Supplementary information

Table S1 Model Feature for each prediction with DeepMet

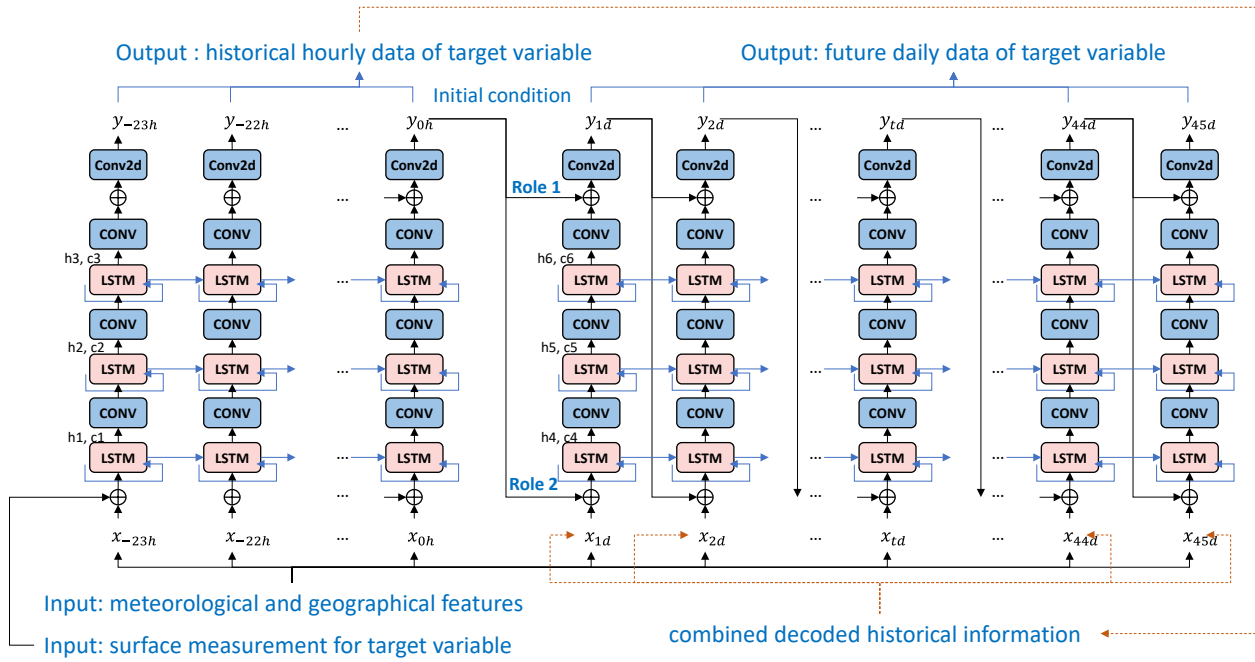| Predict variable	Time independent		Time-dependent	
	Geographic factors	2D features (climatological data)	2D Feature (historical only)	3D feature (historical only)
T2max			UWIND, VWIND, CFRAC, PBL, prep, Q2, RGRND, T2, WSPD10, WSTAR, HFX, LH, USTAR, ZRUF, PRSFC, WBAR, WR, SNOCOV	TA
T2min	DLUSE, HT, LWMASK, MSFX2, LUFRAC, PURB, LAT, LON	LAI, VEG, ALBEDO, SWDNBC		TA
Q2				QC, QV, TA, CFRAC_3D
WSPD10				UW, VW

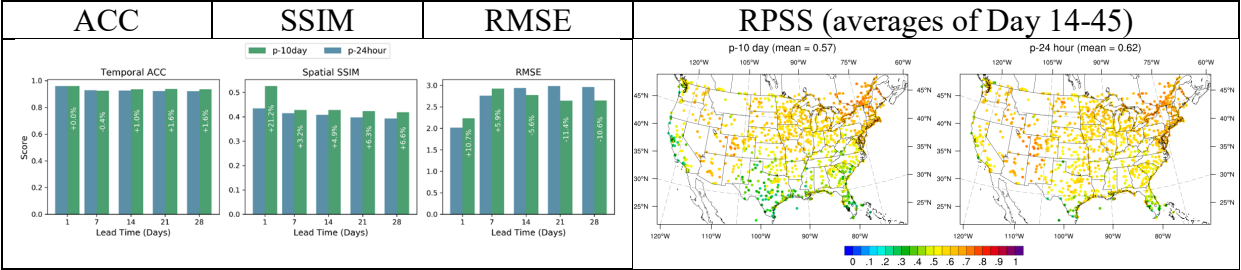**Figure S1.** Model duo-ConvLSTM structure of DeepMet

Figure S2. Comparison of DeepMet performance in predicting following days of T2 with 10-day historical window rather than previous 24 hours
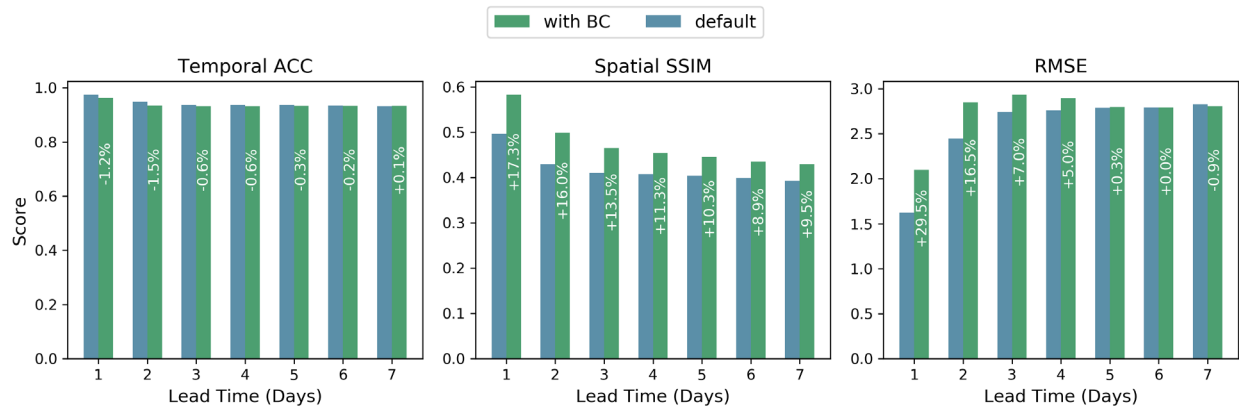
Figure S3. Comparison of DeepMet performance in forecasting 7-day T2 with and without boundary condition (BC) information. (Default refers to predictions without considering future BCs, while "With BC" incorporates BC information inferred from global model forecasts. Results indicate that including BCs notably improves prediction accuracy in the first 4 days, with diminishing impact for longer lead times; testing data for the whole year of 2023)
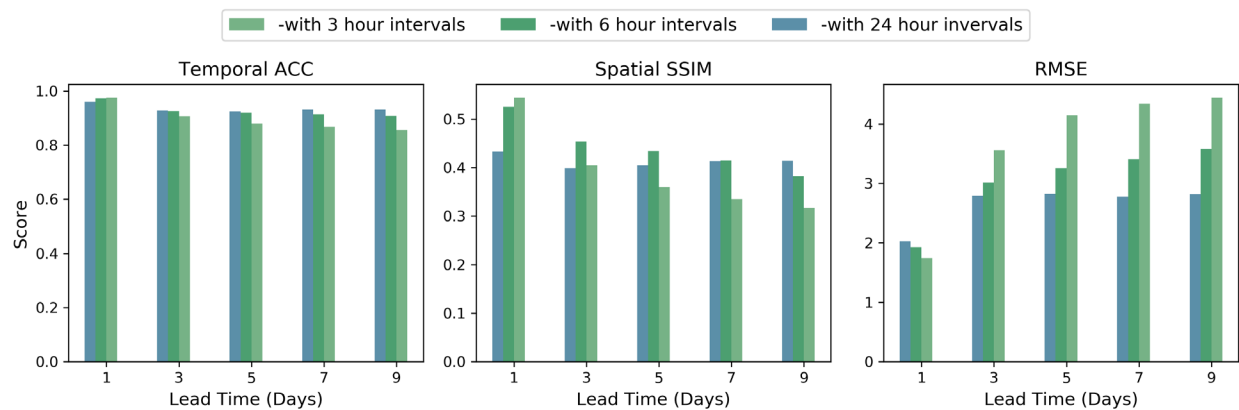
Figure S4. Comparison of DeepMet performance in predicting following days of T2 with different intervals by 3 hour, 6 hour and 24 hour
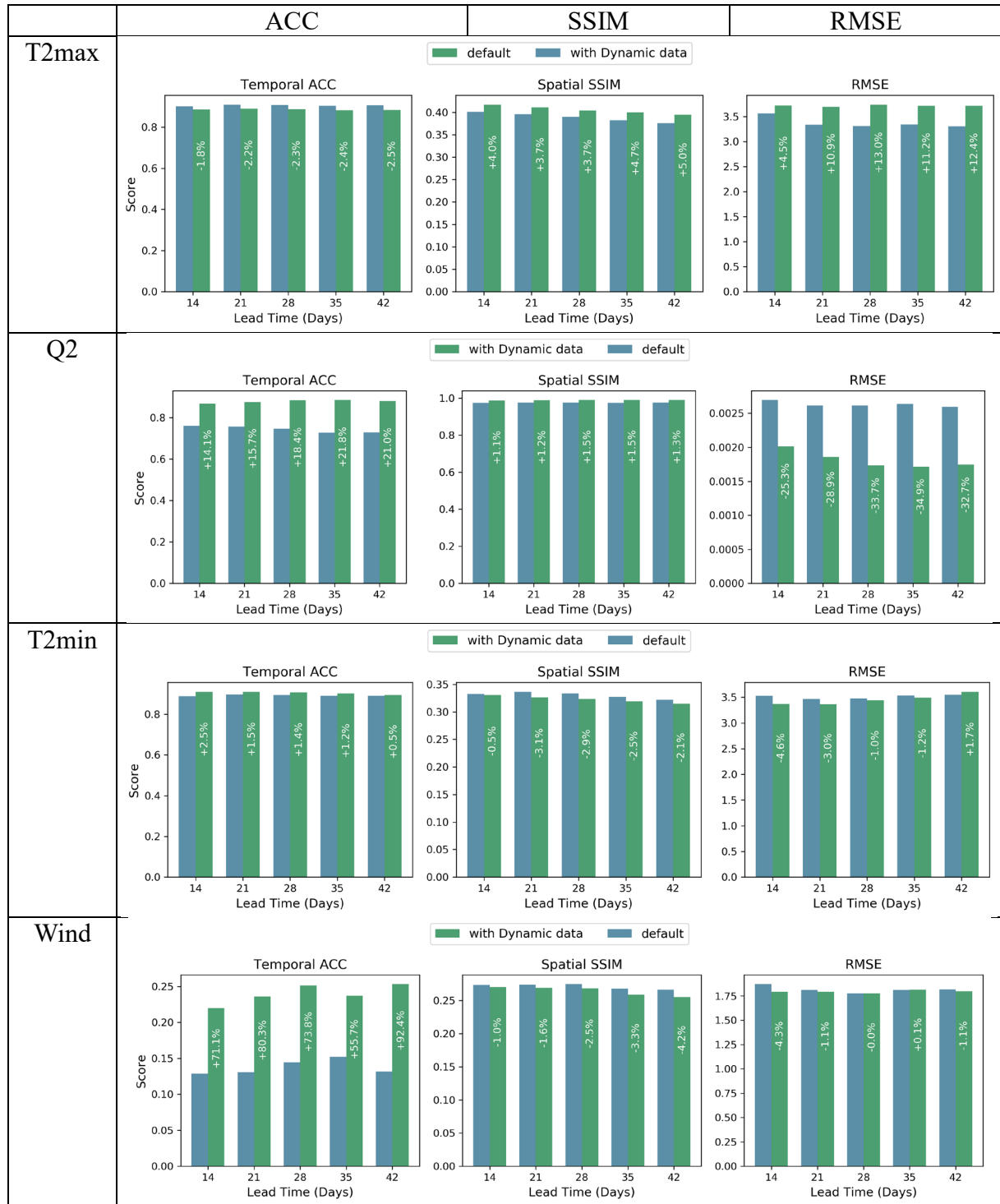
Figure S5. Comparison of DeepMet performance in predicting following days with T2 with better representations of time-series for slowly changing geophysical inputs (e.g., LAI, vegetation cover, albedo, downward shortwave radiation)
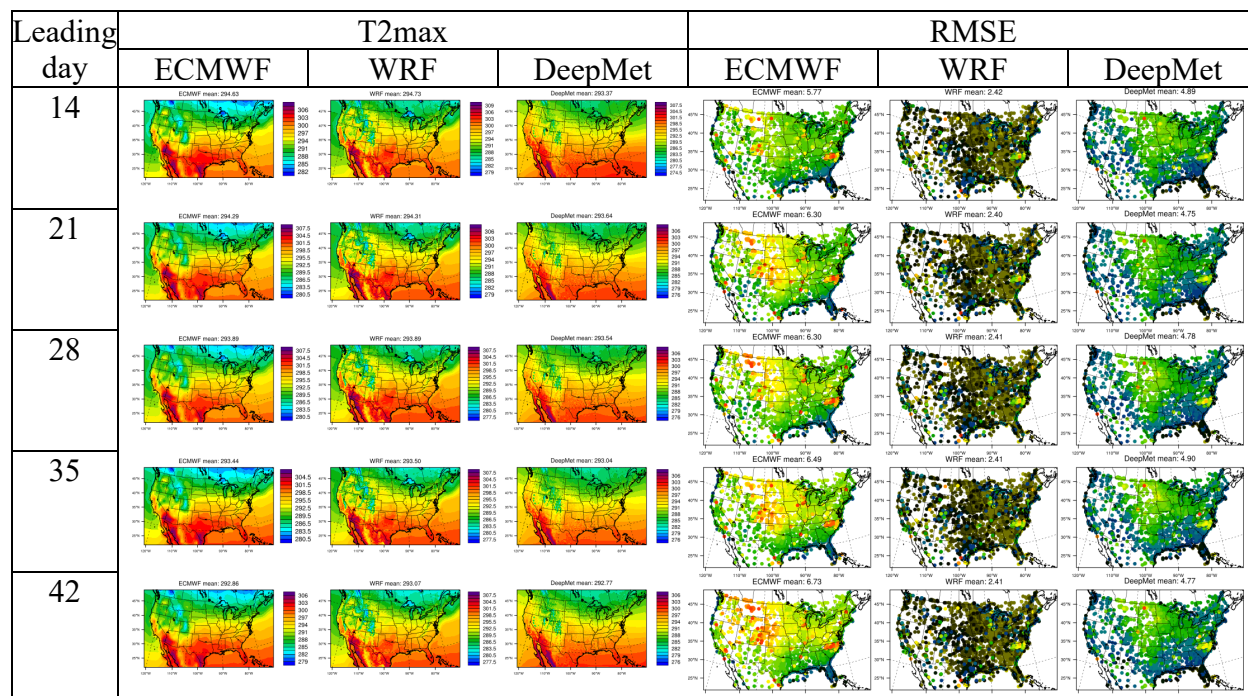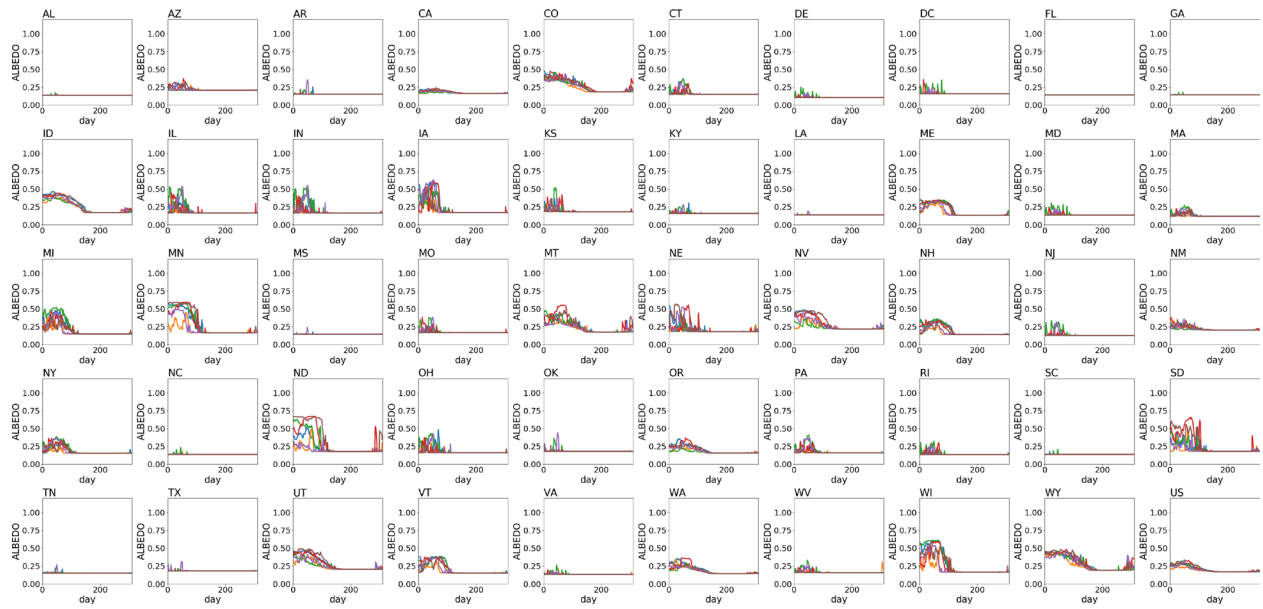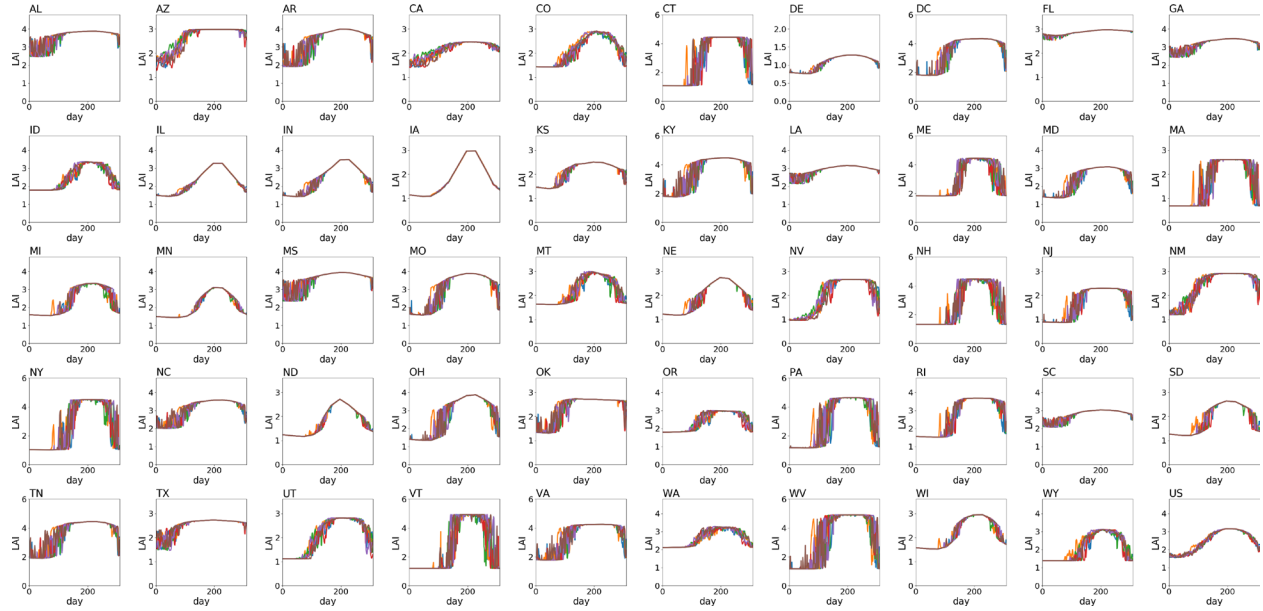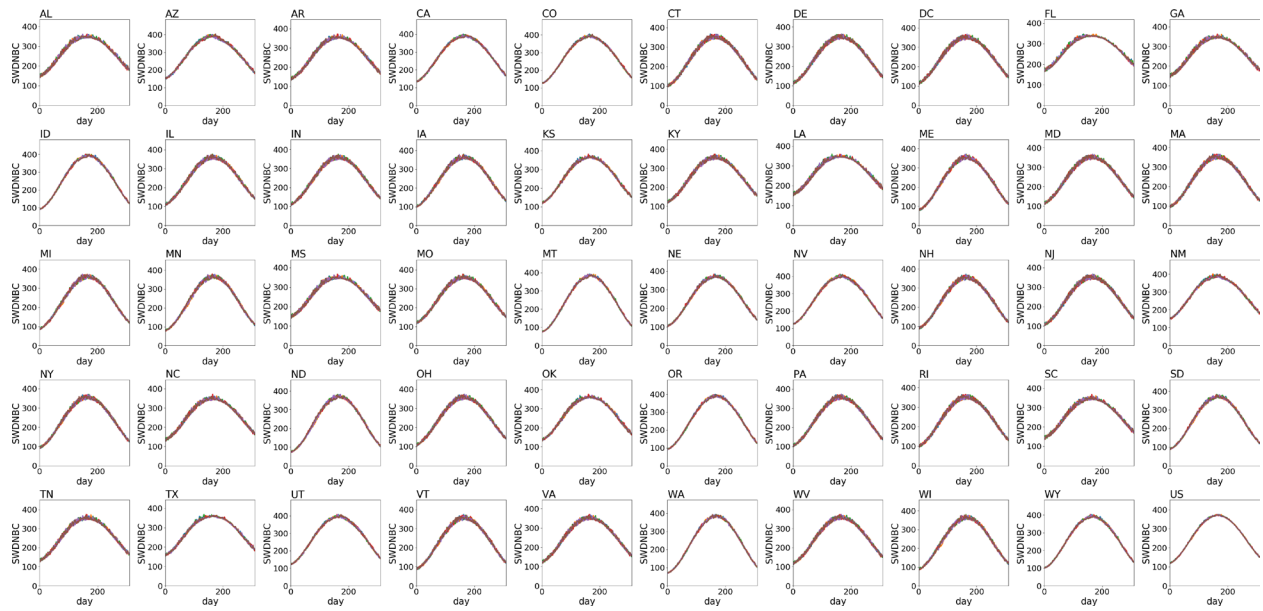
Leading day	T2max			RMSE		
	ECMWF	WRF	DeepMet	ECMWF	WRF	DeepMet
14						
21						
28						
35						
42						

Figure S6. Comparison of ECMWF, WRF, and DeepMet prediction of T2max in following 14-42 days and their comparison with ground measurement from NCDC
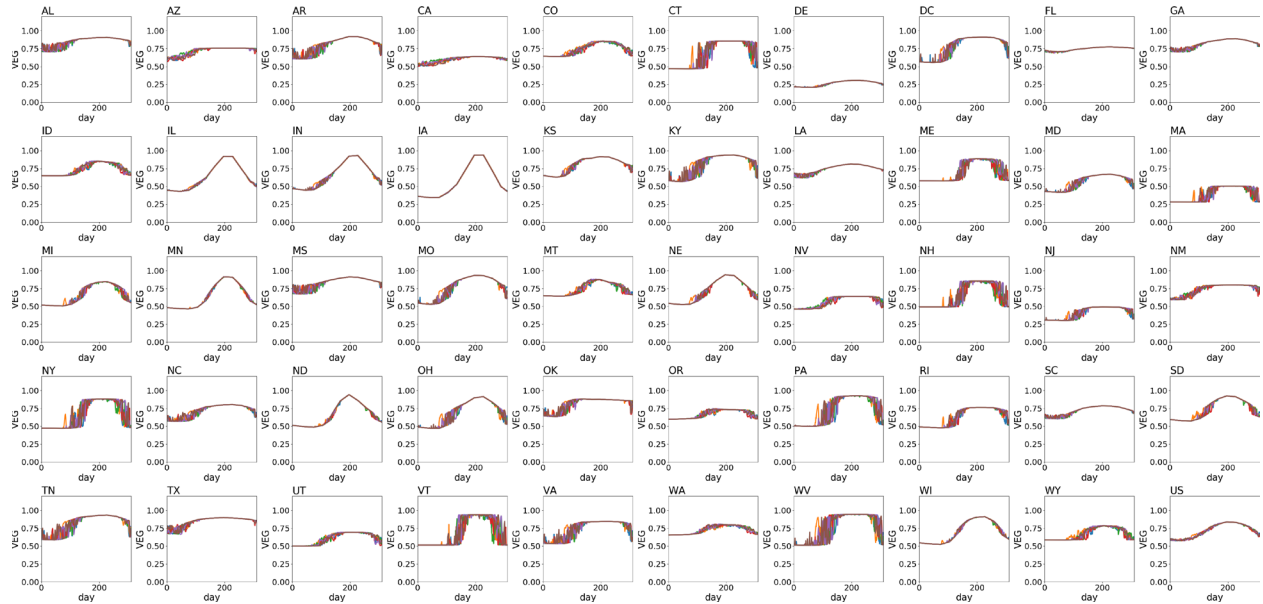
(a) albedo

(b) LAI

(c) SWDNBC

(d) VEG

Figure S7. Comparison of the daily-variation pattern of variables among years of 2008 (blue), 2012 (orange), 2014 (green), 2019 (red), 2021(purple), and 2023(brown)