for details. Hence, there are

still some differences between the two which could further affect the hydrological responses. The two datasets are named as WFD-CRU and WFDEI-CRU respectively for later description, and the time step for all the forcing variables is 3 hours.

The CRU datasets were constructed by monthly observations at meteorological stations across the world's continents. The observations were then interpolated to 0.5-degree longitude/latitude grid cells. Though the CRU compares favorably to some other gridded datasets, it has significant deviations over regions and time periods with sparser observational data (Harris et al., 2014). Moreover, because only the monthly total precipitation was used to correct WFDEI and WFD, it cannot improve the temporal variabilities at smaller time step (i.e., daily or sub-daily). However, the precipitation variations in short period would result in different hydrological responses even in the condition that the total monthly amount remains the same (Potter et al., 2005).

Hence, in additional, the WFDEI-CRU dataset was further corrected by the gridded daily precipitation data from China Meteorological Administration (named as CMA). CMA precipitation product complies 2,416 national meteorological monitoring stations over China, using the climatological optimal interpolation method to generate the gridded 0.5 degree precipitation field from 1951 to 2016 (http://data.cma.cn; Shen et al., 2010). PRISM (Parameter-elevation Regression on Independent Sloped Methods; Daly et al., 2008) was used to lessen the orographic effects (Shen et al., 2010). The density of meteorological stations used in CMA is much higher than that used for CRU. For instance, there are only 6 gauges over Tarim basin in CRU database while 39 gauges are recorded in CMA system (Tao et al., 2011) so that it can to a certain degree improve the data applicability where precipitation is spatially inhomogeneous compared to CRU datasets. Given that the CMA data provides the daily information, it also improves the temporal variations of precipitation rather than using total monthly value in CRU datasets. The corrected atmospheric input dataset is named as WFDEI-CMA in the later context.

### 3.1.3 Glacier melt dataset

As mentioned in Sect.2, glacier melt is a vital water input to the Tarim. However, glacier runoff measurement is so difficult for such large regions that model-based estimates of the glacier melt are necessary. In general, glacier module is not coupled in LSMs (Fraedrich et al., 2005). Hence, rather than building a separated glacier module, we use an independent daily glacier melt dataset obtained from the glacier model called HYOGA2 which has been proven reliable over the globe (Hirabayashi et al., 2013). HYOGA2 is a temperature-index-based model utilizing an extensive global-scale glacier inventory and has several improvements compared to its first version HYOGA in model parameters simulation as well as the temporal extent. More details can be found in the original papers (Hirabayashi et al., 2010, 2013). The glacier melt is added to the rainfall series of WFDEI-CMA as an additional water flux to the system. The melt water is hence participating instantaneously in the water cycle without delays such as stores in ice, glacier pack or groundwater recharge beneath the ice being considered. This method was chosen for its simplicity and the lack of knowledge on the details at the transition between the glacier and the soil. Daily values are uniformly distributed over the 8 time steps per day of WFDEI. By adding the glacier melt, the fourth new forcing dataset is generated as WFDEI-CCG.

### 3.2 Land Surface Model-ORCHIDEE

The land surface model ORCHIDEE (Organizing Carbon and Hydrology In Dynamic EcosystEms) was developed by the Laboratoire de Météorologie Dynamique (IPSL-LMD) (Ducoudré et al., 1993; de Rosnay and Polcher, 1998; D'Orgeval and Polcher, 2008). After more than 20 years' development, ORCHIDEE has been validated from global scale (Alkama et al., 2010) to typical regional cases e.g. tropical rainforest area (Amazon, Guimberteau et al., 2012), semiarid regions (West Africa, D'Orgeval et al., 2008), middle latitude regions (Europe, Tallaksen and Stahl, 2014). Within the ORCHIDEE, only SECHIBA (Schematisation des Echanges Hydriques I'Interface entre la Biosphere et I'atmosphere) which represents the energy and water fluxes between land surfaces and the atmosphere is used in this study. The hydrological module in SECHIBA is based on developments by de Rosnay et al. (2003) and D'Orgeval (2006). 13 types of the vegetation are defined (D'Orgeval and Polcher, 2008) and dynamic LAI is computed to generate the interception and transpiration. The vertical soil water movement is represented by diffusion-type equations resolved on a fine vertical discretization (11 levels) and partitioning between infiltration and surface runoff through a time-splitting procedure (de Rosnay et al., 2002; D'Orgeval and Polcher, 2008; Guimberteau et al., 2012). The routing is conducted based on a linear reservoir concept through redefined routing units which are different from the atmospheric grids (Guimberteau et al., 2012).

#### 3.2.1 Evapotranspiration simulation

On top of precipitation, evapotranspiration and potential evapotranspiration are the two important fluxes and their errors are key to the water cycle modeling. In ORCHIDEE, the evapotranspiration is calculated with energy balance and resistance concepts. The potential evapotranspiration is defined as "the amount of evapotranspiration that would occur if enough water was available at the surface" which was explained in Barella-Ortiz et al. (2013). The $PET$ is computed as the sum of the potential soil evaporation and the potential transpiration from vegetations. For soil evaporation, the diffusive equations are taken with the ratio of the humidity gradient, the aerodynamic resistance and the air density. The virtual surface temperature is used instead of the actual one to compute the saturate humidity, while the virtual surface temperature is calculated through an unstressed surface-energy balance. The method has been proven superior to other diffusive methods in the reference paper Barella-Ortiz et al. (2013). The potential transpiration is driven by the potential evaporation between the evaporating surface and the overlying air, but limited by vegetation resistances. The maximal water loss under stress-free conditions is the potential transpiration Guimberteau et al. (2012). The actual evapotranspiration is a function of the potential evaporation, but is moduled by a series of resistances (canopy and aerodynamics) of the surface layer. The details of the methods in simulating $PET$ and $ET$ can be found in references D'Orgeval (2006); Guimberteau et al. (2012) and Barella-Ortiz et al. (2013).

#### 3.2.2 Snow and soil freezing scheme

There is on key improvement which have been implemented in the current version of ORCHIDEE, that is, the snow and soil freezing scheme. Snow and soil freezing are two crucial water processes in cold regions, snow covers nearly half of land area (Wang et al., 2013), and the frozen soils occupy 55% to 60% of the land surface of the Northern Hemisphere in winter (Zhang

et al., 2003). Snow plays an important role in both the energy and water flux as the snow cover is first an insulation which prevents the heat loss from the soil. It also increases the thermal inertia of the surface by adding a new phase change and acts as a moisture reservoir which stores winter precipitation that is released in spring or early summer. In the old ORCHIDEE version, a constant density and very simple heat capacity are applied for the snow. The snowmelt directly feeds the runoff without refreezing, and the snow layer is mixed with the first soil layer so that they are equal in temperature. While in the new snow scheme, the snow layer is defined and separated from the soil layers. The snowpack is represented in three layers which adequately resolve the snow thermal gradients between the top and base of the snow cover. The energy balance and the temperature of the snow body become more realistic. Refreezing of the snowmelt is allowed which makes the energy changes more reliable. The snow properties are more detailed than before, i.e., the density, albedo and roughness. All the improvements have been validated over France and Northern Eurasia and already implemented in the current ORCHIDEE (Wang et al., 2013).

Soil freezing impedes water infiltration and drainage thus leading to changes in hydrological responses (Woo and Marsh, 2005). At small scales, the soil freezing alters the soil structure and therefore its water capacity, which has a consequence on the water flux between soil and atmosphere, as well as the water availability for plants (Pitman et al., 1999). On the other hand, the frozen soil changes the latent heat exchange, which delays the soil temperature signal (Boike et al., 1998). Soil thermal characteristics are also improved due to the different thermal properties of ice and water. In the old soil thermal equations, thermal advection and phase change of the water are not considered when resolving the latent heat exchanges (Gouttevin et al., 2012). The mechanical effects of soil freezing are therefore ignored. In the new soil freezing scheme, the apparent soil heat capacity can be increased by considering the ice content of the soil layer during a freezing temperature window between 0 $^o$C and -2 $^o$C. A temperature correction is applied if any soil layer is entirely frozen or thawed. Moreover, the soil heat conductivity is changeable according to the ice content in the soil, which affects the thermal propagation in the vertical soil column. Finally, the hydraulic conductivity is reduced as a function of the ice content. Liquid water is not allowed to cross the frozen soil layers. Thus infiltration and drainage are forced to stop. Full descriptions of the new freezing soil equations and the parameterizations setting can be found in Gouttevin et al. (2012).

#### 3.2.3 Human intervention

Irrigation is included in the current version of ORCHIDEE. The irrigation requirement is estimated as the deficit of the available water of the corresponding grid to the potential evapotranspiration. The irrigation extracts water from local grid first and then its neighbouring grids if necessary (Guimberteau et al., 2012). This solution is acceptable for most of the humid regions at a $0.5^o$ resolution since the rivers are very likely existing wintin 100km. Although for dry regions, the Tarim for example, the irrigation area is concentrated, with controlled irrigation infrastructures. The nearby rivers are far from the irrigation area and the irrigated water is not taken directly from the rivers but transported from upstream by channels. Due to the shortcoming of the scheme and the lack of knowledge on the local irrigation, we turned off the irrigation in ORCHIDEE. In the Tarim basin the irrigation accounts for more than 95% of the consumed water (Zhou et al., 2000), so that the difference between the simulated discharge and observations can be attributed to neglecting irrigation in the model.

In conclusion, as shown in Figure 2, four different forcing inputs are prepared to drive the ORCHIDEE simulations, two basic forcing WFD-CRU and WFDEI-CRU, and one after correction by CMA (WFDEI-CMA) and then one after adding glacier melt (WFDEI-CCG). Among them, the input water amounts are different, while WFDEI-CMA and WFDEI-CCG have the same other forcing variables as WFDEI-CRU. Additionally, the experiment with new developed snow and soil freezing scheme is named WFDEI-CCG-SF based on the forcing dataset WFDEI-CCG. Because the WFD covers the overlapping period 1901 to 2002, the WFDEI covers the period from 1979 to 2014, CMA covers years from 1951 to 2016 and the glacier melt dataset covers from 1958 to 2001, all the simulations and analysis in this paper are over the overlap period from 1979 to 2001. The monthly discharge measurements for those chosen hydrological gauges over 1979-2001 is then compiled. Spatial resolution for all the forcing inputs remains 0.5 degree (about 50km at the equator) and time step as 3 hours.

### 3.3 Main biases over the headwater catchments

In the water cycle, all the water entering a specific river basin will become evapotranspiration ($ET$) back to the atmosphere and discharge ($R$) flowing out of the basin as long-term water storage changes are negligible. The underestimation of discharge is thus either attributable to the underestimation of water inputs or overestimation of evapotranspiration. The possible biases from water inputs and the $ET$ estimation are discussed in this section.

#### 3.3.1 Bias in precipitation

The Tarim basin is one of the areas where significant deviations exist among different modeled precipitation estimates and observation-derived datasets because of the sparse observations and orographic effects (Fekete et al., 2004; Wu et al., 2012). Meanwhile, the precipitation bias might be not well addressed in the CMA system because the number of meteorological gauges is still limited in the Tarim region to build reliable interpolation climate field (Shen et al., 2010). Although Xie et al. (2007) have tried to use other gauges outside China, the density of the gauge distribution is extremely low around the boundaries of the Tarim basin where most of the precipitation is generated. Furthermore, due to the orographic effects, the precipitation over the mountainous area is larger and with more significant heterogeneity than that in the plains (Daly et al., 2008; Barry and Chorley, 2009), while the center of the precipitation events is hard to be observed by the nearby gauges. Adam et al. (2006) has pointed out that the orography could cause 41.6% underestimation of the precipitation over the northwestern American mountainous ranges, and the deviations are larger at higher altitude.

#### 3.3.2 Bias in rainfall and snowfall repartition

The differences between WFDEI-CRU and WFDEI-CMA are not only in the total amount of precipitation but also in the proportion of rainfall and snowfall (Table 3). Compared to the rainfall, the snowfall is more difficult to observe and affected by a large uncertainty, hence in the CMA dataset, only the rainfall was recorded and then used to scale the CRU dataset but keeping the relative proportion of liquid and solid precipitation provided by WFDEI-CRU. The energy needs for phrases change is considerably different for the liquid from the solid water. Berghuijs et al. (2014) suggested snow will lead to more

runoff than rain in similar conditions based on the observations over the US and China. The impact of the precipitation type to the evapotranspiration rate is affected by many factors and hard to measure.

#### 3.3.3 Bias in glaciers melt

HYOGA2 is a state of the art global glacier model but which has not been calibrated over the Tarim basin. The general bias for global estimation is around -5.0% compared to the available global glacier mass balance measurements (Hirabayashi et al., 2013). While the estimated annual glacier melt amount is $81.0\times10^8\mathrm{m}^3/\mathrm{yr}$ for the whole Tarim basin (Table 3), significantly lower than previous estimations (Yang, 1991, $133.4\times10^8\mathrm{m}^3/\mathrm{yr}$; Gao et al., 2010, $144.16\times10^8\mathrm{m}^3/\mathrm{yr}$). On the one hand, the difference in the forcing which drives the glacier melting model is probably one of the causes of the deviation. On the other hand, all the glacier melt is evenly distributed in a whole grid. It leads to a higher infiltration ratio and thus feed more evaporation (Berger and Entekhabi, 2001; Potter et al., 2005). This also artificially forces part of the glacier melt flowing out of the grid not belonging to the right basin. Although, it is unable to eliminate this problem with the current gridded concepts. Finer spatial resolution in glacier dataset and model simulation are needed to lessen the impacts of discretization.

#### 3.3.4 Bias in potential evapotranspiration estimation

As described in subsection 3.2.1, the $PET$ estimation is independent on underlying conditions (e.g. topography, vegetation) because enough water is provided. It is therefore determined only by forcing conditions, especially the humidity gradient and aerodynamic conditions (e.g. radiation flux, wind). Temperature also plays a role in its estimation. Thus, the bias in $PET$ is mainly propagated from various forcing variables.

#### 3.3.5 Bias in actual evapotranspiration estimation

Overestimation (underestimation) of the actual evapotranspiration ($ET$) will also result in the discharge underestimation (overestimation). Many processes can cause $ET$ errors either by the biases in $PET$ or the stress functions which limit the potential evaporation. The vegetation fraction, vegetation type, surface slope and soil properties are all the uncertain sources affecting final $ET$ estimation.

#### 3.3.6 Bias sources category

With the main biases listed as above, we consider bias in any processes that changes $P$ or $PET$ as bias from forcing and bias in any processes that directly changes $ET$ as bias caused by model structure. Although the shifts in forcing variables will change the $ET$ estimation, for example, the $P$ restricts the available water for $ET$, this shift is still belonging to forcing category since the relation is indirect. The biases which directly affects $ET$ includes biases in infiltration, soil water movement, snow processes, vegetation representation and many other model processes. And all these are considered as biases caused by model structures.

### 3.4 Budyko hypothesis

Budyko hypothesis is an empirical expression for the coupling of the water and energy balances at the surface. It uses the relation between the water and energy balance equation to partition precipitation ($P$) into evapotranspiration ($ET$) and runoff ($R$). The Budyko curve is the analytical solution to the Budyko hypothesis, expressed as that the evapotranspiration rate ($ET/P$) is a function of the aridity index ($PET/P$) (Budyko, 1974). Many forms of Budyko curves have been developed and can be categorized into non-parameter group and parameter group depending on whether there is an adjustable parameter describing the Budyko shape (Table 1).

The forms without parameters (Formula 1 to 4) are universal for most of the basins, while they are unable to capture the various landscape characteristics across regions (Yang et al., 2007). Regarding the effects of landscape characteristics, adjustable parameters and corresponding formulas were introduced as formula 5 to 8. Although they have different analytical expressions, the shape of these curves is quite similar (Gerrits et al., 2009) and their parameters are highly correlated (Yang et al., 2008). Hence, from the formulas with parameters, the Fu's equation (Formula 6) is chosen in this study as it is more often used in China region.

The ranges of the aridity index $PET/P$ correspond to the regional precipitation feeds and climate types (Table 2). For example, the precipitation for a semiarid region ranges from 400mm/yr to 800mm/y, and the regional aridity index mostly ranges from 2 to 5. Moreover, the Budyko curve is a reflection of the landscape characteristics which can influence the water movement through different hydrological cycle (Dingman, 2015) and thus changes the $ET$ rate. Many surface conditions are related to the Budyko parameter setting. (1) Vegetation. Transpiration accounts for about 42% (25%-64% depends on different models) of the $ET$ over globe (Zhang et al., 2004). Regions with a larger fraction of vegetation cover or covered by vegetation with bigger leaf and deeper roots tend to have larger transpiration rate as well as the $ET$ rate, i.e., the forested catchment tends to show higher evaporation ratio than the grass covered catchments (Zhang et al., 2004; Carmona et al., 2014). (2) Properties of soil determine infiltration rates and the amount of evapotranspirable water. Steeper slopes are more likely to shed surface water as runoff (Yang et al., 2007; Yu et al., 2014). Limit of infiltration ability also matters as intense precipitation rate (Berger and Entekhabi, 2001; Potter et al., 2005) or freezing soils tend to force the water into the surface runoff (the solid frozen soil limits the percolation of infiltrated water, Gouttevin et al., 2012). (3) Ability to transmit or retain infiltrated water, the soil with larger water conductivity is likely to release more subsurface water rather than evaporation (Yang et al., 2007). (4) The soil depth determines the ability to store infiltrated water. Rocky mountains or regions with thin soil would produce more runoff and less $ET$ (Yang et al., 2008; Dingman, 2015).

### 3.5 Bias ~~sources and evaluation ways~~assessment with ORCHIDEE-Budyko framework

The Budyko formulation relates the $ET$ to $PET$ and $P$. Both the biases from the water flux ($P$) and energy flux ($PET$) will propagate to $ET$. The shape parameter of Budyko curve is obtained by fitting the $PET/P$ and $ET/P$ relation, it is thus a reflection of the model if we use the simulated $PET$ and $ET$ fluxes ~~.~~(i.e. ORCHIDEE simulations in this study, red dots and red curve in Figure 3) . However, because of the existing bias in all the three variables, the relations of the three

components may ~~shift~~ have been shifted to an unrealistic state (point A in Figure 3). Therefore, the changes ~~of~~ either in $P$, $PET$ or $ET$ which can shift the system back to a reliable state are considered as the possible bias. The ~~maximum range of these biases for each variable can be provided through the framework. The~~ difference between the unrealistic state with their ~~true~~ corrected values provides the estimation of how the forcing or the model would need to be changed ~~to~~ for model to produce the realistic discharge values. To separate the individual effect of the single water-energy component on the hydrologic cycle, three independent assumptions are made as follows and the illustration can be found in Figure 3:

The red dots in Figure 3 represent the states of the $PET/P$ and $ET/P$ according to ORCHIDEE estimations and forcing inputs for each year. Point A represents the representative state which is the average of the dots locations. It reflects the current model and is probably in an unrealistic state because the modeled discharge ($P - ET$) may be with bias compared to the observations.

Assumption 1. Only the water input ($P$) is uncertain. Becuase the model structure remains unchanged, the relation between $ET/P$ and $PET/P$ is still following the original Budyko line regardless of how $P$ changes. The $PET$ is assumed to be independent of $P$, while the $ET$ is modified as a result of $P$ changes. To meet the deviation between simulated discharge ($RS$) and the observed discharge ($RO$), the representative point (long-term "corrected" annual $ET/P$ against long-term "corrected" annual $PET/P$) should be shifted along the Budyko curve from current point "A" to new point "B" where the difference between the "true" precipitation ($P'$) and the "true" evapotranspiration ($ET'$) equals to the observed discharge ($P' - ET' = RO$). The possible maximum bias in $P$ is calculated as

$$Bias(P) = (P - P')/P' \times 100\%. \tag{1}$$

Assumption 2. Only the $PET$ is uncertain. The $P$ remains the same, while $ET$ is changing because of the changes in $PET$. Under these conditions, the model structure still remains unchanged and so does the Budyko curve. Then the representative point should be shifted along the Budyko curve to point "C" to decrease the $ET$ ratio to meet the discharge observation ($ET' = P - RO$). The $PET$ is changed to a 'true' $PET'$ and the possible maximum bias in $PET$ is calculated as

$$Bias(PET) = (PET - PET')/PET' \times 100\%. \tag{2}$$

Assumption 3. Only the $ET$ is uncertain. $P$ and $PET$ which are mainly linked to the forcing remain the same, while the $ET$ which is significantly affected by the model structure is assumed biased. It is essentially relevant to the model processes rather than the forcing dataset. To compensate the discharge bias, the $ET$ should be decreased to point "D" where the $ET$ equals precipitation minus observed discharge ($ET' = P - RO$). The possible maximum bias in $ET$ is calculated as

$$Bias(ET) = (ET - ET')/ET' \times 100\%. \tag{3}$$

With the target $ET'$, a new Budyko curve can be drawn for new relations between $P$, $PET$ and the new $ET$ (the blue lines in Figure 3). However, all the assumptions are proposed in conditions that only one variable is uncertain, but in reality, any of the three variables can be biased at the same time. The final probable "corrected" state may locate in the shaded area identified by the three states (Figure 3).

## 4 Results and discussion

### 4.1 Forcing and discharge comparison

#### 4.1.1 Forcing inputs comparison among experiments

We specify precipitation ($P$) as the sum of all the water inputs into the system, which include the atmospheric water flux (in its liquid and solid phases) and glacier melt. The precipitation for the three headwater catchments and the upper Tarim are listed in Table 3 for each simulation. The inter-annual variations and the intra-annual cycle of total precipitation over upper Tarim basin are plotted for different forcing in Figure 4.

The annual cycle of the two basic forcing datasets WFD-CRU and WFDEI-CRU are similar while the precipitation in WFDEI-CRU is slightly larger in monthly values (Figure 4-a, red and blue lines). The deviation is the result of their expanding differences after 1990 (Figure 4-b). The precipitation difference is mainly due to that two different versions of CRU (CRU TS 2.10 and CRU TS 3.1) are used for precipitation correction in WFD and WFDEI (See subsection 3.1.2 or Weedon et al., 2014). The precipitation difference between the two CRU dataset is ~~slightly different~~relatively small, while CMA dataset increases the precipitation to a large extent by 37.3% compared to CRU for the Upper Tarim basin (Table 3). The changes are mostly occurring during summer when the peak of CMA precipitation is more than twice as large as precipitation in CRU. The shape of the annual cycle has strongly changed as the timing of the peak is shifted from April to July. However, the precipitation amount in winter (DJF) decreases in CMA, to which the decreasing in snowfall is the major contributor. The changes in rainfall and snowfall are similar in all three headwater catchments (Table 3) and the Upper Tarim basin.

Adding the glacier melt leads to negligible changes during winter and spring, but large increases in the total water inputs in summer (JJA) when the temperature is higher. The estimated glacier melt is 9.1%, 25.8% and 6.1% to the total water inputs for the upper Yarkand, upper Hotan and upper Aksu. It is significantly increasing after 1990 (Figure 5, the trend is $+3.8 \times 10^8 \mathrm{m}^3/\mathrm{yr}^2$, $p = 0.024$), being consistent with its ratio to the total water input ($r = 0.918$, $p < 0.001$). The trend is mainly caused by the climate warming as the glacier melt is highly correlated with the summer temperature ($r = 0.852$, $p < 0.001$). The increasing trend has also been documented in glacier runoff observations (Shangguan et al., 2009).

#### 4.1.2 Assessment of the discharge estimations with observations

Evaluating the bias in precipitation over meteorological rain gauges is not convincing as most gauges are located in lower altitude, which makes it difficult to capture regional patterns as intensive precipitation occurs over higher mountains. Instead, the discharge measurement can serve as a better reference since it integrates the net water flux over the entire basin. Therefore, driven by the forcing, ORCHIDEE was used to simulate the river discharge and compare to *in-situ* observations (Figure 6). The corresponding assessment using criteria for the three headwater catchments are plotted in Figure 7. For the three headwater catchments where most of the discharge of the Tarim basin is generated, the discharge is significantly underestimated, with the underestimation ratio reaching 90% (Figure 6-b,d,f; Figure 7-a) for CRU datasets. Discharge increases after the precipitation are corrected by CMA dataset, with the absolute bias decreasing to around 80%. Adding glaciers melt also increases the

discharge but by a relatively small amount. Changes in the model structure (new snow and soil freezing scheme) further decreases the bias, especially for the Aksu. The final biases of the discharge for the three subbasins are -71.1%, -47.8% and -49.4% respectively. The gradual improvements and corresponding magnitude changes are visible in the annual discharge variability in Figure 6-b,d,f.

Besides the increase in the annual mean discharge, the amplitude of the inter-annual cycle of the discharge is also improved by the progressive changes. The estimated discharge peaks have been shifted from April in CRU simulations to the summer (July or August) by CMA correction and adding glaciers (Figure 6-a,c,e). Correspondingly, the correlation of the annual variability between the estimated and observed discharge has increased above 0.9 for all the three subbasins with the WFDEI-CCG forcing (Figure 7-b). However, contrary to the upper Yarkand and the upper Hotan, the introduction of the new snow and soil freezing scheme decreases the discharge correlation for the upper Aksu from 0.91 to 0.42. An early discharge peak exists in May, while not enough runoff is generated in the summer period (Figure 6-e). Although the correlation decreases, it does not mean the model/simulation deteriorates because correlation only evaluates the similarity of temporal variation but ignores the fact that the discharge amount has been better estimated (Figure 6-e and Figure 7-a).

By extracting the observed discharge in the first and last 5 year from the whole period, we can notice there is an obvious shift of the discharge peak from August to July in the three headwater catchments (Figure 6-a,c,e). The regional precipitation changes largely cause the shift, but the increasing temperature also allows the snow/glacier melting at a higher rate in the most recent period. Besides, increasing trends are detected after the 1990s (Figure 6-b,d,f), as the increasing trend is $1.43\times10^8\mathrm{m}^3/\mathrm{yr}^2$ (or $0.77\times10^8\mathrm{m}^3/\mathrm{yr}^2$, $1.78\times10^8\mathrm{m}^3/\mathrm{yr}^2$) for the upper Yarkand (or Hotan, Aksu). The increasing trends are consistent with the glacier melt, glacier proportion in water input as well as the summer temperature in the same period (Figure 5).

The trends in the estimated discharge are also calculated and compared with the observed trends, expressed as the ratio of the trend in estimation to that in observations (Figure 7-c). For CRU simulations, no increasing trend is detected since the ratio is less than 0. The CMA correction increases the ratio for all three subbasins to around 0.3, which means the precipitation accounts for only around 30% of the discharge increase. Adding the glacier melt increases the ratio for the upper Hotan from 0.31 to 0.76, from 0.35 to 0.54 for the upper Aksu, the improvement of the glacier melting is comparable to the CMA correction. Although, no apparent changes are detected for the upper Yarkand. By comparing the criteria between WFDEI-CMA and WFDEI-CCG, we find that although adding the glacier melt does not change much the absolute amount of discharge or the correlation, the increased trend in discharge has been considerably improved. The increasing glaciers melt is, therefore, one of the contributors to the discharge trend in the Tarim. The modification of the snow and soil freezing scheme increases the trend ratio in the upper Aksu up to 0.72, while slightly decreasing it for the two other catchments.

In brief, the gradual refinement of the forcing datasets (from WFD-CRU to WFDEI-CRU, to WFDEI-CMA, to WFDEI-CCG) is effective for improving the model performance using different criteria (bias, correlation, proportion to the trend) to compare the observed discharge. The three criteria are independent as they stand for the averages, the variation and the trend, which can capture the various aspects of the model agreement to the observations. The responses are similar for different catchments, but in different magnitude at different stages. The correction of CMA dataset is the most significant improvement to all the criteria. The role of glaciers melting is critical for the trends analysis. The modification in snow and soil freezing

scheme increases the total discharge amount but could lead to adverse responses in the correlation and trend simulation. However, the impact of the modification of the model structure is not larger than changes resulting from the forcing biases. From the previous analysis, we can conclude that the simulations of WFDEI-CCG and WFDEI-CCG-SF are comparable in the correlation and trend analysis, while WFDEI-CCG-SF is better regarding the water quantity. Therefore, the further study on the bias is all based on the WFDEI-CCG-SF simulation.

### 4.2 Evidence of the bias in estimated discharge for the headwater catchments

Although the simulations with WFDEI-CCG-SF are better than other experiments, there are still biases compared to the observations (Figure 7-a). In this section, we aim to find evidences of the biases. The annual mean water balance components (rainfall, snowfall, glacier melt, estimated $ET$ and discharge $RS$) of the three upper catchments are plotted as bars with their relations quantified by comparing to the discharge observations ($RO$, red lines in Figure 8-a,c,e) in WFDEI-CCG-SF. Their annual cycles are also plotted as Figure 8-b,d,f over the three headwater catchments. Over a long enough period, the changes in terrestrial water storage are assumed negligible compared to the water fluxes, so that the water input into the system either returns the atmosphere through $ET$ or flows out of the basin as $RS$.

From the left panels of the Figure 8, we have a visual impression of the relative amount of different water inputs and their contribution to the $ET$ or discharge. Note that the sum of the $ET$ and $RS$ is not exactly equal to $P$ because in ORCHIDEE the discharge is represented at the outflow of the grid and not at the confluence point of the analyzed catchment with other tributaries. The largest bias is 8% for the upper Yarkand (Figure 8-a, $ET/P + RS/P = 0.92$), while it is exactly matching for the upper Aksu. The bias can be added to the current $RS$ if necessary.

Among the three headwater catchments, the upper Hotan has the best discharge simulation compared to the observations ($RS/RO = 0.52$). The annual cycle of the water also matches well as all the $P$, $ET$ and the discharge $RO$ or $RS$ have the synchronous peaks in summer (Figure 8-d). There are also deviations between $P - E$ which represents the net water inputs to the system and the estimated discharge (the shaded area with blue lines in Figure 8-d). The deviation implies the regional water storage changes, in summer the soil moisture increases to store the abundant water inputs to release them later in autumn and winter when the drainage rate is larger than infiltration. The water storage is decreasing as a result by then. It is the natural adjustment to the strong seasonality in water inputs.

As the neighboring catchment of the upper Hotan, the upper Yarkand has similar phases of estimated fluxes ratios ($ET/PET$, $ET/P$, $RS/P$) and their inter-annual variations. Although, the estimated discharge rate is smaller ($RS/RO = 0.29$) than that of the upper Hotan. Underestimation in water inputs in summer and autumn is possibly the reason as there is not obvious water storage gaining in summer period and the ratio of observed discharge to regional precipitation is unrealistic high (>0.9, Figure 8-b).

The upper Aksu has different characteristics with the other two regions since it lies in the northern part of the Tarim. It has the larger snowfall proportion in the precipitation. Meanwhile, it has the largest ratio of estimated $ET/PET$, the largest runoff generation ratio ($RS/P = 0.35$) and the least discharge simulation error ($RS/RO = 0.51$). Although, there is certainly large bias in the regional precipitation as the discharge has exceeded the precipitation input in summer period (July and August). The

estimated annual cycle of discharge diverges from the observations (Figure 8-f), as its peak advances by two months and the discharge estimation significantly exceeds the observations in spring (MAM). The runoff generation ratio in summer period is also unrealistic low.

In summary, the biases in discharge estimation exist in terms of the total amount and the annual cycle. The precipitation one of the largest bias source which makes the bias analysis in models more difficult.

### 4.3 ~~Bias sources over the headwater catchments~~

~~As mentioned previously, all the water entering a specific river basin will become evapotranspiration ($ET$) back to the atmosphere and discharge ($R$) flowing out of the basin as long-term water storage changes are negligible. The underestimation of discharge is thus either attributable to the underestimation of water inputs or overestimation of evapotranspiration. The possible biases from water inputs and the $ET$ estimation are discussed in this section.~~

#### 4.2.1 ~~Bias in precipitation~~

~~The Tarim basin is one of the areas where significant deviations exist among different modeled precipitation estimates and observation-derived datasets because of the sparse observations and orographic effects (Fekete et al., 2004; Wu et al., 2012). Meanwhile, the precipitation bias is not well addressed in the CMA system because the number of meteorological gauges is still limited in the Tarim region to build reliable interpolation climate field (Shen et al., 2010). Although Xie et al. (2007) have tried to use other gauges outside China, the density of the gauge distribution is extremely low around the boundaries of the Tarim basin where most of the precipitation is generated. Furthermore, due to the orographic effects, the precipitation over the mountainous area is larger and with more significant heterogeneity than that in the plains (Daly et al., 2008; Barry and Chorley, 2009), while the center of the precipitation events is hard to be observed by the nearby gauges. Adam et al. (2006) has pointed out that the orography could cause 41.6% underestimation of the precipitation over the northwestern American mountainous ranges, and the deviations are larger at higher altitude.~~

#### 4.2.1 ~~Bias in rainfall and snowfall repartition~~

~~The differences between WFDEI-CRU and WFDEI-CMA are not only in the total amount of precipitation but also in the proportion of rainfall and snowfall (Table 3). Compared to the rainfall, the snowfall is more difficult to observe and affected by a large uncertainty, hence in the CMA dataset, and only the rainfall was recorded and then used to scale the CRU dataset but keeping the relative proportion of liquid and solid precipitation provided by WFDEI-CRU. The energy needs for phrases change is considerably different for the liquid from the solid water. Berghuijs et al. (2014) suggested snow will lead to more runoff than rain in similar conditions based on the observations over the US and China. The impact of the precipitation type to the evapotranspiration rate is affected by many factors and hard to measure. However, the mismatch of the discharge peaks for the upper Aksu is very likely the result of the bias in rainfall and snowfall repartition.~~

#### 4.2.1 ~~Bias in glaciers melt~~

~~HYOGA2 is a state of the art global glacier model but which has not been calibrated over the Tarim basin. The general bias for global estimation is around -5.0% compared to the available global glacier mass balance measurements (Hirabayashi et al., 2013). While the estimated annual glacier melt amount is $81.0\times10^8 \mathrm{m}^3/\mathrm{yr}$ for the whole Tarim basin, significantly lower than previous estimations (Yang, 1991, $133.4\times10^8 \mathrm{m}^3/\mathrm{yr}$; Gao et al., 2010, $144.16\times10^8 \mathrm{m}^3/\mathrm{yr}$). On the one hand, the difference in the forcing which drives the glacier melting model is probably one of the causes of the deviation. On the other hand, all the glacier melt is evenly distributed in a whole grid. It leads to a higher infiltration ratio and thus feed more evaporation (Berger and Entekhabi, 2001; Potter et al., 2005). This also artificially forces part of the glacier melt flowing out of the grid not belonging to the right basin. Although, it is unable to eliminate this problem with the current gridded concepts. Finer spatial resolution in glacier dataset and model simulation are needed to lessen the impacts of discretization.~~

~~The glaciers melt in the upper Yarkand is possibly underestimated because the glacier retreat rate in Yarkand is more severe than other regions according to the in-situ observations and different model estimations (Liu et al., 2006; Shangguan et al., 2009). Either the glacier retreat area, the relative area or the estimated and observed glacier runoff in the upper Yarkand is larger than that of the Hotan. The estimated glacier melt obtained from HYOGA2 in Yarkand is 57.3% less than that in Hotan subbasin. As a consequence, the trend changes after adding glacier melt in the forcing dataset is not significant for the upper Yarkand basin (Figure 6-c, with cycle marker). The underestimation of the water input for the upper Aksu in summer can also be probably attributed by the glacier because of its extremely low glacier proportion. However, the underestimation of glacier melt is not the only cause of the final discharge in the two subbasins. Because assuming the bias of glacier melt is 60% (estimated of the glacier to estimation in Yang, 1991 or Gao et al., 2010), the increasing water cannot fill the gap between $RS$ and $RO$ even without consideration of any lose to evaporation (Figure 8-a,e).~~

#### 4.2.1 ~~Bias in evapotranspiration estimation~~

~~Overestimation of the actual evapotranspiration ($ET$) would also result in the discharge underestimation. Many processes can cause $ET$ errors either by the biases in $PET$ or the stress functions which limit the potential evaporation. Radiation fluxes are the major factors in $PET$ estimation, and it is affected by the considerable uncertainty of the energy variables over heterogeneous mountains just as precipitation.~~

~~The Budyko formulation relates the $ET$ to $PET$ and $P$. Thus both the biases from the water flux ($P$) and energy flux ($PET$) will propagate to $ET$. Moreover, as mentioned in Sect. 3.3, the Budyko shape parameter is determined by the regional surface conditions. The vegetation fraction, vegetation type, surface slope and soil properties are all the uncertain sources affecting final $ET$ estimation. However, quantifying the biases is difficult without sufficient measurements and finer gridded datasets.~~

### 4.3 Bias range and possibility analysis for the headwater catchments

Although either $P$ or $PET$ or $ET$ can cause final underestimation of discharge over the three headwater catchments, quantifying the bias ~~from~~ in each flux is challenging and impractical due to the lack of direct measurements and the strong het-

erogeneity over the mountainous area. ~~Separating the individual influence is further complicated by the interactions among the three variables if more than one variable is uncertain.~~ To separate the individual bias, we use the Budyko hypothesis by assuming only one variable is uncertain, while the other two are assumed to have negligible errors, with which we "correct" the model simulation to meet the discharge observation and obtain the possible bias range. Then we evaluate the possibility of rejecting the assumption to find the most likely bias source by checking the status of the water system (i.e., amount of the water-energy components and their relations) in indirect ways. The water-energy components used in the Budyko analysis are all ORCHIDEE outputs of the most satisfactory simulation WFDEI-CCG-SF. The corresponding characteristics of the water-energy components for three headwater catchments are listed in Table 4.

#### 4.3.1 ~~Budyko assumption and bias~~ Bias ranges estimated by the ORCHIDEE-Budyko framework

The ORCHIDEE estimated evapotranspiration rate $ET/P$ against the estimated aridity index $PET/P$ over each subbasin in each year are scattered as red points in Figure 9. The point A represents the Budyko relation between long-term average annual $ET/P$ and the long-term average annual $PET/P$. According to the categories introduced by Ponce et al. (2000), all the three ~~subbasins~~ catchments belong to semiarid climate zones by the definition of the annual average precipitation (Table 2). Hence the aridity index is supposed to be ranging from 2 to 5. Regarding the high elevation and cold temperature, the $PET$ rate is likely to be smaller than the representative climate of this aridity index. Thus the aridity index for the three catchments supposes to be lower than expected. It is realistic for the upper Hotan and the upper Aksu as their aridity is 3.63 and 1.73 respectively. While the aridity index for the upper Yarkand is 5.02 which can be categorized as a semiarid or arid region. It is not very likely since the upper Yarkand is providing water resources for the irrigated area over the lower Yarkand oases (Zhou et al., 2000; Mamitimin et al., 2014).

~~To separate the individual effect of the single water-energy component on the hydrologic cycle, three independent assumptions are made as follows:~~

~~Assumption 1. Only the water input ($P$) is uncertain, the relation between $ET/P$ and $PET/P$ is still following the optimal Budyko line regardless of the $P$ changes as the surface conditions, and thus model structure remains unchanged. The $PET$ is assumed to be independent of $P$, while the $ET$ is modified as a result of $P$ changes. To meet the deviation between simulated discharge ($RS$) and the observed discharge ($RO$), the representative point (long-term~~ However, because there is still bias in the ORCHIDEE discharge estimations with the observations, the current state A is not correct. Based on the assumptions introduced in subsection 3.5, the three possible "corrected" ~~annual $ET/P$ against long-term "corrected" annual $PET/P$)should be shifted along the Budyko curve from current point "A" to new point "B" where the difference between the "true" precipitation ($P'$) and the "true" evapotranspiration ($ET'$) equals to the observed discharge ($P' - ET' = RO$).~~

~~Assumption 2. Only the~~ states by shifting the $P$, $PET$~~is uncertain. The~~, $ET$ respectively are shown in Figure 9 for the three headwater catchments. Taking the upper Yarkand as an example (Figure 9-a), if consider only $P$ is biased (assumption 1), the $P$ has to be shifted from 247.3 mm/yr to 435.4 mm/yr.$PET$ remains the same, ~~while~~ and the $ET$ is changing ~~because of the changes in $PET$. Under these conditions, the model structure still remains unchanged and so does~~ accordingly but the relation

between $ET/P$ and $PET/P$ still follows the Budyko curve. ~~Then the representative point should be shifted along the Budyko curve to point "C" to decrease the $ET$ ratio to meet the discharge observation ($ET' = P - RO$).~~

~~Assumption 3. Only the $ET$ is uncertain.~~ The state is shifted from point A to point B with the $P$ ~~and $PET$ which are mainly linked to the forcing remain the same, while the $ET$ which is significantly affected by the model structure is assumed biased. It is essentially relevant to the model processes rather than the forcing dataset. To compensate the discharge bias, the $ET$ should be decreased to point "D" where the $ET$ equals precipitation minus observed discharge ($ET' = P - RO$). With the target $ET'$, a new Budyko curve can be drawn for new relations between~~ change ratio as 76.1%. Reversely, the possible bias of $P$ ~~,~~ is -43.2% ((247.3-435.4)/435.4*100%). Similarly, to change the $PET$ ~~and the new~~ in order to shift the state to correct (point C, assumption 2), the $PET$ has to be shifted from 1240.4 mm/yr to 225.0 mm/yr. The possible bias in $PET$ is 451.2%. To change the $ET$ ~~(the blue lines in Figure 9). Corresponding changes in the water-energy components can be found in Table 5, in which the bold values are the main changes among the basic water-energy components ($P$~~ in order to shift the state to correct (point D, assumption 3), ~~$PET$ and~~ the $ET$ ~~).~~

~~The corresponding changes ratio of the variables is listed in the last column of~~ has to be shifted from 188.3 mm/yr to 106.9 mm/yr, and the possible bias in $ET$ is 76.1%. The route is same for the other two catchments and their results are listed in Table 5. ~~However~~

Though, the previous analysis is based on the ~~assumption~~ assumptions that the *P* and *PET* is independent and only a single variable is uncertain which might be invalid in reality. However, the three assumptions provide the bias boundaries of each variable and the final system reproducing observed $RO$ is likely to be located within the shaded area shown in Figure 9. Taking the Hotan as an example, to meet the discharge observation, the final changes in $P$ will be 0-26.6%, and decreasing in $PET$ will be 0-46.7%, and the decreasing in $ET$ is ranging 0-16.1%. While in turn, we also conclude that the $P$ is underestimated by 0-21.0%, the $PET$ is overestimated by 87.5% at most and the $ET$ is overestimated by 19.1% at most. If we know the bias for any single variable, the feasible ranges will be narrower than the present.

### 4.3.2 Ranking the bias possibility

Although the Budyko approach provides us with possible ranges for the bias of each variable, it is still difficult to determine the bias source without proper bias measurements for each of the forcing variables. We alternatively compare the regional diagnostics with nearby regions or regions with similar climatic and regional characteristics which have qualified observations in an indirectly way. With which, we can generally rank the occurrence possibility of an uncertain variable.

From multiple model analysis based on CMIP5-GCMs, the estimated $PET$ over the western boundary of the Tarim basin is about the same with the $PET$ over the Tibet Plateau (Scheff and Frierson, 2015). Although, because of the unpredictable biases in GCMs, the absolute values are not highly reliable in their simulation. However, the equivalent relation provides us the ranges of $PET$ over the Tarim headwater catchments by referring to the observations over the Tibet Plateau, where the topography changes are relatively small, and the observations are more abundant. According to Chen et al. (2006) who used site observations from 101 stations over the Tibet Plateau, the annual average $PET$ over the plateau ranges from 580-720mm/yr. Hence the $PET$ over the upper Yarkand and the upper Hotan is probably overestimated (1240.4mm/yr and 1153.7mm/yr).

Therefore, only changes in $P$ or $EP$ are not satisfactory because the $PET$ is unchanged. While for the upper Yarkand, the $PET$ is not the only error source, because to match the discharge deviation by only decreasing $PET$, the $PET$ should be decreased to 225mm/yr, which exceeds the referenced $PET$ range. Moreover, because Yarkand and Hotan are two neighboring regions which have similar climates, the $PET$ should be in the similar amount (615.2mm/yr if only $PET$ is uncertain in upper Hotan, while 225mm/yr in the Yarkand). The estimated $PET$ over the Aksu is realistic since the $PET$ is 631.8mm/yr for the current scenario, but it would decrease to 174.1mm/yr if only the $PET$ is changing, which is too low. Besides the absolute value of $PET$, the ratio $PET/P$ also shifts when $PET$ is changing, which means the climatic types can be changed. By only decreasing the $PET$ in the upper Yarkand and the upper Aksu would cause significant decreases in the aridity index (from 5.02 to 0.91 for the upper Yarkand, from 1.73 to 0.48 for the upper Aksu), which are not realistic for these regions.

$ET$ computation is sensitive to the climatic conditions and the surface conditions, hence the absolute value of $ET$ significantly varies in time and space and its bias very difficult to quantify. While, the evapotranspiration ratio to precipitation ($ET/P$) is typical for specific climatic types or regions with similar land cover types (Yang et al., 2007). Liu et al. (2003) estimated the evapotranspiration ratio to precipitation using remote sensing approach over Canada regions. They concluded that the ratio $ET/P$ is 32% for barren land and 18% for snow/ice land. In general, most of the catchment area of the three headwater catchments consists with barren and snow/ice land. While because of its lower latitude, the $ET$ ratio could be higher but still below the rate for cropland (67%). Therefore, only changes in $P$ is not very likely for Yarkand and Hotan ($ET/P$ is 0.68 for Yarkand and 0.70 for Hotan after the correction). While higher $P$ for the upper Aksu is likely to maintain a realistic $ET/P$ ratio.

The biases of the three variables ($P$, $ET$, $PET$) have relatively weak dependence because they are governed by different processes. $P$ and $PET$ are quite independent because they relate to different forcing variables. Although the $ET$ amount is linked to two other variables, the $ET$ bias is weakly dependent and it also comes more from the surface conditions and model biases. The chances of biases arising from each variable are about the same in theory. However, based on the analysis of the model output and the assumed bias corrected scenarios, there are some priorities for the bias sources over each subbasin. The possibility of biases and the supporting arguments are listed in Table 6 for each headwater catchment. For instance, for the upper Yarkand subbasin, increasing in $P$, especially the glacier melt is necessary because the lower glacier melt ratio compared to the Hotan basin (Sect. 5.1.2.3) and the small trend in model discharge compared to the discharge observation (Sect. 4.2). However, only increasing the precipitation is not sufficient because the current $PET$ is very likely too high compared to the surrounded regions ($PET$=1240.4mm/yr, while $PET$ ranges 580-720mm/yr over the Tibet Plateau). Meanwhile, only decreasing the $PET$ without changing others would cause very low $PET$ rate (225mm/yr) and low aridity index ($PET/P$=0.91) which is not realistic for this region. Modification in $ET$ is possible but not adequate due to the overestimated $PET$. Hence, this error analysis reveals that increasing precipitation over the upper Yarkand subbasin is quite necessary, the overestimation of $PET$ is very likely, while modification in $ET$ estimation is possible but not fundamentally necessary. For the upper Hotan, the most likely biases come from the overestimation of $PET$, while the two other variables are possible. Increasing precipitation and changes in temporal variability are necessary for the upper Aksu subbasin, while it is not the only reason, as either the $PET$ or $ET$ or both are overestimated.

### 4.4 Human intervention in the lower oases

The current ORCHIDEE version does not yet take into account the intensified evapotranspiration caused by human activities especially through irrigation, which is a major process in hydrological cycle transferring water to the atmosphere. As a consequence, in the lower oases the simulated discharge at Alar station is significantly larger than the observations (Figure 6-g,h; $146.59\times10^8\mathrm{m}^3/\mathrm{yr}$ in WFDEI-CCG-SF simulation to $43.34\times10^8\mathrm{m}^3/\mathrm{yr}$ in observation). However, because the biases of the upstream discharge will propagate to the Alar station, the catchment of which includes the three basins discussed above, the currently estimated discharge is thus underestimated compared to the potential river flow at Alar which is the natural river flow without human intervention.

We use two simple approaches to estimate this underestimation. The first one, according to work of Tao et al. (2011), all the water increment of the gauge Alar is caused by the water changes from the three headwater catchments. Hence the underestimation of the discharge to the potential river flow at Alar equals to the underestimation of discharge from those three catchments. The increment at Alar should be $110.25\times10^8\mathrm{m}^3/\mathrm{yr}$ (observed $192.15\times10^8\mathrm{m}^3/\mathrm{yr}$ from the three headwater catchments minus simulated $81.9\times10^8\mathrm{m}^3/\mathrm{yr}$). So that the potential river flow at Alar should be $256.84\times10^8\mathrm{m}^3/\mathrm{yr}$ (=$146.59+110.25$), and the influence of human activities on the increasing of $ET$ can be estimated as $213.50\times10^8\mathrm{m}^3/\mathrm{yr}$ ($=256.84-43.34$), 83.1% of discharge.

A second simple scaling approach is applied for the discharge at Alar to obtain the potential river flow that we assume the model bias for the whole upper Tarim basin is constant over space. The scaling factor is 2.35 ($=192.15/81.9$), hence the potential discharge at Alar should be $343.92\times10^8\mathrm{m}^3/\mathrm{yr}$ (=$146.59\times2.35$). And the influence of human activities on the increase of $ET$ is estimated as $300.58\times10^8\mathrm{m}^3/\mathrm{yr}(=343.92-43.34)$, 87.4% of the discharge. The overestimation over the discharge observation is the amount caused by additional human intervention, especially the irrigation-caused evapotranspiration.

To validate the proposed values, we collected the irrigation area over the Tarim basin using FAO global Map of Irrigation Areas (Siebert et al., 2013). According to which, the total irrigated area for the Upper Tarim basin is 13548.5 $\mathrm{km}^2$. In addition, Zhou et al. (2000) provides the gross irrigation quota as $1.77\times10^6\mathrm{m}^3/\mathrm{km}^2$, hence the total irrigation water consumption will be $239.80\times10^8\mathrm{m}^3/\mathrm{yr}$. Hence the results of the two approaches assessing human net abstraction are -11% and 25% ($213.50\times10^8\mathrm{m}^3/\mathrm{yr}$ and $300.58\times10^8\mathrm{m}^3/\mathrm{yr}$ in two approaches) respectively to the consumption data, which are acceptable because of the unknown biases in irrigation area as well as the gross irrigation quota. The proportion of the consumed water (83.1%-87.4%) is higher than the estimation (74.7%) in 1995 which could be explained by the intra-annual variation of inflow and abstraction.

## 5 Conclusions

In this work, we ~~have progressively improved the atmospheric forcing data derived from the ERA and ERA-Interim reanalysis, by correcting with CMA regional dataset and by adding glacier melt estimation. The snow and soil freezing scheme in ORCHIDEE was also modified for the cold areas, and then the result sensitivity to the forcing changes~~ proposed an ORCHIDEE-Budyko framework which is used to attribute the modeled discharge bias to different sources as the forcing and model structure~~was~~

analyzed by comparing with *in-situ* discharge observations. Results show that the progressive refinement of the forcing datasets and model structure improves the model performance regarding the absolute bias , correlation coefficient and the discharge trend estimation . We note that the improvements resulting from the forcing data changes are more significant than those resulting from model modifications.

Bias in the precipitation ($P$) and any processes related to the potential evapotranspiration ($PET$) is considered as bias from forcing and bias in any processes affecting the actual evapotranspiration ($ET$) estimation is considered as bias from model structures. The discharge simulation was provided by the land surface model-ORCHIDEE with latest developments in its modules and driven by the most improved forcing inputs (WFDEI-CCG-SF). However, underestimation in the discharge still exists for the best simulation (WFDEI-CCG-SF) over the three headwater catchments. The possible causes Tarim headwater catchments, where the biases of $P$ underestimation, $PET$ and $ET$ overestimation are discussed and their contributions estimated using a Budyko approach. The approach works with are analyzed with a Budyko analytical approach. With a set of assumptionswhich allow to isolate the impact of different sources of bias. Information , we isolated the biases in three variables and their possibilities were assessed with information from nearby and hydro-climatic similar regionsare used to overcome the limited set of observations in the Tarim and infer the possible magnitude of error sources and rank their importance. Results show that precipitation (here considered as the sum of rainfall, snowfall and glacier melt) underestimation is highly likely for the upper Yarkand and the upper Aksu, while the overestimation of $PET$ is likely to affect the upper Yarkand and the upper Hotan. The overestimation in $ET$ is a possible but not likely the only cause for the discharge underestimation for all headwater catchments. In the lower oases, humans consume 83.1%-87.4% of the discharge for irrigation usage which is also a bias source in the current version of model. Thus, including detailed human modules is in need for any large scale model.

In this attempt to analyze the performance of a complex land surface model over the Tarim basin, large biases are found in the discharge estimation. Our finding that the bias is most likely caused by the forcing variables rather than the model is probably the reason for the failures of other models in specific regions as well. Our work provides more information about the Tarim basin's water cycle and guidance for future studies that the bias in forcing variables should firstly be assessed and reduced in order to perform pertinent analysis of the regional water cycle. Land surface models are a recommended tool for water cycle analysis because of their independence of calibration and good ability to simulate most variables of the water cycle and their interplay which facilitates the identification of bias sources. This kind of application along with the improvements of forcing data are also important for predicting water resources in the Tarim as well as other high-altitude basins in central Asia in a changing climate.

*Code and data availability.* The land surface model is free to download through website http://forge.ipsl.jussieu.fr/orchidee/. The WFD and WFDEI datasat can be obtained from http:// www.eu-watch.org. The CRU datasets are available through http://data.ceda.ac.uk/badc/cru/data/. The CMA dataset is available through website http://data.cma.cn/

*Author contributions.* J. Polcher initialized the ideas presented in this paper with X. Zhou and T. Yang. X. Zhou prepared the simulations, the figures and the manuscript. T. Yang and Y. Hirabayashi participated in the data preparing. T. Nguyen-Quang participated in the model development. All authors contributed to the discussion and revising the paper.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The work was jointly supported by the International Co-Foundation NSFC-ANR(NSFC-41561134016, ANR-15-CE01-0011), the National Natural Science Foundation of China (Grants 41371051, 51421006), the Chinese Academy of Sciences (Grant KZZD-EW-12), the Fundamental Research Funds for the Central Universities (Grant 2015B31214), the Global Environmental Research Fund (S-14) of the Ministry of the Environment, Japan and the China Scholarship Council (CSC, 201506710042). The work was supported by computing resource of the IPSL ClimServ cluster at École Polytechnique, France. We also thank the editor and two anonymous reviewers for their contribution to the improvement of this paper.

## References


Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of global precipitation products for orographic effects, J. Clim., 19, 15–38, https://doi.org/10.1175/JCLI3604.1, 2006.

Alkama, R., Kageyama, M., and Ramstein, G.: Relative contributions of climate change, stomatal closure, and leaf area index changes to 20th and 21st century runoff change: A modelling approach using the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) land surface model, J. Geophys. Res. Atmos., 115, D17 112, https://doi.org/10.1029/2009JD013408, 2010.

Barella-Ortiz, A., Polcher, J., Tuzet, A., and Laval, K.: Potential evaporation estimation through an unstressed surface-energy balance and its sensitivity to climate change, Hydrol. Earth Syst. Sci., 17, 4625–4639, https://doi.org/10.5194/hess-17-4625-2013, 2013.

Barry, R. G. and Chorley, R. J.: Atmosphere, Weather and Climate, CUP Archive, https://doi.org/10.4324/9780203871027, 2009.

Berger, K. P. and Entekhabi, D.: Basin hydrologic response relations to distributed physiographic descriptors and climate, J. Hydrol., 247, 169–182, https://doi.org/10.1016/S0022-1694(01)00383-3, 2001.

Berghuijs, W. R., Woods, R. a., and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow-supplement, Nat. Clim. Chang., 4, 583–586, https://doi.org/10.1038/NCLIMATE2246, 2014.

Boike, J., Roth, K., and Overduin, P.: Thermal and hydrologic dynamics of the active layer at a continuous permafrost site (Taymyr Peninsula , Siberia), Water Resour. Res., 34, 355–363, https://doi.org/10.1029/97WR03498, 1998.

Budyko, M. I.: Climate and Life, Academic Press, New York, https://doi.org/10.1016/0033-5894(67)90014-2, 1974.

Carmona, A. M., Sivapalan, M., Yaeger, M. A., and Poveda, G.: Regional patterns of interannual variability of catchment water balances across the continental U.S.: A Budyko framework, Water Resour. Res., 50, 9177–9193, https://doi.org/10.1002/2014WR016013, 2014.

Chen, S., Liu, Y., and Axel, T.: Climatic change on the Tibetan Plateau: Potential evapotranspiration trends from 1961-2000, Clim. Change, 76, 291–319, https://doi.org/10.1007/s10584-006-9080-z, 2006.

Chen, Y., Li, W., Xu, C., and Hao, X.: Effects of climate change on water resources in Tarim River Basin, Northwest China, J. Environ. Sci., 19, 488–493, https://doi.org/10.1016/S1001-0742(07)60082-5, 2007.

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

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, Int. J. Climatol., 28, 2031–2064, https://doi.org/10.1002/joc.1688, 2008.

de Rosnay, P. and Polcher, J.: Modelling root water uptake in a complex land surface scheme coupled to a GCM, Hydrol. Earth Syst. Sci., 2, 239–255, https://doi.org/10.5194/hess-2-239-1998, 1998.

de Rosnay, P., Polcher, J., Bruen, M., and Laval, K.: Impact of a physically-based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes, J. Geophys. Res., 107, ACL 3 1–18, https://doi.org/10.1029/2001JD000634, 2002.

de Rosnay, P., Polcher, J., Laval, K., and Sabre, M.: Integrated parameterization of irrigation in the land surface model ORCHIDEE. Validation over Indian Peninsula, Geophys. Res. Lett., 30, 1–4, https://doi.org/10.1029/2003GL018024, 2003.

Dingman, S. L.: Physical hydrology, Waveland Press, 2015.

Döll, P., Fiedler, K., and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, Hydrol. Earth Syst. Sci. Discuss., 6, 4773–4812, https://doi.org/10.5194/hessd-6-4773-2009, 2009.

D'Orgeval, T.: Impact du changement climatique sur le cycle de l'eau en Afrique de l'Ouest: modélisation et incertitudes, Ph.D. thesis, Paris 6, 2006.

D'Orgeval, T. and Polcher, J.: Impacts of precipitation events and land-use changes on West African river discharges during the years 1951-2000, Clim. Dyn., 31, 249–262, https://doi.org/10.1007/s00382-007-0350-x, 2008.

D'Orgeval, T., Polcher, J., and de Rosnay, P.: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, Hydrol. Earth Syst. Sci., 12, 1387–1401, https://doi.org/10.5194/hessd-5-2251-2008, 2008.

Ducoudré, N. I., Laval, K., and Perrier, A.: SECHIBA, a New Set of Parameterizations of the Hydrologic Exchanges at the Land-Atmosphere Interface within the LMD Atmospheric General Circulation Model, 1993.

Federer, C. A., Vörösmarty, C., and Fekete, B.: Intercomparison of methods for calculating potential evaporation in regional and global water balance models, Water Resour. Res., 32, 2315–2321, https://doi.org/10.1029/96WR00801, 1996.

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., and Willmott, C. J.: Uncertainties in precipitation and their impacts on runoff estimates, J. Clim., 17, 294–304, https://doi.org/10.1175/1520-0442(2004), 2004.

Fraedrich, K., Jansen, H., Kirk, E., Luksch, U., and Lunkeit, F.: The Planet Simulator: Towards a user friendly model, Meteorol. Zeitschrift, 14, 299–304, https://doi.org/10.1127/0941-2948/2005/0043, 2005.

Fu, B.: More on the calculation of average total evaporation (in Chinese), Sci. Atmos. Sin,, 5, 23–31, 1981.

Gao, X., Ye, B. S., Zhang, S. Q., Qiao, C. J., and Zhang, X. W.: Glacier runoff variation and its influence on river runoff during 1961-2006 in the Tarim River Basin, China, Sci. China Earth Sci., 53, 880–891, https://doi.org/10.1007/s11430-010-0073-4, 2010.

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, Water Resour. Res., 45, 1–15, https://doi.org/10.1029/2008WR007308, 2009.

Giorgi, F. and Francisco, R.: Evaluating uncertainties in the predicition of regional climate change, Geophys. Res. Lett., 27, 1295–1298, https://doi.org/10.1029/1999GL011016, 2000.

Gouttevin, I., Krinner, G., Ciais, P., Polcher, J., and Legout, C.: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology, Cryosphere, 6, 407–430, https://doi.org/10.5194/tc-6-407-2012, 2012.

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J. M., Prigent, C., Guyot, J. L., Cochonneau, G., Espinoza, J. C., Filizola, N., Fraizy, P., Lavado, W., De Oliveira, E., Pombosa, R., Noriega, L., and Vauchel, P.: Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets, Hydrol. Earth Syst. Sci., 16, 911–935, https://doi.org/10.5194/hess-16-911-2012, 2012.

Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat, P., Tallaksen, L. M., Warnaars, T., Wiberg, D., Weedon, G. P., Lanen, H. V., Ludwig, F., and Haddeland, I.: WATCH: Current Knowledge of the Terrestrial Global Water Cycle, J. Hydrometeorol., 12, 1149–1156, https://doi.org/10.1175/JHM-D-11-024.1, 2011.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, https://doi.org/10.1002/joc.3711, 2014.

Hernández, M. R. and Francés, F.: On the Influence of Error Model in the Good Performance of the Hydrological Model for the Right Reasons, International Conference on Hydroinformatics, 2014.

Hirabayashi, Y., Döll, P., and Kanae, S.: Global-scale modeling of glacier mass balances for water resources assessments: Glacier mass changes between 1948 and 2006, J. Hydrol., 390, 245–256, https://doi.org/10.1016/j.jhydrol.2010.07.001, 2010.

Hirabayashi, Y., Zang, Y., and Watanabe, S.: Projection of glacier mass changes under a high-emission climate scenario using the global glacier model HYOGA2, Hydrol. Res. Lett., 7, 6–11, https://doi.org/10.3178/HRL.7.6, 2013.

Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies, Agric. For. Meteorol., 138, 44–53, https://doi.org/10.1016/j.agrformet.2006.03.009, 2006.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42, 1–10, https://doi.org/10.1029/2005WR004376, 2006.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in combining projections from multiple climate models, J. Clim., 23, 2739–2758, https://doi.org/10.1175/2009JCLI3361.1, 2010.

Liu, J., Chen, J. M., and Cihlar, J.: Mapping evapotranspiration based on remote sensing: An application to Canada's landmass, Water Resour. Res., 39, 1189, https://doi.org/10.1029/2002WR001680,, 2003.

Liu, S., Ding, Y., Shangguan, D., Zhang, Y., Li, J., Han, H., Wang, J., and Xie, C.: Glacier retreat as a result of climate warming and increased precipitation in the Tarim river basin, northwest China, Ann. Glaciol., 43, 91–96, https://doi.org/10.3189/172756406781812168, 2006.

Liu, Z., Xu, Z., Huang, J., Charles, S. P., and Fu, G.: Impacts of climate change on hydrological processes in the headwater catchment of the Tarim River basin, China, Hydrol. Process., 24, 196–208, https://doi.org/10.1002/hyp.7493, 2010.

Mamitimin, Y., Feike, T., Seifert, I., and Doluschitz, R.: Irrigation in the Tarim Basin, China: farmers' response to changes in water pricing practices, Environ. Earth Sci., 73, 559–569, https://doi.org/10.1007/s12665-014-3245-2, 2014.

Mezentsev, V. S.: More on the calculation of average total evaporation, Meteorol. Gidrol, 5, 24–26, 1955.

Ol'dekop, E. M.: On evaporation from the surface of river basins: Transactions on Meteorological Observations, Tech. rep., Univ. of Tartu, Tartu, Estonia, Lur-evskogo (in Russian), 1911.

Pike, J. G.: The estimation of annual run-off from meteorological data in a tropical climate, J. Hydrol., 2, 116–123, 1964.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, Int. J. Climatol., 23, 479–510, https://doi.org/10.1002/joc.893, 2003.

Pitman, A. J., Slater, A. G., Desborough, C. E., and Zhao, M.: Uncertainty in the simulation of runoff due to the parameterization of frozen soil moisture using the Global Soil Wetness Project methodology, J. Geophys. Res. Atmos., 104, 16 879–16 888, https://doi.org/10.1029/1999JD900261, 1999.

Ponce, V., Pandey, R., and Ercan, S.: Characterization of drought across climatic spectrum, J. Hydrol. Eng., 5, 222–224, https://doi.org/10.1061/(Asce)1084-0699(2000)5:2(222), 2000.

Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., and Jakeman, A. J.: Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments, Water Resour. Res., 41, 1–11, https://doi.org/10.1029/2004WR003697, 2005.

Pritchard, H. D.: Asia's glaciers are a regionally important buffer against drought, Nature, 545, 169–174, https://doi.org/10.1038/nature22062, 2017.

Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, Journal of Hydrology, 198, 69–97, https://doi.org/10.1016/S0022-1694(96)03329-X, 1997.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, 1–22, https://doi.org/10.1029/2009WR008328, 2010.

Scheff, J. and Frierson, D. M. W.: Terrestrial aridity and its response to greenhouse warming across CMIP5 climate models, J. Clim., 28, 5583–5600, https://doi.org/10.1175/JCLI-D-14-00480.1, 2015.

Schreiber, P.: Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa, Z. Meteorol, 21, 441–452, 1904.

Shangguan, D., Liu, S., Ding, Y., Ding, L., Xu, J., and Jing, L.: Glacier changes during the last forty years in the Tarim Interior River basin, northwest China, Prog. Nat. Sci., 19, 727–732, https://doi.org/10.1016/j.pnsc.2008.11.002, 2009.

Shen, Y., Feng, M., Zhang, H., and Gao, F.: Interpolation methods of China daily precipitation data, J. Appl. Meteorol. Sci., 21, 279–286, 2010.

Shi, P., Yang, T., Zhang, K., Tang, Q., Yu, Z., and Zhou, X.: Large-scale climate patterns and precipitation in an arid endorheic region: linkage and underlying mechanism, Environ. Res. Lett., 11, 044 006, https://doi.org/10.1088/1748-9326/11/4/044006, 2016.

Siebert, S., Döll, P., Feick, S., Frenken, K., and Hoogeveen, J.: Global Map of Irrigation Areas version 5, March 2013, https://doi.org/10.13140/2.1.2660.6728, 2013.

Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance, Geophys. Res. Lett., 41, 429–434, https://doi.org/10.1002/2013GL058573, 2014.

Tao, H., Gemmer, M., Bai, Y., Su, B., and Mao, W.: Trends of streamflow in the Tarim River Basin during the past 50years: Human impact or climate change?, J. Hydrol., 400, 1–9, https://doi.org/10.1016/j.jhydrol.2011.01.016, 2011.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections., Philos. Trans. A. Math. Phys. Eng. Sci., 365, 2053–75, https://doi.org/10.1098/rsta.2007.2076, 2007.

Trenberth, K. E., Fasullo, J., Smith, L., Qian, T., and Dai, A.: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data, J. Hydrometeorol. - Spec. Sect., 8, 758–769, https://doi.org/10.1175/JHM600.1, 2006.

Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, J. Geophys. Res. Atmos., 118, 6064–6079, https://doi.org/10.1002/jgrd.50395, 2013.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, https://doi.org/10.1002/2014WR015638, 2014.

Weiß, M. and Menzel, L.: A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments, Adv. Geosci., 18, 15–23, https://doi.org/10.5194/adgeo-18-15-2008, 2008.

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C. Y.: Calibration of hydrological models using flow-duration curves, Hydrology and Earth System Sciences, 15, 2205–2227, https://doi.org/10.5194/hess-15-2205-2011, 2011.

Woo, M. K. and Marsh, P.: Snow, frozen soils and permafrost hydrology in Canada, 1999-2002, Hydrol. Process., 19, 215–229, https://doi.org/10.1002/hyp.5772, 2005.

Wu, Y.-p., Shen, Y.-p., and Larry Li, B.: Possible physical mechanism of water vapor transport over Tarim River Basin, Ecol. Complex., 9, 63–70, https://doi.org/10.1016/j.ecocom.2011.12.002, 2012.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A Gauge-Based Analysis of Daily Precipitation over East Asia, J. Hydrometeorol., 8, 607–626, https://doi.org/10.1175/JHM583.1, 2007.

Yang, D., Sun, F., Liu, Z., Cong, Z., Ni, G., and Lei, Z.: Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of China using the Budyko hypothesis, Water Resour. Res., 43, 1–12, https://doi.org/10.1029/2006WR005224, 2007.

Yang, H., Yang, D., Lei, Z., and Sun, F.: New analytical derivation of the mean annual water-energy balance equation, Water Resour. Res., 44, 1–9, https://doi.org/10.1029/2007WR006135, 2008.

Yang, T., Wang, X., Yu, Z., Krysanova, V., Chen, X., Schwartz, F. W., and Sudicky, E. A.: Climate change and probabilistic scenario of streamflow extremes in an alpine region, Journal of Geophysical Research: Atmospheres, pp. 8535–8551, https://doi.org/10.1002/2014JD021824.Received, http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50740/abstract, 2014.

Yang, T., Zhou, X., Yu, Z., Krysanova, V., and Wang, B.: Drought projection based on a hybrid drought index using Artificial Neural Networks, Hydrol. Process., 29, 2635–2648, https://doi.org/10.1002/hyp.10394, 2015.

Yang, Z.: Glacier Water Resources in China (in Chinese), Lanzhou: Gansu Science and Technology Press, 1991.

Yu, Z.: Assessing the response of subgrid hydrologic processes to atmospheric forcing with a hydrologic model system, Global and Planetary Change, 25, 1–17, https://doi.org/10.1016/S0921-8181(00)00018-7, 2000.

Yu, Z., Lakhtakia, M. N., Yarnal, B., White, R. A., Miller, D. A., Frakes, B., Barron, E. J., Duffy, C., and Schwartz, F. W.: Simulating the river-basin response to atmospheric forcing by linking a mesoscale meteorological model and hydrologic model system, Journal of Hydrology, 218, 72–91, https://doi.org/10.1016/S0022-1694(99)00022-0, 1999.

Yu, Z., Pollard, D., and Cheng, L.: On continental-scale hydrologic simulations with a coupled hydrologic model, Journal of Hydrology, 331, 110–124, https://doi.org/10.1016/j.jhydrol.2006.05.021, 2006.

Yu, Z., Lu, Q., Zhu, J., Yang, C., Ju, Q., Yang, T., Chen, X., and Sudicky, E. A.: Spatial and Temporal Scale Effect in Simulating Hydrologic Processes in a Watershed, Journal of Hydrologic Engineering, 19, 99–107, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000762, 2014.

Zhang, L., Dawes, W. R., and Walker, G. R.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water Resour. Res., 37, 701–708, https://doi.org/10.1029/2000WR900325, 2001.

Zhang, L., Hickel, K., Dawes, W., Chiew, F., Western, A., and Briggs, P.: A rational function approach for estimating mean annual evapotranspiration, Water Resour. Res., 40, W025 021–W02502 114, https://doi.org/10.1029/2003WR002710, 2004.

Zhang, T., Heginbottom, J., Barry, R., and J, B.: Distribution of seasonally and perennially frozen ground in the Northern Hemisphere, Permafrost, pp. 1289–1294, 2003.

Zhou, H., Song, Y., Hu, S., Hongwei, Z., Yudong, S., and Shunjun, H.: Irrigated agriculture and sustainable water management strategies in the Tarim Basin, in: Proc. Work. "New Approaches to water" Manag. Cent. Asia, Aleppo, Syria, 6–11 Nov, https://doi.org/10.1002/2013WR014696.Received, 2000.

Zomer, R. J., Trabucco, A., Straaten, O. V., and Bossio, D. A.: Trees and Water: Smallholder Agroforestry on Irrigated Lands in Northern India, Colombo, Sri Lanka: International Water Management Institute, 2007.

Zomer, R. J., Trabucco, A., Bossio, D. A., and Verchot, L. V.: Climate change mitigation : A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation, Agric. Ecosyst. Envir, 126, 67–80, https://doi.org/10.1016/j.agee.2008.01.014, 2008.

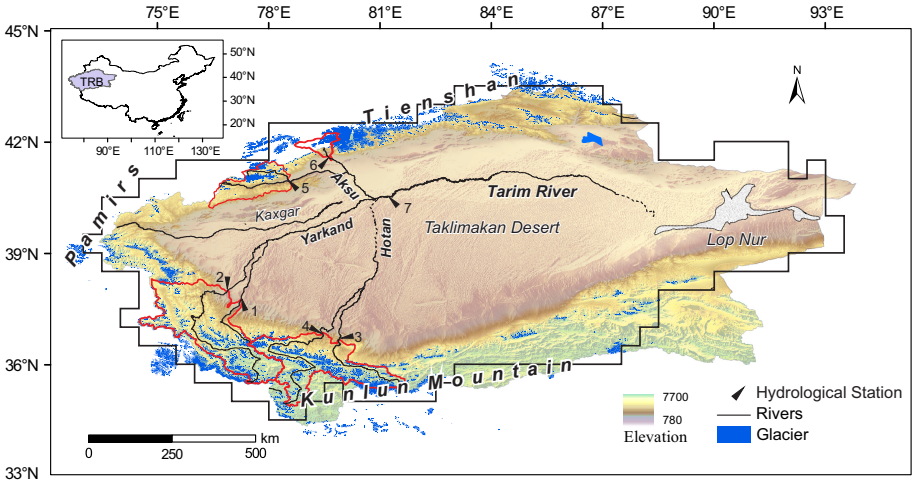


**Figure 1.** The major rivers and the glacier distribution in the Tarim basin. The upper Yarkand catchment is defined by hydrological gauges 1-JK and 2-KQ; the upper Hotan catchment is defined by hydrological gauges 3-TGZLK and 4-WLWT; the upper Aksu catchment is defined by hydrological gauges 5-SLGLK and 6-XHL. The Upper Tarim basin is defined by hydrological gauge 7-Alar.

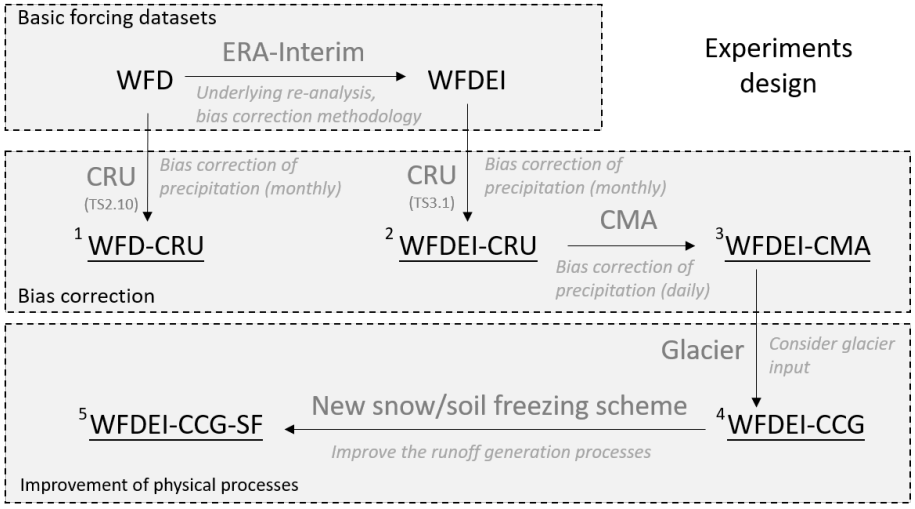


**Figure 2.** The flowchart of the 5 experiments designed for driving the ORCHIDEE in this study. WFD and WDDEI are two basic forcing datasets. The underlined terms with numbers are five experiments while the grey arrows represent the development of the forcing compared to their previous ones.

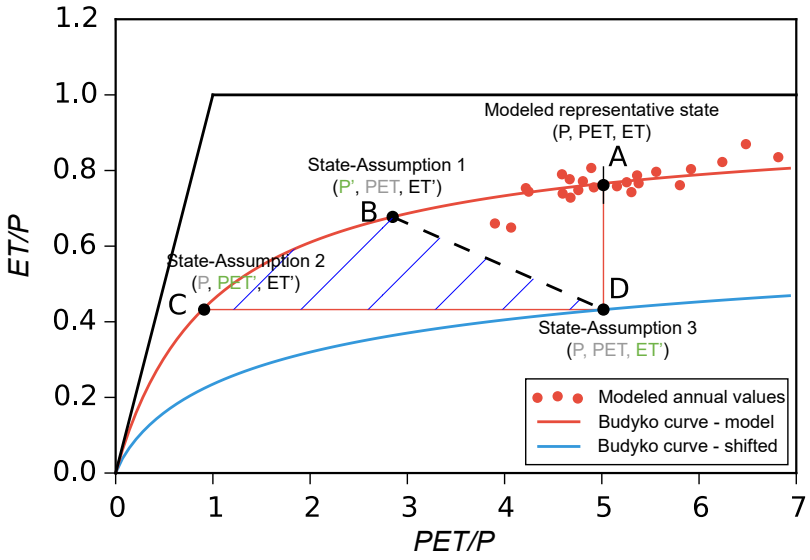


**Figure 3.** The illustration of the ORCHIDEE-Budyko framework. Point A represents the average state among the modeled annual values (with land surface model-ORCHIDEE), and the red curve is the simulated Budyko curve following the modeled state. Point B, C, D represent the representative state with shifting the $P$, $PET$ and $ET$ respectively with different assumptions to meet the discharge observations. A shifted Budyko curve (blue) is obtained crossing the point D which indicates a new state of model structure. The new points of B and C still stay on the original Budyko curve indicating that the model structure remain the same and the changes only relate to forcing variables. The shade area is the area among the three shifted states.

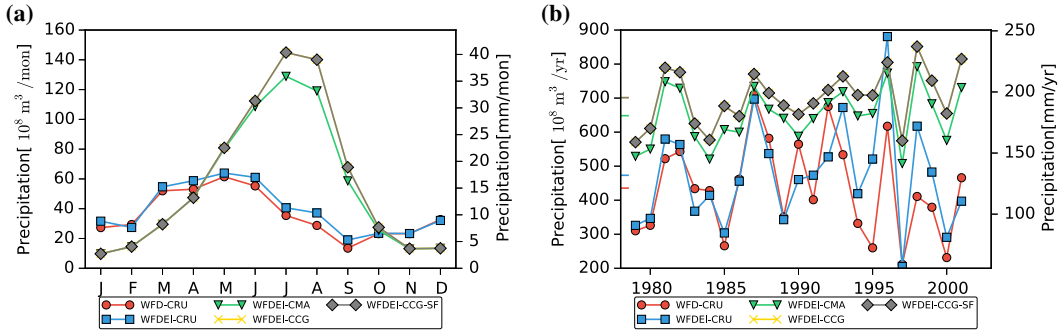


**Figure 4.** The inter-annual cycle and intra-annual series of the precipitation (including rainfall, snowfall and glacier) in difference simulations for the Upper Tarim basin.

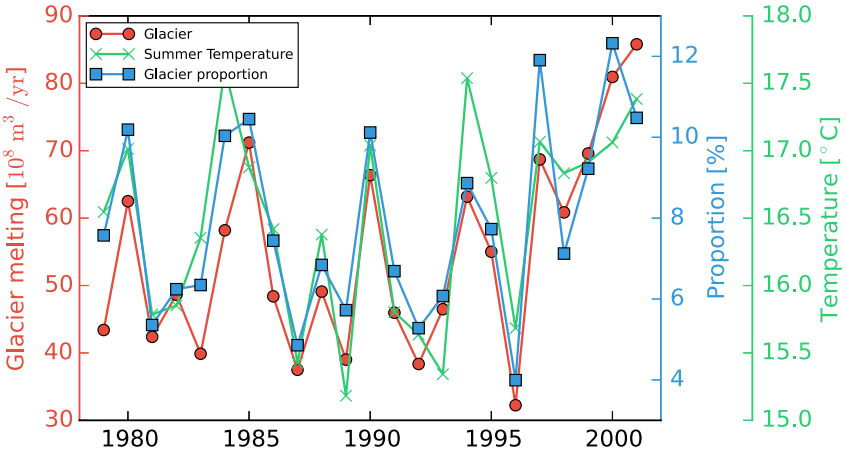


**Figure 5.** The temporal variations of the glacier melt, the proportion of glacier melt in the water input and the average summer temperature over the Upper Tarim basin. They are in good correlation and have a consistent increasing trend after 1990.

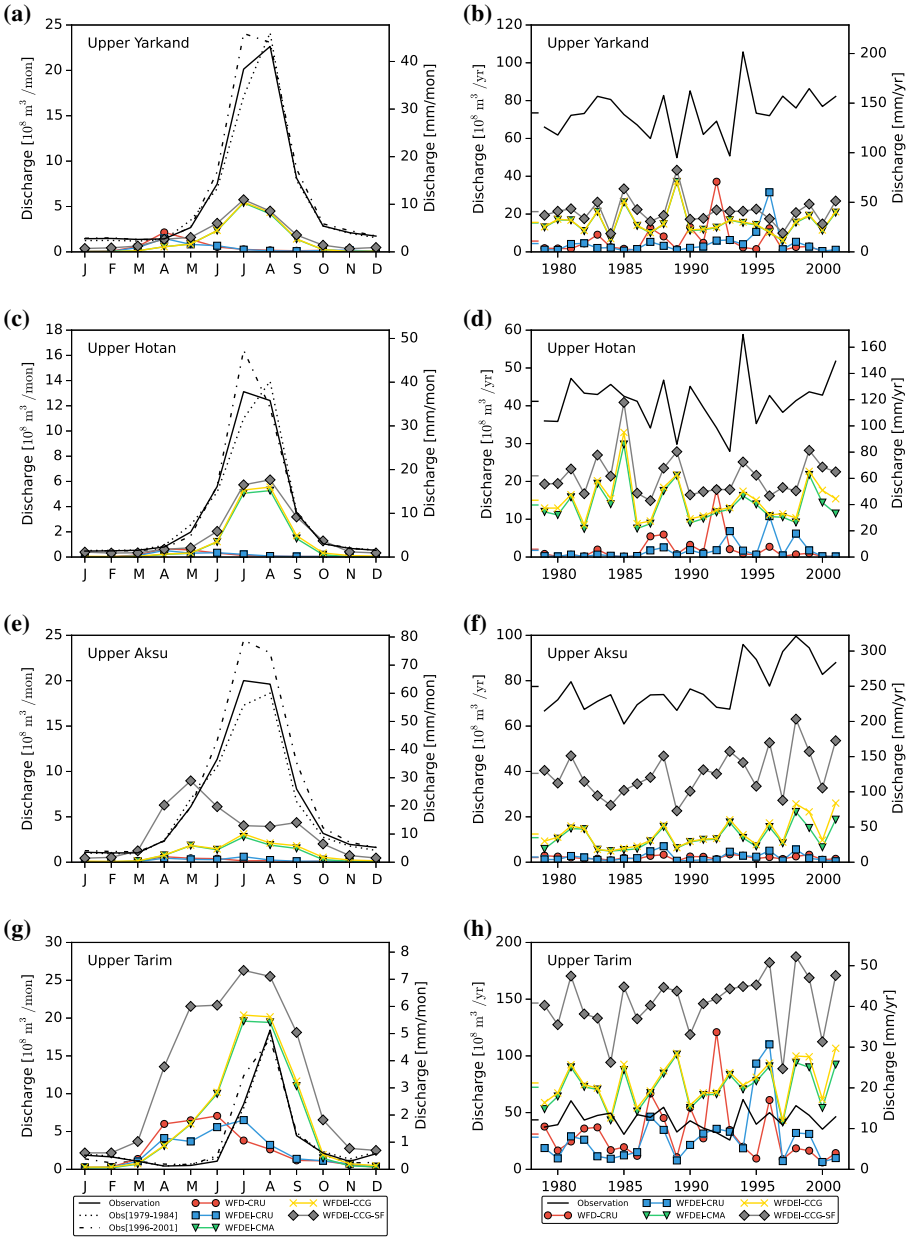


**Figure 6.** The inter-annual cycle and intra-annual series of the discharge simulation for three headwater catchments and the Upper Tarim basin, a,b-Yarkand, c,d-Hotan, e,f-Aksu and g,h for the Upper Tarim. Observed discharge for each subbasin was aggregated by the measurements at separated discharge gauges and shown as the black solid line. The simulated discharges at the corresponding grids were extracted from each experiment and plotted as the color lines with markers. The dotted line and dotdash line in the inter-annual cycle plots represent the observed discharge in different periods.

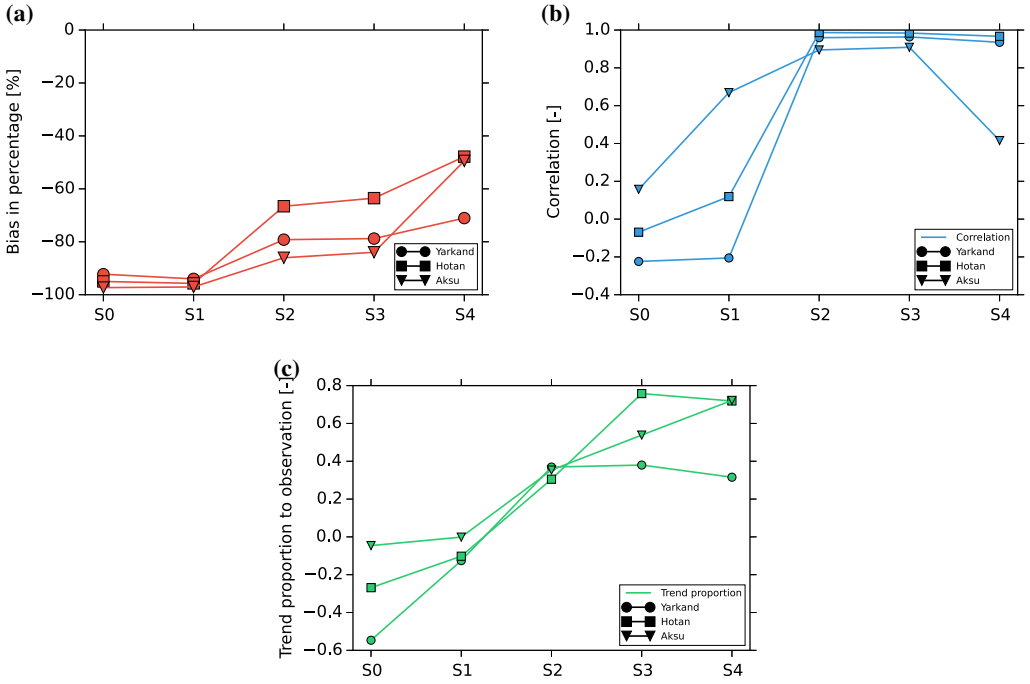


**Figure 7.** The discharge diagnostics of different experiments for three catchments (round-Yarkand, square-Hotan and triangle-Aksu). a) represents the absolute bias in percentage, b) represents the correlation of the inter-annual cycle and c) represents the ratio of the trend in estimated discharge to that in observed discharge for period 1990-2001. S0 to S4 correspond to the experiments WFD-CRU, WFDEI-CRU, WFDEI-CMA, WFDEI-CCG and WFDEI-CCG-SF respectively.

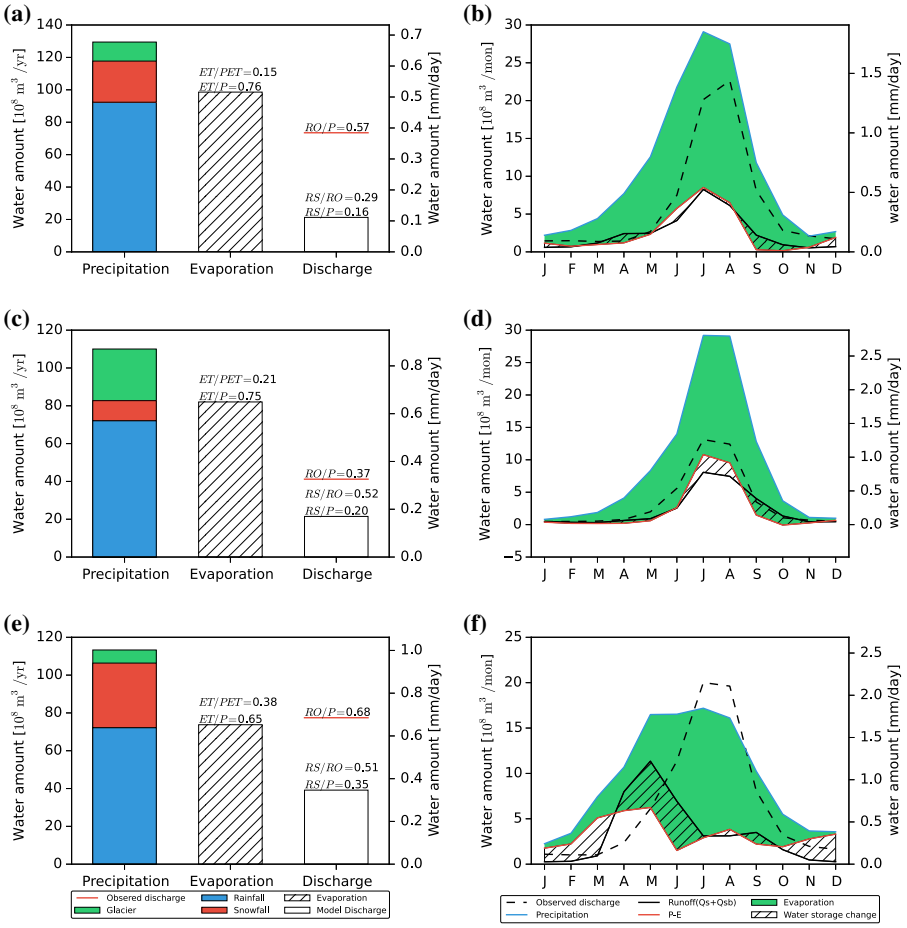


**Figure 8.** The water input components (rainfall, snowfall, glacier melt), evapotranspiration and discharge for three headwater catchments (a-Yarkand, c-Hotan and e-Aksu from the top to the bottom) in the WFDEI-CCG-SF simulation. In the left panels, the amounts of different variables are plotted as bars, while the average mean of the observed discharge (RO) is plotted as the red line. $RS$ denotes the simulated discharge by ORCHIDEE. In the right panels, the annual cycle of the water inputs (blue line), evapotranspiration (green shadow), estimated discharge (runoff plus drainage, solid black line), observed discharge (dashed black line) and the changes in terrestrial water storage (TWS, slashed area) changes are plotted as b,d and f. The green slashed area represents decreasing in TWS and the white one slashed area represents increasing in TWS.

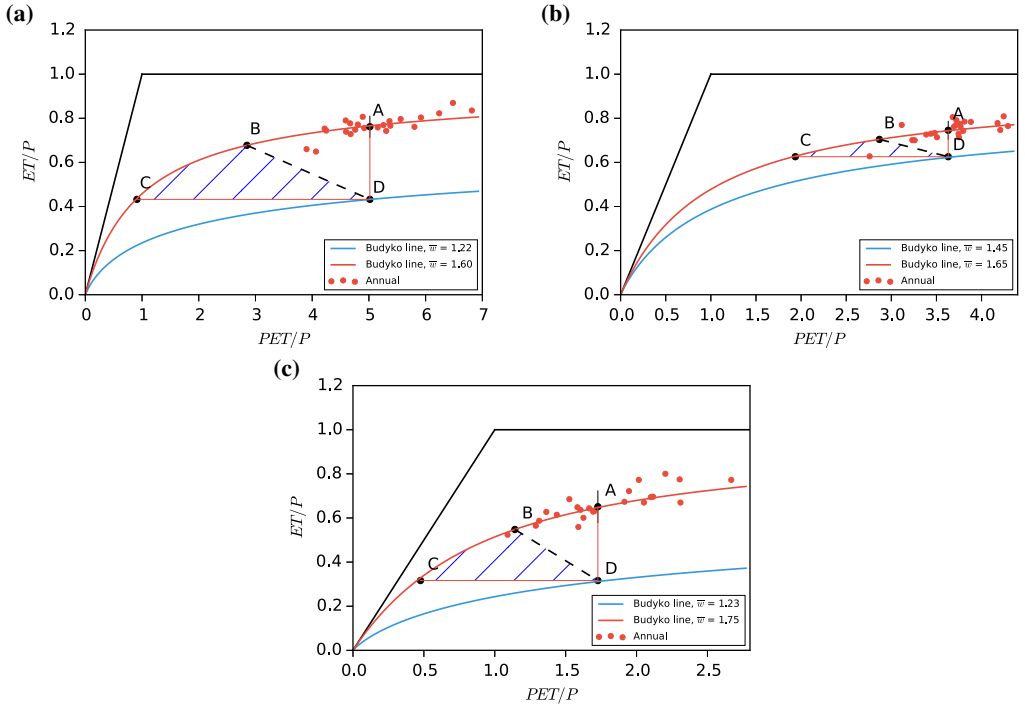


**Figure 9.** Budyko relation for three headwater catchments (a-Yarkand, b-Hotan, c-Aksu). The red points represent the values for each year, the $P$ is obtained from the forcing, while $ET$ and $PET$ are obtained from the model. Point A represents the long-term average Budyko relation and red lines are the optimal fitted Budyko line. Either the water input $P$, the potential evapotranspiration $PET$ or the actual evapotranspiration $ET$ can be modified to meet the observed discharge, which correspondingly shift the representative points from A to B, C or D. B and C stay in the original Budyko curve while a new optimal fitted Budyko curve through point D can be built after the changes in $ET$. The shaded area is the most likely area when not only single variable is changing.

**Table 1.** Different formulas of Budyko curves. Note that the aridity index is expressed as $\emptyset = PET/P$.

| No. | Parameter | Formula | Reference |
|---|---|---|---|
| 1 | Non-parameter | $ET/P = 1 - e^{(-\emptyset)}$ | Schreiber, 1904 |
| 2 | Non-parameter | $ET/P = \emptyset \tanh(1/\emptyset)$ | Ol'dekop, 1911 |
| 3 | Non-parameter | $ET/P = 1/(1 + (1/\emptyset)^2)^{0.5}$ | Pike, 1964 |
| 4 | Non-parameter | $ET/P = \{[1 - e^{-\emptyset}] \cdot \emptyset \cdot \tanh(1/\emptyset)\}^{0.5}$ | Budyko, 1974 |
| 5 | Parameter n | $ET/P = 1/[1 + (1/\emptyset)^n]^{1/n}$ | Mezentsev, 1955 ; Choudhury, 1999 ; Yang et al., 2008 |
| 6 | Parameter $\varpi$ | $ET/P = 1 + \emptyset - [1 + (\emptyset)^{\varpi}]^{1/\varpi}$ | Fu, 1981 |
| 7 | Parameter $\varpi$ | $ET/P = (1 + \varpi\emptyset)/(1 + \varpi\emptyset + 1/\emptyset)$ | Zhang et al., 2001 |

**Table 2.** The definition of climate types by precipitation and aridity index (Ponce et al., 2000).

| Climatic types | $P$(mm/yr) | Aridity index ($PET/P$) |
|---|---|---|
| Superarid | <100 | >30 |
| Hyperarid | 100-200 | 12-30 |
| Arid | 200-400 | 5-12 |
| Semiarid | 400-800 | 2-5 |
| Subhumid | 800-1600 | 0.75-2 |
| Humid | 1600-3200 | 0.375-0.75 |
| Hydrperhumid | 3200-6400 | 0.1875-0.375 |
| Superhumid | >6400 | <0.1875 |

**Table 3.** The five simulations in this study and basic diagnostics of the water inputs over three headwater catchments and the Upper Tarim basin. Units: $10^8 \mathrm{m}^3/\mathrm{yr}$.

| Subbasins | Area ($\mathrm{km}^2$) | Simulations | Precipitation | Rainfall | Snowfall | Glacier |
|---|---|---|---|---|---|---|
| Upper Yarkand | 55637 | WFD-CRU | 80.19 | 11.16 | 69.03 | - |
| | | WFDEI-CRU | 82.04 | 20.82 | 61.22 | - |
| | | WFDEI-CMA | 117.73 | 92.35 | 25.38 | - |
| | | WFDEI-CCG | 141.21 | 104.09 | 25.38 | 11.74 |
| | | WFDEI-CCG-SF | | | | |
| Upper Hotan | 34557 | WFD-CRU | 31.83 | 6.63 | 25.20 | - |
| | | WFDEI-CRU | 33.34 | 4.37 | 28.97 | - |
| | | WFDEI-CMA | 82.71 | 72.13 | 10.58 | - |
| | | WFDEI-CCG | 137.34 | 99.45 | 10.58 | 27.31 |
| | | WFDEI-CCG-SF | | | | |
| Upper Aksu | 31982 | WFD-CRU | 76.11 | 49.06 | 27.05 | - |
| | | WFDEI-CRU | 78.03 | 41.83 | 36.19 | - |
| | | WFDEI-CMA | 106.38 | 72.22 | 34.16 | - |
| | | WFDEI-CCG | 120.23 | 79.15 | 34.15 | 6.93 |
| | | WFDEI-CCG-SF | | | | |
| Upper Tarim | 359022 | WFD-CRU | 435.94 | 221.55 | 214.39 | - |
| | | WFDEI-CRU | 473.31 | 246.35 | 226.96 | - |
| | | WFDEI-CMA | 649.71 | 524.08 | 125.63 | - |
| | | WFDEI-CCG | 758.74 | 578.60 | 125.62 | 54.52 |
| | | WFDEI-CCG-SF | | | | |

**Table 4.** The water-energy components for the headwater catchments in WFDEI-CCG-SF simulation. Units: mm/yr.

| | Precipitation | Rainfall | Snowfall | Glacier | $ET$ | $RS$ | $PET$ | $RO$ |
|---|---|---|---|---|---|---|---|---|
| Upper Yarkand | 247.3 | 198.8 | 48.5 | 22.4 | 188.3 | 59.1 | 1240.4 | 140.4 |
| Upper Hotan | 317.8 | 287.2 | 30.6 | 78.9 | 236.9 | 81.3 | 1153.7 | 118.9 |
| Upper Aksu | 365.8 | 255.5 | 110.3 | 22.4 | 238.1 | 128.0 | 631.8 | 250.1 |

**Table 5.** The annual average values for different water-energy components ($P$, $ET$, $PET$; units in mm/yr) and their relations ($P$-$ET$, $PET/P$ and $ET/P$) for the three upstream subbasins. The scenarios correspond to the diagnostics of current model (A) and three bias assumptions list above from B to D. The bold values are the main factors changed within the three basic water-energy components. The changing ratio indicates the ratio of the changing value to the original value (unit in %). While the bias range implies the bias of the values in the current variables compared to the values they should be (unit in %).

| | | $P$ | $PET$ | $ET$ | $P-ET$ | $PET/P$ | $ET/P$ | Factor | Changing ratio | Bias Range |
|---|---|---|---|---|---|---|---|---|---|---|
| Upper Yarkand | A | 247.3 | 1240.4 | 188.3 | 59.0 | 5.02 | 0.76 | - | - | - |
| | B | **435.4** | 1240.4 | **294.9** | 140.5 | 2.85 | 0.68 | $P$ | 76.1% | -43.2% |
| | C | 247.3 | **225.0** | **106.9** | 140.4 | 0.91 | 0.43 | $PET$ | -81.9% | 451.2% |
| | D | 247.3 | 1240.4 | **106.9** | 140.4 | 5.02 | 0.43 | $ET$ | -43.2% | 76.1% |
| Upper Hotan | A | 317.8 | 1153.7 | 236.9 | 80.9 | 3.63 | 0.75 | - | - | - |
| | B | **402.3** | 1153.7 | **283.3** | 119.0 | 2.85 | 0.70 | $P$ | 26.6% | -21.0% |
| | C | 317.8 | **615.2** | **198.8** | 118.9 | 1.94 | 0.63 | $PET$ | -46.7% | 87.5% |
| | D | 317.8 | 1153.7 | **198.8** | 118.9 | 3.63 | 0.63 | $ET$ | -16.1% | 19.1% |
| Upper Aksu | A | 365.8 | 631.8 | 238.1 | 127.7 | 1.73 | 0.65 | - | - | - |
| | B | **553.3** | 631.8 | **303.1** | 250.2 | 1.14 | 0.55 | $P$ | 51.3% | -33.9% |
| | C | 365.8 | **174.1** | **115.7** | 250.1 | 0.48 | 0.32 | $PET$ | -72.4% | 262.4% |
| | D | 365.8 | 631.8 | **115.7** | 250.1 | 1.73 | 0.32 | $ET$ | -51.4% | 105.8% |

**Table 6.** Summary of the possible causes of the underestimation in discharge and the corresponding arguments. Three levels of the possibility are presented as Yes: with direct argument, likely: with indirect argument, No: with negative argument.

| Subbasin | Variable | Is it a factor? | Is it the only factor? |
|---|---|---|---|
| Upper Yarkand | $P$ underestimation | YES: Glacier (low glacier ratio, smaller trend in discharge simulation) | Likely NO (very high $PET$=1240.4 mm/yr; high $ET/P$ = 0.68) |
| | $PET$ overestimation | Likely YES: Very high $PET$=1240.4 mm/yr | NO (very low $PET$=225.0 mm/yr; low $PET/P$=0.91) |
| | $ET$ overestimation | | Likely NO (very high $PET$=1240.4 mm/yr; high $PET/P$=5.02) |
| Upper Hotan | $P$ underestimation | | Likely NO (very high $PET$=1153.7 mm/yr; high $ET/P$=0.70) |
| | $PET$ overestimation | Likely YES: Very high $PET$=1153.7 mm/yr | |
| | $ET$ overestimation | | Likely NO (very high $PET$=1153.7 mm/yr) |
| Upper Aksu | $P$ underestimation | YES: $P < R$ in summer | Likely NO (low $PET/P$=1.14) |
| | $PET$ overestimation | | NO (very low $PET$=174.1 mm/yr, very low $PET/P$=0.48) |
| | $ET$ overestimation | | |