Supplement of Hydrol. Earth Syst. Sci., 30, 3647–3673, 2026
https://doi.org/10.5194/hess-30-3647-2026-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Continental-scale prediction of hydrologic signatures and processes

Ryoko Araki et al.

Correspondence to: Ryoko Araki (raraki8159@sdsu.edu, raraki@ucsb.edu) and Hilary K. McMillan (hmcmillan@sdsu.edu)

Disclaimer: Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

Table S1. Number of watershed samples from the Caravan and GAGES-II datasets, along with a summary of the landscape attributes used for random forest (RF) training and prediction in this study. See Figure S1 for the subset structure.

Dataset	Subset		Signature values	Attributes used for RF training	Attributes used for RF prediction	Number of samples
Caravan US subset (#)	Passed quality control	Only available in Caravan	Calculated from observed data	-	-	2,717
		Overlapping with GAGES-II (*)	Calculated from observed data	Caravan attributes, substituted with GAGES-II equivalent (same as Table x)	-	4,748
			Not Predicted	-	Some attributes are not available	11
	Did not pass quality control	Only available in Caravan	Predicted	-	Caravan attributes	2,424
			Not Predicted	-	Some attributes not available	206
		Overlapping with GAGES-II (**)	Predicted	-	Caravan attributes, substituted with GAGES-II equivalent	618
	Not within the Continental United States		Not included	-	-	1,479
	Overlapping CAMELS and HYSETS gauges		Not included	-	-	641
GAGES-II	Not overlapping with Caravan (##)		Calculated from observed data	-	-	2,807
			Predicted	-	Caravan attributes, substituted with GAGES-II equivalent. Climate attributes were calculated from gridMET data. PET and aridity were bias-corrected to match values based on ERA FAO-PM using linear regression.	843
			Not predicted	-	Some attributes are not available	40
	Overlapping with Caravan; the sum of (*) and (**)		See above rows	See above rows	See above rows	5,377

Total gages plotted on the map	Signatures other than overland flow; the sum of (#) and (##)	See above rows	See above rows	See above rows	The sum of (#) and (##) is 14,403. However, the signature calculation returned errors or attributes not available for prediction for 257 watersheds; and excluding insufficient datasets; therefore, results were obtained for a **total of 14,146 watersheds**
	Overland flow signatures (excluding snow-dominated watersheds, where the snow fraction > 0.2)	See above rows	See above rows	See above rows	**10,432**

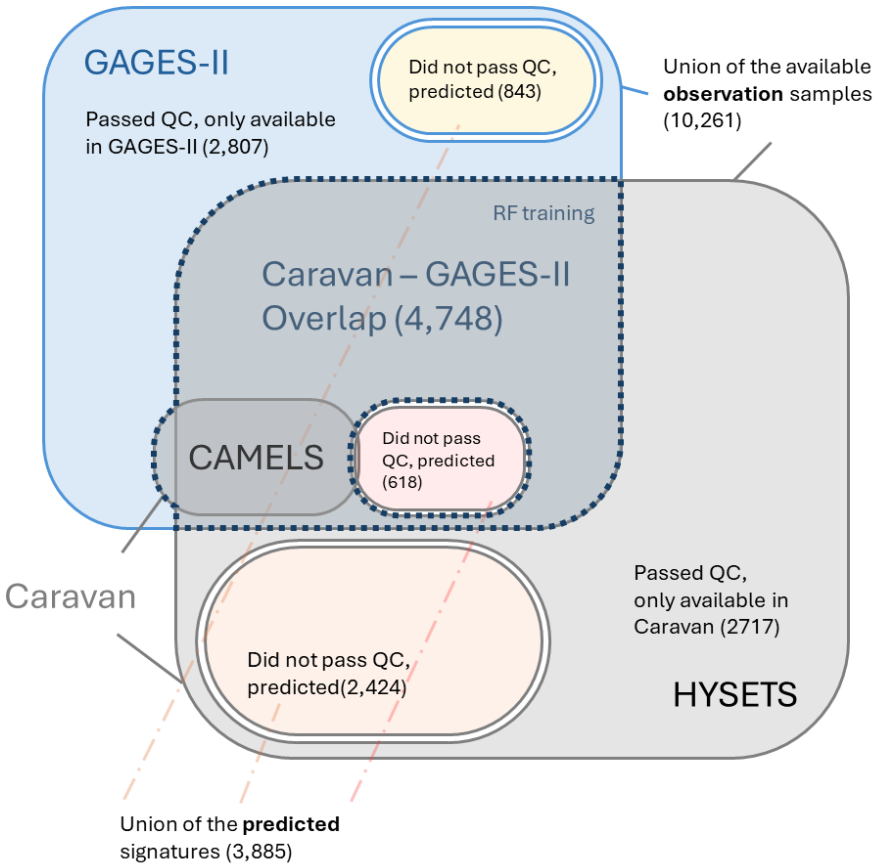


Figure S1: Venn diagram of samples used in this study. Watershed samples outside of CONUS are excluded from the figure for clarity. The color scheme matches that of Figure 1 in the manuscript.

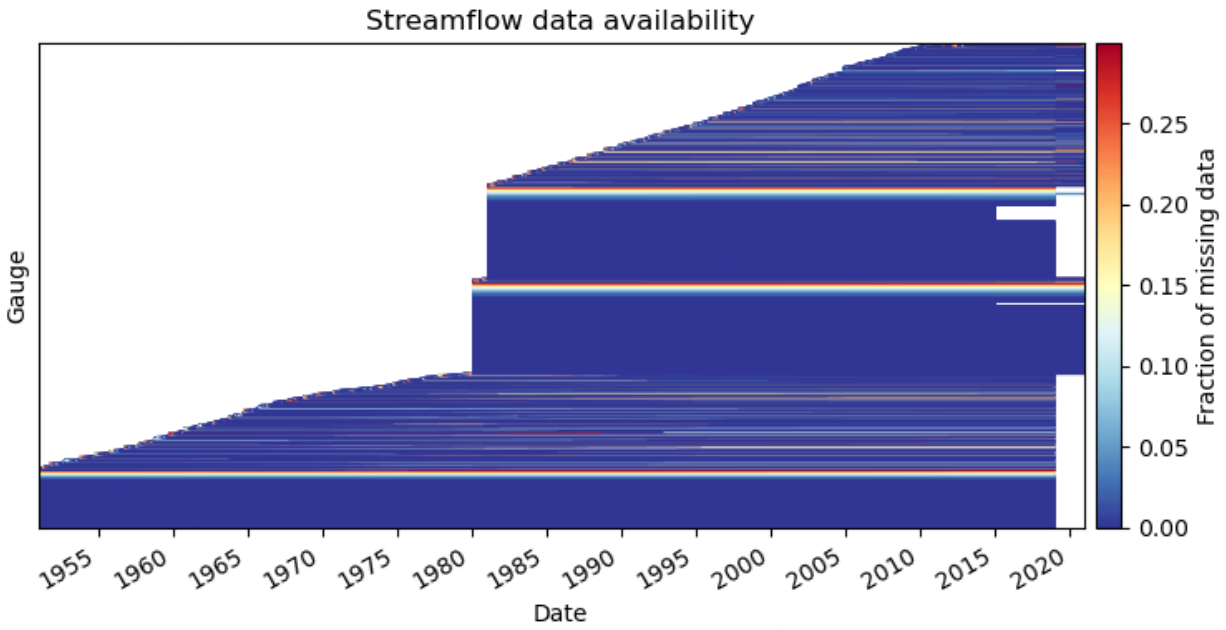


Figure S2: Record availability and data completeness for all gauge stations eligible for signature calculations, across both data sources (Caravan and GAGES-II). Each horizontal bar represents the available record period of an individual gauge, spanning from its start to end date. Bar color indicates the fraction of missing data within the record period.

Table S2: TOSSH parameters were tuned based on the USGS gauge ID to capture overland flow signatures (used as input for an event separation algorithm, *util_EventSeparation.m*: https://tosshtoolbox.github.io/TOSSH/_static/matlab/TOSSH_code/TOSSH/TOSSH_code/utility_functions/util_EventSeparation.html)

The first 2 digits of the USGS gauge ID	min_termination [hours]	Min_duration [hours]	min_intensity_day [mm/day]	min_intensity_day_during [mm/day]	Max_recessiondays [days]
01	72	24	7.2	7.2	8
02	72	24	7.2	7.2	8
03	72	24	7.2	7.2	8
04	48	24	4.8	4.8	8
05	48	24	4.8	4.8	8
06	48	24	4.8	4.8	8
07	48	24	4.8	4.8	8
08	48	24	4.8	4.8	8
09	48	24	2.4	2.4	8
10	48	24	2.4	2.4	8
11	48	24	2.4	2.4	8
12	72	24	4.8	4.8	8
13	72	24	4.8	4.8	8
14	48	24	4.8	4.8	8
18	48	24	9.6	9.6	8
20	48	24	4.8	4.8	8
22	48	24	4.8	4.8	8
26	48	24	4.8	4.8	8
27	72	24	4.8	4.8	8
28	72	24	4.8	4.8	8
30	48	24	4.8	4.8	8
36	48	24	4.8	4.8	8
39	48	24	4.8	4.8	8
90	48	24	4.8	4.8	8

Table S3: TOSSH parameters were tuned based on the USGS gauge ID for the recession delineation algorithm (used as input for a function called *util_RecessionSegments.m*: https://tosshtoolbox.github.io/TOSSH/_static/matlab/TOSSH_code/TOSSH/TOSSH_code/utility_functions/util_RecessionSegments.html). If the 95th percentile of daily streamflow is below 1 mm day^{-1}, the flow is classified as low; otherwise, it is classified as normal.

flow	**Recession_length [days]**	**N_start [timestep]**	**eps [-]**	**filter_par [-]**
normal	5	0	0.08	0.925
low	10	0	0.02	0.925

Table S4: Climate attributes from the Caravan and GAGES-II datasets used for clustering and defining climate regions. The result of clustering is shown in Figure S2.

Variable name	Description	Source
PPTAVG_BASIN	Mean annual precipitation	GAGES-II
T_AVG_BASIN	Mean annual air temperature	GAGES-II
PET	Mean annual potential evapotranspiration rate estimated from mean monthly air temperature and latitude using Hamon's (1961) equation.	GAGES-II
RH_BASIN	Mean Relative Humidity	GAGES-II
ARIDITY_GAGES2	Aridity index calculated from GAGESII attributes; ratio of mean PET and mean precipitation	GAGES-II
moisture_index	Mean annual moisture index (Knoben et al., 2018) based on monthly aridity describing the water- or energy-limited conditions	Caravan1.4/ERA-5
SNOW_PCT_PRECIP	Mean snow percent of the total precipitation estimate	GAGES-II
PRECIP_SEAS_IND	Precipitation seasonality index (Markham, 1970; Dingman, 2002)	GAGES-II
input_seasonality	Seasonality index (Markham, 1970; Dingman, 2002) calculated for snow water equivalent	(Hammond et al., 2023)
seasonality	Largest changes in monthly moisture index in a year (Knoben et al., 2018)	Caravan1.4/ERA-5
input_PET_synchrony	Water and energy synchrony estimated from a correlation between mean monthly surface water input and mean monthly PET, derived from gridMET and University of Arizona SWE product 1991-2020	(Hammond et al., 2023)
WD_BASIN	Mean annual number of days of measurable precipitation	GAGES-II
high_prec_freq	Frequency of high precipitation days, where precipitation ≥5 times the mean daily precipitation	Caravan1.4/ERA-5
high_prec_dur	Average duration of low precipitation events (number of consecutive days where precipitation ≥5 times mean daily precipitation)	Caravan1.4/ERA-5
low_prec_freq	Frequency of low precipitation days, where precipitation <1 mm/day	Caravan1.4/ERA-5
low_prec_dur	Average duration of low precipitation events (number of consecutive days where precipitation <1 mm/day)	Caravan1.4/ERA-5
FST32F_BASIN	Mean day of the year of the first freeze	GAGES-II
LST32F_BASIN	Mean day of the year of the last freeze	GAGES-II

Section S1: Excluded Caravan landscape attributes in the study. We removed strongly correlated variables (Spearman's $\rho > 0.8$ or < -0.8). Rationales are given in the following.

- **ele_mt_sav / ele_mt_smn / ele_mt_smx:** The average, maximum, and minimum elevation are strongly correlated with each other ($|\rho|>0.8$). Picked the average elevation (ele_mt_sav) as a representative of the watershed elevation characteristics.
- **sgr_dk_sav / slp_dg_sav:** Stream gradient (sgr_dk_sav) and terrain slope (slp_dg_sav) are strongly correlated ($|\rho|>0.8$). Picked the terrain slope (slp_dg_sav) as a stronger predictor of hydrologic process control.
- **cly_pc_sav / snd_pc_sav / slt_pc_sav:** There is a strong correlation ($|\rho|>0.8$) between sand fraction (snd_pc_sav) and clay fraction (cly_pc_sav), as well as between sand fraction (snd_pc_sav) and silt fraction (slt_pc_sav). Picked clay fraction (cly_pc_sav) and silt fraction (slt_pc_sav) as a representative of a soil texture, as information on clay and silt content constrains sand content and is more relevant to water retention capacity.
- **ppd_pk_sav / pop_ct_usu / urb_pc_sse / nli_ix_sav / hft_ix_s93/ hft_ix_s09 / rdd_mk_sav:** There is a strong correlation ($|\rho|>0.8$) among these attributes representing anthropogenic impacts. Removed human footprint indices (hft_ix_s93 & hft_ix_s09) because they are primarily defined based on land-covers, whose information is already contained in and interferes with land-cover attributes. Road density (rdd_mk_sav), nighttime light (nli_ix_sav), and urban extents (urb_pc_sse) are considered as a result of high population density (ppd_pk_sav) and high population count (pop_ct_usu). Picked population density (ppd_pk_sav) as a primary representative of anthropogenic impacts on the watershed through urbanization. Note that gross domestic product (gdp_ud_sav) and the human development index (hdi_ix_sav) were excluded from the analysis due to their weak association with hydrologic responses.
- **aridity/ moisture_index:** There is a strong correlation ($|\rho|>0.8$) between moisture index (moisture_index) and aridity index (aridity). Picked aridity index (aridity) because it is widely used and shown to be a strong predictor of hydrologic processes (Budyko & Miller, 1974; Meira Neto et al., 2020).
- **low_prec_dur / high_prec_dur:** Low and high precipitation duration are strongly correlated ($|\rho|>0.8$). Picked low precipitation duration (low_prec_dur) because it characterizes the extended dry period.
- **prm_pc_sse**: Permafrost extent was excluded from the analysis because its values were zero in most watersheds over the continental United States, and it appeared to be the least important attribute across signatures and regions.

Section S2: Preliminary random forest experiments with Caravan dataset subsets

To evaluate the impact of dataset selection on random forest training, we conducted preliminary experiments using subsets of the Caravan dataset. In the Fig. S3,

- **All Caravan**: All watersheds within the Caravan dataset with sufficient data quality (gauges with insufficient data removed).
- **CAMELS subset:** Caravan subset of CAMELS watersheds; watersheds within the CAMELS samples. Trained using Caravan hydroclimatic data and attributes. CAMELS is widely recognized as a benchmark for high-quality, natural watersheds.
- **GAGES2 subset:** Caravan subset of GAGES-II watersheds; Watersheds overlapping between Caravan and GAGES-II, using Caravan hydroclimatic data and attributes. USGS's GAGES-II is developed to represent and investigate natural and human-altered flow regimes.
- **GAGES2 subset + attrs:** Identical to the GAGES2 subset in terms of watershed selection, but model training employed Caravan-equivalent GAGES-II attributes (Table 1) rather than the standard Caravan attributes.
- **GAGES2-Ref subset:** Watersheds overlapping between Caravan and GAGES-II, but restricted to those designated as Reference by GAGES-II classification.

It should be noted that the reported R^2 values reflect training performance and does not represent true predictive skill. Key findings from these experiments are as follows:

- Training on *All Caravan* watersheds resulted in relatively poor performance
- The *CAMELS subset* yielded the strongest performance
- The *GAGES2 subset* performed slightly worse than the *CAMELS subset*, but remained comparable.
- Incorporating Caravan-equivalent GAGES-II attributes in the GAGES2 subset improved performance by several percentage points (*GAGES2 subset + attrs* vs. *GAGES2 subset*)

On the basis of these observations, we elected to train our random forest models using the *GAGES2 subset + attrs*, as reported in the main manuscript. We hypothesize that the better performance of the GAGES-II subset may reflect higher-quality streamflow data leading to more accurate signature extraction, or stronger signature–attribute relationships in these watersheds (e.g., less impacted by human influences, climatic influences).

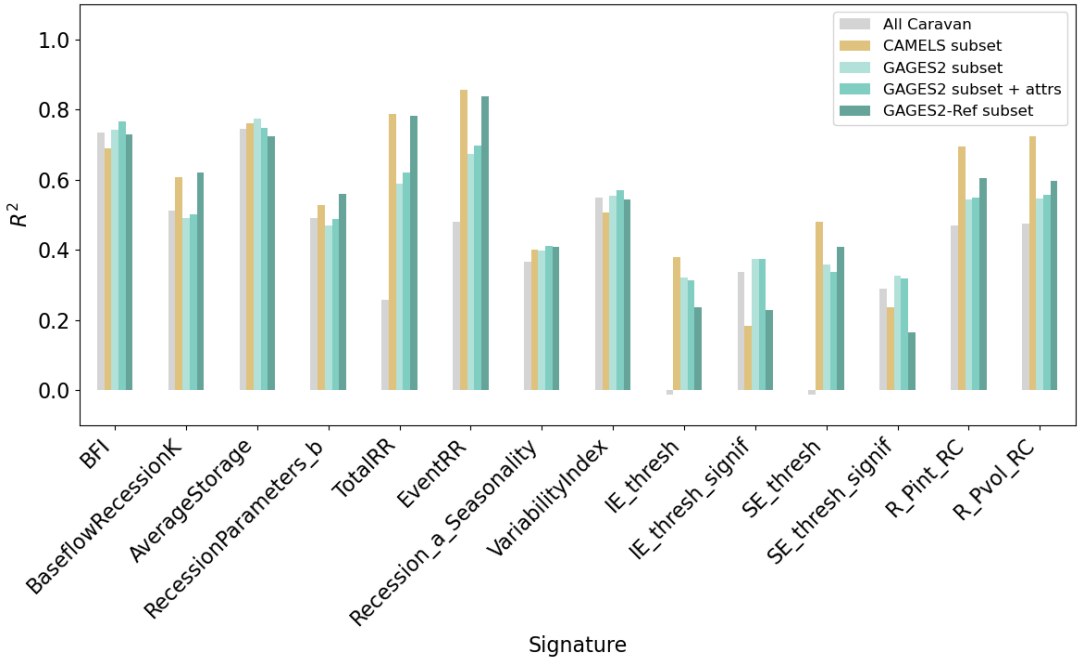


Figure S3: Goodness of fit of random forest models (predicted vs. observed signatures) from sub-setting experiments. The subset names correspond to the explanation in Section S2.

Figure S4: (a) Map of U.S. hydroclimate clusters used in the Random Forest regional experiments. Six climate regions were identified using a Gaussian mixture model in Scikit-learn (Pedregosa et al., 2011) based on hydroclimate attributes from Caravan, GAGES-II, and Hammond et al. (2023) listed in Table S4. Separate Random Forest models were then trained for each region. (b) Climate characteristics of the six clusters. Values are standardized by the overall dataset mean and standard deviation. Blue indicates that all quantiles (25th, 50th, and 75th percentiles) are above average, while pink indicates that they are below average.

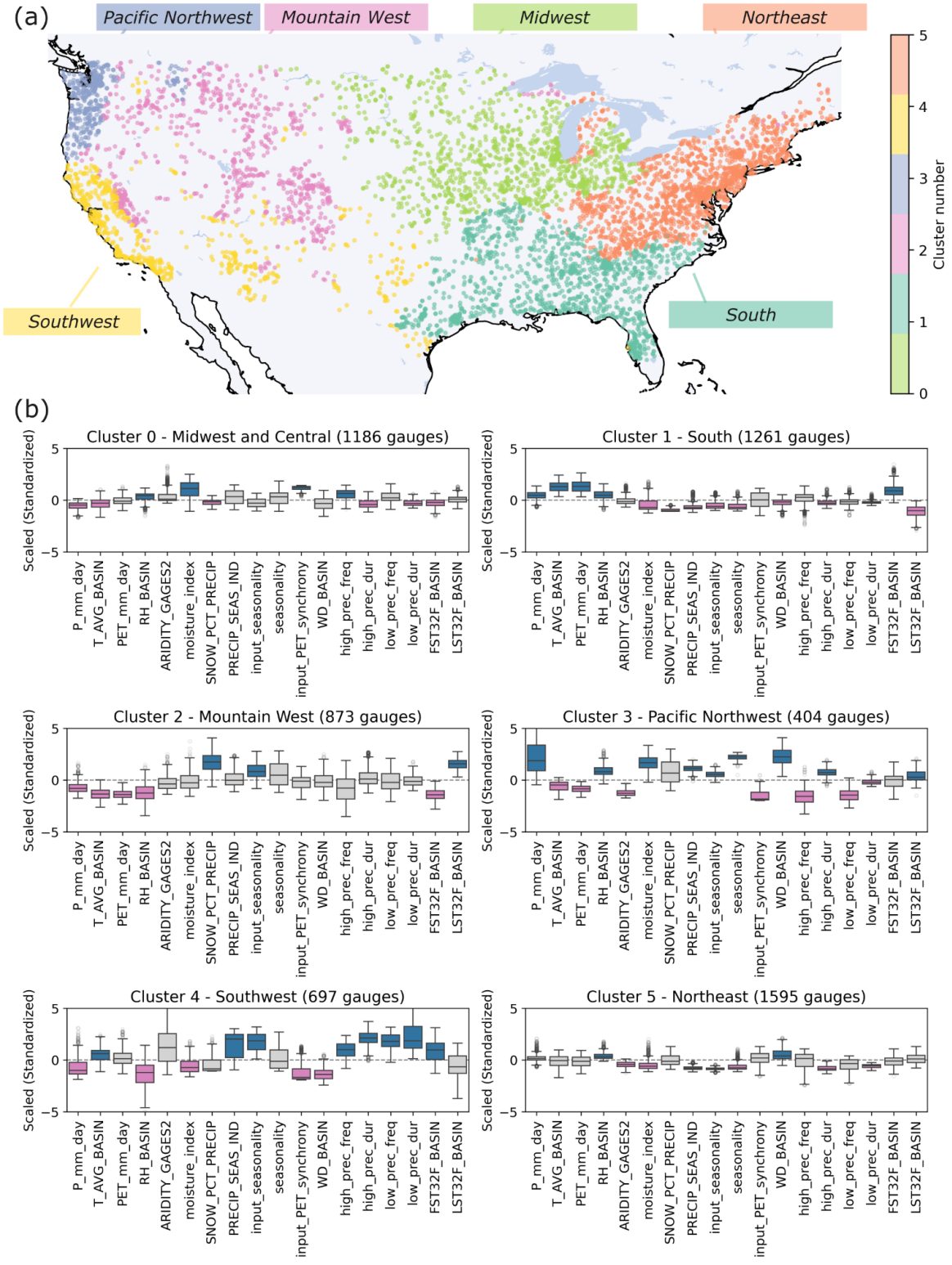
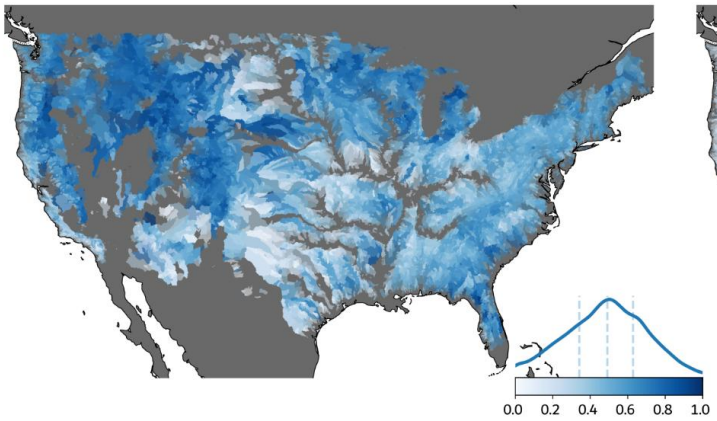
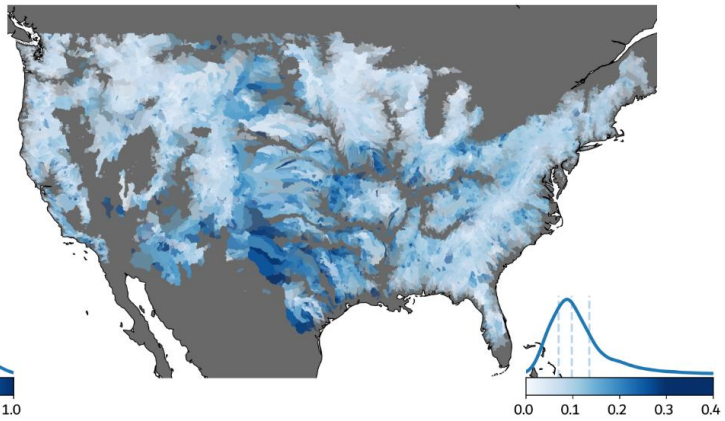


Figure S5. Hydrologic signature values analyzed to generate Figures 2 and 3 (a–c). The chart above the color bar shows the distribution of signature values. Dashed lines indicate the 0.25, 0.5, and 0.75 quantiles.

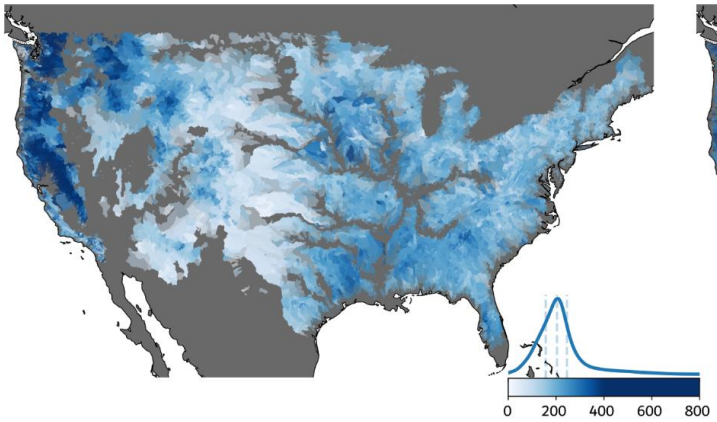
(a.1) Baseflow - BFI



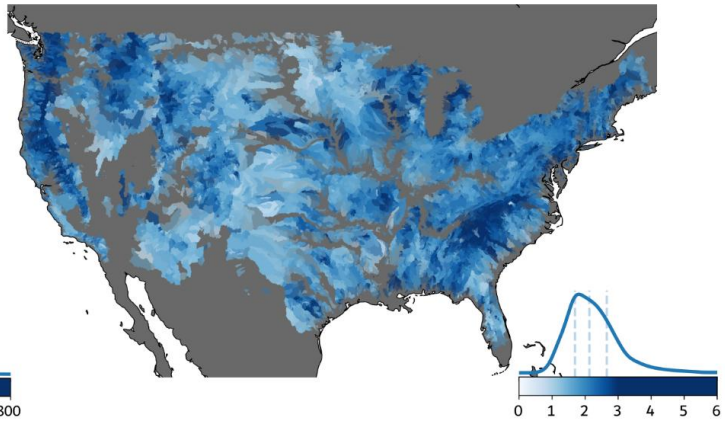
(a.2) Baseflow - RecessionK



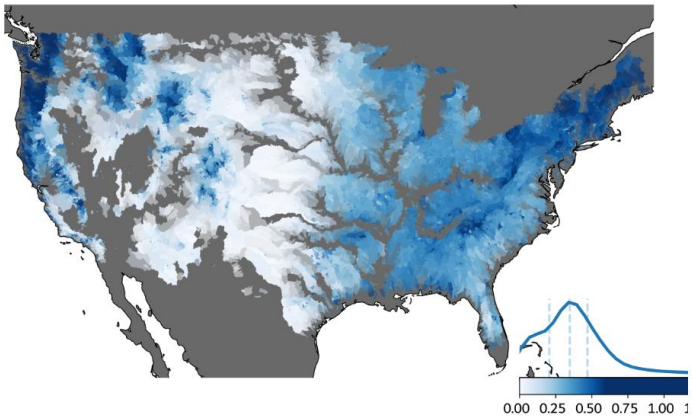
(b.1) Storage capacity and retention - Average Storage



(b.2) Storage capacity and retention - RecessionParameters_b



(c.1) Water balance - TotalRR



(c.2) Water balance - EventRR

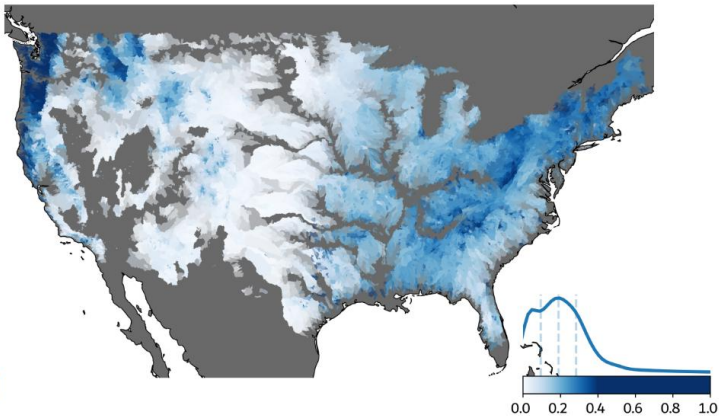


Figure S6. Hydrologic signature values used to generate Figures 2 and 3 (d-f). The chart above the color bar shows the distribution of signature values. Dashed lines indicate the 0.25, 0.5, and 0.75 quantiles.

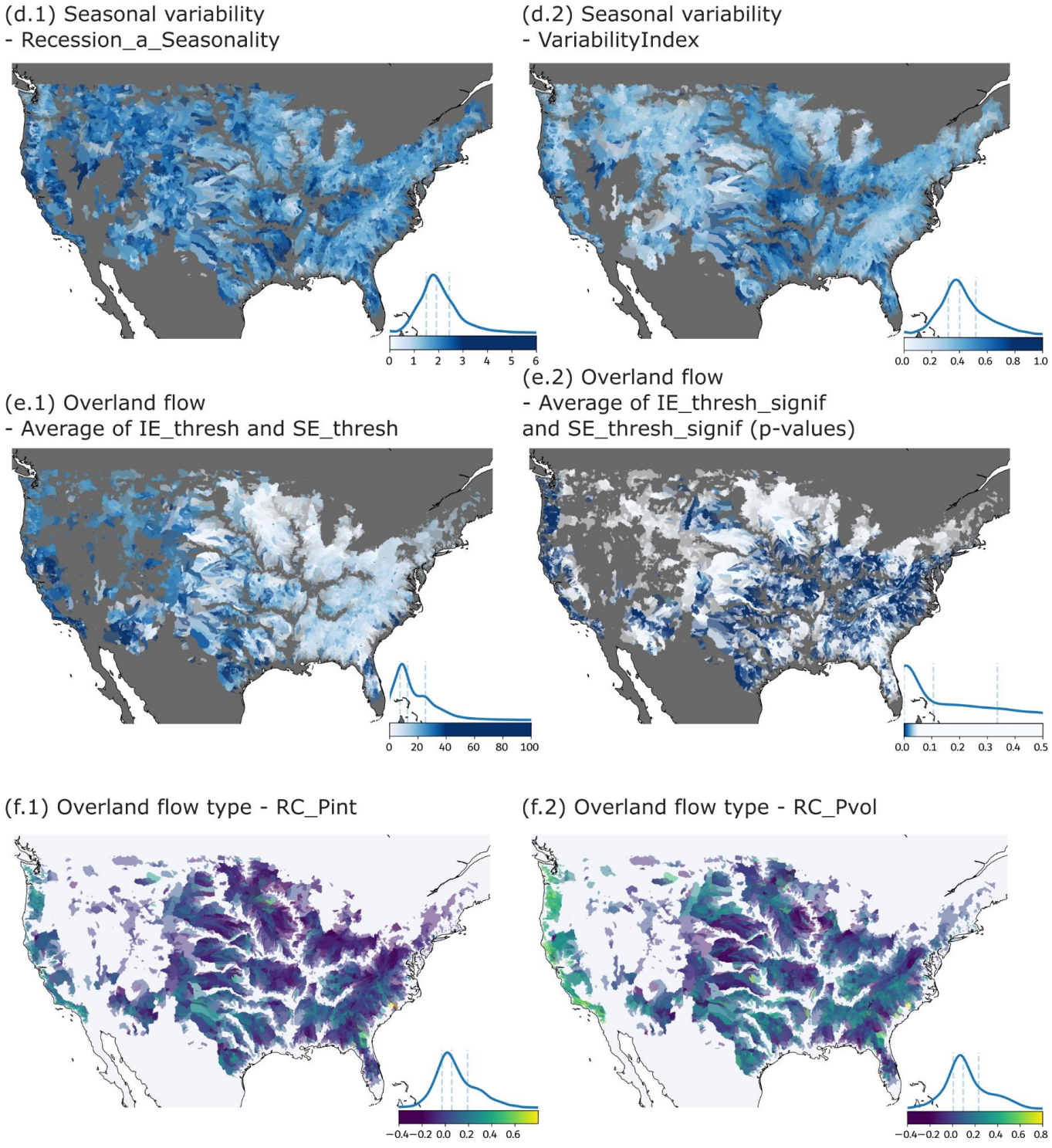


Figure S7: Maps showing climate attributes' values (right column) and Shapley values (left column) for a selected signature.

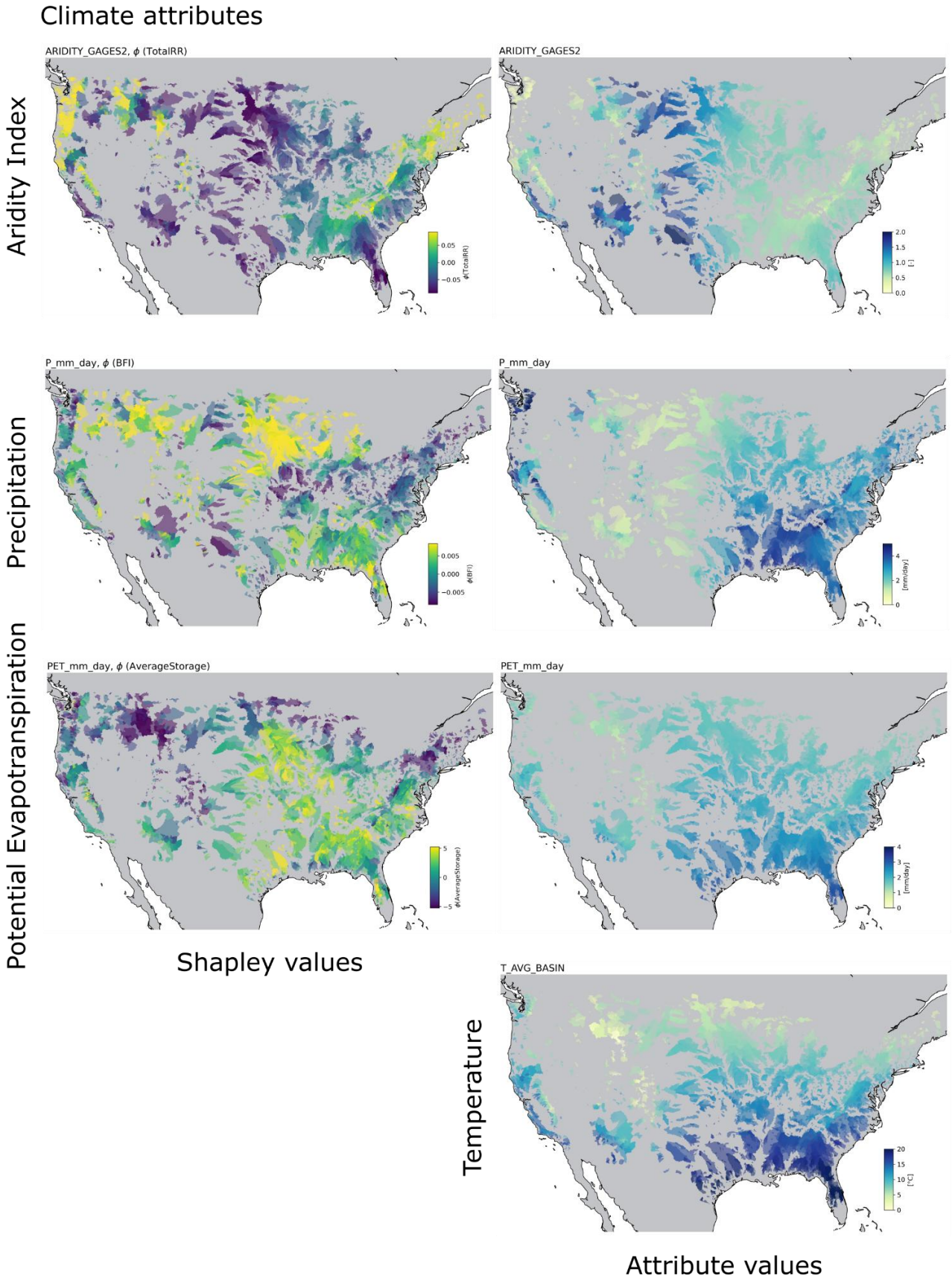


Figure S8: Maps showing climate attributes' values (right column) and Shapley values (left column) for a selected signature. The Figure continues to introduce Soil and Geologic attributes.

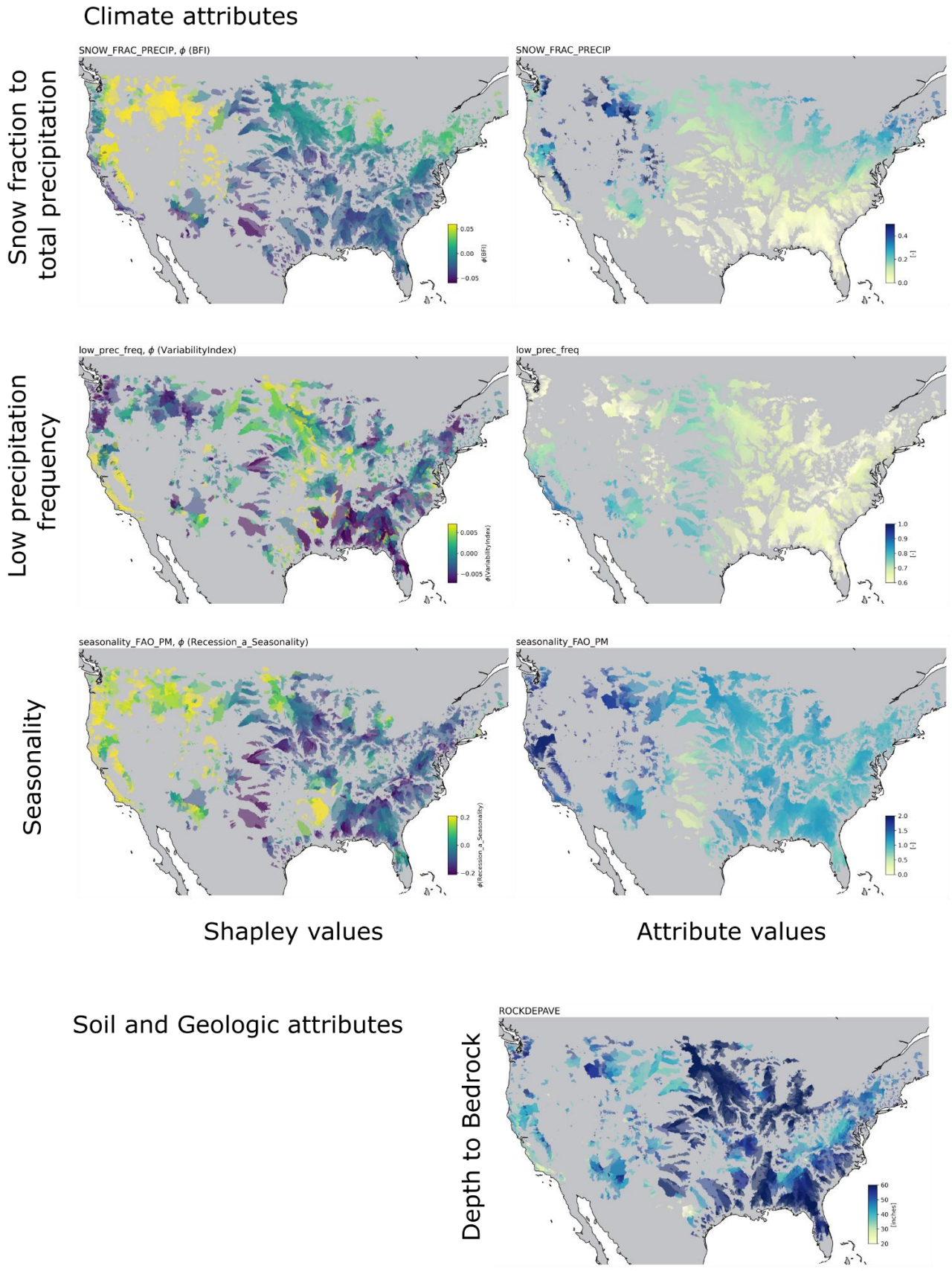


Figure S9: Maps showing soil and geologic attributes' values (right column) and Shapley values (left column) for a selected signature.

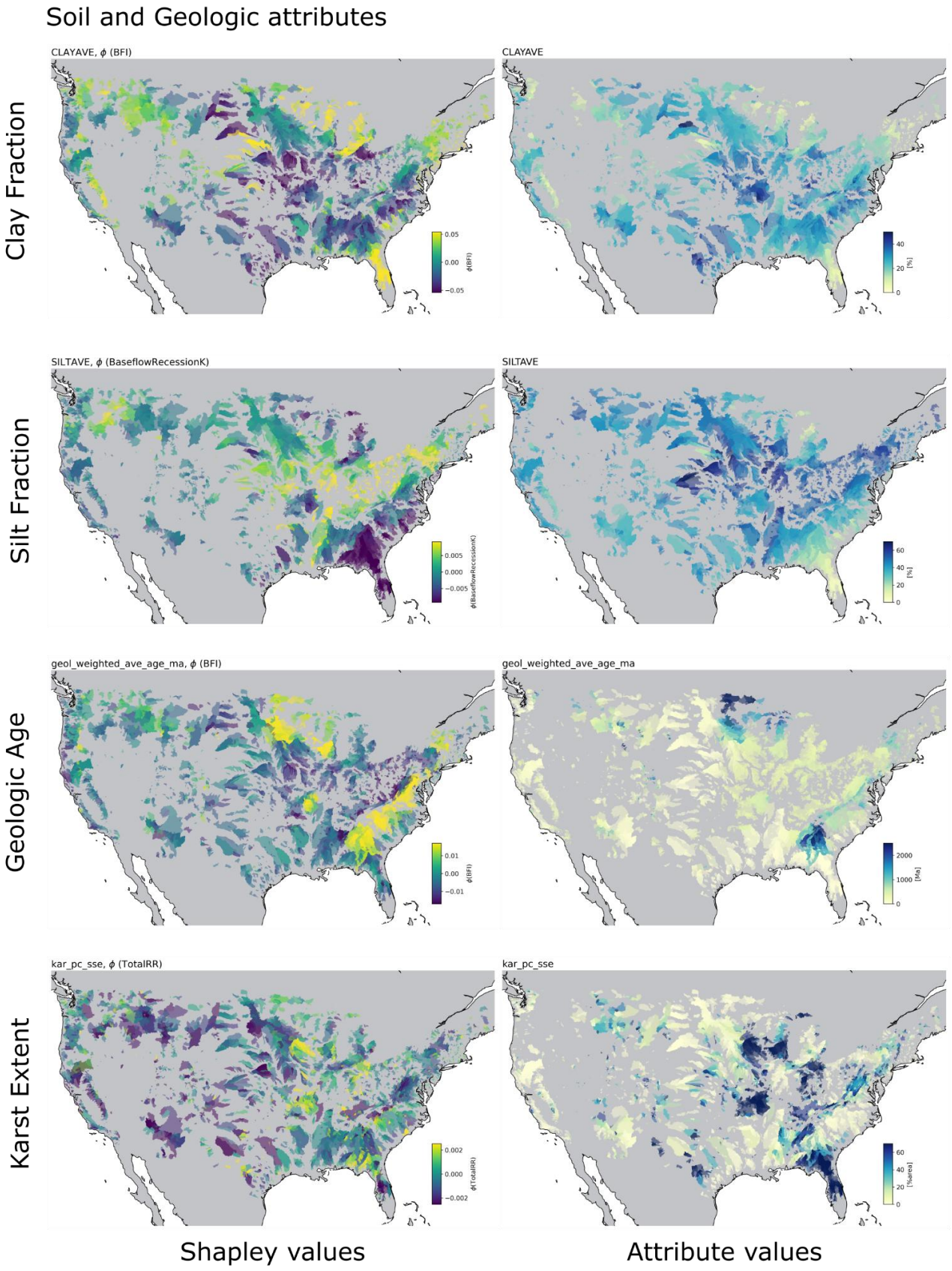


Figure S10: Maps showing land cover attributes' values (right column) and Shapley values (left column) for a selected signature.

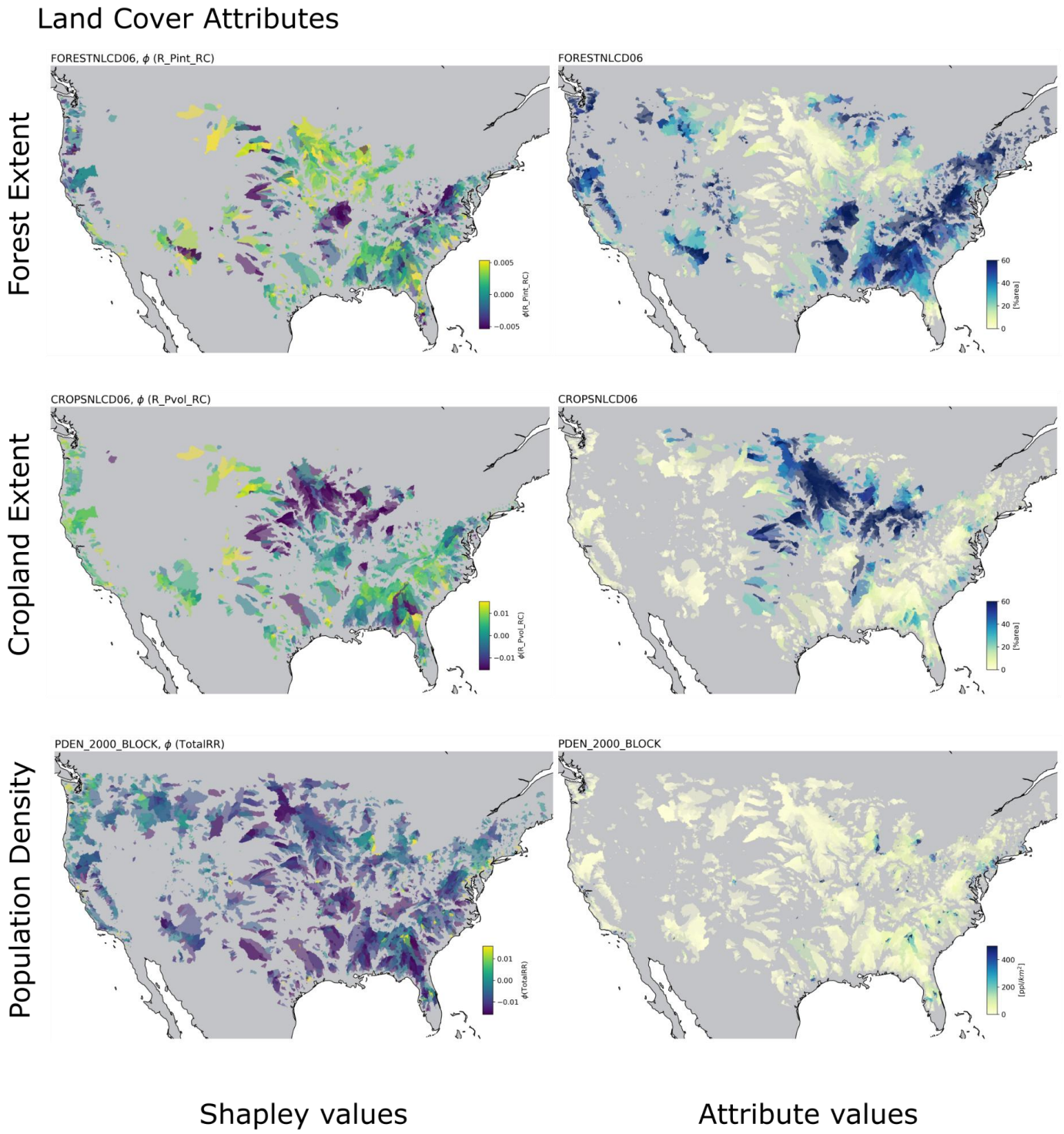


Figure S11: Maps showing topographic attributes' values (right column) and Shapley values (left column) for a selected signature.

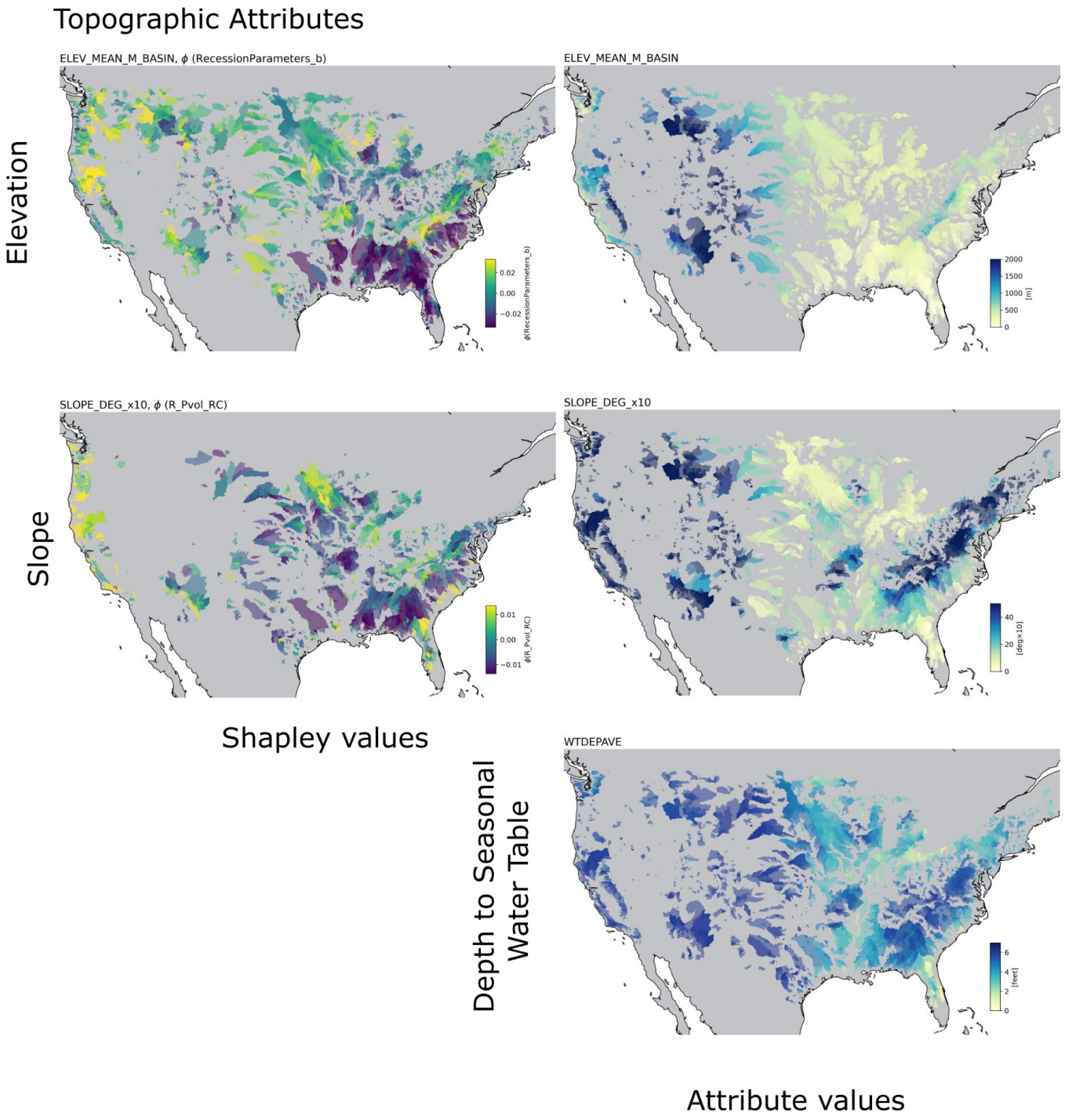


Figure S12: Goodness of fit of random forest models (predicted vs. observed signatures) from regional experiments. The region names correspond to the climate clusters defined in Figure S1.

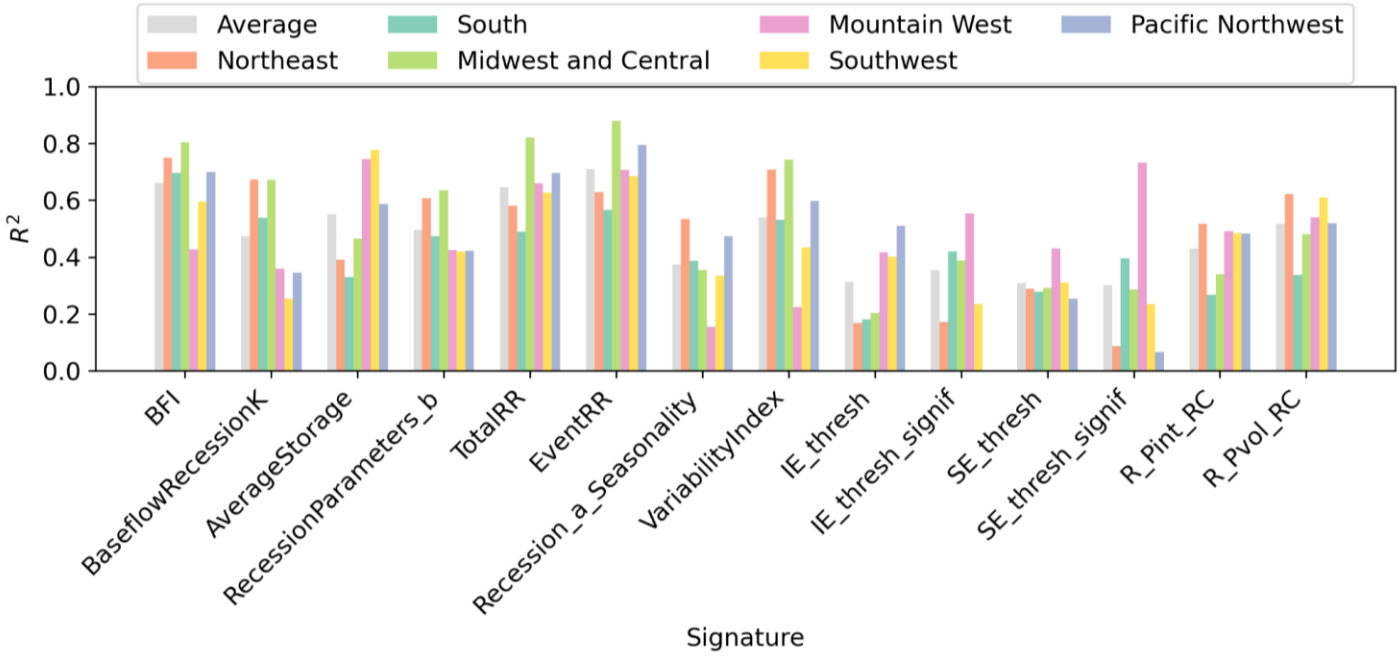
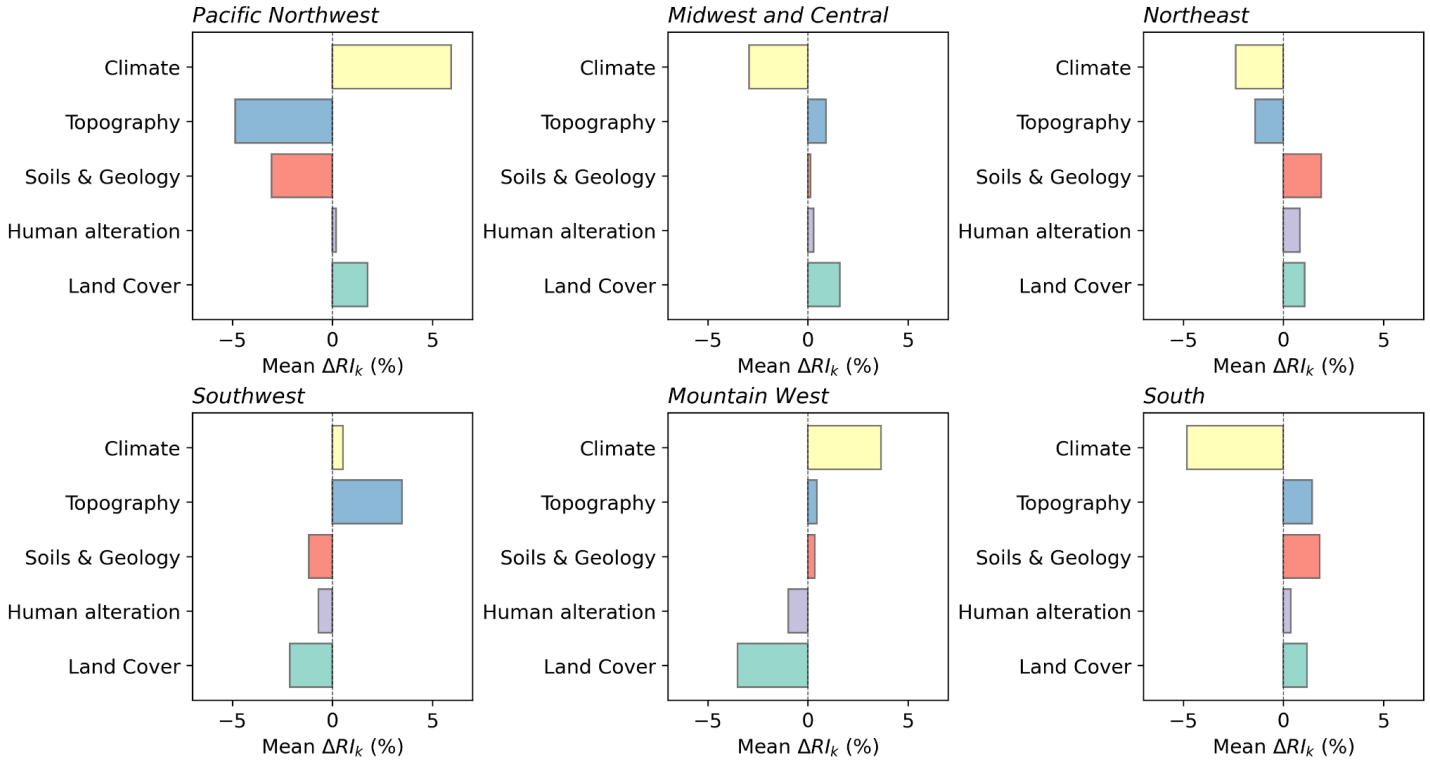


Figure S13: Bar plot showing changes in relative variable importance (IncMSE%) of landscape attribute categories in each of the regional experiments compared to the average of all regional experiments.



First, we calculated the relative importance of category k for a given signature y. To do this, we summed the variable importance of all attributes x within category k ($x \in k$) and normalized it by the total contribution of all attributes ($x \in A$).

$$RI_k^y = 100 \times \sum_{x \in k} I_x^y / \sum_{x \in A} I_x^y$$

This value, RI_x^y, represents the relative contribution of category k to signature y. For example, $RI_{landcover}^{BFI}$ can be interpreted as *"Land cover accounts for $RI_{landcover}^{BFI}$% of the variable importance in predicting the baseflow index signature (BFI)."*

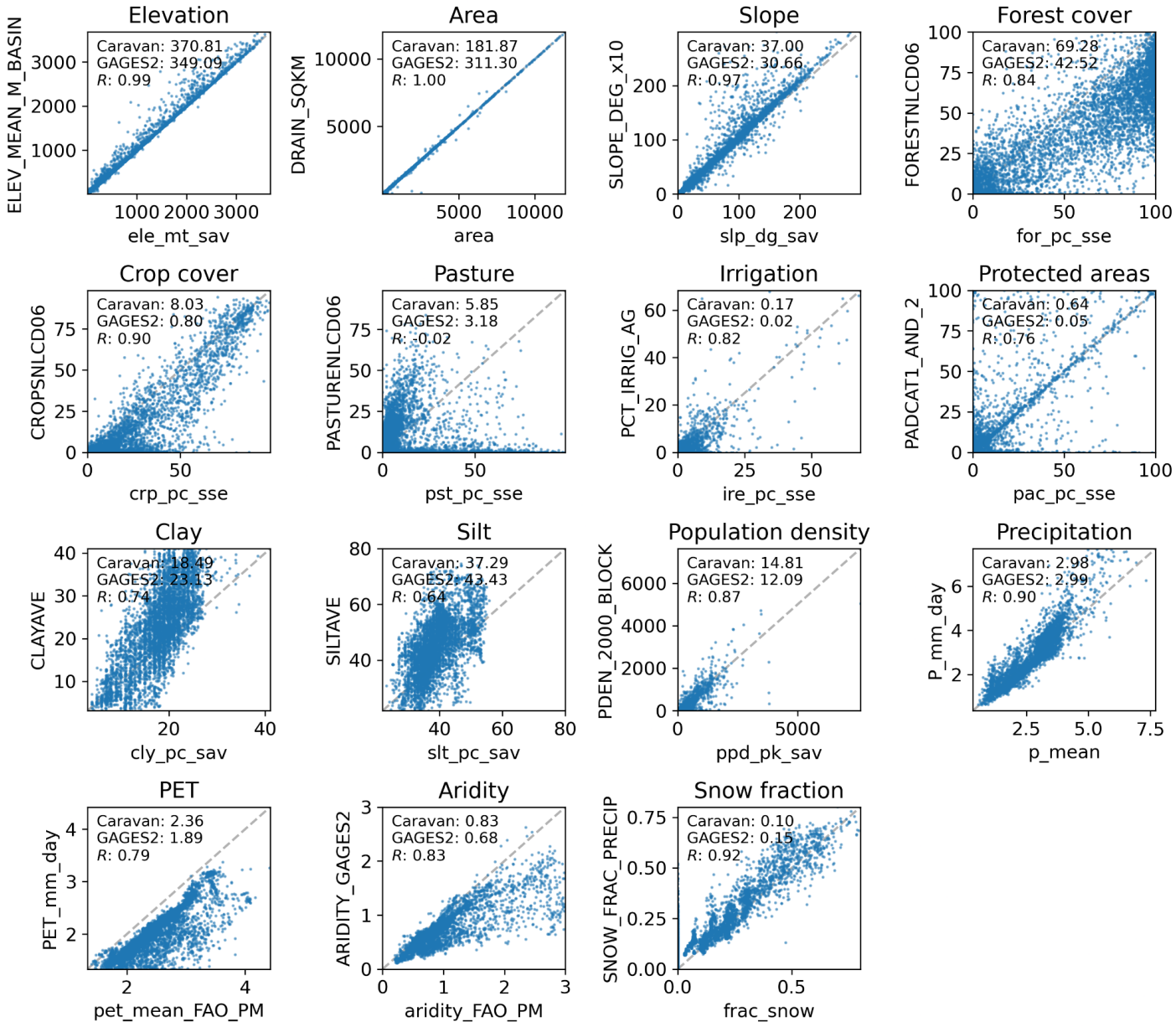
Next, we compared the relative importance from the regional random forest model to that from the average of all regional models:

$$\Delta RI_k^y(r) = RI_k^y(r) - \frac{1}{R}\sum_{r \in R} RI_k^y(r)$$

, where r is a region and R is all six regions. For example, this comparison tells us: *"Land cover is $\Delta RI_{landcover}^{BFI}$% more important in the Midwest regional model than the average for predicting BFI."*

Finally, we computed the mean $\Delta RI_k^y(r)$ across signatures y ($y \in Y$), which defines the x-axis in the plot: $Mean\ \Delta RI_k$ *(%)*.

Figure S14: Comparisons of Caravan (*x*-axis) and GAGES-II equivalent attributes (*y*-axis) after unit conversion (units and attribute descriptions are in Table 2). Median values of each attribute and the correlation coefficient *R* are shown in the legend. The correlation coefficients are $R > 0.7$ for most of the attributes, with the exception of silt content ($R = 0.64$).



References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International journal of climatology*, *33*(1), 121-131.

Arsenault, R., Brissette, F., Martel, J. L., Troin, M., Lévesque, G., Davidson-Chaput, J., ... & Poulin, A. (2020). A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Scientific Data*, *7*(1), 243.

Broxton, P. D., Van Leeuwen, W. J., & Biederman, J. A. (2019). Improving snow water equivalent maps with machine learning of snow survey and lidar measurements. *Water Resources Research*, *55*(5), 3739-3757.

Budyko, M. I., & Miller, D. H. (1974). *Climate and Life*. Bondi, NSW, Australia: Geniza. https://doi.org/10.1016/s0074-6142(09)x6001-8

Falcone, J.: GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow, https://doi.org/10.5066/P96CPHOT, 2011.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., and Meador, M. R.: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, Ecology, 91, 621–621, https://doi.org/10.1890/09-0889.1, 2010.

Hammond, J. C., Sexstone, G. A., Putman, A. L., Barnhart, T. B., Rey, D. M., Driscoll, J. M., et al. (2023). High resolution SnowModel simulations reveal future elevation-dependent snow loss and earlier, flashier surface water input for the upper Colorado river basin. *Earth's Future*, *11*(2). https://doi.org/10.1029/2022ef003092

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*(7), 5088–5109. https://doi.org/10.1029/2018wr022913

Meira Neto, A. A., Roy, T., de Oliveira, P. T. S., & Troch, P. A. (2020). An aridity index-based formulation of streamflow components. *Water Resources Research*, *56*(9), e2020WR027123. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027123?casa_token=mNQrAGZMPOUAAAAA:TiiRtOCg33enyfxAmyEGeOTytcxfqsca7BEXQcPgsHGWL3fYYneuJNyXQoom7PISbjo26d-O1DE4_cQ

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.