Supplemental Material For

Emulating Dynamically Downscaled Snow Water Equivalent Across the Western United States: Training Data Composition Is Essential for Robust Climate Projections

Fanghe Zhao[1], Stefan Rahimi[1], Melissa Bukovsky[1,2], Karandeep Singh[3], Chad Thackeray[3], Jesse Norris[3], Alex Hall[3], Weichen Liu[1]

[1]Department of Atmospheric Science, University of Wyoming, WY, USA
[2]Haub School of Environment and Natural Resources, University of Wyoming, WY, USA
[3]Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA

Table of Contents:

Figure S1. WRF domain, evaluation regions, and input grid.

Figure S2. SWEDE training and validation diagnostics.

Figure S3. April 1st snow water volume by subregion.

Figure S4. Yearly snow water volume by subregion.

Figure S5. Three-way comparison of SWE.

Figure S6. Quantile comparison of non-zero SWE.

Figure S7. Seasonal cycle of maximum regional SWE.

Figure S8. Pixel-level SWE prediction performance.

Figure S9. Regional MAE across ESM simulations.

Figure S10. Regional correlation across ESM simulations.

Figure S11. SWE predictions for the FGOALS-g3 target.

Figure S12. SWE predictions for the CanESM5 target.

Table S1. Model input variables and resolutions.

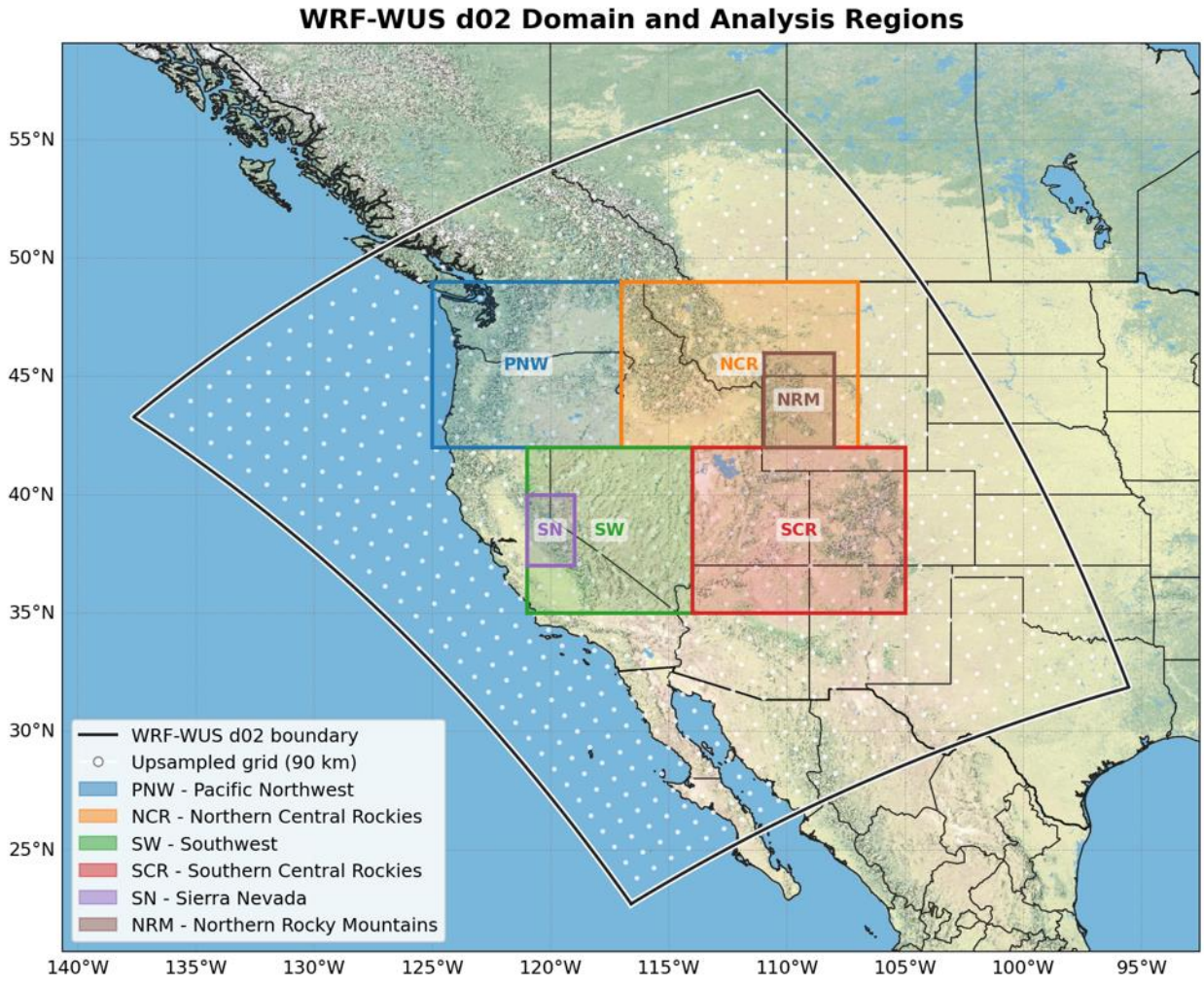


Figure S1. The WUS-D3 WRF simulation domain (d02, black outline) at 9 km resolution and the six evaluation subregions used in this study: Pacific Northwest (PNW), Northern Central Rockies (NCR), Southwest (SW), Southern Central Rockies (SCR), Sierra Nevada (SN), and Northern Rocky Mountains (NRM), following the regional definitions of Stefan et al. (2019). White circles denote the coarsened input grid at approximately 90 km resolution, obtained via block averaging.

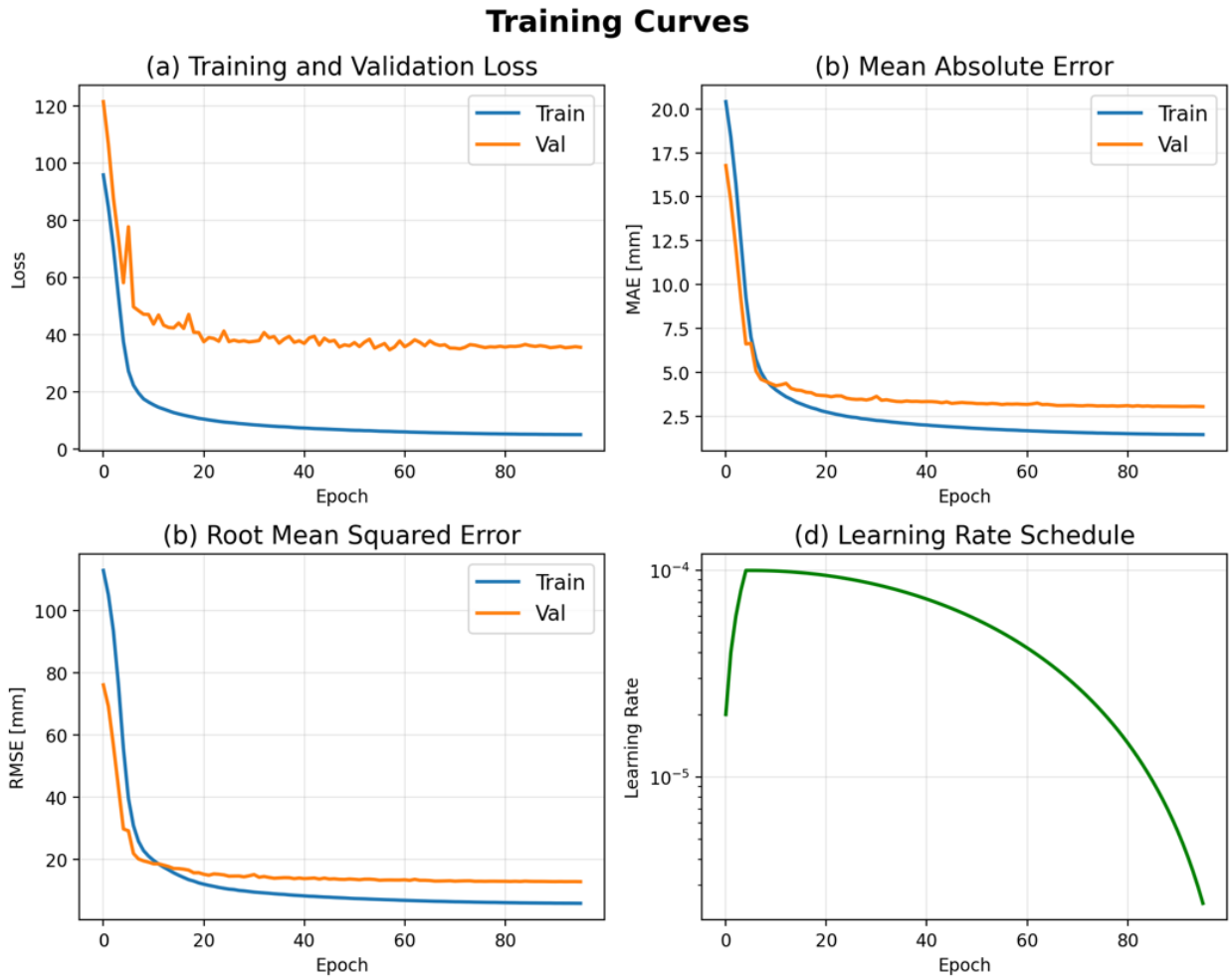


Figure S2. Training diagnostics for the SWEDE model over 100 epochs, showing (a) composite training and validation loss, (b) mean absolute error (MAE, mm), (c) root mean squared error (RMSE, mm), and (d) the cosine annealing learning rate schedule with 5-epoch linear warmup.

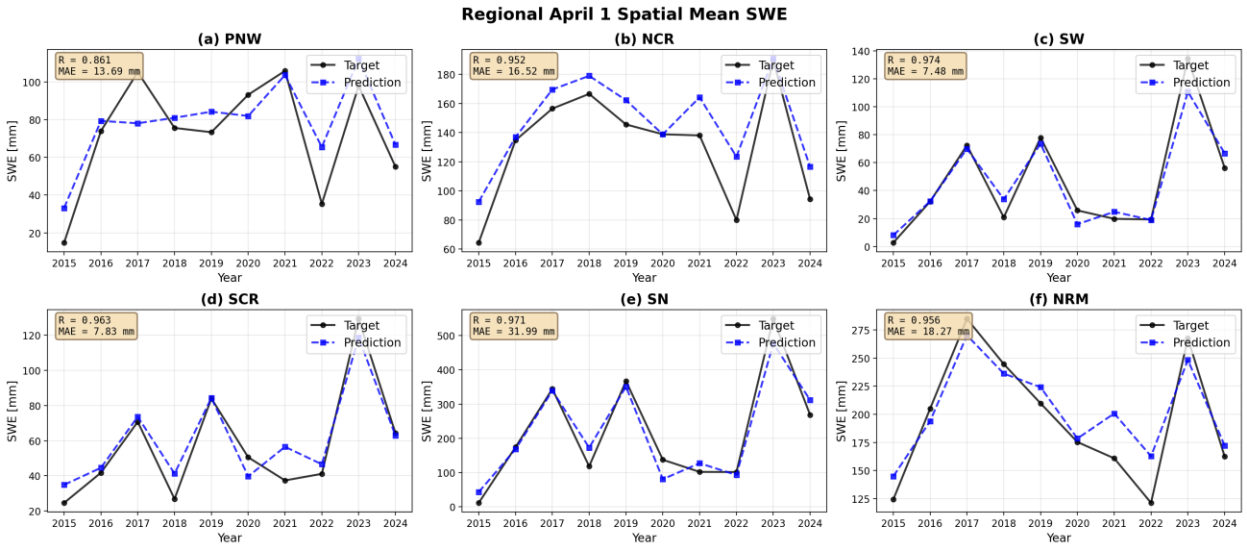


Figure S3. Regional April 1st of spatial mean SWE (mm) for the six evaluation subregions during the WRF-ERA5 held-out period (2014-2024). Black lines denote the WRF target and blue dashed lines denote the SWEDE prediction. Pearson correlation (R) and mean absolute error (MAE, mm) are annotated in each panel.

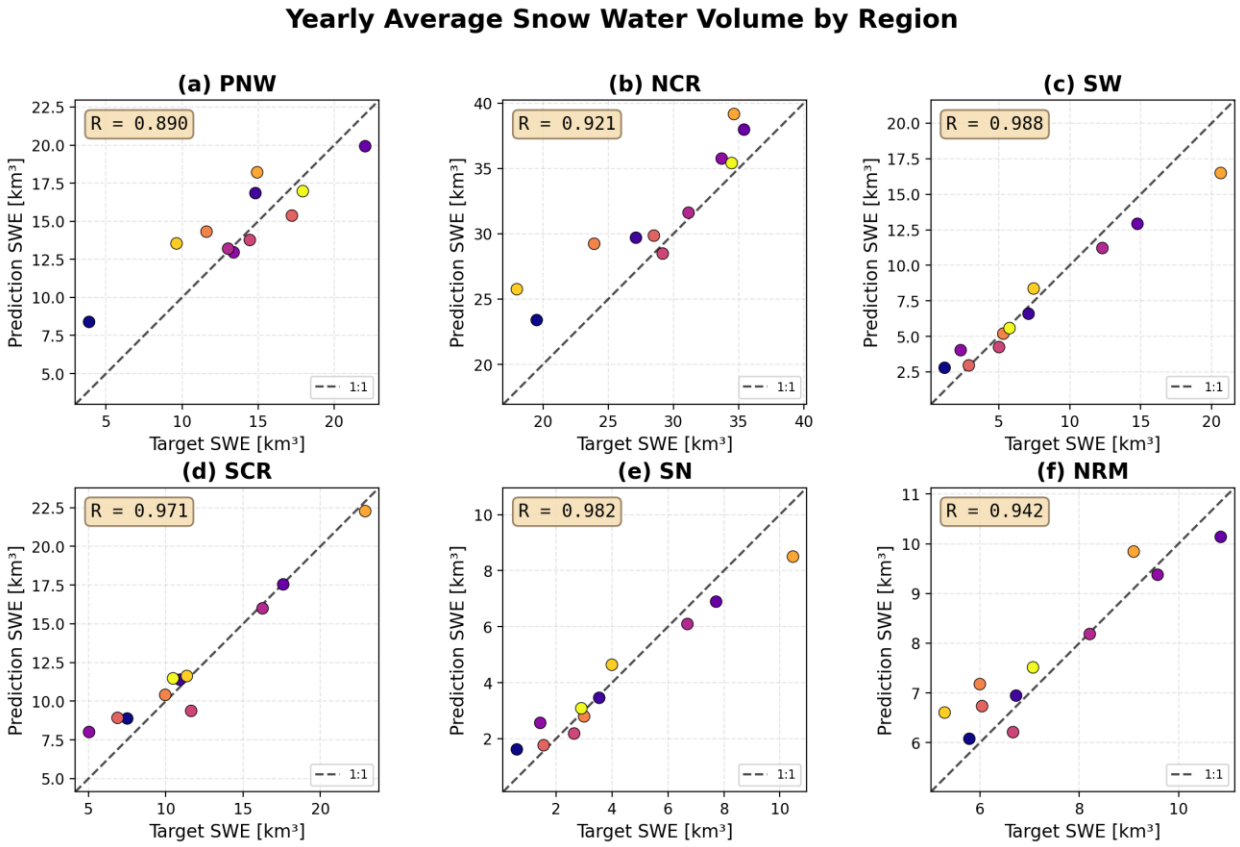


Figure S4. Yearly average snow water volume (km³) by subregion during the WRF-ERA5 held-out period (2014-2024), shown as scatter plots of SWEDE prediction against WRF target. Points are colored by year, with the 1:1 reference line shown as a dashed gray line. Pearson correlation (R) is annotated in each panel.

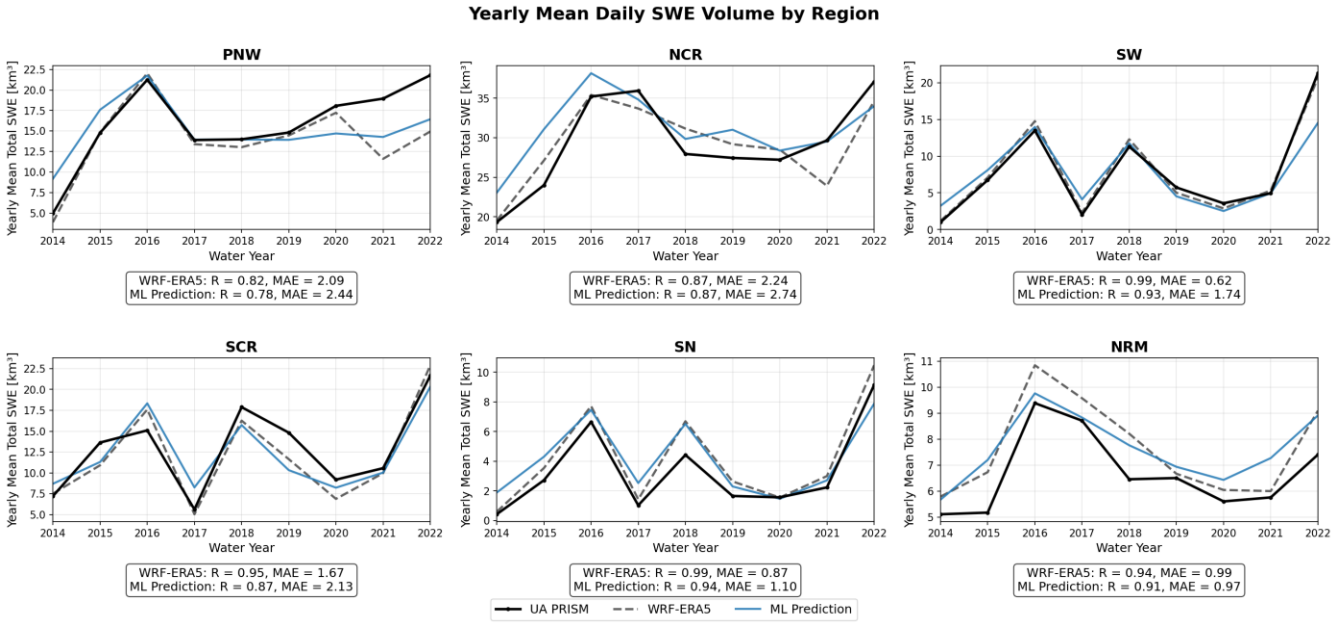


Figure S5. Three-way comparison of SWE among SWEDE (ML Prediction, blue), WRF-ERA5 (ERA5, gray dashed), and the University of Arizona PRISM-based SWE product (UA PRISM, black) for the six evaluation subregions. (a) Yearly mean daily SWE volume (km³) for water years 2014-2022, with Pearson correlation (r) and MAE reported for both WRF-ERA5 and SWEDE against UA PRISM.

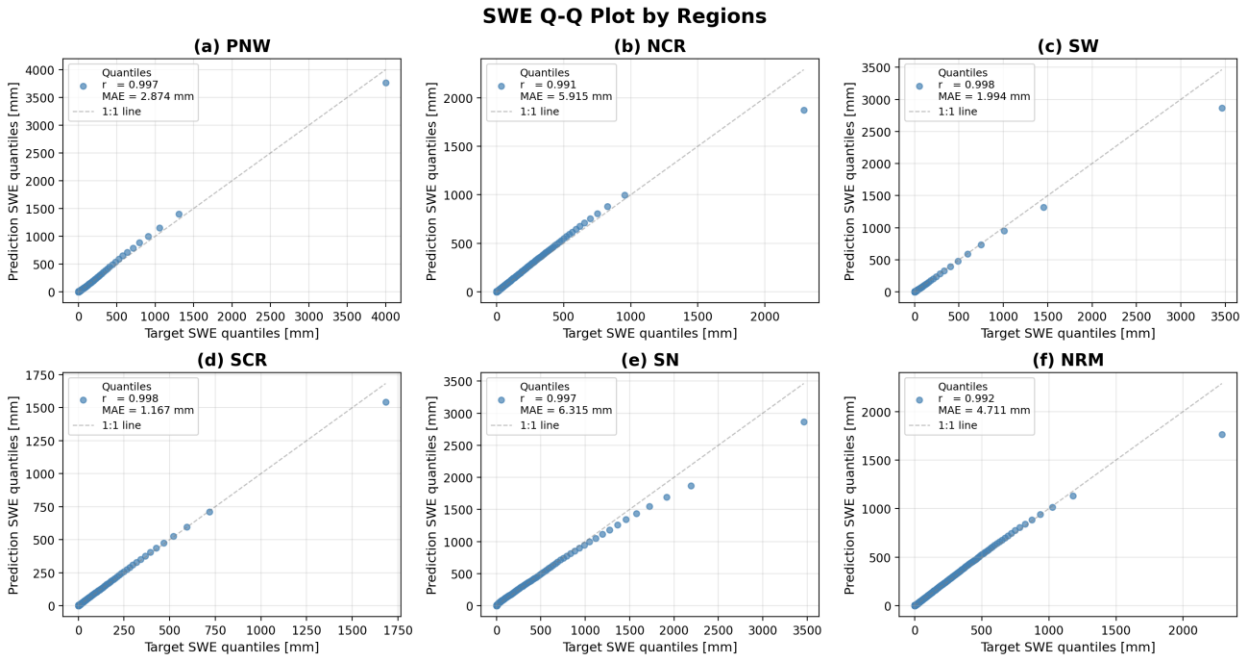


Figure S6. Quantile-quantile plots of non-zero SWE (mm) for the six evaluation subregions during the WRF-ERA5 held-out period (2014-2024). Each panel compares the quantiles of SWEDE predictions against the WRF target, with the 1:1 reference line shown as a dashed gray line. Pearson correlation (r) of the quantile pairs is annotated in each panel.

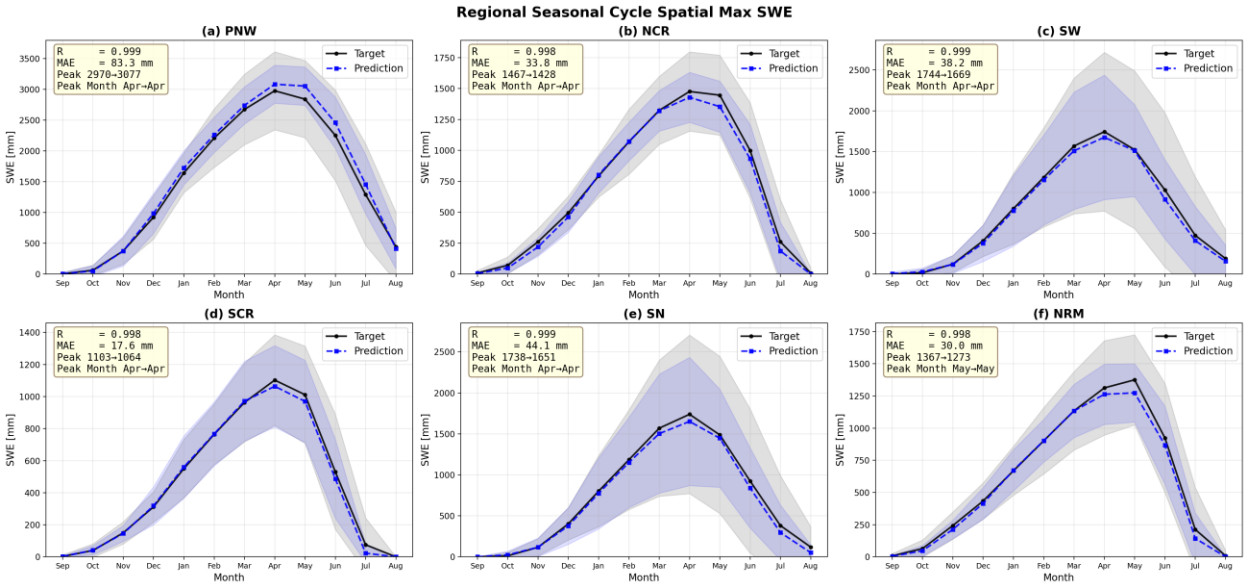


Figure S7. As in Figure 4, but for the spatial maximum SWE (mm) within each subregion. The spatial maximum peaks approximately one month later than the regional mean in most subregions.

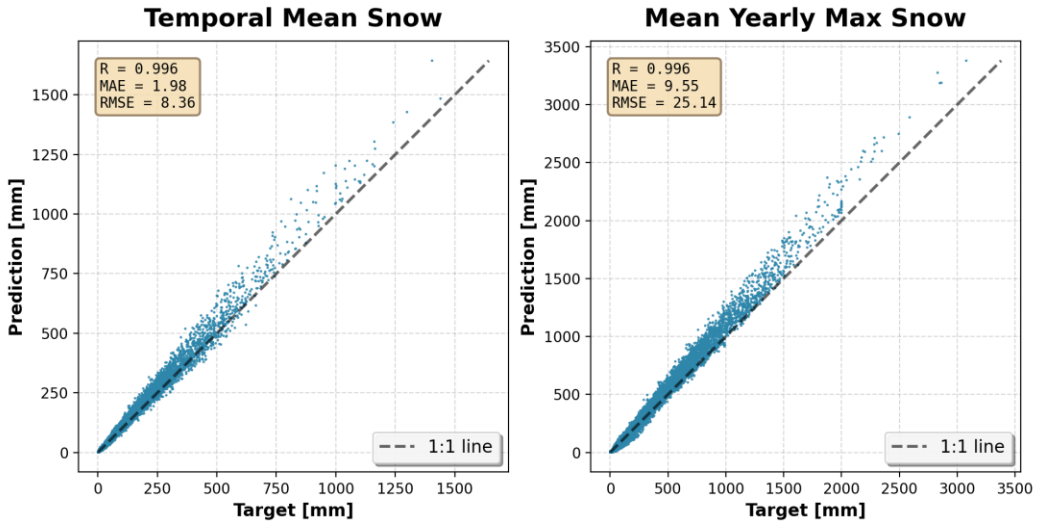


Figure S8. Pixel-level scatter analysis of SWEDE predictions against the WRF-ERA5 held-out period (2014-2024): (a) temporal mean SWE (mm) and (b) mean yearly maximum SWE (mm) across all grid cells in the domain. Pearson correlation (r), MAE (mm), and RMSE (mm) are annotated in each panel. The 1:1 reference line is shown as a dashed gray line.

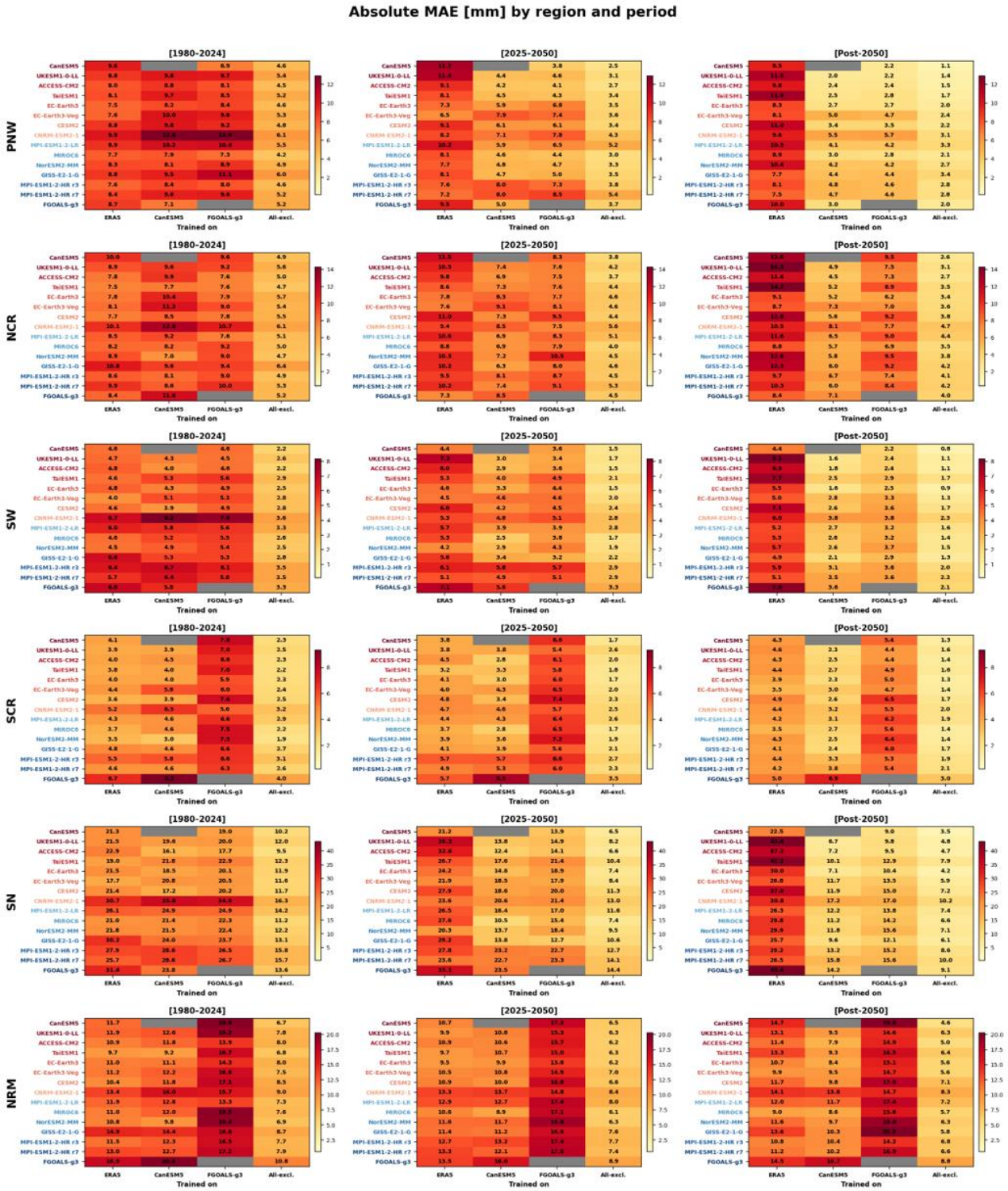


Figure S9. Absolute MAE (mm) for SWEDE evaluated on each of the 15 ESM-driven WRF simulations, shown separately for the six evaluation subregions (PNW, NCR, SW, SCR, SN, NRM) across three time periods: 1980-2024, 2025-2050, and post-2050. Columns correspond to training configurations (WRF-ERA5, CanESM5, FGOALS-g3, and All-excl.), and rows

correspond to the evaluation ESM. Diagonal entries for single-ESM training configurations are excluded. Lower MAE (lighter shading) indicates better performance. This figure provides the regional decomposition of the whole-domain results shown in Figure 14.

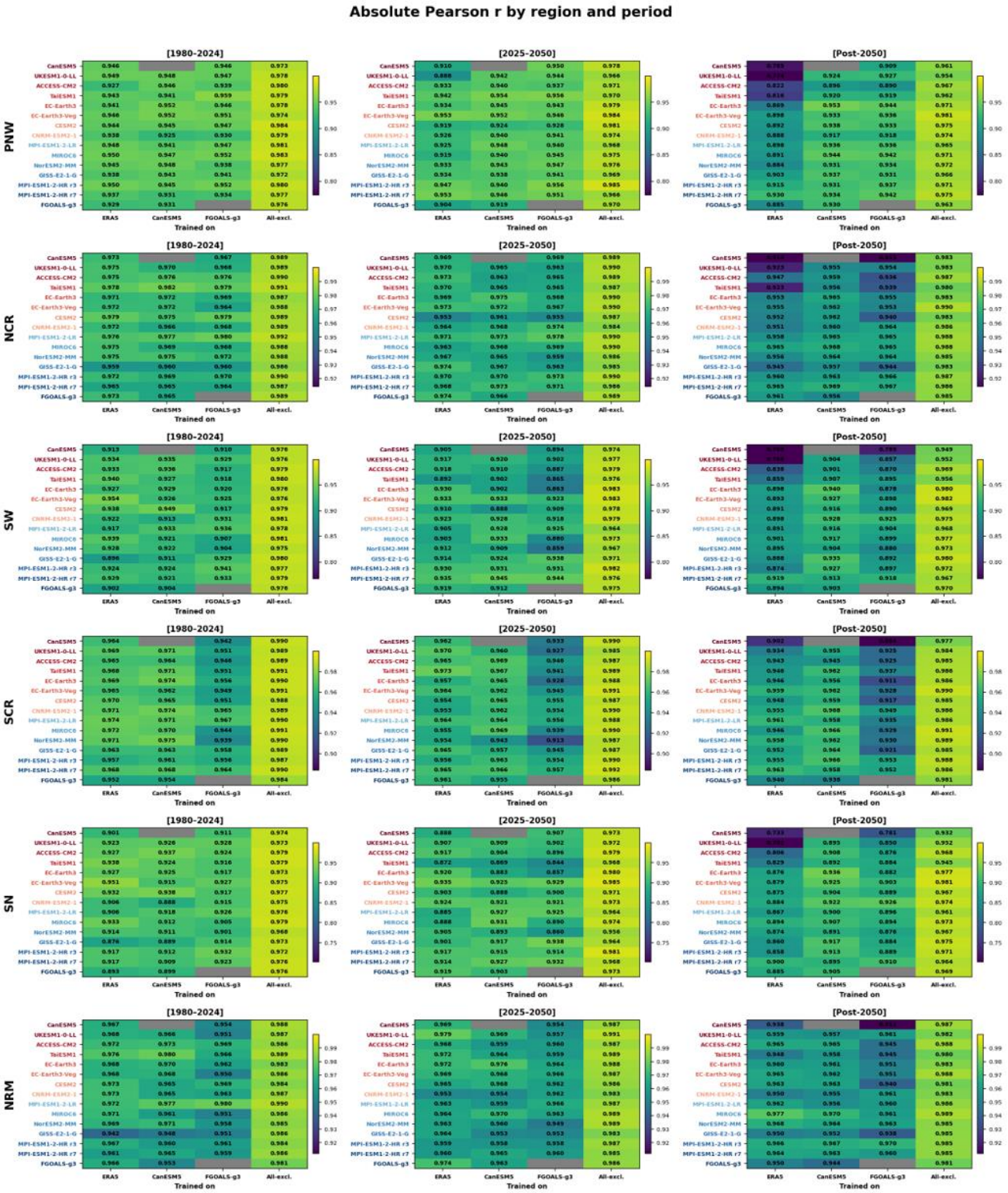


Figure S10. As in Figure S9 but showing Pearson correlation (r) for each subregion. Higher r (darker shading) indicates better performance. This figure provides the regional decomposition of the whole-domain results shown in Figure 15.

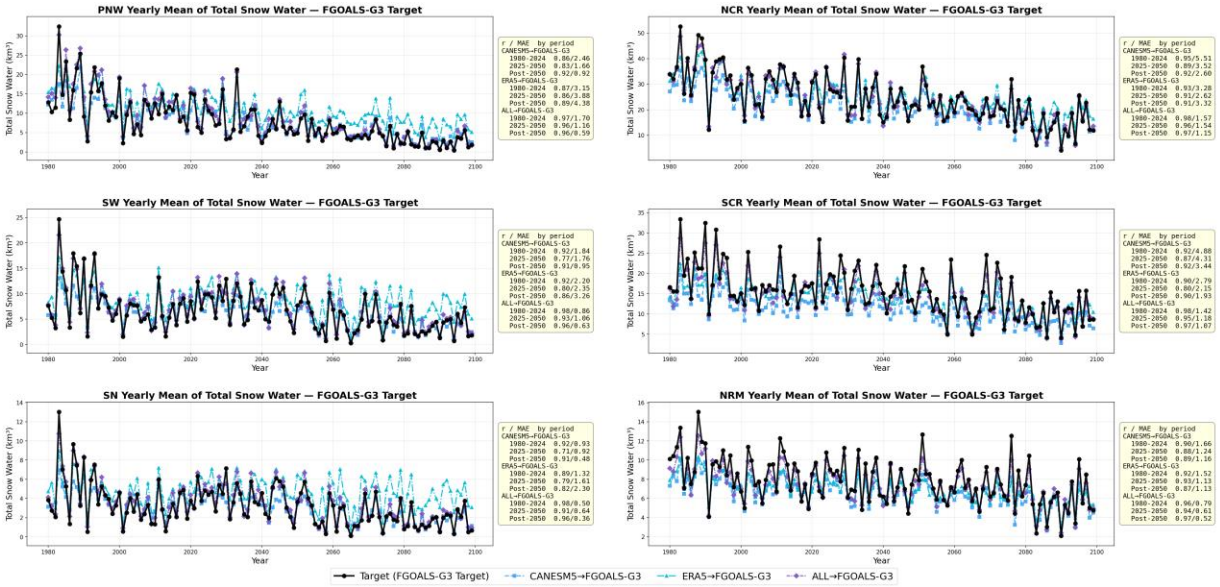


Figure S11. Yearly mean total snow water (km³) for the six evaluation subregions with FGOALS-g3 as the evaluation target. Blue line denotes the WRF-downscaled FGOALS-g3 target, and colored lines denote SWEDE predictions from three training configurations: CanESM5-trained, WRF-ERA5-trained, and All-excl. (trained on all ESMs excluding FGOALS-g3). Pearson r and MAE are annotated for each training configuration across the three evaluation periods.

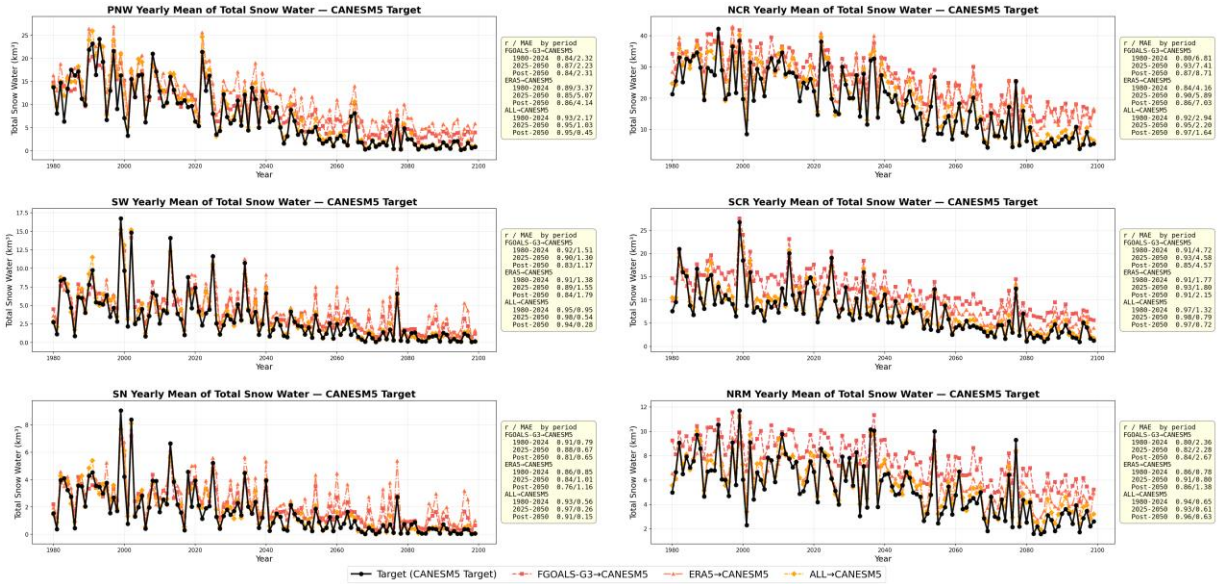


Figure S12. As in Figure S11, but with CanESM5 as the evaluation target. Training configurations shown are FGOALS-g3-trained, ERA5-trained, and All-excl. (trained on all ESMs excluding CanESM5).

Table S1. Static and dynamic input variables used in the Dual-Encoder Swin U-Net. Static variables are provided on the 9 km target grid, whereas dynamic variables are provided at approximately 90km resolution

Branch	Variable	Description
Static	LANDMASK	Binary land/water mask
Static	LU_INDEX	USGS land use category
Static	HGT	Terrain elevation (m)
Static	VAR	Sub-grid terrain variance
Static	GRADHGT_X	Zonal terrain gradient
Static	GRADHGT_Y	Meridional terrain gradient
Dynamic	t2	2-m air temperature (K)
Dynamic	t2min	Daily minimum temperature (K)
Dynamic	t2max	Daily maximum temperature (K)
Dynamic	q2	2-m specific humidity (kg kg^{-1})
Dynamic	psfc	Surface pressure (Pa)
Dynamic	prec	Daily total precipitation (mm)
Dynamic	lw_dwn	Downward longwave radiation (W m^{-2})
Dynamic	sw_dwn	Downward shortwave radiation (W m^{-2})
Dynamic	t2_lag90d	90-day mean of t2
Dynamic	prec_lag110d	110-day mean of precipitation
Dynamic	lw_dwn_lag80d	80-day mean of downward longwave radiation
Dynamic	sw_sfc_lag130d	130-day mean of surface shortwave radiation
Dynamic	q2_lag45d	45-day mean of q2