Supplement of Hydrol. Earth Syst. Sci., 26, 1089–1109, 2022
https://doi.org/10.5194/hess-26-1089-2022-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

The importance of vegetation in understanding terrestrial water storage variations

Tina Trautmann et al.

Correspondence to: Tina Trautmann (ttraut@bgc-jena.mpg.de)

Supplement

S1 Spatial Pattern of scaled parameters in VEG

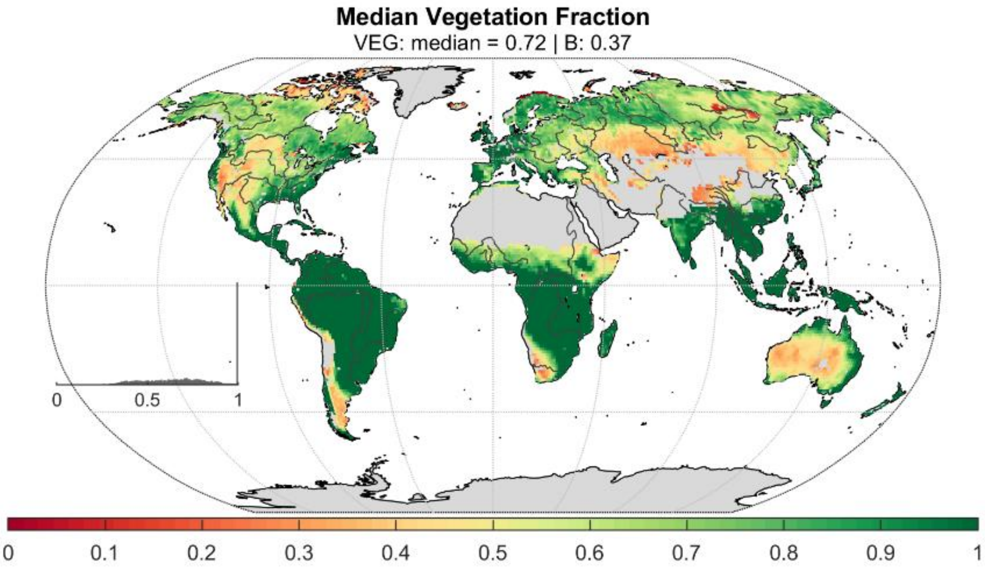

Figure S1 Global distribution of the median vegetation fraction *p_{Veg}* after calibration of the VEG experiment.

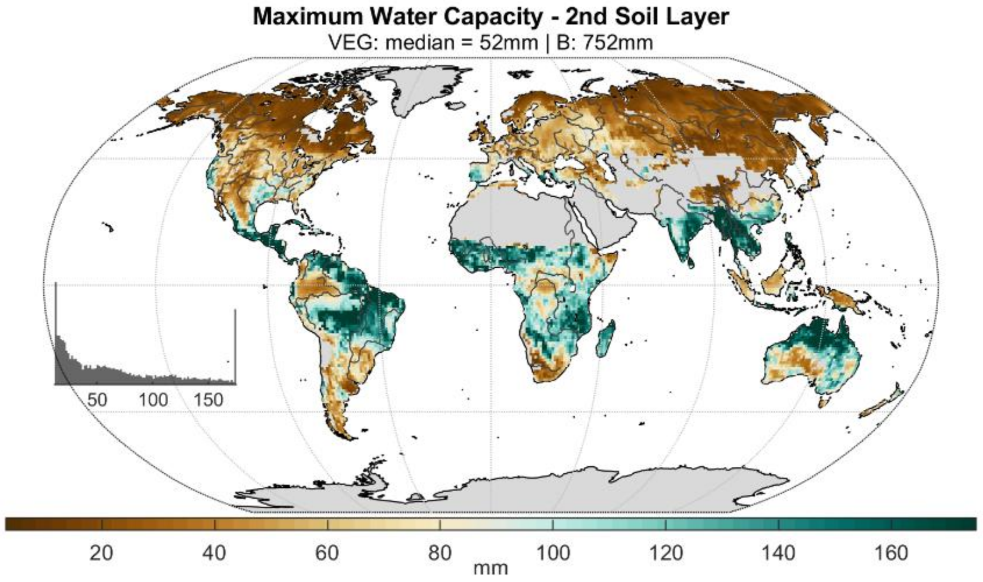

Figure S2 Global distribution of the maximum water capacity of the 2nd soil layer *$wSoil_{max(2)}$* after calibration of the VEG experiment.

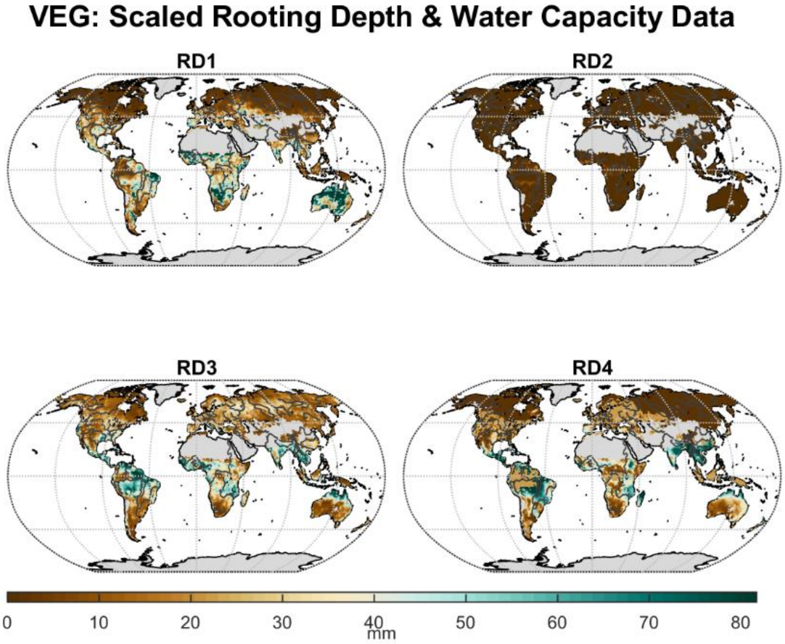

Figure S3 Global distribution of the maximum water capacity of the 2nd soil layer contributed by each data stream after calibration of their scaling parameters in the VEG experiment. RD1 = maximum rooting depth by Fan et al. 2017; RD2 = effective rooting depth by Yang et al. 2016; RD3 = maximum soil water capacity by Wang-Erlandsson et al. 2016; RD4 = plant available water capacity by Tian et al. 2019.

S2 Effective Alpha Coefficient

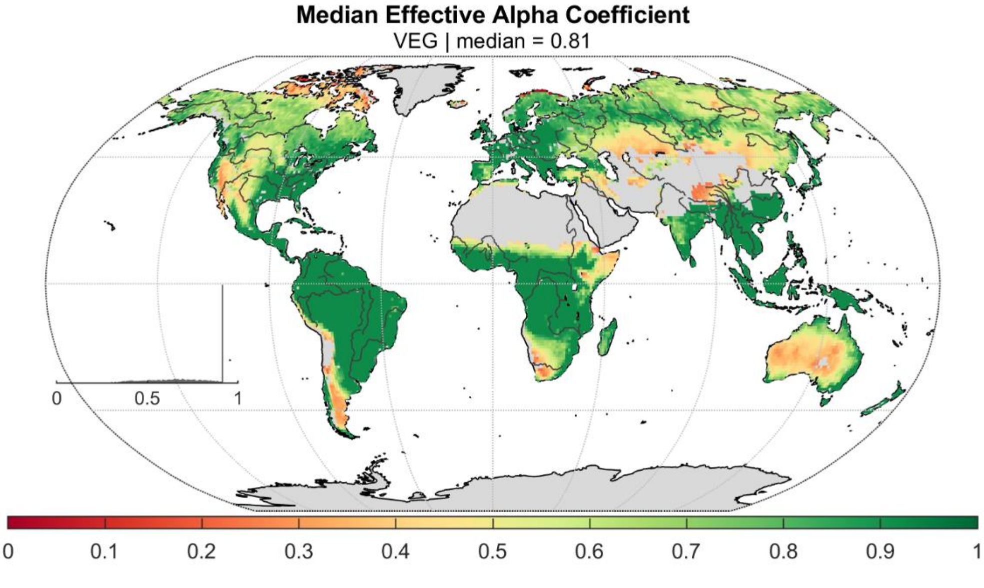

Figure S4 Global distribution of the median effective alpha coefficient (α_{Veg} * p_{Veg}) in the Priestley-Taylor formula after calibration of the VEG experiment.

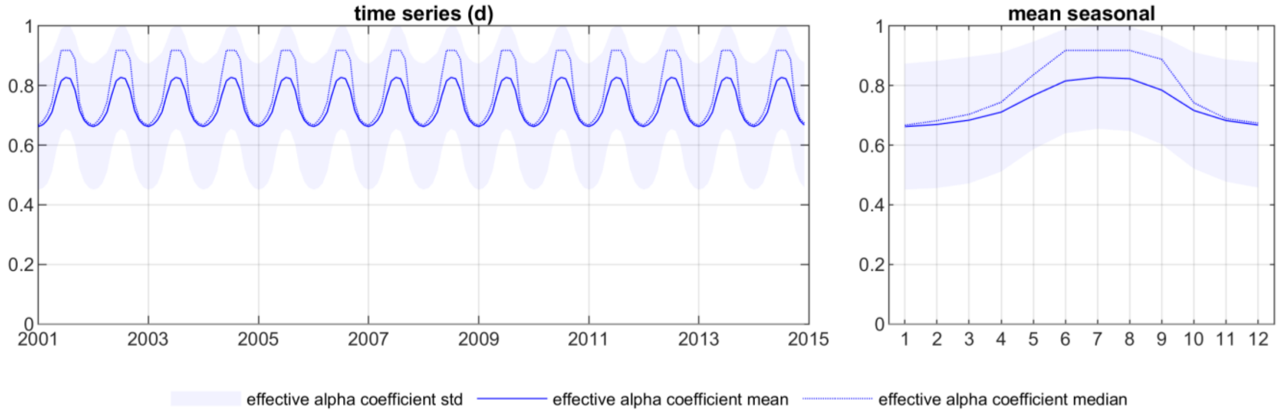

Figure S5 Daily time series and mean seasonal dynamics of the area weighted average, median and standard deviation of the grid-wise effective alpha coefficient in the Priestley-Taylor formula of the calibrated VEG experiment.

S3 Parameter Correlation

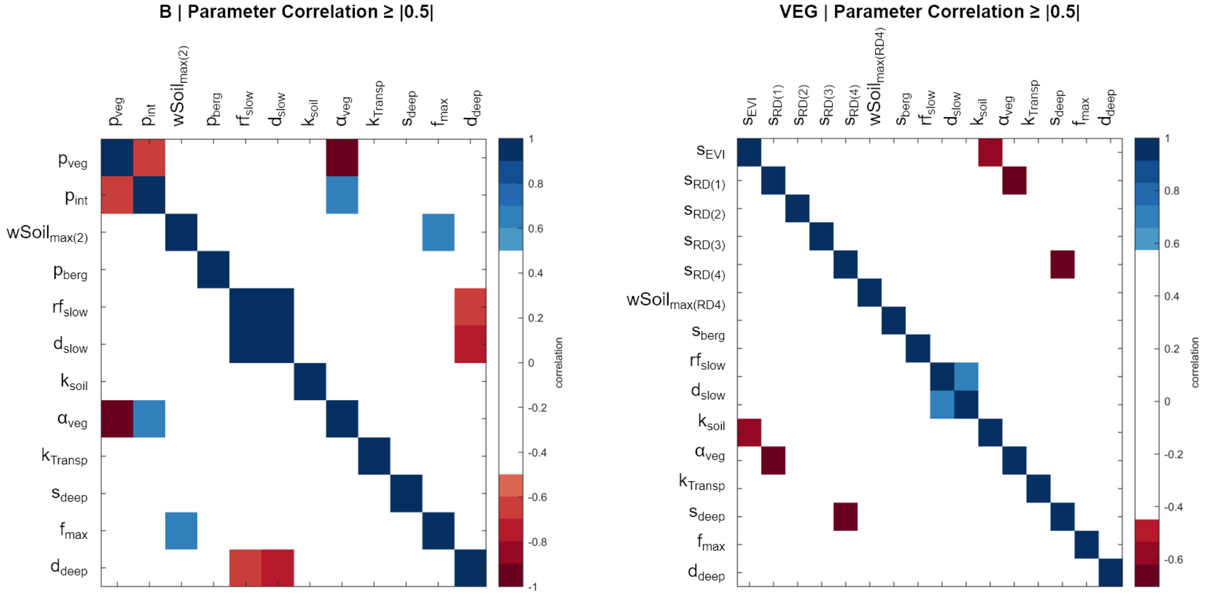

Figure S6 Correlation (≥ |0.5|) between model parameters for the B and VEG experiment.

S4 Regional IAV TWS Composition

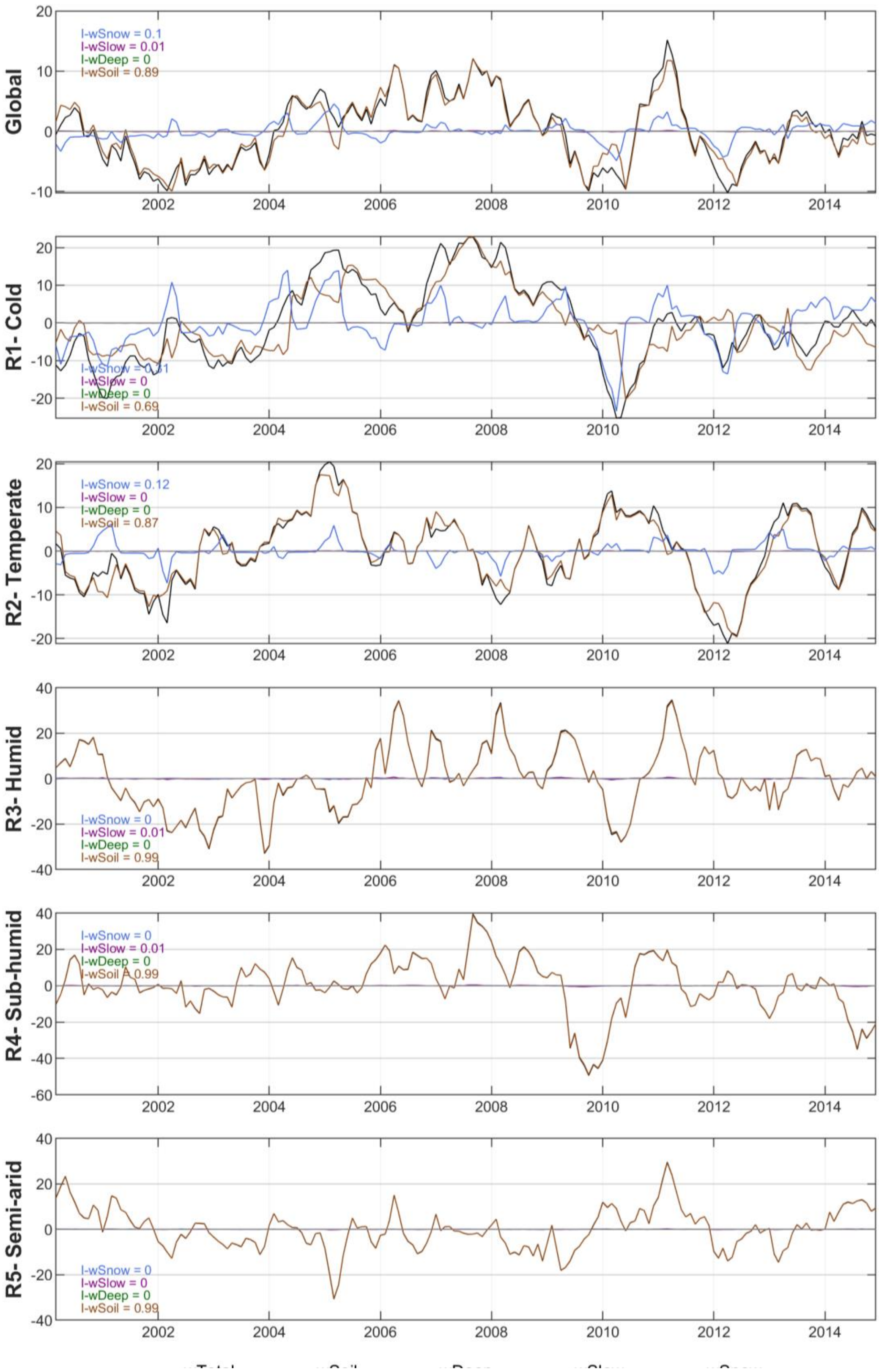

Figure S7 Global and regional average inter-annual variability of simulated total water storage (*wTotal*) and its components (*wSoil*, *wDeep*, *wSlow*, *wSnow*) for B, including the regional Impact Index *I* for each storage.

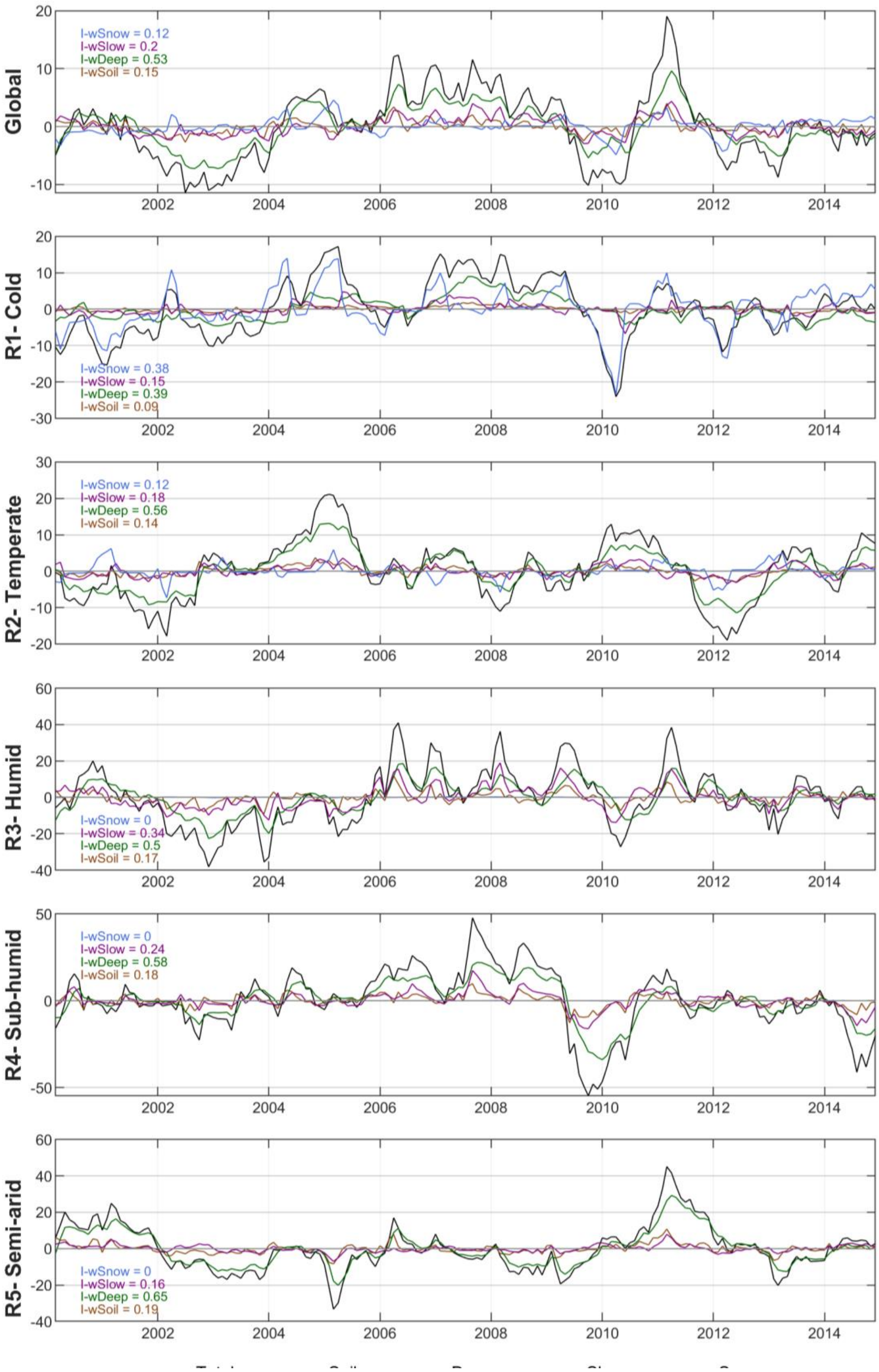

Figure S8 Global and regional average inter-annual variability of simulated total water storage (*wTotal*) and its components (*wSoil*, *wDeep*, *wSlow*, *wSnow*) for VEG, including the regional Impact Index *I* for each storage.

S5 T over ET

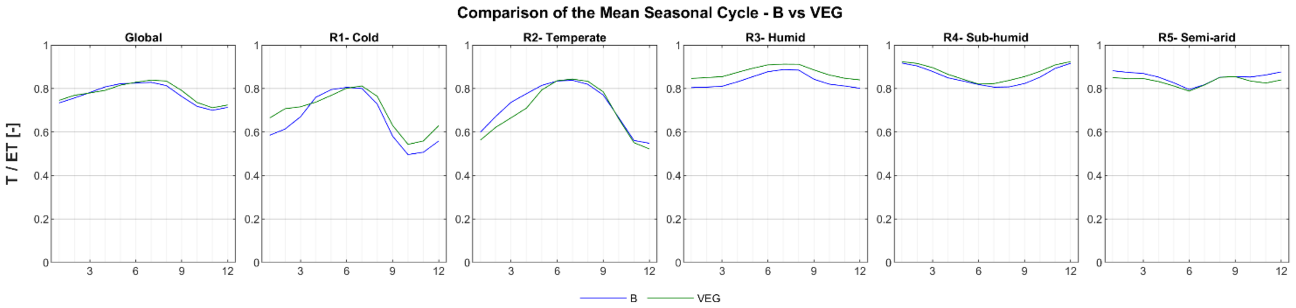

Figure S9 Global and regional average mean seasonal cycles of modelled transpiration (*T*) over evapotranspiration (ET) for B and VEG experiments.

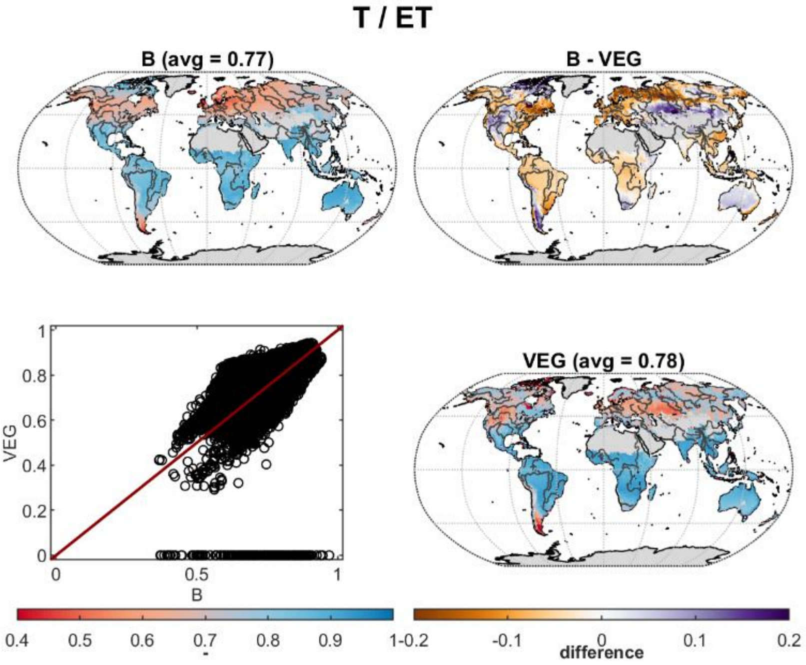

Figure S10 Global distribution of modelled transpiration (*T*) over evapotranspiration (ET) for B and VEG experiments, as well as the difference between both (lower right).

S6 Q Components

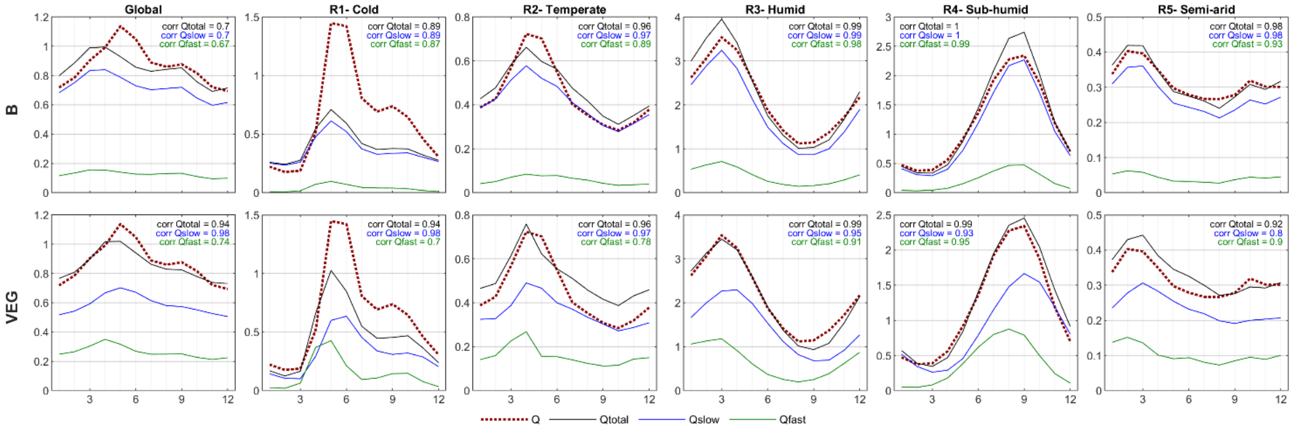

Figure S11 Global and regional average mean seasonal cycle of observed grid-wise runoff from GRUN (Q) and simulated total runoff (*Qtotal*), as well as its components *Qslow* and *Qfast*, for the B and VEG experiments. corr is the Pearson correlation coefficient of the respective simulation with observed Q.

S7 Comparison of VEG & VEG without capillary rise

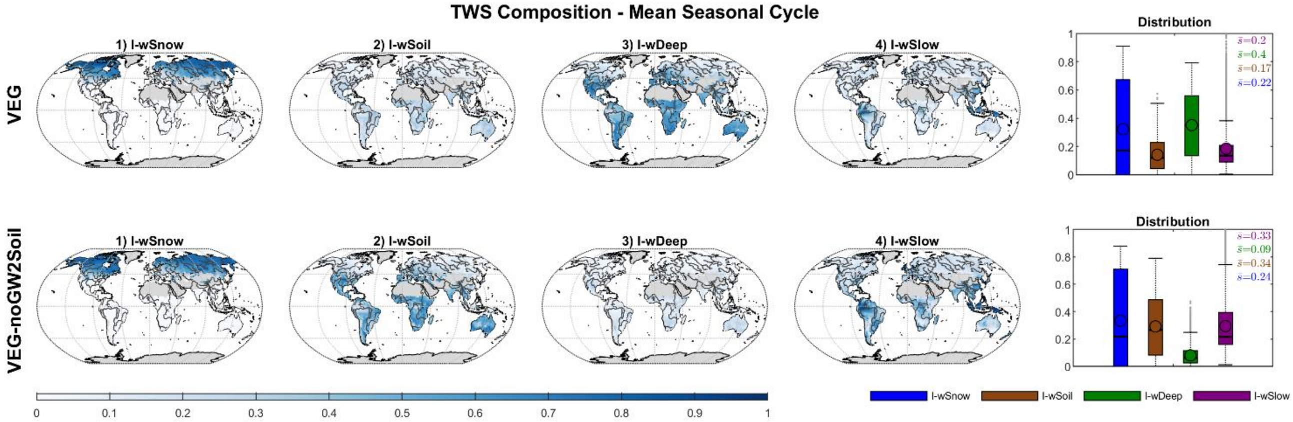

Figure S12 Global distribution of the Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the mean seasonal cycle of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from *wDeep* to *wSoil* is turned off prior to model calibration.

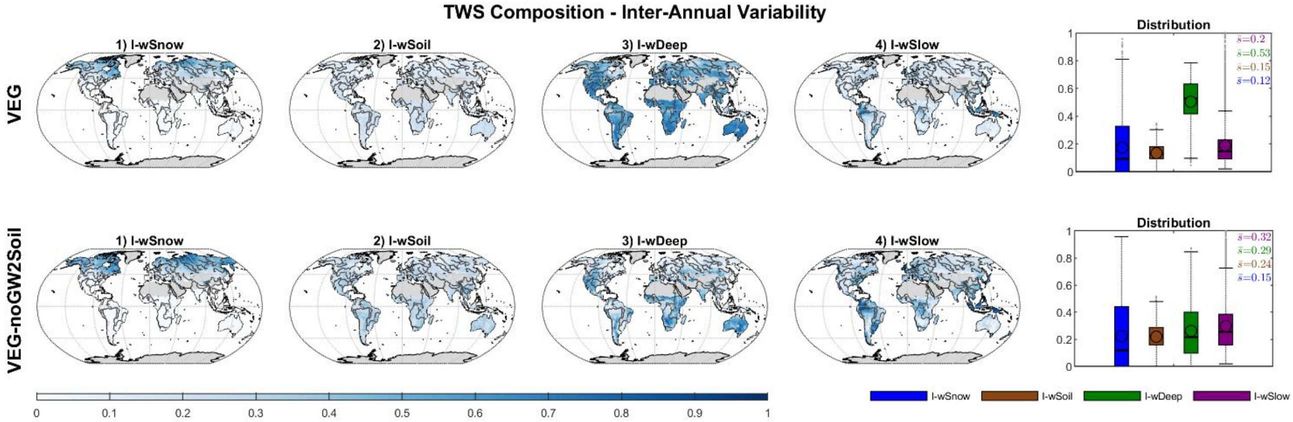

Figure S13 Global distribution of the Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the inter-annual anomalies of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from *wDeep* to *wSoil* is turned off prior to model calibration.

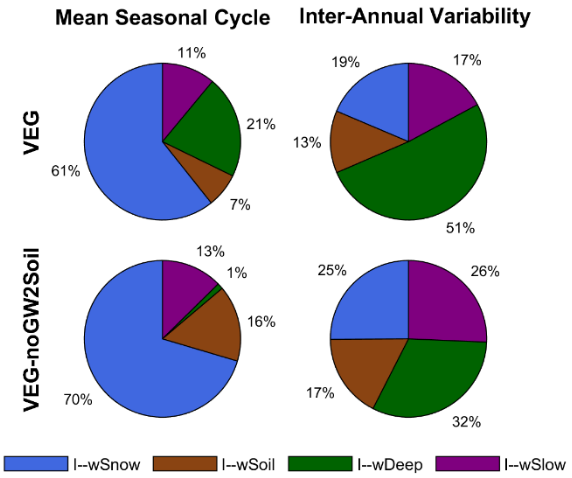

Figure S14 Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the global average mean seasonal cycle and inter-annual variability of total water storage, for VEG and VEG-noGW2Soil, which is a variant of the VEG experiment, in with the capillary rise from *wDeep* to *wSoil* is turned off prior to model calibration.

S8 Comparison of VEG & VEG with fixed k_{Transp} at 0.05

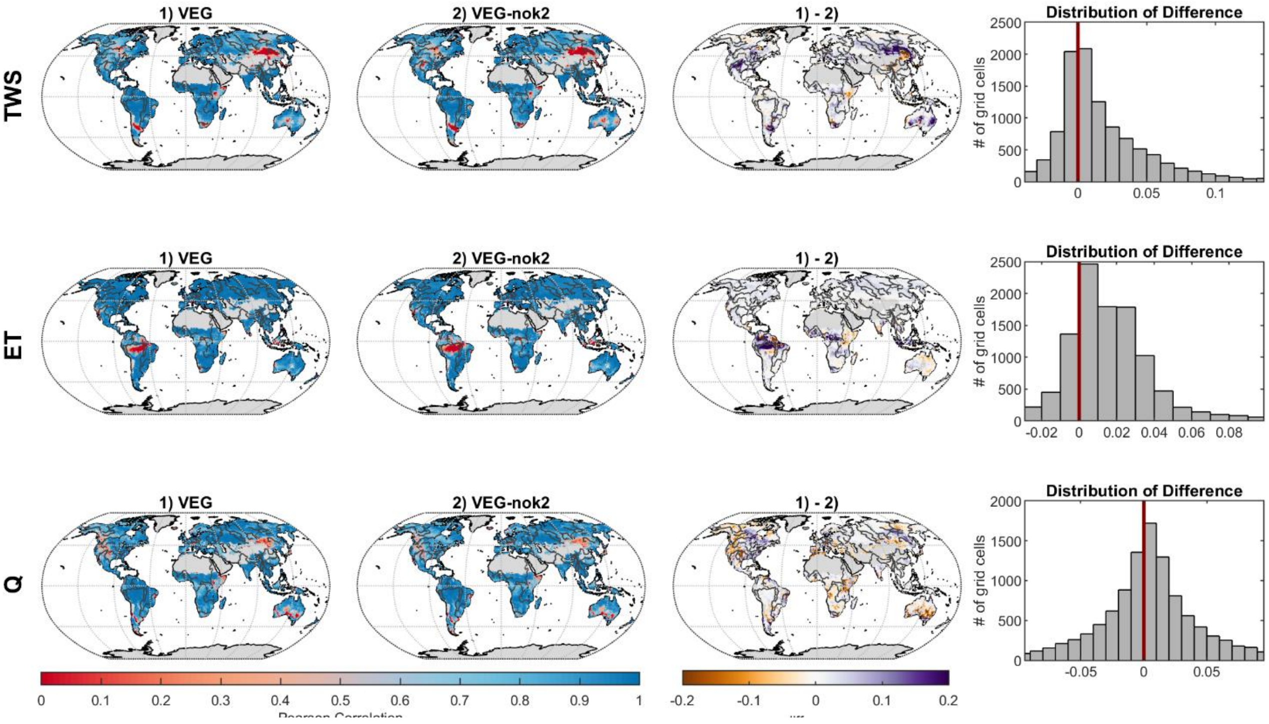

Figure S15 Grid-wise Pearson's correlation coefficient for total water storage (TWS), evapotranspiration (ET) and runoff (Q) between 1) observations and VEG, and 2) observations and VEG-nok2, as well as differences between 1) and 2) (brown color, i.e., negative values indicate higher correlations for VEG-nok2, while purple color, i.e., positive values indicate better correlation values for VEG). VEG-nok2 is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

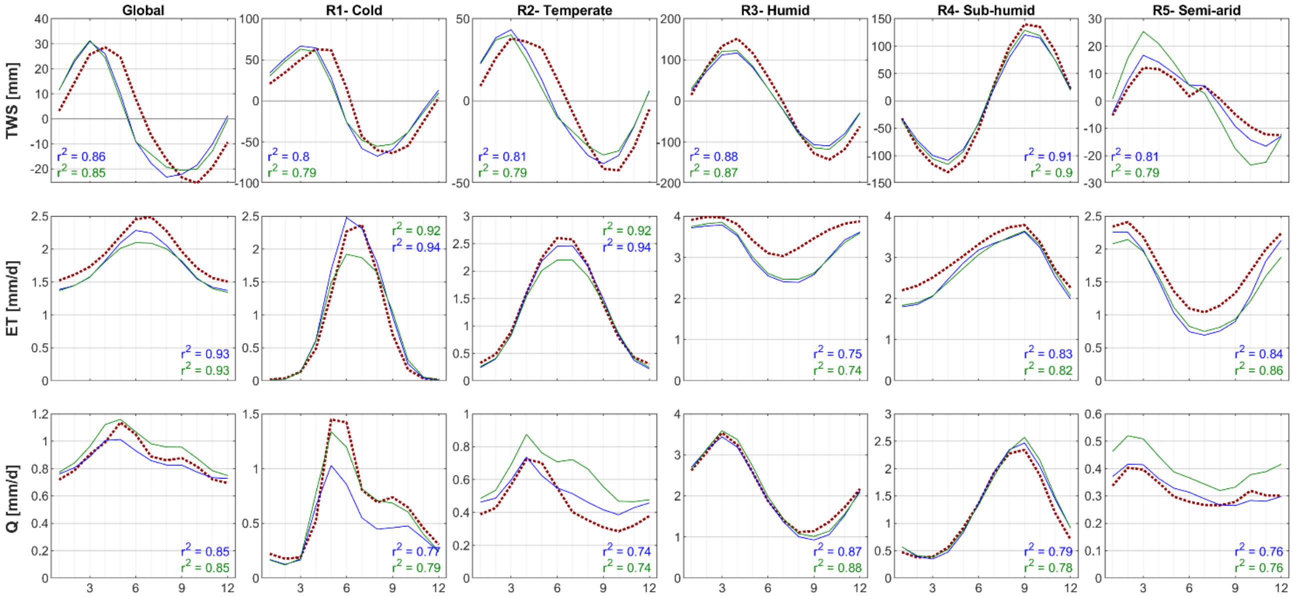

Figure S16 Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05, compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q).

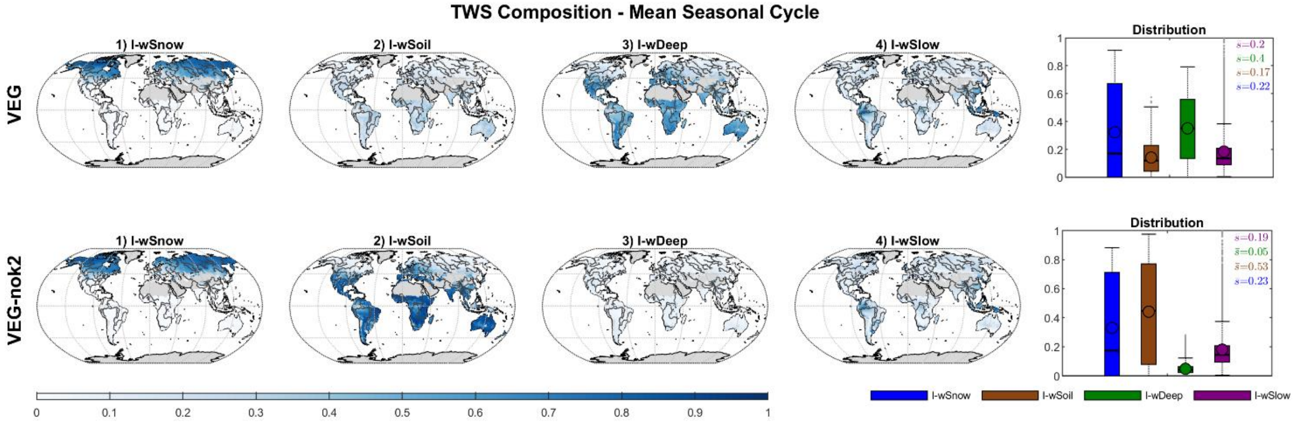

Figure S17 Global distribution of the Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the mean seasonal cycle of total water storage, for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

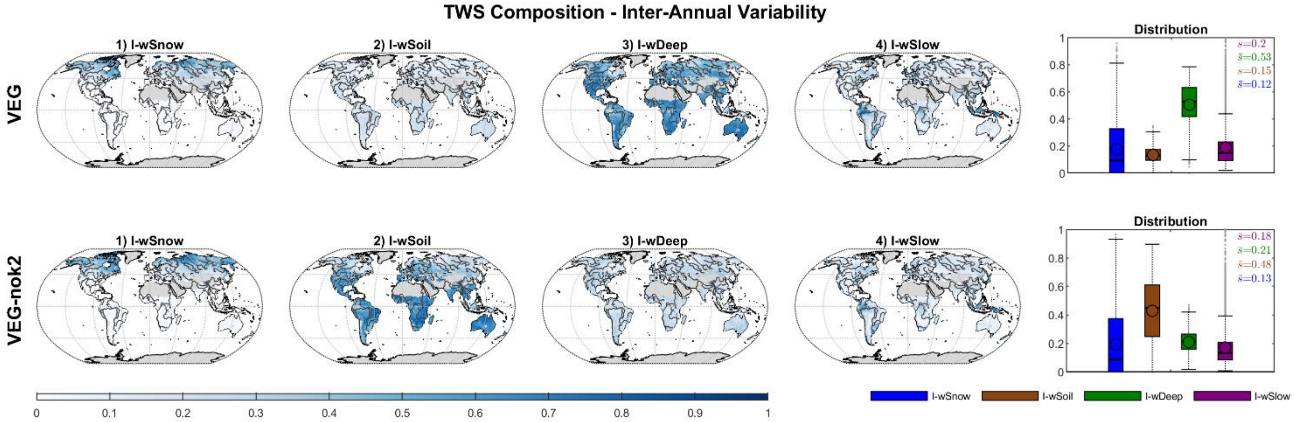

Figure S18 Global distribution of the Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the inter-annual anomalies of total water storage, for VEG and VEG-VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

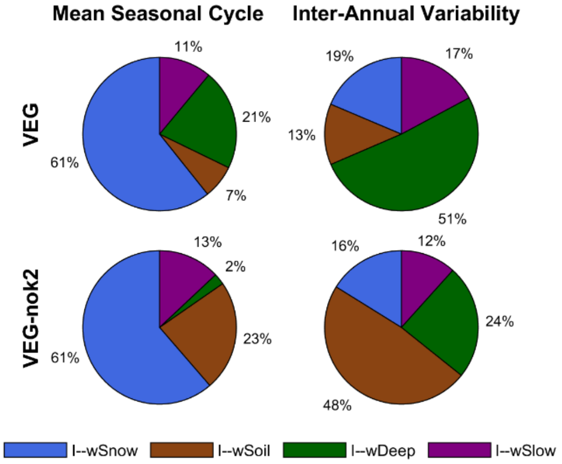

Figure S19 Impact Index *I* for the contribution of simulated snow (*wSnow*), soil (*wSoil*), deep water storage (*wDeep*) and delayed water storage (*wSlow*) to the global average mean seasonal cycle and inter-annual variability of total water storage, for VEG and VEG-nok2, which is a variant of the VEG experiment, in which the k_{Transp} parameter is not calibrated but fixed at a low value of 0.05.

S9 PFT experiment

The following shows an experiment similar to the traditional approach of global hydrological models, in which vegetation-dependent parameters are defined and calibrated for different plant-functional type (PFT) classes separately and then model performance and TWS composition is analyzed in comparison to the **B** and **VEG** experiments. The results show that the larger number of parameters (due to different sets for different PFT) does not lead to marked improvements of model performance, but instead increases parameter uncertainty possibly due to overparameterization. In terms of TWS composition, we see substantial differences in the **PFT** experiment compared to **B** and **VEG,** which underlines our conclusions that the representation of vegetation in GHMs is critical for interpreting TWS variations.

Based on the GSWP2 land cover classification (Dirmeyer et al. 2006), we consider 12 PFT classes (Fig. S20), for which we define individual values of $wSoil_{max(2)}$ (maximum available water capacity of the 2nd soil layer) and s_{berg} (scaling parameter to derive the runoff/infiltration coefficient). Since state-of-the-art global hydrological models (GHMs) usually include seasonal dynamics of leaf area index (LAI) to calculate, e.g., transpiration, we decided to keep the definition of the active vegetation fraction as a function of seasonal EVI data as in the **VEG** experiment. For the **PFT** experiment, we focus (i) on $wSoil_{max(2)}$ because GHMs usually apply a PFT specific rooting depth, and (ii) on s_{berg} because this is similar to the runoff coefficient γ which is tuned in some GHMs (e.g., the WaterGAP model (Müller Schmied et al. 2021)).

When considering these 12 PFT classes, the number of calibration parameters increases from 12 (in **B**) and 16 (in **VEG**) to 34 (in **PFT**). Analysis of parameter uncertainty shows high uncertainties for a set of parameters common with **B**, while optimized parameter values are between those of **B** and **VEG** (Table S1). Additionally, and unlike **B** and **VEG**, **PFT** has high uncertainty of $wSoil_{max(2)}$ for all PFT classes, and high correlation between each PFT's $wSoil_{max(2)}$ and s_{berg} (Fig. S21). High uncertainty of $wSoil_{max(2)}$ is an indication that having one $wSoil_{max(2)}$ per PFT may not explain the within-PFT variability. On the other hand, high correlation between each PFT's $wSoil_{max(2)}$ and s_{berg} is systematic, as both parameters are based on the same spatial distribution of PFT classes - and highlights an advantage of the **VEG** experiment, in which both are based on independent data sets.

In terms of model performance, Fig. S22 shows a partial improvement for TWS and ET in the **PFT** experiment. Especially in the *Humid* and *Sub-humid* regions, TWS simulation in **PFT** matches GRACE observations better. These regions include tropical regions, where data for maximum plant available water capacity by Tian et al. 2019 (RD4) are not available. While we filled the missing values for tropical regions with the same $wSoil_{max(RD4)}$ value as in the Northern latitudes, the improved performance in the **PFT** experiment suggests that at least 2 different $wSoil_{max(RD4)}$ fill values seem necessary for different climate regions. In contrast to TWS and ET, **PFT** performance of Q is poorer than in **B** and **VEG**, with a clear underestimation of the seasonal variability. To consider model performance in relation to the number of calibration parameters, we calculated the Akaike information criterion (AIC). Since low values of AIC indicate better performance compared to the other experiments, **PFT** only performs superior regarding ET, while the increased number of model parameters isn't advantageous regarding TWS and Q simulations. Also, note that the increased number of model parameters comes at an additional computational cost.

Furthermore, the results of the **PFT** experiment confirm that changing the representation of vegetation has a marked impact on the simulated TWS composition (Fig. S23-S25). In **PFT**, among the liquid water storages wSoil contributes most to mean seasonal TWS variability, with Impact Index values between those of **B** and **VEG** (Fig. S23, Fig. S25). Compared to **VEG**, wSlow is in general less important in **PFT**, while wDeep has a less impact on mean seasonal TWS, but its contribution to inter-annual TWS variability increases.

Overall, this analysis underlines that including continuous fields of vegetation parameters is preferable than the 'traditional' PFT-based approaches of defining parameters for distinct PFT classes (and their calibration) - in terms of model calibration and the uncertainty of calibrated model parameters, but also regarding model performance in relation to the number of model parameters. Furthermore, we could highlight that the representation of vegetation in hydrological models is crucial for the partitioning of simulated TWS.

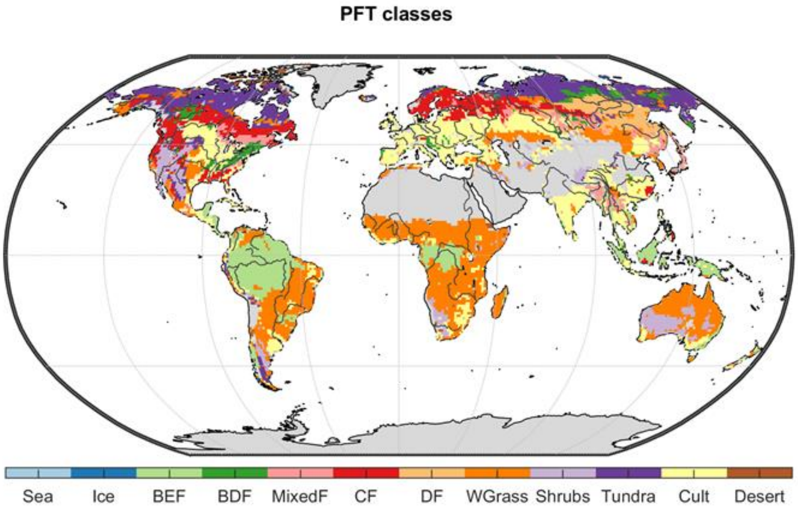

Figure S20 Classes of plant functional type used in the PFT experiment. (Sea (PFT0); Ice=Continental Ice (PFT1); BEF=Broadleaf Evergreen Forest (PFT2); BDF=Broadleaf Deciduous Forest & Woodland (PFT3); MixedF=Mixed Coniferous & Broadleaf Deciduous Forest & Woodland (PFT4); CF= Coniferous Forest & Woodland (PFT5); DF=High Latitude Deciduous Forest & Woodland (PFT6); WGrass= Wooded C4 Grassland (PFT7); Shrubs=Shrubs & Bare Ground (PFT8); Tundra (PFT9); Cult=Cultivation (PFT10); Desert (PFT11)).

Table S1 Calibrated parameter values and their uncertainty for B, VEG and PFT. Red font indicates a calibrated parameter that hits the parameter bounds, and red background indicates parameter uncertainty ≥ 20%.

parameter	**calibrated values ± uncertainty**					
	B		**VEG**		**PFT**	
vegetation fraction						
p_{veg}	0.37	± 0.05				
s_{EVI}			3.89	± 0.05	3.75	± 0.03
evapotranspiration						
p_{int}	1	± 0.08	0.6	± 0.02	0.71	± 0.02
k_{Soil}	0.1	± 0.01	0.4	± 0.08	0.27	± 0.04
α_{veg}	2.25	± 0.15	0.92	± 0.00	0.87	± 0
k_{Transp}	0.12	± 0.32	0.48	± 1.76	0.5	± 4.32
deep soil						
s_{Deep}	9.1	± 461317	5.6	± 0.21	8.48	± 0.24
f_{max}	1.5	± 0.00	5.1	± 0.01	11.77	± 0.02
d_{Deep}	1	± 5.61	0.01	± 0.00	0.03	± 0
delayed water storage						
rf_{Slow}	0.78	± 1.72	0.68	± 0.01	0.62	± 0.05
d_{Slow}	1	± 2329	0.02	± 0.03	0.03	± 0.19
infiltration/runoff						
p_{berg}	1.32	± 0.02				
s_{berg}			3.08	± 0.02		
s_{berg_PFT0}					3.7	± 0.45
s_{berg_PFT1}					3.11	± 0.32
s_{berg_PFT2}					1.87	± 0.01
s_{berg_PFT3}					2.57	± 0.09
s_{berg_PFT4}					2.04	± 0.03
s_{berg_PFT5}					4.31	± 0.05
s_{berg_PFT6}					0.5	± 0.01
s_{berg_PFT7}					2.9	± 0.03
s_{berg_PFT8}					0.48	± 0.01
s_{berg_PFT9}					0.69	± 0.01
s_{berg_PFT10}					1.36	±0.01
s_{berg_PFT11}					2.5	± 0.11
soil moisture						
$wSoil_{max(2)}$	752	± 0.02				
$s_{RD(1)}$			0.01	± 0.00		
$s_{RD(2)}$			0	± 0.00		
$s_{RD(3)}$			0.15	± 0.06		
$s_{RD(4)}$			0.15	± 0.07		
$wSoil_{max(RD4)}$			145	± 0.08		
$wSoil_{max\_PFT0}$					1.57	± 8.94
$wSoil_{max\_PFT1}$					0.78	± 10.23
$wSoil_{max\_PFT2}$					1.01	± 0.41
$wSoil_{max\_PFT3}$					1.27	± 1.42
$wSoil_{max\_PFT4}$					0.5	± 0.5
$wSoil_{max\_PFT5}$					0.54	± 0.32
$wSoil_{max\_PFT6}$					0.85	± 2.53
$wSoil_{max\_PFT7}$					01.01	± 0.57
$wSoil_{max\_PFT8}$					1.45	± 2.72
$wSoil_{max\_PFT9}$					0.56	± 1.07
$wSoil_{max\_PFT10}$					0.39	± 0.2
$wSoil_{max\_PFT11}$					0.7	± 3.23

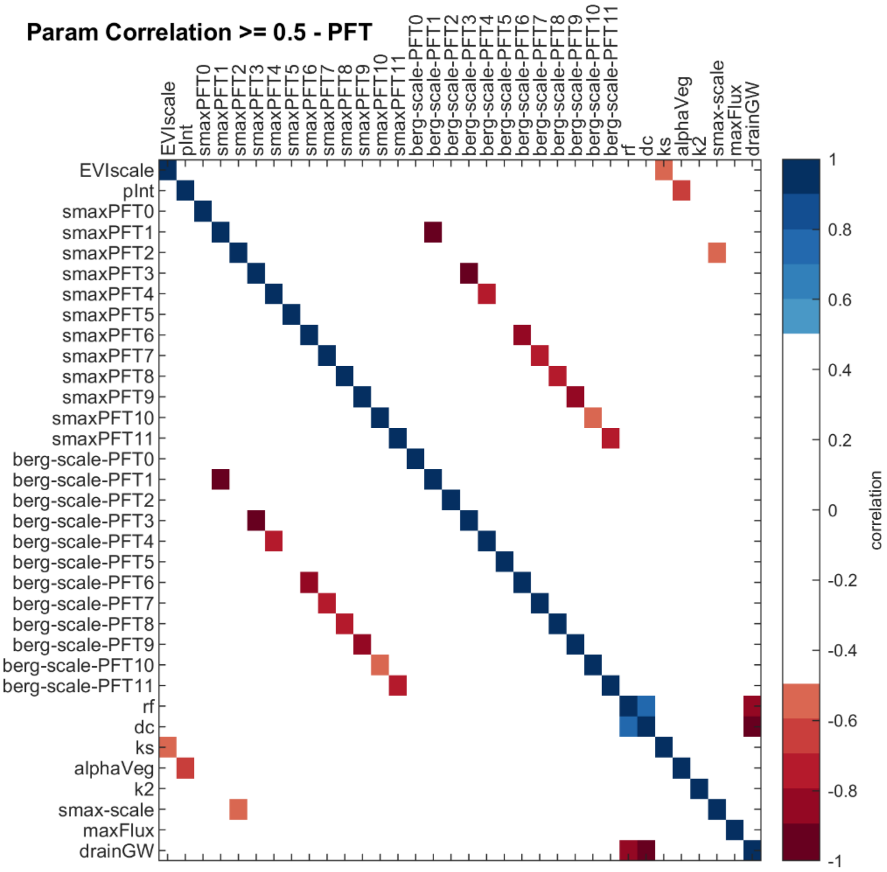

Figure S21 Correlation of calibrated parameters for the PFT experiment. Shown are only correlation coefficients |r|≥0.5.

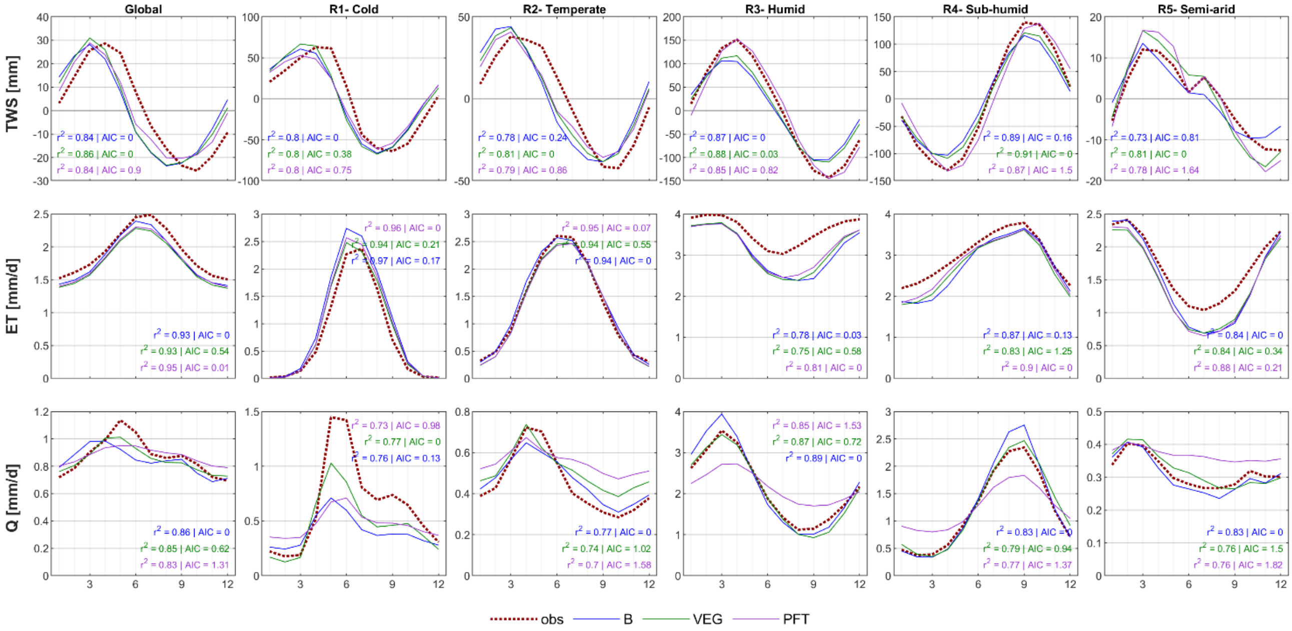

Figure S22 Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for the B, VEG and PFT experiments compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q). For each, the Pearson correlation (r^2) and Akaike information criterion (AIC) are calculated to compare model performance in terms of seasonal dynamics and of mean standard error in relation to the number of calibration parameters.

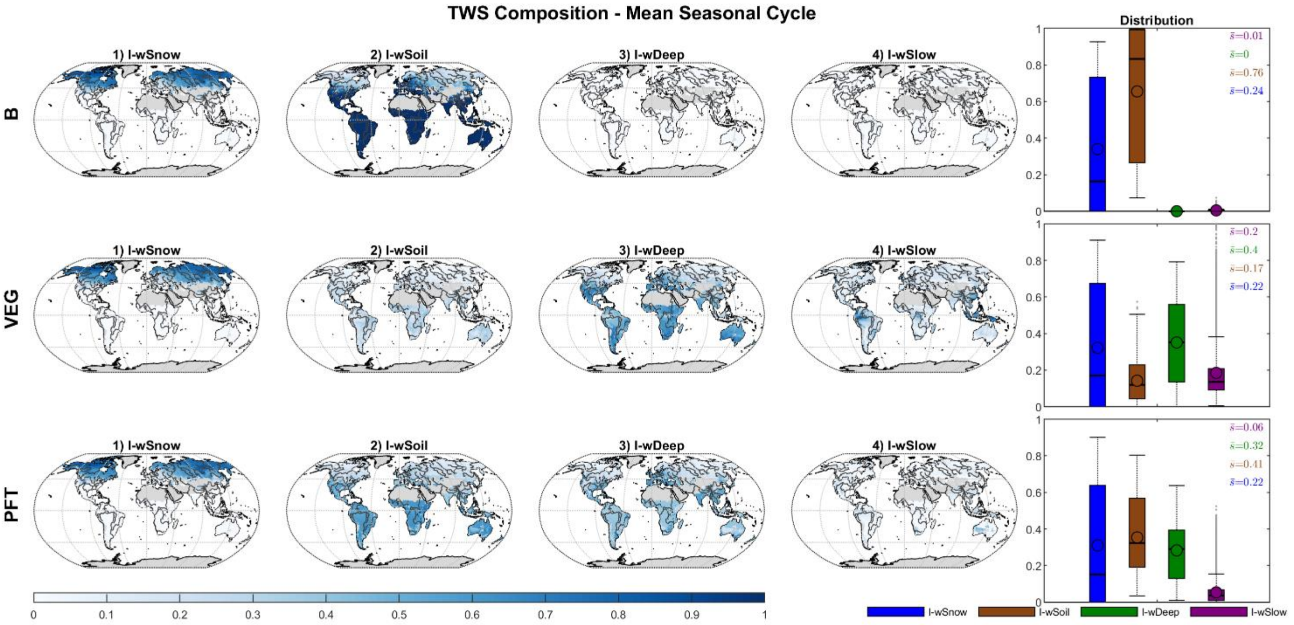

Figure S23 Global distribution of the Impact Index *I* for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the mean seasonal cycle of total water storage, for B, VEG and PFT.

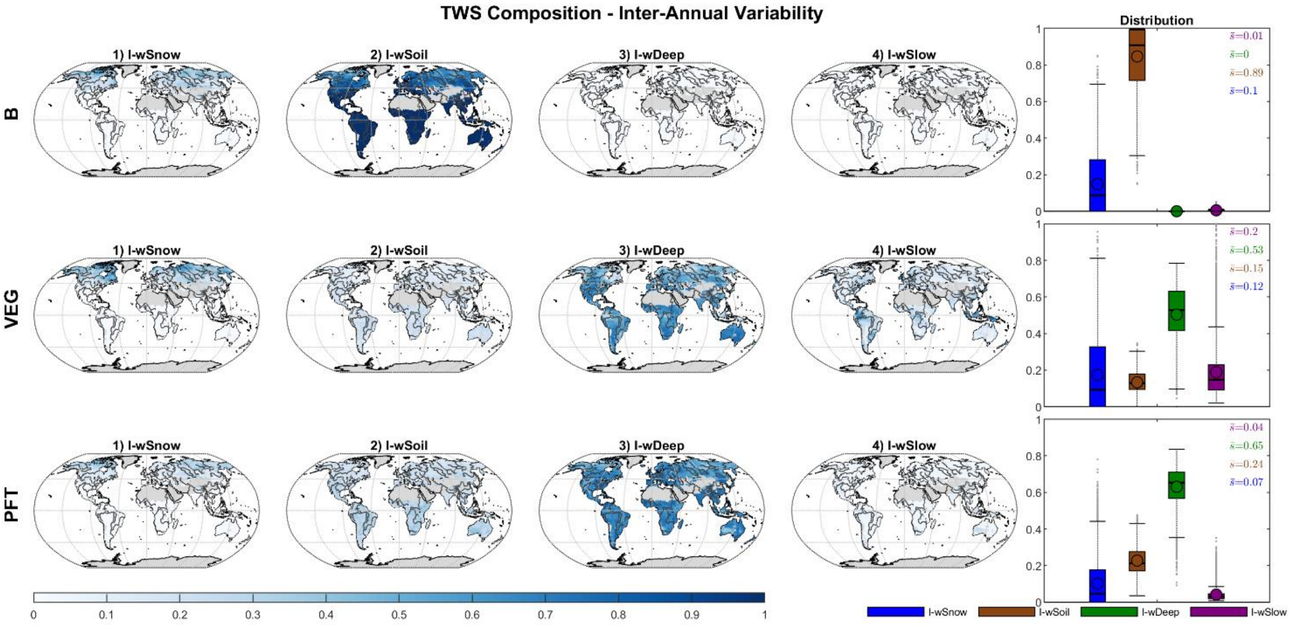

Figure S24 Global distribution of the Impact Index *I* for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the inter-annual variability of total water storage, for B, VEG and PFT.

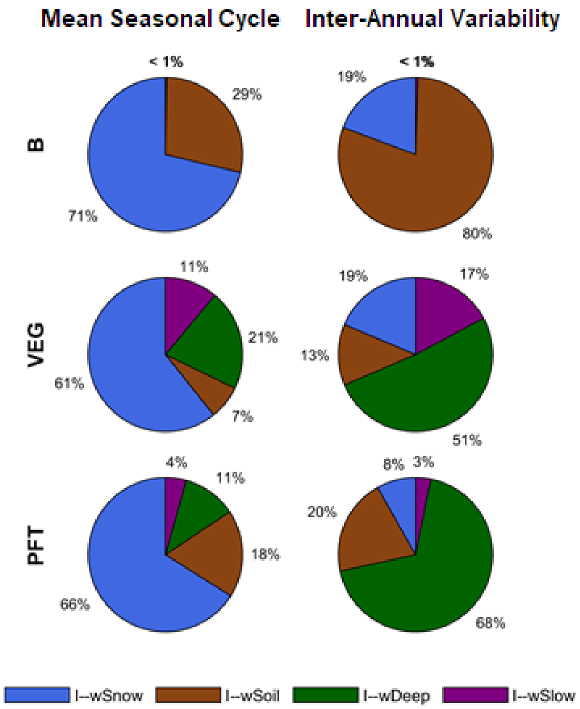

Figure S25 Impact Index *I* for the contribution of simulated snow (wSnow), soil (wSoil), deep water storage (wDeep) and delayed water storage (wSlow) to the global average mean seasonal cycle and inter-annual variability of total water storage, for B, VEG and PFT.

S10 Consistency Check of Observational Data

In the following, we check for possible inconsistencies between the different observational data products. Similar to Rodell et al. 2015, we calculate the monthly water (im)balance, WB, from the observations for the period 01/2004-11/2010 (the time period in which none of the observation data has missing monthly values):

$$WB = P_{GPCP1DD} - ET_{FLUXCOM} - Q_{GRUN} - dS_{GRACE} \qquad \text{Eq. (S1)}$$

with ideally WB = 0.

Fig. S26 shows the average monthly water imbalance scaled by each grid's average monthly precipitation $P_{GPCP1DD}$. While regionally large differences exist, the global mean and median are around 0. The global mean value of - 0.05 corresponds to a water balance residual of ~ 5% of precipitation - which is similar to the global residual of 4.3 % of precipitation reported in Rodell et al. 2014. Also temporally, the global average (Fig. S27) varies around 0, suggesting no major systematic inconsistency at the global scale, yet with a small imbalance with a tendency to negative values. This suggests that more water leaves the system than comes in when looking at the observational data. In comparison, there is obviously no imbalance for the simulations from **B** and **VEG** as they close the water balance by definition of the model - which represents the major advantage of using models instead of observational based data from different sources.

We also calculated each variable in Eq. (S1) by solving the water balance with the other observed components and we compared the resulting water-balance-derived variable with the actual observed one. Differences between both indicate inconsistencies between a particular observed variable and the remaining observational variables. For ET, Q and TWS, we additionally plot the modelled fluxes and storage changes from **B** and **VEG** to evaluate the effect of observational inconsistencies on model simulations (Fig. S27). The modelled fluxes are smoother and closer to the observations than the same estimate of the variable from the water balance. Therefore, we find that the model allows to potentially bridge the inconsistencies between the different data products. However, for dS, **B** and **VEG** show a time shift compared to the observed storage change, that is not reflected in dS calculated from P, ET and Q observations. Accordingly, this underlines that the phase lag between observed and modelled TWS variations is not caused by data inconsistencies, but rather related to the potential deficiencies in the model structure, as already discussed in the main text of the manuscript.

Fig. S28 compares the residuals of the simulated and observed ET, Q and dS (*mod-obs*), and the ones of the water-balance derived and the observed variables (*WB-obs*). Large residuals of *WB-obs* point again to data inconsistencies among the observed variables. When the residuals *WB-obs* and *mod-obs* in a region agree, it implies that the multi-criteria calibration approach prevents overfitting of the model(s) to an observed variable that is inconsistent with the remaining observed variables. Therefore, the model performance in these regions might be relatively poor in view of the inconsistent data streams, which is in fact a desirable behavior in the model calibration (e.g., ET in the *Semi-arid* region and dS in *Temperate* and *Humid* region).

When the residuals of *mod-obs* are considerably smaller than *WB-obs*, the model fits an observed variable well although it is inconsistent with the remaining observed variables (e.g., *Q* and dS in the *Semi-arid* region). Further, when the residuals of *mod-obs* are large but *WB-obs* does not indicate data inconsistencies, it points to issues related to model structure and parameter identifiability (e.g., Q in the *Cold* region, where the model(s) lacks the representation of permafrost, freeze/thaw dynamics and ice jam in rivers).

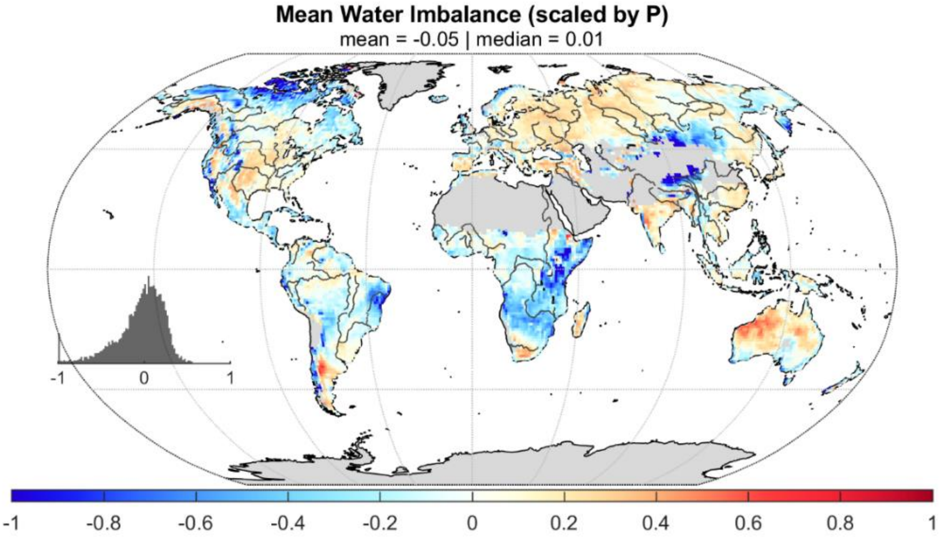

Figure S26 Mean water imbalance scaled by mean precipitation.

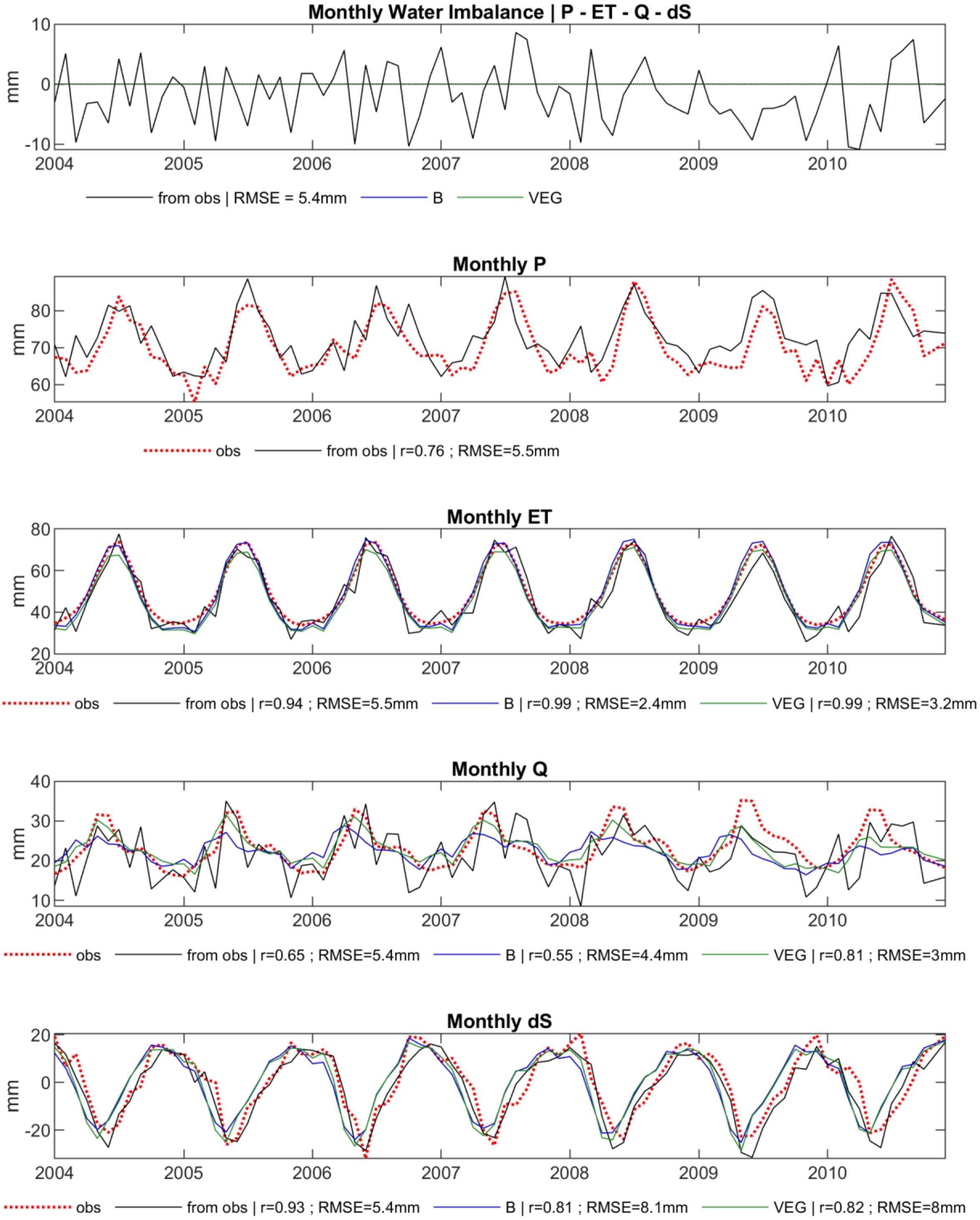

Figure S27 Global average time series of the water imbalance calculated from the observations (top row), and of water balance variables calculated from the other observations by resolving the water balance equations (from obs) vs the observed variable (obs) vs the simulated variable of the B and VEG simulations.

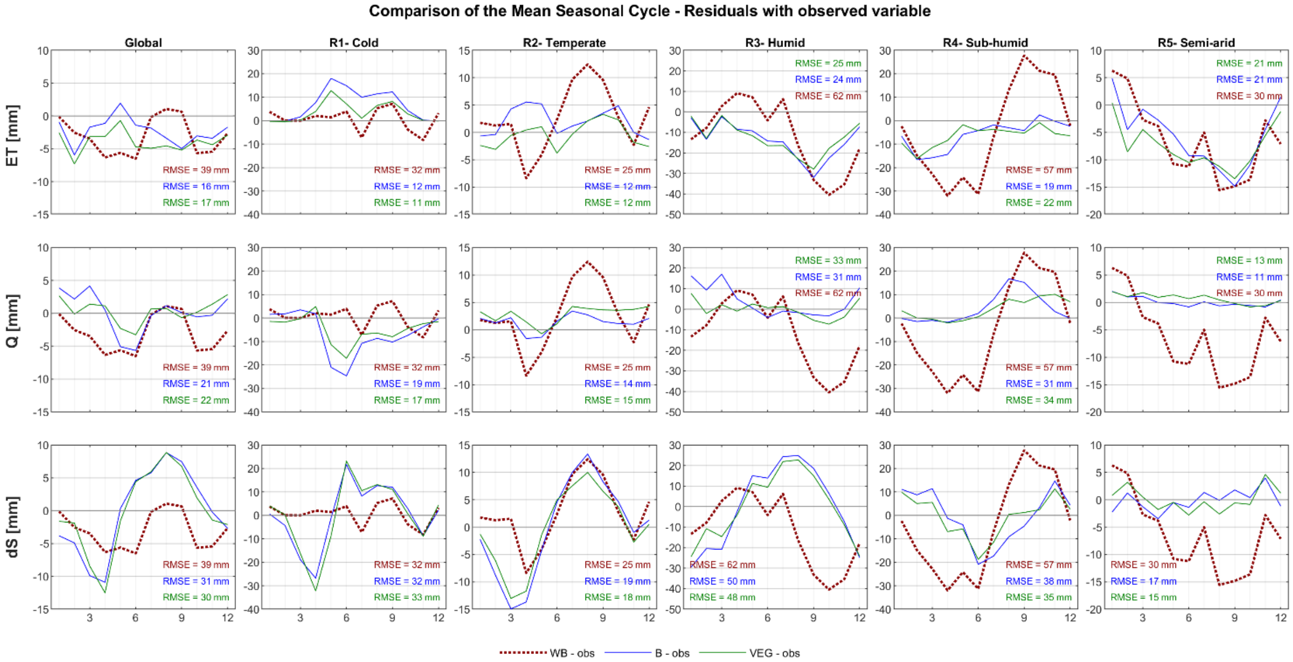

Figure S28 Global and regional mean seasonal cycles of the difference between the simulations of ET, Q and dS with B and VEG and the respective observations (B-obs, VEG-obs), as well as difference between observed variables and the same variables calculated via the water balance from the other observations (WB-obs).

S11 Analysis for Koeppen-Geiger Zones

Additionally to the hydroclimatic cluster analysis shown in the main text, we performed a regional analysis for Koeppen-Geiger climate zones. To do so, we aggregated Koeppen-Geiger subgroups considering the main climate group and distinguishing between humid and semi-arid conditions. The resulting zones are shown in Fig. S29. Fig. S30 evaluates model performance for the Koeppen-Geiger regions and Fig. S31 shows the composition of seasonal TWS variations therein. Note that most parts of the *Polar* and *Boreal* Koeppen-Geiger (KG) zone are included in the *Cold* region (R1) of the hydroclimatic cluster classification. We find that the regional averages are very similar for both classification schemes in terms of model performance and composition of seasonal TWS variations.

The Northern Hemisphere *Temp* and *Boreal-sa* KG zones are both included in the *Temperate* hydroclimatic region (R2). *Temp* KG and the *Temperate* region (R2) agree well regarding model performance and seasonal cycles, although we see a slightly better performance for the *Temp* KG regarding TWS and Q. In the *Boreal-sa* KG, **B** and **VEG** do not reproduce the spring peak of Q and precede the observed TWS significantly, decreasing model performance slightly when combining the *Temp* and *Boreal-sa* KG zones in one hydroclimatic region. Therefore, it would make sense to further split up the *Temperate* hydroclimatic cluster region. However, *Boreal-sa* KG spans Northern China, where poorer model performance is also evident from the performance maps in Fig. 4 of the main manuscript.

However, as mentioned in the main text, the advantage of the hydroclimatic cluster regionalization becomes obvious when interpreting results of the *Arid* and *Temp-sa* KG zones. This is because these climate zones are distributed across the Southern and Northern Hemisphere, causing 2 peaks in the regional seasonal cycles for TWS, ET and Q, due to opposing seasonal dynamics. The *Arid* KG zone includes the *Semi-arid* cluster regions (R5) in the Southern Hemisphere, as well as parts of the *Temperate* region (R2) (mainly in North America). The *Temp-sa* KG zone covers a rather small fraction of the study area, that is spread over the *Temperate* region (R2) in the Northern Hemisphere and the *Semi-arid* region (R5) of the Southern Hemisphere.

The effect of opposing seasonal cycles also exists in the *Tropic* KG zone, although less pronounced due to the proximity to the equator where the climate is more homogeneous and seasonality is low. The *Tropic* KG corresponds to the *Humid* cluster region (R3) on the Southern Hemisphere, and parts of the *Sub-humid* region (R4) on the Northern Hemisphere. Compared to the hydroclimatic cluster regions, the *Tropic* KG has less seasonal variation (a smaller amplitude) of TWS, ET and Q, due to its larger area North and South of the equator. Both, **B** and **VEG** underestimate the ongoing depletion of TWS from September to December in *Tropic* KG, which is likely related to the opposing seasonal cycles of TWS in the *Humid* (R3) and the *Sub-humid* (R4) cluster regions. In the *Tropic* KG, Q peaks in March (as in *Humid* (R3)) and has a second, smaller peak in September (when Q peaks in the *Sub-humid* region (R4)). However, model performance is very similar for *Tropic* KG and the *Humid* and *Sub-humid* cluster regions.

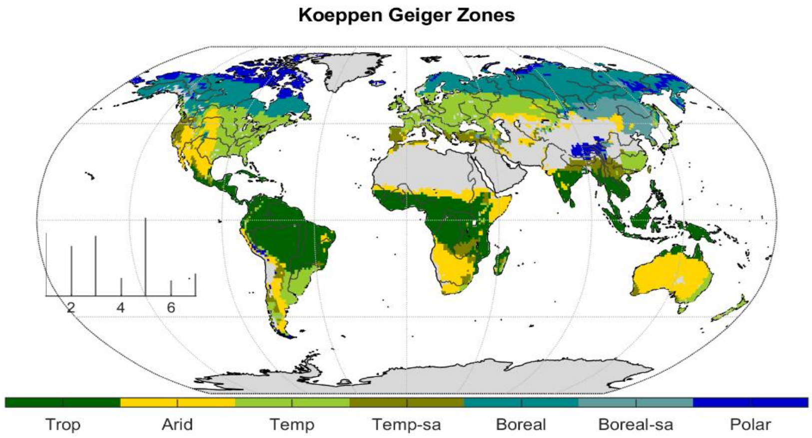

Figure S29 Regions based on Koeppen-Geiger climate zones (Trop = Af, Am, As, Aw; Arid = BSh, BSk, BWh, BWk; Temp = Cfa, Cfb, Cfc, Dfa, Dfb; Temp-sa = Csa, Csb, Csc, Cwa, Cwb, Cwc; Boreal = Dfc, Dfd; Boreal-sa = Dsa, Dsb, Dsc, Dwa, Dwb, Dwc, Dwd; Polar = EF, ET).

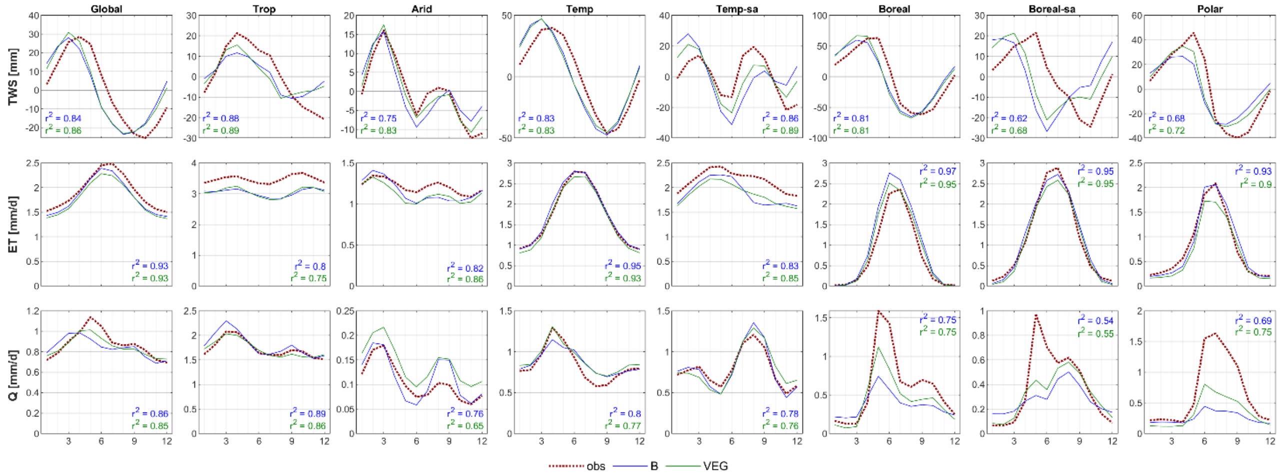

Figure S30 Global and regional mean seasonal cycles of total water storage (TWS), evapotranspiration (ET) and runoff (Q) for the B and VEG experiments compared to the observational constraints by GRACE (TWS), FLUXCOM (ET) and GRUN (Q).

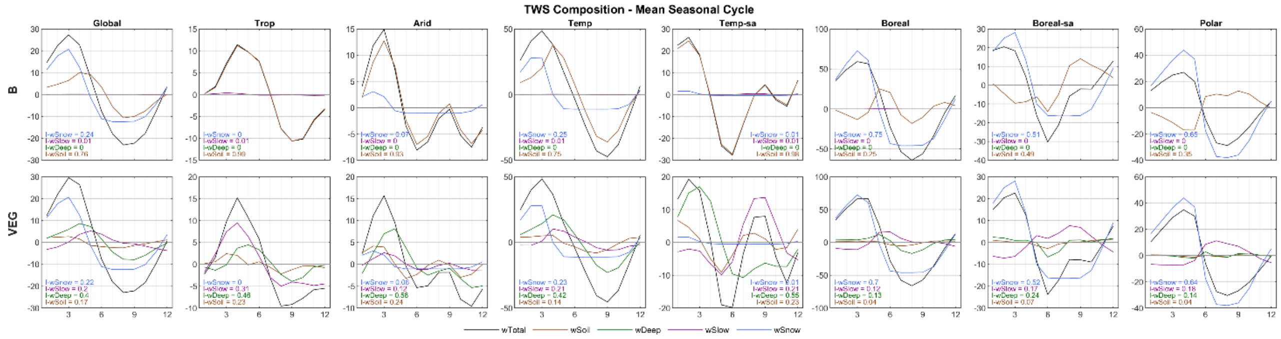

Figure S31 Global and regional mean seasonal cycles of simulated total water storage and its components for B and VEG, including the regional Impact Index *I* for each storage.