Supplement of Hydrol. Earth Syst. Sci., 28, 479–503, 2024
https://doi.org/10.5194/hess-28-479-2024-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

On the need for physical constraints in deep learning rainfall–runoff projections under climate change: a sensitivity analysis to warming and shifts in potential evapotranspiration

Sungwook Wi and Scott Steinschneider

Correspondence to: Sungwook Wi (sw2275@cornell.edu)

Text S1: Adjustments to Static Attributes

In the primary article, we describe two sets of scenarios for the deep learning models used in this work: 1) one in which changes are only made to the dynamic inputs features of each model, and 2) one with changes to both dynamic features and to static features that depend on those dynamic features. Here we describe in more detail the adjustments made to the static features for each site, which include: pet_mean, aridity, t_mean, frac_snow (see Table S1 below for the definition of these features). Importantly, these are the static features that are dependent on temperature and PET, the two dynamic inputs adjusted in our analysis.

To adjust t_mean, we use the full time series of daily maximum and minimum temperature (on which t_mean was originally based), and shift those time series upward by 4°C. Using those adjusted series, we calculate daily average temperature as the mean of maximum and minimum temperature on each day, and then calculate the long-term mean of daily average temperature to develop an updated estimate of t_mean.

To adjust frac_snow, we first calculate the adjusted time series of daily average temperature based on the time series of daily maximum and minimum temperature shifted upward by 4°C. Then, we count all days in the record when precipitation occurs and this adjusted time series of daily average temperature is below 0°C, and divide this number by the total number of days of non-zero precipitation in the record. The resulting value is the updated value for frac_snow.

We develop two versions of adjusted pet_mean, one based on Hamon PET and the other for Priestley-Taylor PET. The adjusted Hamon PET is based entirely on the series of daily maximum and minimum temperature shifted by 4°C. We use Eqs. 7-8 in the main article to calculate daily Hamon PET under warming. We then take the long-term mean of this time series to develop an updated estimate of pet_mean. Similarly, for Priestley-Taylor PET, we couple the warmed temperature time series with the unadjusted time series of net shortwave radiation, and then use the approach in Eq. 9 in the main article to calculate a daily time series of Priestley-Taylor PET. We again take the long-term mean of this time series to develop an updated estimate of pet_mean.

Finally, we develop two versions of adjusted aridity, one based on Hamon PET and the other for Priestley-Taylor PET. In both cases, we calculate adjusted aridity as the ratio of the updated values for pet_mean under warming and the unadjusted value for long-term mean precipitation (another static input to the models).

Table S1. Static watershed attributes that are adjusted in a subset of scenarios used in this analysis.

Attribute	Description
pet_mean	Mean daily potential evapotranspiration
aridity	Ratio of mean PET to mean precipitation
t_mean	Mean of daily maximum and daily minimum temperature
frac_snow	Fraction of precipitation falling on days with mean daily temperatures below 0°C

Additional Supporting Tables

Table S2. Range of values considered in the grid search during hyper-parameter tuning.

Hyper-parameter	**Values Tested**
Number of Hidden Layer Nodes	64, 96, 128, 256
Mini-Batch Size	64, 128, 256, 512
Learning Rate	0.0001, 0.0005, 0.001, 0.005
Number of Epochs	30, 50
Dropout Rate*	0, 0.2, 0.4

Table S3. Additional details for gauges highlighted in Figures 5 and 6 of main article.

Gauge ID	**Country**	**Site Name**	**Drainage Area (km^2)**
02ED032	Canada	Willow Creek near Minesing	231
02GG013	Canada	Black Creek near Bradshaw	213
02HJ003	Canada	Ouse River near Westwook	283
04126740	United States	Platte River at Honor, MI	324
04220045	United States	Oak Orchard Creek near Shelby NY	378
04168400	United States	Lower River Rouge at Dearborn, MI	236

Additional Supporting Figures

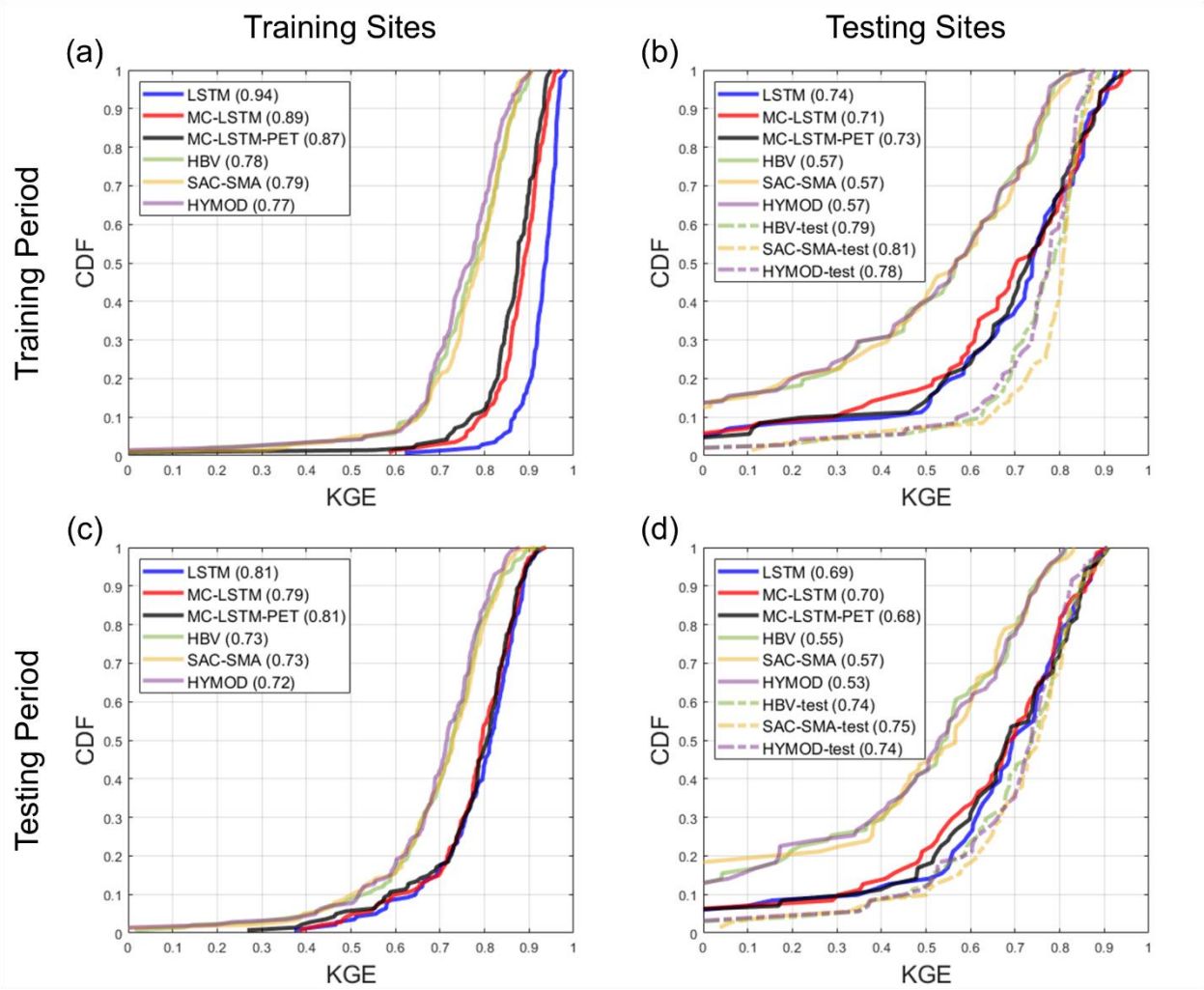

Figure S1. The distribution of Kling-Gupta efficiency (KGE) for streamflow estimates across sites from each model at the (a) the 141 training sites and (b) 71 testing sites for the training period. Similar results for the testing period are shown in panels (c) and (d), respectively. For the process models fit to the testing sites (denoted "-test"), no performance results are available at the training sites. All models are trained using Hamon PET.

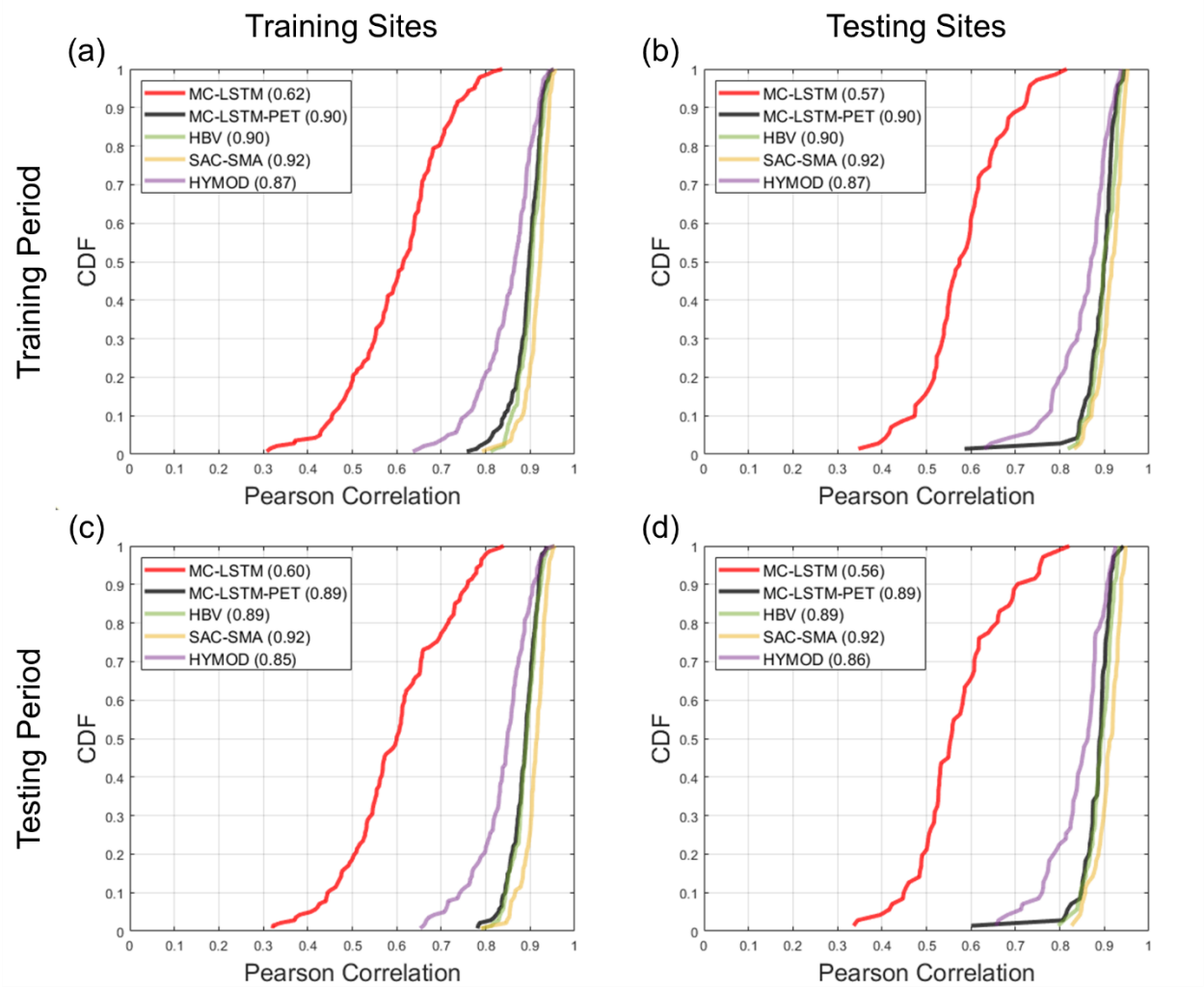

Figure S2. The correlation between model estimated and GLEAM AET from each model at the (a) the 141 training sites and (b) 71 testing sites for the training period. Similar results for the testing period are shown in panels (c) and (d), respectively. The LSTM is not included in this comparison. All models are trained using Priestley-Taylor PET.

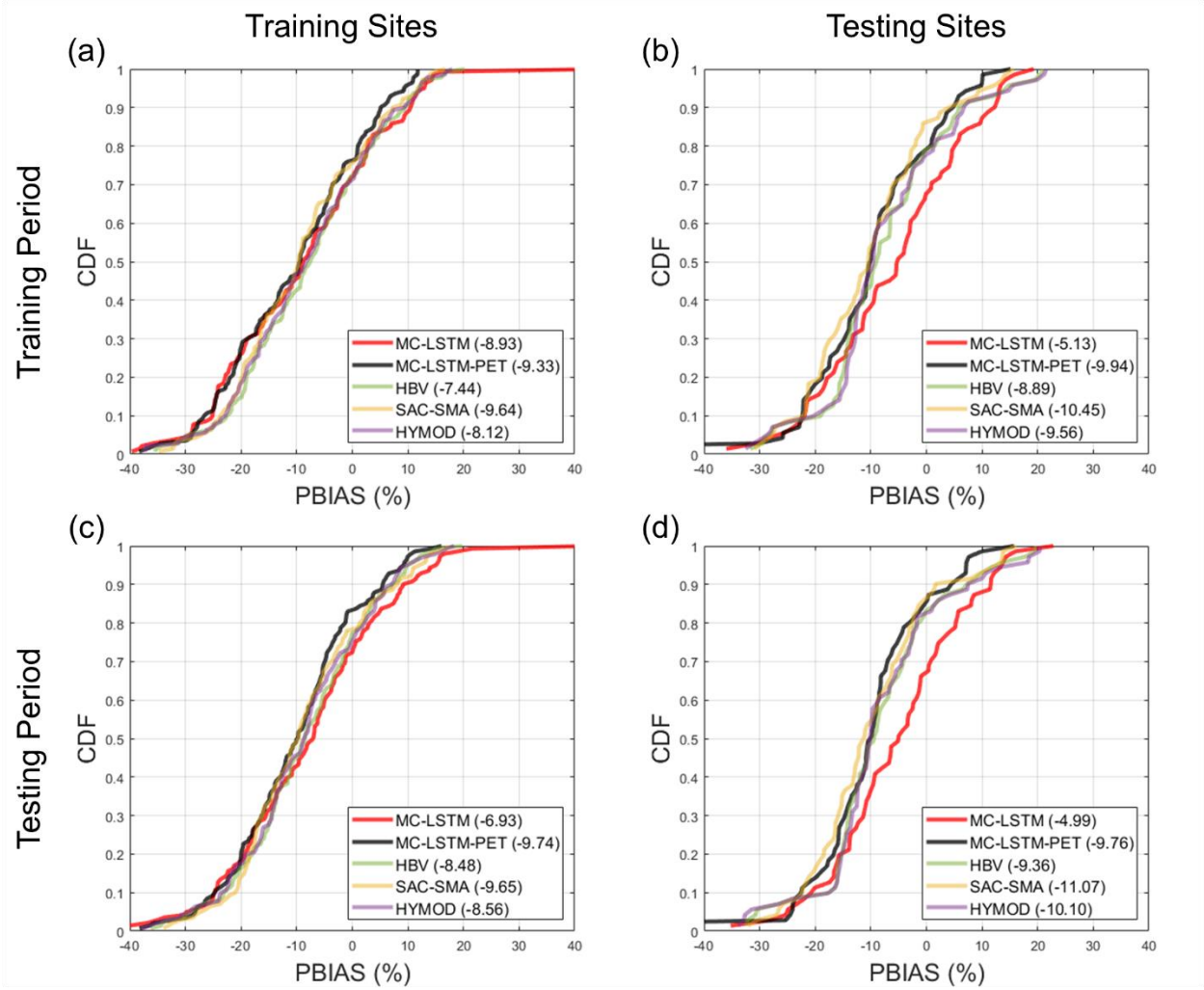

Figure S3. The PBIAS between model estimated and GLEAM AET from each model at the (a) the 141 training sites and (b) 71 testing sites for the training period. Similar results for the testing period are shown in panels (c) and (d), respectively. The LSTM is not included in this comparison. All models are trained using Priestley-Taylor PET.

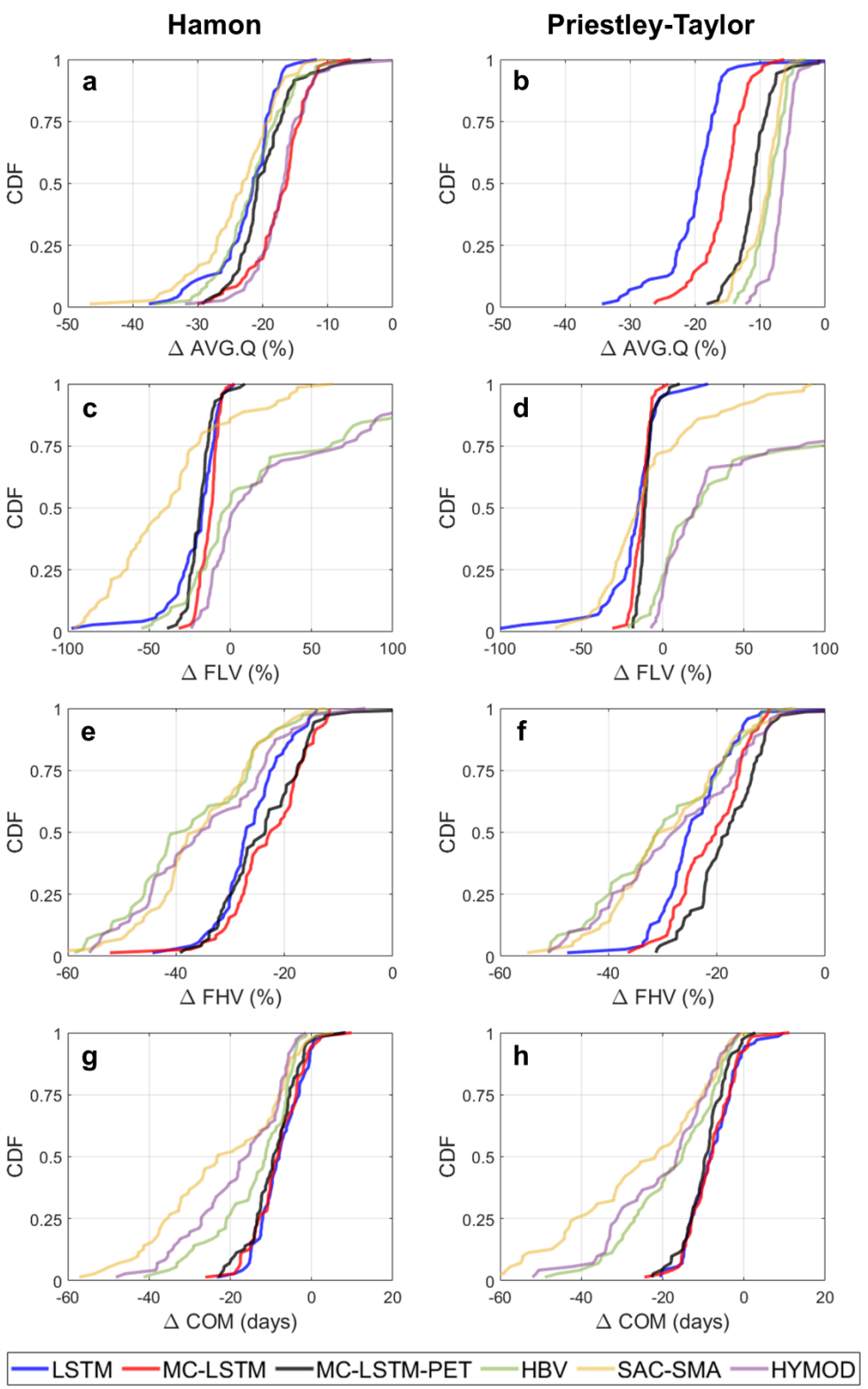

Figure S4. The distribution of change in (a,b) AVG.Q, (c,d) FLV, (e,f) FHV, and (g,h) COM across the 71 testing sites and all models under a scenario of 4°C warming using (a,c,e,g) Hamon PET and (b,d,f,h) Priestley-Taylor PET. For the DL models, changes were only made to the dynamic inputs (i.e., no changes to static inputs).

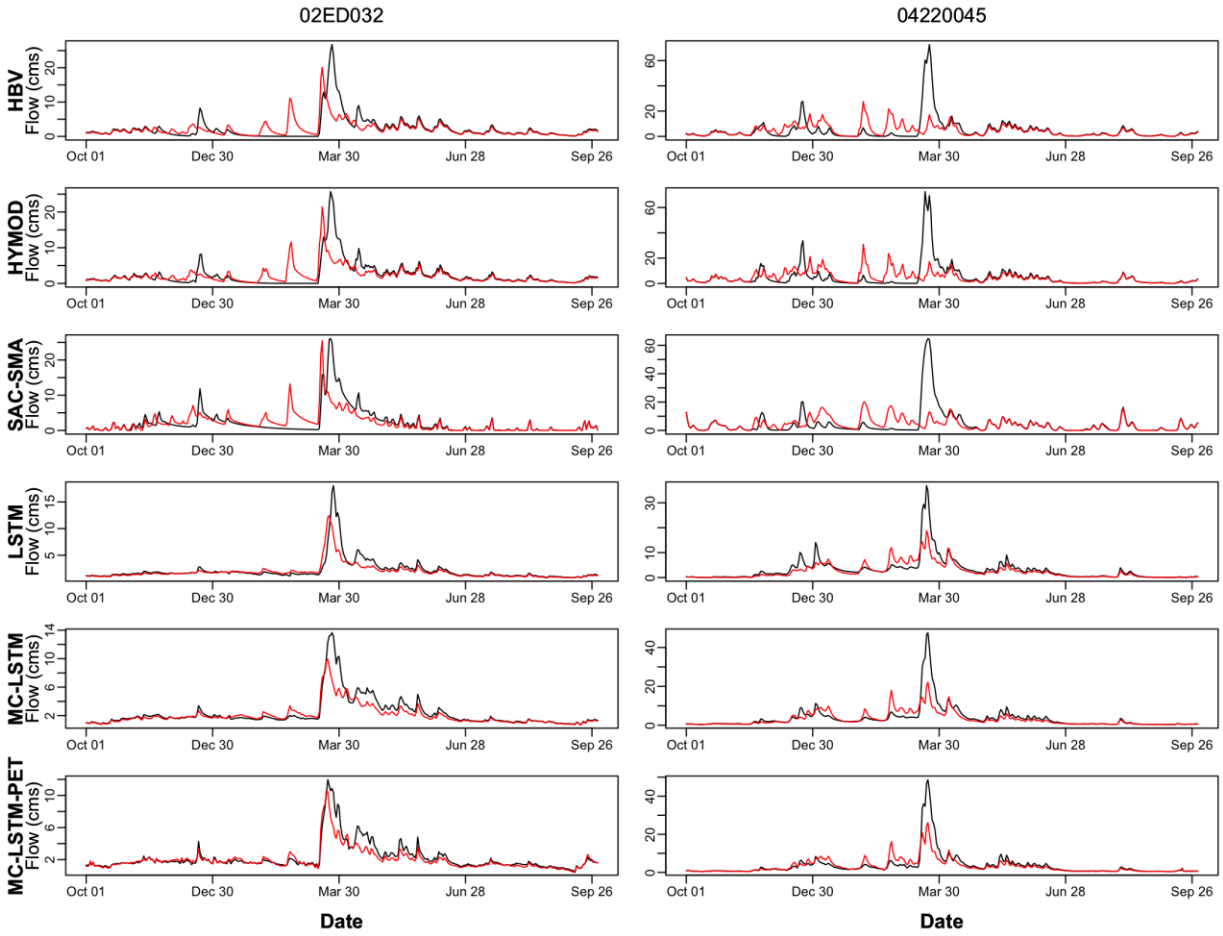

Figure S5. Daily streamflow hydrograph for one water year (2002 October- 2003 September) across the three different process-based models (HBV, HYMOD, SAC-SMA) and deep-learning models (LSTM, MC-LSTM, MC-LSTM-PET) under 0°C warming (black) and 4°C warming (red). Results are shown for two sites (highlighted in Figure 1 of the main article), and are constructed with models using Priestley-Taylor PET.

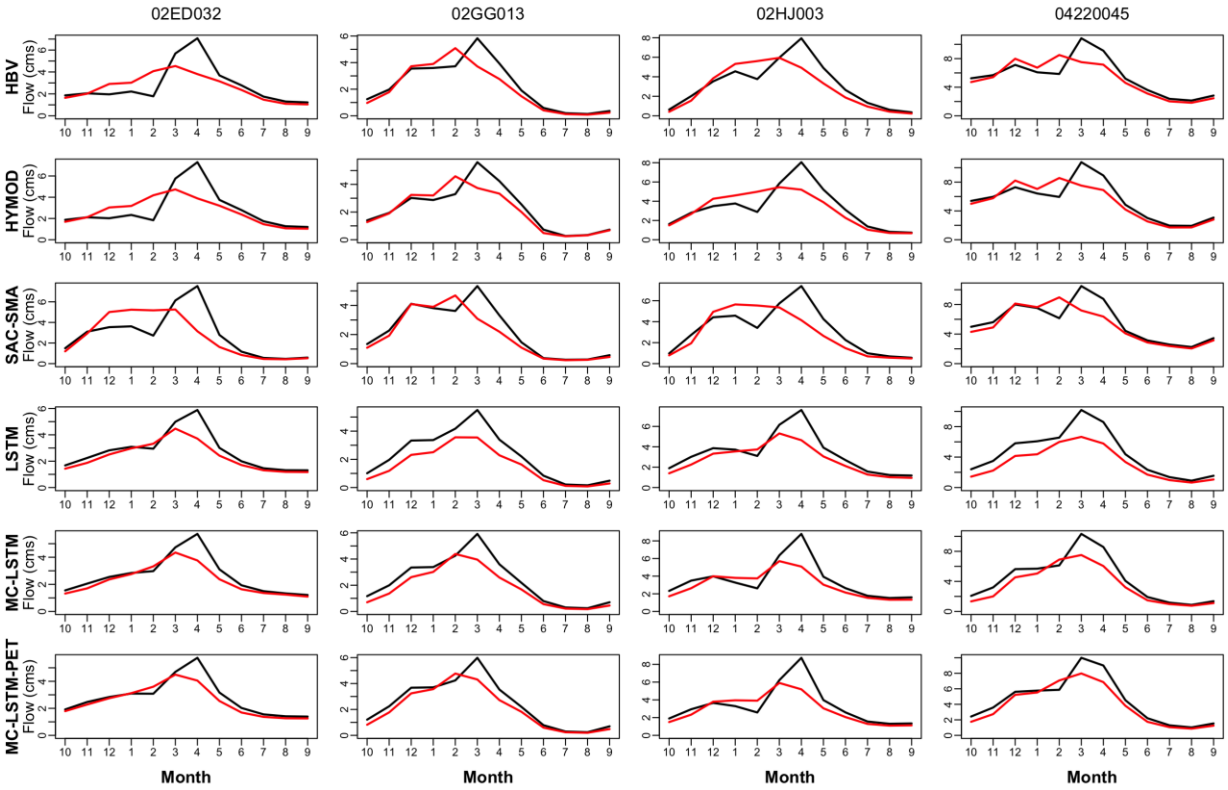

Figure S6. Mean monthly streamflow averaged across the entire record, shown for the three different process-based models (HBV, HYMOD, SAC-SMA) and deep-learning models (LSTM, MC-LSTM, MC-LSTM-PET) under 0°C warming (black) and 4°C warming (red). Results are shown on a water year basis (October-September) for four sites (highlighted in Figure 1 of the main article), and are constructed with models using Priestley-Taylor PET.

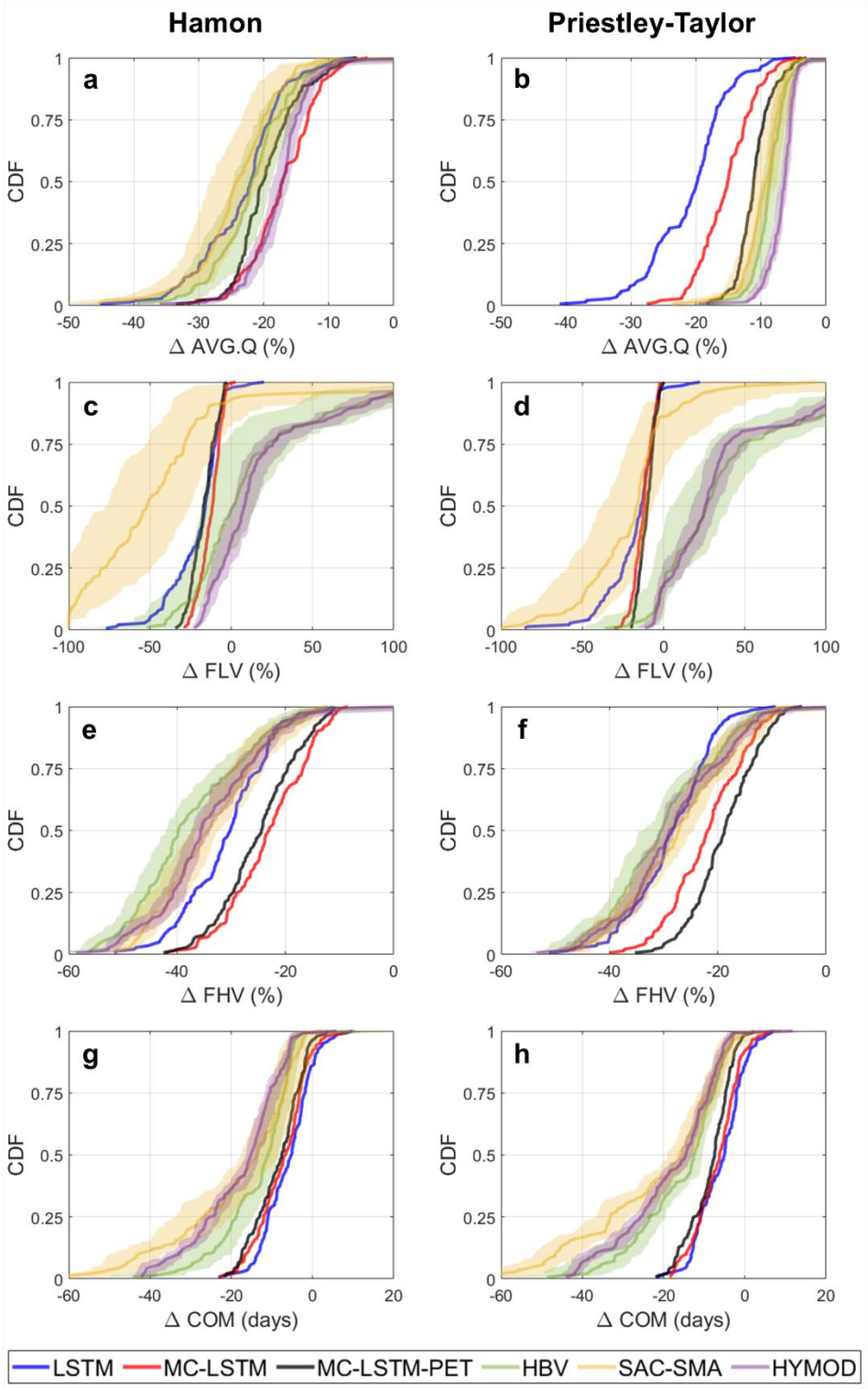

Figure S7. The distribution of change in (a,b) long term mean daily flow (AVG.Q), (c,d) low flows (FLV), (e,f) high flows (FHV), and (g,h) seasonal streamflow timing (COM) across the 141 training sites and all models under a scenario of 4°C warming using (a,c,e,g) Hamon PET and (b,d,f,h) Priestley-Taylor PET. For the deep learning models, changes were only made to the dynamic inputs (i.e., no changes to static inputs). For the process models, the uncertainty in the change in each streamflow attribute across 10 different training trails is shown as translucent shading.

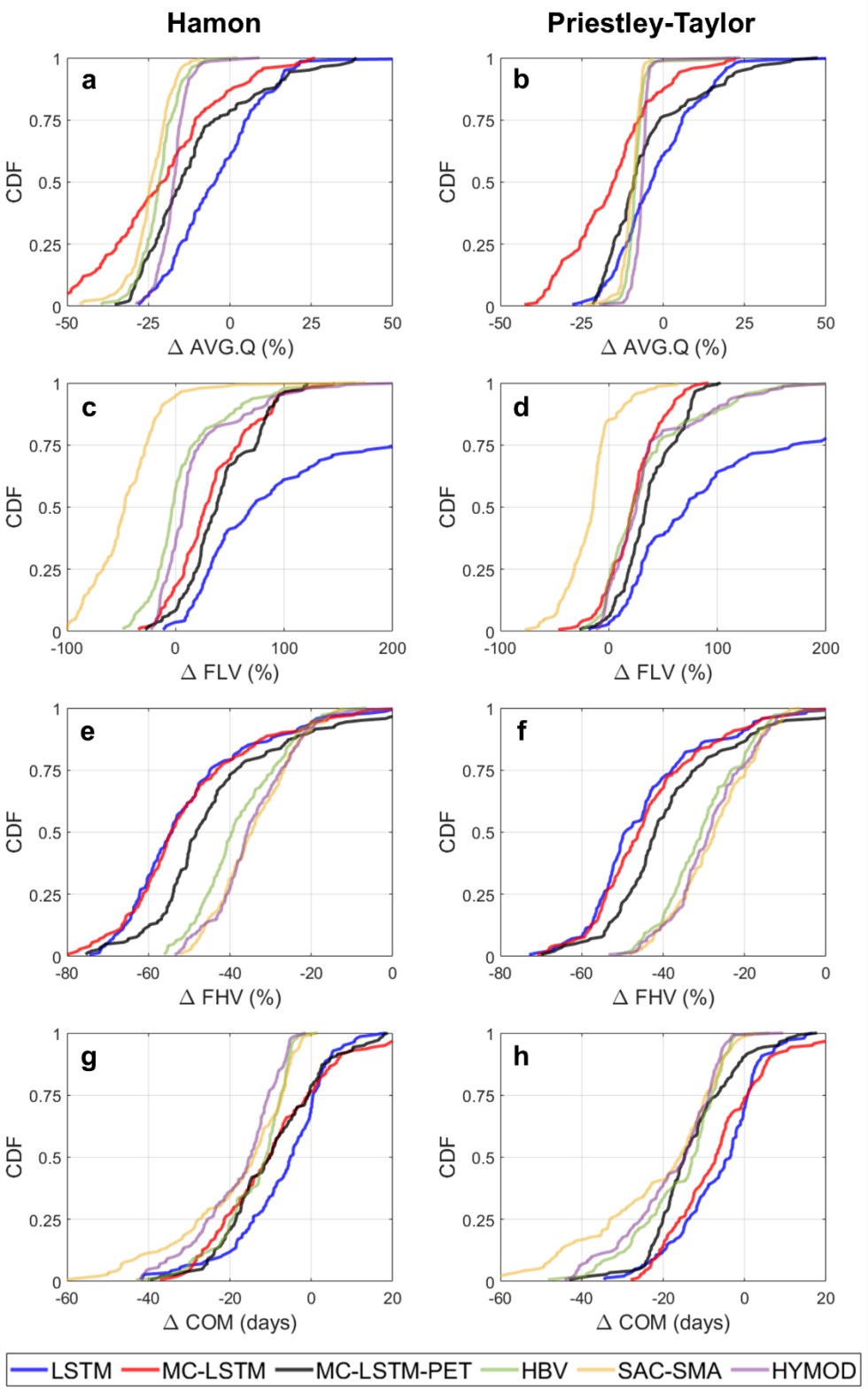

Figure S8. The distribution of change in (a,b) AVG.Q, (c,d) FLV, (e,f) FHV, and (g,h) COM across the 141 training sites and all models under a scenario of 4°C warming using (a,c,e,g) Hamon PET and (b,d,f,h) Priestley-Taylor PET. For the DL models, changes were made to both the dynamic and static inputs.

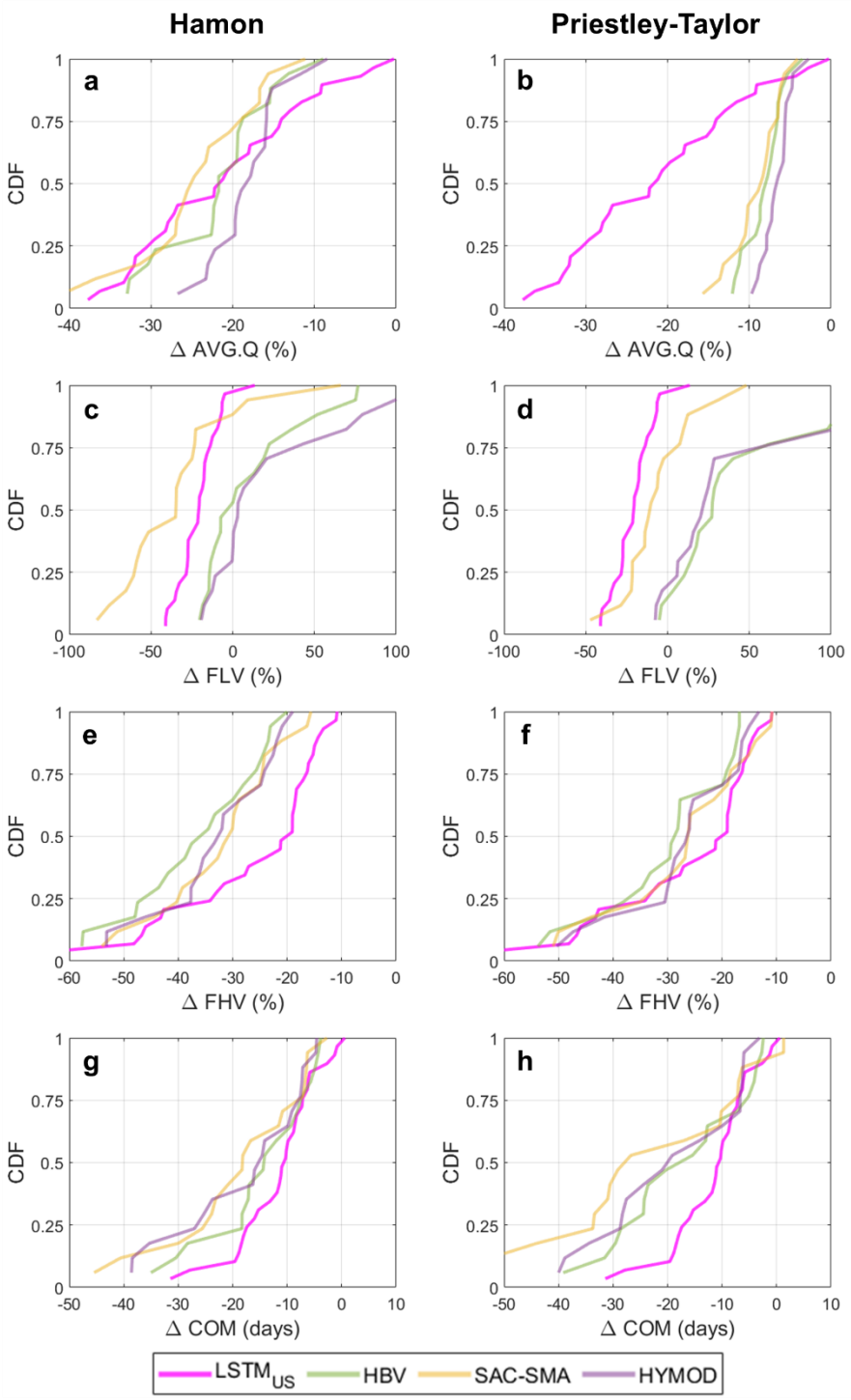

Figure S9. The distribution of change in (a,b) AVG.Q, (c,d) FLV, (e,f) FHV, and (g,h) COM across 29 CAMELS sites within the Great Lakes basin under the National LSTM, as well as for 17 of those 29 sites from the Great Lakes process models, under a scenario of 4°C warming. For the process models only, results differ when using (a,c,e,f) Hamon PET and (b,d,f,h) Priestley-Taylor PET. For the National LSTM, changes were made to both the dynamic and static inputs.