

We would like to thank Referee #3 for his interest in this topic and for the valuable comments to improve our manuscript. Based on the comments additional calculations have been performed. Our point-by-point response to the comments is given in the following (Comments in black, Answers in blue and the content related to the changes in the revised manuscript are marked in orange.):

General comments

#1 The manuscript presents an interesting extension of the FLEX model with enhanced capability for root zone storage simulation at the global scale. Root zone storage capacity is an Achilles heel in global hydrological modelling that is crucial for determining water stress, but most often dependent on highly uncertain soil and rooting depth data. Thus, the authors are addressing an important issue of high relevance for the hydrological modelling community. However, among other improvement possibilities, I think that the analyses need to be more systematic and rigorous, and the manuscript need to better communicate the motivations underlying the developers' choices. I think the manuscript merits to be published after a major revision. My main concerns are the following:

We would like to thank the referee for assessing the quality of the paper and for providing very constructive and valuable comments. Indeed, the root zone storage capacity (RZSC) is a persistent weakness in global hydrological modeling, while it is crucial for water fluxes partitioning. This is the primary motivation of our work to integrate an advanced RZSC dataset into a hydrological model and to test the capacity of the model for root zone water storage (RZWS) simulation. Based on the referee's comments, we have performed additional computation and analyses to make the results more systematic and rigorous. A point-to-point reply to each specific comment is provided below.

#2 The manuscript could benefit from clearer descriptions of rationale and motivations for the model development, the analyses performed and other choices made. For example, why was runoff selected for evaluation against ERA-Interim/Land and the non-calibrated ISIMIP simulations? Why not use gauged data for a selection of basins and the ERA-Interim/Land and ISIMIP for global gridded comparisons? Were other variables and potentially better datasets considered and rejected for which reasons? How come capillary rise is disregarded on the basis of "lack of information at the global scale" is there are other models that take it into account? Why was Penman-Monteith FAO 56 PM method used (P6L25)? What were the considerations? Etc. Reviewers and readers will always have different views on preferred evaluation datasets and equations, but a clear description of the underlying rationale and motivation could help bridging differences in perspective if choices can be well-justified.

Thank you for the comments. The corresponding responses are provided below. Since there are a numerous questions in this comment, we have repeated the specific comment before the detailed response.

#2-1 For example, why was runoff selected for evaluation against ERA-Interim/Land and the non-calibrated ISIMIP simulations?

The gridded data set ERA-Interim/Land is selected for model evaluation mainly because the current version of the WAYS model does not include a runoff routing module on the global scale. Therefore, the results are not comparable to the observed gauged data. The ERA-Interim/Land data set is a global land surface reanalysis data. It is well assessed and has been used as reference data for many studies (Alfieri et al., 2013; Orth and Seneviratne, 2015; Reichle et al., 2017). Thus, the evaluation of WAYS against ERA-Interim/Land is well-justified.

Since WAYS uses the same driving data as the ISIMIP2a models and the ISIMIP2a simulations are widely discussed in many studies, we believe that the additional comparison between WAYS and the ISIMIP2a models can provide added-value for evaluating our model. Therefore, the ISIMIP2a simulations are also shown in the results section together with the ERA-Interim/Land data. We did mention the purpose of inclusion of the ISIMIP2a simulations in the results (page 13, line 15: “Since WAYS uses the same driving data as the ISIMIP2a models and the ISIMIP2a simulations have been widely discussed in many studies (Schewe et al., 2014; Müller Schmied et al., 2016; Gernaat et al., 2017; Zaherpour et al., 2018), we also perform a comparison between WAYS and the ISIMIP2a models to further evaluate our model.”). However, we did not mention this in the validation strategy section in the manuscript. We have now further clarified this issue in the revised manuscript (please see “Authors’ change in the manuscript.”).

Indeed, the ISIMIP2a models are not calibrated. We have mentioned this issue in the manuscript (page 13, line 30: This result occurs partly because some of the ISIMIP2a models are not calibrated at all (Zaherpour et al., 2018), whereas WAYS is calibrated to a Composite Monthly Runoff data set that assimilates the monitored river discharge (Fekete et al., 2011).). We have now revised the captions of related figures (Figures 2 and 3 in the revised manuscript) to note this issue.

Authors’ change in the manuscript.

Page 11, Line 5: (the changes is marked as blue)

In this study, the ERA-Interim/Land runoff data are used for validation of the runoff simulation, and the Normalized Difference Infrared Index (NDII) is used for the validation of the WAYS model for root zone water storage simulation. Considering the time period of coverage of both data sets (ERA-Interim/Land: 1979-2010, NDII: 2000-present) and the study period (1971-2010) of this work, the period 2001-2010 is selected as the validation period. For runoff evaluation, ISIMIP2a simulations are also included, as they use the same climate forcing as our study in the same period. The purpose of inclusion of the ISIMIP2a simulations for comparison can be found in the model evaluation section (see Section 4).

Page 11, Line 9: (the changes is marked as blue)

ERA-Interim/Land is a global land surface reanalysis data set produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Balsamo et al., 2015). The gridded data set ERA-Interim/Land is selected for model evaluation mainly because the current version of the WAYS model does not include a runoff routing model on the global scale. Therefore, the results are not comparable with observed gauge data. Since the ERA-

Interim/Land data set is well assessed with a quality check through comparison with ground-based and remote sensing observations, it has been used as reference data for many studies (Xia et al., 2014; Dorigo et al., 2017).

Page 13, Line 15: (the changes is marked as blue)

“Since the ISIMIP2a simulations are widely discussed in many studies (Schewe et al., 2014; Müller Schmied et al., 2016; Gernaat et al., 2017; Zaherpour et al., 2018), the comparison between WAYS and the ISIMIP2a models can provide added-value for evaluation in addition to examine only with the reference data.” is changed to “Since WAYS uses the same driving data as the ISIMIP2a models and the ISIMIP2a simulations have been widely discussed in many studies (Schewe et al., 2014; Müller Schmied et al., 2016; Gernaat et al., 2017; Zaherpour et al., 2018), we also perform a comparison between WAYS and the ISIMIP2a models to further evaluate our model.”

#2-2 Why not use gauged data for a selection of basins and the ERA-Interim/Land and ISIMIP for global gridded comparisons?

The simulated results are not compared to the gauged data because the current version of WAYS does not include a runoff routing module on the global scale. Therefore, the results are not comparable to the observed gauged data.

The evaluation of runoff is performed on the basin scale rather because it is difficult to show the global gridded comparisons for the time series simulated runoff. Thus, ten major basins are selected for the validation based on the coverage of the two RZSC datasets ($S_{R,CRU-SM}$ and $S_{R,CHIRPS-CSM}$). However, we agree with the reviewer that a comparison with gauged data is important. Thus, additional calculations have been performed. In the revised manuscript, we have evaluated our results with observed discharges from the Global Runoff Data Centre (GRDC). Since WAYS does not have a native runoff routing module at the moment, a third-part runoff routing tool CaMa-flood is applied to route the WAYS simulated runoff (Yamazaki et al., 2011). Given that the manuscript is already quite extensive, the discharge comparison is not a direct evaluation to WAYS but an evaluation of both WAYS and CaMa-Flood. We have included this information in Supplementary Information (SI).

Authors' change in the manuscript.

Page 18, Line 4: (the following paragraph is inserted in the end of runoff evaluation section)

The performance of WAYS is further evaluated against the gauge observations. Since WAYS does not have a native runoff routing module at the moment, a third-part runoff routing tool, CaMa-flood, is applied to route the WAYS simulated runoff (Yamazaki et al., 2011). The evaluation results can be found in the Supplementary Information (SI).

The following part is put in the SI.

To further evaluate the model performance, we have evaluated our results with observed discharges from the Global Runoff Data Centre (GRDC). The CaMa-flood model is the only available open-source global runoff routing model (<http://hydro.iis.u->

tokyo.ac.jp/~yamadai/cama-flood/) that is capable of simulating backwater effects, which is important for plain regions, making it a popular choice for many studies (Hirabayashi et al., 2013; Mateo et al., 2014; Pappenberger et al., 2012).

The GRDC stations along a river were selected with interstation areas larger than 7000 km² to omit catchments with hydrological processes that are not properly represented by global hydrological models operating at a 0.5° resolution (Hunger and Döll, 2008). In total, 154 stations are selected for major river basins worldwide. For discharge simulation, the CaMa-flood is run at a 0.5° resolution to maintain consistency with the WAYS simulated runoff. The WAYS_CRU simulation is used for routing due to global coverage of the data. The discharge is simulated for the 1971-2010 period.

For the evaluation, the simulated discharge is compared with the GRDC data at each selected station depending on the data availability. Since the observations provided by GRDC are on a monthly time scale, the simulated data are also aggregated to the monthly scale for the comparison. The correlation coefficient and Nash-Sutcliffe efficiency coefficient are calculated, while the correlation coefficient between the simulated discharges and GRDC station records are visualized in Figure S1.

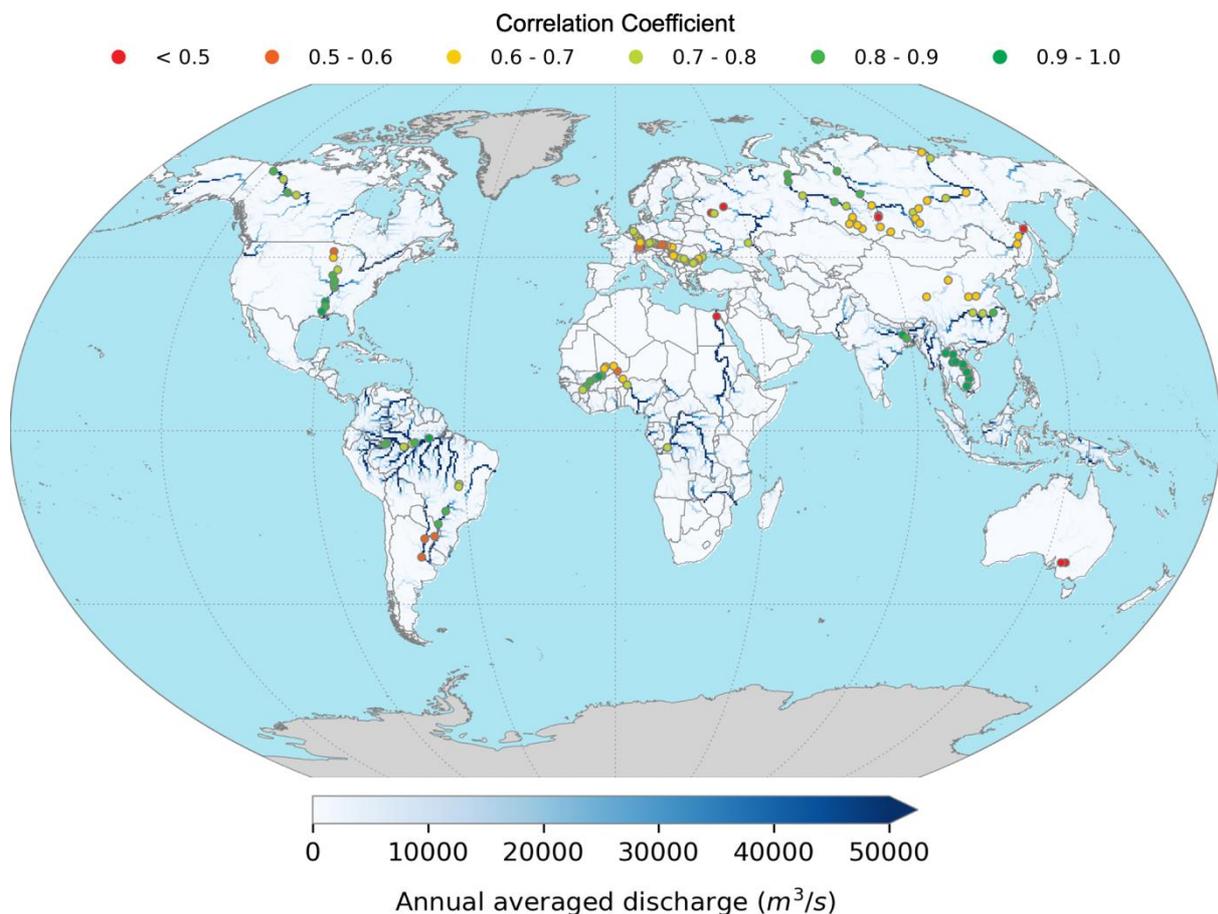


Figure S1. The evaluation of simulated discharge by comparison with the GRDC observations. The discharge is simulated by the CaMa-flood model, and the WAYS simulated runoff based on the RZSC data ($S_{R,CRU-SM}$) is used as the input data for routing. The background of the figure is the annual averaged discharge for the 1971-2010 period. The point indicates the correlation coefficient between simulated discharge and GRDC observations. The location of the points

implies the location of the GRDC station. Different colors at the points represent the magnitudes of the correlation coefficient.

The simulation shows a generally good correlation with the GRDC observations, while poor performance in the discharge simulation is also found in a few stations. The errors between the simulated discharge and observations could be caused by both the WAYS model for runoff simulation as well as the CaMa-Flood model for runoff routing, as the CaMa-Flood model itself also shows different performances in basins across the world (Yamazaki et al., 2011). The relatively low performance of WAYS is found in middle-high latitudes compared with low-middle latitude regions. This result could be explained by the relatively simple snow-melt module in the WAYS model, which thus could consequently produce low-quality runoff for river routing in cold regions. In Australia, only two GRDC stations in the Murray Darling basin are selected for the evaluation, and the correlation coefficient between simulated discharge and GRDC station is less than 0.5, indicating the large difference between them.

Figure S2 shows the histogram of the data points within different intervals of the correlation coefficient. Only in 7.2% of the stations are the correlations between simulation and observation less than 0.5. For more than half the stations, the correlations are higher than 0.7. The results show a generally good correspondence between the simulated and observed discharge. The generally good performance in the discharge simulation confirms the strong capacity of WAYS for runoff generation.

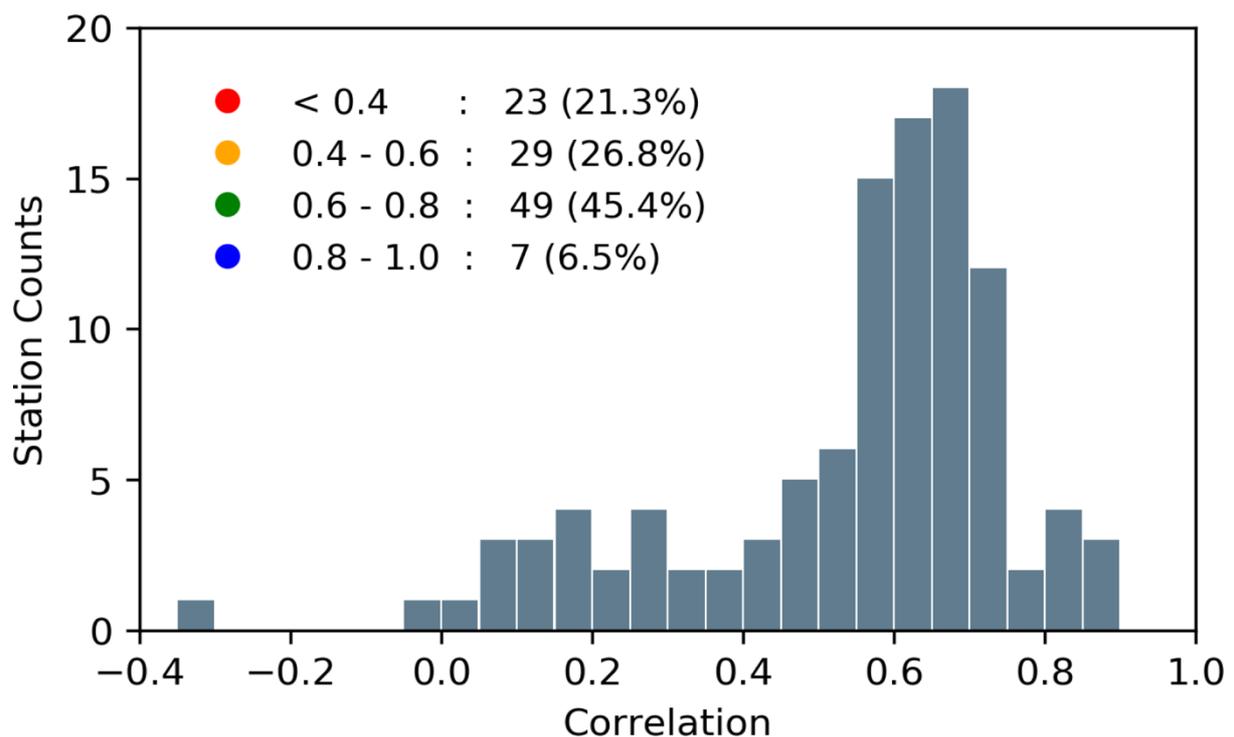


Figure S2. Histogram showing the percentage of data points within different intervals of the correlation coefficient.

#2-3 Were other variables and potentially better datasets considered and rejected for which reasons?

In addition to ERA-Interim/Land, other reanalysis runoff data are available, such as ERA-Interim, GLDAS, and NECP, among others. However, they show low robustness based on the available research results. For instance, GLDAS v1.0-CLM is found to overestimate runoff globally, and GLDAS v1.0-Noah generated more surface runoff over northern middle-high latitudes (Lv et al., 2018). GLDAS v2.0-Noah showed a significant underestimation trend in exorheic basins (Wang et al., 2016). The snowmelt-runoff peak magnitude simulated by GLDAS v2.1-Noah was found to be excessively high in June and July (Lv et al., 2018). NECP runoff is found to be too high during the winter and too low during the summer in the Mississippi River Basin (Roads and Betts, 2000). ERA-Interim is found to be less close to the observed stream flows compared with ERA-Interim/Land (Balsamo et al., 2015).

The ERA-Interim/Land is well assessed with quality checks by comparison with ground-based observations (GRDC observation) and is widely used as benchmark data (Alfieri et al. 2013; Balsamo et al. 2015; Orth and Seneviratne 2015; Reichle et al. 2017; Wang-Erlandsson et al. 2014). Therefore, it is selected as the reference data for this study for runoff comparison.

Authors' change in the manuscript.

Page 11, Line 20: (the following paragraph is inserted in the end of section "3.3.1 ERA-Interim/Land Runoff Data")

It should be noted that there are other reanalysis runoff data available, such as ERA-Interim, GLDAS and NECP. However, they show low robustness based on the available research results. For instance, GLDAS v1.0-CLM was found to overestimate runoff globally, and GLDAS v1.0-Noah generated more surface runoff over the northern middle-high latitudes (Lv et al., 2018). GLDAS v2.0-Noah showed a significant underestimation trend in exorheic basins (Wang et al., 2016). The snowmelt-runoff peak magnitude simulated by GLDAS v2.1-Noah was found to be excessively high in June and July (Lv et al., 2018). NECP runoff was found to be too high during the winter and too low during the summer in the Mississippi River Basin (Roads and Betts, 2000). ERA-Interim was found to be less close to the observed stream flows compared with ERA-Interim/Land data (Balsamo et al., 2015).

#2-4 How come capillary rise is disregarded on the basis of "lack of information at the global scale" is there are other models that take it into account?

WAYS contains the capillary module, which is adopted from the FLEX model. At the current stage, it is, however, disabled due to the lack of global information on the groundwater table, which could affect the simulated results in this work, e.g., evaporation and RZWS. We decided to disable the capillary module based on our experimental analysis.

We set up two experimental runs for WAYS to check the impact of the capillary module in the current version by switching it on/off. Since there is no observed groundwater table information to constrain the capillary rise amount, switching on the capillary module significantly overestimates the evaporation globally. The global averaged annual evaporation reaches 697 mm/year. Switching off the capillary module reduces the evaporation to 513 mm/year. A merged benchmark synthesis product of evaporation, i.e.,

LandFluxEVAL data, shows only 491 mm/year, which is much closer to the value without the capillary module. Thus, the capillary module is temporary disabled in WAYS until the global information on groundwater table is available. We have mentioned this issue in the revised manuscript. A response to the similar comment can be found in “the response letter to Referee #1” (Comment 6).

In fact, many models ignore the capillary at the global scale due to the absence of groundwater table information (Döll et al., 2003; De Graaf et al., 2015; Hanasaki et al., 2018). Consideration of the capillary in hydrological simulation is more popular in regional studies, mainly due to the local groundwater data availability (Gao et al., 2014; Vergnes et al., 2014).

Authors' change in the manuscript.

Page 25, Line 4: The following paragraph is inserted in the discussion part

Moreover, the current study does not consider the groundwater access and irrigation mainly due to the lack of global information. The groundwater table information is crucial for capillary rise simulation (Vergnes et al., 2014). Capillary rise simulation without proper water table information could significantly overestimate the evaporation. Thus, the capillary rise flux is ignored in this study. A similar strategy has also been applied by other works due to the absence of the information on the global water table (Döll et al., 2003; De Graaf et al., 2015; Hanasaki et al., 2018). Observations of irrigation on the global scale are also not available (Leng et al., 2015). Although there are simulated irrigation data available on the global scale, the inherent uncertainties could be propagated in our model simulation. Therefore, irrigation is also not considered at this time. However, this neglect could potentially introduce biases into the model simulation in irrigated areas and deep rooted plant-distributed regions, as both irrigation and capillary rise are an additional supply of soil water recharge. The biases may cause an underestimation of evaporation, especially in the dry summertime (Vergnes et al., 2014). This underestimation could consequently affect the simulation of RZWS and runoff because of the interlinkage of these three elements (Rockström et al., 1999). It is found that ignoring the capillary rise could reduce soil water content in the root zone (RZWS), while the runoff will also be reduced (Vergnes et al., 2014). However, these shortcomings can be simply overcome once the global data are available.

#2-5 Why was Penman-Monteith FAO 56 PM method used (P6L25)? What were the considerations?

Indeed, many methods are available to estimate potential evapotranspiration (PET) from standard meteorological observations. The Penman-Monteith FAO 56 PM method is recommended by FAO and other studies based on their thorough analysis in PET method intercomparisons (Allen et al., 1998; Lu et al., 2005; Vörösmarty et al., 1998). The Penman-Monteith FAO 56 PM method is based on fundamental physical principles and is found to be the most reliable method for potential evapotranspiration estimation where sufficient meteorological data exist (Chen et al., 2005; Kingston et al., 2009).

We would like to mention that FLEX uses the Hamon method for PET estimation. However, the Hamon method is found to have less robustness in different climatic conditions as well as drawbacks in terms of the daily variability of PET simulation (Bai et al., 2016; Droogers and Allen, 2002). Therefore, we have used the Penman-Monteith FAO 56 PM method in our study.

Authors' change in the manuscript.

Page 7, Line 4: (the changes is marked as blue)

Potential evapotranspiration is derived by the Hamon equation (Hamon, 1961) in the FLEX model, and it is now replaced by the using the Penman-Monteith FAO 56 PM method (Allen et al., 1998) for the following reason. The Hamon method is found to have less robustness in different climatic conditions as well as drawbacks in the daily variability of the PET simulation due mainly to the relatively simple equation in the Hamon method, as it only employs the average air temperature as an input (Bai et al., 2016; Droogers and Allen, 2002). In contrast, the Penman-Monteith FAO 56 PM method is based on fundamental physical principles and is found to be the most reliable method for potential evapotranspiration estimation when sufficient meteorological data exist (Chen et al., 2005; Kingston et al., 2009). The Penman-Monteith FAO 56 PM method is recommended by FAO and other studies based on thorough analyses of PET method intercomparisons (Allen et al., 1998; Jian biao et al., 2005; Vörösmarty et al., 1998).

#3 The analyses could be better designed to facilitate understanding of how and why WAYS perform in certain ways, and thus, give more insight into how various components of the model affect the root zone storage and runoff simulation? For example, can the authors show how results are affected by e.g., use of root zone storage capacity derived from uncertain root depth and soil data versus the root zone storage capacity from Wang-Erlandsson et al., 2016? Can the authors perform some sensitivity analyses to highlight model structure and parameter sensitivity?

Thank you for the comment. We agree with the referee that our study could benefit from a better design of the experiment. To facilitate understanding of how RZSC could affect the model simulation, we have additionally conducted a simulation of WAYS with RZSC derived from an uncertain root depth and soil data. The simulated results are then compared between two runs, i.e., one with RZSC from Wang-Erlandsson (2016) and the other with uncertain RZSC. Since this part could support the conclusion of this paper regarding the importance of correct representation of RZSC in models, we include it in the revised manuscript by adding some figures in the SI.

In addition, we have also performed a sensitivity test to highlight the model structure and parameter sensitivity. Since this part is not directly related to the main conclusion of the manuscript but important to demonstrate the model robustness, we have included it in the SI.

Authors' change in the manuscript.

Page 21, Line 26: (the following paragraphs are added)

RZSC is a key parameter of the WAYS model. Therefore, it is important to investigate how RZSC could affect the model simulation. In addition to the model simulated with satellite data-derived RZSC products ($S_{R,CHIRPS-CSM}$ and $S_{R,CRU-SM}$), we have additionally conducted WAYS simulations with RZSC derived from uncertain root depth and soil data. The uncertain RZSC ($S_{R,LOOKUP-TABLE}$) is derived based on literature values of root depth and soil texture data (Müller Schmied et al., 2014; Wang-Erlandsson et al., 2016). Due to the global coverage of the RZSC data ($S_{R,CRU-SM}$), only the simulation with $S_{R,CRU-SM}$ is used for comparison. The spatial distribution of the uncertain RZSC is shown in Figure S17, and the differences between $S_{R,CRU-SM}$ and $S_{R,LOOKUP-TABLE}$ are shown in Figure S18. It can be seen that there are large differences between the two RZSC products. The simulation with uncertain RZSC $S_{R,LOOKUP-TABLE}$ shows overestimation globally except for some regions around low-middle latitudes. The latitudinal averaged RZSC further confirms the overestimation of $S_{R,LOOKUP-TABLE}$ at middle-high latitudes (Figure S19).

The large differences between these two RZSC data sets also introduce differences in simulated hydrological elements. Figure S20 shows the impacts of RZSC on the model simulation, including runoff, evaporation and RZWS. A blue color (decrease of RMSE and increase of the ranked correlation) indicates an improvement of the simulated results by replacing the uncertain RZSC ($S_{R,LOOKUP-TABLE}$) with satellite data-derived RZSC ($S_{R,CRU-SM}$), while a red color implies the opposite. For comparison, reference data are used for different variables. For runoff, evaporation and RZWS, the reference data are ERA-Interim/Land (2001-2010, monthly), LandFluxEVAL (1989-2005, monthly) and NDII (2001-2010, 8-days), respectively. Generally, the model simulations are improved by using the RZSC $S_{R,CRU-SM}$. This result emphasizes the importance of an appropriate representation of RZSC in WAYS. A decline of the model performance is also found in some regions at high latitudes and low latitudes. This result can be partially explained by the inherent uncertainty in the $S_{R,CRU-SM}$ data, as they are derived from other data sets. The RZSC derivation method itself as well as the input data can also introduce biases (Wang-Erlandsson et al., 2016).

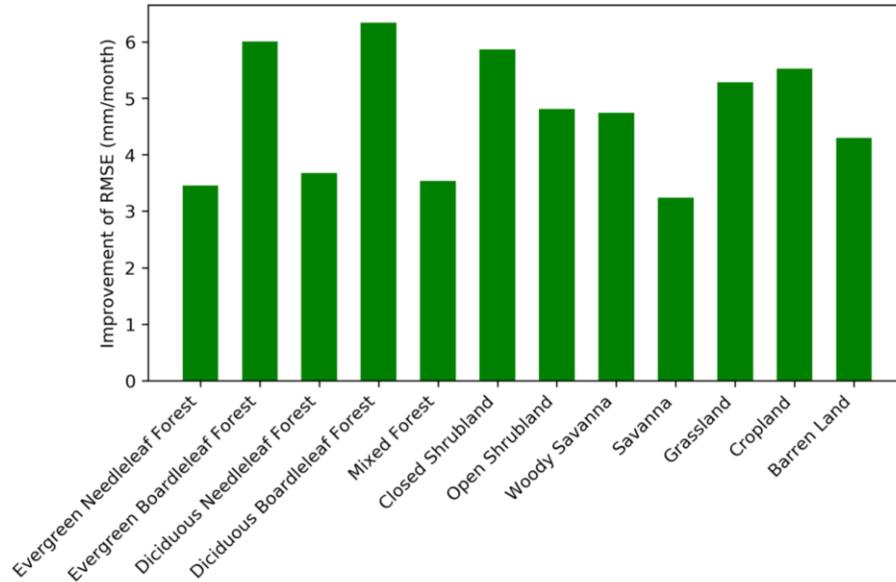
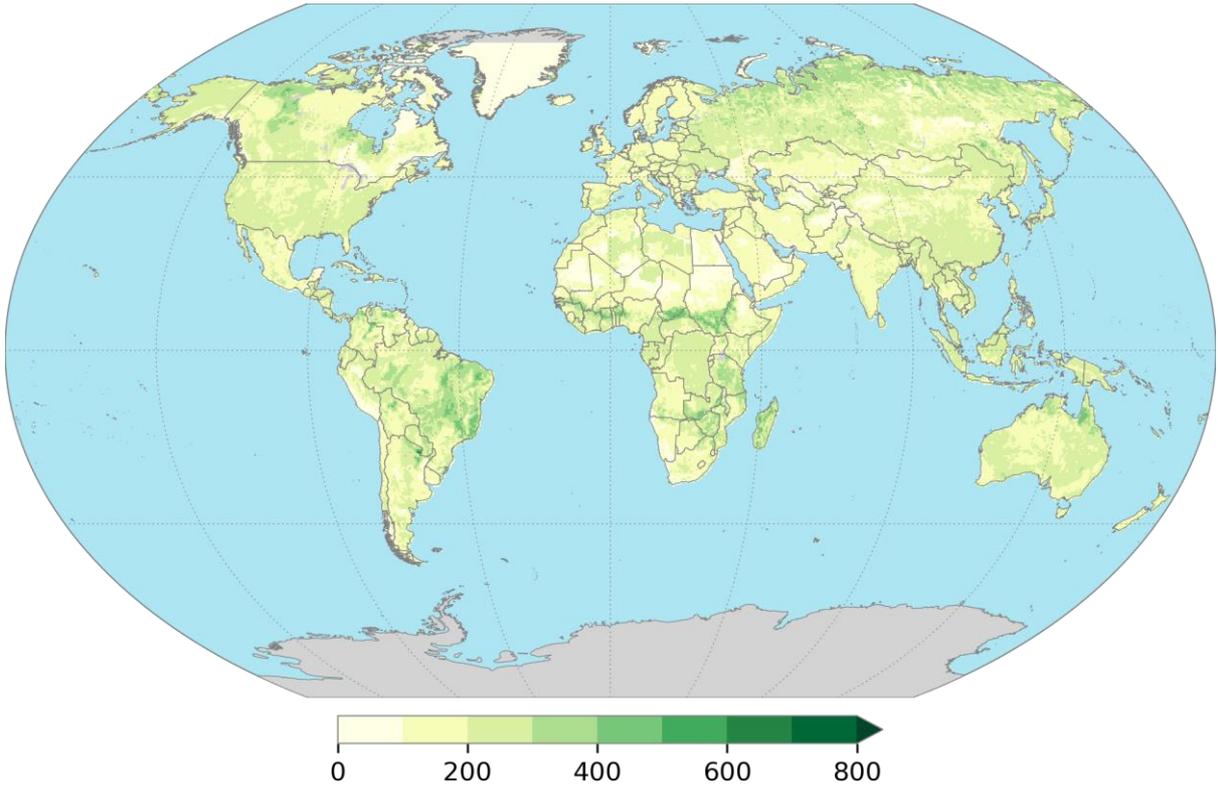


Figure 7. The improvement of RMSE in evaporation simulations for different land covers by using the satellite data-derived RZSC ($S_{R,CRU-SM}$) instead of the uncertain RZSC ($S_{R,LOOKUP-TABLE}$).

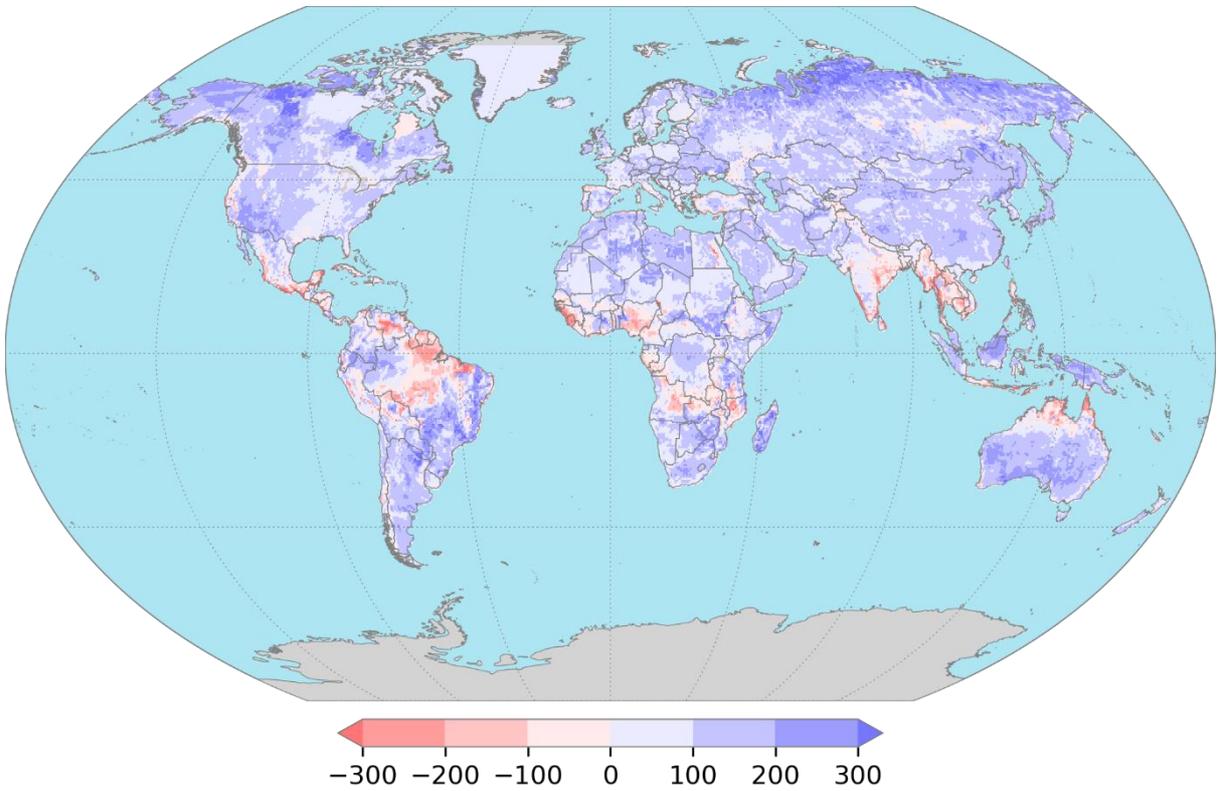
Figure 7 shows the RMSE improvements of simulated monthly evaporation for different land covers obtained by implementing the satellite data-derived RZSC ($S_{R,CRU-SM}$) instead of the uncertain RZSC ($S_{R,LOOKUP-TABLE}$). The analysis reveals that the satellite data-derived RZSC ($S_{R,CRU-SM}$) has great potential to improve the evaporation simulation for all kinds of land covers. The largest improvements are found in broadleaf forests. The improvements in the needleleaf forest, mixed forest and savanna are relatively low. The findings also resonate with another work that used a simple terrestrial evaporation to atmosphere model (STEAM) for evaporation simulation (Wang-Erlandsson et al., 2016).

The following figures can be found in the SI of this paper.



Root Zone Storage Capacity (mm)

Figure S3. Spatial distribution of uncertain RZSC ($S_{R,LOOKUP-TABLE}$).



Root Zone Storage Capacity (mm)

Figure S4. The difference between $S_{R,CRU-SM}$ and $S_{R,LOOKUP-TABLE}$ ($S_{R,LOOKUP-TABLE} - S_{R,CRU-SM}$).

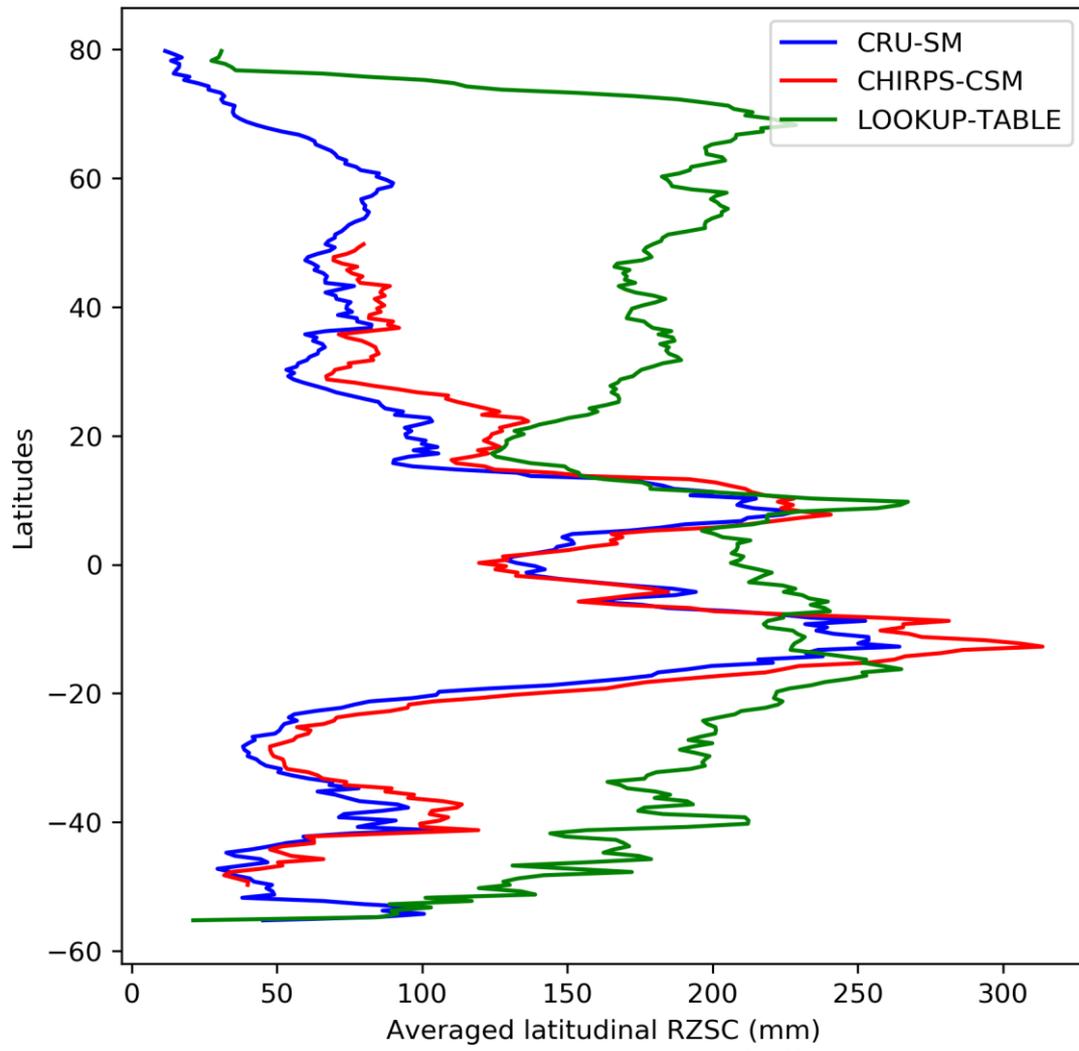


Figure S5. Latitudinal averaged RZSC of different products.

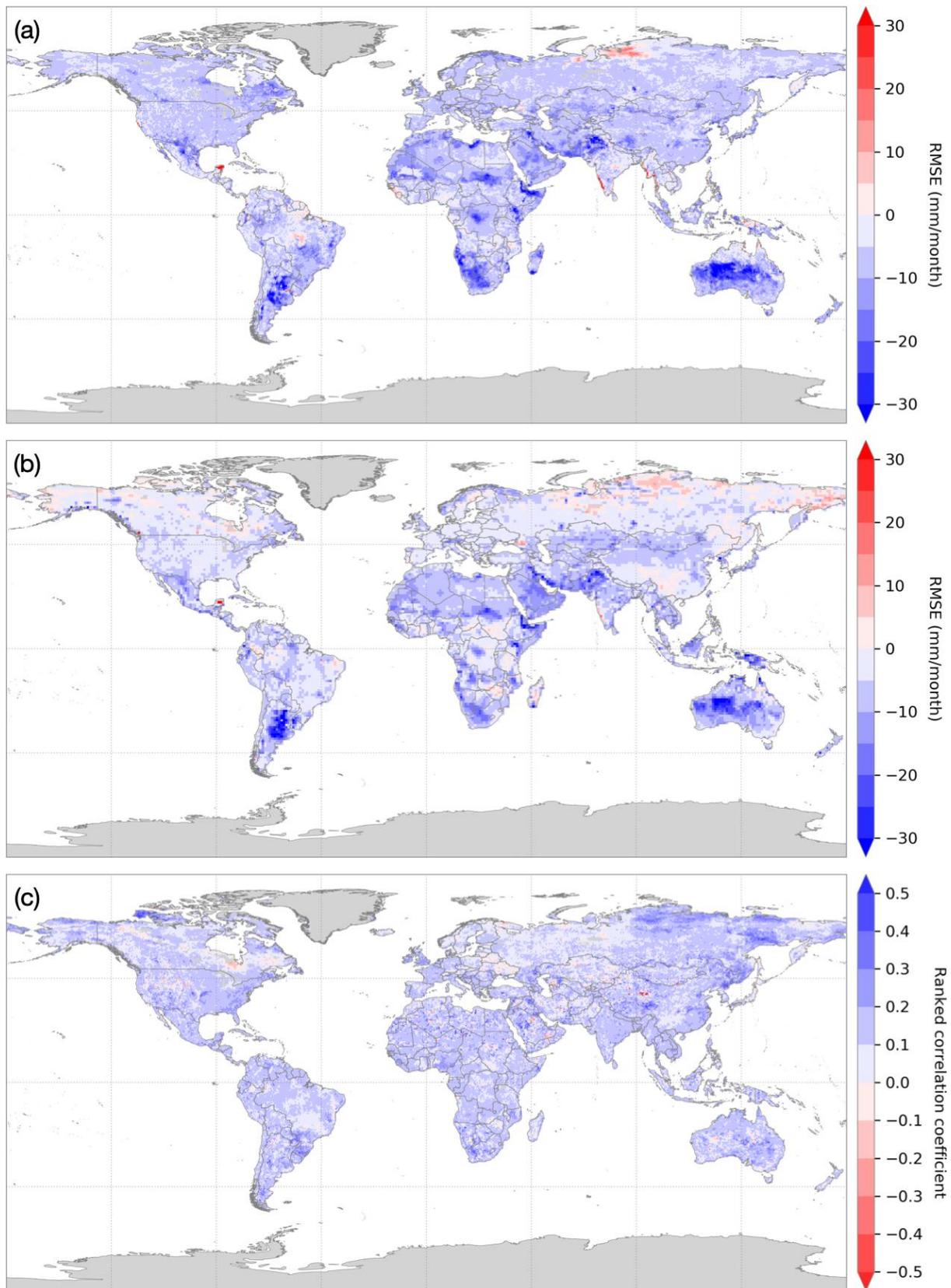


Figure S6. The impacts of RZSC on the model simulation. Blue color indicates the improvement of the simulated results by replacing the uncertain RZSC ($S_{R,LOOKUP-TABLE}$) with satellite data-derived RZSC ($S_{R,CRU-SM}$), while red color implies the opposite. (a) The result for runoff and the reference data for comparison is ERA-Interim/Land data (2001-2010, monthly), (b) the result for evaporation and the reference data for comparison is

LandFluxEVAL data (1989-2005, monthly), and (c) the result for RZWS and the reference data is NDII data (2001-2010, 8 days).

The following part is put in the SI.

In the WAYS mode, β and C_e are two crucial parameters that control the partitioning of precipitation into evaporation and runoff, thus affecting the water balance. Due to the incredibly high computation cost, only the sensitivity of the model simulation to these two parameters are tested. First, a pixel is selected randomly from the domain to demonstrate the impacts of parameter perturbation on simulated evaporation, runoff and RZWS. For each experiment, only one parameter is perturbed, and the other one is set to the calibrated value. The calibrated value for β and C_e is 0.17 and 1.67, respectively. The parameter is perturbed within the range randomly 1000 times during the experiment. Simulations are executed from 2009 to 2010 on a daily scale, while the results are shown on a monthly scale (see Figure S7). The model is more sensitive to parameter C_e than parameter β . The uncertainties caused by the parameter C_e are generally larger than those caused by the parameter β , especially for RZWS. These two parameters also have complementary effects on the model simulation, causing larger uncertainties for the simulation than one parameter.

To further investigate the uncertainties stemming from parameters on a global scale, a Monte Carlo simulation of 1000 samples is performed by perturbing the two parameters simultaneously. For both parameters, the normal distribution is used for the Monte Carlo perturbation. Simulations are executed from 2001 to 2010 on a daily scale. The coefficient of variation (CV) for each pixel is then calculated, which reflects the uncertainties (De Graaf et al., 2015). A high value of CV indicates relatively higher uncertainty caused by the parameters, while a low value of CV implies the opposite. Figure S8 shows that parameter-induced uncertainties of evaporation and runoff have similar patterns, while the magnitude is slightly higher for the runoff globally. This finding is consistent with the pixel-based sensitivity test (see Figure S7). The simulated RZWS has the largest uncertainties with the Monte Carlo simulation. Additionally, the uncertainties of RZWS show the opposite trend to the uncertainties of evaporation and runoff. In the northern part of Africa, the Arabian Peninsula, northwest of China and southern part of Australia, the uncertainties in evaporation and runoff are low. However, the uncertainties in RZWS are quite large in these regions.

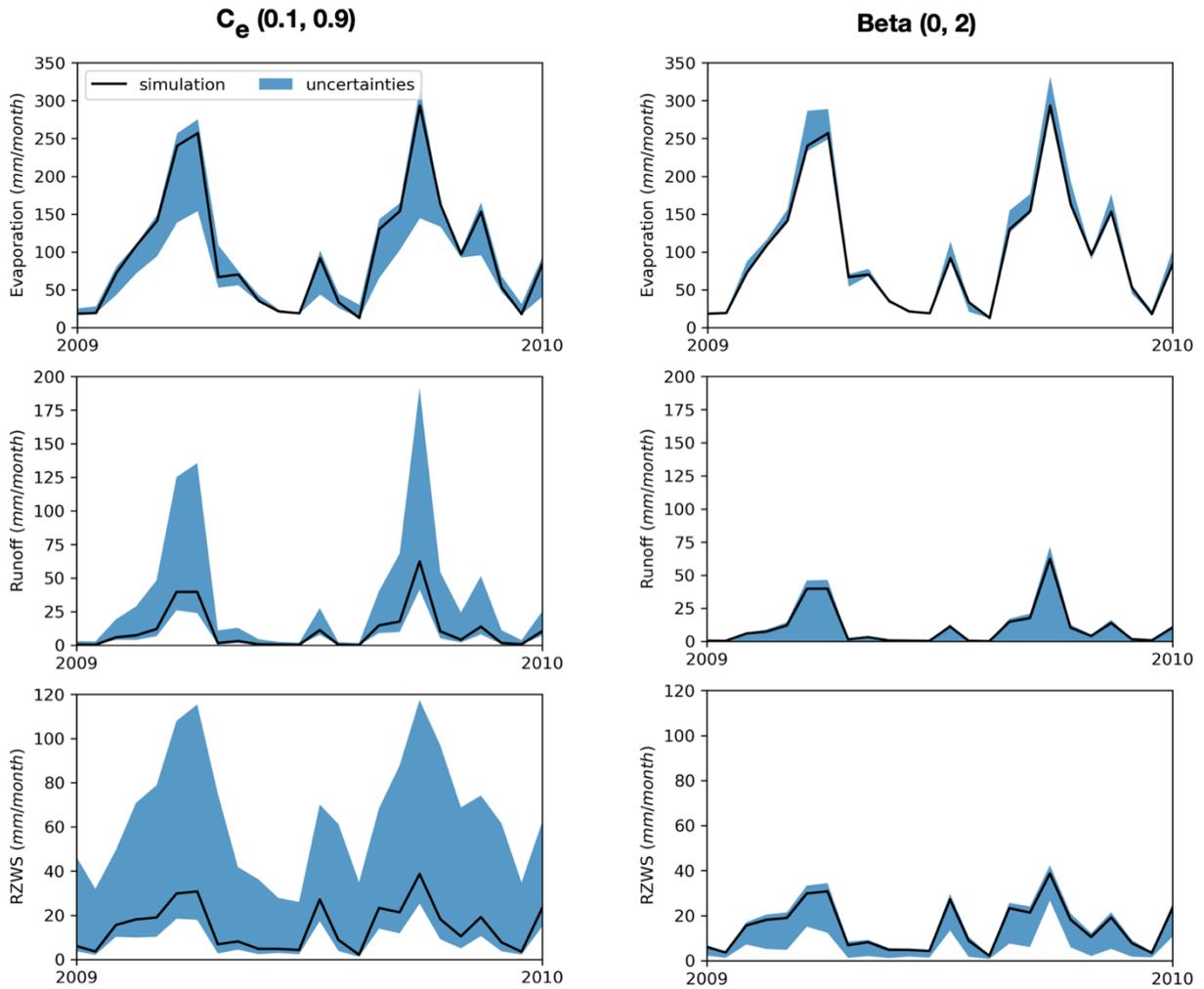


Figure S7. Sensitivity of simulated evaporation (top), runoff (middle) and RZWS (bottom) to parameter β and C_e in a randomly selected pixel within the domain. The black solid line represents the simulation based on the calibrated value. The blue area indicates the uncertainties induced from the perturbation of the parameter 1000 times.

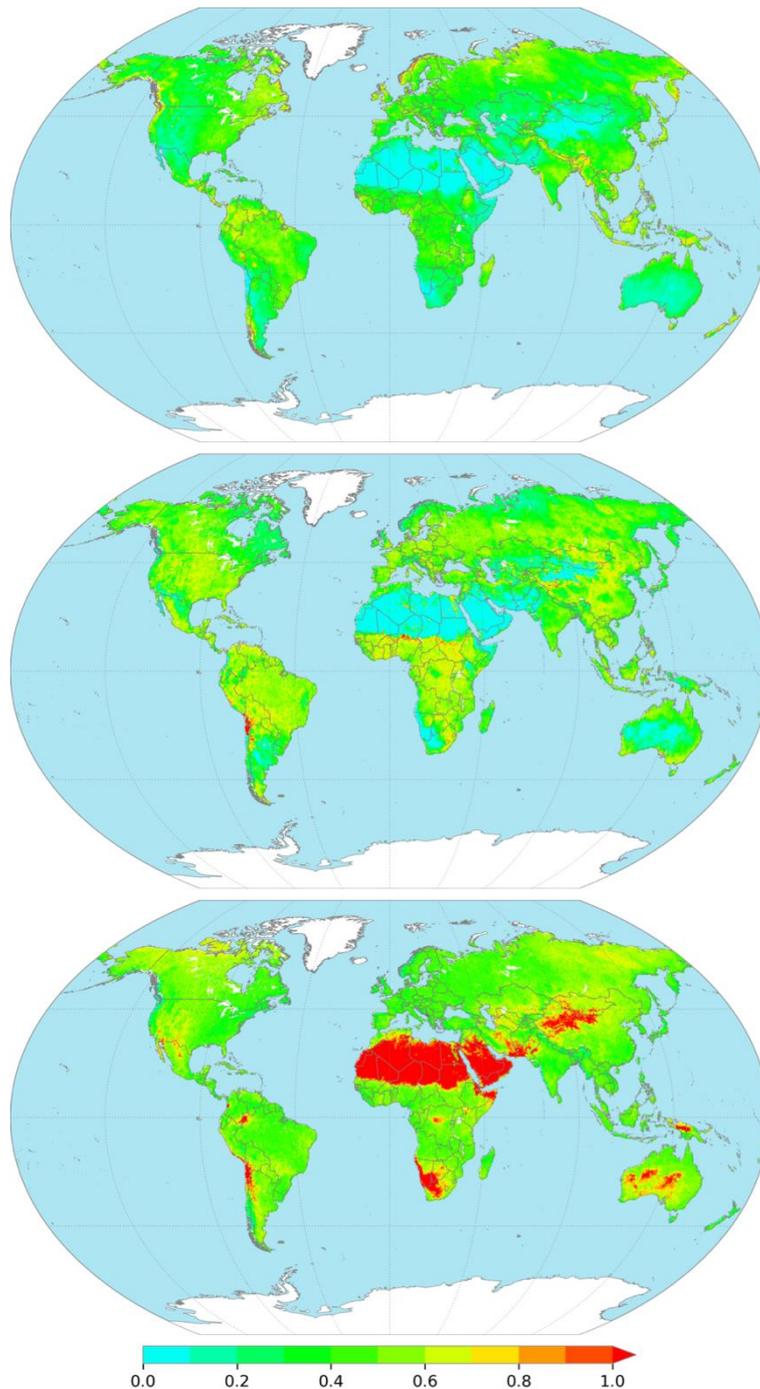


Figure S8. Coefficient of variation of model-simulated evaporation (top), runoff (middle) and RZWS (bottom) from 1000 Monte Carlo simulations with different parameter settings for β and C_e .

#4 The WAYS model is developed based on essential features of the FLEX model (P3L14), and as such I would (1) suggest the authors to present an overview of the similarities and differences between the two and (2) to retain “FLEX” in the model naming (e.g., FLEX-WAYS). Retaining FLEX in the name benefits the model developers that do not need to explain the model roots and will have an easier time communicating the new model developments that builds on an existing well-established mode, and would also be a nice acknowledgement of the earlier FLEX model developments. The practice of name roots

exists in the modelling community, and e.g., the models LPJmL and LPJ-GUESS show through their names that they share the same roots.

Thank you for the comment. To present the similarities and differences between FLEX and WAYS, we have updated Table 1 in the revised manuscript. The last column named “Reference” in Table 1 highlights the sources of equations that are adopted from FLEX or from the literature.

Regarding the model naming, however, we would prefer to keep it as WAYS for the following reasons: (1) WAYS is expected to be further developed by integrating new features, such as a water quality module that allows for environmental impact studies or an economic module to connect the physics of water and virtual water. In this case, WAYS needs its own postfix to identify different features, e.g., WAYS-WQ or WAYS-ECO. A prefix of FLEX will make the model name too complicated. (2) The FLEX model itself actually has different branches, e.g., FLEX-Topo indicates a topography-driven FLEX model, and FLEX^D represents a semidistributed FLEX model while FLEX^{T0} stands for FLEX-Topo without constraints (Gao et al., 2014). All the different types of FLEX have the same equations for all hydrological processes but with different model structures during the application. WAYS has replaced many equations from FLEX to enhance the capacity of the model for global simulation. A prefix of FLEX could cause confusion as the FLEX branches seldom change the equations. (3) The application of name roots to models is a good strategy for the models that share the core structure and equations but with different features as added functions. For instance, the VIC model is developed based on a small-scale distributed model Xinanjiang (Zhao, 1992). It has its own name and has been further developed by including additional features, e.g., VIC-CropSyst-v2, that simulate the nexus of climate, hydrology, cropping systems, and human decisions.

Authors’ change in the manuscript.

Page 8: Table 1 is updated. (the changes are marked as blue)

Table 1. Water balance and constitutive equations used in WAYS

Reservoirs	Water balance equations	Constitutive equations	Reference
		$P_{tj} = \max(0, P_r - (S_{i,max} - S_i))$ (2)	-
Interception reservoir	$\frac{dS_i}{dt} = P_r - E_i - P_{tj}$ (1)	$E_i = E_p \left(\frac{S_i}{S_{i,max}} \right)^{2/3}$ (3)	Deardorff (1978)
		$S_{i,max} = m_c L$ (4)	Wang-Erlandsson et al. (2014)
Snow reservoir	$\frac{dS_w}{dt} = \begin{cases} -M & \text{if } T > T_t \\ P_s & \text{if } T \leq T_t \end{cases}$ (5)	$M = \begin{cases} \min(S_w, F_{DD}(T - T_t)) & \text{if } T > T_t \\ 0 & \text{if } T \leq T_t \end{cases}$ (6)	Rango and Martinec (1995)
		$P_e = P_{tj} + M$ (8)	-
Root zone reservoir	$\frac{dS_{rz}}{dt} = P_e - R - E_a$ (7)	$\frac{R}{P_e} = 1 - \left(1 - \frac{S_{rz}}{(1 + \beta)S_{rz,max}} \right)^\beta$ (9)	Sriwongsitanon et al. (2016)
		$E_a = (E_0 - E_i) \cdot \min\left(1, \frac{S_{rz}}{C_c S_{rz,max}(1 + \beta)}\right)$ (10)	Sriwongsitanon et al. (2016)
Slow response reservoir	$\frac{dS_s}{dt} = R_s - Q_s$ (11)	$R_s = \min(f_s R, R_{s,max})$ (12)	Döll and Fiedler (2008)
		$Q_s = S_s / K_s$ (13)	Döll et al. (2003)
		$R_f = R - R_s$ (15)	-
Fast response reservoir	$\frac{dS_f}{dt} = R_f - Q_{ff} - Q_f$ (14)	$Q_{ff} = \max(0, S_f - S_{ft}) / K_{ff}$ (16)	-
		$Q_f = S_f / K_f$ (17)	-

Note: all the time scale-dependent parameters need to be divided by Δt to make the equations dimensionally correct and suitable for any other time scales.
- in the reference column indicates that the formula is taken from the FLEX model.

Page 26, Line 9: (the changes are marked as blue)

This study was supported by the National Natural Science Foundation of China (Grant No. 41625001), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060402) and the National Natural Science Foundation of China (Grant No. 41571022). [We would like to acknowledge the authors of the FLEX model for their great help during the development of WAYS.](#)

#5 The WAYS performance evaluation in terms of root zone storage moisture is highly dependent on the comparison with NDII, which weakens the conclusions, since also further work is still needed to robustly establish the relationship between NDII and soil moisture at the global scale. It is after all only recently suggested by Sriwongsitanon et al. (2016) – a study in a river basin in Thailand – that NDII can have the potential to be used as a proxy for catchment scale root zone storage capacity. The authors could potentially strengthen their conclusions by evaluating model simulation outputs with additional sources of data/methods, such as FLUX-tower, evaporation, EVI etc. Summarizing evaluation figures can be shown in the main manuscript, and others could be included in Supplementary Information. A more detailed list of the equations and calibration process could also be included in Supplementary Information for transparency.

Thank you for the comment. Based on the referee's suggestion, we have performed an additional evaluation on our model simulation to strengthen the conclusions. Since RZWS has close links to the total evaporation, we have compared the WAYS simulated evaporation to FLUX-tower observations (FLUXNET2015) as well as a merged benchmark synthesis product of evaporation (LandFluxEVAL).

The referee also suggested that we compare our model simulation to EVI. However, as we stated in our manuscript (page 2, line 22), EVI as well as NDVI are the most widely used vegetation indices, which have strong links to root zone soil moisture but cannot reflect the dynamics of the water content in the root zone layer (Santos et al. 2014). However, NDII determines the water stress of plants in the root zone by taking advantage of the property of shortwave infrared reflectance, thus possessing the intuitive advantage of reflecting the dynamics of RZWS than EVI (Sriwongsitanon et al. 2016). As we have already compared our model results to NDII, we prefer to skip the comparison between our model simulation with EVI.

Since this evaporation evaluation can support the conclusion of this paper regarding the capacity for hydrological cycle simulation, we include the work of evaporation evaluation against FLUX-tower data in the main text of the revised manuscript and the rest in SI. Please refer to the changes in the revised manuscript.

Regarding "A more detailed list of the equations and calibration process could also be included in Supplementary Information for transparency.", we have further updated the table with model equations by adding necessary references to each equation and one more table to illustrate the model parameters as well as the parameter ranges. Since the model equations and parameters are important to a study on model development, we have retained these two tables in the main text of the manuscript. In addition, the calibrated parameters will be uploaded as supplements in terms of netCDF files.

Authors' change in the manuscript.

Page 20, Line 26: The following paragraphs are inserted

RZWS has a close link to the total evaporation, as RZWS represents the available water that plants can use. In this section, the performance of WAYS in evaporation simulation is evaluated against the FLUXNET2015 data. FLUXNET2015 is a global network of micrometeorological flux measurement sites that measure the exchange of CO₂, water vapor and energy between the biosphere and the atmosphere (Pastorello et al., 2017). The tower-measured latent heat flux (LF, W/m²) is converted to ET (mm/day) using the proportionality parameter between energy and depth units of ET (Velpuri et al., 2013) as follows:

$$ET = \frac{LE}{\lambda}$$

The results are shown in Figure S15. The background is the annual averaged evaporation from WAYS for the period 1971-2010. The points indicate the comparison results between the flux tower and WAYS simulation. The locations of the points indicate the locations of the flux towers, and the colors indicate the correlation coefficient. WAYS is found to have relatively better performance in America, Europe and China than in Africa and Australia.

However, a few stations near the boundary of America and Europe also show weak correlations between the simulations and flux tower data.

Figure S16 shows the percentage of data points within different intervals of the correlation coefficient. The calculated correlation coefficient is crowded in the interval of 0.6-0.8, while more than half of the stations (56%) show a correlation coefficient of more than 0.6. The relatively poor performance of the model in some regions could be partially explained by the following reason. FLUXNET2015 corresponds to point-based observation data, while WAYS simulates the evaporation on grid cells with a 0.5 degree spatial resolution. For the comparison, the model simulation in a certain pixel is selected based on the distance between the flux tower and the center of the pixel. The model simulation actually represents an averaged value for a 0.5 x 0.5-degree pixel. This averaging will inherently introduce errors when comparing the simulation to station-based data. Similar results are also found in other studies comparing FLUXNET2015 data to either model simulations or remote sensing-derived evaporations (Lorenz et al., 2014; Velpuri et al., 2013).

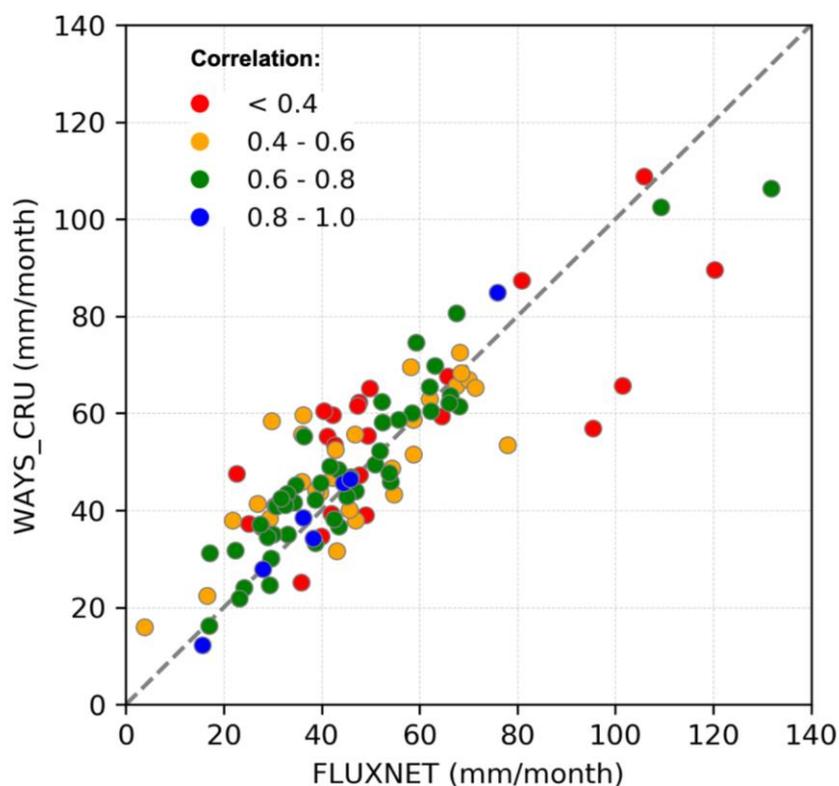


Figure 6. Averaged monthly evaporation of WAYS simulation (WAYS_CRU) against the FLUXNET data.

Furthermore, the average monthly evaporation is compared to the FLUXNET2015 data at each flux tower, and the results are shown in Figure 6. Good correspondence between the model simulation and flux tower data can be found by visual inspection. The points with a higher correlation coefficient show a better relationship between the model simulation and flux tower observation and are distributed closer to the diagonal. The evaluation results confirm the generally good performance of WAYS in monthly evaporation simulation. The detailed results on evaporation evaluation against FLUXNET2015 are provided in the SI as Excel files. In addition, an evaluation of the evaporation simulation is further conducted

against LandFluxEVAL, a merged benchmark synthesis product of evaporation at the global scale (Mueller et al., 2013). The results can be found in the SI.

The following part is put in the SI.

The model simulation is further compared to a gridded data set, LandFluxEVAL data, for evaporation evaluation. The LandFluxEVAL data are a merged benchmark synthesis product of evaporation on a global scale and a combination of land-surface model simulations, remote sensing products, reanalysis data and ground observation data (Mueller et al., 2013). The LandFluxEVAL data are used in many studies as reference data for evaporation evaluations (Lorenz et al., 2014; Martens et al., 2017; Wartenburger et al., 2018). Since the LandFluxEVAL data are only available at 1-degree spatial resolution, the WAYS simulated evaporation is aggregated to 1 degree to match the resolution of the reference data. The evaluation is executed for 1989-2005 based on the availability of the LandFluxEVAL data. For the spatial evaluation, the WAYS simulation based on RZSC ($S_{R,CRU-SM}$) is used due to the global coverage of the RZSC product. For latitudinal comparison, both runs of WAYS simulated evaporation are used.

A promising relationship between WAYS simulated evaporation and LandFluxEVAL evaporation is found both in spatial pattern and in latitudinal average (see Figure S9). The generally high correlation coefficient (Figure S9, a) confirms the good performance of the WAYS model. However, relatively poor performance is also found in some regions in Europe, North America and South America (Amazon basin). It can also be seen that the spatial pattern of WAYS simulated annual averaged evaporation follows that of LandFluxEVAL data, while overestimations are found in regions, e.g., the Amazon basin and southeast Asia. The latitudinal evaluation shows that both WAYS simulations (WAYS_CRU and WAYS_CHIRPS) display a slight overestimation.

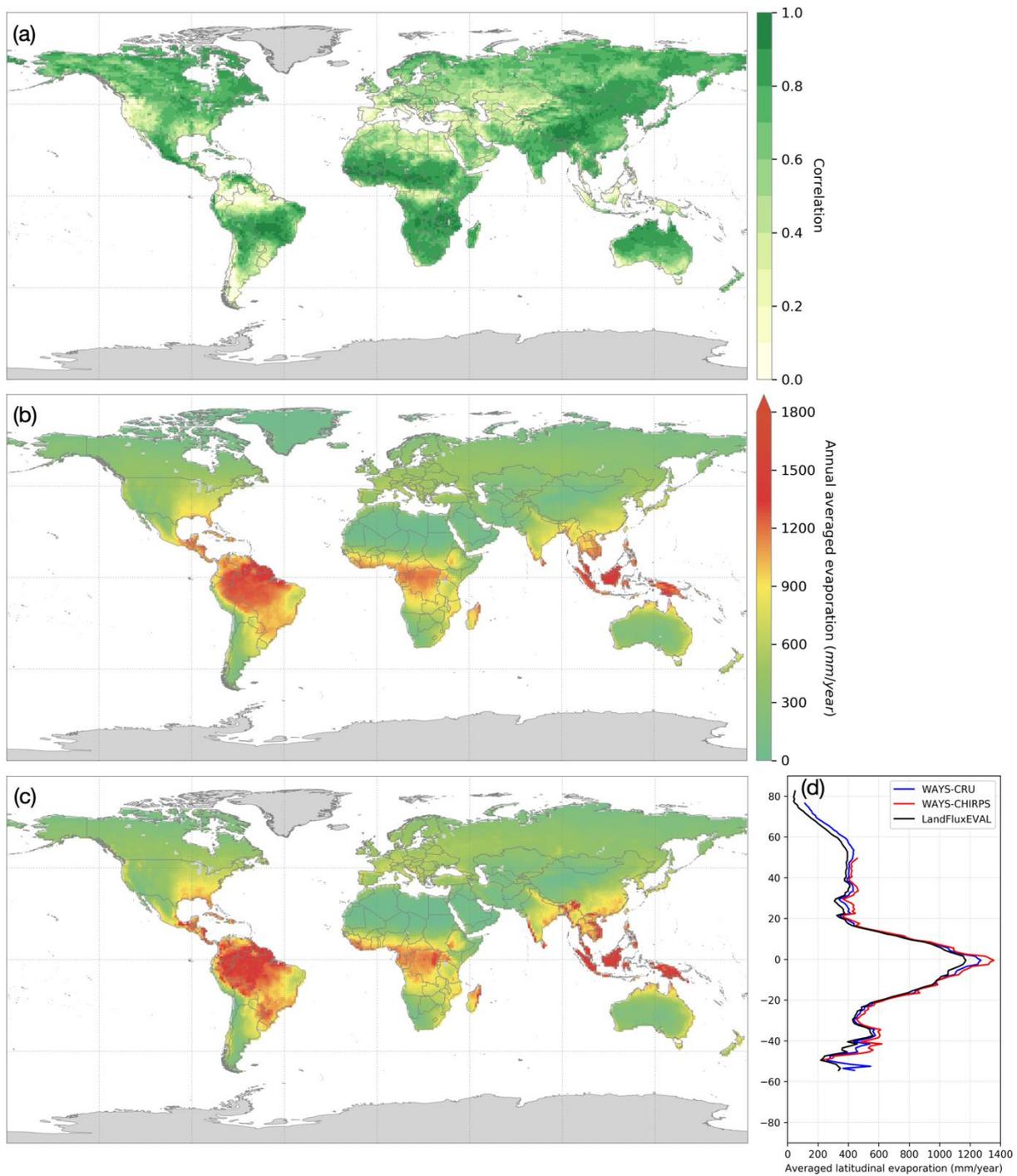


Figure S9. Validation results of the evaporation of WAYS simulation against the LandFluxEVAL data (1989-2005). (a) The calculated correlation coefficient between LandFluxEVAL data and WAYS simulation, (b) the annual averaged evaporation of LandFluxEVAL data, (c) the annual averaged evaporation of WAYS simulation based on RZSC $S_{R,CRU-SM}$, and (d) the comparison of the averaged latitudinal evaporation for WAYS model runs as well as the LandFluxEVAL data.

#6 Wang-Erlandsson et al., 2016 found that normalizing the root zone storage capacity using the Gumbel distribution by land cover type further improves performance, and recommended the use of Gumbel distribution. Please consider applying the Gumbel normalization to the root zone storage capacity data.

We incorporate the suggestion in the revised manuscript. We have now updated the RZSC data based on the Gumbel normalization. The optimized RZSC is calculated based on the suggested return period for each land cover by Wang-Erlandsson et al. (2016). Figure shows the flow chart of updating the RZSC data based on the Gumbel normalization with a different optimized return period. Since RZSC is a key parameter in the WAYS model, the model is recalibrated, and the simulations are also updated accordingly due to the change of RZSC.

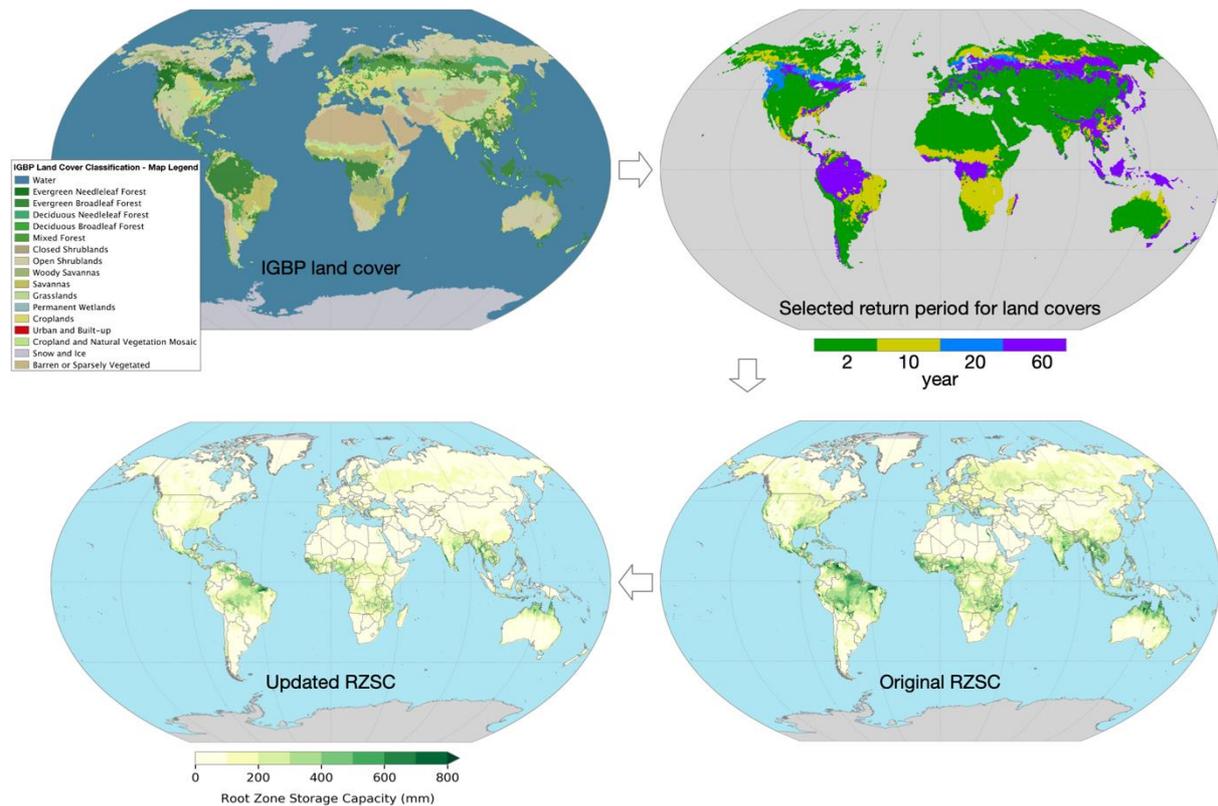


Figure 1. Flow chart of updating RZSC based on the Gumbel normalization

Authors' change in the manuscript.

Page 10, Line 15: (the changes are marked as blue)

Since Wang-Erlandsson et al. (2016) suggested that a Gumbel normalization of RZSC by land cover types with different return periods could further improve the model performance, we have accordingly adjusted the RZSC in this study. The two selected global root zone storage capacity products are shown in Figure S13, and their mean latitudinal values are shown in Figure S14.

#7 Please consider discussing how and where the results might be influenced by groundwater access and irrigation, noting that the root zone storage capacity in Wang-Erlandsson et al., 2016 was adjusted for irrigation but not access to groundwater, while WAYS do not account for either groundwater or irrigation.

Thank you for the comment. Indeed, the WAYS model does not consider the groundwater access and irrigation at the current stage. Although the capillary module is included in

WAYS, it is currently disabled due to the lack of information on the global groundwater table. The same strategy is also applied in other works, especially for hydrological simulation on the global scale (Döll et al., 2003; De Graaf et al., 2015; Hanasaki et al., 2018). Based on the experimental test on the capillary module, activation of the capillary module without groundwater information for capillary flux constrain is found to significantly increase the simulated evaporation. Nevertheless, ignorance of the capillary rise and irrigation are shortcomings of our study. This issue is now discussed in the revised manuscript.

Authors' change in the manuscript.

Page 25, Line 4: The following paragraph is inserted in the discussion part

Moreover, the current study does not consider the groundwater access and irrigation mainly due to the lack of global information. The groundwater table information is crucial for capillary rise simulation (Vergnes et al., 2014). Capillary rise simulation without proper water table information could significantly overestimate the evaporation. Thus, the capillary rise flux is ignored in this study. A similar strategy has also been applied by other works due to the absence of the information on the global water table (Döll et al., 2003; De Graaf et al., 2015; Hanasaki et al., 2018). Observations of irrigation on the global scale are also not available (Leng et al., 2015). Although there are simulated irrigation data available on the global scale, the inherent uncertainties could be propagated in our model simulation. Therefore, irrigation is also not considered at this time. However, this neglect could potentially introduce biases into the model simulation in irrigated areas and deep rooted plant-distributed regions, as both irrigation and capillary rise are an additional supply of soil water recharge. The biases may cause an underestimation of evaporation, especially in the dry summertime (Vergnes et al., 2014). This underestimation could consequently affect the simulation of RZWS and runoff because of the interlinkage of these three elements (Rockström et al., 1999). It is found that ignoring the capillary rise could reduce soil water content in the root zone (RZWS), while the runoff will also be reduced (Vergnes et al., 2014). However, these shortcomings can be simply overcome once the global data are available.

#8 Please provide the source code, and not only by request.

Thank you for the suggestion. We will make the code of the WAYS model, including all the parameters, freely available after the manuscript is accepted.

Specific comments

- 1. P1L10: state what was used for evaluating root zone storage (i.e., NDII) in the abstract.**

Thank you. The reference data for RZWS evaluation is stated in the abstract.

Authors' change in the manuscript.

Page 1, Line 10: (the changes is marked as blue)

The results show the ability of the model to mimic RZWS dynamics in most of the regions through comparison with proxy data, the Normalized Difference Infrared Index (NDII).

2. **P1L10: “many applications”: please provide concrete examples.**

Thank you. It is addressed accordingly.

Authors' change in the manuscript.

Page 1, Line 11: (the changes is marked as blue)

Compared to existing hydrological models, WAYS's ability to resolve the field-scale spatial heterogeneity of RZSC and simulate RZWS may offer benefits for many applications, e.g., agriculture and land-vegetation-climate interaction investigations.

3. **P1L11: “attention needs to also. . .”: hardly the most important limitation, please consider rather listing the more pressing future model developments needs and emphasize the key contribution of this model in comparison to other existing global hydrological models.**

Thank you. We have incorporated the suggestion of the referee and have emphasized the key contribution of this model in comparison to other existing global hydrological models. We have further listed the pressing future research needs in the abstract accordingly.

Authors' change in the manuscript.

Page 1, Line 11: (the changes is marked as blue)

Compared to existing hydrological models, WAYS's ability to resolve the field-scale spatial heterogeneity of RZSC and simulate RZWS may offer benefits for many applications, e.g., agriculture and land-vegetation-climate interaction investigations. However, the results from this study suggest an additional evaluation of RZWS is required for the regions where the NDII might not be the correct proxy.

4. **Please point out that Sriwongsitanon et al. (2016) is a study in a river basin in Thailand and not a global study.**

Thank you. It is addressed accordingly.

Authors' change in the manuscript.

Page 2, Line 26: (the changes is marked as blue)

Recently, Sriwongsitanon et al. (2016) investigated the relation between root zone water storage and the Normalized Difference Infrared Index and found a promising correspondence between them in a river basin in Thailand, especially in the dry seasons, where water stress exists. However, a global scale study has been absent in the literature.

5. **P6L27 “no information is available at the global scale”: Please consider including a few more lines describing the issues related to capillary rise modelling in global scale models and include related references, such as (Vergnes, Decharme, and Habets 2014) and references within.**

Thank you. It is addressed accordingly.

Authors' change in the manuscript.

To avoid repetition, please to see the changes in response to comment 7

6. **P8L28, "it has been well-justified (de Boer-Euser et al., 2019)": please consider specifying what is justified and add other relevant sources, e.g. "the method has been shown to increase model performance at both basin and global scale (e.g., de Boer-Euser et al., 2016, 2019, Gao et al. 2014, Wang-Erlandsson et al., 2016, Nijzink et al., 2016)".**

Thank you. It is addressed accordingly.

Authors' change in the manuscript.

Page 9, Line 28: (The changes are marked in blue)

This method has been well justified (de Boer-Euser et al., 2019) and overcomes the shortcomings of the traditional methods (look-up table approach; field observation-based approach) at the global scale, such as data scarcity, location bias, and risks of unlikely vegetation and soil combinations due to data uncertainty (Feddes et al., 2001). **The method has been shown to increase the model performance at both the basin and global scales (Gao et al., 2014b; Nijzink et al., 2016; Wang-Erlandsson et al., 2016). Moreover, it has been proven to be able to produce plausible root zone storage capacity in boreal regions by investigating the relationship between RZSC and numerous environmental factors, including climate variables, vegetation characteristics, and catchment characteristics (de Boer-Euser et al., 2019).**

7. **P14L11, "reported in his work": please change to "reported in their work".**

Thank you. It is corrected.

8. **P22L6 "DNII", should be NDII.**

Thank you. It is corrected.

References

Vergnes, J.-P., B. Decharme, and F. Habets. 2014. "Introduction of Groundwater Capillary Rises Using Subgrid Spatial Variability of Topography into the ISBA Land Surface Model." *Journal of Geophysical Research: Atmospheres* 119(19): 11,065-11,086.

Cited in the revised manuscript.

All the references are included in the manuscript.