Global patterns in vegetation accessible subsurface water storage emerge from spatially varying importance of individual drivers

van Oorschot, Fransje; Hrachowitz, Markus; Viering, Tom; Alessandri, Andrea; van der Ent, Ruud J

LETTER • OPEN ACCESS

Global patterns in vegetation accessible subsurface water storage emerge from spatially varying importance of individual drivers

To cite this article: Fransje van Oorschot *et al* 2024 *Environ. Res. Lett.* **19** 124018

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Global patterns in vegetation accessible subsurface water storage emerge from spatially varying importance of individual drivers

Fransje van Oorschot[1,2,*] ⓘ, Markus Hrachowitz[1] ⓘ, Tom Viering[3], Andrea Alessandri[2] ⓘ and Ruud J van der Ent[1] ⓘ

1 Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
2 Institute of Atmospheric Sciences and Climate, National Research Council of Italy (CNR-ISAC), Bologna, Italy
3 Pattern Recognition Laboratory, Faculty of Electrical Engineering Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
* Author to whom any correspondence should be addressed.

E-mail: f.vanoorschot@tudelft.nl

Abstract

Vegetation roots play an essential role in regulating the hydrological cycle by removing water from the subsurface and releasing it to the atmosphere. However, the present understanding of the drivers of ecosystem-scale root development and their spatial variability globally is limited. This study investigates the varying roles of climate, landscape, and vegetation on the magnitude of root zone storage capacity (S_r) worldwide, which is defined as the maximum volume of subsurface moisture accessible to vegetation roots. To this aim, we quantified S_r and evaluated 21 possible climate, landscape, and vegetation controls for 3612 river catchments worldwide using a random forest machine learning model. Our findings reveal climate as primary, but spatially varying, driver of ecosystem scale S_r with landscape and vegetation characteristics playing a minor role. More specifically, we found the mean inter-storm duration as most dominant control of S_r globally, followed by mean temperature, mean precipitation, and mean topographic slope. While the inter-storm duration, temperature, and slope exhibit a consistent relation with S_r globally, the relation between precipitation and S_r varies spatially. Based on this spatial variability, we classified two different regimes: precipitation driven and energy limited. The precipitation-driven regime exhibits a positive relation between precipitation and S_r for precipitation of up to $3 \, \mathrm{mm \, d^{-1}}$, above which the relation flattens and eventually becomes negative. The energy-limited regime exhibits a strictly negative relation between precipitation and S_r. Using the random forest model based on these three dominant climate variables and the landscape variable slope, we generated a global gridded dataset of S_r, which closely resembles other global datasets of root characteristics. This suggests that our parsimonious approach based on four globally available variables to estimate S_r on a global scale has the potential to be readily and easily integrated into the parameterization of S_r in global hydrological and land surface models. This may enhance the accuracy of global predictions of land–atmosphere exchange fluxes and hydrological extremes by providing a robust representation of both spatial and temporal variability in vegetation root characteristics.

1. Introduction

Vegetation continuously adjusts to the prevailing climate and landscape characteristics ensuring optimal functionality (Gentine *et al* 2012, Fan *et al* 2017). One of the properties identified as adaptive in both space

and time are vegetation root systems, that are shaped in a way to provide both anchoring in the subsurface (Read and Stokes 2006), as well as access to sufficient nutrients and water (Zhang *et al* 2019, Oldroyd and Leyser 2020, Maan *et al* 2023). Water uptake by roots of vegetation regulates vegetation transpiration,

globally the largest water flux released from terrestrial systems (Schlesinger and Jasechko 2014), and the associated latent heat flux into the atmosphere. In spite of its importance for the global water and energy budgets, direct large scale (i.e. beyond lab scale or individual plants) observations of root systems and the related water uptake do not exist.

Therefore, several indirect methods have been developed to represent vegetation root characteristics on large (here: global) scales. Schenk *et al* (2009) provided global estimates of the soil depths that contain 95% of the roots (i.e. the 95% rooting depth), extrapolated from a sample of several hundred direct point-scale root observations of individual plants (Schenk and Jackson 2003). Other studies followed inverse methods based on optimality principles to infer root characteristics at a global scale. For example, Kleidon (2004) maximized net primary production, while Yang *et al* (2016) used balances of carbon cost and benefits, and Fan *et al* (2017) derived depths of root water uptake from balances of water supply and demand. Similar water supply and demand considerations were also used by Wang-Erlandsson *et al* (2016) and Stocker *et al* (2023) to estimate global distributions of root zone storage capacity S_r (mm). S_r is defined as the maximum volume of subsurface moisture accessible to vegetation roots, representing all sources of water within the reach of roots, including unsaturated soil, deep and shallow groundwater (Gao *et al* 2014). S_r is a fundamental characteristic of terrestrial hydrological systems as it regulates not only water budgets by partitioning precipitation into drainage and evaporation, but also energy budgets over the associated latent heat flux (Zhang *et al* 2001, Donohue *et al* 2012, Wang-Erlandsson *et al* 2016). Many studies have suggested that, on ecosystem scale, S_r is mainly shaped by climate and in particular by the interplay of the temporal dynamics of water and energy availability, as vegetation optimizes its root system to sustain water demand (Kleidon 2004, Laio *et al* 2006, Guswa 2008, Gentine *et al* 2012, Gao *et al* 2014, 2023, De Boer-Euser *et al* 2016). Consequently, ecosystem disturbances such as climate change and human land-use change also influence the evolution of S_r, as demonstrated by multiple studies (Nijzink *et al* 2016, Liu *et al* 2020, Hrachowitz *et al* 2021, Bouaziz *et al* 2022, Tempel *et al* 2024, Wang *et al* 2024). Thus, insight in the specific controls of S_r is essential for predicting how different ecosystems will respond to such disturbances.

A range of other recent studies has explored how climate variables influence the extent of S_r across different regions in varying climatic zones. Gao *et al* (2014) identified precipitation inter-storm duration and seasonality index as key controls of S_r in Thailand and the United States. Inter-storm duration

is an indicator for the length of dry periods during which vegetation relies on its subsurface water buffer for transpiration and was linked to the size of vegetation root systems in multiple previous studies (e.g. Gentine *et al* 2012, Sivandran and Bras 2013). Conversely, de Boer-Euser *et al* (2019) observed a strong positive relationship between mean temperature and S_r in Finland, along with a positive correlation between aridity index and S_r, as also found by Zhao *et al* (2016) in China. Also Gao *et al* (2014) noted that, on average, drier regions have larger S_r than wetter regions in Thailand and the United States, but eco-region classes Tropical Savanna and Semi-arid Prairies deviated from this trend. Contrasting signals were also reported by Singh *et al* (2020), who showed that increased aridity in tropical forests leads to decreased tree cover but increased S_r due to the remaining scarcer vegetation investing more in roots to create a water buffer for drier periods. However, in drier savanna-grasslands, S_r decreases with increased aridity (Singh *et al* 2020). Yet, other studies do highlight that vegetation (de Boer-Euser *et al* 2019) and landscape characteristics such as soil properties (Laio *et al* 2006, Collins and Bras 2007) and geology (Hahm *et al* 2019, 2024, McCormick *et al* 2021) can play a relevant role at regional scales. Furthermore, vegetation and landscape characteristics in practice effectively determine S_r in most land surface models (Liu *et al* 2020, Van Oorschot *et al* 2021, Wang *et al* 2021).

Overall, previous studies suggest that the influence of climate variables on S_r, considering both their magnitude and direction, is not consistent across different regions (Gao *et al* 2014, Zhao *et al* 2016, de Boer-Euser *et al* 2019, Singh *et al* 2020). Similarly, the spatially varying role of landscape and vegetation characteristics versus climate has not been systematically quantified and analyzed on a global scale. Thus, it remains unclear which aspects of climate, landscape and vegetation are the most important controls on S_r on global scale, and how these controls vary in space.

Here we bridge this knowledge gap by quantitatively characterizing, for the first time, how different climate, landscape and vegetation variables control the magnitude of S_r at the global scale, and how these controls vary spatially. Based on historical long-term water balance data, we estimate S_r in 3612 catchments worldwide using the memory method as in Van Oorschot *et al* (2024). We then test a wide range of climate, landscape, and vegetation variables to quantify their influence on S_r in different regions using a random forest model. Having identified a set of first order controls, we then use this model to extrapolate the catchment S_r estimates to a global gridded map of S_r ensuring coverage of regions where insufficient water balance data are available to directly estimate S_r. Finally, we evaluate how these estimates relate to

Table 1. Data sources used for the hydrological variables daily precipitation (P (mm d^{-1})), daily potential evaporation (E_p (mm d^{-1})), and annual mean discharge ($\overline{Q}$ (mm d^{-1})).

Variable	Data source
P (mm d^{-1})	Global Soil Wetness Project Phase 3 (GSWP-3) (Dirmeyer *et al* 2006, Lange and Büchner 2020)
E_p (mm d^{-1})	Global Land Evaporation Amsterdam Model version 3.5a (GLEAMv3.5a); based on the Priestley–Taylor approach (Miralles *et al* 2011, Martens *et al* 2017)
$\overline{Q}$ (mm d^{-1})	GSIM (Do *et al* 2018, Gudmundsson *et al* 2018), LamaH-CE (Klingler *et al* 2021), CAMELS Australia (Fowler *et al* 2021), CAMELS US (Addor *et al* 2017), and EStreams (do Nascimento *et al* 2024)

other global datasets of root characteristics (Schenk and Jackson 2003, Kleidon 2004, Schenk *et al* 2009, Wang-Erlandsson *et al* 2016, Yang *et al* 2016, Fan *et al* 2017, Stocker *et al* 2023).

2. Methods

2.1. Catchment data
Following the data and methods from Van Oorschot *et al* (2024), we estimated catchment-scale root zone storage capacity S_r in the 3612 study catchments using catchment-averaged daily precipitation data P (mm d^{-1}), daily potential evaporation E_p (mm d^{-1}), and annual mean discharge $\overline{Q}$ (mm d^{-1}) from sources documented in table 1. We selected 3612 catchments based on the following four criteria: (1) at least 10 years of overlap between Q, P and E_p data; (2) catchment not exceeding the water limit, i.e. $\overline{Q} < \overline{P}$; (3) catchment not exceeding the energy limit, i.e. annual mean actual evaporation ($\overline{E_a} = \overline{P} - \overline{Q}$) $< \overline{E_p}$; (4) catchment area $<10\,000$ km^2 to limit the heterogeneity within catchments.

To investigate the controls on S_r, we selected 21 catchment-averaged variables (table 2), subdivided into three categories: climate, landscape, and vegetation. All variables were obtained from global datasets to ensure data consistency across catchments and in view of extrapolation from catchments to global scale. These variables were selected based on three main criteria: (1) globally available data that is representative for catchment scales; (2) variables with limited mutual interactions and (3) variables with limited assumptions on vegetation root characteristics (e.g. root depth) that are based on scarce observations that are not representative for ecosystems at landscape scales (Van Oorschot *et al* 2021).

2.2. Root zone storage capacity estimation
Using the memory method, a term introduced by Van Oorschot *et al* (2021) and also known as water balance method and related to the mass curve technique,

root zone storage capacity S_r (mm) is derived from root zone storage deficits (S_d (mm)), e.g. Gao *et al* 2014, Wang-Erlandsson *et al* 2016, Dralle *et al* 2021, Van Oorschot *et al* 2021, 2024). Based on long-term precipitation, potential evaporation and river discharge data we here computed long-term time-series of catchment S_d following the methodology described in Van Oorschot *et al* (2024). These time-series reflect both water supply to and water uptake by the vegetation's roots, leaving the vegetation as transpiration. Despite the inherent limitations of this method (Van Oorschot *et al* 2021, 2024) the S_r estimates it produces have been shown to closely align with those derived from hydrological model calibration, providing independent confirmation of their accuracy (Gao *et al* 2014). After fitting the S_d time-series with the Gumbel distribution, S_r was estimated based on the extreme S_d values with a 20 year return period to represent the memory of vegetation to past water deficit conditions. The extreme value analysis was done to generalize the results as the time-series of the catchments have different lengths and do not necessarily overlap, and to represent the timescale of vegetation adaptation. Previous studies have shown that low vegetation adapts its S_r to droughts occurring with relatively low return periods (<10 years) and high vegetation to >40 years (Wang-Erlandsson *et al* 2016). For the aim of this study a-priori differentiation between land cover types is not desirable, and therefore a fixed 20 year return period was selected for all catchments following Singh *et al* (2020) and Bouaziz *et al* (2020). Full details of the memory method can be found in the supplementary material S1.

2.3. Random forest model
We used a random forest regression model to predict catchment S_r and identify the dominant controls on S_r using the variables presented in table 2. A random forest model was selected because it can represent the non-linear interactions between catchment variables and S_r which appeared during the studies of Gao *et al* (2014), Zhao *et al* (2016), de Boer-Euser *et al* (2019), and during exploratory analyses on our dataset. The model was trained by minimizing the mean absolute difference (MAD (mm)) between root zone storage capacity from the memory method ($S_{r,M}$) and from the random forest model predictions ($S_{r,P}$). We applied a five-fold cross validation in order to fairly estimate the generalization performance. This implies that we performed model training and testing five times on a different subset of the total dataset, with for each fold 80% of the catchments (2890) were used for model training and 20% of the catchments (722) for model testing. Model performance is quantified by the mean and the standard deviation of the MAD between the $S_{r,M}$ and $S_{r,P}$ over the five cross-validation folds. Supplementary section S2 describes the details of the random forest model.

Table 2. Catchment variables used for the random forest models (section 2.3). All variables represent a single value per catchment, obtained through averaging grid cells that lie within the catchment boundary. For each catchment time series matching the available discharge data were used.

	Name	Description	Units	Data source
Climate	$\overline{P}$	Mean precipitation	mm d^{-1}	GSWP-3 (Dirmeyer *et al* 2006, Lange and Büchner 2020)
	$\overline{E_{\mathrm{p}}}$	Mean potential evaporation	mm d^{-1}	GLEAMv3.5a (Miralles *et al* 2011, Martens *et al* 2017)
	$\overline{T}$	Mean temperature	°C	GSWP-3
	$\overline{t_{\mathrm{IS}}}$	Mean inter-storm duration	d	GSWP-3
	$I_{\mathrm{S,P}}$	Seasonality index of precipitation (Gao *et al* 2014)	—	GSWP-3
	$I_{\mathrm{S,E_p}}$	Seasonality index of potential evaporation (Gao *et al* 2014)	—	GLEAMv3.5a
	T_{d}	Temperature difference, defined as the difference between the monthly mean maximum and minimum temperature	°C	GSWP-3
	I_{AS}	Asynchronicity index between monthly mean precipitation and potential evaporation (Feng *et al* 2019)	—	GSWP-3 and GLEAMv3.5a
	$\bar{f}_{\mathrm{snow}}$	Mean snow cover fraction	—	MOD10A1 (Hall and Riggs, 2021)
	$f_{\mathrm{snow,v}}$	Variability of snow cover fraction defined as $f_{\mathrm{snow,s}}/\bar{f}_{\mathrm{snow}}$ with $f_{\mathrm{snow,s}}$ the standard deviation of monthly mean f_{snow}	—	MOD10A1
Landscape	e	Elevation	m	HydroSHEDS Hydrologically Conditioned DEM (Lehner *et al* 2008) and Multi-Error-Removed Improved-Terrain (MERIT) DEM for latitudes >60° (Yamazaki *et al* 2017)
	s	Slope	%	Same as for elevation.
	d_{b}	Depth to bedrock	m	SoilGrids250m (Hengl *et al* 2017)
	f_{clay}	Fraction of soil clay content for 0–200 cm depth	—	SoilGrids250m
	f_{sand}	Fraction of soil sand content for 0–200 cm depth	—	SoilGrids250m
Vegetation	f_{tree}	Tree cover fraction	—	MOD44B.006 (DiMiceli *et al* 2015)
	f_{nontree}	Non tree cover fraction	—	MOD44B.006
	f_{nonveg}	Non vegetation fraction defined as $1 - f_{\mathrm{tree}} - f_{\mathrm{nontree}}$	—	MOD44B.006
	$\overline{\mathrm{LAI}}$	Mean leaf area index	—	CGLS (Verger *et al* 2019)
	v_{LAI}	Variability of leaf area index defined as $\mathrm{LAI}_s/\overline{\mathrm{LAI}}$ with LAI_s the standard deviation of monthly mean LAI	—	CGLS
	I_{a}	Irrigated area fraction	—	(Siebert *et al* 2015)

With the above described procedure we built a model with 21 predictor variables (21-variable model) (table 2), which was used to investigate which catchment variables are a dominant control on S_{r}. The degree of control, i.e. the variable importance, is quantified by the permutation feature importance, which represents the decrease in model performance (ΔMAD) when the values of this single variable are randomly shuffled, while keeping the values of other variables (Breiman 2001). Based on the variable importance of the 21-variable model and cross-correlations between individual variables, we selected

a subset of four predictor variables used in a second 4-variable model. The reduced number of variables in the 4-variable model allows for in-depth model interpretation, and is simpler for future applications than the 21-variable model as it requires less data.

2.4. Model interpretation methods

For the 4-variable model, we used individual conditional expectation (ICE) curves to analyze how individual variables influence $S_{\mathrm{r,P}}$ in each catchment (Goldstein *et al* 2015). In the ICE curves, S_{r} is predicted by the 4-variable model for each catchment

Figure 1. (a) Memory method root zone storage capacity ($S_{r,M}$ (mm)). (b) Difference between the 4-variable model predicted and memory method: $\Delta S_r = S_{r,P} - S_{r,M}$ of the test data, represented by the combined test results of the five-fold cross validation models. Relative differences between $S_{r,M}$ and $S_{r,P}$ are shown in figure S3. Dots represent the catchment outlets.
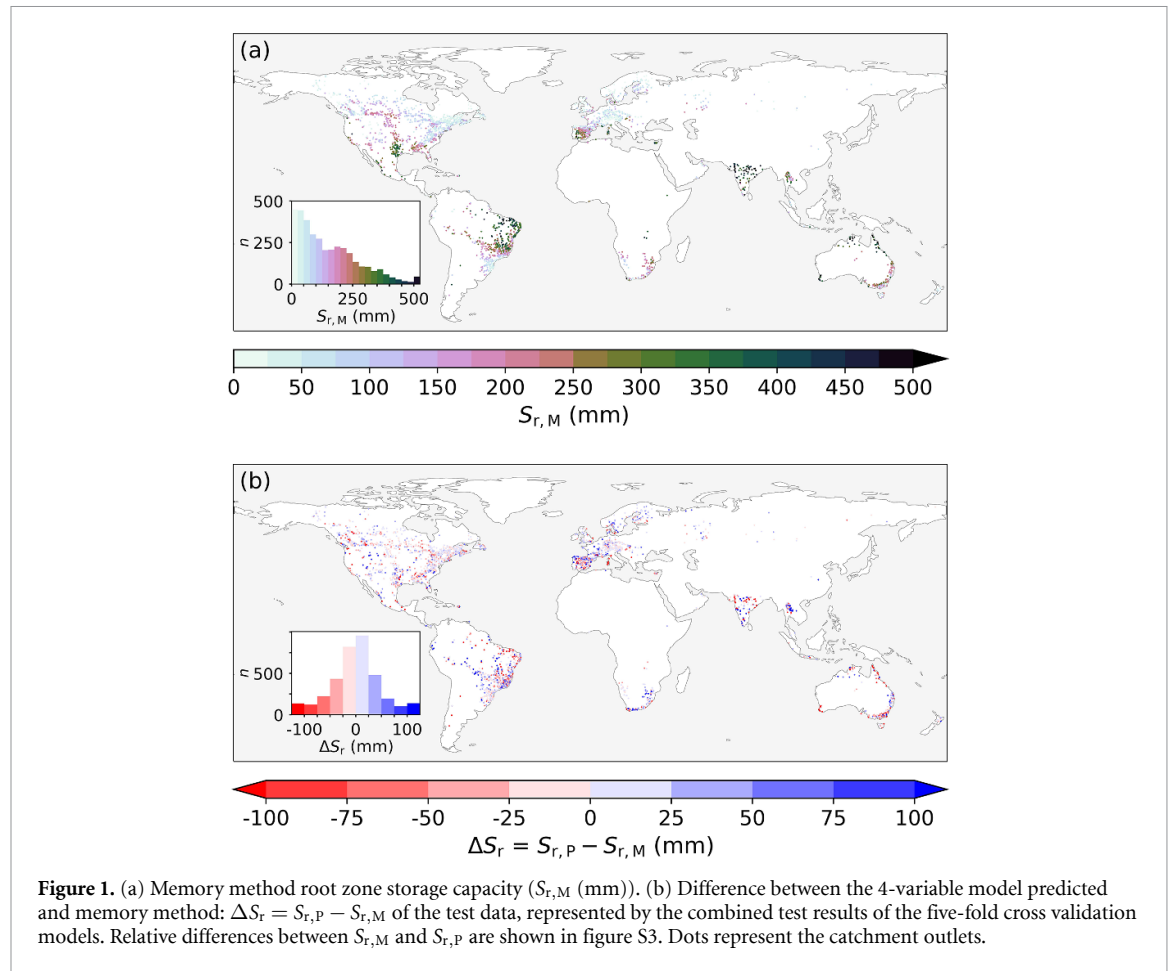
with a range of hypothetical values for one of the variables, while keeping the original values of the other variables. This way we obtain a curve for each variable, for each catchment, that represents the dependence between the model $S_{r,P}$ and that variable for a single catchment. The 4-variable model is then also used to globally estimate $S_{r,P}$ at a 0.5° latitude × 0.5° longitude spatial resolution. This global S_r map is compared to other global datasets of root characteristics in terms of spatial patterns and Spearman rank correlations (r). Details of these global datasets are provided in table S3.

3. Results and discussion

3.1. Root zone storage capacity prediction
The median $S_{r,M}$ estimated from the memory method in the 3612 catchments reached 120 mm (5–95th percentiles: 10–390 mm). The lower $S_{r,M}$ values are concentrated in cool-temperate humid regions while higher values are scattered around warmer, more arid regions (figure 1(a)). These magnitudes and pattern are broadly consistent with previous regional S_r estimates based on the memory method (Gao *et al* 2014, De Boer-Euser *et al* 2016, Zhao *et al* 2016, Singh *et al* 2020). Using these estimates to train the 21-variable model then resulted in a MAD

$= 13 \pm 0.2$ mm ($R^2 = 0.97$) (figure 2(a)). The test sets of the cross-validation analysis exhibited with 35 ± 2 mm ($R^2 = 0.81$) a larger scatter (figure 2(b)), but the limited performance fluctuations between the individual cross validation folds on the unseen test data (table S2) indicate that the model is robust. The sequence of variable importance in the 21-variable model suggests that climate variables play by far the largest role, while landscape and vegetation variables are less important for explaining $S_{r,P}$ globally (figure 2(c)). We found that the inter-storm duration ($\overline{t_{IS}}$) is the most important variable in the 21-variable model. This can be seen in figure 2(c) by the increase of MAD by ≈ 25 mm when the catchment values of the mean inter-storm duration ($\overline{t_{IS}}$) are randomly shuffled between catchments, while keeping the other variables unchanged. The variable importance of mean inter-storm duration $\overline{t_{IS}}$ is followed by the climate variables mean snow cover $\overline{f}_{snow}$ (ΔMAD ≈ 13 mm), mean temperature $\overline{T}$ (ΔMAD ≈ 13 mm), and mean precipitation $\overline{P}$ (ΔMAD ≈ 7 mm) (figure 2(c), table 2). While the most important landscape variables slope (s) and elevation (e) play a moderate role (ΔMAD ≈ 5 mm), all other landscape variables and all vegetation variables are characterized by much lower ΔMAD < 2 mm. This indicates that they only have minor explanatory power for S_r.
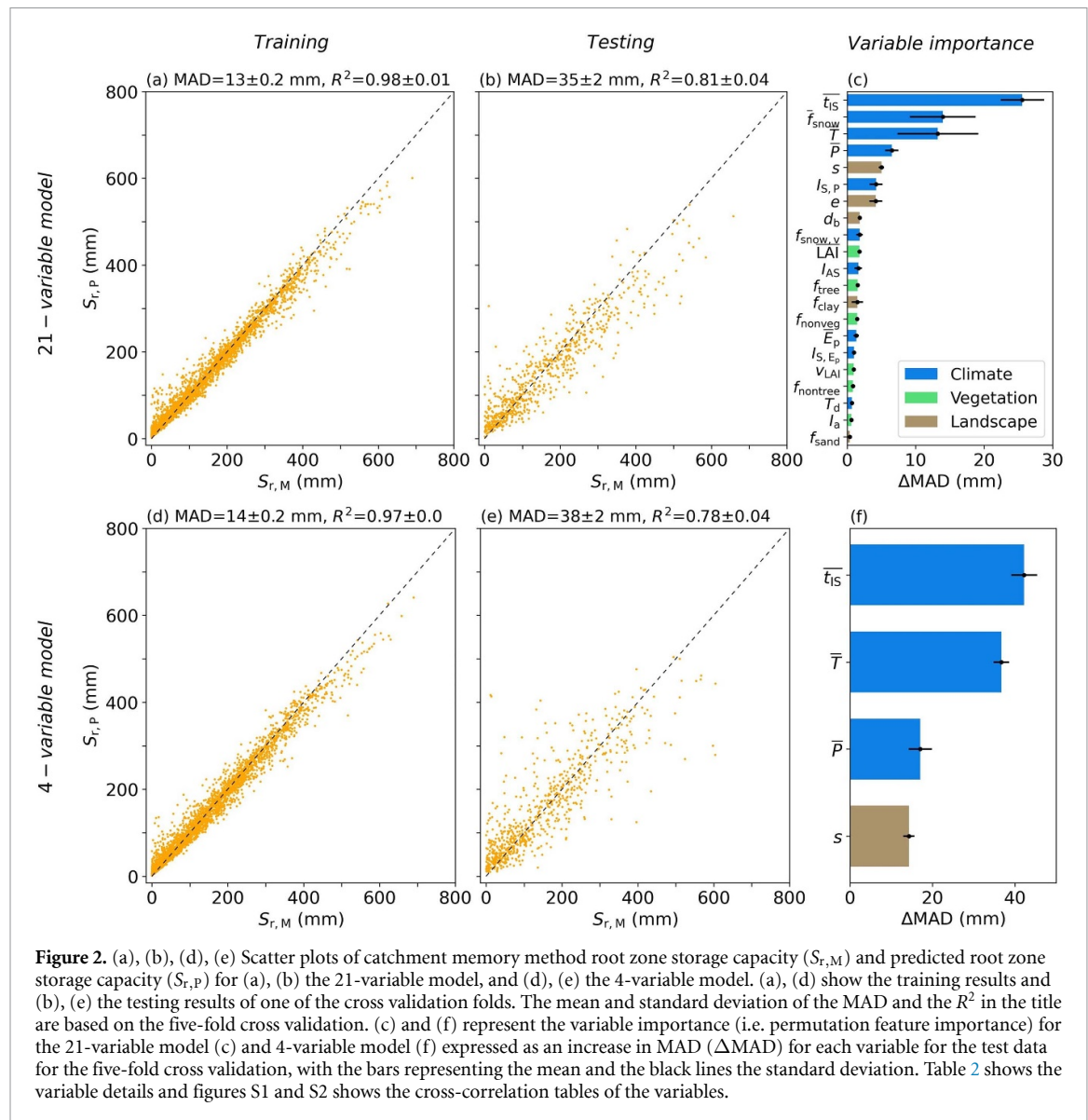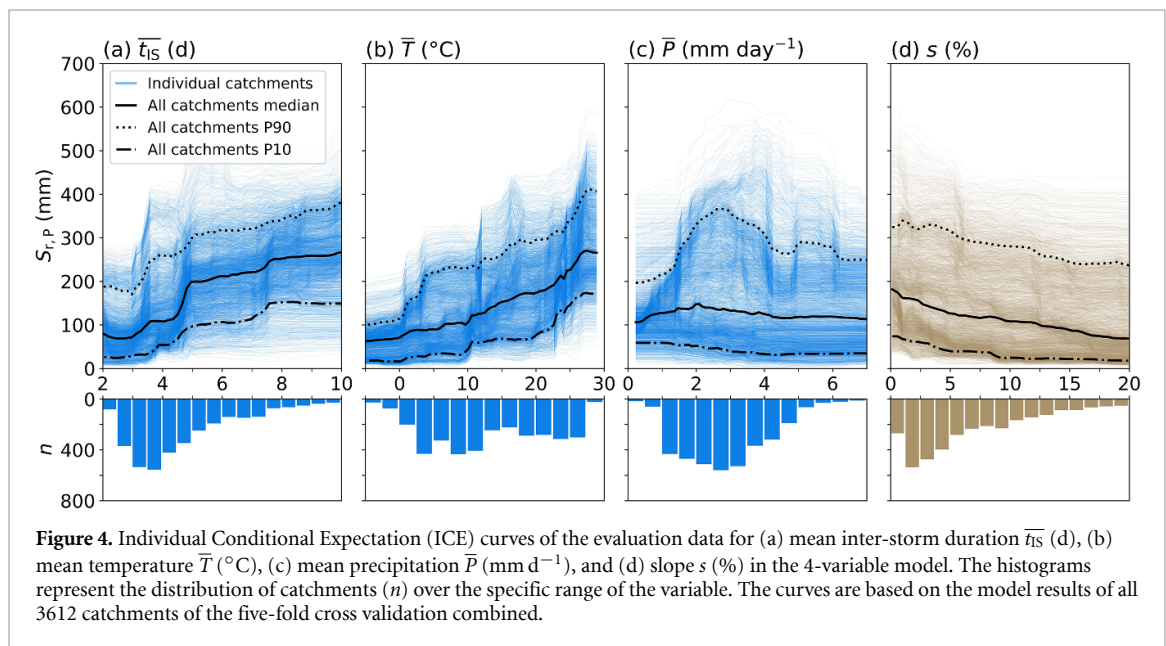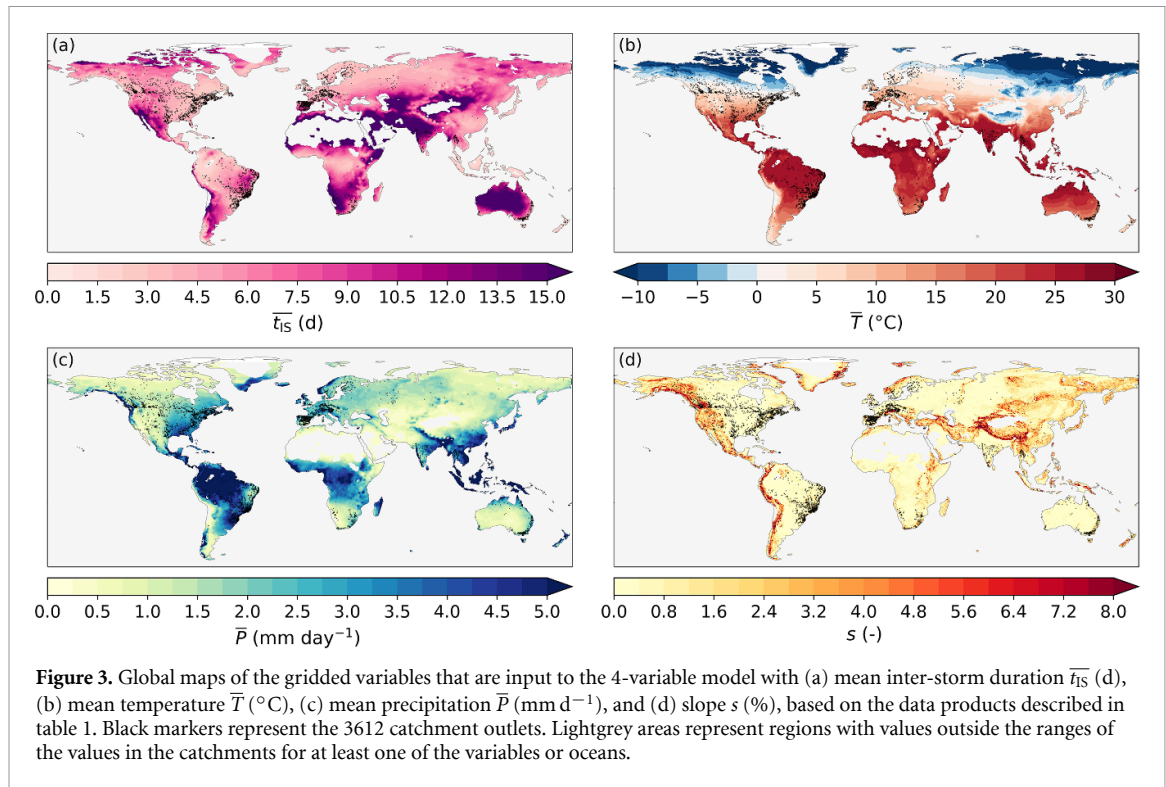
Figure 2. (a), (b), (d), (e) Scatter plots of catchment memory method root zone storage capacity ($S_{r,M}$) and predicted root zone storage capacity ($S_{r,P}$) for (a), (b) the 21-variable model, and (d), (e) the 4-variable model. (a), (d) show the training results and (b), (e) the testing results of one of the cross validation folds. The mean and standard deviation of the MAD and the R^2 in the title are based on the five-fold cross validation. (c) and (f) represent the variable importance (i.e. permutation feature importance) for the 21-variable model (c) and 4-variable model (f) expressed as an increase in MAD (ΔMAD) for each variable for the test data for the five-fold cross validation, with the bars representing the mean and the black lines the standard deviation. Table 2 shows the variable details and figures S1 and S2 shows the cross-correlation tables of the variables.

Based on these results we have then removed variables with low variable importance as well as correlated variables (see figures S1 and S2) to obtain an interpretable model, while maintaining high model performance. The resulting parsimonious 4-variable model uses $\overline{t_{IS}}$, $\overline{T}$ (excluding $\overline{f_{snow}}$ due to its strong correlation with $\overline{T}$), $\overline{P}$, and s (figure 3). With a training MAD $= 14 \pm 0.2$ mm ($R^2 = 0.97$) (figure 2(d)), this reduced model predicts $S_{r,P}$ for the unseen test data with a MAD $= 38 \pm 2$ mm ($R^2 = 0.78$) (figure 2(e)), which is very close to the test performance of the full 21-variable model (figure 2(b)). Here $\overline{t_{IS}}$ also emerges as the most important variable, followed by $\overline{T}$, $\overline{P}$, and s in the same hierarchy as in the full 21-variable model (figure 2(f)). Figure 1(b) shows that $S_{r,P}$ as predicted by the 4-variable model captures well the general pattern of $S_{r,M}$ with only rather limited deviations across all regions. Overall, ΔS_r remains within ± 50 mm for 3173 (88%) catchments. Larger ΔS_r,

both positive and negative, are found in regions such as India, Spain, and Northern Australia (figure 1(b)). However, in these regions $S_{r,M}$ is relatively large, thus resulting in relatively minor relative differences (figure S3).

The dominance of climate variables as primary controls on S_r corresponds well with previous studies (e.g. Gao *et al* 2014, De Boer-Euser *et al* 2016, and Yang *et al* 2016). Our findings also show the relevance of topography, albeit to a lesser degree compared to climate, in shaping S_r, which was also highlighted by Fan *et al* (2017). However, other studies also emphasized the importance of other landscape characteristics such as geology and soil textures for S_r (Laio *et al* 2006, Hahm *et al* 2019, 2024, McCormick *et al* 2021). In particular, Hahm *et al* (2024) demonstrated that in regions with limited variability in climate characteristics, geologic factors can become a stronger control on S_r. This also holds for the relevance of irrigation,

Figure 3. Global maps of the gridded variables that are input to the 4-variable model with (a) mean inter-storm duration $\overline{t_{IS}}$ (d), (b) mean temperature $\overline{T}$ (°C), (c) mean precipitation $\overline{P}$ (mm d^{-1}), and (d) slope s (%), based on the data products described in table 1. Black markers represent the 3612 catchment outlets. Lightgrey areas represent regions with values outside the ranges of the values in the catchments for at least one of the variables or oceans.

Figure 4. Individual Conditional Expectation (ICE) curves of the evaluation data for (a) mean inter-storm duration $\overline{t_{IS}}$ (d), (b) mean temperature $\overline{T}$ (°C), (c) mean precipitation $\overline{P}$ (mm d^{-1}), and (d) slope s (%) in the 4-variable model. The histograms represent the distribution of catchments (n) over the specific range of the variable. The curves are based on the model results of all 3612 catchments of the five-fold cross validation combined.

that was previously found to be regionally influencing S_r (Van Oorschot *et al* 2024). Besides that, the limited impact of the degree of irrigation and other vegetation variables in shaping S_r directly results from their inherent dependence on climate conditions. The results of our analysis above are largely consistent with previous findings as they suggest that landscape characteristics are important as a secondary control at regional scales where differences of climate factors are more limited. For ecosystem scales in a global context, and thus a wide range in climates, climate is the clear first order control on S_r.

3.2. Relation between predictor variables and root zone storage capacity

The individual influences of the four variables on $S_{r,P}$ in the 4-variable model are represented by the ICE curves in figure 4. Generally, as mean inter-storm duration $\overline{t_{IS}}$ and mean temperature $\overline{T}$ increase, $S_{r,P}$ also increases (figures 4(a) and (b)). Thus, overall, $S_{r,P}$ is higher in warmer regions with longer dry periods. However, it can also be observed that $\overline{t_{IS}}$ has more effect on $S_{r,P}$, i.e. a steeper slope in the ICE curve, for values between 3–5 days than for higher $\overline{t_{IS}}$, i.e. longer periods without rainfall (figure 4(a)). As

Table 3. Characteristics of the two groups with distinct relations between mean precipitation $\overline{P}$ (mm d^{-1}) and root zone storage capacity S_r (mm) as identified in figure 5, with values for mean inter-storm duration $\overline{t_{IS}}$ (d), mean temperature $\overline{T}$ (°C), mean precipitation $\overline{P}$ (mm d^{-1}), slope s (%), and tree dover fraction f_{tree} (−) representing the median and inter quartile range (25th percentile–75th percentile) of the catchments in the specific group (table 2).

Group name	$\overline{t_{IS}}$ (d)	$\overline{T}$ (°C)	$\overline{P}$ (mm d^{-1})	s (%)	f_{tree} (−)	S_r (mm) increasing with $\overline{P}$ (mm d^{-1})	S_r (mm) decreasing with $\overline{P}$ (mm d^{-1})
Precipitation driven	7 (6–8)	20 (14–24)	2.3 (1.6–3.4)	5 (3–9)	0.13 (0.08–0.20)	$\overline{P} \approx 0$–3	$\overline{P} > 3$
Energy limited	4 (3–4)	9 (4–14)	3.0 (2.3–3.8)	6 (3–12)	0.39 (0.27–0.49)	—	$\overline{P} > 0$

$\overline{t_{IS}}$ increases, vegetation needs to invest more in roots in order to maintain transpiration during dry periods up to that value before it levels off and $S_{r,P}$ does not significantly increase anymore at higher $\overline{t_{IS}}$, which is related to lower vegetation density in drier regions with longer dry periods. The modelled $S_{r,P}$ has a stronger response to changes in $\overline{T}$ for $\overline{T} > 10\,°C$ than in colder regions with $\overline{T} < 10\,°C$. With increasing $\overline{T}$, there is more energy available for transpiration, and, as long as water is available, vegetation will transpire more under higher temperatures, leading to larger S_r (figure 4(b)). On the other hand, $S_{r,P}$ consistently decreases with increasing slope s (figure 4(d)). Steeper slopes are typically found in regions with higher elevations, which are associated with lower temperatures and reduced vegetation density, leading to lower S_r. In contrast with the almost monotonically increasing behaviour of $\overline{t_{IS}}$ and $\overline{T}$, and monotonically decreasing behaviour of s in relation to $S_{r,P}$ (figures 4(a), (b) and (d)), the mean precipitation $\overline{P}$ exhibits a more complex relationship with $S_{r,P}$. For $\overline{P} < 2\,\text{mm d}^{-1}$, the median relationship shows an increase in $S_{r,P}$, while for $\overline{P} > 2\,\text{mm d}^{-1}$ this relationship is inverted and $S_{r,P}$ decreases with increasing $\overline{P}$. In addition, while for $\overline{t_{IS}}$, $\overline{T}$, and s, the general shapes of the ICE curves of the individual catchments largely resemble each other, considerable differences from the average pattern are observed for $\overline{P}$, as indicated by the curves of the 10th and 90th percentiles of the curves in figure 4(c).

To further investigate these diverging patterns in the relationship between $\overline{P}$ and $S_{r,P}$, we disentangled the individual curves of figure 4(c) and grouped them based on their overall trajectories (considering slopes and peaks of the curves) in figure 5. This resulted in two distinct groups with each group showing similar influence of $\overline{P}$ on $S_{r,P}$, characterizing different regions globally (figure 5, table 3). Overall, the precipitation-driven group shows the largest influence of changes in $\overline{P}$ on S_r (figure 5(a)). In these regions, $S_{r,P}$ strongly increases with increasing $\overline{P}$ up to ∼3 mm d^{-1}. This is related to the lower vegetation density and thus less actively transpiring plants in these dry regions (figure 6(a), table 3). Here, increases in $\overline{P}$ directly lead to more vegetation activity and/or cover, and thus

higher S_r. Hence, the development of S_r is mostly precipitation driven. Above $\overline{P} > 3\,\text{mm d}^{-1}$, the systems experience transitions into systems that are not water limited anymore and where additional water input does not result in more root development and transpiration. This pattern dominates in regions that are characterized by relatively high temperatures and high rainfall seasonality (indicated by a large $\overline{t_{IS}}$), typical of tropical monsoon and (semi)-arid climates (figures 3 and 5(c), table 3), for example North-Eastern Brazil, India, and Northern Australia. For the energy-limited group, an opposite signal compared to the precipitation-driven group is found, with $\overline{P}$ negatively influencing $S_{r,P}$ (figure 5(b), table 3). This pattern is found in energy-limited regions with year-round rainfall and relatively high tree cover, such as Europe and Canada, as well as tropical rainforests in the Amazon and Indonesia (figures 5(c) and 6(b), table 3). Here, sufficient water is available throughout much of the year to satisfy vegetation water demand. Therefore, an increase in water availability here does not lead to denser vegetation, and existing vegetation can reallocate resource investment into above-surface growth instead of extending its root system. As a consequence, additional water input, i.e. increase in $\overline{P}$, and thus frequent water re-supply to the root zone results in a reduced ecosystem-scale S_r (figure 5(b)): vegetation can access sufficient water with the need for deeper roots.

The spatial variability of the relation between $\overline{P}$ and S_r as shown in figure 5 is in line with previous regional studies by Gao *et al* (2014), de Boer-Euser *et al* (2019), and Singh *et al* (2020). Specifically, the contrasting $\overline{P} - S_r$-relation between tropical rainforests (energy limited) and the savanna or prairie grassland regions (precipitation driven), correspond to the water stress and vegetation regimes in rainforest-savanna transitional regions found by Singh *et al* (2020). Also Guswa (2008) found similar contrasting relations between wetness and rooting depth for energy-limited vs. water-limited regions. The relations between our selected variables and S_r presented in figures 4 and 5 can represent how ecosystems can be plausibly expected to respond in
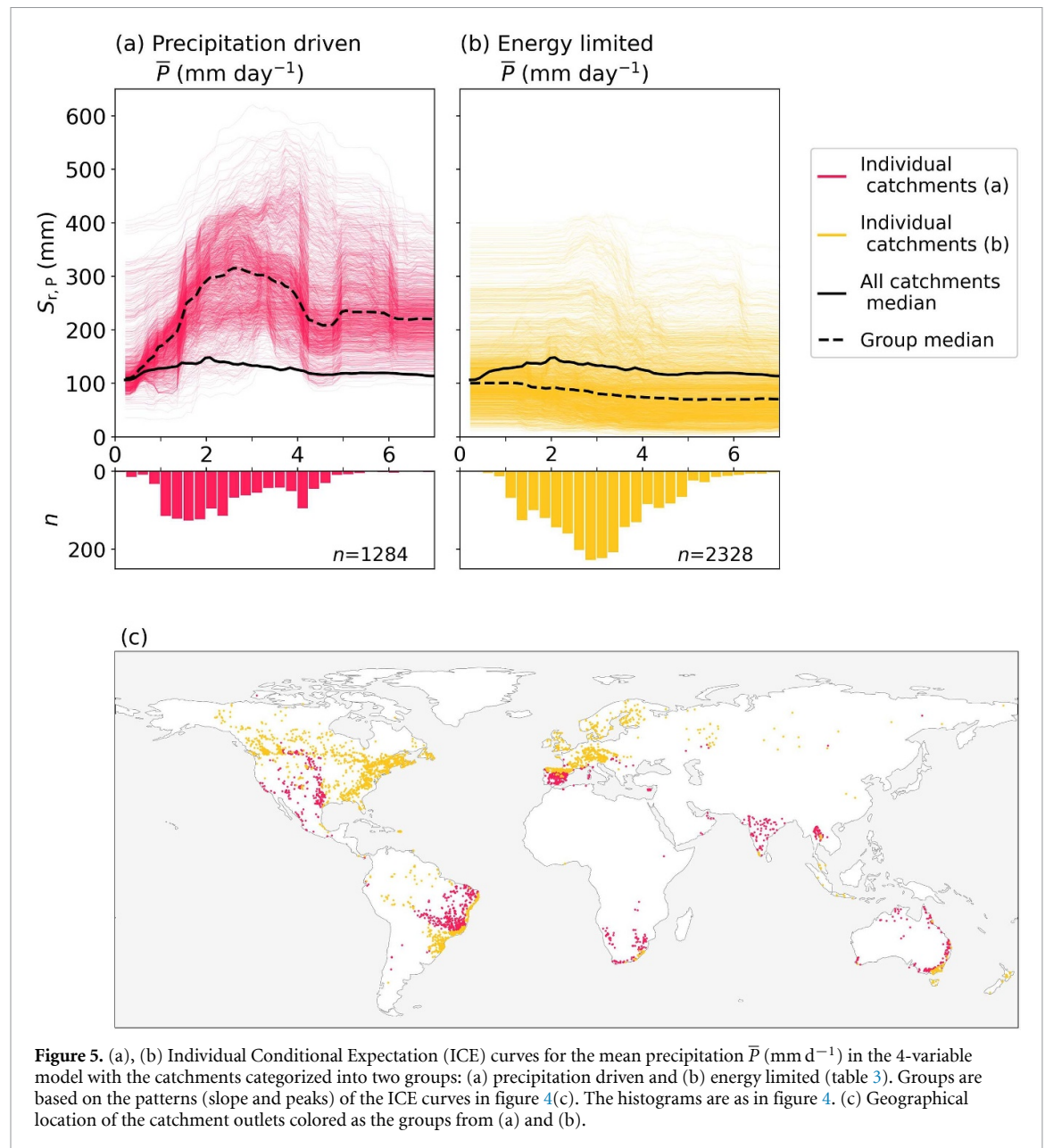
Figure 5. (a), (b) Individual Conditional Expectation (ICE) curves for the mean precipitation $\overline{P}$ (mm d^{-1}) in the 4-variable model with the catchments categorized into two groups: (a) precipitation driven and (b) energy limited (table 3). Groups are based on the patterns (slope and peaks) of the ICE curves in figure 4(c). The histograms are as in figure 4. (c) Geographical location of the catchment outlets colored as the groups from (a) and (b).

terms of S_r to changes in the three climatic variables over time, resulting from internal variability and climate change. However, it should be noted that this only holds for relatively small changes in the variables, while larger changes may be accompanied by a transition of the entire ecosystem into a different state and the associated changes in vegetation composition. Furthermore, we assumed that the patterns are largely climate driven, but human influences can have major effects as well (e.g. Grill *et al* 2019, Hrachowitz *et al* 2021).

3.3. Global comparison of root characteristics
Here we used the 4-variable model and the data from figure 3 to create a global gridded map of $S_{r,P}$ (figure 7(a)), extrapolating beyond regions with available discharge observations. To place our results into a wider context, we compared them to seven

other global estimates of root characteristics obtained with a wide range of different approaches (table S4) (Schenk and Jackson 2003, Kleidon 2004, Schenk *et al* 2009, Wang-Erlandsson *et al* 2016, Yang *et al* 2016, Fan *et al* 2017, Stocker *et al* 2023). Note that this comparison can only consider the general spatial pattern, while the absolute magnitudes are not necessarily comparable between these studies due to differences in methods, assumptions and underlying data.

Overall, the global distribution of $S_{r,P}$ broadly corresponds with the spatial patterns of other studies reflecting the hydrologically active rootzone in terms of root zone storage capacities reported by Stocker *et al* (2023) ($r = 0.72$) and Wang-Erlandsson *et al* (2016) ($r = 0.48$), and in terms of the optimized hydrologically active rooting depth by Kleidon (2004) ($r = 0.61$) (figures 7(a)–(d) and 8(a)–(d)). Major differences are observed in the La Plata basin, where
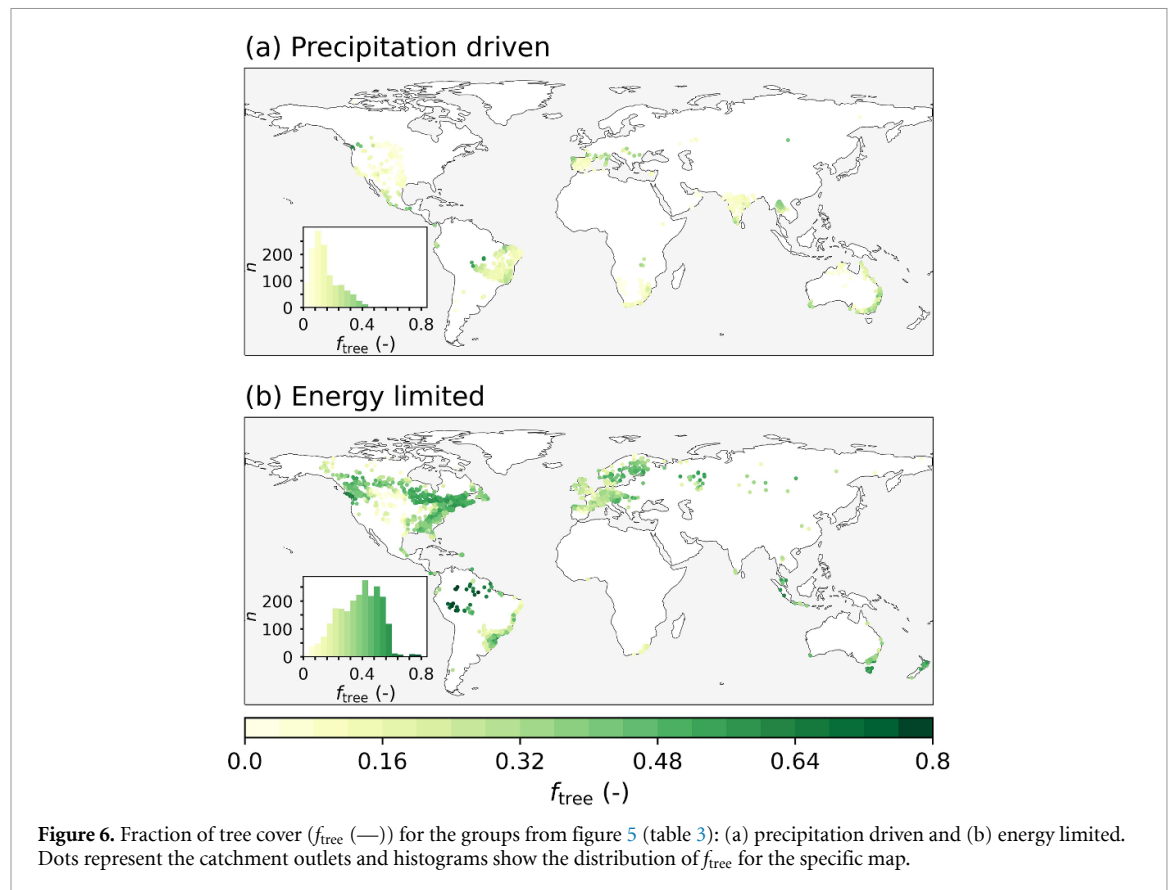
Figure 6. Fraction of tree cover (f_{tree} (—)) for the groups from figure 5 (table 3): (a) precipitation driven and (b) energy limited. Dots represent the catchment outlets and histograms show the distribution of f_{tree} for the specific map.

our $S_{r,P}$ estimates and Kleidon (2004) show relatively large values, while the S_r values from Stocker *et al* (2023) and Wang-Erlandsson *et al* (2016) are relatively small (figures 7(a)–(d)). Other differences are found in relatively arid regions of the Central US, and Central Asia (figures 7(a)–(d)), in which our 4-variable model $S_{r,P}$ estimates are higher compared to Wang-Erlandsson *et al* (2016), Stocker *et al* (2023), and Kleidon (2004) (figure S4).
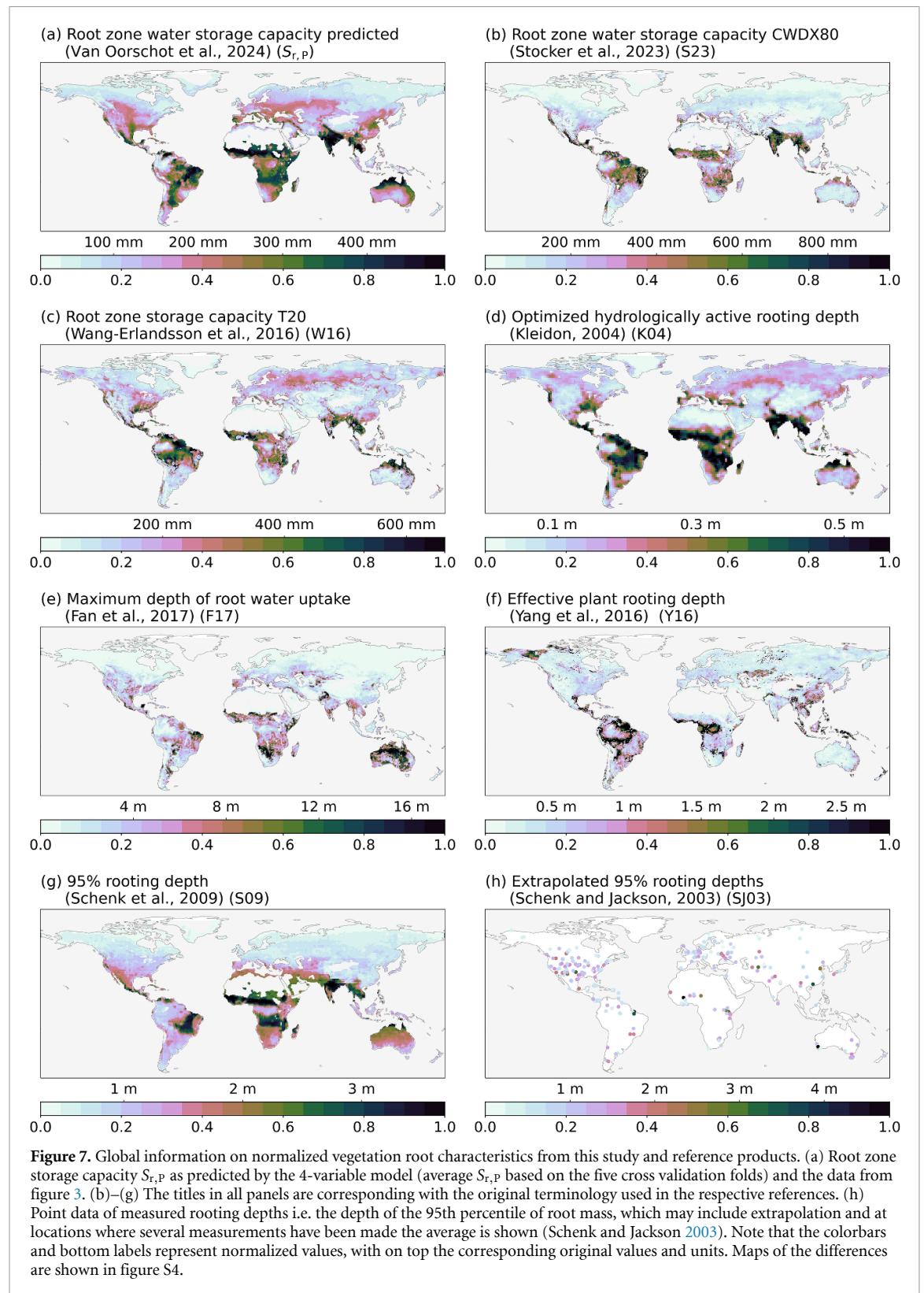
It should be noted that the rooting depth characteristics presented in figures 7(e)–(h) are not necessarily proportional to the root zone storage capacity S_r due to the fact that rooting depth is a single plant property compared to S_r being an ecosystem property. Moreover, while rooting depth only represents the vertical extension of the roots, S_r accounts for the entire root profile, including lateral root extent and root density. Nevertheless, the rooting depth related products of Fan *et al* (2017) and Schenk *et al* (2009) (figures 7(e) and (g)) broadly resemble the spatial patterns of $S_{r,P}$, with $r = 0.70$ and 0.73, respectively (figures 8(a), (e) and (g)). However, considerable differences are observed between the maximum depth of root water uptake in figure 7(e) and our $S_{r,P}$ (figure 7(a)) in Australia and Southern Africa, where the maximum depth of root water uptake (Fan *et al* 2017) is considerably larger than $S_{r,P}$ (figure S4). In these arid regions, the maximum depth of root water uptake is high because individual trees have

deep roots to access sufficient water. However, the ecosystem S_r is low because of low vegetation density.

The spatial pattern of our $S_{r,P}$ in North-America with relatively low values in the Midwest US and relatively high values in the Great Plains corresponds well with the results of Fan *et al* (2017) (figure 7(e)), Schenk *et al* (2009) (figure 7(g)) and Gao *et al* (2014), but less with the other datasets (figures 7(b)–(d) and S4). These differences arise likely from dry season/summer dormancy that is common for this C3-grass dominated part in the prairies of the Great Plains (e.g. Ke *et al* 2013). In the memory method used in this study (Van Oorschot *et al* 2024) as well as by Gao *et al* (2014), transpiration is defined as a fraction of potential evaporation based on the long-term mean actual evaporation from the water balance. During the dry season when potential evaporation is typically high and grasses go dormant, this methodological assumption likely leads to an overestimation of transpiration, and thus S_r. The root characteristics from Kleidon (2004), Wang-Erlandsson *et al* (2016), and (Stocker *et al* 2023) are derived from datasets that directly represent vegetation activity, and are, therefore, less exposed to this limitation. This possibly also explains the differences in arid regions discussed before.

Our $S_{r,P}$ estimates contrast with the effective plant rooting depth values from Yang *et al* (2016) in many regions, but it should be noted that Yang *et al* (2016)

Figure 7. Global information on normalized vegetation root characteristics from this study and reference products. (a) Root zone storage capacity $S_{r,P}$ as predicted by the 4-variable model (average $S_{r,P}$ based on the five cross validation folds) and the data from figure 3. (b)–(g) The titles in all panels are corresponding with the original terminology used in the respective references. (h) Point data of measured rooting depths i.e. the depth of the 95th percentile of root mass, which may include extrapolation and at locations where several measurements have been made the average is shown (Schenk and Jackson 2003). Note that the colorbars and bottom labels represent normalized values, with on top the corresponding original values and units. Maps of the differences are shown in figure S4.

similarly and markedly contrasts with all other datasets, with $r < 0.08$ (figures 8(a) and (f)). Also the comparison of observed rooting depths from Schenk and Jackson (2003) (figure 7(h)) with the other datasets reveals limited similarity. However, amongst all these global datasets our modeled $S_{r,P}$ exhibits the strongest correlation with these rooting depths, with $r = 0.23$ (figures 8(a) and (h)).

Overall, the major features in the spatial pattern of $S_{r,P}$ in our study mirror those in the other compared seven datasets. In addition, $S_{r,P}$ correlates better with most of the other datasets than
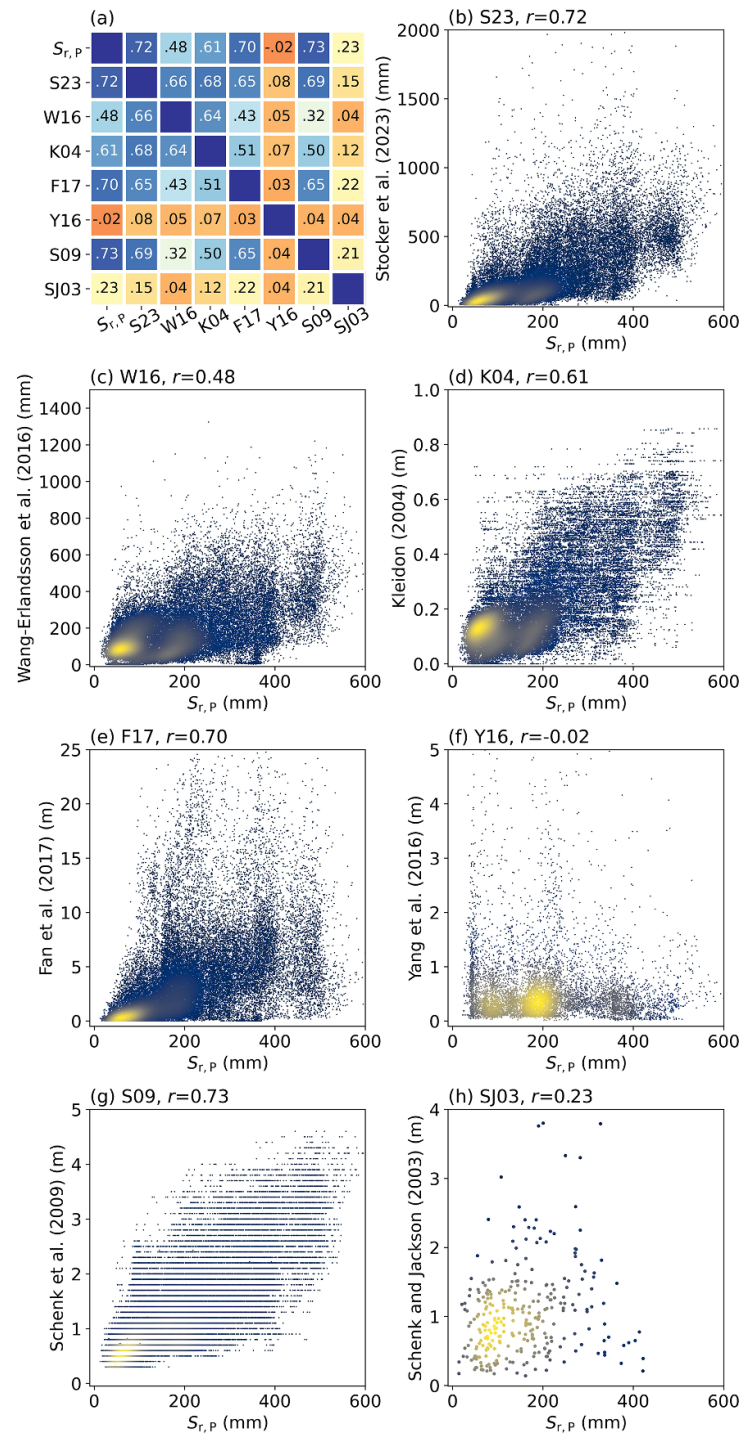
Figure 8. (a) Spearman rank correlation coefficients between reference root characteristics from figures 7(b)–(h) and the root zone storage capacity $S_{r,P}$ as predicted by the 4-variable model and the data from figure 2 (figure 7(a)). (b)–(h) Scatterplots of $S_{r,P}$ and the reference root characteristics. Note that the variables and the units are not the same for each product. The numbers in the title indicate the Spearman rank correlation coefficients, and the marker colors represent the point density. See figure 7 for the abbreviations used.

these datasets among themselves (figure 8(a)). This is a strong indication that using a parsimonious 4-variable model based on globally available hydro-climatic indices and the topography represented by slope, produces global pattern of $S_{r,P}$ that are at least as good as those of alternative approaches.

These alternative approaches rely on higher dimensional models to estimate evaporation as compared to our 4-variable model, which is grounded in discharge observations. Furthermore, the four variables are readily represented and simulated by Earth system models (ESMs), therefore

enabling the possibility of an interactive representation of S_r as a dynamically evolving variable in ESMs.

4. Conclusions

In our analysis we quantified for the first time the controls of ecosystem-scale root zone storage capacity S_r at the global scale using a random forest model. From this analysis, hydro-climatic variables emerged as the most dominant controls. Topographic slope also influences S_r, though to a lesser extent than climate. Other landscape and vegetation characteristics were found to play a minor role. More specifically, we found that inter-storm duration, temperature, precipitation, and topographic slope are the most dominant controls of S_r globally. We further found that inter-storm duration and temperature exhibit a near-monotonic positive relations with S_r, and that the slope is consistently negatively related to S_r. In contrast, the relation between precipitation and S_r varies in space. The emerging pattern suggests that while precipitation is strongly positively correlated with S_r in relatively dry regions with low vegetation cover, energy-limited regions are rather characterized by a negative relation. This highlights the distinct roles of precipitation for vegetation water-use in different climatic settings.

Our global S_r predictions, based on random forest models driven with the above variables, correspond closely with other global datasets of rootzone characteristics, which typically rely on more complex data sources and computations. In contrast, our model predicts S_r from the long-term means of three hydro-climatic variables ($\overline{P}, \overline{T}, \overline{t_{IS}}$), for which both historical data as well as future projections are readily available, and from topographic slope (s) that can be assumed as temporally invariant. The strength of our S_r prediction is further highlighted by its stronger correlation with most other datasets compared to the correlations observed between those datasets themselves.

For future studies, our approach opens the possibility to formulate and implement S_r as dynamically-evolving prognostic variable in large-scale land surface and hydrological models, mimicking the continuous evolution of S_r over time (Wang *et al* 2024) and replacing the current static representation of this parameter in most of these models. As such, we emphasize that the methodology to estimate S_r provided here can readily be applied in land surface and hydrological models to improve their global representation of the coupling of root zone storage capacity with hydro-climatic variability (Abramowitz *et al* 2024). This does not only have the potential to improve predictions of extreme river flow and seasonal water supply but, perhaps more importantly, also transpiration and thus latent heat fluxes representation (Van Oorschot *et al* 2021, Giardina *et al* 2024), which as a knock-on effect is expected to

have major implications for climate predictions and projections.

ORCID iDs

Fransje van Oorschot ● https://orcid.org/0000-0002-8811-0620
Markus Hrachowitz ● https://orcid.org/0000-0003-0508-1017
Andrea Alessandri ● https://orcid.org/0000-0002-2153-7961
Ruud J van der Ent ● https://orcid.org/0000-0001-5450-4333

References

Abramowitz G *et al* 2024 On the predictability of turbulent fluxes from land: PLUMBER2 MIP experimental description and preliminary results *EGUsphere* **2024** 1–47

Addor N, Newman A J, Mizukami N and Clark M P 2017 The CAMELS data set: catchment attributes and meteorology for large-sample studies *Hydrol. Earth Syst. Sci.* **21** 5293–313

Bouaziz L J E, Aalbers E E, Weerts A H, Hegnauer M, Buiteveld H, Lammersen R, Stam J, Sprokkereef E, Savenije H H G and Hrachowitz M 2022 Ecosystem adaptation to climate change: the sensitivity of hydrological predictions to time-dynamic model parameters *Hydrol. Earth Syst. Sci.* **26** 1295–318

Bouaziz L J, Steele-Dunne S C, Schellekens J, Weerts A H, Stam J, Sprokkereef E, Winsemius H H, Savenije H H and Hrachowitz M 2020 Improved understanding of the link between catchment-scale vegetation accessible storage and satellite-derived soil water index *Water Resour. Res.* **56** 1–22

Breiman L 2001 *Mach. Learn.* **45** 5–32

Collins D B and Bras R L 2007 Plant rooting strategies in water-limited ecosystems *Water Resour. Res.* **43** 1–10

De Boer-Euser T, McMillan H K, Hrachowitz M, Winsemius H C and Savenije H H 2016 Influence of soil and climate on root zone storage capacity *Water Resour. Res.* **52** 2009–24

de Boer-Euser T, Meriö L-J and Marttila H 2019 Understanding variability in root zone storage capacity in boreal regions *Hydrol. Earth Syst. Sci.* **23** 125–38

Delft High Performance Computing Centre (DHPC) 2022 *DelftBlue Supercomputer (Phase 1)* (available at: www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase1)

DiMiceli C, Carroll M, Sohlberg R, Kim D, Kelly M and Townshend J 2015 *MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006* (https://doi.org/10.5067/MODIS/MOD44B.006)

Dirmeyer P, Gao X, Zhao M, Guo Z, Oki T and Hanasaki N 2006 GSWP-2: multimodel analysis and implications for our perception of the land surface *Bull. Am. Meteorol. Soc.* **87** 1381–98

Do H X, Gudmundsson L, Leonard M and Westra S 2018 The Global Streamflow Indices and Metadata Archive (GSIM)—part 1: the production of a daily streamflow archive and metadata *Earth Syst. Sci. Data* **10** 765–85

do Nascimento T V, Rudlang J, Höge M, van der Ent R, Chappon M, Seibert J, Hrachowitz M and Fenicia F 2024 EStreams: an integrated dataset and catalogue of streamflow, hydro-climatic variables and landscape descriptors for Europe (https://doi.org/10.31223/X5M39F)

Donohue R J, Roderick M L and McVicar T R 2012 Roots, storms and soil pores: incorporating key ecohydrological processes into Budyko's hydrological model *J. Hydrol.* **436–437** 35–50

Dralle D N, Hahm W J, Chadwick K D, McCormick E and Rempe D M 2021 Technical note: accounting for snow in the estimation of root zone water storage capacity from precipitation and evapotranspiration fluxes *Hydrol. Earth Syst. Sci.* **25** 2861–7

Fan Y, Miguez-Macho G, Jobbágy E G, Jackson R B and Otero-Casal C 2017 Hydrologic regulation of plant rooting depth *Proc. Natl Acad. Sci.* **114** 10572–7

Feng X, Thompson S E, Woods R and Porporato A 2019 Quantifying asynchronicity of precipitation and potential evapotranspiration in mediterranean climates *Geophys. Res. Lett.* **46** 14692–701

Fowler K J, Acharya S C, Addor N, Chou C and Peel M C 2021 CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia *Earth Syst. Sci. Data* **13** 3847–67

Gao H, Fenicia F and Savenije H H G 2023 HESS opinions: are soils overrated in hydrology? *Hydrol. Earth Syst. Sci.* **27** 2607–20

Gao H, Hrachowitz M, Schymanski S J, Fenicia F, Sriwongsitanon N and Savenije H 2014 Climate controls how ecosystems size the root zone storage capacity at catchment scale *Geophys. Res. Lett.* **41** 7916–23

Gentine P, D'Odorico P, Lintner B R, Sivandran G and Salvucci G 2012 Interdependence of climate, soil and vegetation as constrained by the Budyko curve *Geophys. Res. Lett.* **39** 2–7

Giardina F, Padrón R, Stocker B D, Schumacher D and Seneviratne S 2024 Dry biases in land water storage and excessive soil moisture limitation in CMIP6 models (https://doi.org/10.21203/rs.3.rs-4086740/v1)

Goldstein A, Kapelner A, Bleich J and Pitkin E 2015 Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation *J. Comput. Graph. Stat.* **24** 44–65

Grill G *et al* 2019 Mapping the world's free-flowing rivers *Nature* **569** 215–21

Gudmundsson L, Do H X, Leonard M and Westra S 2018 The Global Streamflow Indices and Metadata Archive (GSIM)—part 2: quality control, time-series indices and homogeneity assessment *Earth Syst. Sci. Data* **10** 787–804

Guswa A J 2008 The influence of climate on root depth: a carbon cost-benefit analysis *Water Resour. Res.* **44** 1–11

Hahm W J, Dralle D N, Lapides D A, Ehlert R S and Rempe D M 2024 Geologic controls on apparent root-zone storage capacity *Water Resour. Res.* **60** e2023WR035362

Hahm W J, Rempe D M, Dralle D N, Dawson T E, Lovill S M, Bryk A B, Bish D L, Schieber J and Dietrich W E 2019 Lithologically controlled subsurface critical zone thickness and water storage capacity determine regional plant community composition *Water Resour. Res.* **55** 3028–55

Hengl T *et al* 2017 SoilGrids250m: global gridded soil information based on machine learning *PLoS One* **12** 1–40

Hrachowitz M, Stockinger M, Coenders-Gerrits M, Van Der Ent R, Bogena H, Lücke A and Stumpp C 2021 Reduction of vegetation-accessible water storage capacity after deforestation affects catchment travel time distributions and increases young water fractions in a headwater catchment *Hydrol. Earth Syst. Sci.* **25** 4887–915

Ke Y, Leung L R, Huang M and Li H 2013 Enhancing the representation of subgrid land surface characteristics in land surface models *Geosci. Model Dev.* **6** 1609–22

Kleidon A 2004 Global datasets of rooting zone depth inferred from inverse methods *J. Clim.* **17** 2714–22

Klingler C, Schulz K and Herrnegger M 2021 LamaH-CE: LArge-SaMple DAta for hydrology and environmental sciences for Central Europe *Earth Syst. Sci. Data* **13** 4529–65

Laio F, D'Odorico P and Ridolfi L 2006 An analytical model to relate the vertical root distribution to climate and soil properties *Geophys. Res. Lett.* **33** L18401

Lange S and Büchner M 2020 ISIMIP2a atmospheric climate input data (https://doi.org/10.48364/ISIMIP.886955)

Lehner B, Verdin K and Jarvis A 2008 New global hydrography derived from spaceborne elevation data *EOS Trans. Am. Geophys. Union* **89** 93–94

Liu X, Chen F, Barlage M and Niyogi D 2020 Implementing dynamic rooting depth for improved simulation of soil moisture and land surface feedbacks in Noah-MP-Crop *J. Adv. Model. Earth Syst.* **12** e2019MS001786

Maan C, ten Veldhuis M-C and van de Wiel B J H 2023 Dynamic root growth in response to depth-varying soil moisture availability: a rhizobox study *Hydrol. Earth Syst. Sci.* **27** 2341–55

Martens B, Miralles D G, Lievens H, Van Der Schalie R, De Jeu R A, Fernández-Prieto D, Beck H E, Dorigo W A and Verhoest N E 2017 GLEAM v3: satellite-based land evaporation and root-zone soil moisture *Geosci. Model Dev.* **10** 1903–25

McCormick E L, Dralle D N, Hahm W J, Tune A K, Schmidt L M, Chadwick K D and Rempe D M 2021 Widespread woody plant use of water stored in bedrock *Nature* **597** 225–9

Miralles D G, De Jeu R A, Gash J H, Holmes T R and Dolman A J 2011 Magnitude and variability of land evaporation and its components at the global scale *Hydrol. Earth Syst. Sci.* **15** 967–81

Nijzink R *et al* 2016 The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change? *Hydrol. Earth Syst. Sci.* **20** 4775–99

Oldroyd G E D and Leyser O 2020 A plant's diet, surviving in a variable nutrient environment *Science* **368** eaba0196

Read J and Stokes A 2006 Plant biomechanics in an ecological context *Am. J. Bot.* **93** 1546–65

Schenk H and Jackson R 2003 *Global Distribution of Root Profiles in Terrestrial Ecosystems* (http://doi.org/10.3334/ORNLDAAC/660)

Schenk H, Jackson R, Hall F, Collatz G, Meeson B, Los S, Colstoun E D and Landis D 2009 *ISLSCP II Ecosystem Rooting Depths* (http://doi.org/10.3334/ORNLDAAC/929)

Schlesinger W H and Jasechko S 2014 Transpiration in the global water cycle *Agric. For. Meteorol.* **189-190** 115–7

Siebert S, Kummu M, Porkka M, Döll P, Ramankutty N and Scanlon B R 2015 A global data set of the extent of irrigated land from 1900 to 2005 *Hydrol. Earth Syst. Sci.* **19** 1521–45

Singh C, Wang-Erlandsson L, Fetzer I, Rockström J and Van Der Ent R 2020 Rootzone storage capacity reveals drought coping strategies along rainforest-savanna transitions *Environ. Res. Lett.* **15** 124021

Sivandran G and Bras R L 2013 Dynamic root distributions in ecohydrological modeling: a case study at Walnut Gulch Experimental Watershed *Water Resour. Res.* **49** 3292–305

Stocker B D, Tumber-Dávila S J, Konings A G, Anderson M C, Hain C and Jackson R B 2023 Global patterns of water

storage in the rooting zones of vegetation *Nat. Geosci.* **16** 250–6

Tempel N T, Bouaziz L, Taormina R, van Noppen E, Stam J, Sprokkereef E and Hrachowitz M 2024 Vegetation response to climatic variability: implications for root zone storage and streamflow predictions *EGUsphere* **2024** 1–42

Van Oorschot F, Van Der Ent R J, Alessandri A and Hrachowitz M 2024 Influence of irrigation on root zone storage capacity estimation *Hydrol. Earth Syst. Sci.* **28** 2313–28

Van Oorschot F, Van Der Ent R J, Hrachowitz M and Alessandri A 2021 Climate-controlled root zone parameters show potential to improve water flux simulations by land surface models *Earth Syst. Dyn.* **12** 725–43

Verger A, Baret F and Weiss M 2019 Vegetation and energy algorithm theoretical basis document *Copernicus Global Land Operations* pp 1–93 (available at: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_LAI1km-V2_I1.41.pdf)

Wang S, Hrachowitz M and Schoups G 2024 Multi-decadal fluctuations in root zone storage capacity through vegetation adaptation to hydro-climatic variability have minor effects on the hydrological response in the Neckar River basin, Germany *Hydrol. Earth Syst. Sci.* **28** 4011–33

Wang Y, Zeng Y, Yu L, Yang P, Van der Tol C, Yu Q, Lü X, Cai H and Su Z 2021 Integrated modeling of canopy photosynthesis, fluorescence and the transfer of energy, mass and momentum in the soil–plant–atmosphere continuum (STEMMUS–SCOPE v1.0.0) *Geosci. Model Dev.* **14** 1379–407

Wang-Erlandsson L, Bastiaanssen W G, Gao H, Jägermeyr J, Senay G B, Van Dijk A I, Guerschman J P, Keys P W, Gordon L J and Savenije H H 2016 Global root zone storage capacity from satellite-based evaporation *Hydrol. Earth Syst. Sci.* **20** 1459–81

Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O'Loughlin F, Neal J C, Sampson C C, Kanae S and Bates P D 2017 A high-accuracy map of global terrain elevations *Geophys. Res. Lett.* **44** 5844–53

Yang Y, Donohue R J and McVicar T R 2016 Global estimation of effective plant rooting depth: implications for hydrological modeling *Water Resour. Res.* **52** 8260–76

Zhang J, Wang J, Chen J, Song H, Li S, Zhao Y, Tao J and Liu J 2019 Soil moisture determines horizontal and vertical root extension in the perennial grass *Lolium perenne* L. growing in Karst soil *Front. Plant Sci.* **10** 629

Zhang L, Dawes W R and Walker G R 2001 Response of mean annual evapotranspiration to vegetation changes at catchment scale *Water Resour. Res.* **37** 701–8

Zhao J, Xu Z and Singh V P 2016 Estimation of root zone storage capacity at the catchment scale using improved Mass Curve Technique *J. Hydrol.* **540** 959–72