Supplement of Earth Syst. Sci. Data, 14, 5267–5286, 2022
https://doi.org/10.5194/essd-14-5267-2022-supplement

Supplement of

A 1 km daily soil moisture dataset over China using in situ measurement and machine learning

Qingliang Li et al.

Correspondence to: Wei Shangguan (shgwei@mail.sysu.edu.cn)

Section S1.

Fig. S1(a) shows that stations are dense in the east part of China, but sparse in the west part. Fig. S1(b) represents that the sample size varies with soil depth, and large numbers of missing values exist at 70 and 90 cm soil depths. From Fig. S1 (c), we could see that the values of the *in-situ* SM at all soil depths were mainly concentrated in the range from 0.2 to 0.4 m^3/ m^3. Fig. S1(d) denotes that the data number in low standard deviation (0~0.05 m^3/ m^3) is smaller than that in high standard deviation (0.05~0.07 m^3/ m^3) from at 10 to 40 cm soil depths. But the opposite conclusion can be drawn from 50 to 100 cm soil depths (larger data number in low standard deviation is than that in high standard deviation). Meanwhile, Fig. S1(d) also hints that the standard deviation of SM at deeper soil depth (except that at 100 cm soil depth) is lower than that at upper soil depth. Decreasing standard deviation with increased soil depth denoted that *in-situ* SM is more stable in deep soil depth, which is consistent with the previous studies (Gao and Shao 2012; Wang et al. 2013). From Fig. S1(e), we could see that the stations have 8 climate types, most observations belong to temperate climate with dry winter (Cw), temperate climate, fully humid (Cf) and snow climate with dry winter (Dw), and the data with tropical monsoon climate (Am) and snow climate, fully humid (Df) are sparse, which occupy only small parts of China.

Reference

Gao, L. and Shao, M.: Temporal stability of shallow soil water content for three adjacent transects on a hillslope, Agr. Water Manage., 110, 41–54, https://doi.org/10.1016/j.agwat.2012.03.012, 2012.

Wang, X.., Pan, Y.., Zhang, Y.., Dou, D., Hu, R., and Zhang, H.: Temporal stability analysis of surface and subsurface soil moisture for a transect in artificial revegetation desert area, China, J. Hydrol., 507, 100–109, https://doi.org/10.1016/j.jhydrol.2013.10.021, 2013.

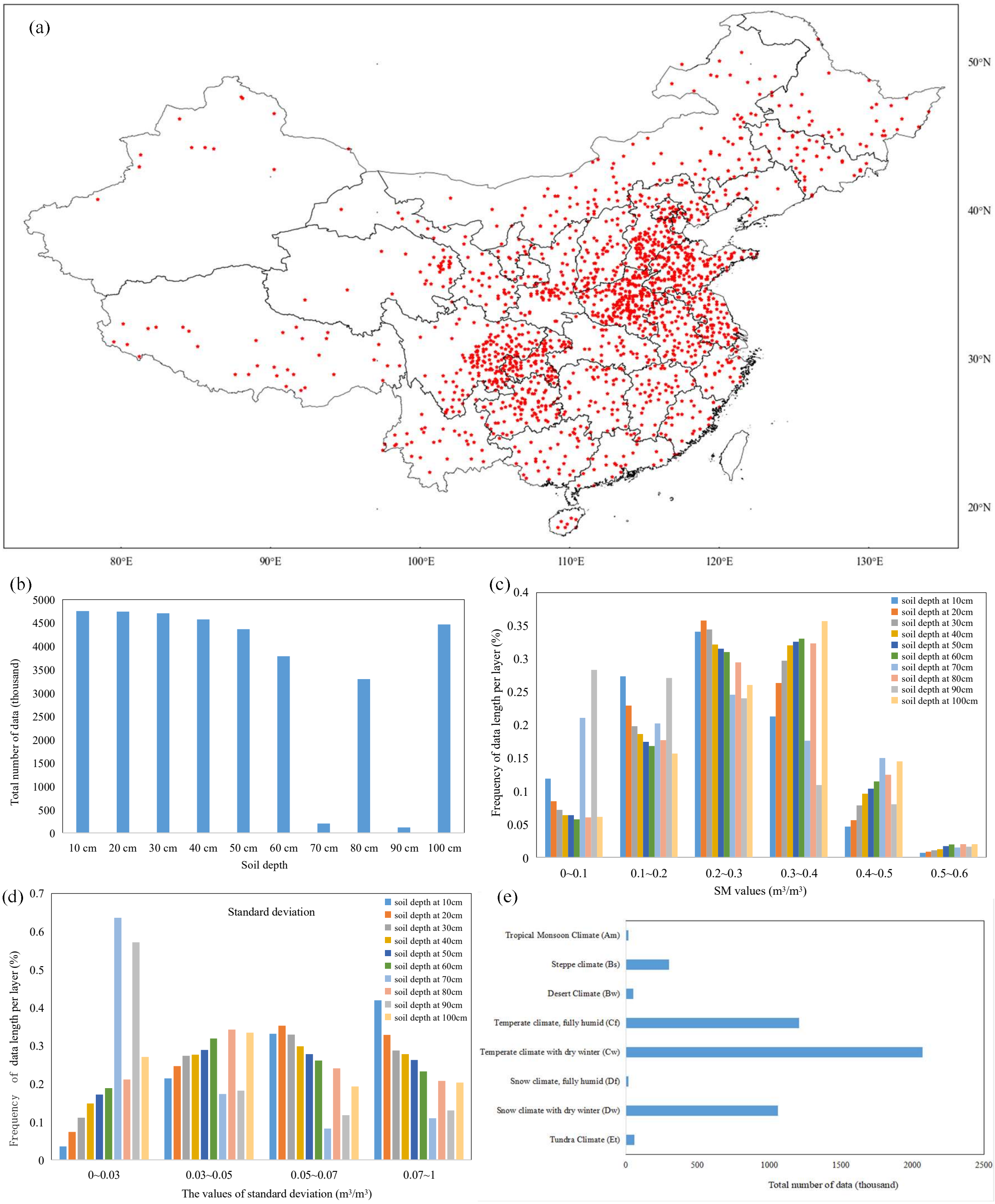


Figure S1: (a) The locations of all stations in China; (b) Total data number per soil depth; (c) Frequency of data length per layer for SM values; (d) Frequency of data length per layer for standard deviation; (e) Total data number per climate zone.

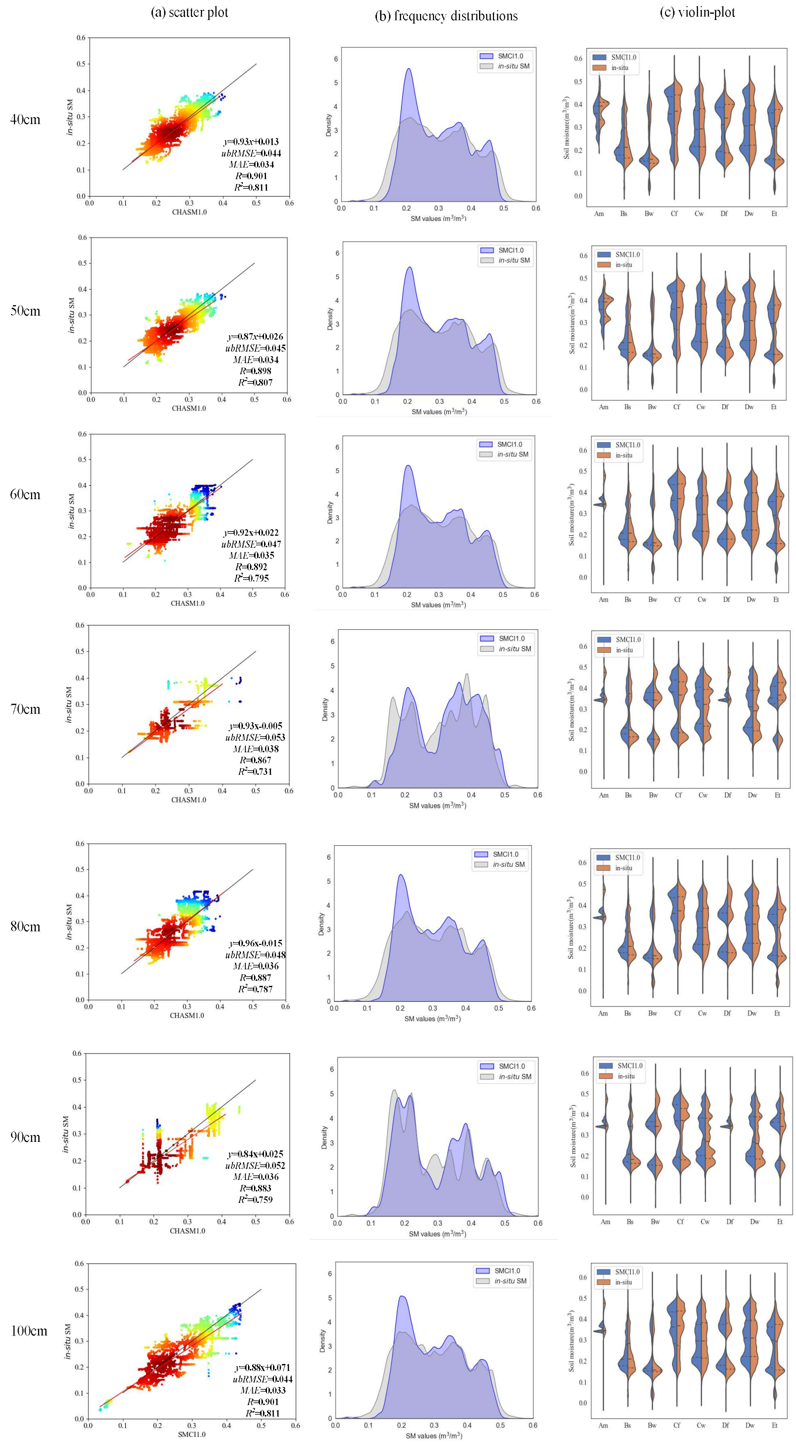


Figure S2. Comparisons between SMCI1.0 and in-situ SM from 10 to 30 cm soil depth: comparison of (a) the scatter plot between the mean of CHASM1.0 and that of in-situ SM at each station, (b) the frequency distributions of the whole SM values in SMCI1.0 and that in in-situ measurement networks, (c) the violin-plot for the distribution of daily SM from stations for each climate type.

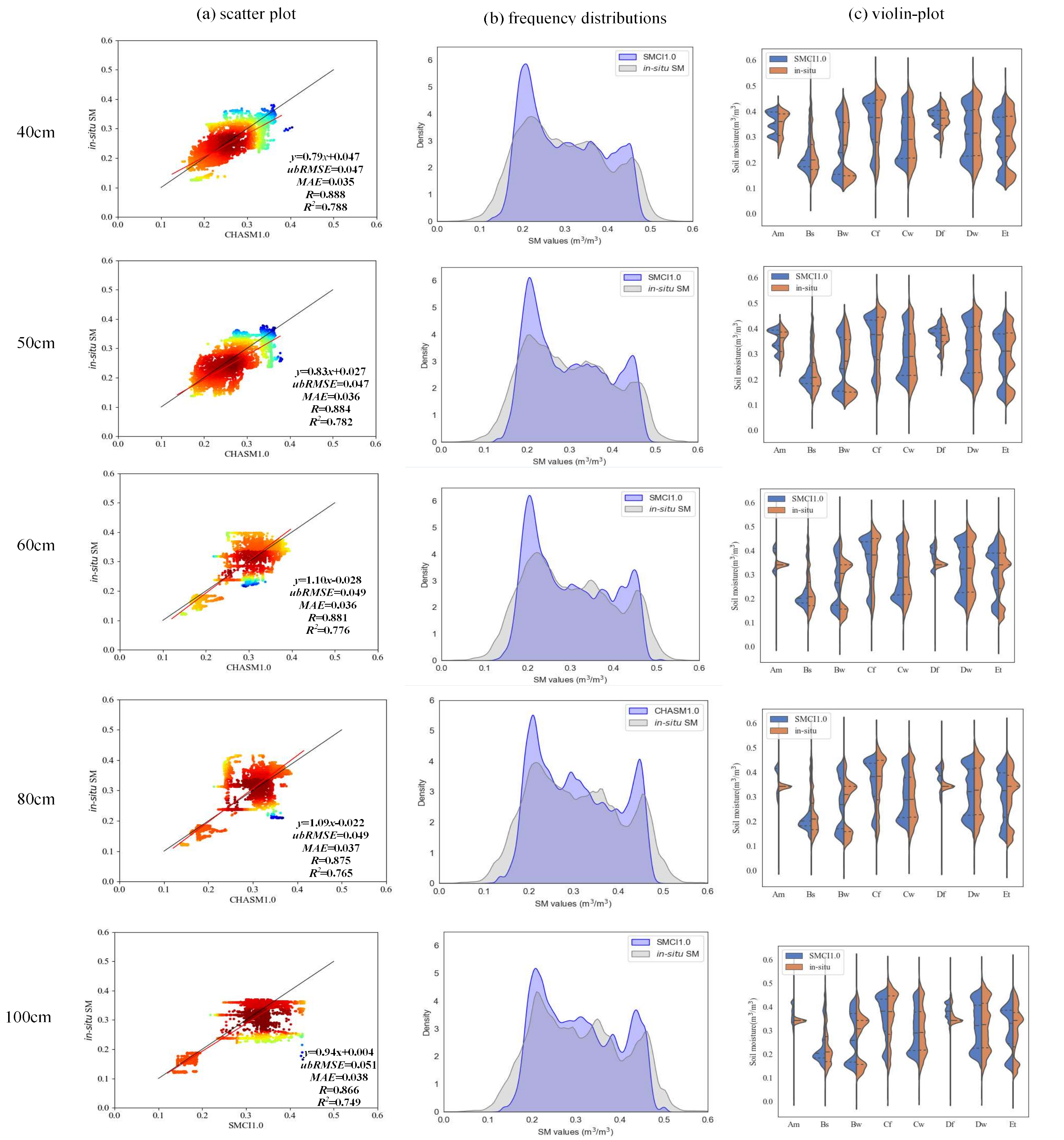
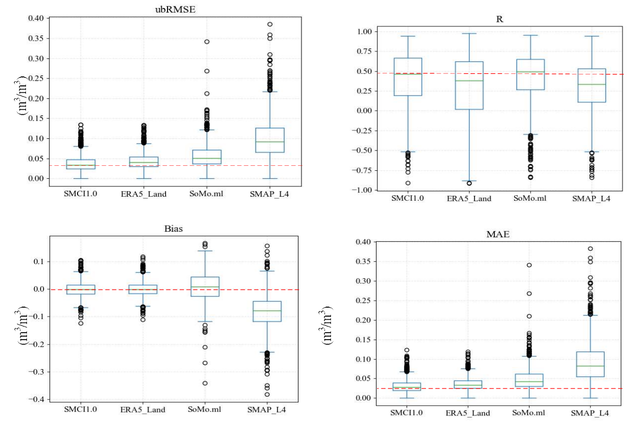
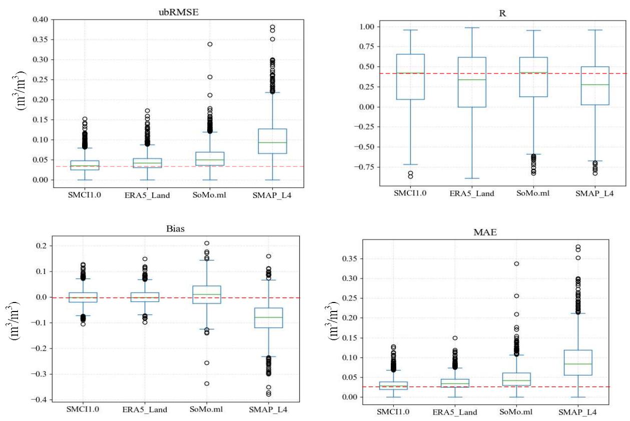


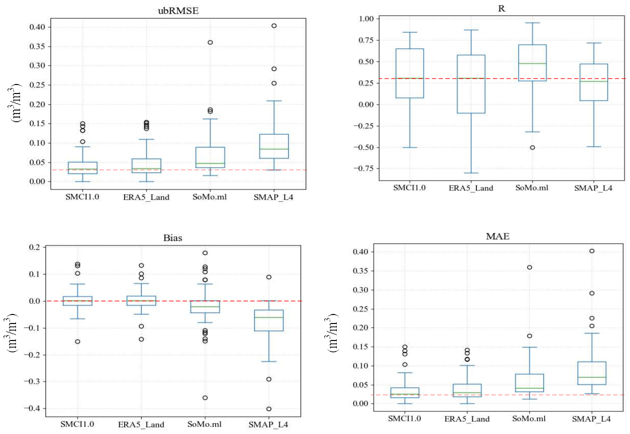
Figure S3. Same as Fig. S1 but for station-to-station mappings. There is no test set at 70 and 90 cm due to the few in-situ observations.



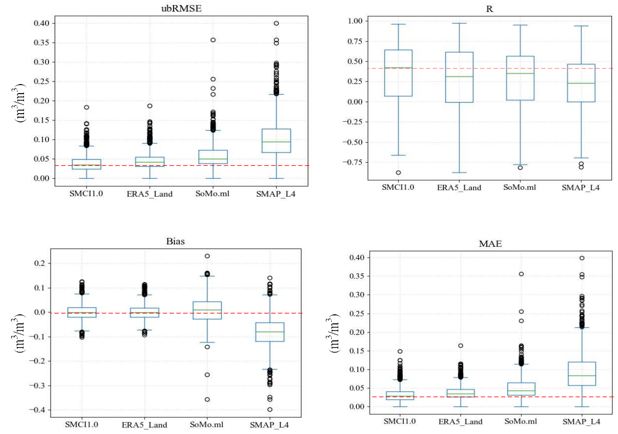
(a) 50 cm soil depth



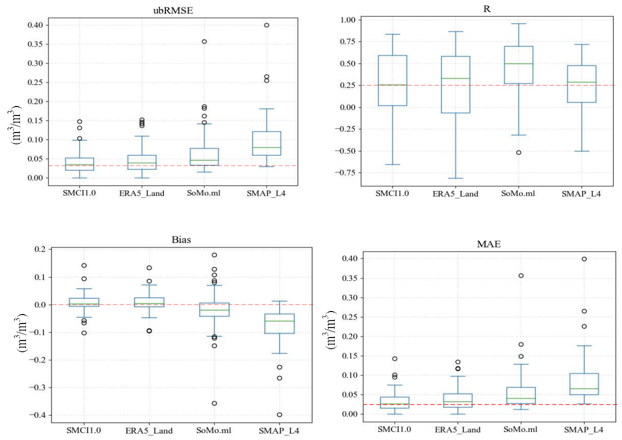
(b) 60 cm soil depth



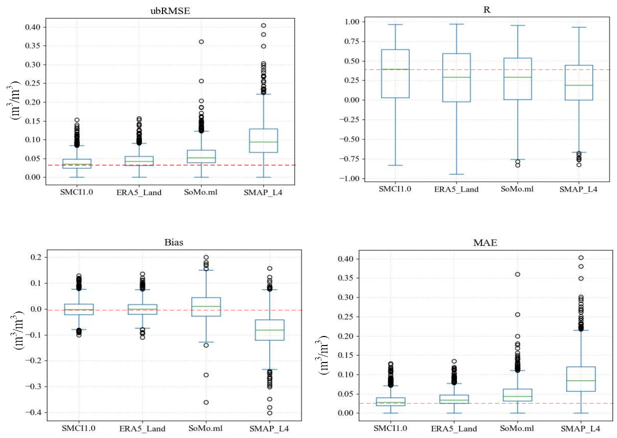
(c) 70 cm soil depth



(d) 80 cm soil depth



(e) 90 cm soil depth



(f) 100 cm soil depth

Figure S4. Comparison with other gridded datasets (a) at 50 cm soil depth, (b) at 60 cm soil depth, (c) at 70 cm soil depth, (d) at 80 cm soil depth, (e) at 90 cm soil depth, (f) at 100 cm soil depth.

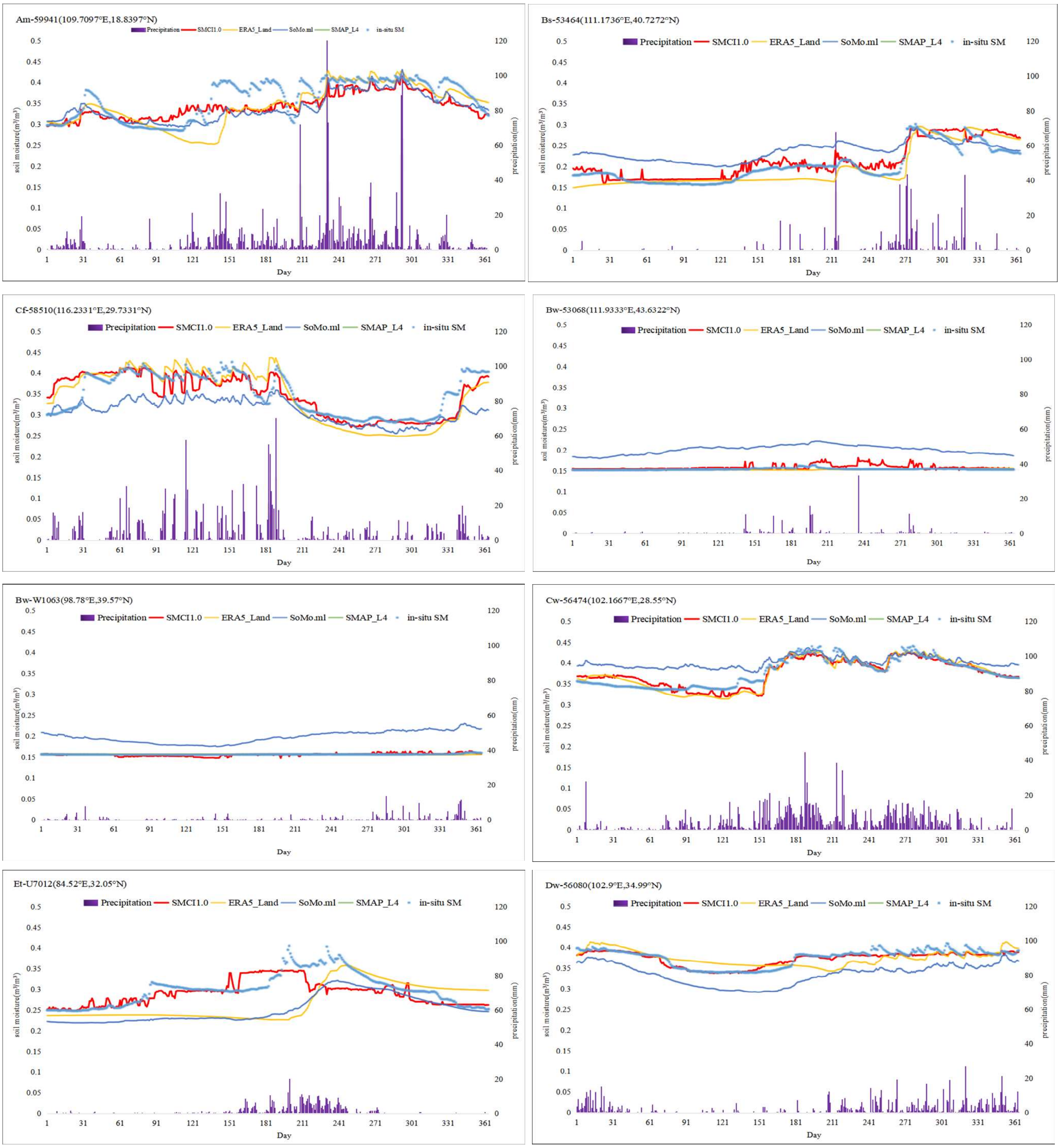


Figure S5. Time series of 30 cm soil depth in different climate regions. Each plot contains in-situ and estimated SM along with daily precipitation.

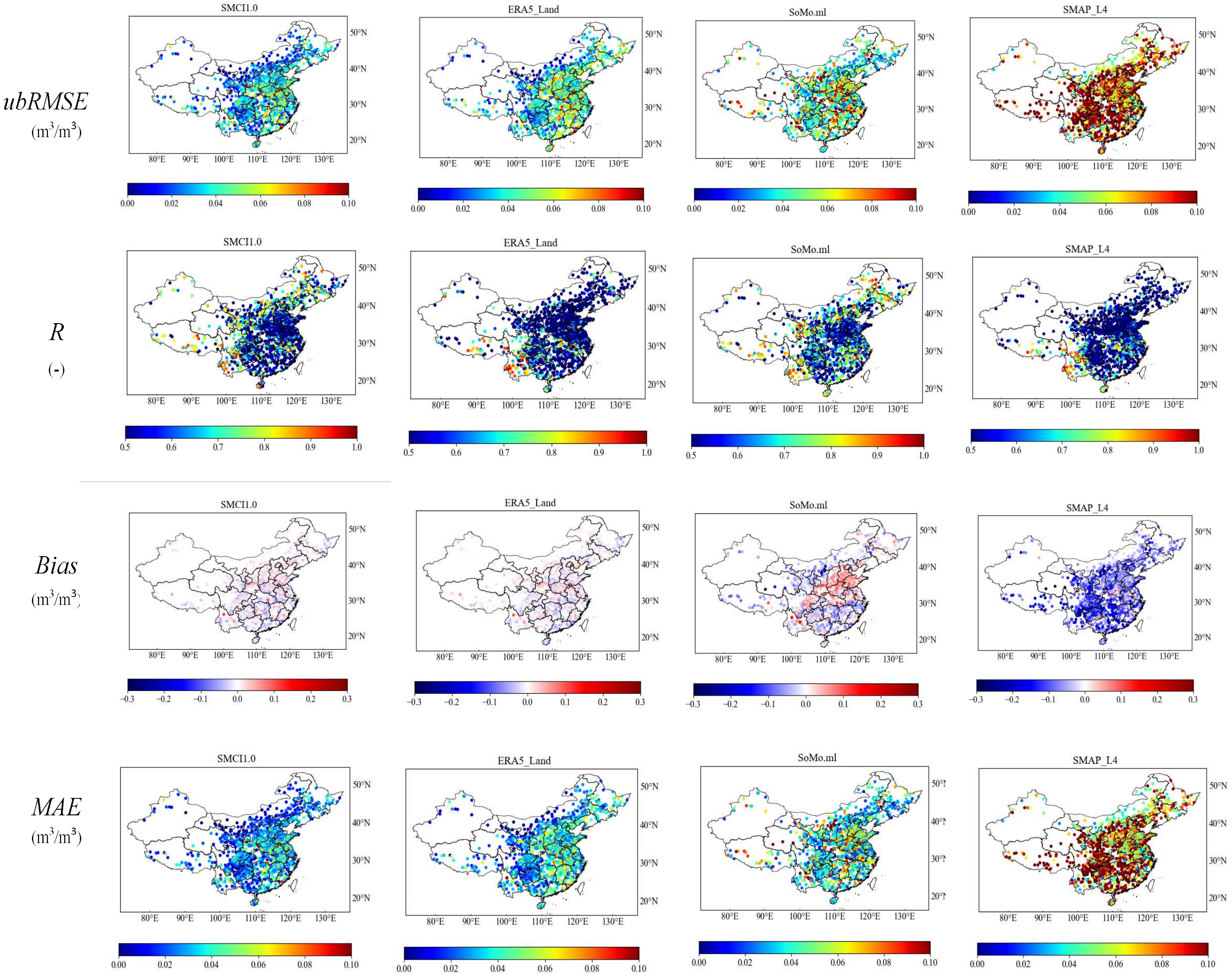
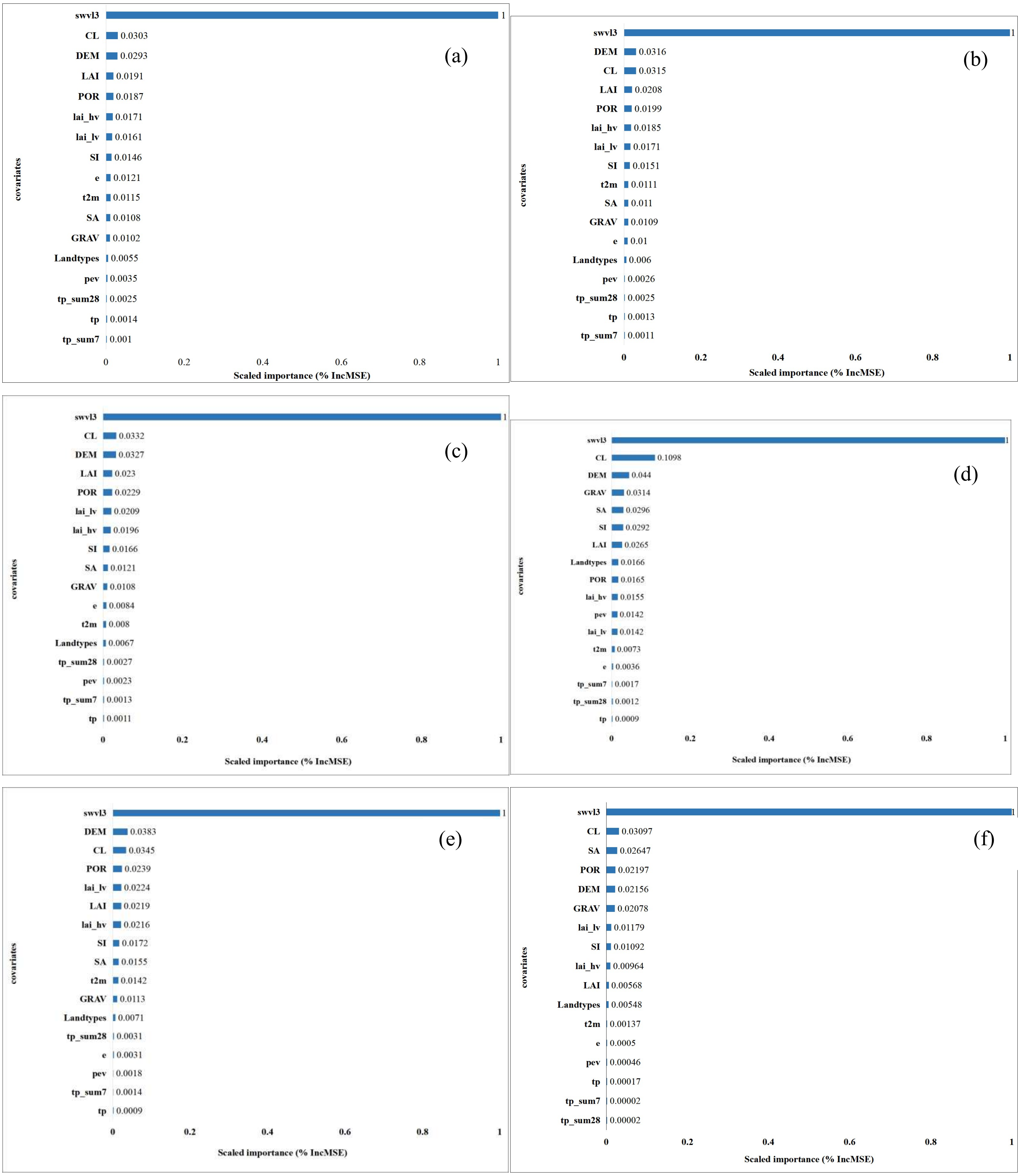
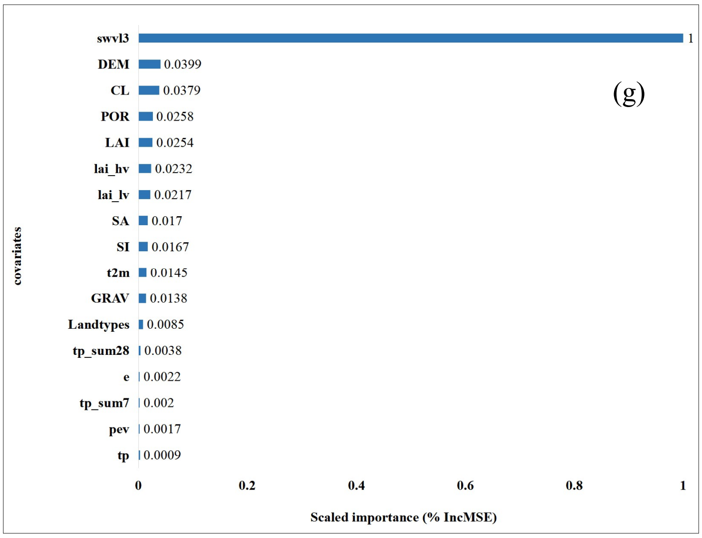


Figure S6. Goodness of fit statistics (ubRMSE, R, Bias, and MAE) at 30 cm soil depth during the tested period.


(a)
covariates
swvl3 1
CL 0.0303
DEM 0.0293
LAI 0.0191
POR 0.0187
lai_hv 0.0171
lai_lv 0.0161
SI 0.0146
e 0.0121
t2m 0.0115
SA 0.0108
GRAV 0.0102
Landtypes 0.0055
pev 0.0035
tp_sum28 0.0025
tp 0.0014
tp_sum7 0.001
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)
(b)
covariates
swvl3 1
DEM 0.0316
CL 0.0315
LAI 0.0208
POR 0.0199
lai_hv 0.0185
lai_lv 0.0171
SI 0.0151
t2m 0.0111
SA 0.011
GRAV 0.0109
e 0.01
Landtypes 0.006
pev 0.0026
tp_sum28 0.0025
tp 0.0013
tp_sum7 0.0011
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)
(c)
covariates
swvl3 1
CL 0.0332
DEM 0.0327
LAI 0.023
POR 0.0229
lai_lv 0.0209
lai_hv 0.0196
SI 0.0166
SA 0.0121
GRAV 0.0108
e 0.0084
t2m 0.008
Landtypes 0.0067
tp_sum28 0.0027
pev 0.0023
tp_sum7 0.0013
tp 0.0011
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)
(d)
covariates
swvl3 1
CL 0.1098
DEM 0.044
GRAV 0.0314
SA 0.0296
SI 0.0292
LAI 0.0265
Landtypes 0.0166
POR 0.0165
lai_hv 0.0155
pev 0.0142
lai_lv 0.0142
t2m 0.0073
e 0.0036
tp_sum7 0.0017
tp_sum28 0.0012
tp 0.0009
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)
(e)
covariates
swvl3 1
DEM 0.0383
CL 0.0345
POR 0.0239
lai_lv 0.0224
LAI 0.0219
lai_hv 0.0216
SI 0.0172
SA 0.0155
t2m 0.0142
GRAV 0.0113
Landtypes 0.0071
tp_sum28 0.0031
e 0.0031
pev 0.0018
tp_sum7 0.0014
tp 0.0009
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)
(f)
covariates
swvl3 1
CL 0.03097
SA 0.02647
POR 0.02197
DEM 0.02156
GRAV 0.02078
lai_lv 0.01179
SI 0.01092
lai_hv 0.00964
LAI 0.00568
Landtypes 0.00548
t2m 0.00137
e 0.0005
pev 0.00046
tp 0.00017
tp_sum7 0.00002
tp_sum28 0.00002
0 0.2 0.4 0.6 0.8 1
Scaled importance (% IncMSE)



60 **Figure S7. Relative importance of covariates for the RF model from 40 to 100 cm soil depth. (a) at 40 cm soil depth, (b) at 50 cm soil depth, (c) at 60 cm soil depth, (d) at 70 cm soil depth, (e) at 80 cm soil depth, (f) at 90 cm soil depth, (g) at 100 cm soil depth.**

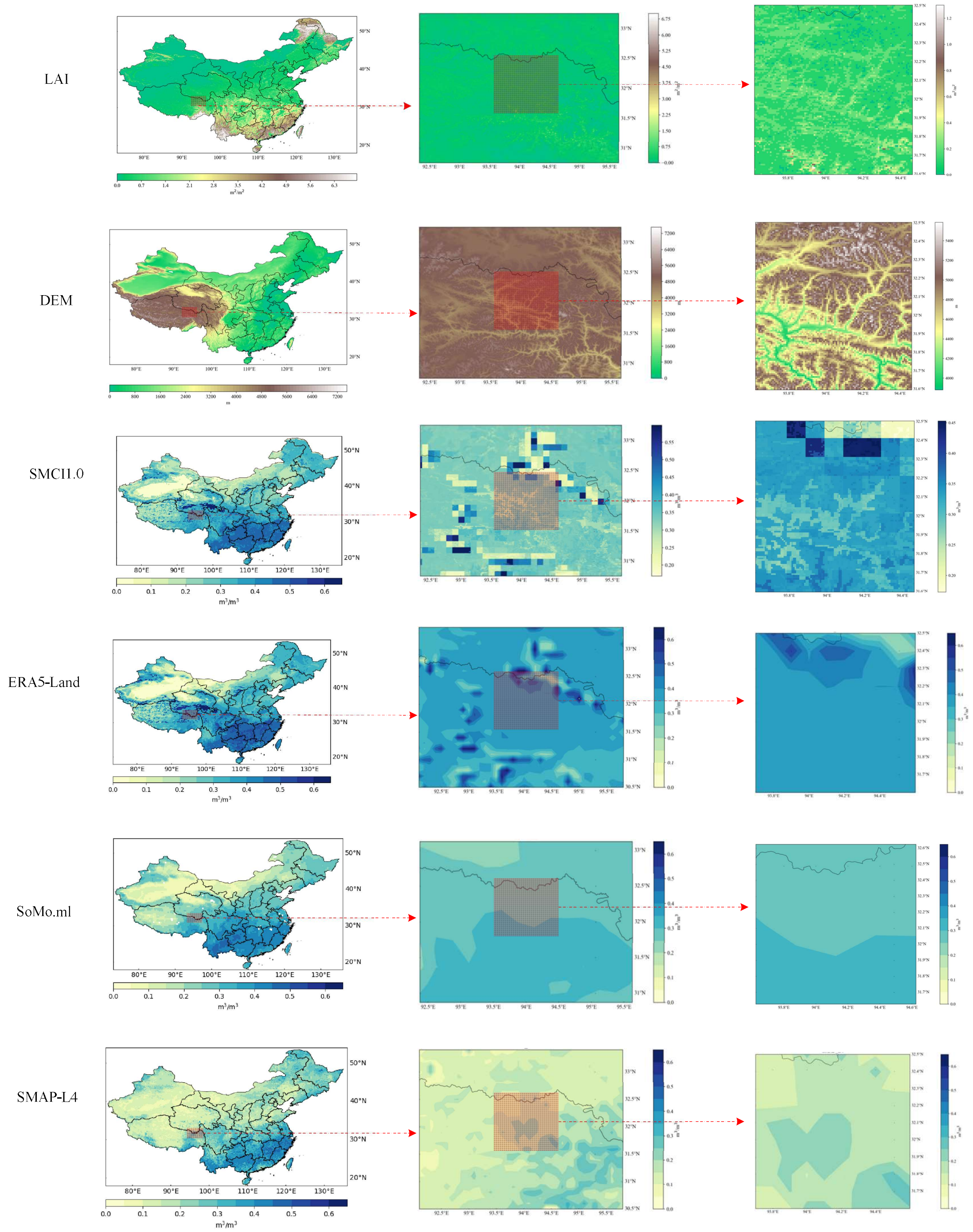


Figure S8. Soil moisture maps from different products on 1st January 2016 over Qinghai-Tibet Plateau region. The resolution is 1km for SMC1.0, 9km for ERA5-land and SMAP-L4 and 0.25 degree for SoMo.ml.

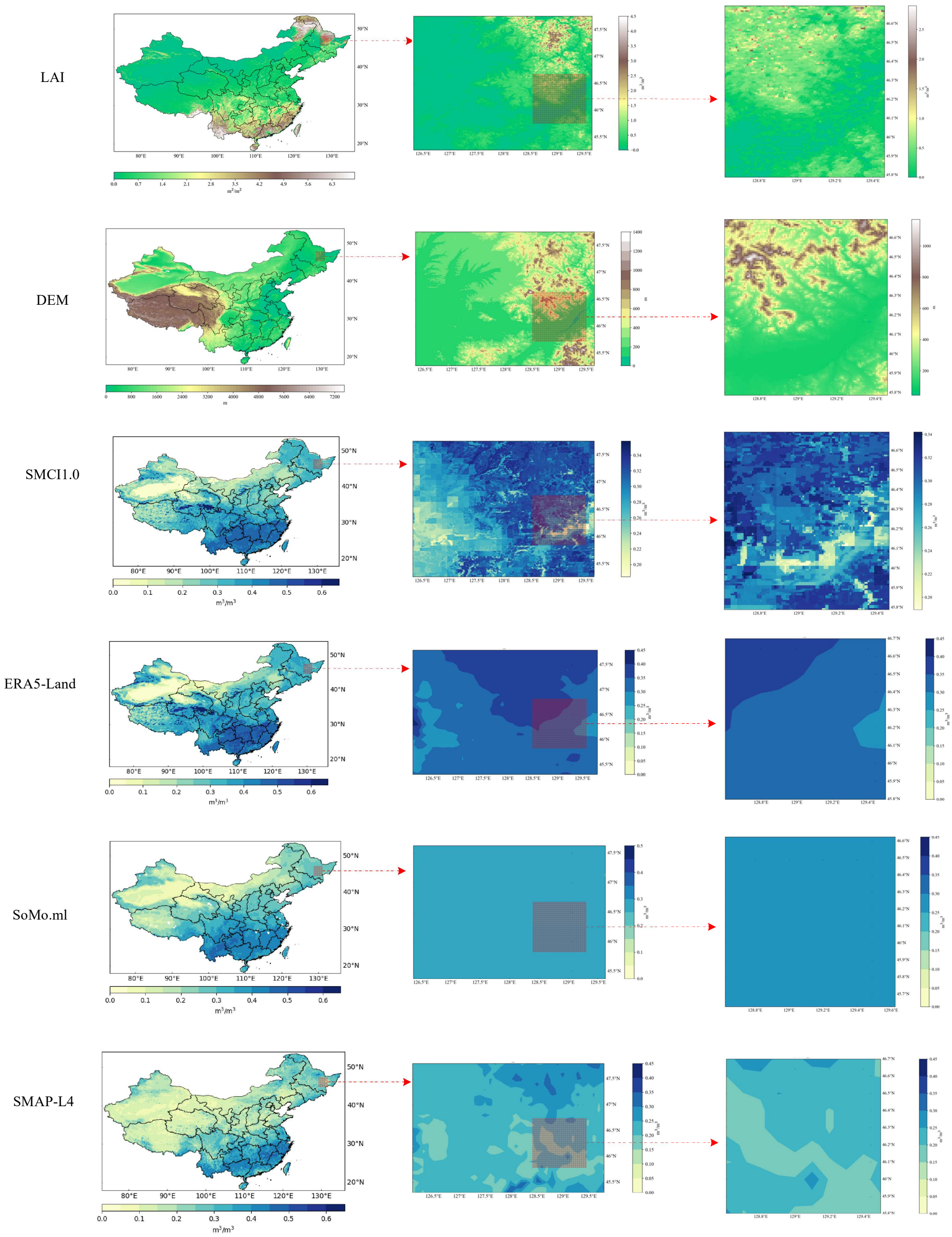


Figure S9. Soil moisture maps from different products on 1st January 2016 over Northeast region. The resolution is 1km for SMC1.0, 9km for ERA5-land and SMAP-L4 and 0.25 degree for SoMo.ml.

Table S1. The accuracy of the RF models with all hyper-parameters at 10 cm soil depth based on grid hyper-parameters method (the best hyper-parameter is shown in bold font).

min_samples_leaf / *max_features*	5	10	15	20	25
1	0.0608	0.0603	0.0602	**0.0601**	0.0602
2	0.0621	0.0614	0.0611	0.0610	0.0608
3	0.0625	0.0619	0.0616	0.0614	0.0613
5	0.0626	0.0622	0.0618	0.0616	0.0615
6	0.0629	0.0624	0.0621	0.0619	0.0617
7	0.0629	0.0624	0.0621	0.0619	0.0618
8	0.0631	0.0626	0.0623	0.0621	0.0620
9	0.0631	0.0626	0.0623	0.0622	0.0620
10	0.0631	0.0626	0.0623	0.0622	0.0620
11	0.0631	0.0627	0.0624	0.0623	0.0621
12	0.0632	0.0627	0.0625	0.0622	0.0622
13	0.0632	0.0628	0.0625	0.0623	0.0622
14	0.0633	0.0628	0.0625	0.0624	0.0622
15	0.0634	0.0628	0.0626	0.0624	0.0622
16	0.0634	0.0629	0.0626	0.0624	0.0623
17	0.0634	0.0629	0.0627	0.0625	0.0624
18	0.0633	0.0629	0.0627	0.0625	0.0624
19	0.0634	0.0629	0.0627	0.0626	0.0625
20	0.0635	0.0630	0.0627	0.0626	0.0625
21	0.0635	0.0630	0.0628	0.0626	0.0625
22	0.0635	0.0631	0.0628	0.0626	0.0626
23	0.0636	0.0631	0.0629	0.0627	0.0625
24	0.0636	0.0632	0.0629	0.0627	0.0626
25	0.0637	0.0632	0.0630	0.0628	0.0627