Supplement

Tables

Table S1. Overview of CMIP6 climate models utilised in this study, including their corresponding atmosphere models, land surface models and associated Member IDs.

No.	Institution; Country	Model	Atmosphere Model	Land Surface Model	Member ID	Ref.
1	BCC; China	BCC-CSM2-MR	BCC_AGCM3_MR	BCC_AVIM2.0	r1i1p1f1	Wu et al. (2019)
2	CCCMa; Canada	CanESM5	CanAM5	CLASS3.6, CTEM1.2	r1i1p1f1	Swart et al. (2019)
3	NCAR; USA	CESM2	CAM6	CLM5	r11i1p1f1	Danabasoglu et al. (2020)
4	CMCC; Italy	CMCC-CM2-SR5	CAM5	CLM4.5	r1i1p1f1	Cherchi et al. (2019)
5	CNRM/CERFACS; France	CNRM-CM6-1	ARPEGE-Climate v6.3 + SURFEX v8.0	ISBA + CTRIP	r1i1p1f2	Voldoire et al. (2019)
6	CNRM/CERFACS; France	CNRM-ESM2-1	ARPEGE-Climate v6.3 + SURFEX v8.0	ISBA + CTRIP	r1i1p1f2	Séférian et al. (2019)
7	IPSL; France	IPSL-CM6A-LR	LMDZ6A	ORCHIDEEv2	r1i1p1f1	Boucher et al. (2020)
8	MIROC; Japan	MIROC-ES2L	MIROC-AGCM + SPRINTARS	VISIT-e and MATSIRO6	r1i1p1f2	Hajima et al. (2020)
9	MPI-M; Germany	MPI-ESM1-2-LR	ECHAM6.3	JSBACH3.2	r11i1p1f1	Mauritsen et al. (2019)
10	NCC; Norway	NorESM2-MM	Modified CAM6	CLM5	r1i1p1f1	Seland et al. (2020)
11	MOHC; UK	UKESM1-0-LL	Unified Model + UKCA	JULES-ES-1.0	r1i1p1f2	Sellar et al. (2019)

Table S2. Global mean BGWS [%], precipitation [mm day^{-1}], runoff [mm day^{-1}], and transpiration [mm day^{-1}] for the historical period, the SSP3-7.0 scenario, and the change (SSP3-7.0 – Historical) based on 11 CMIP6 ESMs.

	BGWS [%]			**Precipitation [mm day^{-1}]**			**Runoff [mm day^{-1}]**			**Transpiration [mm day^{-1}]**		
Model	**Historical**	**SSP3-7.0**	**Change**	**Historical**	**SSP3-7.0**	**Change**	**Historical**	**SSP3-7.0**	**Change**	**Historical**	**SSP3-7.0**	**Change**
BCC-CSM2-MR	13.61	13.01	-1.11	2.48	2.57	0.09	1.06	1.13	0.07	0.49	0.48	-0.01
CESM2	4.44	7.05	2.48	2.48	2.53	0.06	1.02	1.06	0.03	0.68	0.58	-0.10
CMCC-CM2-SR5	7.75	6.97	-0.10	2.60	2.77	0.17	1.52	1.65	0.13	0.78	0.84	0.07
CNRM-CM6-1	-5.80	-3.33	2.48	2.35	2.44	0.10	0.93	1.02	0.08	0.81	0.78	-0.02
CNRM-ESM2-1	-5.79	-2.55	3.34	2.32	2.42	0.09	0.96	1.05	0.09	0.79	0.72	-0.07
CanESM5	10.26	8.01	-2.08	2.50	2.74	0.24	0.90	1.04	0.14	0.41	0.45	0.04
IPSL-CM6A-LR	6.05	7.83	1.84	2.80	2.99	0.20	1.16	1.36	0.20	0.70	0.71	0.01
MIROC-ES2L	2.25	1.24	-1.00	2.79	2.92	0.13	0.68	0.75	0.06	0.66	0.69	0.03
MPI-ESM1-2-LR	-5.94	-6.87	-0.93	2.23	2.35	0.11	0.51	0.57	0.06	0.62	0.64	0.02
NorESM2-MM	3.22	5.73	2.09	2.34	2.36	0.02	0.95	0.96	0.01	0.67	0.57	-0.11
UKESM1-0-LL	1.71	6.66	5.02	2.40	2.57	0.17	0.88	1.02	0.14	0.79	0.71	-0.08
Ensemble mean	3.37	4.54	1.15	2.48	2.61	0.13	0.98	1.08	0.10	0.68	0.66	-0.02

Table S3. Global fraction of land area (%) based on their historical values of the BGWS metric (positive or negative historical state) and their projected trends under the SSP3-7.0 scenario (positive or negative change) using on 11 CMIP6 ESMs.

Model	**Historical State Positive**		**Historical State Negative**	
	Change Positive	**Change Negative**	**Change Positive**	**Change Negative**
BCC-CSM2-MR	32.98%	35.53%	12.34%	18.35%
CESM2	25.26%	27.91%	32.63%	13.89%
CMCC-CM2-SR5	25.39%	26.67%	19.33%	27.45%
CNRM-CM6-1	27.26%	16.80%	38.38%	17.53%
CNRM-ESM2-1	20.78%	21.03%	36.17%	21.96%
CanESM5	16.10%	37.70%	17.88%	28.09%
IPSL-CM6A-LR	34.82%	22.06%	23.05%	19.98%
MIROC-ES2L	16.53%	35.00%	26.89%	21.59%
MPI-ESM1-2-LR	10.76%	21.12%	27.16%	40.97%
NorESM2-MM	25.20%	27.47%	31.89%	15.04%
UKESM1-0-LL	30.45%	18.86%	42.84%	7.62%
Ensemble mean	24.21%	23.48%	30.91%	21.40%

Table S4. Fraction of land area (%) that experiences a BGWS shift from a blue water to a green water regime or vice versa based on 11 CMIP6 ESMs.

Model	**Positive to Negative (%)**	**Negative to Positive (%)**
BCC-CSM2-MR	4.58%	3.21%
CESM2	4.97%	8.93%
CMCC-CM2-SR5	7.59%	2.80%
CNRM-CM6-1	1.85%	5.12%
CNRM-ESM2-1	3.95%	6.00%
CanESM5	9.06%	5.04%
IPSL-CM6A-LR	3.76%	5.90%
MIROC-ES2L	6.30%	5.34%
MPI-ESM1-2-LR	3.53%	3.06%
NorESM2-MM	4.57%	7.54%
UKESM1-0-LL	2.38%	10.31%
Ensemble mean	4.09%	6.31%

Figures

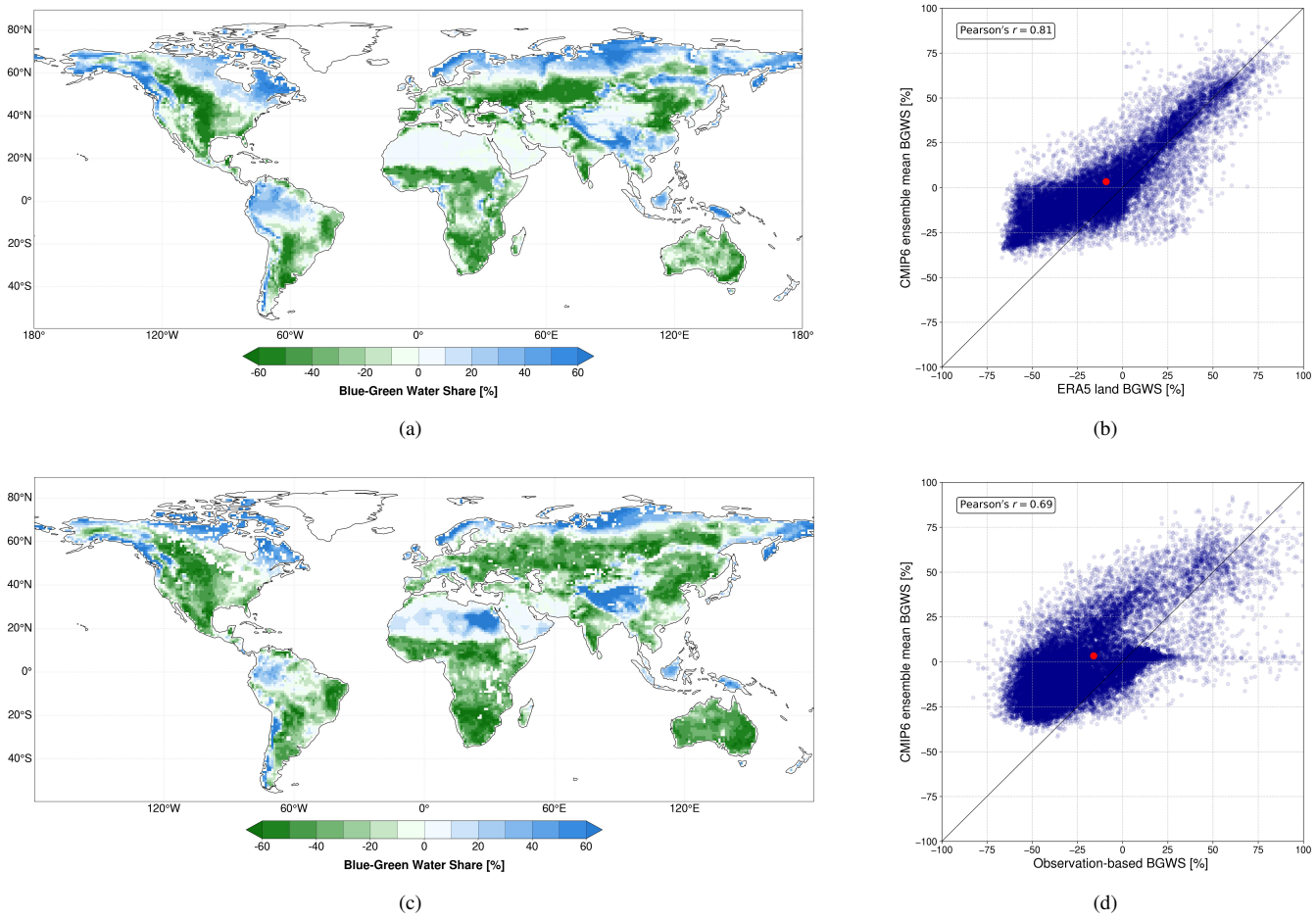


Figure S1. Comparison of historical (1985-2014) BGWS across reanalysis, observation-based data, and ensemble mean of 11 CMIP6 ESMs. Global BGWS maps for **a**, ERA5 land, and **b**, observation-based data. Scatter plots comparing **c**, ERA5 land, and **d**, observation-based data, with the ensemble mean. Density of data points in the scatter plots is represented by varying transparency, with higher opacity indicating areas of greater data density. The red dot marks the global mean for each dataset. Observation-based data products are GPCC for precipitation, G-RUN for runoff, and GLEAM for transpiration.

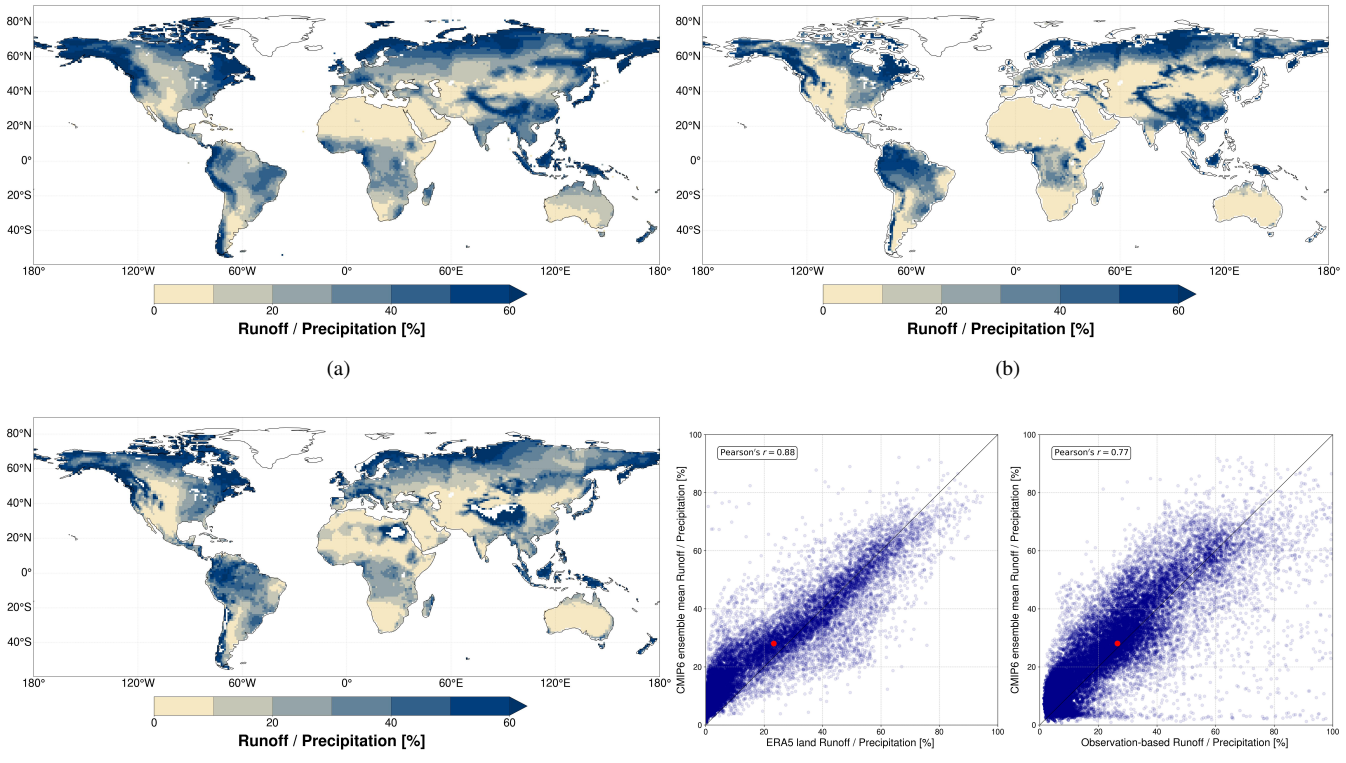


Figure S2. Comparison of historical (1985-2014) runoff/precipitation ratio across reanalysis, observation-based data, and ensemble mean of 11 CMIP6 ESMs. Global runoff/precipitation ratio maps for a, ensemble mean based on 11 CMIP6 ESMs, b, ERA5 land, and c, observation-based data. Scatter plots comparing d, ERA5 land, and e, observation-based data, with the ensemble mean. Density of data points in the scatter plots is represented by varying transparency, with higher opacity indicating areas of greater data density. The red dot marks the global mean for each dataset. Observation-based data products are GPCC for precipitation and G-RUN for runoff.

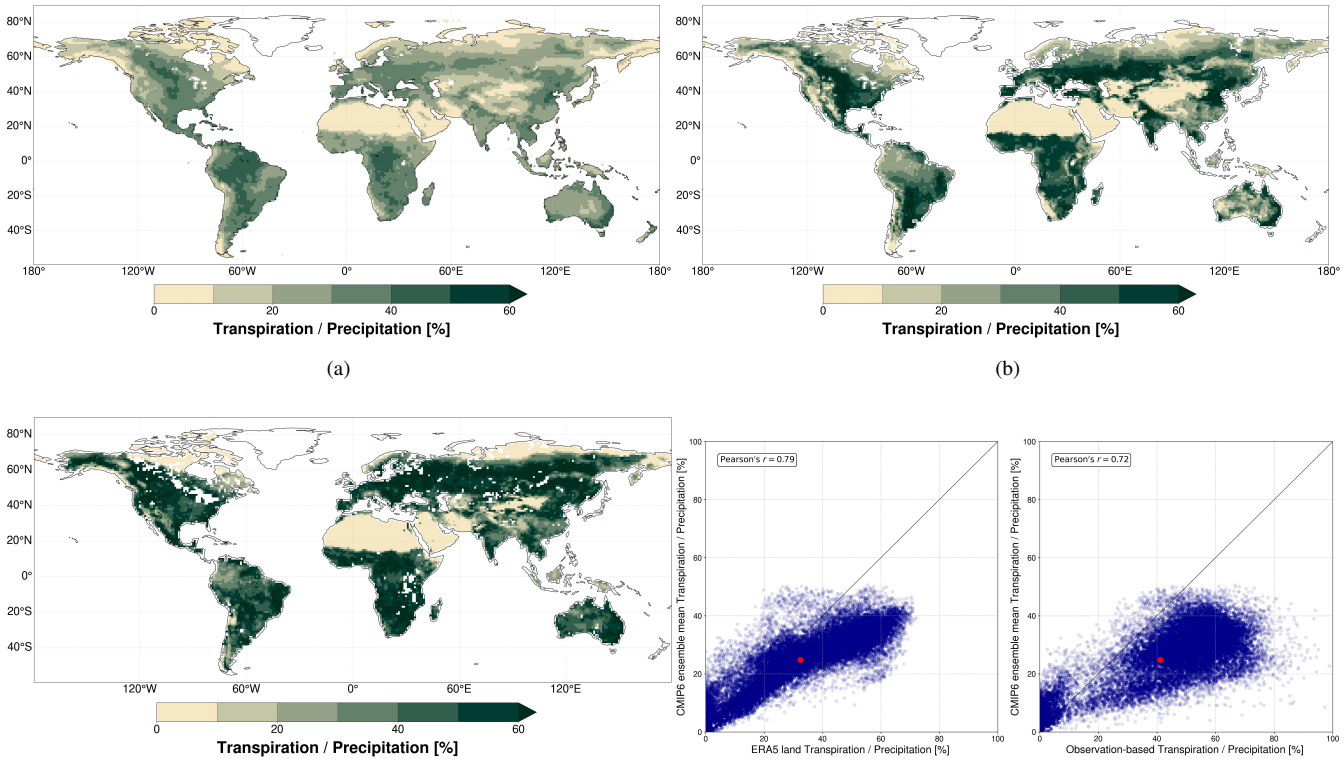


Figure S3. Comparison of historical (1985-2014) transpiration/precipitation across reanalysis, observation-based data, and ensemble mean of 11 CMIP6 ESMs. Global transpiration/precipitation ratio maps for a, ensemble mean based on 11 CMIP6 ESMs, b, ERA5 land, and c, observation-based data. Scatter plots comparing d, ERA5 land, and e, observation-based data, with the ensemble mean. Density of data points in the scatter plots is represented by varying transparency, with higher opacity indicating areas of greater data density. The red dot marks the global mean for each dataset. Observation-based data products are GPCC for precipitation and GLEAM for transpiration.

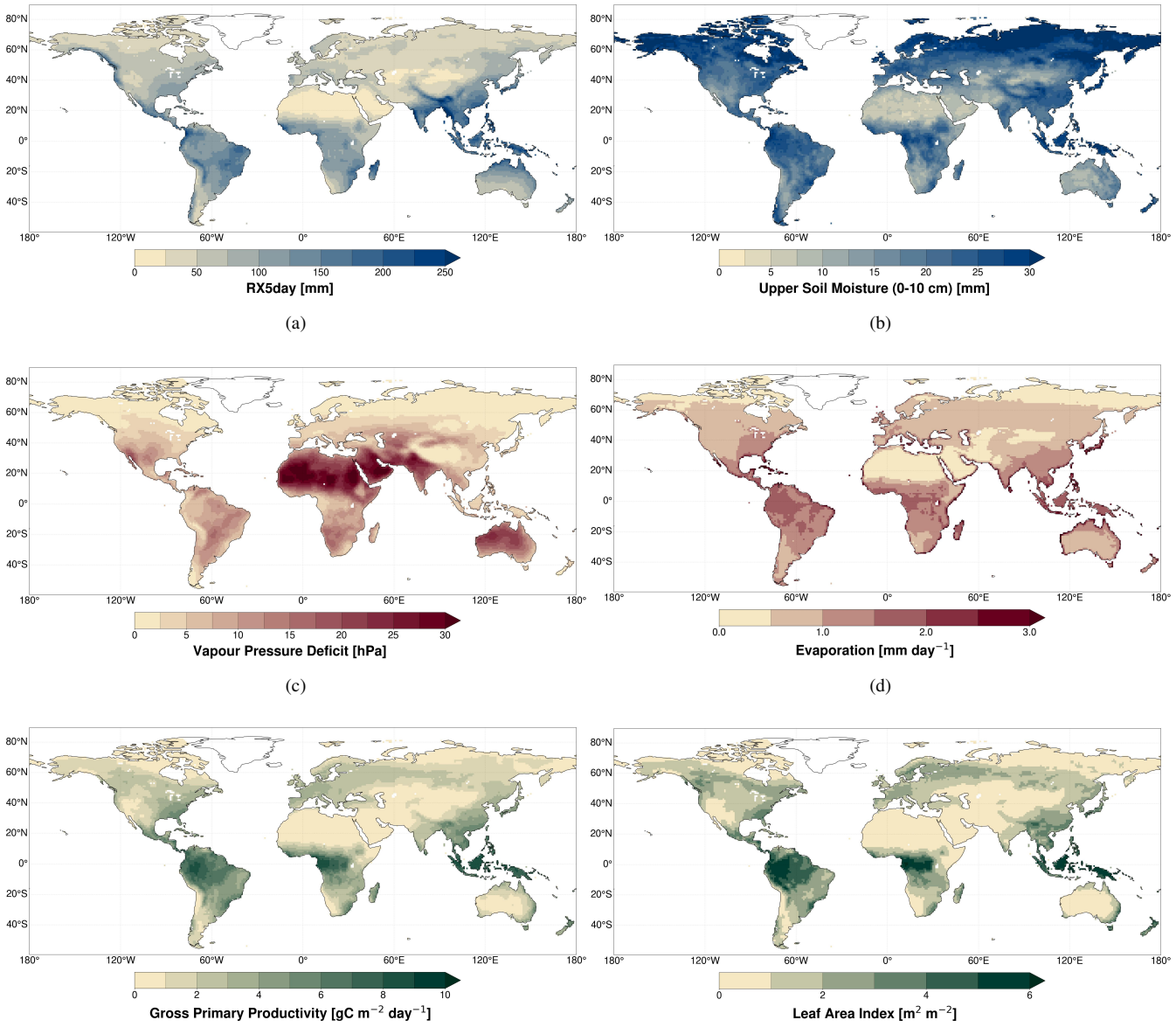
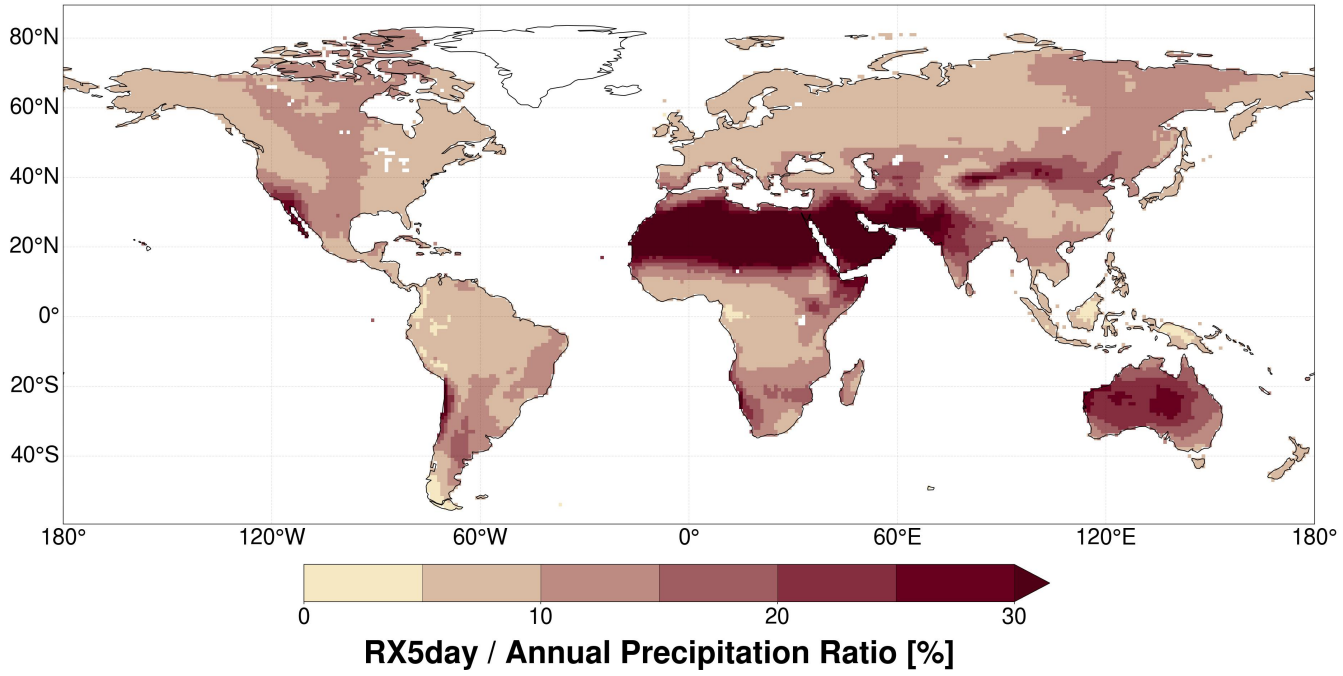
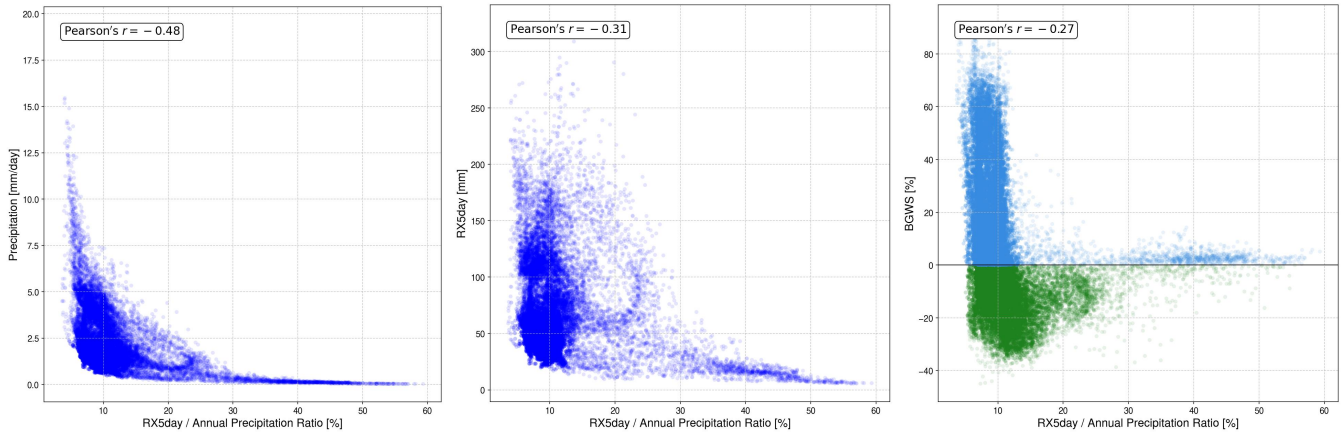


Figure S4. Historical ensemble mean (1985-2014) based on 11 CMIP6 ESMs. a, Annual maximum five-day consecutive precipitation (RX5day) [mm], b, upper soil moisture (0-10 cm) [mm], c, vapour pressure deficit [hPa], d, evaporation [mm day^{-1}], e, gross primary productivity [gC m^{-2} day^{-1}], and f, leaf area index [m^2 m^{-2}].

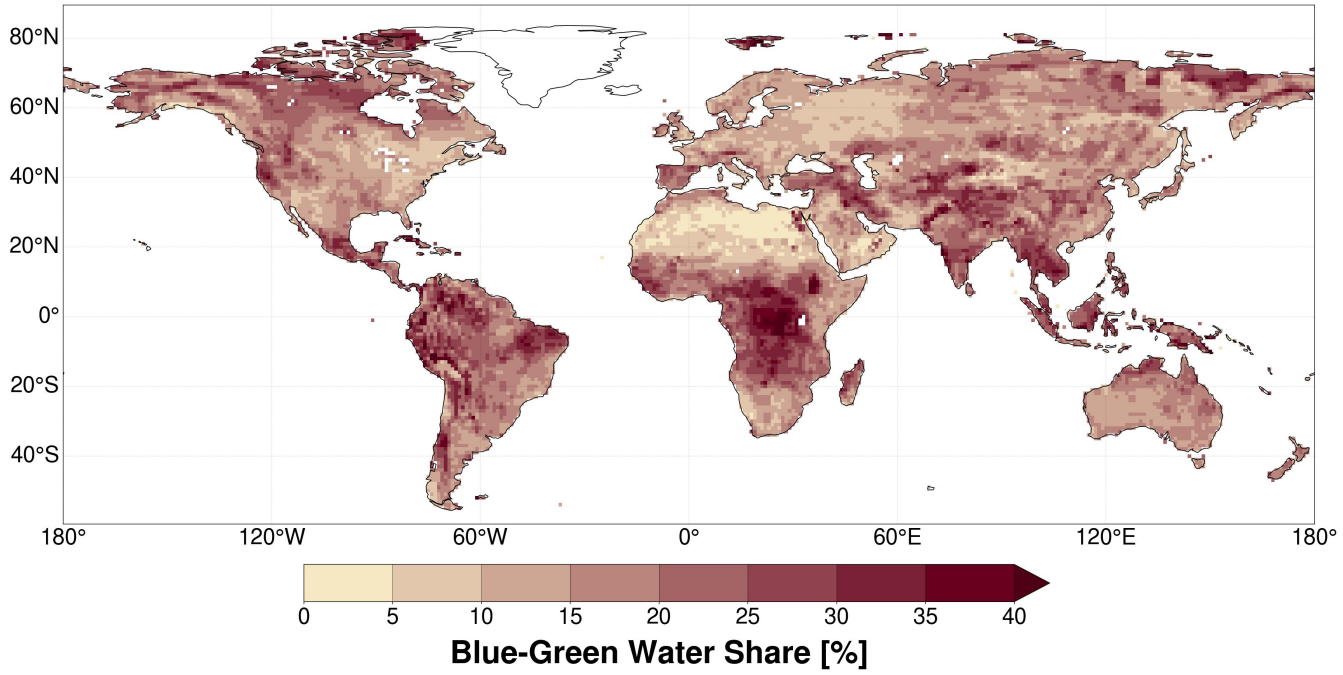


(a)

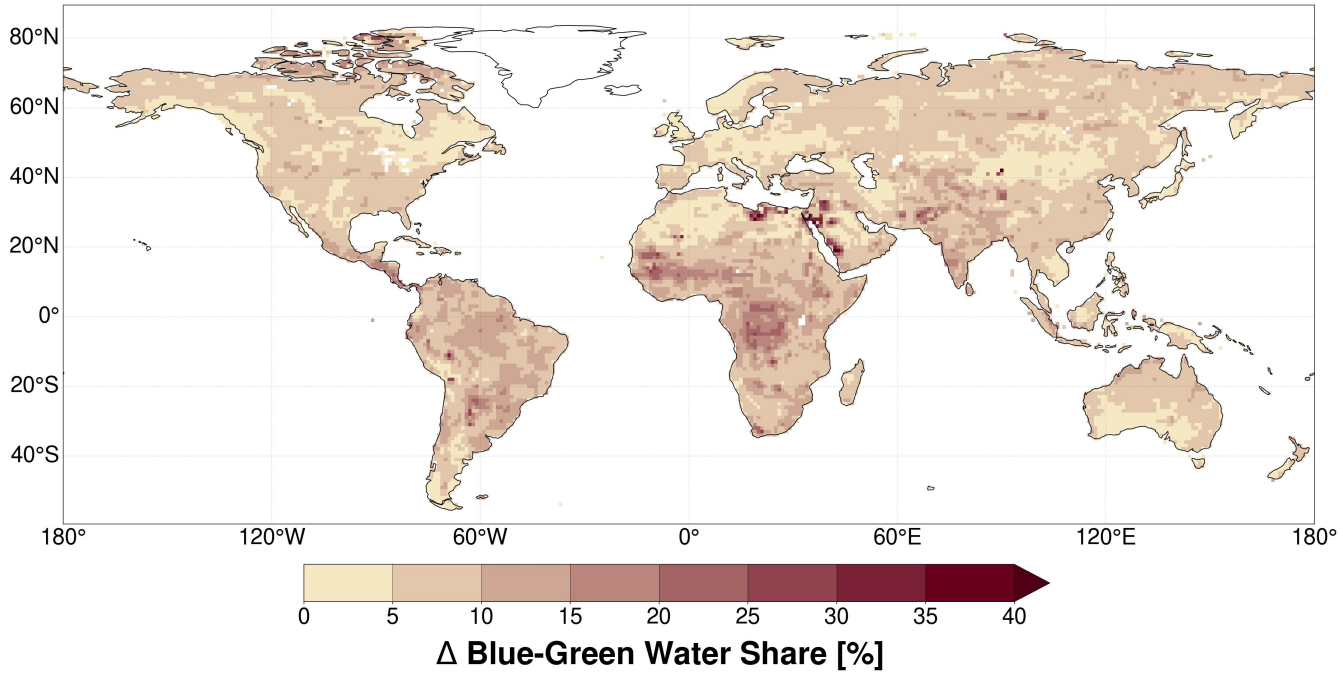


(b) (c) (d)

Figure S5. Historical (1985–2014) RX5day ratio (RX5day/annual precipitation) and its relationships with precipitation metrics and the BGWS based on 11 CMIP6 ESMs. a, Ensemble mean map of the RX5day ratio [%]. b-d, Scatter plots showing the relationship of the RX5day ratio with b, mean precipitation, c, RX5day, and d, BGWS. In d, blue points representing positive BGWS values and green points representing negative values. Pearson's correlation coefficient (r) is indicated in each scatter plot.



(a)



(b)

Figure S6. Ensemble standard deviation of BGWS [%] based on 11 CMIP6 ESMs. a, Historical period (1985–2014), and b, projected change (2071–2100 minus 1985–2014) under the SSP3-7.0 scenario. Each grid cell represents the standard deviation of BGWS values across the model ensemble.

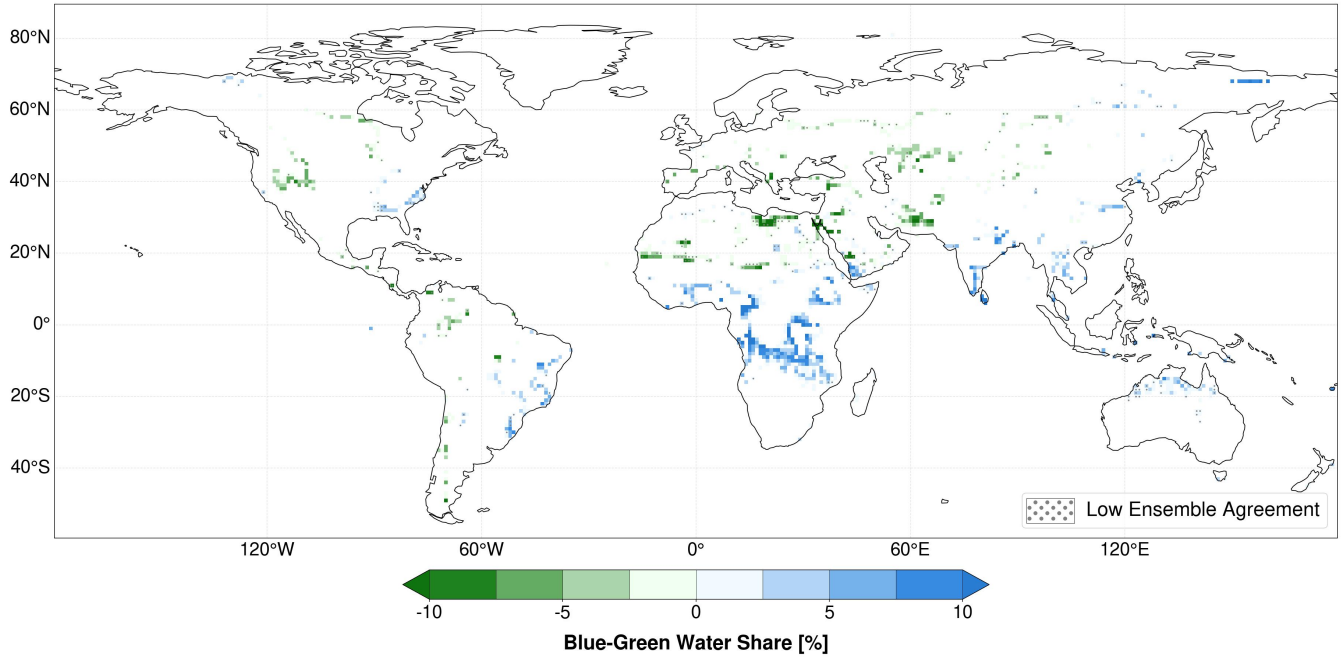


Figure S7. Ensemble mean BGWS in the far future (2071-2100) under the SSP3-7.0 scenario based on 11 CMIP6 ESMs. Only grid cells that experience a shift from a blue water to a green water regime or vice versa are coloured. Stipplings denote regions of low ensemble agreement where less than 70% of the eleven ensemble members agree on the sign of the change (SSP3-7.0 – historical).

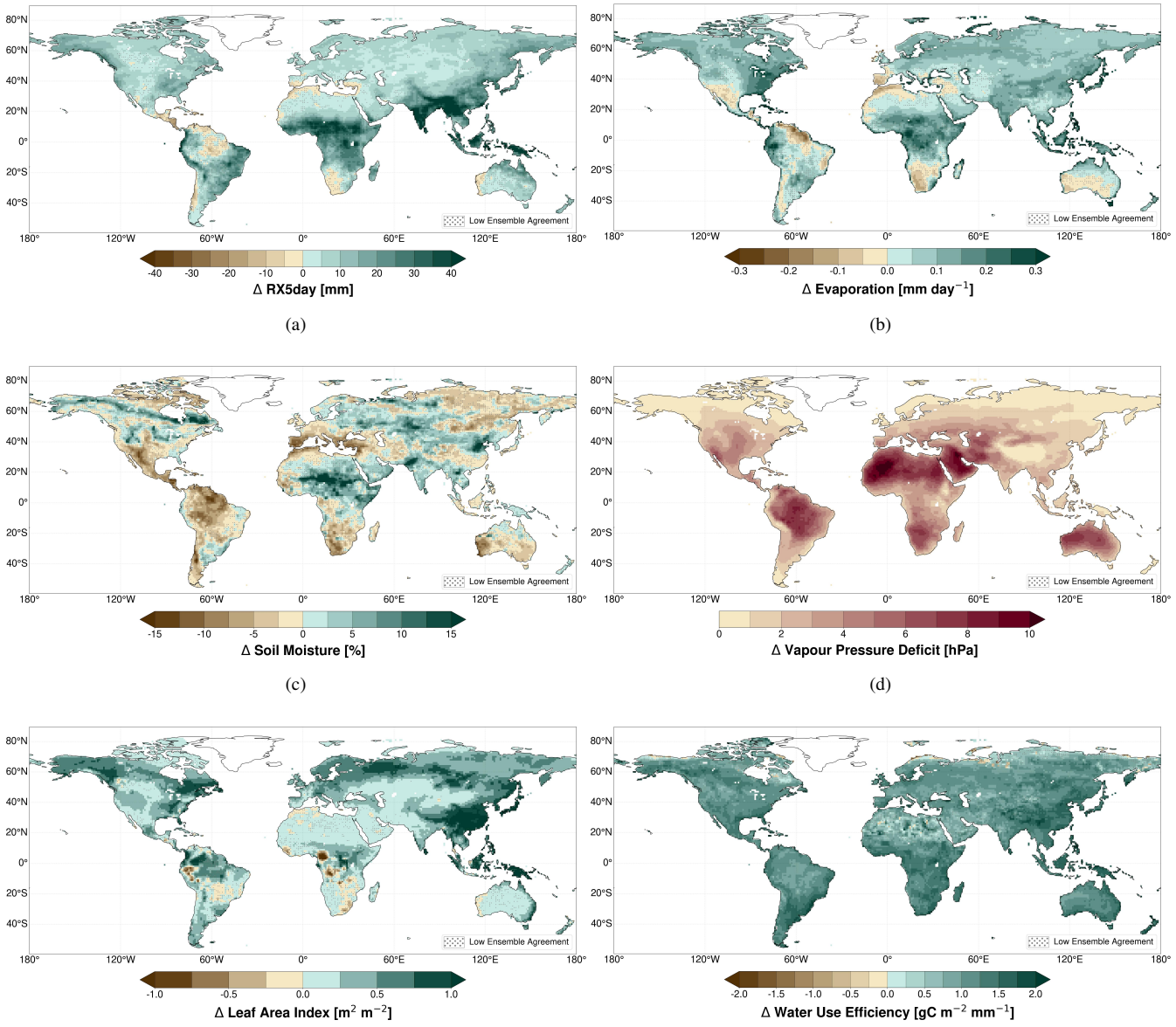
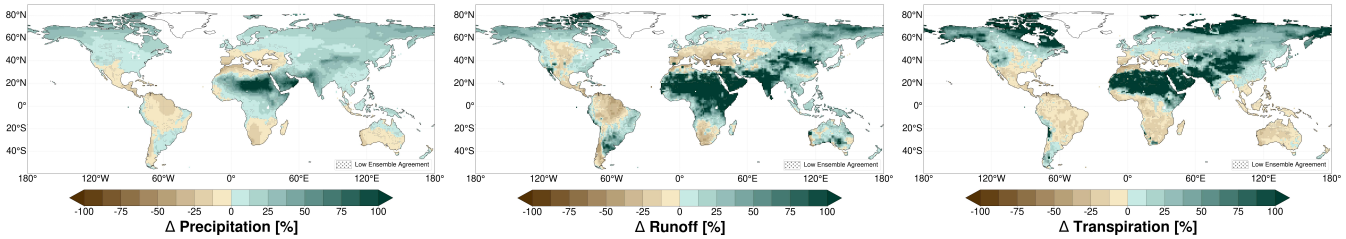
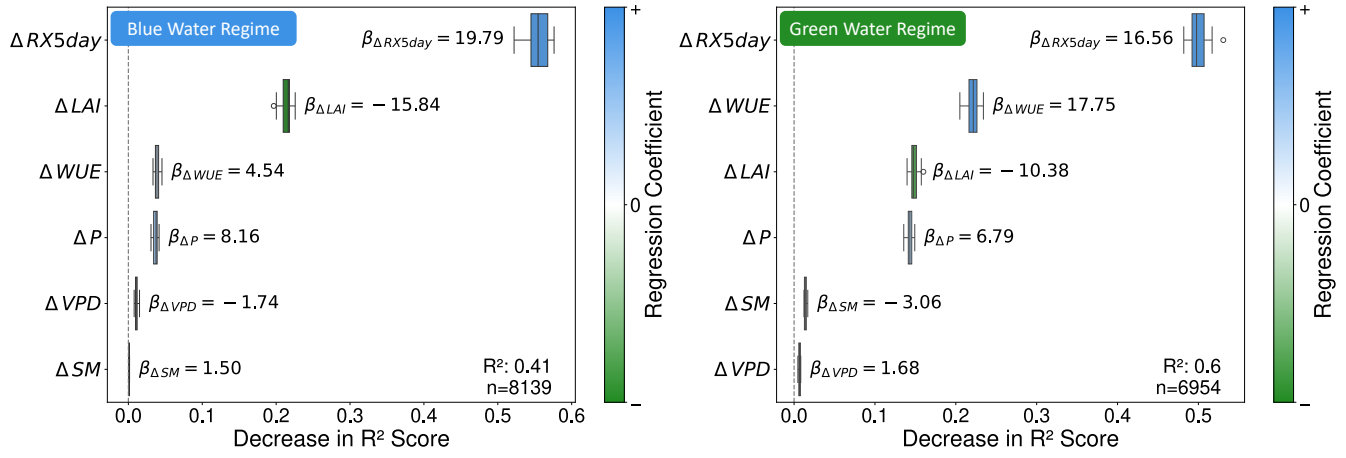


Figure S8. Ensemble mean change (2071-2100 minus 1985-2014) under the SSP3-7.0 scenario based on 11 CMIP6 ESMs. a, Δ annual maximum five-day consecutive precipitation (RX5day) [mm], b, Δ evaporation [mm day^{-1}], c, Δ soil moisture [%], d, Δ vapour pressure deficit [hPa], e, Δ leaf area index [m^2 m^{-2}], and f, Δ water-use efficiency [gC m^{-2} mm^{-1}]. Stipplings denote regions of low ensemble agreement where less than 70% of the eleven ensemble members agree on the sign of the change.



(a) (b) (c)

Figure S9. Ensemble mean relative change (2071-2100 minus 1985-2014) under the SSP3-7.0 scenario based on 11 CMIP6 ESMs. a, Δ precipitation [%], b, Δ runoff [%], and ca, Δ transpiration[%]. Stipplings denote regions of low ensemble agreement where less than 70% of the eleven ensemble members agree on the sign of the change.



(a) (b)

Figure S10. Variable importance of the training data across the blue and green water regimes based on 11 CMIP6 ESMs. Variable importance plots present the decrease in accuracy score of selected predictors (Methods) based on the permutation importance analysis for the global, a, blue, and b, green water regime. Colour of boxes indicate the regression coefficient with blue presenting a positive and green a negative value. Model performance on training dataset (R^2) and amount of grid cells considered in this analysis (n) are indicated.

Random Forest Analysis

We employ Random Forest (RF) regression to investigate the relationships between predictor variables and the target variable (ΔBGWS), leveraging its capability to handle non-linear interactions and complex variable relationships. This analysis serves as a complementary approach to our primary method, multiple linear regression (MLR), allowing us to compare variable importance rankings derived from SHAP (SHapley Additive exPlanations) values to those obtained through permutation importance in MLR. The similarity in results between these two fundamentally different approaches supports the robustness of our findings (Fig. 3a,b, S10, and S11).

As with the MLR analysis, the dataset is divided into 70% training and 30% testing subsets, with the RF model trained solely on the training data. Hyperparameter tuning is performed, evaluating combinations of key parameters such as maximum tree depth, minimum samples per leaf, minimum samples per split, and the number of trees in the forest. To prevent overfitting, we define a threshold of 10% for the difference between training and testing (R^2) scores, excluding models exceeding this threshold. Separate hyperparameter grids are defined for the blue and green water regimes, reflecting their distinct characteristics. For the blue water regime, the optimal configuration includes a maximum depth of 20, a minimum of 10 samples per leaf, a minimum split size of 2 samples, and 200 estimators. For the green water regime, the configuration is adjusted to a maximum depth of 15, a minimum of 5 samples per leaf, a minimum split size of 2 samples, and 300 estimators.

To evaluate variable importance in our Random Forest models, we compute SHAP values (Lundberg and Lee, 2017). SHAP values are based on Shapley values from cooperative game theory and uniquely satisfy local accuracy, consistency, and missingness properties, making them a robust framework for interpreting machine learning models. These values assign each predictor an additive contribution to the model's output, enabling both local (individual prediction) and global (model-wide) interpretability. We calculate SHAP values for both training and testing datasets to ensure that patterns in variable importance are consistent across seen and unseen data. SHAP summary plots are employed to visualise the direction and magnitude of each predictor's contribution to the model output, with mean absolute SHAP values summarising overall variable importance (Fig. S11).

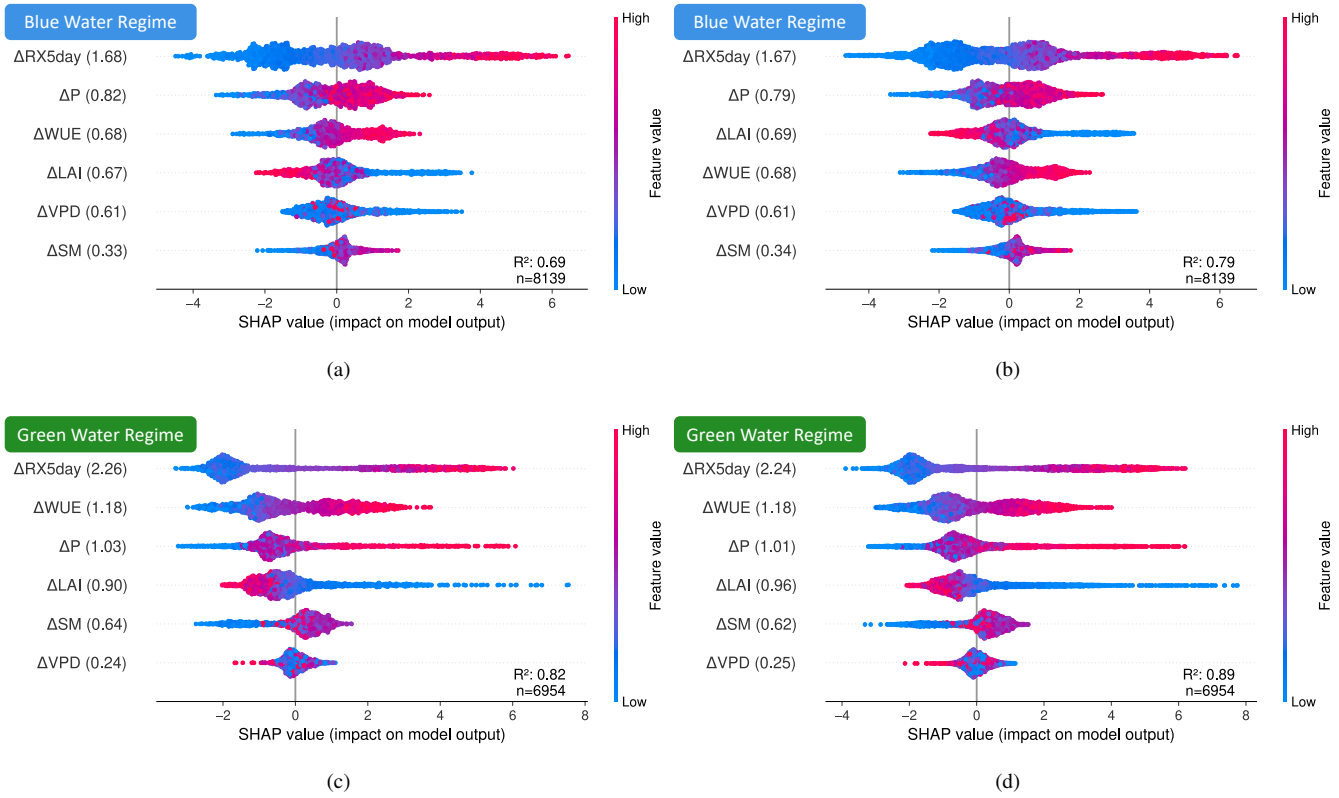


Figure S11. Random Forest-based variable importance of predictors for blue and green water regimes based on 11 CMIP6 ESMs. a, b, Variable importance based on SHAP values for the blue water regime, shown for the testing dataset, a, and the training dataset, b. c,d, Corresponding results for the green water regime, shown for the testing dataset, c, and the training dataset,d. Positive SHAP values indicate a positive impact on model output, with dot colours representing feature values; note that low feature values do not necessarily correspond to negative values, as changes are non-uniform, or for variables like VPD, strictly positive. Numbers in brackets after variable names indicate the mean absolute SHAP value, representing each variable's overall impact on the model output. Model performance (R^2) and the number of grid cells analysed (n) are provided for each dataset.