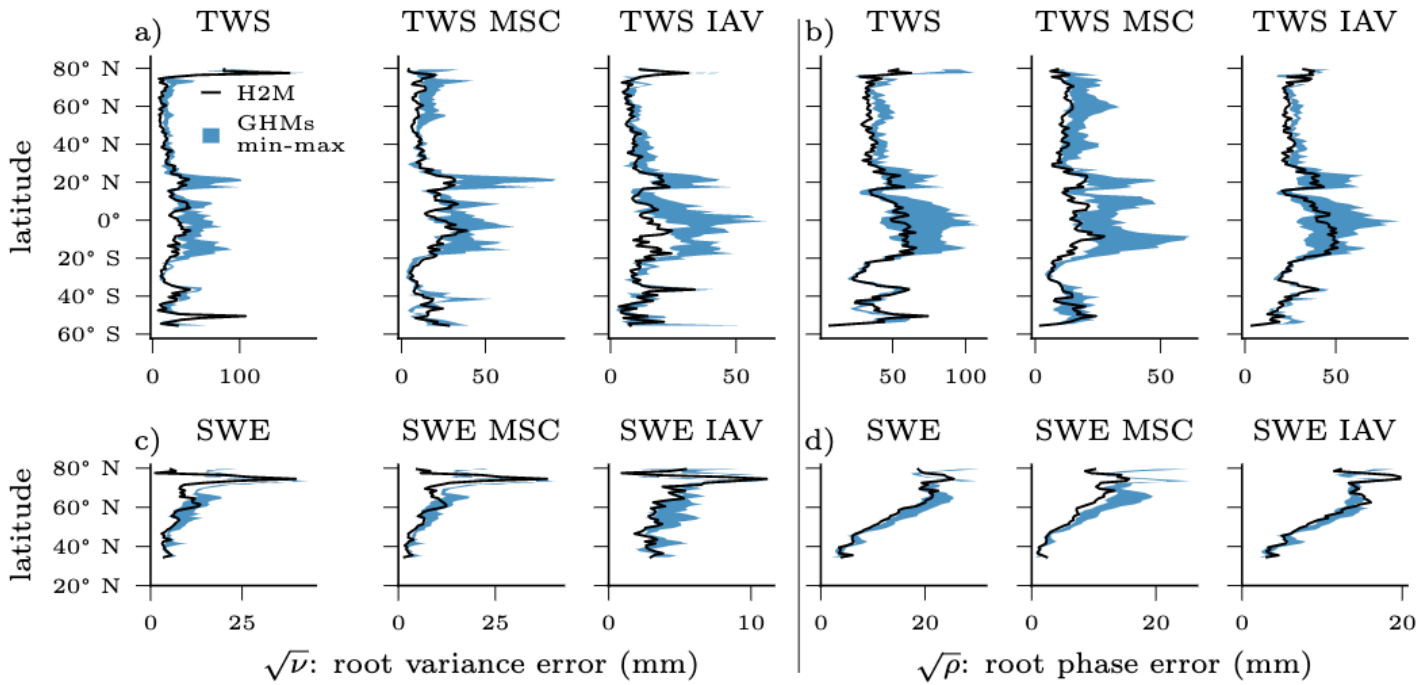

Figure 4. Comparison of the root phase and variance error of the hybrid hydrological model (H2M) to the process-based global hydrological models (GHMs) for terrestrial water storage (TWS) and the snow water equivalent (SWE) time-series, their seasonality (MSC) and interannual variability (IAV). Plot a) shows the TWS variance error, b) the TWS phase error, c) the SWE variance error, and d) the SWE phase error. The black line represents the mean latitudinal error (average across longitudes) of the H2M and the shaded area is the minimum to maximum error of the GHMs. The metrics are calculated for the test period from 2003 to 2012. Note that x-scale differs between plots.

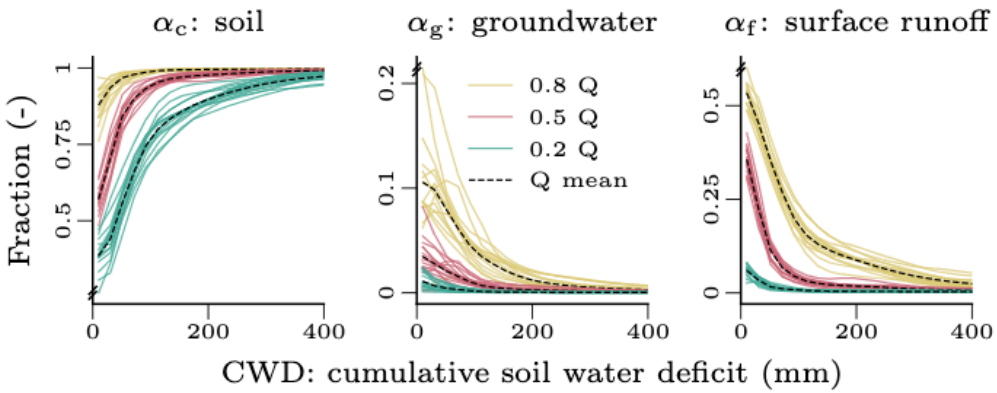

Figure 7. Relationship between the water input partitioning fractions for soil (α_c), groundwater (α_g), and fast runoff (α_f) and the cumulative soil water deficit (CWD) as learned by the neural network. The colored lines represent the 0.2, 0.5, and 0.8 quantiles of the spatio-temporal distribution for different cross-validation runs to show the robustness of the simulations. The dashed, dark lines are the average across the runs per quantile. The plots are based on global daily cell-timesteps from 2009 to 2014. Note that the y-scale differs between plots.

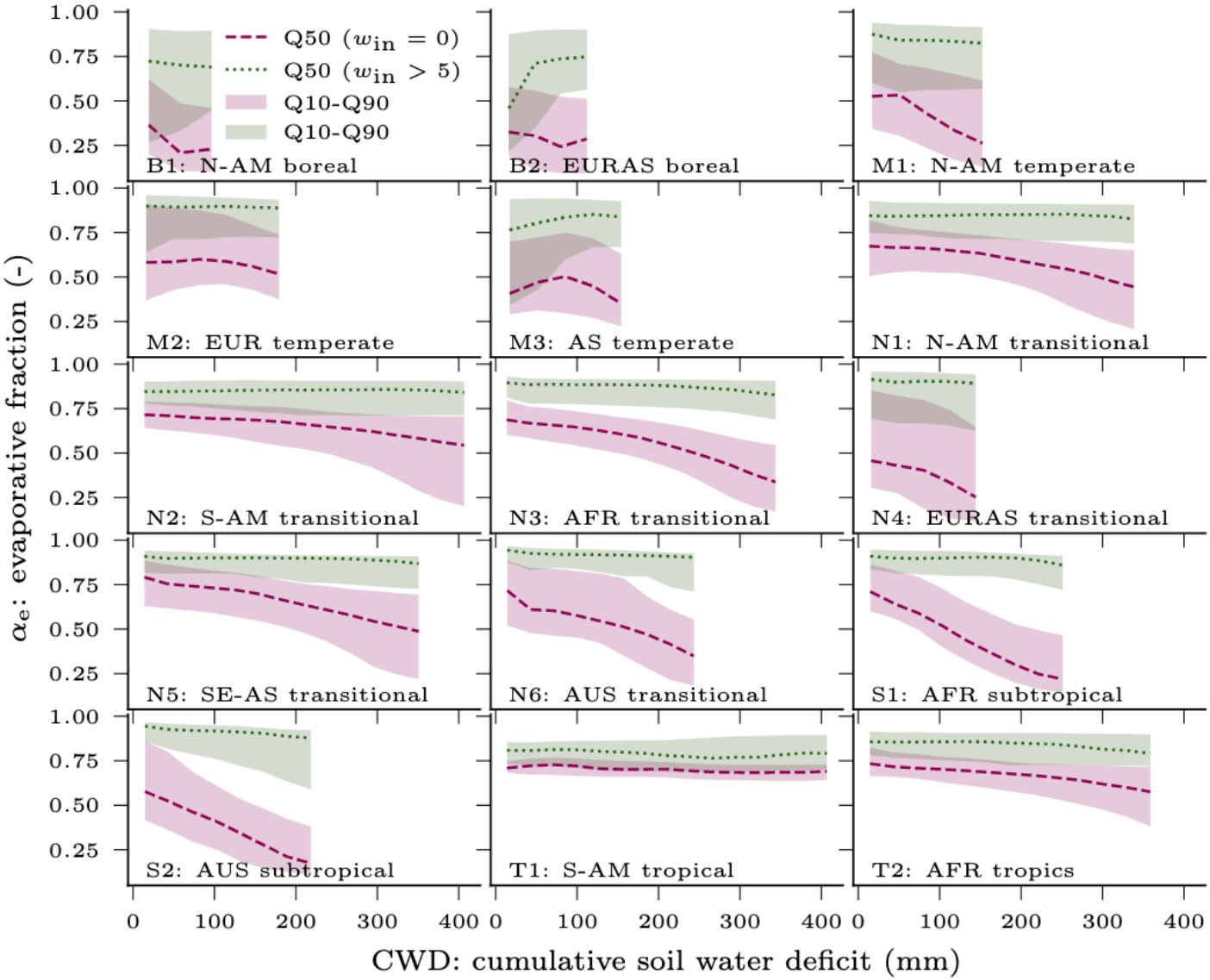

Figure 8. Relationship between evaporative fraction (α_e) and cumulative soil water deficit (CWD) for different hydroclimatic regions. The median and 0.1–0.9 quantile range is shown for conditions without water input ($w_{in} = 0$ mm), i.e., no precipitation or snowmelt, and with high water input ($w_{in} > 5$ mm). Note that the CWD minimum was subtracted per grid cell. To exclude cells with a low CWD variability, only the cells in the top 60 percent maximum CWD were used.

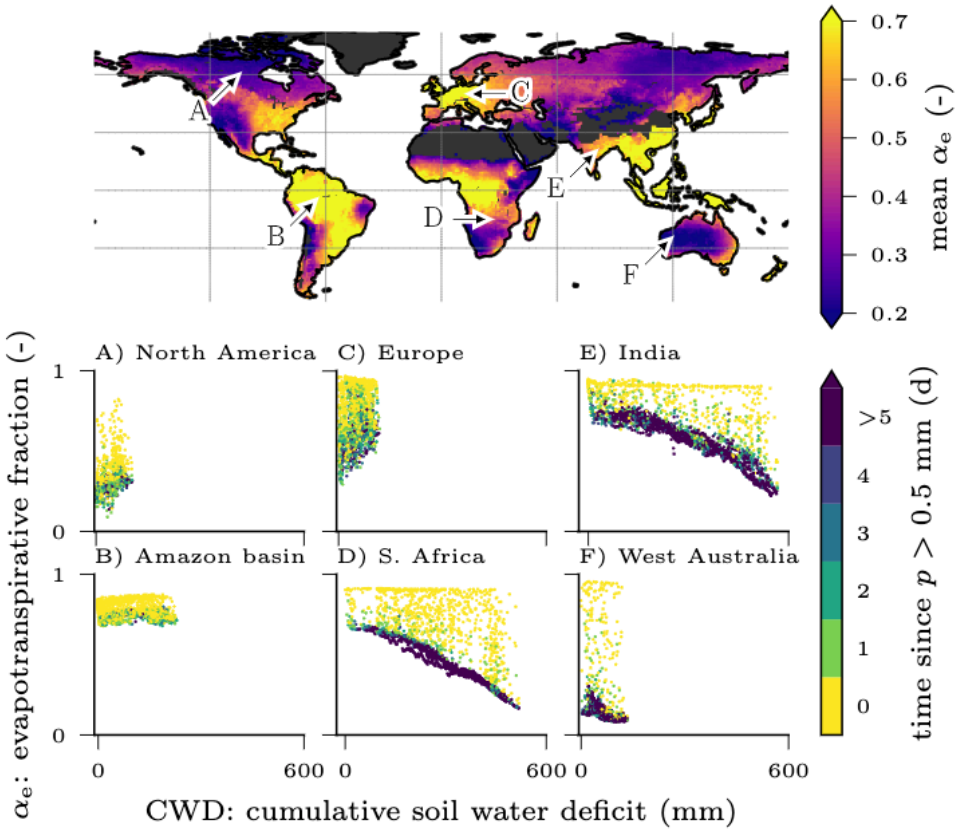

Figure 9. The map shows the mean evaporative fraction (α_e) and scatterplots display the relationship between (α_e) and the cumulative water deficit (CWD), colored by days since last precipitation ($p > 0.5$ mm). The plots are based on global daily cell-timesteps, filtered for positive air temperatures and net radiation, from 2009 to 2014.

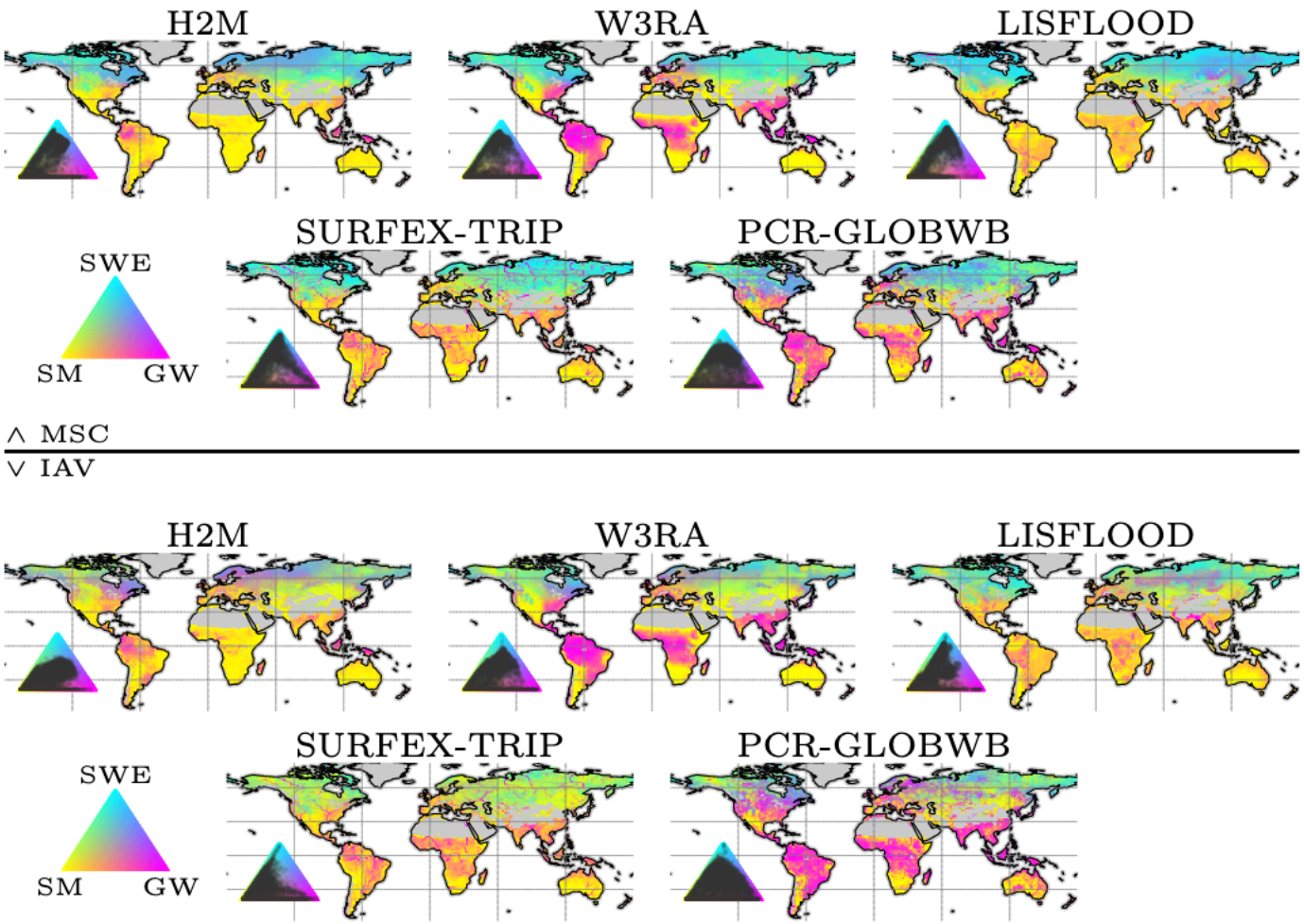

Figure 10. Terrestrial water storage (TWS) variation partitioning into cumulative water deficit (CWD), groundwater (GW), and snow water equivalent (SWE) variation based on the validation period for the hybrid hydrological model (H2M) and a set of process-based global hydrological models (GHMs). The top panels show the partitioning of the mean seasonal cycle (MSC), the bottom the interannual variability. The map colors correspond to the mixture of the contributions of the two variables, the inset ternary plots reflect the density of the map points projected onto the components. The contribution is calculated as the sum of the bias-removed absolute deviance of a component from the mean, divided by the contribution of all components. Note that surface storage is included in the groundwater component for the models SURFEX-TRIP and PCR-GLOBWB.

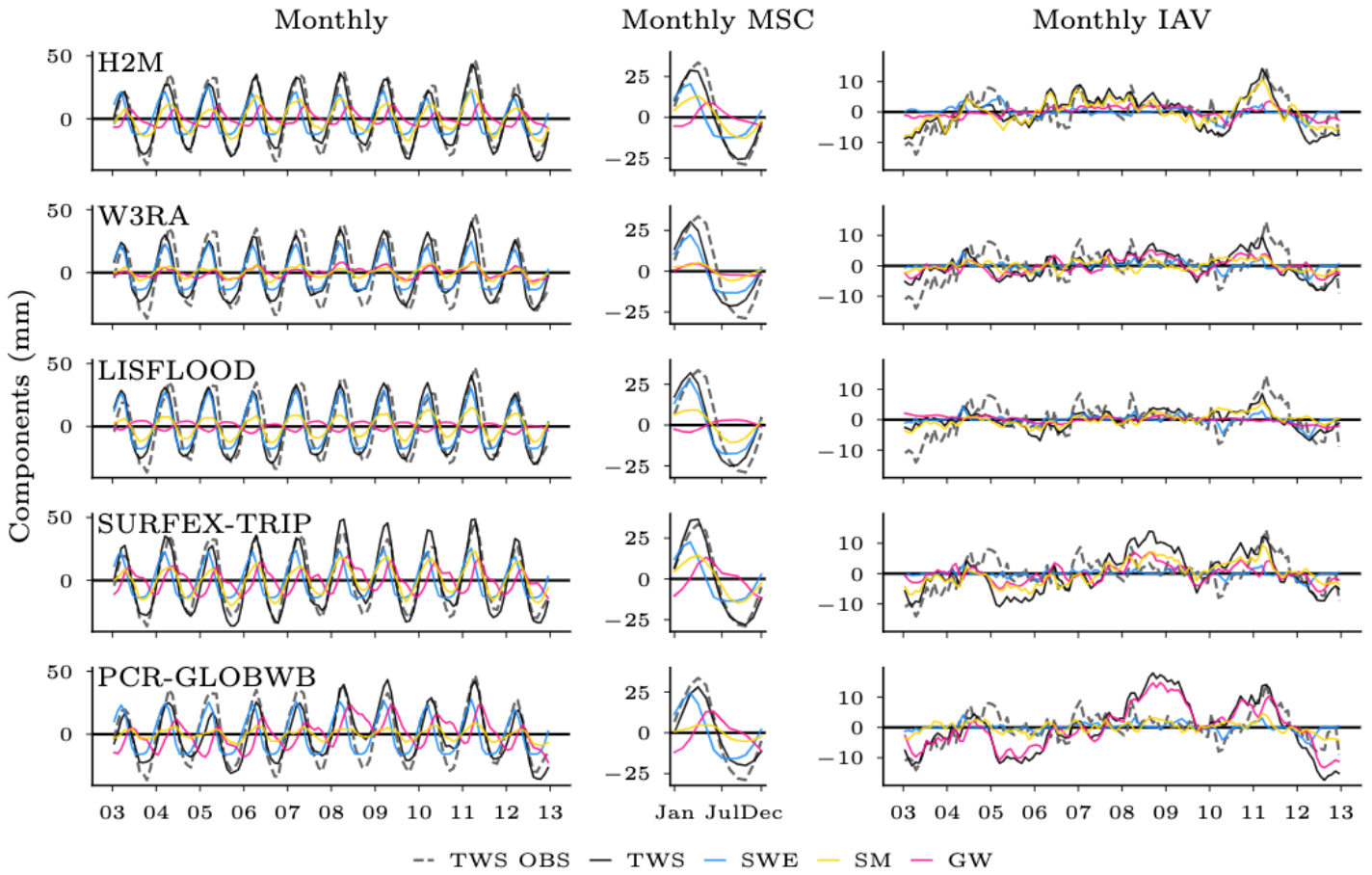

Figure 11. Global average variability of the terrestrial water storage (TWS) and the components snow water equivalent (SWE), soil moisture (SM), and groundwater (GW) for the hybrid hydrological model (H2M) and the process-based global hydrological models (rows). For reference, the TWS observations are shown (TWS OBS). The monthly signal (left) and its decomposition into the mean seasonal cycle (MSC, center) and the interannual variability (IAV, right) are shown (columns). The time-series were aggregated using the cell size weighted average, only cell-timesteps present in all model simulations were used. The y-scale is consistent in columns but varies across the signal components. The training and test period is shown for the complete years 2003 to 2012. Note that surface storage is included in the groundwater component for the models SURFEX-TRIP and PCR-GLOBWB.

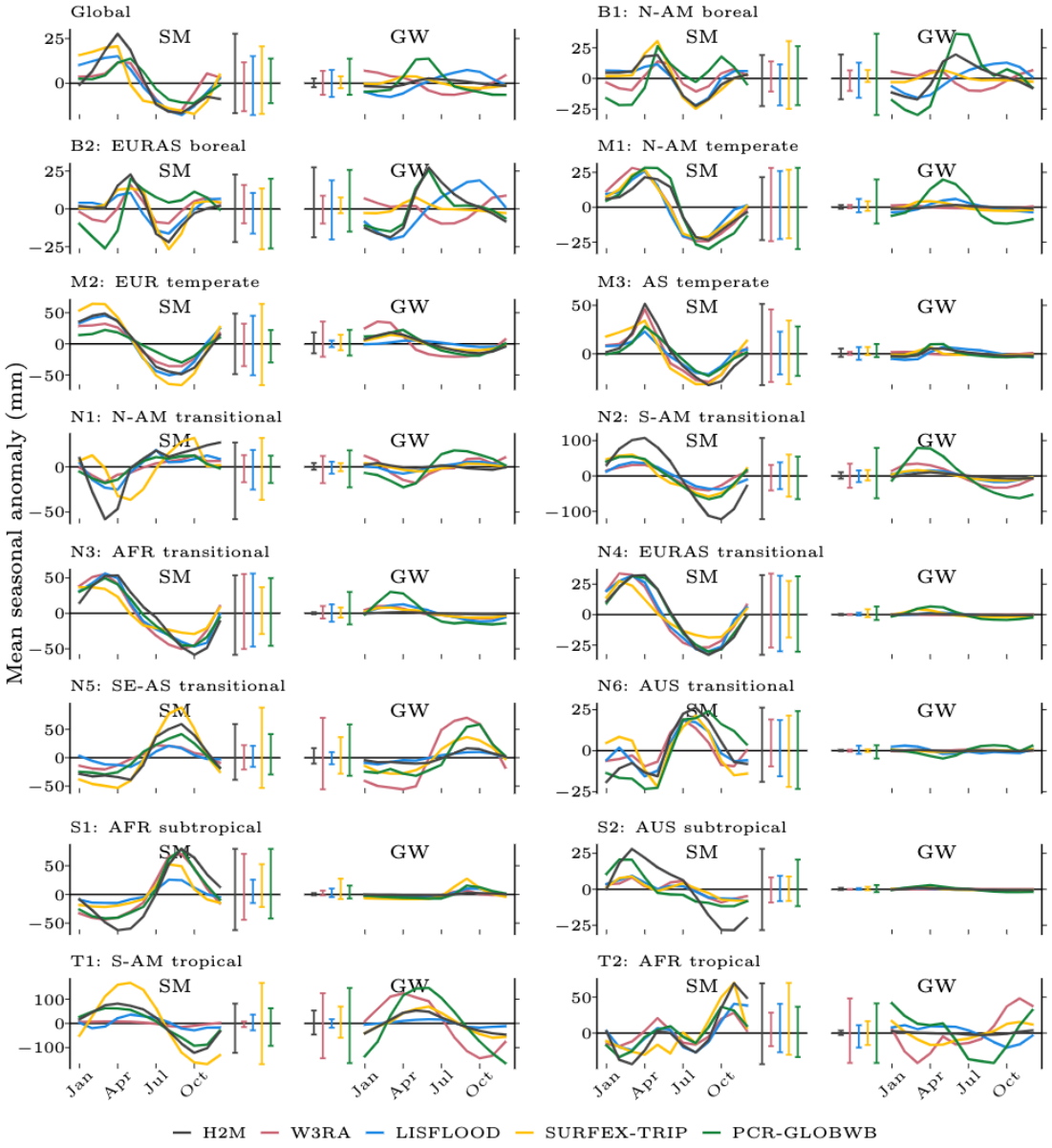

Figure 12. Global and regional mean seasonal anomalies of soil moisture (SM) and groundwater (GW) for the hybrid model (H2M) and the process-based global hydrological models. Ranges from the minimum to the maximum value per model are shown next to the seasonal cycle as vertical lines. The regions are shown in Figure 2. Surface storage is included in the groundwater component for the models SURFEX-TRIP and PCR-GLOBWB. The plots are based on global daily cell-timesteps from 2009 to 2014. Note that the y-scale is consistent within, but differs across regions.

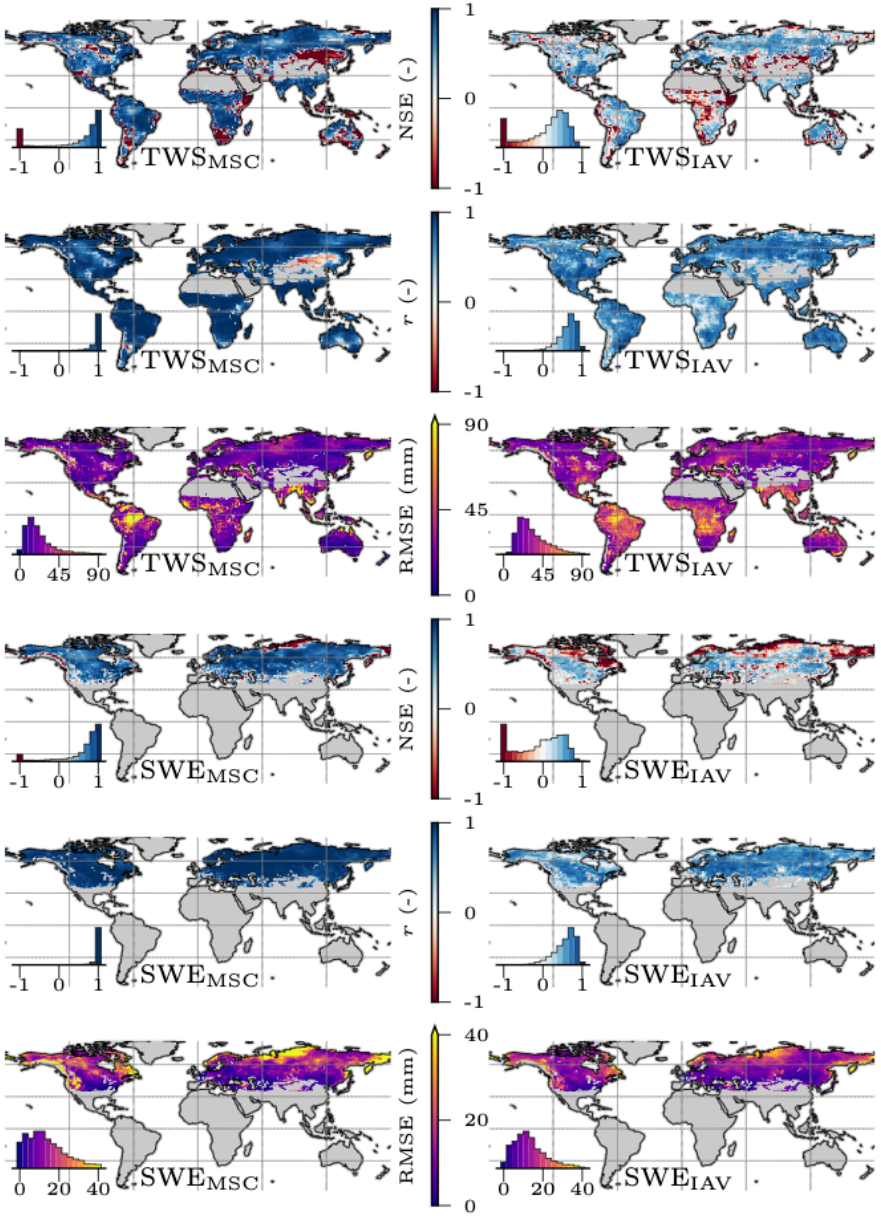

Figure A1. Local model performance for terrestrial water storage (TWS) and snow water equivalent (SWE) on the mean seasonal cycle (MSC) and the interannual variability (IAV) within the test period. The Nash–Sutcliffe model efficiency (NSE), Pearson correlation (r) and Root Mean Square Error (RMSE) are shown. Negative NSE have been remapped to a range -1 to 0 using the hyperbolic tangent function to avoid large negative values. The inset plots show the cell-area weighted histogram of the map values.

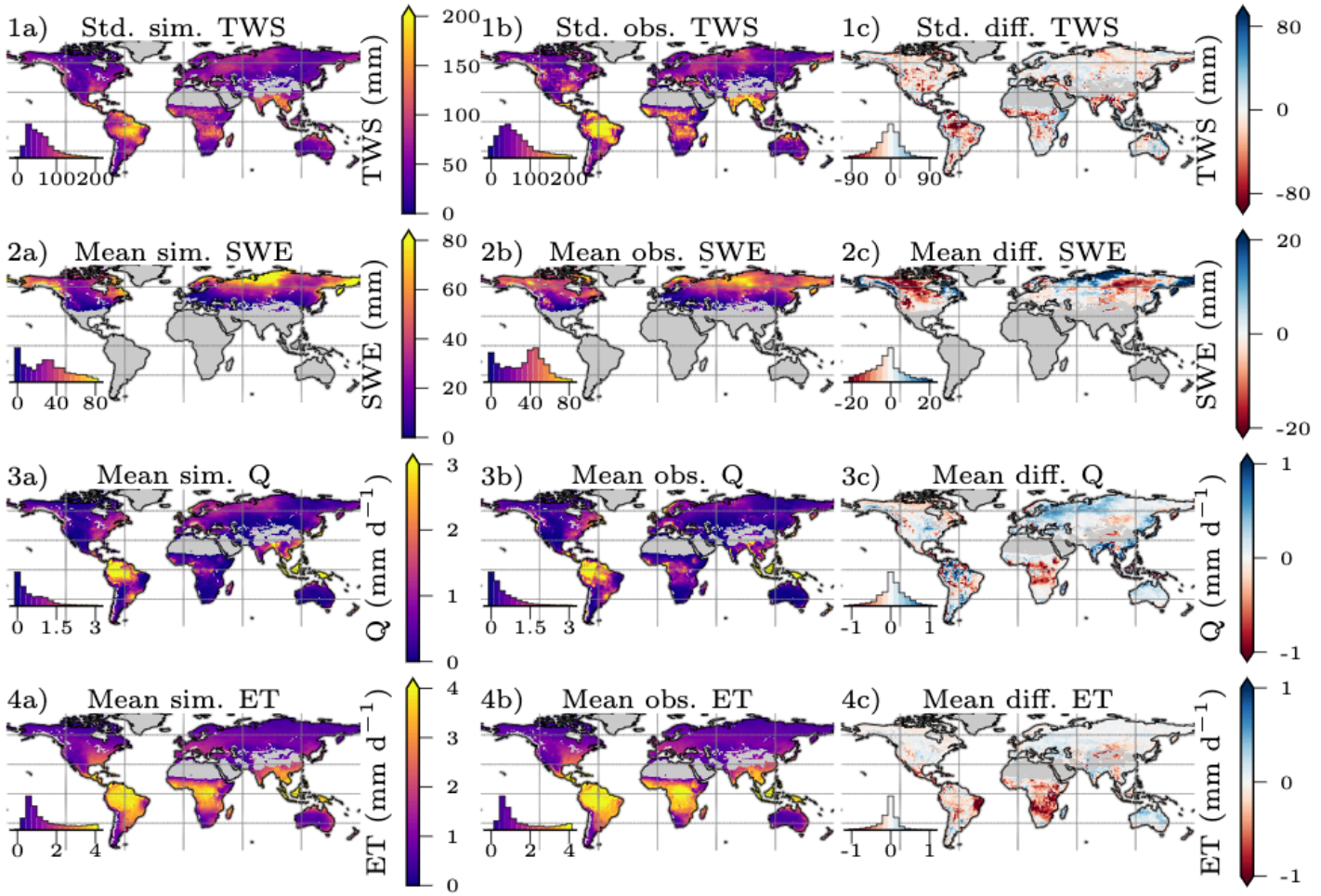

Figure A2. Mean a) simulated, b) observed, and c) difference of simulated - observed (positive means simulated is larger) terrestrial water storage (TWS, 1a–c), snow water equivalent (SWE, 2a–c), total runoff (Q, 3a–c), and evapotranspiration (ET, 4a–c). Note that for the TWS, the standard deviation is shown as the values represent variations around the mean. The inset histograms represent the map value distributions, the mean for the test period (2009 to 2014) is shown.

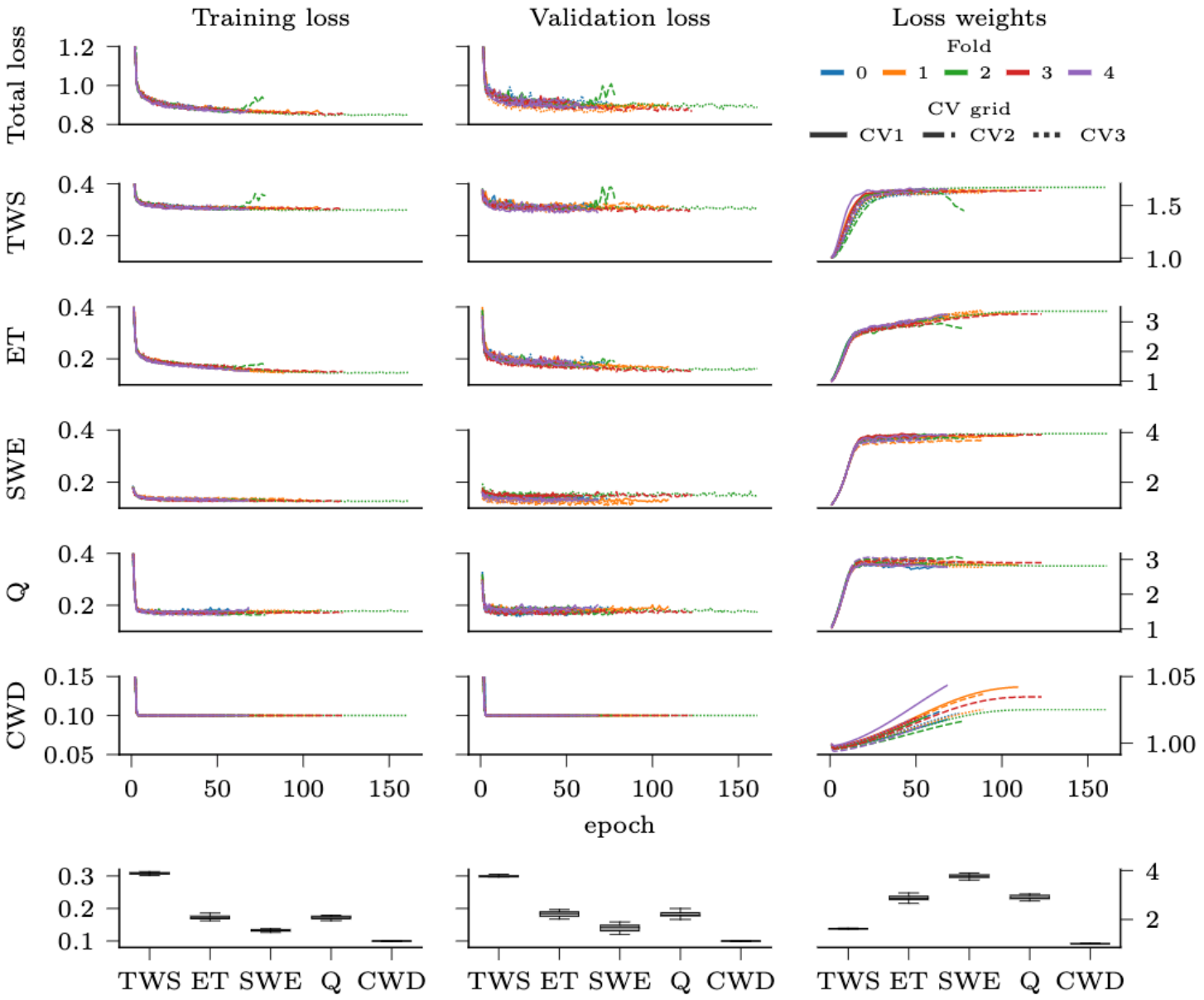

Figure B1. Model training process for the cross validation runs. The left and central column represent the unweighted total and variable-specific MSE loss. The right column shows how the variable weights developed over training time. The x-axis represents the number of iterations through the training set ("epochs"). The bottom row contains the column-wise distribution of the variables losses (or weights) at the end of the model optimization. Note that for the soft constraint on CWD, a bias of 0.1 was added, i.e., 0.1 is the optimum.

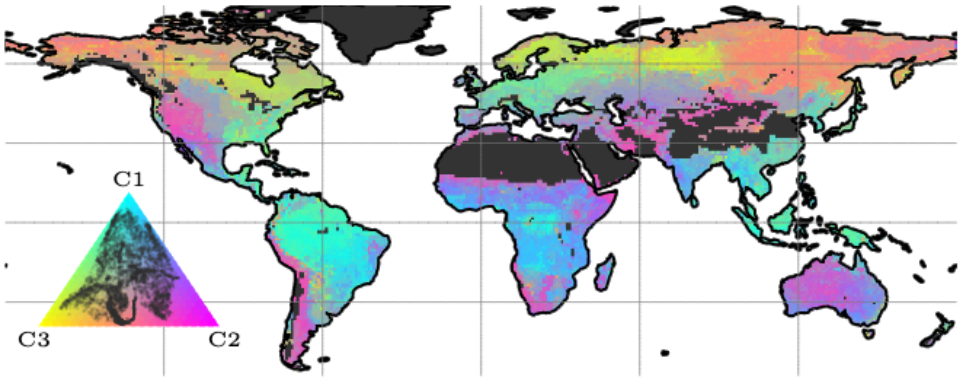

Figure B2. The t-distributed stochastic neighbor (t-SNE) reduction to 3 dimensions of static variable encoding (originally 12 dimensions) of one cross-validation run. The encoding is a low-level representation of the static inputs, i.e., soil and land-cover properties, learned by a neural network. The inset ternary plots show the distribution of the map values.